

Overcoming current and emerging challenges of the open data ecosystem for organisations

Link Digital launches its first series of papers exploring the aspects of the open data and digital ecosystem.

Phone: +61 2 6111 2907

Post: GPO Box 199, Canberra, ACT 2601, Australia

Email: info@linkdigital.com.au

Web: <https://linkdigital.com.au/>

Andrew Nette

Technical Writer/ Policy Analyst

Link Digital

December 2023

About the document

Today, Link Digital launched its first in a series of white papers, delving into the intricacies of the open data and digital ecosystem.

Titled "Overcoming current and emerging challenges of the open data ecosystem for organisations," serves two purposes. Firstly, it provides a comprehensive understanding of the present state of open data and data sharing. Secondly, it introduces Link Digital's expertise and growing influence within this data ecosystem. We offer trusted guidance and expertise for organisations seeking to enter this space or enhance their existing activities within it.

In recent years, much has been written about the dot-com boom's decline in 2000 and the rise of industry giants such as Google, eBay, and Amazon. However, one aspect that often goes unnoticed is how organisations have restructured their engagement with data and the digital landscape. Over the past decade, we have seen a convergence of previously independent IT, marketing, and communication functions into digital teams that possess a more sophisticated and integrated approach to data utilization. Simultaneously, governments have undergone a significant transformation in data collection, use for service delivery, and policy formulation. The global pandemic has further accelerated this shift, with citizens increasingly accessing government information and services digitally.

Our white paper examines these changes and explores the growing recognition of data as more than just an economic driver. We believe data has the power to be open and shared, thereby improving government policy-making, enhancing services, and supporting a well-informed citizenry and civil society. Alongside highlighting the benefits of open data and data sharing, we dive into the challenges hindering the growth and realization of the open data movement's full potential. We discuss frameworks such as the Third wave of Open Data, as well as the efforts to standardize data sharing through the adoption of FAIR principles. These principles ensure that public data is easily accessible and promotes interoperability between data sharing portals. Additionally, we explore the global significance of the Comprehensive Knowledge Archive Network (CKAN), an open-source software that, when deployed and hosted on a web server, becomes a powerful open data platform that Link Digital proudly co-stewards.

Our white paper not only showcases Link Digital's trajectory as a leading digital service organisation operating at the intersection of open data and open-source technology but also envisions our evolving role within the data ecosystem. We are committed to leveraging open data in various domains, including corporate environmental, social, and corporate governance. We strive to assist in shaping government policies that address the needs of our rapidly evolving society, combat disinformation, build trust in government and public institutions, and develop artificial intelligence models that are trustworthy, safer, diverse, and socially beneficial.

We encourage you to join us in this journey of collaboration and co-creation as we navigate the challenges and opportunities of the open data ecosystem. Together, we can create a better future powered by open and shared knowledge.

¹ Open Knowledge Foundation, "We are open advocates," accessed July 24, 2023, <https://okfn.org/who-we-are/>

I. The importance of knowing your data better

It is now a widely accepted proposition that data is the main currency of the 21st century. This trend is being driven by the increasing digitisation of nearly all key facets of life and the ability of technology to help capture data. Most of us now, sometimes almost on an intuitive level, have become used to the existence of data all around us and to deploying it in our everyday lives. And governments, corporations and the public now accept the proposition that data is worth something, either financially or because it brings the promise of greater knowledge. But do most people really understand what data is and, more importantly, how to refine or unlock its value, whether it be for greater government or corporate transparency, a better-informed citizenry or improved public policy?

The economic impact arising from the growth of data was first articulated by [British mathematician and entrepreneur Clive Humby](#) in a 2006 talk to the [American Association of National Advertisers](#), titled 'Data is the new oil'. What Humby meant by this is that raw data, just like crude oil, has very little value in and of itself unless it is processed. Humby's idea emerged in the aftermath of the dot-com boom which ended in 2000. The boom saw a significant investment of venture capital into various technology start-ups and an expansion of Internet use. The subsequent bust resulted in a massive decline in the value of tech stocks. But as the tech sector gradually stabilised, two important developments occurred. The first was the emergence of companies such as [Google](#), [eBay](#) and [Amazon](#). A second, less commented on shift was a gradual change in the way many organisations structured their interactions with the data and digital space. In the very early 2000s, this relationship was often configured around having separate IT, marketing and communications departments, each with different types of managers, work tasks and reporting functions. Over the last two decades, however, many of the functions of these previously separate units, particularly IT and marketing, have converged into a digital team which has a different and more sophisticated relationship with data and its use. Parallel to this has been an evolution in how governments in many countries view their data and digital operations, and a corresponding shift to citizens accessing more government information and services digitally. Most service organisations operating in this space had previously existed to provide either IT support or marketing and communications support. But in line with these trends, these organisations shifted their focus to provide a broader suite of digital services.

[Link Digital](#) is one such agency operating in this space. First established as Link Web Services in 2001, it initially provided a mix of technical and design assistance related to digital products such as multimedia websites and CD-ROMs. But from its earliest origins, the company aimed to be a pure digital agency, providing both the technical and creative capabilities that are now commonly found in similar organisations world-wide. As the concept of the digital agency matured, in 2004 Link Web Services updated its trading name to Link Digital and moved to take advantage of various opportunities that were then emerging in the data and digital space

[Amazon Web Services](#) established a presence in Australia in 2012 and Link Digital was one of its first local [Advanced Tier Consulting Partners](#). The same year, Link Digital oversaw the design of the new website for Australia's then Department of Human Services (now known as [Services Australia](#)), and the design, hosting and supporting of one of the Australian national government's early open data portals, [The Australian Government Open Data Portal](#), established in 2013. Link Digital has gone on to work on state-based government data portals in New South Wales ([Data NSW](#)), South Australia ([Data SA](#)), Western Australia ([Data WA](#)), and Victoria ([Data Vic](#)). The organisation has also designed data portals for individual government agencies, and a large range of non-profits.

These portals all use the [Comprehensive Knowledge Archive Network \(CKAN\)](#), a piece of open-source software launched in 2006 that can be configured and set up to function as an open data platform once it is deployed and hosted on a web server. In addition to a growing installation base, CKAN boasts a community of users and contributors, which gives it wide potential for adoption and significant capabilities, something that will be discussed in further detail later. Link Digital is what is referred to as a 'co-steward' of CKAN. This is not to be confused with another term associated with the open data ecosystem, that of a data steward. As a co-steward of CKAN, Link Digital serves to support the community of users and contributors that collaborate with the project. 'Data stewards may be either individuals or groups of individuals within an organisation who are dedicated employees that initiate and contribute to sustainable data collaboration,' which may or may not involve utilising CKAN. Link Digital was also an early participant in the community around [Drupal](#), an open source content management software platform, and in 2012 Link Digital Executive Director Steven De Costa ran for the board of the [Drupal Association](#). Drupal became more generally accepted within government during the presidency of Barack Obama, when the [Whitehouse website](#) used Drupal and released its code, a development that played a major role in mainstreaming the transformation of government as a user of open source technology. This was part of a wider plan that was first signalled in a [January 2009 memorandum](#) to heads of United States federal executive departments and agencies to prepare a strategy for more open government. Specifically, it committed the Obama administration to 'creating an unprecedented level of openness in Government. We will work together to ensure the public trust and establish a system of transparency, public participation, and collaboration. Openness will strengthen our democracy and promote efficiency and effectiveness in Government.'²

¹ The Third Wave of Open Data, "Primer #1: Creating and Empowering Chief Data Stewards."

² "Transparency and Open Government: Memorandum for the heads of executive departments and agencies," The White House, January 21, 2009, <https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>

Link Digital is now firmly established in the evolving data ecosystem, with a focus on supporting open data, digital service design and citizen centred design practices that have come to the fore ever since the late 2010s. Link Digital currently has offices in Australia, Vietnam, the Philippines, Ukraine, and recently incorporated a subsidiary entity in Canada – Datashades LTD trading as Link Digital. We aim to deepen our current trajectory as a digital agency working at the intersection of open data and open source technology by bringing enterprise data ecosystems to government, non-profits, academic and corporate clients, with one particular focus of future expansion being the role of open data in corporate environmental, social, and corporate governance (ESG).

In addition to being a major and trusted participant in the open source software community, especially in relation to CKAN, Link Digital is active in a range of related activities.

We have been a long term supporter of [GovHack](#), an annual open data competition held in Australia and New Zealand, first run in 2009 and which now draws participants from around the world. We have also been involved in consultations around the Australian government's proposed [Data and Digital Government Strategy](#): the data and digital visions for a world leading APS to 2030, a vision to deliver simple, secure, and connected public services for all people and business through world class data and digital capabilities.

Internationally, Link Digital is a participant in the [Open Government Partnership](#) (OGP) and the [Open Knowledge Foundation](#) (OKF). Established in 2004, the OKF is dedicated to 'advancing open knowledge as a design principle beyond just data.'³ The OGP is an international grouping of national and local governments and civil society organisations aimed at promoting transparent and open government. De Costa participated in the most recent [OGP Summit in Tallinn, Estonia in September 2023](#), on the theme of open government in the digital age and the potential for policy making to be more open, transparent and democratic. And as part of our initial orientation of the Canadian open data market, Link Digital has developed strong relations with an array of groups, including the [Canadian Open Data Society](#), a non-profit that hosts the annual [Canadian Open Data Summit](#), and [Open North](#), a non-profit working on data governance and digital strategy with governments, civic-focused organisations, and business.

Link Digital's mature understanding of the current state of play in relation to open data is very much in line with global trends, particularly what is referred to as the '[Third Wave of Open Data](#)'.⁴ Central to the Third Wave is the proposition that the open data ecosystem has moved beyond a focus on simplifying data and how it impacts our society and economy; to how data can better be systematised and how governments, private corporations and civil society organisations can better know and understand the data they have and share it to provide improved products and services, and a more informed citizenry.

The following paper seeks to provide a clearer understanding of the current state of play and future trends relating to open data and data sharing, particularly the importance of the Third Wave. This is as a way of introducing Link Digital's expertise and growing position in this data ecosystem, and its value proposition in terms of providing expertise and trusted guidance for organisations seeking to enter this space. Link Digital has consciously positioned itself as a supplier of data and digital services with an ongoing and proven concern for supporting civil society through strengthening public institutions. Link Digital is a one stop digital agency that can provide expertise and guidance for organisations seeking to engage with the data ecosystem, or those who are already in it but seeking to refine or build on their position. Link Digital's services include open source/data cataloguing and web development, as well as the skills to assist organisations in the ever evolving area of data governance.

Data can be refined and in doing so it can contribute new knowledge or shift our understanding of prior information by providing new insights and contributing to and/or refining existing systems of knowledge. But there is so much data now that many organisations seeking to utilise it are often left with little better than a veritable data swamp, from which it is impossible to derive any real insights or benefit. As a trusted enabler of open data initiatives Link Digital can support organisations to deploy the policy and broader public good gains from the data they have and create meaningful change.

³ Open Knowledge Foundation, "We are open advocates," accessed July 24, 2023, <https://okfn.org/who-we-are/>

⁴ For a more thorough discussion of the Third Wave of Open Data, see Open Data Policy Lab, [The Third Wave of Open Data Toolkit: Operational Guidance on Capturing the Institutional and Social Value of Data Re-Use](#) (New York: Tandon School of Engineering, New York University, 2023).

II. Open data: the current state of play

Defining some frequently used terms is a good way to start any discussion of the state of play as regards the evolving open data ecosystem. 'Data' refers to 'observations that have been converted into a digital form that can be stored, transmitted or processed and from which knowledge can be drawn.'⁵

Discussing the extraordinary growth of data, Australia's [Productivity Commission](#) notes the way in which data generation and usability has enabled a range of new products and services, which in turn has driven more data creation.⁶

'Open data' is defined by the OKF as meaning that 'anyone can freely access, use, modify, and share [the data] for any purpose (subject, at most, to requirements that preserve provenance and openness).'⁷ This definition can be extended to commercial and research organisations that want to better understand their data resources.

Data sharing is not necessarily the same thing as open data and usually involves the sharing of data between agencies that is not public.⁸ It requires an understanding of who the data is being shared between and the conditions under which it is being shared. For example, only a certain portion of a data set may be shared with certain staff or departments in an organisation. The sharing may be done based on stipulations regarding how the data is to be used, such as a negotiated timetable, and the data access might be revocable under certain circumstances.

These conditions are often set out in an increasingly important piece of governance known as a data contract. According to the data democracy company [Atlas](#), based in New Delhi India, a data contract or agreement as it is sometimes referred to, 'outlines how data can get exchanged between two parties. It defines the structure, format, and rules of exchange in a distributed data architecture. These formal agreements make sure that there aren't any uncertainties or undocumented assumptions about data.'⁹ In addition to general agreements about intended use, ownership and provenance of the data concerned, data contracts include agreements about the data's schema, associated metadata, semantics, and service level agreements.

The [European Commission](#) (EC) projections suggest there will be a 530 per cent increase in the volume of global data over the period 2018 to 2025.¹⁰ [Statista](#) notes that the amount of data created, captured, copied, and consumed globally reached 64.2 zettabytes in 2020, and was projected to grow to more than 180 zettabytes in the years leading to 2025 (A zettabyte is equivalent to a trillion gigabytes).¹¹ As the [Open Data Policy Lab](#) puts it:

Data—which can be generated from things like smartphones, scientific studies, and financial transactions—has proliferated and is now an instrumental part of our modern world. Much of the increase in data is the result of datafication, the surge in data collection and storage that has resulted from exponential increases in networked computing and the rise of mobile phones and other digital devices with embedded processors.¹²

While governments have always collected and used data, they are now doing so more than ever. According to [David Gruen](#), an statistician and Agency Head of the [Australian Public Service Data Profession](#), the rapid growth in the availability and range of data sources over the last two decades due to the digital revolution is changing the way governments formulate policy, a development which, in turn, necessitates a level of data literacy on the part of public service staff that was not required a few years ago.¹³

These trends were accelerated by the COVID-19 pandemic. The pandemic boosted the pace of digitisation and increased data creation as more people worked, learned and shopped from home and used home entertainment options more often. In many countries it also played a major role in increasing data sharing between and within the different levels of government and the digitisation of government services. An historically unique occurrence, the pandemic necessitated a response that went outside the regular functions and information and administration silos of government, the research community, not for profit organisations and the corporate sector. It demanded both a major shift in the use of data and an unprecedented level of data sharing to facilitate effective responses, for example, contact tracing.

From the [European Union](#) [EU] setting up a regionwide interoperability gateway to enable a secure information exchange between different national contact-tracing apps, to the [US National Institutes of Health](#) establishing a centralized repository of COVID-19 health records for facilitating research and discovery, the rapid sharing of data proved instrumental to the public sector pandemic response.¹⁴

Of particular significance in the context of COVID-19 was the level of data sharing that took place between scientists to develop workable vaccines. This led to what were by scientific standards, major breakthroughs relatively quickly, with the first vaccines available in the United States by December 2020, only a year into the pandemic. As a joint report by [KPMG](#) and [Microsoft](#) puts it:

In 2020, the power of open data was highlighted by the efforts of [Johns Hopkins University](#), which brought together diverse datasets to visualize the global impact of COVID-19. Beyond retrospective reporting, there are now growing calls to make critical healthcare data more accessible, so agencies and governments can pre-empt future outbreaks and inform policy decisions that mitigate their impact.¹⁵

⁵ "Measuring investment in data, databases and data science: Conceptual framework," Statistics Canada, June 24, 2019, <https://www150.statcan.gc.ca/n1/pub/13-605-x/2019001/article/00008-eng.htm>.

⁶ "2017 Inquiry into Data Availability and Use," Productivity Commission, accessed June 15, 2023, <https://www.pc.gov.au/inquiries/completed/data-access/report>

⁷ "The Open Definition," Open Knowledge Foundation, accessed July 5, 2023, <https://opendefinition.org/>.

⁸ KPMG and Microsoft, Unleash the full potential of open data in the Public Sector: How the right data foundations will set you on the path to AI, 2021, <https://kpmg.com/ae/en/home/insights/2021/04/unleash-the-full-potential-of-open-data-in-the-public-sector.html>.

⁹ "Data Contracts: The Key to Scaling Distributed Data Architecture and Reducing Data Chaos," Atlas, last updated April 20, 2023, <https://atlas.com/data-contracts/>.

¹⁰ "European Data Strategy," European Commission, accessed June 15, 2023 https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en.

¹¹ "Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025," Statista, accessed October 3, 2023, <https://www.statista.com/statistics/871513/worldwide-data-created/>.

¹² Open Data Policy Lab, The Third Wave of Open Data Toolkit: Operational Guidance on Capturing the Institutional and Social Value of Data Re-Use (New York: Tandon School of Engineering, New York University, 2023), 6.

¹³ David Gruen, "Strengthening APS Data Capability," Vimeo, Institute of Public Administration Australia, Australian Government Data Forum: Pathways, Perspective and Practice, May 17, 2023, speech, <https://vimeo.com/829313165/ea2042fa97?share=>

¹⁴ Jueren Klenk, Tasha Austin, Costi Perricos, and Muthukumarana, "Fluid data dynamics: Generating greater public value from data," Deloitte Insights, March 4, 2021, <https://www2.deloitte.com/us/en/insights/industry/public-sector/government-trends/2021/fluid-government-data-dynamics.html>.

¹⁵ KPMG and Microsoft.

Although the private sector remains less convinced of its advantages, data sharing within and between governments and between governments and the public is now viewed as an undisputed positive in the context of improving government services and supporting civil society and a more informed citizenry. According to the [Open Data Charter](#), a collaboration between 170 governments and organisations working to make data more freely available:

Open data allows users to compare, combine, and follow the connections among different datasets, tracing data across a number of programs and sectors. When data can be effectively combined and compared, it can help highlight trends, identify social and economic challenges and inequities, and benchmark progress in public programs and services.¹⁶

The Open Data Charter goes on to state that data ‘enables governments, citizens, and civil society and corporates to make better informed decisions. Effective and timely access to data helps individuals and organisations develop new insights and innovative ideas that can generate social and economic benefits.’¹⁷ As Australia’s Productivity Commission puts it, ‘The substantive argument for making data more available is that opportunities to use it are largely unknown until the data sources themselves are better known, and until data users have been able to undertake discovery of data.’¹⁸

Open data has been referred to as a new form of social contract between government and citizens, and as akin to a social movement with its own principles and political aims. These chiefly revolve around the assertion that the digital transformation currently taking place, and the possibilities for open data and data sharing presented by it, have ‘enormous potential to foster more transparent, accountable, efficient, responsive, and effective governments and civil society and private sector organisations, and to support the design, delivery, and assessment of sustainable development goals at a global scale.’¹⁹ The open data movement sees itself as a key component in rebuilding the public trust in government that has been steadily eroding since the 1970s.²⁰ [A 2023 survey based on field research in 28 countries](#) notes an increasingly polarised world, caused by growing lack of trust in government institutions, the disintegration of a shared media/news environment, and a mass class divide, with people in the top quartile of income living ‘in a different trust reality than those of the bottom quartile.’²¹ Connected to this, there is a rapidly emerging role for open data and the movement around it in developing and maintaining confidence in the data used by corporations as part of their ESG reporting. All these ambitions, which are rooted in the notion of an information commons – a land or resource that belongs to the entire community – begin from the premise that ‘data collected from the public, often using public funds or publicly funded infrastructure, should also belong to the public or, at the very least, be made broadly accessible to those pursuing public-interest goals.’²²

The main form of data sharing occurs through what has already been referred to as open data portals. These are front-line interfaces that bridge data providers and users, helping to ensure that data is not just accessible but usable.

Other forms of data sharing involve business to government, or what is sometimes referred to as ‘B2G’, which the EC defines as ‘a collaboration in which a company or other private organisation makes available its data (or insights) to the public sector (local, regional, national or EU) for a public interest purpose’²³; and internal asset registers (IAR). In their traditional form, an IAR is a detailed list of business assets, including physical and fixed assets, IT and digital assets, or assets relating to the running of a specific business, for example a university or a hotel. This is compiled with a view to giving the organisation concerned clear information about the status, procurement date, location, cost, depreciation, and current value of each of its assets. In line with the realisation about the need to use data more effectively, organisations have increasingly been moving to compile IARs of their digital assets, i.e., the information assets in their possession. This has usually taken the form of an information asset register, an internal data portal which is used to corral and manage data internally. Over the last decade, Link Digital has had experience with all three of these forms of data sharing.

It is important to stress that alongside recognition of the benefits of open data and data sharing, there is a growing recognition of the problems and the barriers preventing the open data movement from growing and reaching its full potential. While more data than ever before is being created and shared, there are also ‘tremendous inequities and asymmetries when it comes to data access,’ both within countries and between them.²⁴ As a 2023 report by the think tank [New America](#) states:

Digital divides encompass disparities in infrastructure and investment in digital technologies between the rural and the urban, between developing and wealthy nations, and among different socioeconomic and identity groups within societies. While wealthy countries push forward into new digital frontiers such as AI [artificial intelligence], developing economies still have limited access to digital markets, technologies, and broadband, as well as slower internet speed and a lack of opportunities for digital entrepreneurship.²⁵

Open data and data sharing has undoubtedly created new opportunities in terms of keeping governments accountable and improving policy outcomes, but it has also exposed people to a plethora of new risks related to data privacy, consent, surveillance and identity theft. These problems are encapsulated in what has been referred to as ‘the mosaic effect’. This is ‘derived from the [mosaic theory of intelligence gathering](#), in which disparate pieces of information become significant when combined with other types of information.’²⁶

This occurs when multiple datasets are linked to reveal new information. Even if data is appropriately anonymized, and efforts are made to remove personal identifiers, if there are multiple datasets containing similar or complementary information, it’s possible to determine identity based on the various data combined across the datasets such as gender, location, educational status, etc.

¹⁶ Open Data Charter, International Open Data Charter, September 2015.

¹⁷ Ibid.

¹⁸ Productivity Commission.

¹⁹ International Open Data Charter, 1.

²⁰ ‘Trust in public institutions: Trends and implications for economic security,’ United Nations Department of Economic and Social Affairs, July 20 2021, <https://www.un.org/development/desa/dspd/2021/07/trust-public-institutions/>.

²¹ Edelman, 2023 Edelman Trust Barometer: Global Report, February 8, 2023, 4, <https://www.edelman.com.au/trust/2023/trust-barometer>.

²² International Open Data Charter, 3.

²³ ‘Business-to-government data sharing: Questions and answers,’ European Commission, accessed June 20, 2023, <https://digital-strategy.ec.europa.eu/en/faqs/business-government-data-sharing-questions-and-answers> accessed June 20

²⁴ Stefaan, G. Verhulst, Andrew Young, Andrew J. Zahuranec, Susan Ariel Aaronson, Anie Calderon, and Matt Gee, The Emergence of a Third Wave of Open Data: How to Accelerate the Re-Use of Data for Public Interest Purposes While Ensuring Data Rights and Community Flourishing (New York: Tandon School of Engineering, New York University, 2020), 4.

²⁵ Gordon LaForge and Patricia Gruver, Governing the Digital Future (New America, Washington D.C., 2023), 18, <https://www.newamerica.org/planetary-politics/reports/governing-the-digital-future/>

²⁶ Melissa Edmiston, Stephanie Coker, Stephanie Jamilla, and Thembellhle Tshabalala, ‘The Pros and Cons of Open Data,’ MERL Centre, accessed August 2, 2023, <https://merlcenter.org/guides/pros-and-cons-of-open-data/>

And despite the massive advances in digitisation that have taken place over the last 20 years, the 'rules and regulations dictating data usage remain sparse.'²⁷

While this failure offers additional flexibility to some, the broader effect is greater uncertainty about risk. Organisations do not understand how to respond coherently when data is abused or misused. Ad hoc responses to daily challenges have eroded public trust [in open data] and damaged institutional credibility.

This is an issue that is transnational in scope.

It will take governance systems to mitigate the risks and harms of the digital revolution. Yet so far, global digital governance is incoherent and patchwork—fractured along technical, national, geographic, and sectoral lines. Countries impose domestic regulations, but cyberspace is transnational, and digital technologies proliferate at astonishing speed. These technologies challenge the centuries-old notion of sovereignty as distinctly territory-bound, a consensus that has underpinned the international order for centuries. The sovereignty of nation-states still depends on control over physical terrain, but in the theoretically borderless landscape of cyberspace, sovereignty is unbound from conventional geography.²⁸

All these issues have been exacerbated as open data and data sharing ecosystems rush to embrace AI. Often lost in the growing debate about the problems, risks and benefits associated with AI is the fact that machine learning models are only as good or bad as the data they are trained on. This is especially the case for 'generative AI', an umbrella term for machine learning algorithms, such as ChatGPT, which generate artificial digital content. 'Rather than merely processing existing data, generative AI entails the creation of machines or models that can generate new content, such as images, videos, music, or text.'²⁹ Among the key concerns around AI are its implications for privacy, security, reliability, accuracy, fairness and inclusion. In addition to longer term ethical questions, these issues have very real and immediate implications. As the Canadian Government's recent Directive on Automated Decision Making puts it: 'An AI tool could, for example, decide whether someone is eligible for a service, determine the level of benefit someone is entitled to, or process survey data to inform policy direction.' Central to this concern is AI's capacity to exhibit bias. This might occur in content generated by ChatGPT or it might be a service decision influenced by an AI algorithm, either of which might be discriminatory, non-representative or based on stereotypes, including racial identity and gender.

Link Digital understands that open source data is essential to create what has been referred to as 'ethical AI'.³⁰ This is AI that is more trustworthy, safer, diverse and socially beneficial. In the words of one commentator:

Open-source data science offers one of the most promising models for minimizing AI bias because it enables collaboration, trust, and transparency. In closed or proprietary systems, the engineer has complete control – and complete responsibility – over how the model behaves. But in an open system, engineers benefit from the perspectives, insights, and contributions of others working on similar problems.³¹

In this respect, however, it is not enough for data to be open and shareable, it also needs to be high quality. Given that most AI systems are only as good as the data they are trained on, many organisations have developed laws and suggested protocols to help make data of higher quality. For example, the European Parliament is about to adopt new regulations that stipulate that foundation AI model providers, such as Google and ChatGPT owner OpenAI, must describe data sources, including potentially copyrighted data, that have been used to train their AI models. Guidelines recently released by UNESCO stress the vital importance of effective governance frameworks to ensure data collected is of sufficient quality. This includes ensuring that the data is not outdated, is comprehensive and collected from credible sources. It should have been collected 'with consent only and not in privacy-invasive ways. Data about people should be disaggregated where relevant, ideally by income, sex, age, race, ethnicity, migratory status, disability and geographic location.'³²

Another key barrier in the way of the open data movement is the reluctance of the private sector to embrace open data and data sharing initiatives, with the result that most of its data remains unavailable and used purely for commercial gain. This is despite the considerable potential advantages of open data and data sharing initiatives for the private sector. According to a 2021 report by Frontier Economics for the Open Data Institute, 'Scaling to the GDP of the 20 largest economies in 2019, estimates suggest that data sharing could unlock between 700 billion and 1.75 trillion US\$ in value.' 'People spend 60% to 80% of their time trying to find data and dealing with data-quality issues to get it ready for analysis,' Dan Vesset, group vice president of the International Data Corporation was quoted as saying in a 2021 article.³³ 'It's a huge productivity loss when you are spending a substantial amount of time recreating data assets that exist somewhere else in the organisation because you simply lack internal visibility.' Some commentators argue that many in the private sector are approaching data the wrong way, as something purely to be extracted, amassed, used and, sometimes, abused.

As we look to the new data economy, we need a new guiding metaphor: a way of thinking about data that focuses less on simply exploiting a resource, and more on building equitable and sustainable relationships with consumers. The reality is that data isn't a resource to be passively extracted from consumers – it's more like a currency that consumers actively invest in in order to unlock specific benefits.³⁴

²⁷ "Primer #5: Establishing Governance Frameworks," The Third Wave of Open Data Toolkit.

²⁸ LaForge and Gruver, 10

²⁹ United Nations Educational Scientific and Cultural Organisation, Open data for AI: What now? (UNESCO, Paris, 2023), 39, <https://www.unesco.org/en/articles/open-data-ai-what-now>

³⁰ Natalie Williams, "AI and data trends for 2023," University of Technology Sydney, November 23, 2022, <https://open.uts.edu.au/insights/professional-development/ai-and-data-trends-2023/#>

³¹ Sam Babic, "Is Open-Source Data Science the Key to Unbiased AI?" Spiceworks, February 10, 2023, <https://www.spiceworks.com/tech/artificial-intelligence/guest-article/is-open-source-data-science-the-key-to-unbiased-ai/>

³² Open data for AI: What now?, 17.

³³ Beth Stackpole, "Making the business case for a chief data officer," MIT Management Sloan School, February 8, 2021, <https://mitsloan.mit.edu/ideas-made-to-matter/making-business-case-a-chief-data-officer>

³⁴ Jon Suarez-Davis, "Data isn't 'the new oil' – its way more valuable than that," The Drum, December 12, 2022, <https://www.thedrum.com/opinion/2022/12/12/data-isn-t-the-new-oil-it-s-way-more-valuable>

While this paper has a significant focus on public sector government initiatives such as data portals, it is important to stress that private sector activities in the data sharing space, including initiatives to better organise data internally and for public presentation, will grow in importance. For example, Link Digital believes that the connection between open data and data sharing activities and maintaining corporate competitive advantage is something that will only become stronger in the years ahead and is likely to emerge as an increasingly important driver of share price. This is especially the case for corporates who have an interest in being good actors over and above the level of regulation they are officially subject to. Such corporates often report data related to their ESG in their annual report, but there is pressure for them to deliver more information, in a timelier manner. Climate risk is another area where analysts are increasingly wanting to use up to date data to make market decisions.

III. The Third Wave of Open Data

As has been previously discussed, the COVID-19 pandemic merely accelerated the general trend towards open data and sharing data that was already underway, particularly on the part of governments seeking to respond to complex public policy challenges. Government activity in this space has taken three main forms.

The first has involved the establishment of public portals to share data with other government agencies, as well as the public and private sector.

The second has been efforts to standardise data and data sharing by adopting the [FAIR principles](#) to ensure public data can be accessed efficiently and to allow greater interoperability between data sharing portals. First enunciated in a 2016 article in the open access journal, *Scientific Data*, the FAIR principles are Findability, Accessibility, Interoperability, and Reuse of digital data.³⁵ The FAIR principles have gradually gained traction amongst the international research community and are now widely seen as a crucial component of data accessibility. An April 2022 article in the journal *Nature* by physician [Matthias Scheffler](#) and several colleagues, asserted the importance of FAIR principles for successful open data infrastructure, particularly with the uptake of AI, thus:

The enormous amount of research data produced every day... represents a gold mine of the twenty-first century. This gold mine is, however, of little value if these data are not comprehensively characterized and made available. How can we refine this feedstock; that is, turn data into knowledge and value? For this, a FAIR (findable, accessible, interoperable and reusable) data infrastructure is a must. Only then can data be readily shared and explored using data analytics and artificial intelligence (AI) methods.³⁶

The FAIR principles emphasise what is called machine actionability – “the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention because humans increasingly rely on computational support to deal with data as a result of the increase in volume, complexity, and creation speed of data.”³⁷ Interoperability is another important component. The [Australian Government Records Interoperability Framework](#), for example, focuses on semantic interoperability to allow humans and machines to clearly understand information’s context. Another practical example of the FAIR principles in action is the [EU’s Interoperability Framework](#). The Framework aims to provide leaders of EU local administrations with a common model of open data use to facilitate delivery of services to the public across domains, cities, regions and borders. First developed in 2017, the Framework was recently redesigned to include participation by a wider range of stakeholders, including corporates, non-profit and community organisations, as well as to bring into the frame the human, social, cultural, legal and ethical aspects of large scale data usage.

As the FAIR data principles are increasingly applied within the data ecosystem, it is expected that standards in data management will become higher. Accordingly, the third form of government activity related to open data and data sharing has been redesigning data governance. [The University of Queensland](#) defines data governance as involving ‘the organisation of people, processes, and technology to enable effective data management’.³⁸ This includes roles and responsibilities, the technologies, systems and tools to enable data discovery, and policies and procedures to support security, privacy, ethical use and data quality. Creating governance frameworks to enable effective data enabled operations and data sharing is important not only to ensure successful service delivery but to build and maintain public trust.

Link Digital’s service offerings encompass all these government activities: the development of open data portals, an activity which is still highly relevant and in which we remain heavily engaged in; efforts to standardise data and data sharing by adopting the FAIR principles; and data governance. Our position as a data and digital agency with deep roots in the open source software community is also completely in sync with the Third Wave.

The First Wave of Open Data relied on Freedom of Information laws and utilised regulation and legislation to access certain requested data. The Second Wave, which coincided with the dot-com boom, began with the arrival of open source software and Web 2.0 in the late 1990s and relied primarily on open government data (though some private sector data was also included) and sought to make data open by default. It often opened datasets without a clear understanding or mapping of how they could best be used.³⁹

³⁵ Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, and Niklas Blomberg, et al, “The FAIR Guiding Principles for scientific data management and stewardship,” *Scientific Data* 3, 160018 (2016): 1–9, <https://doi.org/10.1038/sdata.2016.18>

³⁶ Matthias Scheffler, Martin Aeschlimann, Martin Albrecht, Tristan Bereau, Hans-Joachim Bungartz, Claudia Felser, et al, “FAIR data enabling new horizons for materials research,” *Nature*, Vol 604, 2022, 635.

³⁷ “Fair Principles,” Go Fair, accessed June 21, 2023, <https://www.go-fair.org/fair-principles/>

³⁸ “Data Governance: what is it and how can it help you?,” The University of Queensland, April 27, 2023, <https://data.uq.edu.au/data.uq.edu.au/blog/data-governance>

³⁹ The Emergence of a Third Wave of Open Data, 5–6.

The Third Wave builds on and enhances these previous activities. Foremost, this is by focusing on the creation of quality data. [The World Wide Web Foundation's](#) Open Data Barometer noted as early as 2017 that a great deal of government data is incomplete and can be of low quality, which poses a threat to the long term support for open data initiatives.⁴⁰ The poor quality of much of the data available on the open Internet has also been touched on in the context of the data used to train AI, and which contain biases that AI simply perpetuates. In addition to the implications for individuals or groups of people, inaccurate, out of date, incomplete data or biased data can have significant implications for organisations and businesses, by leading to incorrect assumptions and decisions and undermining trust in those providing digital services.

Linked to this, many governments increasingly view quality open data as central not just to making more informed public policy decisions but inoculating public institutions from the threat of disinformation. 'Unfortunately, information pollution can make it near impossible for regular folk to inform themselves,' writes Link Digital's De Costa.⁴¹

It is even worse if we don't have a clear and trusted voice coming from our public institutions. With questionable information being put in front of us, we risk becoming questionable actors in our own society...With strengthened public institutions, data and digitally capable of meeting the needs of our communities, we will continue to raise up our children, ourselves and our society.⁴²

Canada, for example, explicitly places open data and digital government as part of a broader mission to safeguard democracy and protect public institutions from the threat of deliberate disinformation. Canada's 2022-24 National Action Plan on Open Government, developed as part of the country's membership of the OGP, includes a [Digital Citizen Contribution Program](#) that supports research on 'democracy and social inclusion in Canada by enhancing and/or supporting efforts to counter online disinformation and other online harms and threats.' The Canadian Government also has an explicit strategy of using data to foster equity, fairness and inclusion. [Statistics Canada's Disaggregated Data Action Plan](#), created under the aegis of the [2023-2026 Data Strategy for the Federal Public Service](#), includes the ambitious aim to improve the quality of data on specific population groups, with a current focus on Indigenous Peoples, women and people with disability. Addressing gender gaps, racism and systemic barriers to equity requires data to be used in such a way that increases understanding of these issues and measures progress to ensure those gaps are closed.

The Third Wave stresses the adoption of 'a more purpose-driven approach', seeking 'not simply to open data for the sake of opening, but to focus on impactful re-use, especially through inter-sectoral collaborations and partnerships.'⁴³ As part of their attempts to make data more publicly accessible, for example, there is now greater cooperation between different departments and administrative functions within government, and between government and civil society. Such initiatives are increasingly taking place on a subnational level, i.e., a province, state, and city level. [Deloitte's Government Trends 2023](#) cites a wide range of examples, including specialised open data portals and intergovernmental portals to encompass everything from cities now funneling data into a single command point to enable greater efficiency and better decision making, to its use in collecting data for air traffic management to hotel bookings.⁴⁴

Another ambition expressed by the Third Wave is a need to increase corporate involvement in the open data and data sharing ecosystem, particularly on the part of small and medium-sized enterprises, who need help 'to realise the benefits it [data] can bring to drive business innovation.'⁴⁵ Deloitte argues that businesses need 'to willingly relinquish control over some data and work with policymakers to design viable solutions.'⁴⁶ To do this, however, government needs to work on developing 'a value proposition for the private sector: In creating plans for a data-sharing ecosystem generating public benefits, government leaders should showcase business value to encourage private-sector participation.'⁴⁷ This is particularly important given what Edelman's 2023 Trust Barometer notes as the increasing trust being placed in business compared to government globally, which is putting it 'under pressure to step into the void left by government.'⁴⁸

Other major Third Wave foci are:

- The 'broader technical, social, political and economic context within which data is produced and consumed.'⁴⁹
- Open data initiatives that encompass the private sector.
- The social impact of open data initiatives.
- Engagement on the part of open data practitioners with data end users and trying to involve a greater diversity of users.
- Building capacity and meeting open data demand at the subnational level, i.e., the public sector and other institutions in cities, municipalities, states, and provinces.
- Prioritising data responsibility and data rights.
- Articulating value and building an impact evidence base to encourage more players, including corporates, to become involved in open data activities.
- Establishing governance frameworks and seeking regulatory clarity.

⁴⁰ World Wide Web Foundation, Open Data Barometer, Global Report, 4th edition, 2017, 14-15, <https://webfoundation.org/research/open-data-barometer-fourth-edition/>

⁴¹ Steven De Costa, "Data is living rent free in your head," Link Digital, May 19, 2023 <https://linkdigital.com.au/news/2023/05/data-is-living-rent-free-in-your-head/>

⁴² Ibid.

⁴³ The Third Wave of Open Data, 5-6.

⁴⁴ Ramani Moses, Abrar Khan, Arpan Kumar Saha, Emma Downey, Aparna Prusty, and Shambhavi Shah. Government Trends 2023 (Deloitte Centre for Government Insights, 2023), 30.

⁴⁵ Gurin, Joel, Carla Bonina, Stefaan Verhulst. 'Private Sector', in The State of Open Data: Histories and Horizons, eds, Tim Davies, Stephen B. Walker, Mor Rubinstein, and Fernando Perini (Cape Town and Ottawa: African Minds and International Development Research Centre, 2019), 418.

⁴⁶ Government Trends, 30.

⁴⁷ Ibid, 34.

⁴⁸ Edelman, 4.

⁴⁹ The Third Wave of Open Data, 9.

IV. Open data state of play: the example of the Australian Government

There is a significant amount of literature examining the major progress made on open data initiatives in jurisdictions such as the United States, Canada and the EU, to name just a few. In terms of focusing attention on the advances in this area, however, it may be useful to focus on what is happening in one national jurisdiction, Australia.

The last few years have seen Australia undertake a range of open data initiatives. A [Digital Transformation Agency](#) was founded in 2016 under the oversight of the Minister for Finance. The Agency oversees strategy around the provision of digital services, what is known as the [Australian Government Architecture](#), 'a collection of digital artifacts and guidance materials that can be used as a guide to inform users how to align to the digital direction of government,' and national ICT investments. The Government's decision to align its data and digital activities with the FAIR principles has been backed up by developments such as the establishment of the previously mentioned Australian Government Records Interoperability Framework in 2019. More recent strategy has been driven by a 2020 document, [The Foundational Four: Starting an ongoing data improvement Journey](#). This emphasises the need for national government agencies to have a senior leader responsible for data management, including a clear strategy and objectives, well-articulated governance mechanisms, and a focus on asset discovery. In 2022, the [Data Availability and Transparency Act](#) was passed by Australia's national parliament to establish best practice for sharing government data. Overseen by the [National Data Commissioner](#), it sets out a regulation framework for the sharing of data from all Australian Government entities save those excluded on national security grounds. The Act also sets out further steps to make it easier to use government data, including the creation of data inventories to make data known and discoverable.⁵⁰

At the May [2023 Australian Open Data Forum in Canberra](#), organised by the [Institute of Public Administration](#), [Jenny Wilkinson](#), Secretary of the [Commonwealth Department of Finance](#) and co-chair of the [Secretaries Digital and Data Committee](#), stated the government is embarking on further reforms to maximise the benefits of data and data sharing to deliver better services.⁵¹ She noted that many of these have been foreshadowed in the draft Data and Digital Government Strategy prepared by the Department of Finance Digital Transformation Agency. The Data and Digital Government Strategy sets out many of the key aims, skills and aptitudes necessary for the Government to take its data and digital capability to the next level, including:

- A stated focus on digital inclusion and accessibility in service design and orientation.
- A 'digital by design' approach to deliver policies and services for people and business.
- A commitment to make more data assets held by the APS discoverable and further improvements to the competencies and capabilities of the APS.
- Reference to seeking to engage with additional actors in the data and digital space, including public and private sector bodies, researchers and the community.
- The stated emphasis on the need to build and maintain public trust in the Government's use of data and digital technologies. This includes improving the management of data incidents and breaches.

As Wilkinson stated it at the Australian Open Data Forum:

It is just so clear to me that there are a huge number of benefits for the government and the APS [Australian Public Service] in continuing to expand the ways in which data are used to provide insights into the economy and society.⁵²

As was the case in many countries, the COVID-19 pandemic saw [Australia's Treasury](#) make extensive use of "new and novel" sources of digital data "to understand the evolution of the economy during the pandemic."

This included using new data sources, like bank data transactions to understand spending trends, using Google mobility data to understand behavioural patterns, and using ATO [[Australian Tax Office](#)] data on payrolls to understand at a local government level the impact that the pandemic restrictions were having on labour markets across the country. All of these data were more granular and more timely than most of the standard data sources we had recently had available to us. They helped inform critical decisions on policy interventions... and, frankly, I cannot imagine how we would have navigated the pandemic without some of these new and novel data sources.⁵³

⁵⁰ Data inventories are essentially a form of Internal Asset Registers that Link Digital has been including as part of its recommended solution for government and enterprise private clients since 2015.

⁵¹ Jenny Wilkinson, "A data-driven digitally-enabled government," Vimeo, Institute of Public Administration Australia. Australian Government Data Forum: Pathways, Perspective and Practice, Date Showcase Three, May 17, 2023, <https://vimeo.com/829267591/5b27b780a6?share=copy>

⁵² Ibid, 7:15-7:27.

⁵³ Ibid, 12:44-13:44.

Another example she cited was the joint work done in 2020 by the [Department of Health and Aged Care](#) and the [Australian Institute of Health and Welfare](#) in relation to rapid access to mental health data to track the impact of COVID-19 lockdown policies initiated by state and territory governments during the pandemic. This enabled the national government to get an accurate picture on what is happening with mental health across the country during the pandemic, which in turn was used to better inform decision making.⁵⁴

Speaking at the same Open Data Forum, Gruen noted that the Australian government is also increasingly using multi-agency data sets to inform public policy.⁵⁵ He highlighted work by [Australian Bureau of Statistics](#) and [Reserve Bank](#) that uses a data set that contains information on rents for 600,000 rental properties across regional and capital cities in Australia, updated monthly. “With 600,000 data points you can derive detailed insights... This is getting really down and granular and its dramatic what you can find out when you have that much data.” Gruen also pointed to the APS’s [Data Capability Framework](#), developed in 2021, which sets out 26 capabilities for working with data in the public services, each of which having three levels: foundational, intermediate and advanced. This is part of a wider APS data workforce strategy that includes encouraging a diverse data workforce, career pathways for people in data roles, and a focus on the ethical adoption of AI and data security. Underlining these workforce initiatives, in late 2022, the APS introduced a range of data job role profiles, which include key descriptors and technology specialisations and relationships relevant to the role.⁵⁶ As has been discussed previously in this paper, Australian governments at the national, state and territory level are also utilising a range of open data portals.

V. The role of the Comprehensive Knowledge Archive Network Software

Central to the current global development of open data initiatives is CKAN. This is [open source software](#) for which the original source code has been made freely available and can be redistributed and modified. The guiding principle behind CKAN is that open data is a public good that should be non-right restricted and easily transferable, and that creates more value the more it is shared. There are essentially two types of CKAN projects, both of which Link Digital has extensive experience in: what we might call traditional open data programs, such as the data portals; and other data products that may be open or partly open, but for which the technical solution concerned is built to more targeted outcomes. For example, corporates can use CKAN to register and manage internal data assets. Non-government organisations can use it as a free and open source tool to publish operational data and broaden their user communities. And the scientific community can utilise it to enable public data repositories in areas of active research.

There are many ways to measure CKAN's impact. One is obviously usage. If software is being used it means people need it and find value in it. The growing role of CKAN in the digital ecosystem is evidenced by the fact that it serves as the key component of some of the largest and most successful open data portals around the world. [Recent research identified nearly 400 regularly updated CKAN data portals of different sizes in 59 countries.](#)⁵⁷ These are spread across six major continents, with Brazil and the United States having the largest prevalence, followed by Argentina, Australia, Canada, Germany, the UK, Spain, Italy, and Japan. The most significant open data portals that CKAN is a part of include:

- [The EU's Open Data Portal](#): open data from EU institutions, agencies and other bodies, in all over one and a half million datasets, searchable by country and a broad series of 176 catalogues.
- [The United States Open Data Portal](#): over 250,000 data sets held by agencies across the federal government.
- [Open Africa](#): the largest repository of data on the African continent.
- [Canadian Government Open Data Portal](#): the official website of the Canadian Government.
- [National Health Service](#): the UK's biggest health website. Millions of people access NHS content through partner websites, apps, internet-connected devices. More than 2,000 organisations share NHS content, including other NHS websites, local authorities, charities and commercial organisations, from start-ups to large technology leaders.
- [Opendata.swiss](#): the central open data portal of the Swiss Government.

Data management platforms based on open source technology such as CKAN have the potential to provide real world benefits over the long term as digital public infrastructure. They are a viable, low risk and long term investment for government and civil society organisations, something that is uncommon in the dynamic world of technology. And they provide a tangible way for civil society to keep governments open and ensure they are accountable for the decisions they make. In recognition of its real world impact CKAN was recently awarded the status of being a [Digital Public Good \(DPG\)](#) by the [Digital Public Goods Alliance](#). This was a recognition of CKAN's value as a driver of positive change in helping to attain many of the Sustainable Development Goals put forward by [United Nations \(UN\)](#). And CKAN is currently used by many multilateral organisations like the [UN Refugee Agency](#), the [United Nations Office for the Coordination of Humanitarian Affairs](#), [World Food Program](#) and the [Food and Agriculture Organisation](#). The recognition of CKAN as a DPG is also associated with its use by philanthropic organisations and by grassroots civil society movements.

⁵⁴ Ibid, 13.49-15.25.

⁵⁵ Gruen, Strengthening APS Data Capability, 18.56-19.05.

⁵⁶ "Data job role personas," Australian Public Service Commission, December 19, 2022, <https://www.opsc.gov.au/initiatives-and-programs/aps-professional-streams/aps-data-profession/job-personas>

⁵⁷ "CKAN Instance Analysis," Pathways to Enable Open-Source Ecosystems, accessed August 3, 2023, <https://civildataecosystem.org/2023/06/19/analysis.html>

CKAN has the following features, which in turn are built into the data portals listed above and others:

- It aligns with the FAIR data principles of Findability, Accessibility, Interoperability and Reusability.
- It has rich data metadata features that enhance discoverability.
- It has an interoperability system that allows for cross-organisation and cross jurisdictional data sharing in a secure manner. Key to this is architecture based on application programming interfaces, referred to as APIs. These are software intermediaries that allow two applications to talk to each other, and hence to extract and share data within and across organisations.
- It has the flexibility and ability to support and manage a large range of diverse datasets.
- It allows organisations to tailor their open data platform according to their particular requirements and scale up as their data needs grow.
- It is non-commercial which allows the project to maintain its direction without being swayed by external financial interests.

Open source software such as CKAN has a large and expanding user base, including people drawn from private enterprise, academia and the community sector. In this way it can keep pace with changes and continuously improve due to the broad range of people contributing to its design, development and defect correction.⁵⁸ It is this 'absence of hierarchies' which 'enables ingenuity to proliferate in a compounded, decentralised fashion.'⁵⁹ The CKAN software is supported by a core development team that contributes to keep it in good shape. This is work that essentially never stops, meaning that CKAN is on a continuous improvement journey, with each successive iteration of the software looking at improving its efficiency, performance, stability, interaction, and usability.

VI. Link Digital's value proposition in the Third Wave of Open Data

Link Digital posits the following simple but powerful value proposition: that as a trusted and experienced purveyor of data competence we can help organisations, whether they be governments, non-profits or corporates to build data enabled operations in such a way to maximise the public benefit of data, while keeping transaction costs as low as possible.

Link Digital's expertise, firmly embedded in the data and digital ecosystem that emerged after 2010, is now fully in sync with the very latest thinking in relation to FAIR, the Third Wave, and the importance of open data to meet challenges around ethical AI and greater government and corporate accountability. In addition to being a CKAN co-steward and an active participant in the vibrant open source community around it, we are active in innovating with open data via supporting GovHack, active in policy direction in government through our contributions to the Data and Digital Government Strategy, and active in international efforts surrounding open government practices via our involvement in OGP events.

Clearly, open data and data sharing initiatives will continue to grow in significance across the spectrum of applications identified in this paper: open data portals, B2G data sharing, and IARs. Given this and the context of the Third Wave more generally, it is worth briefly reiterating some of the challenges inherent in this rapidly evolving field, both for organisations embarking on open data initiatives for the first time and participants already involved in the field but seeking to expand or maximise their activities. This serves to reinforce Link Digital's value proposition and how it can add value and assist organisations to take advantage of the potential offered by the Third Wave.

These challenges can be summarised as follows:

- There is more data than ever before, and it is growing all the time, meaning that the open data ecosystem is expanding in size and complexity.
- While open data portals will remain a key form of open data infrastructure, they will by necessity become more multifaceted and increasingly focus on enhanced data transferability and usability features. As the [Third Wave of Open Data Tool Kit](#) puts it: 'While the open data portal format will likely remain a common piece of technical infrastructure, new and sophisticated technological developments could facilitate greater collaboration and responsibility in data re-use.'⁶⁰ Future open data portals will become more complex and interactive, with features enabling users to contribute data, collaborate on open data projects, and engage in real time discussion about data. This will require an enhanced level of design and service support, to ensure portals remain user-friendly across a broad spectrum of levels of technological competence.
- Given that resources are limited it is important for organisations to identify priority areas where open data activities can provide the most benefit. This underscores the vital importance of engaging with relevant stakeholders to understand these needs.
- 'One key challenge is that much—possibly the majority—of generated data today resides in the private sector, hidden away in silos collected, controlled, and often monetized by companies and other entities.'⁶¹ Unlocking this will require new approaches and models of collaboration.

⁵⁸ Datashades.info, "CKAN portals across the global," accessed July 13, 2023, <https://datashades.info/>

⁵⁹ Ritika Puri, "Why Governments need open source more than ever," The Overflow, February 26, 2023, <https://stackoverflow.blog/2023/02/26/why-governments-need-open-source-more-than-ever/>

⁶⁰ The Third Wave of Open Data Tool Kit, "Primer #6: Creating the Technical Infrastructure for Re-Use."

⁶¹ The Third Wave of Open Data Tool Kit, 4.

As a developer and contributor to CKAN since 2013, Link Digital's core service offering is providing expert consulting, design and build capability for customised data repository management based on CKAN and other open source software. In addition, our service suite includes helping organisations identify priority areas where open data can provide the most benefit and engaging with relevant stakeholders to better comprehend their data needs and determine what types of capacity could add value. We understand the importance of data quality and useability and can help organisations enhance data discoverability, improve data interoperability, and promote data reuse.

We also appreciate that open source data projects are not just about technical or software problems. Attention needs to be paid to the human, cultural, legal and ethical aspects of large scale open data and data sharing projects. As the open data ecosystem increases in size and complexity, arguably so too do the consequences of mistakes and missteps, particularly in relation to the mishandling of data, making data privacy and security increasingly important. All these facets highlight the importance of what have been referred to as 'new data intermediaries,' partners already active in the ecosystem who can lower transaction costs in relation to collaborative relationships and negotiating the necessary legal agreements.⁶² Link Digital's expertise extends to helping organisations navigate data security, governance and policy implications such as how open data interfaces with organisational change, and the balance and diversity of skills, as well as how organisations relate to shifting conceptions of government and civil society.

Link Digital now has over two decades of experience working with government, non-profit and corporate clients globally to help them make meaningful change through open data. The fact that many of these clients have been with Link Digital for a decade is a testament to how we operate in line with the fundamental values of the open knowledge movement: transparency, ethical operation, and trust. Trust, or as the Third Wave of Open Data Tool Kit puts it, 'obtaining a social license'⁶³ is increasingly important in ensuring that open data re-use initiatives create public good. Link Digital's value proposition is validated through various successful projects we've delivered and our continued co-stewardship of CKAN. We are not just a service provider; we are contributors to the open-source community, helping shape the future of CKAN and open source data initiatives going forward. This stewardship reflects our dedication to the open data cause and ensures our knowledge and skills are constantly updated, thereby enabling Link Digital to deliver highly relevant solutions for clients. This holistic approach positions Link Digital not just as a service provider, but as a trusted advisor and implementation partner in open and shared data initiatives.

VII. Link Digital's value proposition in action: the Pacific Data Hub

Perhaps the best proof of Link Digital's value proposition is the projects it has been involved in. The open data portals Link Digital has designed and hosted for Australian governments at a national and state level demonstrate the trust placed in Link Digital by all levels of government is just one example of our reliability and expertise as a service provider in the open data space. Another Link Digital project that is worth mentioning is the [Pacific Data Hub](#) (PDH). Launched in December 2020 as a gateway to information from across the Pacific, the PDH illustrates CKAN's ability to store and organise a diverse array of data from different sources. It also underscores Link Digital's expertise in successfully bringing all the technical elements of open data dissemination together in one portal and assisting with the broader complexities involved in sourcing and organising data sets from across multiple geographic and administrative jurisdictions. The PDH is overseen by the [Pacific Community](#) (SPC), the principal scientific and technical organisation for the Pacific. According to [Sioeli Tonga, Program Manager at SPC](#), the portal's inception arose from the following questions:

How do we ensure that the data under our stewardship is protected, shared responsibly and shared ethically? And how do we ensure that it continues to be accessible beyond project funding horizons? And, also, how do we ensure that data can be used to inform policy developments and decision making that results in positive outcomes for Pacific people?

The PDH serves as a gateway to the most comprehensive data collection of relevance to the 22 Pacific Island nations and territories, including information on climate change, disaster preparedness, food security, gender equality, human rights, and employment.⁶⁵

The portal also pulls in related data from SPC's metropolitan members, Australia, New Zealand, the United States and France, and from the organisations such as the [World Bank](#), the UN, and the [Organisation for Economic Cooperation and Development](#). The portal is not only a natural extension of SPC's role as a steward of Pacific data. It assists the SPC to undertake critical work, including the compilation, updating and dissemination of the [Pacific Sustainable Development Indicators](#) to assist countries in meeting their UN's Sustainable Development Goals reporting requirements. In addition to helping to enable regional governments and bodies to create more informed and 'Pacific centric' solutions, the portal provides a sustainable open data infrastructure for member states, some of whom lack the means to establish and maintain one on their own.

⁶² The Third Wave of Open Data Tool Kit, "Primer #4: Supporting New Data Intermediaries."

⁶³ The Third Wave of Open Data Tool Kit, "Primer #7: Fostering Public Data Competence."

⁶⁴ Sioeli Tonga, "Establishing a Sustainable Data Infrastructure for the Pacific," YouTube, CKAN Monthly Live #5, November 17, 2022, <https://www.youtube.com/watch?v=PgICooCAMUs>, 4:55-5:24.

⁶⁵ "About Us," Pacific Data Hub, accessed June 20, 2023, <https://pacificdata.org/about-us>

The portal seamlessly combines CKAN with Drupal to present an open data portal that is attractive and 'highly functional to the public' in its design, and is fully usable for desktop, mobile and tablet devices.⁶⁶ It allows data to be searched by country or subject, and presented in different visual ways, including spatially and over time, and links data back to its source. The technical aspects of the project were complex; not only pulling in and presenting data collected over 70 years by SPC and stored in different tools and platforms, but data held by other organisations stored in their own individual platforms and systems. An even bigger challenge was establishing the trust of stakeholders, in other words, obtaining the social license. The portal took Link Digital two years to mature, from conception to approval of proof of concept, and to gathering a critical mass of data, features and context into the catalogue. 'That initial step of developing a prototype that we could use to demonstrate value quickly was really important, that got us the buy-in we needed to continue the journey,' maintains Tonga. 'I can't stress enough how important that partnership with Link Digital for us was to make that happen, because we certainly did not have the capacity or capability internally to do that ourselves quickly.'⁶⁷

In the context of this shifting terrain Link Digital's value proposition is that as a trusted and experienced purveyor of data competence, it can help organisations build data enabled operations in such a way to keep maximising the public benefit of data, while keeping transaction costs as low as possible. Link Digital is deeply knowledgeable of this space and the tools that can assist organisations to navigate and make the most of the opportunities presented by it. The key one of these is CKAN, a free, community led, open source software application, increasingly used by governments and other organisations around the world to make large scale data more usable and discoverable, and to enable independent analysis in ways that go far beyond the data's original purpose.

As the open data landscape evolves, Link Digital is prepared to adapt and grow, ensuring that our services continue to meet our clients' needs and contribute meaningfully to open data initiatives. It is well placed not only to help organisations navigate the Third Wave but the possibilities and challenges that lie beyond.

VI. Conclusion

Not only is data increasing. How we think about data is undergoing a fundamental transformation, as governments, researchers, civil society organisations, and the private sector increasingly realise that data needs to be open and shared to unlock its full benefits.

Powered by technological development and the broad uptake of the FAIR principles, the opportunities emerging from open data to improve the quality of government decision making, increase transparency and rebuild trust in government have never been more apparent. But so too are the challenges and risks that accompany this potential: inequality of data access and use, the threat to privacy and identity theft, and the growing complexity of the open data ecosystem. These opportunities and challenges require an approach to open data and data sharing that can encompass high quality data, enhanced transferability and useability, sound data governance, and fluency with new forms of working and collaborating.

While this paper has primarily focused on open data portals and B2G initiatives, Link Digital understands and is also well placed to meet the challenge of data driven activities of the private sector, such as IARs and efforts to use data to meet the growing demands for stronger corporate ESG standards. And Link Digital is now turning its attention to how the open data ecosystem will interface with broader questions involving human/digital interactions proposed by advances in AI and cybernetics. Not only is Link Digital across the technical and software aspects of open data curation and customising and building large scale open data portals, it understands the human, social, cultural, legal and ethical aspects of open data and data sharing projects. This is part of Link Digital's larger focus of assisting organisations to think differently about data and its use. Its services also cover hosting, maintenance, and support, to provide a comprehensive solution for organisations embarking on or enhancing their open data initiatives.

Research and authored by

Andrew Nette, Technical Writer/ Policy Analyst

Proofread by

Taleatha Franks, ANZ Marketing Manager

Layout and design by

Kate Sampher, Global Marketing Manager

Reviewed by

Marc Evans, CEO

Reviewed and approved by

Steven De Costa, Chairman

⁶⁶ "Pacific Data Hub", Link Digital, accessed June 21, 2023, <https://linkdigital.com.au/project/pacific-data-hub/>

⁶⁷ Tonga, "Establishing a Sustainable Data Infrastructure for the Pacific," 33.50-34.24.

Bibliography

Austin, Tasha, Kara Busath, Allie Diehl, and Pankaj Kamleshkumar Kishani. "Trustworthy open data for trustworthy AI: Opportunities and risks of using open data for AI." Deloitte Insights, December 10, 2021. <https://www2.deloitte.com/xe/en/insights/industry/public-sector/open-data-ai-explainable-trustworthy.html>

Australian Bureau of Statistics, APS Data Capability Framework, Version 1. Canberra: Commonwealth Government of Australia, 2022.

Babic, Sam. "Is Open-Source Data Science the Key to Unbiased AI?" Spiceworks, last updated February 10, 2023. <https://www.spiceworks.com/tech/artificial-intelligence/guest-article/is-open-source-data-science-the-key-to-unbiased-ai/>

Digital Transformation Agency. Data and Digital Government Strategy: The data and digital vision for a world leading-APS to 2030. Canberra: Commonwealth Government of Australia, 2021.

Davies, Tim, Stephen B. Walker, Mor Rubinstein, and Fernando Perini. The State of Open Data: Histories and Horizons. Cape Town and Ottawa: African Minds and International Development Research Centre, 2019.

Edelman. 2023 Edelman Trust Barometer: Global Report. February 8, 2023. <https://www.edelman.com.au/trust/2023/trust-barometer>

Edmiston, Melissa, Stephanie Coker, Stephanie Jamilla, and Thembelihle Tshabalala. "The Pros and Cons of Open Data." MERL Centre, accessed August 2, 2023. <https://merlcenter.org/guides/pros-and-cons-of-open-data/>

Frontier Economics. Economic Impacts of Trust, Report Prepared for the ODI, Frontier Economics. February 2021. https://theodi.org/wp-content/uploads/2021/03/RPT_Trust-in-data-ecosystems-23.02.21-STC-final-report.pdf

Gurin, Joel, Carla Bonina, and Stefaan Verhulst. "Private Sector," in The State of Open Data: Histories and Horizons, edited by Davies, Tim, Stephen B. Walker, Mor Rubinstein, and Fernando Perini, 418-429. Cape Town and Ottawa: African Minds and International Development Research Centre, 2019.

Klenk, Jueren, Tasha Austin, Costi Perricos, and Muthukumarana. "Fluid data dynamics: Generating greater public value from data." Deloitte Insights, March 4, 2021. <https://www2.deloitte.com/us/en/insights/industry/public-sector/government-trends/2021/fluid-government-data-dynamics.html>

KPMG and Microsoft. Unleash the full potential of open data in the Public Sector: How the right data foundations will set you on the path to AI. 2021. <https://kpmg.com/ae/en/home/insights/2021/04/unleash-the-full-potential-of-open-data-in-the-public-sector.html>

LaForge, Gordan, and Patricia Gruver. Governing the Digital Future. Washington D.C.: New America, 2023 <https://www.newamerica.org/planetary-politics/reports/governing-the-digital-future/>

Open Data Charter. International Open Data Charter. September 2015

Moses, Ramani, Abrar Khan, Arpan Kumar Saha, Emma Downey, Aparna Prusty, and Shambhavi Shah. Government Trends 2023. Deloitte Centre for Government Insights, 2023.

Office of National Data Commissioner. The Foundational Four: Starting an ongoing data improvement strategy. Canberra: Commonwealth Government of Australia, 2020.

Open Data Policy Lab. The Third Wave of Open Data Toolkit: Operational Guidance on Capturing the Institutional and Social Value of Data Re-Use. New York: Tandon School of Engineering, New York University, 2023.

Puri, Ritika. "Why Governments need open source more than ever." The Overflow, February 26, 2023. <https://stackoverflow.blog/2023/02/26/why-governments-need-open-source-more-than-ever/>

Suarez-Davis, Jon. "Data isn't 'the new oil' - its way more valuable than that." The Drum, December 12, 2022. <https://www.thedrum.com/opinion/2022/12/12/data-isn-t-the-new-oil-it-s-way-more-valuable>

Scheffler, Matthias, Martin Aeschlimann, Martin Albrecht, Tristan Bereau, Hans-Joachim Bungartz, Claudia Felser, et al, "FAIR data enabling new horizons for materials research," Nature, Vol 604, 2022, 635-642.

Treasury Board of Canada. Canada's National Action Plan on Open Government 2022-24. <https://open.canada.ca/en/content/national-action-plan-open-government>

"Trust in public institutions: Trends and implications for economic security." United Nations Department of Economic and Social Affairs. July 20, 2021. <https://www.un.org/development/desa/dspd/2021/07/trust-public-institutions/>

United Nations Educational Scientific and Cultural Organisation. Open data for AI: What now? Paris: UNESCO, Paris, 2023.

Verhulst, Stefaan, G., Andrew Young, Andrew J. Zahuranec, Susan Ariel Aaronson, Anie Calderon, and Matt Gee. The Emergence of a Third Wave of Open Data: How To Accelerate the Re-Use of Data for Public Interest Purposes While Ensuring Data Rights and Community Flourishing. New York: Tandon School of Engineering, New York University, 2020.

Wilkinson, Mark, D, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, and Niklas Blomberg, et al. "The FAIR Guiding Principles for scientific data management and stewardship". Scientific Data 3, 160018 (2016): 1-9, <https://doi.org/10.1038/sdata.2016.18>

Williams, Natalie. 'AI and data trends for 2023. University of Technology, November 23, 2022. <https://open.uts.edu.au/insights/professional-development/ai-and-data-trends-2023/>

World Wide Web Foundation. Open Data Barometer: Global Report, Fourth Edition. 2017, <https://webfoundation.org/research/open-data-barometer-fourth-edition/>