Assemblage theory, data systems and data ecosystems: The data assemblages of the Irish planning system

Rob Kitchin[1], Juliette Davret[2], Carla Maria Kayanan[3] and Samuel Mutter[4]

Maynooth University Social Sciences Institute, Iontas Building, North Campus, Maynooth University, County Kildare, Ireland

1. Rob.Kitchin@mu.ie, ORCID: 0000-0003-4458-7299
2. Juliette.Davret@mu.ie, ORCID: 0009-0005-7778-6595
3. Carla.Kayanan@mu.ie, ORCID: 0000-0002-4359-3534
4. Samuel.Mutter@mu.ie, ORCID: 0000-0002-9246-7147

Data Stories Working Paper 4
2 December 2024
datastories.maynoothuniversity.ie

The Data Stories project is funded by the European Research Council, grant no. 101052998.

Abstract

The work of the state has been transformed by digitalisation and datafication, with paper systems increasingly complemented and replaced by digital data systems. While much research has been undertaken on the effects of these data systems on e-government and e-governance, relatively little critical attention has been paid to the ontological nature of the data systems themselves, how they come into being, operate, iterate, and work in conjunction with other systems. In this paper, we build on work that understands the nature of data systems as data assemblages; that is, as being thoroughly socio-technical in nature, emerging and unfolding contingently, contextual and relationally through material and discursive practices and processes. We extend Kitchin (2014) and Kitchin and Lauriault's (2018) notion of a data assemblage through an engagement with assemblage theory, and detail how data systems are articulated and scaffolded into being, and scaled into a functioning data ecosystem, which is conceptualised as an assemblage of data assemblages. We illustrate the utility of this reconceptualization through an analysis of the development and control function of the Irish planning system, which is composed of a number of interlinked data systems used for managing the entire planning process from seeking planning permission, monitoring construction, to completed property.

Key words: data assemblage, data ecosystem, data system, digitalisation, assemblage theory, e-government

Introduction

Since the birth of modern digital computing in the 1950s, the state has been undergoing a process of digitalisation as bureaucratic and operational practices transfer from paper-based systems to digital systems. This process of digitalisation is still underway, but states are now thoroughly digital enterprises, with systems of e-government and processes of e-governance core modes of operation (Dunleavy et al., 2006; Falk et al., 2017). Accompanying this digital turn has been a 'data revolution', wherein digital data have become a vital asset for public bodies in managing their internal affairs, performing their operations, engaging with clients/customers, and coordinating their relationships with other organisations (Mayer-Schönberger and Cukier, 2013; Kitchin, 2014). Indeed, the past thirty years have witnessed large-scale datafication across all aspects of state work with the extensive, routine capture of fine-grained digital data across services and functions, the adoption of data infrastructures and the construction of data ecosystems, the rollout of big data systems and analytics, the digitisation of paper-based data, and the use of surveillance and dataveillance to manage and govern populations and organisations (van Dijck, 2013; Datta, 2023).

These twin processes of digitalisation and datafication have attracted significant critical attention in recent years, notably through the digital turn in disciplines such as Anthropology, Media Studies, Geography, Science and Technology Studies, and Sociology. However, the range of new concepts introduced to make sense of mass datafication and the proliferation of data-driven systems across all aspects of state services and operations (as well as business and civil society) are still in their infancy and typically constitute minor theory focused on specific aspects or issues (Kitchin, 2024). As such, they provide useful initial theoretical tools, but still require further development and elaboration and detailed empirical grounding. In this paper, we seek to advance and entwine the concepts of data assemblages and data ecosystems to: (1) advance a new over-arching conceptualisation of the organisation and operation of data-driven state work within and between public sector bodies and other entities at various scales; (2) map out the socio-technical construction and operation of this data landscape and how it forms and is held together and functions in practice. We illustrate our argument through a case study of the planning system in Ireland, with a particular focus on the development and control pipeline: making and processing planning applications, undertaking stakeholder feedback and public consultation, appealing decisions, monitoring construction and building control, and creating official statistics.

In the first part of the paper, we consider how to make sense of the data landscape of state work. We contend that data systems are socio-technical in nature (Bijker and Law, 1994; Feenberg, 1999), constituting data assemblages. Here, we use data system as a descriptive term for defining an IT system designed primarily to capture, manage, process and share data, with data assemblage being a conceptual term for explaining the produced nature of a data system. In other words, a data system is understood as being an assembled entity; as constituting a data assemblage. We outline Kitchin (2014) and Kitchin and Lauriault's (2018) initial conceptualisation of data assemblages and extend it: (1) through the application of assemblage theory as conceived by Nail's (2017) reading of Deleuze and Guattari (1987); (2) by examining how data systems become interconnected to form a data ecosystem that discursively and materially shapes how data work is undertaken within a domain. This data ecosystem is itself an assemblage composed of many interconnected data systems. We illustrate the utility of this assemblage thinking for making sense of state data work using our Irish planning case study. We detail how processes of articulation and scaffolding are employed to initiate, assemble and implement data systems, and how they continue to iterate in response to prevalent conditions, actors and objects, by examining the establishment and on-going operation of the Irish Building Control Management System (BCMS). Finally, we chart how several data systems are topologically arranged and interconnected through various seams and data mobilities to form a planning data ecosystem that sustains various forms of data-driven work along the entire development and control pipeline.

Understanding data systems as data assemblages

Kitchin (2014) defines a data assemblage as a complex socio-technical arrangement composed of many apparatus, actors and practices whose central concern is the production, management, analysis, and sharing of data. In Kitchin's (2014) formulation, a data system consists of the thorough intermeshing of a technical and contextual stack that frame and compose the data assemblage. The technical stack is, in effect, the embodiment of the data architecture – the component technologies (e.g., network, hardware, operating system, database, software, interface) that comprise the instrumental means by which data are generated, processed, stored, shared, analysed and experienced. The contextual stack consists of all the discursive and material elements that shape how a data system is built, operates and is maintained over time (e.g., systems of thought, forms of knowledge, finance, governmentalities, individual actors and communities, marketplace). In other words, all kinds of material apparatus and discursive elements are assembled together within multiple, overlapping contexts to produce and maintain a data system, defining what is 'possible, desirable and expected of data' in relation to a task or domain (Kitchin, 2014: 24).

Kitchin's formulation of a data assemblage draws inspiration from Foucault's (1980: 194) concept of the *dispositif*, which refers to a 'thoroughly heterogeneous ensemble consisting of discourses, institutions, architectural forms, regulatory decisions, laws, administrative measures, scientific statements, philosophical, moral and philanthropic propositions' that enhance and maintain the exercise of power within society. The dispositif of a data system produces what Foucault terms 'power/knowledge,' that is, knowledge that fulfils a strategic function: 'the apparatus is thus always inscribed in a play of power, but it is also always linked to certain coordinates of knowledge which issue from it but, to an equal degree, condition it. This is what the apparatus consists in: strategies of relations of forces supporting, and supported by, types of knowledge' (Foucault, 1980: 196). Data systems are expressions of power/knowledge: they shape what can be done, how it is performed, and for what ends. In other words, data systems are never a neutral, essential, and objective means of capturing, processing and sharing data, but are bundles of contingent and relational processes that do work in the world (Star and Ruhleder, 1996; Kitchin, 2014). They are complex socio-technical systems that are embedded within a larger institutional landscape of institutions and corporations (Ruppert, 2012). Data systems then are the product of data politics and data power (Kennedy et al., 2015; Ruppert et al., 2017), and they reinscribe and reproduce these relations.

Kitchin and Lauriault (2018: 9) observe that data assemblages are not discretely bounded, fixed and stable, but rather are emergent, contingent on an ever-shifting context and the mutability of action across actors and actants (non-human components). This contingency is present with respect to its day-to-day form and operation, but also its longer term development, with data assemblages constantly evolving 'as new ideas and knowledges emerge, technologies are invented, organizations change, business models are created, the political economy changes, regulations and laws are introduced and repealed, skill sets develop, debates take place, and markets grow or shrink' (Kitchin and Lauriault, 2018: 9). Data assemblages then are 'never entirely coherent, and [are] always being re-made' (Allen and Vollmer, 2018: 23). In other words, data systems designed to perform the same roles and tasks vary in their form, conditions, actors and use across organisations and context.

Kitchin and Lauriault (2018) used the ideas of Ian Hacking (1982, 1986, 1991), who likewise drew inspiration from Foucault's thinking, to explain how a data assemblage is constantly unfolding and how it gains legitimacy. Hacking posits that in the creation and use of a data system there are two interrelated processes at work: a 'looping effect' and 'engines of discoverability'. The looping effect concerns how data are classified and organized, how a data ontology comes into existence and its use reinforced, and how it can reshape that which has been classified (e.g., how planning data ontologies affect the practices of planning). Hacking (1986) explains that it has five stages: classification (identifying shared and desirable characteristics); objects of focus (wherein people come to understand and act toward the objects according to their classification); institutions (institutionalisation of classifications and management of data systems/infrastructures); knowledge (the formalisation of knowledge about and based on the data); and experts (those that implement and use data systems/infrastructures). Through this looping effect a process of 'making up' a domain occurs wherein the data system reshapes the structures and procedures of the domain in its image. For example, the workflow of operational planning tasks become aligned with fulfilling the data requirements of planning data systems. 'Engines of discoverability' are the means by which data are measured, turned into information and knowledge, and acted upon, including processes of counting, quantifying, correlation, establishing norms, creating bureaucratic institutions and procedures, and taking action (Hacking 1986). Such engines legitimate, reproduce and reinforce the data assemblage and its work (Kitchin and Lauriault 2018).

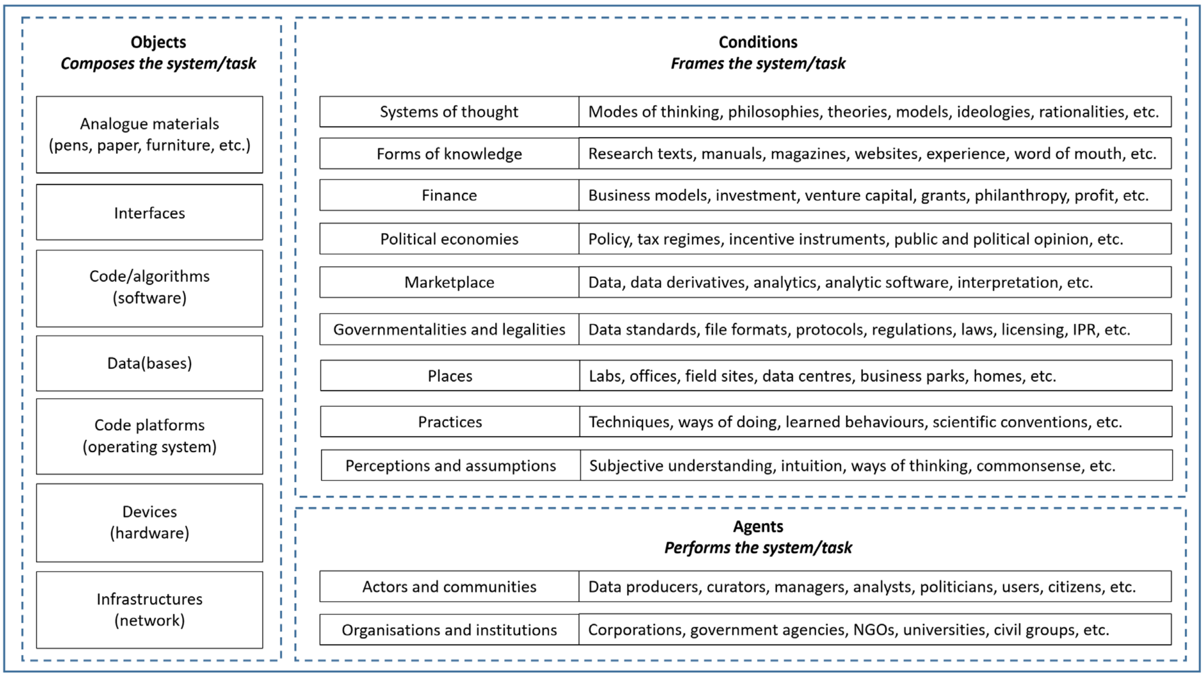
While Kitchin and Lauriault's formulation of a data assemblage is rooted in the ideas of Foucault and Hacking, its core ideas and logics (the assembly of material and discursive elements and processes; its dynamic, contingent, and relational nature), aligns with assemblage theory as devised by Deleuze and Guattari (1987) and extended and reworked by DeLanda (2006, 2016) and others. Here, we consider assemblage theory through Nail's (2017) reading of Deleuze and Guattari's (1987) formulation. Nail notes that Deleuze and

Guattari's vision of assemblage theory considers how a heterogeneous set of elements – that can be loosely categorised into three basic types (conditions, objects and agents, or what they call 'abstract machine', 'concrete assemblage' and 'personae') – are entwined through a set of contingent, relational arrangements to (re)produce an assemblage. As Allen and Vollmer (2018: 27) note: ''Conditions' are the abstract, governing ideas and sets of relations that connect objects and agents in meaningful ways. 'Objects' are the concrete parts that get arranged in particular ways. 'Agents' are those people who arrange the objects according to the prevailing conditions.' All three types of constituent elements are necessary for an assemblage to be produced, sustained, and to operate, and '[e]ach one presupposes and is immanent to the other' (Nail, 2017: 28).

Deleuze and Guattari (1987) identify four kinds of assemblage in which the dominant driver of how the assemblage is constituted and operates varies between with the balance of elements: territorial (objects), state (conditions), capitalist (actors), and nomadic (can shift between the three elements). While an element might dominate in each case, there is always a mix of drivers in operation. As we illustrate below, the data assemblages of the planning system are predominately state in form. As Nail (2017: 30) explains, '[s]tate assemblages are arranged in such a way that the conditioning relations attempt to unify or totalize all the concrete elements and agencies in the assemblage' as a means to manage populations and social and economic activity and coordinate and deliver services. These processes of management and coordination are stratified, organised vertically across scales and horizontally across organisations, with Deleuze and Guatarri (1987) identifying three main forms of stratification: binary (two way relations that might be hierarchical), circular (multiple relations that are entangled), and linear (one way relations along a dominant path, with possible deviations and side-loops) (Nail 2017). Assemblages and their stratifications continually reproduce or transform themselves, being subject to processes of change, or what Deleuze and Guatarri (1987) terms deterritorialization and reterritorialization.

Following Allen and Vollmer (2018), we think it is productive to rework Kitchin's (2014) data assemblage, reconfiguring the technical and contextual stacks and their component parts into the three elements of conditions, objects and agents (see Figure 1). In this framing, conditions (the abstract machine) are identified as broadly relating to knowledge (systems of thought, forms of knowledge), economics (finance, political economy, marketplace), governance (governmentalities and legalities), sites (places), and human action (practices, perceptions and assumptions), with their combination contextualising and shaping the production and functioning of a data assemblage. This formulation of a data assemblage aligns with Muller (2015) and Nail's (2017) observations that assemblages are relational (dependent on connections between entities to form new entities), productive (they produce new effects in the world), heterogeneous (composed of many apparatus and elements), dynamic (constantly mutating, transforming and breaking up), desired (they are designed to solve tasks), and they constitute a multiplicity being 'neither a part nor whole' (Nail, 2017: 23) and they can be disassembled and reassembled differently (unlike a body, which is a unity, an organic whole). In this sense, an assemblage is a 'fragmentary whole', since elements can be 'added, subtracted, and recombined with one another ad infinitum without ever creating or destroying an organic unity' (Nail, 2017: 23).

Figure 1: The conditions, objects and actors that make-up a data assemblage



Source: A re-organisation and extension of Kitchin (2014) and Kitchin and Lauriault (2018) to align with Deleuze and Guattari's assemblage theory.

Data systems assembled as a data ecosystem

While the concept of data assemblages has utility for understanding the development, constitution and operation of data systems, it also potentially places a boundary/scalar limit on this endeavour. Indeed, data systems are never fully isolated but are scaffolded together in order to ensure that a suite of tasks within a domain can be performed. In other words, data systems are themselves assembled together to form data ecosystems bound together by data mobilities (Bates et al., 2016; Kitchin et al., 2024) and socio-technical arrangements (contracts, governance arrangements, regulations, standards). In DeLanda's (2006) terms a data ecosystem is itself an assemblage, which in turn can be part of another data assemblage composed of multiple data ecosystems: society in his terms is composed of a multitude of interconnecting and nested assemblages that span scales in a flat ontology from local to global. Or to put it another way, the world is composed of 'assemblages all the way down', with '[e]ach component of an assemblage … also an assemblage' (Ball, 2018).

Much of the literature on data ecosystems (e.g., Oliveira and Loscio, 2018; Gelhaar et al., 2021) describe them on the same terms as Tarantino's (2019) notion of a datascape; that is, composed of a number of actors whose data systems interact with each other in order to exchange, produce and consume data around a common endeavour. Here, a set of technical, social and governance/legal interrelationships and interdependencies have been established between actors and systems that determines the constitution of the data ecosystem and how it operates (Scassa, 2019). There is little notion of the biological roots of the ecosystem metaphor evident in this framing. However, others have used the metaphor more literally, casting the notion of an assemblage as ecologies of interdependent relations. For example, van Schalkwyk et al. (2016) draw directly on the language and ideas of ecosystems theory to discuss how data ecosystems involve co-determinate and symbiotic relationships between mutually interacting organisms (firms, institutions, customers, etc.), including 'keystone species' such as data intermediaries that create enabling conditions for successful collaborations by providing services that add value (e.g., research and training consultancies). Organisms are complexly arranged, with movements between them cyclical and reinforcing, and interdependencies existing between organisms and resources, which together enables adaptation and creates resilience (van Schalkwyk et al., 2016). A literal application of the ecosystem metaphor might be appealing, but when applied in practice it soon becomes strained. Nonetheless, the ecosystem metaphor is useful for designating a set of interlinked data systems that collectively perform the data work that enables a domain, such as the planning system, to operate, but in our view is best conceived as an 'assemblage of assemblages' (DeLanda, 2016).

A core aspect of the functioning of a data ecosystem is the sharing and mobility of data within and between data systems, binding them together through co-dependent interconnections (Bates et al., 2016). Such data mobility is facilitated by seams (links and interfaces between systems) and shared metadata, standards, protocols (Inman and Ribes

2018) and hindered by frictions and vulnerabilities (e.g., access controls; incompatible data formats and standards; mistakes and glitches; resistance by actors; costs and skills capacities; and regulatory and legal limitations) (Edwards, 2010; Bates, 2018). Without these data mobilities the tasks being performed within a data ecosystem cannot be realised as downstream processes are reliant on upstream outcomes and input data (Kitchin et al., 2024). Data frictions that limit data mobility thus produce fragmented data ecosystems (Kitchin and Moore-Cherry, 2021). Data ecosystems are sustained through shared practices of maintenance and repair, adapt to prevalent conditions, and evolve through the introduction of new actors, innovations and relationships. They are held together through shared goals and legislative mandates and supported by institutional (e.g., strategic partnerships, licensing, memoranda of understanding) and technical (e.g., protocols, data standards) arrangements, guided by regulations, values and norms. At the same time, data ecosystems are full of data politics and data power and are open to exogenous forces that can render relations precarious or conflictual requiring mediation and active management, or threaten to destroy the ecosystem (van Schalkwyk et al., 2016).

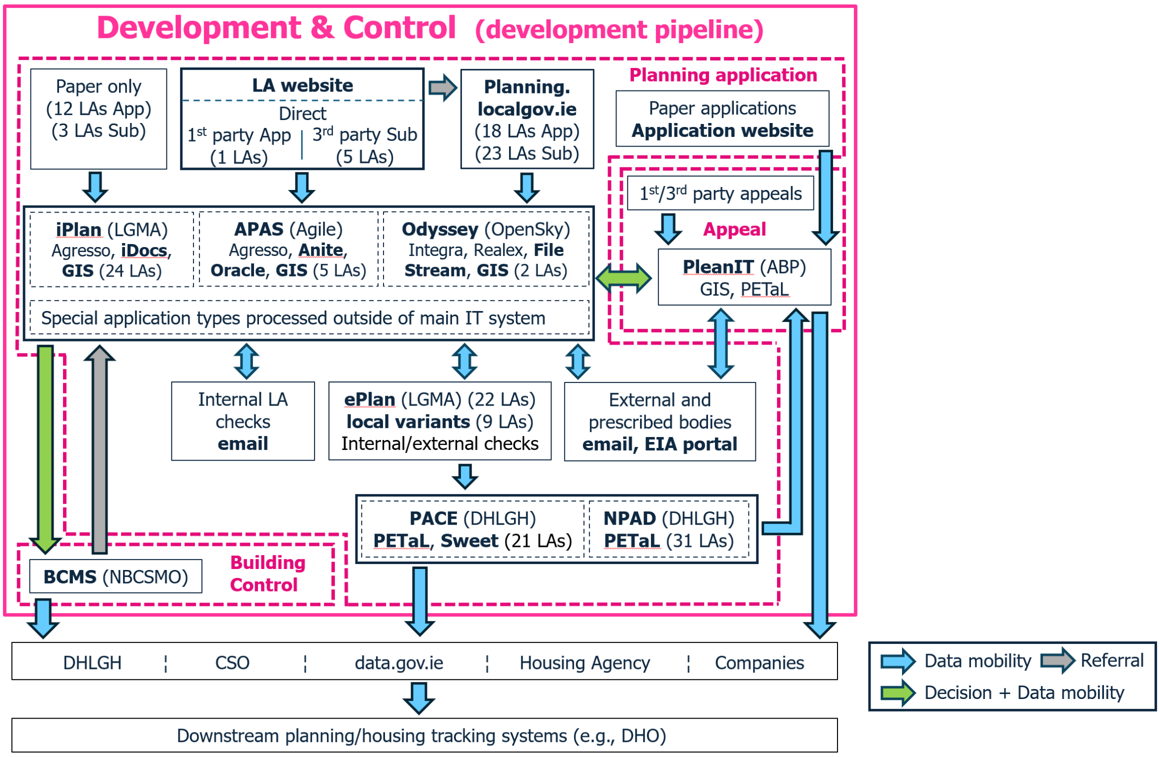
In the following sections, we illustrate our extension to the notion of data assemblage and data ecosystem with respect to the Irish planning system based on empirical research that sought to chart the various data systems and data infrastructures used by planning authorities and related actors to manage and fulfil their statutory functions and perform their various roles and responsibilities, document the data generated and how they are used, and map the relationships and data mobilities within and between each system and infrastructure across multi-tiered system of governance. Undertaken between June and August 2023, the fieldwork consisted of a complementary set of methods. Interviews were conducted with 29 public sector officials involved in data work within the planning system at local, regional and national scale. A number of the interviews were of a walk-through nature, with the interviewee demonstrating and explaining the workflow related to the use of a data system. This was supplemented by a close reading of the user manuals for these systems. A full data audit was undertaken of five systems: three planning application management systems (iPlan, APAS and Odyssey) used by LAs, the BCMS, and planning.localgov.ie (online application system). In addition, the data variables available in a number of downstream open data sites and planning/housing data tracking tools (e.g., Dublin Housing Observatory, Housing Delivery Tracker, Housing for All dashboard) were documented. This suite of methods enabled an understanding of the data architecture of each system, as well as how they were interconnected to form a wider data ecosystem.

The formation and on-going emergence of a data assemblage

Since the early 2000s the planning and control function of the Irish planning system, a role undertaken by local authorities (LAs), has been undergoing a process of digitalisation as the tasks and processes of planning assessment has increasingly become digitally mediated. In Dublin, the four LAs adopted the use of APAS, a commercial planning application management system, in 2000 to manage and process applications for planning permission. Other LAs in Ireland adopted iPlan from 2002 onwards, a management system developed by the Local Government Computer Services Board (LGCSB) for the same purpose. Planning applications were submitted to a LA as paper documents and staff would then scan them and enter key data into APAS or iPlan, which were used to process applications, store data and record decisions. While these systems could be used to manage most application types, in a number of specialised cases applications had their own procedures due to limitations in data system functionality. These specialised cases were typically handled using spreadsheets and/or GIS. By 2004, both APAS and iPlan had been complemented with online eplanning platforms that provided access to the public to selected components of planning applications so they could view and assess them, though submissions supporting or opposing developments had to be submitted on paper, which were then scanned and logged. Subsequently, several other data systems have been adopted to complement APAS, iPlan, and eplanning websites, to manage, assess and monitor the development pipeline from application to construction to turn-key to form a relatively complex planning and control data ecosystem (Author et al., forthcoming; see Figure 2).

Each of these data systems comprise a data assemblage that consists of a unique collection of objects, agents and conditions. Each is composed of a constellation of infrastructure, hardware, databases, software, interfaces and other elements, assembled into a data architecture designed to undertake designated tasks. Each has been assembled by a set of actors (software engineers, IT specialists, data stewards, policy makers, consultants, domain specialists) and stakeholders (LAs, government departments, state agencies, IT companies). This assembling has taken place within a context (conditions) that has shaped intention, design thinking and implementation, including forms of knowledge, established ways of undertaking work, political pressure, policy directives, financing, regulations, law, contracts, and so on, with these being negotiated within and across actors. A core conditioning factor is the relationship between data systems and on-going processes of digitalisation; for example, the extent to which a data system complements or replaces a paper-based system, or how it relates to, and is expected to interconnect with, other data systems. In the case of data systems such as iPlan, used by 24 LAs, while the data architecture is the same across LAs, workflows and work practices and cultures relating to its use can vary locally (it has its own constellations of local agents and conditions). With respect to APAS, the data architecture can also vary and in the Irish case the five APAS instances are quite different in their form.

Figure 2: The data assemblages and data ecosystem for terrestrial planning development and control in Ireland, August 2023



Bolded text = Planning data system or digitally mediated means of capturing planning data.
Anite, iDocs, File Stream, Oracle are document filing systems. GIS are used to view mapping data. Agresso, Integra, Realex are financial systems used for paying planning fees. EIA portal is the Environmental Impact Assessment system. PETaL is an extract, transform, load utility. Sweet is an ArcGIS solution for data collection.

As already noted, data systems do not come into the world ready-made (Bowker and Star, 1999), but ‘emerge through an incremental process of enacting, extending, standardising and embedding technical and social practices in specific contexts for unique needs’ (Aula

2019: 2). The notions of articulation and scaffolding provide a useful means for describing how a data system is made-up; how Hacking's (1986) looping effect and engines of discoverability are envisioned and actualised. Articulation refers to the identification, scheduling, coordinating and monitoring of all necessary tasks – and the steps in these tasks – to complete a job (Kaltenbrunner, 2015; Tanweer et al., 2016). In essence, articulation involves planning the workflow and resources needed, and aligning relevant actors, to undertake tasks and to construct and maintain systems and processes (Nadim, 2016). Each stage in the making up of a data system might be separately articulated, with the stages then meshed together. Articulation is not routinized, everyday production work, but rather the mapping out and aligning of the processes needed for such work (Kaltenbrunner, 2015). A key component of articulation is working through a course of action to reach completion despite problems encountered, such as glitches and unanticipated events. Scaffolding involves assembling the resources needed – data, technologies, finance, governance, and personal and institutional relationships – and to construct the bureaucratic structures and procedures and technical systems to fulfil these articulations (Halfmann, 2020). Scaffolding helps realise articulation work by providing the means to achieve its ambition; to scaffold the engines of discoverability together to form a data system. Over time, new rounds of articulation and scaffolding are employed as data systems and their operations are reviewed and revisions and upgrades are devised.

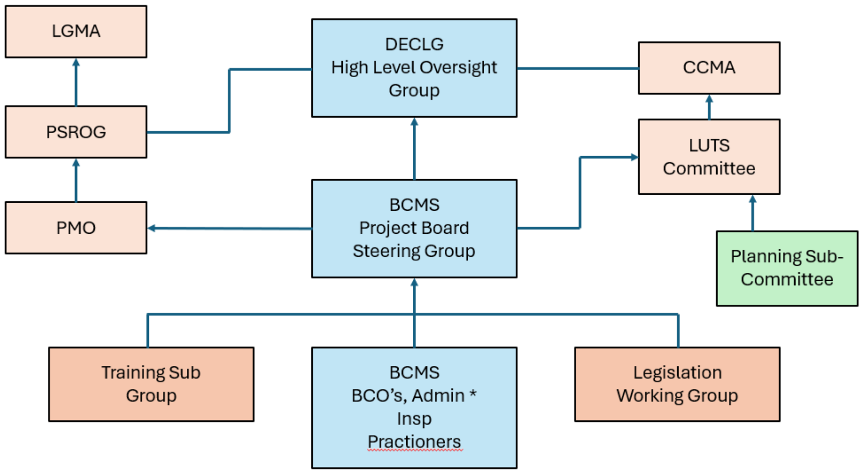
How objects, actors and conditions interact to articulate and scaffold a data assemblage is evident in the case of the BCMS, first introduced in 2014. The rationale for the BCMS was to provide a single, digital nationwide system, for use by all LAs, for tracking the pipeline of construction from commencement to completion and to monitor compliance with building control requirements. The initial idea for the project was driven by the introduction of new regulations (S.I.9 of the Building Control Amendment Regulations 2014) concerning the commencement and certification of construction works, which came into effect in March 2014; this included a need for completion certificates to be lodged in a new National Statutory Building Control Register (Dwyer-Bond et al., 2019). The new regulations were deemed necessary due to the failings of existing building control monitoring during the Celtic Tiger property boom which led to thousands of buildings with poor build quality, pyrite contamination, and inadequate fire safety measures (Ahern et al., 2018).

Prior to 2014, building control was monitored separately by the 34 building control authorities (which were the LAs; since reduced to 31 due to local government reform). Each

LA had its own system that differed in its forms and processes (e.g., paper and/or spreadsheet based), monitoring was a time and resource intensive task, and it was troublesome to try to produce harmonised data and a national picture of construction activity (Dwyer-Bond et al., 2019). In addition, it was possible for unqualified staff to certify compliance. BCMS would digitalise the monitoring of building control and be the new statutory national register, providing a standardised process and harmonised data across all building control authorities that could only be uploaded by accredited professionals. It would be delivered as a shared service for all LAs, administered by the new National Building Control and Market Surveillance Office (NBCMSO) hosted by Dublin City Council, following the model for IT development advocated in the Programme for Government, the Public Services Reform Plan, and Construction 2020 policy (O'Dowd, 2016). Rather than the initiative being driven by a single stakeholder, BCMS was a shared project of the Local Government Management Agency (LGMA), An Bord Pleanála (ABP, national planning appeals agency), the 31 LAs, and the Department of the Environment, Community and Local Government (DECLG), and it actively consulted with industry to try to ensure its proposed design was efficient and cost effective for those uploading data and documents (Dwyer-Bond et al., 2019). On initiation a relatively complex project governance structure was put in place to help guide the articulation and scaffolding of the system, including an oversight (strategic direction) and steering (project management) groups (see Figure 3).

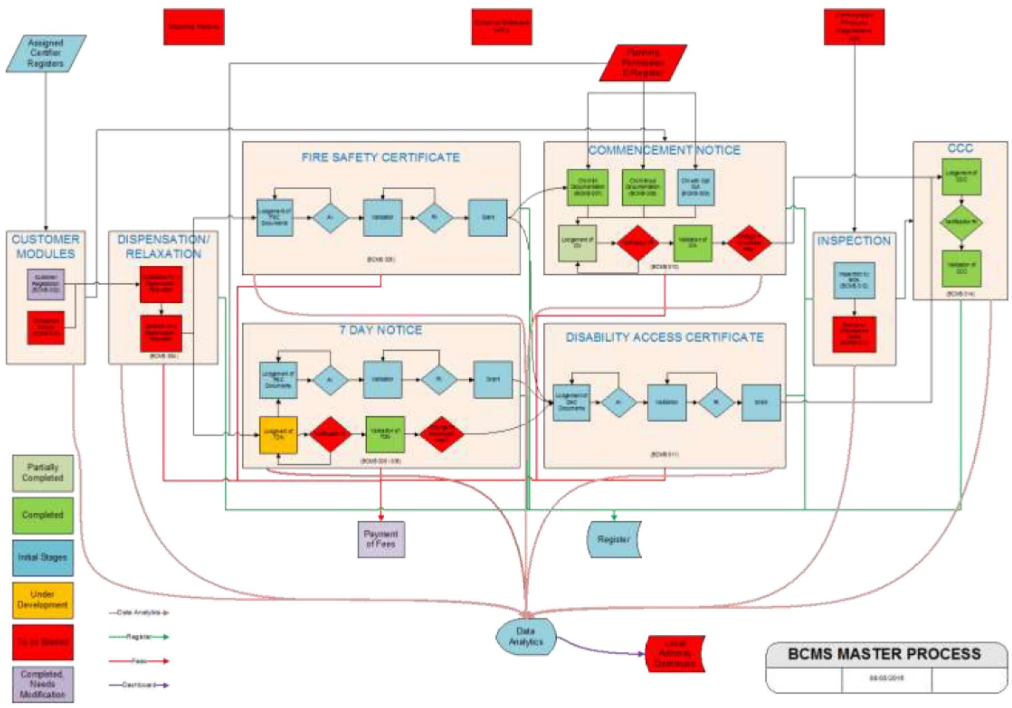
The articulation work undertaken by this governance structure and its various actors was contested and negotiated, shifting through various iterations, in order to try to gain consensus on the vision, functions and data architecture of the BCMS, ultimately leading to a project master plan (Figure 4) and an infrastructure master plan (Figure 5). These master plans provided the roadmap for development, with project staff then working to scaffold these into operation by assembling relevant resources within conditioning constraints (e.g., available finance, successful tendering, tendered company having sufficient skilled staff, knowledge, etc.). The tendered company was expected to work to a project timetable and milestones (see Table 1). This process of articulation and scaffolding continued for a couple of years, with an initial launch followed by continued development and rollout of additional functionality. It also had to deal with changing conditions, such as the intervention of the then Minister for DECLG, Alan Kelly, who subsequently altered the system specification by providing an exemption option for recording data in BCMS for those building one-off houses and housing extensions to avoid a 'cost burden' (Kelly, 2015; indicated by the inclusion of an 'opt out module' in Table 1).

Figure 3: BCMS project governance



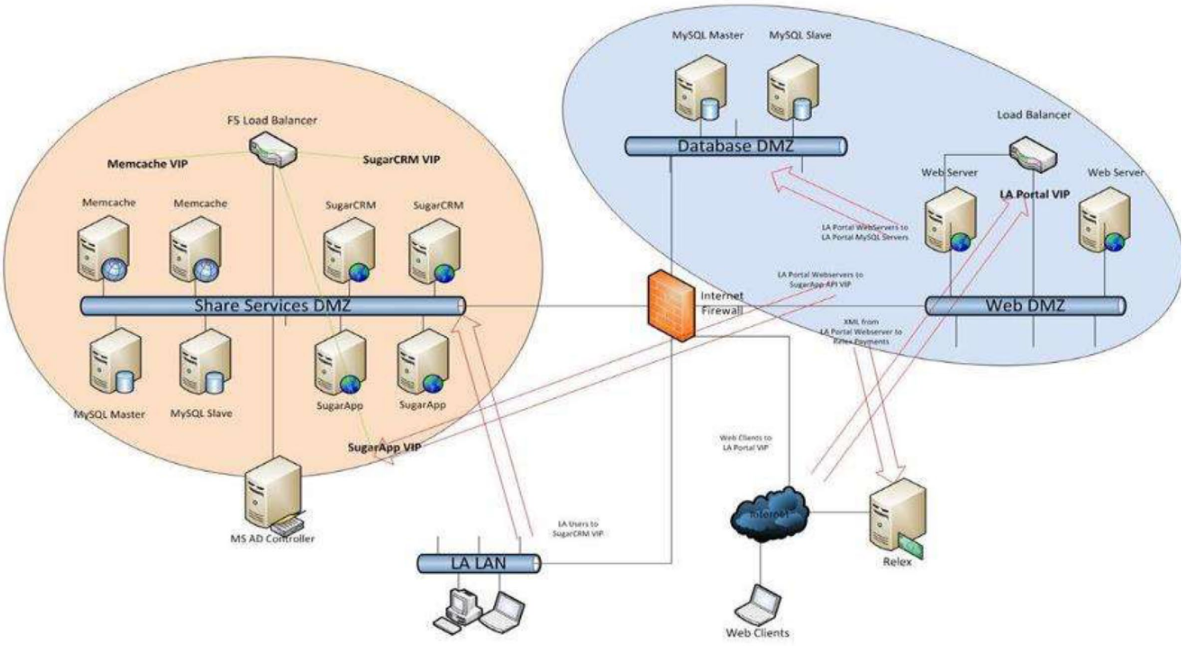
Source: Redrawn from O'Dowd (2016)
CCMA = City and County Management Association (a non-statutory body composed of the chief executives of the county and city councils).
PSROG = Public Sector Reform Oversight Group (it is a sub-committee of the LGMA board and is composed of senior representatives of the CCMA, LGMA, DECLG and the private sector).
PMO = Local Government Programme Management Office.
LUTS = Land Use and Transportation

Figure 4: BCMS project master plan



Source: O'Dowd (2016), original published quality.

Figure 5: BCMS IT infrastructure master plan



Source: O’Dowd (2016), original published quality.

Table 1: Timeline of projected project development for 2016

December 2015:
- Tender Review
- Review and Update of FAQ’s

January 2016:
- Successful tenderers informed
- Campaign reminder email issued
- Review of Framework for Building Control Authorities drafted
- Results of Resources Survey drafted
- February 2016
- CCC Module goes live

Launched 19th February 2016
- New front end interface
- User Guides for front and back end
- Commencement & Completion process fully online
- Opt Out Module

2nd Quarter 2016
- Re-design of Register
- Project Assessment
- 7 Day Notice functionality ‘switched on’
- Inspection / Enforcement Module
- Data Analytics
- Business Plan Review

3rd Quarter 2016
- Fire Safety Certificate Module
- Disability Access Certificate Module
- CN / 7 DN Revised Information function

4th Quarter 2016
- Links to Professional Registers
- Companies Module
- Dispensation / Relaxation Module
- Document Management solution
- System Architecture review and software updates

In subsequent years, the BCMS has been tweaked through processes of maintenance and repair, upgraded to new software and migrated onto new hardware, and extended through additional development work in response to user feedback and changes in regulations and legislation. Its design and operation is presently under review to address data quality issues and improve the ability to track developments across data systems. In other words, BCMS has constantly undergone a process of being reterritorialized; it is an emergent socio-technical system, shaped by shifting prevalent conditions, the work of diverse actors, and new technical innovations. An important aspect of this assemblage and its emergence is its relationship to and interactions with other data assemblages and its position and role in the wider planning development and control data ecosystem.

The data ecosystem of the Irish planning system

The delivery of the development and control function of the Irish planning system has, since its inception in 1963, consisted of an amalgam of inter-related and inter-connected data systems that form a data ecosystem. Data was initially captured in paper form, requests were sent to different internal and external units for feedback, the public's opinions were sought, decisions were made and shared. Data were stored in ledgers, folders, and file cabinets of different officials and were passed between teams and institutions. The processes of digitalisation and evolution of digital technologies, and the desire for institutional and regulatory control and the ability to formulate and track evidence-informed policy, have radically increased the volume of data captured, the degree of interconnection between data assemblages and the scope of data sharing, and by extension the complexity and scale of the data ecosystem.

Figure 2 charts the development and control data ecosystem as of August 2023, its various data assemblages, the actors primarily responsible for their operation and management, and their interconnections and data mobilities. The chart is loosely organised sequentially from top-to-bottom, with the exception of the appeal process, with different data assemblages being used to manage different stages of the development and control process from application, through assessment, appeal, construction, and the production of open data. As the various arrows indicate, each system is dependent on data mobilities and data being shared with other data systems in order to perform key tasks and collectively deliver the development and control functions as statutorily required by legislation (Kitchin et al., 2024). Interestingly, the data ecosystem operates without any one entity being in charge, or its actors necessarily knowing how it all fits together and operates. In fact, most of the actors we spoke with did not know the full extent of their own organisation's data work, and certainly had little knowledge of the data ecosystem as a whole.

At the top of the chart is the process of applying for planning permission. In August 2023, the digitalisation process remained partial. 12 LAs still required paper applications, and three only accepted third party submissions as a written letter. Likewise, ABP only accepted paper applications, though for large developments it requested the applicant to set up a website that hosted all relevant documents in digital form. For legal reasons, ABP printed out any digital material sent to them, such as emails, and added them to their paper case files. 18 LAs had adopted planning.localgov.ie as a portal for digital submission of applications, and one LA used their own portal. Data submitted via these portals are imported directly into the planning application management system. While the planning application management systems are used to process the vast majority of applications, limitations in their functionality meant that a number of specialised applications were handled outside them, including those made under Sections 5, 35, 42, 44, 57, 247 and Parts V, VII and XI of the Planning Act.

At the application stage, a check is made with payment systems (Agresso for iPlan and APAS; Realex or Integra for Odyssey) regarding whether the planning fee had been paid. After initial checking of submitted materials, selected details of the application are passed into an ePlan system, which allows the public to inspect the proposed development. Feedback is sought on every application from internal LA units (e.g., transportation, environment, and archaeology and heritage departments) and selected external bodies (e.g., Office of the Planning Regulator) and prescribed bodies (e.g., Department of Housing, Local Government and Heritage, DHLGH; Transport Infrastructure Ireland; Health Service Executive) via email, ePlan or EIA (Environmental Impact Assessment) portal. Selected data (up to 25 fields) are extracted from each LA's ePlan site using an automated ETL (extract, transform, load) process (PETaL) and imported into NPAD (National Planning Applications Database) and PACE (Planning Application Capture Environment). NPAD provides an open, online, interactive mapping tool to view all planning applications made to each of the 31 LAs since 2012 and is entirely dependent on source data from ePlan sites to function. If a planning application is unsuccessful then an appeal can be made to ABP who will request the transfer of associated data and documents from the relevant LA and, once a decision has been made, pass back the outcome and its conditions. Once permission is given for a development to proceed, the building control phase is initiated by notifying the BCMS of commencement.

Selected data are also shared with third parties for the production of official statistics and use as open data. In the case of the Central Statistics Office (CSO), each LA and ABP submit, on a monthly basis, 14 data fields relating to each granted planning permission in order to comply with the statutory provision under the Short Term Statistics Regulation (EC) Number 1882/2003 to supply Eurostat with data necessary to compile variables 411 and 412 of Annex B (Construction). Other selected data are shared with DHLGH (Department of Housing, Local Government and Heritage) and the Housing Agency, some of which is made openly available on their data hubs and via data.gov.ie (the national open data site). These data are imported into a number of planning and housing tracking systems that provide open dashboard visualisations and interactive maps for monitoring key performance indicators. These include:

- DHLGH Housing Delivery Tracker (https://storymaps.arcgis.com/stories/ab12ed6d50a540e2891170c24955ff49)
- Housing for All dashboard (https://public.tableau.com/app/profile/statistics.unit.housing/viz/HousingforAll/0_Overview)
- OPR Digital Planning Hub (https://opr-hub-oprgis.hub.arcgis.com/)
- DHO (https://airomaps.geohive.ie/dho/)
- Regional Development Monitor (https://rdm.geohive.ie/)
- Dublin Housing Task Force mapper (https://housinggovie.maps.arcgis.com/apps/View/index.html?appid=3fa56a71ee774f9487d14a9e5336b00c)

In addition, data held in ePlan and BCMS are scraped by private companies on a daily (Construction Information Services Ireland – CIS) and weekly (Building Information Ireland - BII) basis and converted into commercial data products. CIS and BII clean and wrangle the data into more useable forms, validate the data and link it to other datasets (such as procurement data scraped from e-tender portals), and produce analysis tools that enable site development to be tracked from permission to completion. Through the provision of open data and commercial services the reach and utility of the data ecosystem is massively extended.

While this data ecosystem functions largely as intended, in that it enables informed decisions to be made regarding planning applications and for building control to be monitored, it is not optimal or fully comprehensive in its scope or functionality and several data frictions exist that hinder the interconnections between data assemblages (Author et al., forthcoming). The continued use of paper and manual data re-entry is inefficient and weakens data quality through mistyping and miscodings. The lack of a consistent ID reference number across systems, with a planning application receiving unique IDs at pre-planning, planning, appeals, and construction phases, impedes the ability to track the development pipeline. Several types of application have to be processed outside of the planning application management systems. Each data system possesses its own data architecture (with its own database design, data standards, data ontologies, data dictionaries) reducing data compatibility and interoperability and limiting the ability to produce harmonious national-scale datasets and official statistics (e.g., iPlan and Odyssey make extensive use of free text fields, and Odyssey and APAS make strong use of check boxes and drop-down selections; iPlan has 65 required fields, whereas Odyssey has 40 and APAS 21) (Author et al. forthcoming).

Conclusion

The work of the state has been transformed by digitalisation and datafication. While paper remains an important media in state bureaucracy, it is rapidly being complemented and replaced by digital data systems. While much research has been undertaken on the effects of these data systems on e-government and e-governance, relatively little critical attention has been paid to the ontological nature of the data systems themselves, how they come into being, operate, iterate, and work in conjunction with other systems. Most research focusing on the nature of data systems is technically framed, mapping data architectures or mechanisms of operation, from a computer or information science, or human-computer interaction, perspective. However, as we have argued and illustrated in this paper, data systems are thoroughly socio-technical in nature, the product of objects, conditions and agents. They emerge and unfold contingently, contextually and relationally through material and discursive practices and processes, and they are saturated in politics and power. Data systems are assembled and continually reterritorialized, and they perform diverse work in the world. They are never constituted and work alone, but are always interconnected and interdependent on other data assemblages. Data systems are meshed together to form evolving data ecosystems.

We have sought to make sense of the data landscape of state work by extending Kitchin (2014) and Kitchin and Lauriault's (2018) notion of a data assemblage through an engagement of Nail's (2017) reading of Deleuze and Guatarri's (1987) assemblage theory. Using the example of BCMS and the wider development and control functions of the Irish planning system, we have illustrated how a data system is assembled through processes of articulation and scaffolding, and how several data systems are assembled into a functioning, emergent data ecosystem to mediate development and control processes and practices from planning permission to turn-key. Our reworking of the notion of data assemblages acknowledges that there are different types of assemblages (territorial, state, capitalist and nomadic) that are the unfolding product of prevalent conditions (e.g., systems of thought, forms of knowledge, finance, political economy, governmentalities and legalities), agents (e.g., actors, communities, organisations) and objects (e.g., infrastructures, hardware, software, data, interfaces). Data assemblages can themselves be progressively scaled (Massey, 1993) into larger assemblages. That is, they are relationally placed and connected to other entities, with data systems bound together through shared goals, institutional arrangements, legislative mandates, infrastructure, and data mobilities into data ecosystems. In other words, while it is feasible and legitimate to examine a single data system to understand its constitution and operation, this scale of analysis can obscure the position, relations and role of a data system in a wider data ecosystem.

While we believe that our conceptualisation of data assemblages and data ecosystem has wide utility for making sense of the processes of digitalisation, datafication and data-driven operations, more research is required to further develop its tenets and test its robustness. This includes exploration and application with respect to other forms of assemblage (territorial, capitalist and nomadic) and other domains (e.g., health, transport, education, etc.), and paying closer attention to the specific processes of articulation, scaffolding, and iteration at work.

Acknowledgements

The research reported in this paper was funded by the European Research Council (no. 101052998) and the Local Government Management Agency. The research received ethical approval from Maynooth University Research Ethics Committee and the Ethics Review of the European Research Council. During the research regular meetings were held with the LGMA Research Advisory Group and we are thankful to Dr. Bernie O'Donoghue Hynes, Tom Elders, Tara Kerrins, and Holly Morrin for facilitating access and providing direction and constructive advice and feedback. We are grateful to all the interviewees for spending time to explain and demonstrate their work and the systems they use. Thanks to Mairéad Phelan and Ronan Glynn of the National Building Control and Market Surveillance Office for insights into BCMS. Thanks to Oliver Dawkins and Danielle Hynes for feedback on an initial draft.

References

Ahern, E., Bescos, C.J. and Desai, A. (2018) Evaluating how Ireland has improved Building Regulations Compliance and Energy Efficiency. In Elsharkawy, H., Zahiri, S. and Clough, J. (eds) *Proceedings of International Conference for Sustainable Design of the Built Environment SDBE 2018*, London. pp. 421-432. https://aru.figshare.com/articles/conference_contribution/Evaluating_how_Ireland_has_improved_Building_Regulations_Compliance_and_Energy_Efficiency/23777493/1

Allen, W.L. and Vollmer, B.A. (2018) Clean skins: Making the e-Border security assemblage. *Environment and Planning D: Society and Space*, 36(1): 23–39.

Aula, V. (2019) Institutions, infrastructures, and data friction – Reforming secondary use of health data in Finland, *Big Data & Society* 6(2), 1-13.

Ball, A. (2018) Manuel DeLanda, Assemblage Theory. *Parrhesia* 29: 241-247.

Bates, J. (2018) 'The politics of data friction', *Journal of Documentation*, 74(2): 412–429.

Bates, J., Lin, Y.-W. and Goodale, P. (2016) 'Data journeys: Capturing the socio-material constitution of data objects and flows', *Big Data & Society*, 4(2): 1–12.

Bijker, W.E. and Law, J. (eds.) (1994) *Shaping Technology/Building Society: Studies in Sociotechnical Change*. MIT Press, Cambridge, MA.

Bowker, G. and Star, L. (1999) *Sorting Things Out: Classification and Its Consequences*. MIT Press, Cambridge, MA.

Datta, A. (2023) The digitalising state: Governing digitalisation-as-urbanisation in the global south. *Progress in Human Geography* 47(1) 141–159.

DeLanda, M. (2006) *A New Philosophy of Society: Assemblage Theory and Social Complexity*. London: Continuum.

DeLanda, M. (2016) *Assemblage Theory*. Edinburgh University Press, Edinburgh.

Deleuze, G. and Guattari, F. (1987) *A Thousand Plateaus: Capitalism and Schizophrenia*. Trans Massumi, B. University of Minnesota Press, Minneapolis.

Dunleavy, P., Margetts, H., Bastow, S. and Tinkler, J. (2006) *Digital Era Governance: IT Corporations, the State, and E-Government*. Oxford: Oxford University Press.

Dwyer-Bond, E., Casey, B. and Coates, D. (2019) Developing data sources for a Commercial Property Statistical System (CPSS) in Ireland, International Conference on Real Estate Statistics, 20-22 February 2019, Luxembourg. https://www.cso.ie/en/media/csoie/methods/methodologicalresearch/commercialproperty/ICREA_Ireland_2019.pdf

Edwards, P.N. (2010). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: The MIT Press.

Falk, S., Rommele, A. and Silverman, M. (eds) *Digital Government*. Springer.

Feenberg, A. (1999). *Questioning Technology*. Routledge.

Foucault, M. (1980). The confession of the flesh. In C. Gordon (Ed.) *Power/Knowledge*. New York, Pantheon Books, pp. 194–228.

Gelhaar, J., Groß, T. and Otto, B. (2021) A taxonomy of data ecosystems. In Bui, T.X. (ed) *Proceedings of the 54th Hawaii International Conference on System Sciences*. HICSS, Honolulu, pp. 6113-6122.

Hacking, I. (1982) Biopower and the avalanche of numbers. *Humanities in Society* 5(3&4): 279-295.

Hacking, I. (1986) Making up people, in Heller, T. et al. (eds) *Reconstructing Individualism*, Stanford University Press, Stanford. pp. 222-236.

Hacking, I. (1991) A tradition of natural kinds. *Philosophical Studies,* 61(1/2): 109-126.

Halfmann, G. (2020) 'Material origins of a data journey in Ocean Science: How sampling and scaffolding shape data practices', in Leonelli, S. and Tempini, N. (eds) *Data Journeys in the Sciences*. Springer Open, Cham, Switzerland. pp. 27-44.

Inman, S. and Ribes, D. (2018) 'Data streams, data seams: Toward a seamful representation of data interoperability'. Paper presented at Design Research Society, University of Limerick 25-28 Sept. https://dl.designresearchsociety.org/drs-conference-papers/drs2018/researchpapers/16/

Kelly, O. (2015) Builders of one-off homes can opt out of regulations from Tuesday. Irish Times, 31 Aug. https://www.irishtimes.com/news/environment/builders-of-one-off-homes-can-opt-out-of-regulations-from-tuesday-1.2334297

Kennedy, H., Poell, T. and van Dijck, J. (2015) 'Data and agency', *Big Data & Society*, 2(2): 1–7.

Kitchin, R. (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Sage, London.

Kitchin, R. (2024) *Critical Data Studies: An A to Z Guide to Concepts and Methods*. Polity Press, Cambridge.

Kitchin, R., Davret, J., Kayanan, C. and Mutter, S. (2024) *Data mobilities: Rethinking the movement and circulation of data*. Data Stories Working Paper 3, Maynooth University.

Kitchin, R. and Lauriault, T. (2018) 'Towards critical data studies: Charting and unpacking data assemblages and their work', in Thatcher, J., Eckert, J. and Shears, A. (eds), *Thinking Big Data in Geography*. University of Nebraska Press, Lincoln (2018), pp. 3–20.

Kitchin, R. and Moore-Cherry, N. (2021) Fragmented governance, the urban data ecosystem and smart cities: the case of Metropolitan Boston. *Regional Studies* 55(12): 1913-1923.

Massey, D. (1993) Power geometry and a progressive sense of place. In J. Bird et al., eds., *Mapping the Futures*. Routledge, London, pp. 60–70.

Mayer-Schönberger, V. and Cukier, K. (2013) *Big Data: A Revolution that will Change How We Live, Work and Think*. John Murray, London.

Muller, M. (2015) Assemblages and actor-networks: Rethinking socio-material power, politics and space. *Geography Compass* 9(1): 27-41.

Nadim, T. (2016) 'Data labours: How the sequence databases GenBank and EMBL-Bank make data', *Science as Culture*, 25(4): 496-519.

Nail, T. (2017) What is an assemblage*? SubStance* 46(1): 21-37.

O'Dowd, E. (2016) Building Control Management System: Supporting Building Control Compliance. Paper presented at IBCI Conference, Sligo. https://www.ibci.ie/docs/conferences/2016/2_Eoin_ODowd_BCMS_Supporting_Building_Control_Compliance.pdf

Oliveira, M.L.S. and Loscio, B.F. (2018) What is a data ecosystem? In Zuiderwijk, A. and Hinnant, C.C. (eds) *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*. ACM, New York. Article 74.

Ruppert, E. (2012) 'The governmental topologies of database devices', *Theory, Culture & Society*, 29: 116–136.

Ruppert, E., Isin, E. and Bigo, D. (2017) 'Data politics', *Big Data & Society*, 5(2): 1–7.

Scassa, T. (2019) 'Ownership and control over publicly accessible platform data', *Online Information Review*, 43(6): 986–1002.

Star, S.L. and Ruhleder, K. (1996) 'Steps toward an ecology of infrastructure: Design and access for large information spaces', *Information Systems Research*, 7(1): 111–134.

Tanweer, A., Fiore-Gartland, B. and Aragon, C. (2016) 'Impediment to insight to innovation: Understanding data assemblages through the breakdown–repair process', *Information, Communication & Society*, 19(6): 736-752.

Tarantino, M. (2019) 'Navigating a datascape: Challenges in automating environmental data disclosure in China', *Journal of Environmental Planning and Management* 61(3): 67-86.

van Dijck, J. (2014) ‘Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology’, *Surveillance & Society*, 12(2): 197–208.

van Schalkwyk, F., Willmers, M. and McNaughton, M. (2016) ‘Viscous open data: The roles of intermediaries in an open data ecosystem’, *Information Technology for Development*, 22(S1): 68–83.