Data mobilities: Rethinking the movement and circulation of digital data

Rob Kitchin[1], Juliette Davret[2], Carla 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 3
8 July 2024
datastories.maynoothuniversity.ie

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

Abstract

The mobility of data has been variously described as data: flows, streams, journeys, threads, transfers, exchanges, and circulation. In each case, the mobility of data is conceived as a movement from here to there; that data moves along a chain of receivers and senders. However, we contend that the metaphor of a data flow (or stream, journey, etc.) does not reflect well the sharing and circulation of digital data. Rather, data replicate (copies), with the original source retaining the data and a new source gaining it, and data proliferates (multiplies) and diffuses across systems and sites when made openly available. As data replicate and proliferate they are transformed through processes of data cleaning, data wrangling, data fusion and enrichment, producing new incarnations of the source data. Moreover, data does not replicate and diffuse alone, but with companions, such as other data (e.g., metadata, derived data) and information (e.g., documentation, visualisations). The replication, proliferation and diffusion of data is facilitated by seams (interface/connection points between systems) and aided by metadata, standards and protocols, and hindered by frictions and vulnerabilities. We illustrate our argument through a case study of the planning data ecosystem in Ireland.

Key words: data mobilities, data journeys, replication, diffusion, data flow, data frictions, data seams

Introduction

Over the past two decades, with the rise of mass datafication (van Dijck 2014), the growth of big data (Mayer-Schönberger and Cukier 2013) and open data (Davies et al. 2019), the proliferation of data infrastructures (Kitchin 2022), and the enormous expansion of data capitalism and data markets and services (Sadowski 2019), significant conceptual and empirical work has been directed at the production and use of data. Some of the attention paid to data has been focused on their movement and circulation, with a range of concepts forwarded to help make sense of data mobilities, including: data flows (McNally et al. 2012; Hoeyer et al. 2016), data streams (Hilgartner and Brandt-Rauf 1994; Dourish and Gomez Cruz 2018), data journeys (Bates et al. 2016; Leonelli 2020), data threads (White 2017), data transfers (Glouftsios and Leese 2023), data exchanges (Weltevrede and Jansen, 2019), data arrows (Flensburg and Lai 2023), data circulation (Beer 2016; Pelizza 2016), and data sharing (Borgman 2015; Davies et al. 2019).

While these conceptualisations of data mobility have utility our contention is that their rootedness in metaphors of movement is misleading and needs rethinking. As we illustrate through our case example of data mobilities in the operation of a planning system, digital data do not move in the conventional sense of a journey – leaving one place to travel to another. Rather digital data replicate (copies), with the original source retaining the data (though it might be deleted once copied) and a new source gaining it. Data proliferate and diffuse, and they mutate and transform as they multiply, altered by data cleaning, data wrangling, and their mediation by technologies, protocols, and practices. Numerous versions of the original source data can thus be produced. For example, a popular dataset on an open data site might be downloaded thousands of times, then be cleaned and wrangled in multiple ways, with these new versions shared through data repositories and publications (Thylstrup 2022). In other words, digital data circulate in a quite different way to other materials given their non-rivalrous (more than one entity can possess the same data) and non-excludable (it is easily copied and it takes effort to block sharing) nature, and the cost of reproduction has a zero marginal cost (it is effectively free to copy) (Floridi 2010).

In the next section, we discuss and critique how the movement and circulation of data has been conceived to date and set out our new theoretical framework for making sense of data mobilities. This is then followed by the methodology utilised to uncover and document the data mobilities at play and an overview of our case study of the Irish planning system, how it is constituted as a data ecosystem of interlinked data assemblages and its associated data mobilities. Drawing on the case study, we then detail the nature of data mobilities with respect to forms and entanglements of replication, the processes of transformation and proliferation, the effects of data frictions, and transitions in the constitution and operations of data mobilities. We conclude by arguing for more sustained theoretical and empirical attention to data mobilities.

Rethinking data mobilities

The mobility of data is vital to much of the work undertaken, and value produced, with data. The sharing of data enables the data lifecycle to occur, data enrichment and data fusion to take place, re-use and repurposing, and for new data products and services to be created. Analogue data, such as paper records, can be rivalrous or more easily constrained in its sharing unless digitised. Digital data, however, are generally non-rivalrous and are easily shared between media and locales (Floridi 2010). Here, we focus on the mobilities of digital data, which is the nature of the vast bulk of data presently in existence (Mayer-Schönberger and Cukier 2013). Digital data and their mobility are not incidental to the functioning of IT systems and data ecosystems, but are a constituent feature. Just as the circulation of blood is vital to a body, data mobilities are vital to the operations of data assemblages, with constituent elements being bound together and interacting through datafied connections and data sharing. The importance of mobilities to data work and to the value of data has led to a number of ways of conceiving of, and accounting for, the mobility of data.

Data flow is mostly used in the literature as a metaphorical descriptor to denote a liquid-like movement of data from one place to another. Data is said to flow between nodes, moving along a path or through a network (McNally et al. 2012; Tarantino 2019). As McNally et al. (2012) detail, these flows have variable temporalities (e.g., duration, rhythms, synchronisation, prioritisation), with the flow rate altering with context; data might flow freely or be viscous, moving slowly, if at all (Van Schalkwyk et al. 2016). Flows also vary in replicability; that is, the extent to which their topographies (e.g., routes, sequences) are stable over time and with context (McNally et al. 2012). Not all data necessarily flows, forming what Hoeyer et al. (2016) call 'nonflows' (e.g., confidential data with restricted access); or data can 'overflow', which they describe as that which travels unintentionally with data (such as decisions/consequences, potential judgement/stigma, interpretation).

A related concept, which also uses a liquid metaphor, is *data pipeline*, which refers to the ordered sequence through which data are generated and processed, with data being channelled along a workflow in a controlled manner, mediated by various actors and systems, as they move through different stages of production. The pipeline metaphor indicates the inherent movement of data between sites and systems of production towards use and wider circulation (e.g., the field, the lab, a database, analytics software, publication software, an open data repository) (Platin 2018). As the data moves it is transformed and extended through processes of cleaning, wrangling (e.g., formatting, generalisation, standardisation), data fusion and enrichment, review, verification, adding metadata, and analysis. Passage along the pipeline is not simply technical in nature, consisting of the rote implementation of a set of defined techniques and practices, but is socio-technical, mediated by people, cultural norms, learned behaviours, scientific conventions, regulations and laws, and other factors.

For others, the notion of a data pipeline is too static and fixed; instead they contend that data and their production are a *data stream*. Hilgartner and Brandt-Rauf (1994) posit that data and its infrastructure should not be understood as discrete, fixed entities, but are fluid in nature, with the production and movement of data consisting of a stream of processes that are never-ending. For example, the work of producing official statistics never stops, with a regular schedule of monthly, quarterly or annual releases, and the processes used are constantly reviewed and refined (Ruppert and Scheel 2021). Similarly, real-time data compose an endless data stream. The stream metaphor then captures the constant movement of data, both in terms of the data flow never ceasing, and the stream shifting its route and nature (as with water streams that migrate across the landscape, with their position, width and depth varying over time). What such a conceptualisation makes clear is that data mobilities are spatio-temporal in nature and charting a single flow along a static pipeline fails to capture this dynamism.

In contrast to conceptualisations that conceive of data as liquid-like, others contend that data do not have liquid qualities (Borgman 2015; Bates 2018) and do not pass along frictionless pipes or streams, disconnected from the politics and praxes of data assemblages and the infrastructures that connect them (Bates et al. 2016; Pelizza 2016). Pelizza (2016) traces the idea of data flow to information theory, wherein the formal model of communication consists of six elements – information source, a message, a transmitter, a channel, a receiver, a destination – with data (as units of information) flowing in a linear, stable fashion along a channel. Some flows are instigated by the transmitter (e.g., a person emails a dataset to someone else), and some by the receiver (e.g., a person downloads data from an open data site), and some are *data exchanges* between transmitter and receiver wherein data flows both ways (Weltevrede and Jansen 2019). However, as Pelizza (2016) and Bates et al. (2016) contend, the movement of data involves more than opening and closing a valve. Instead, movement is subject to choices, negotiations and decisions by transmitter, receiver and middle agents, and is mediated by a complex set of factors and processes – such as data practices, protocols and regulations, infrastructural supports, and data politics – that can create data frictions that slow and block its progress. Moreover, data mutate as they move, transformed at points along the route. In other words, the movement of data is more akin to a *data journey* than a flow (Bates et al. 2016).

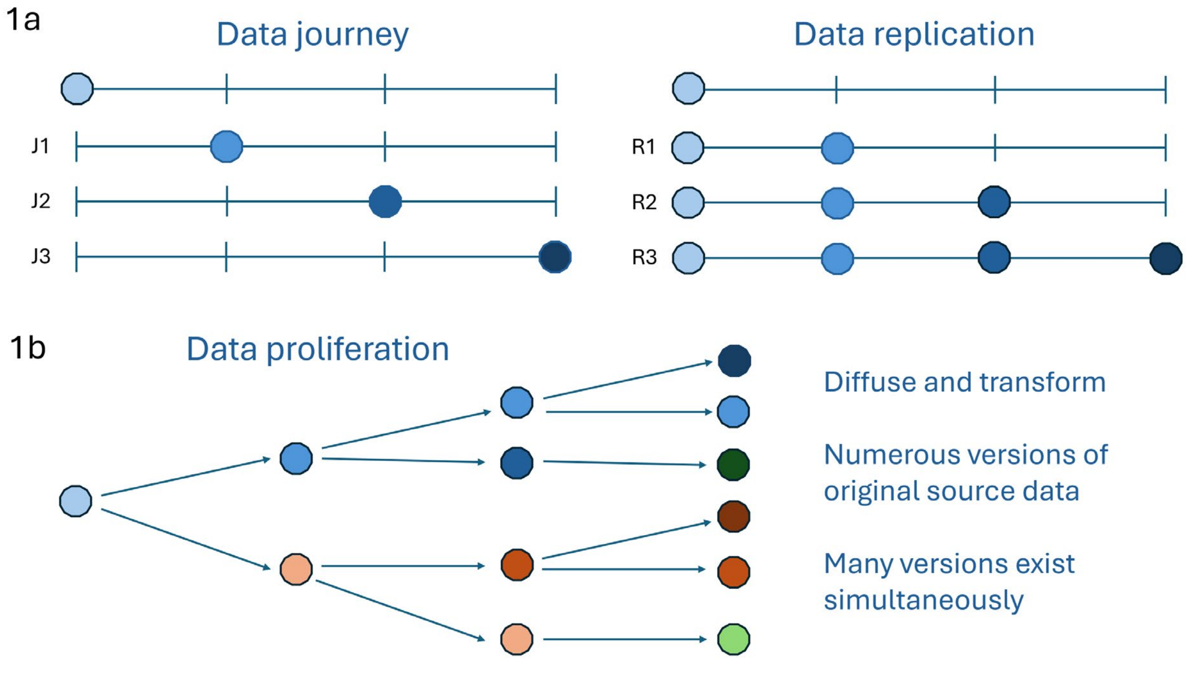
There are two strands of thinking and empirical work concerning the notion of a data journey. The first strand is rooted in Philosophy and Social Studies of Science and can be traced back to a programme of research on the circulation of facts (see Howlett and Morgan 2011) and subsequently developed through a number of further empirical studies (see Leonelli and Tempini 2020). In general, this body of work focuses on the conceptual conditions of data movement, and the data practices, data infrastructures, protocols and governance arrangements that facilitate and mediate data movement and data re-use, mainly with respect to knowledge production (Leonelli 2020). The second strand has its origins in Communications and Media Studies, and in particular in the work of Jo Bates and colleagues (Bates 2018; Bates et al. 2016, 2019). This strand of work conceives of data in more socio-material terms, with its production and movement bound up within the nature and workings of data assemblages (see Kitchin and Lauriault 2014), and shaped by the materialities and contingencies of mediating technologies and their attendant data politics and data power (Bates et al. 2016). Although developed separately, both strands theorise data journeys as contingent, relational and contextual in nature, and reject the notion of data as immutable mobiles. In related work, others have noted that as data journeys along a path and through sites and systems they leave *data traces*; residues in the form of derived data, metadata, and information, the choices and decisions they inform, and the services and products they help create and enact, that present clues as to their nature and their path (Milan 2018, Thylstrup 2022).

The notion of data journeys has been critiqued by White (2017), who instead proposed the concept of *data threads*. Data journeys, in his view, gives the impression that data movement is a largely linear, sequential process along a singular path from origin to destination. Instead, he posits, the path is more complex, with data taking circuitous routes, encountering dead-ends, and looping back on themselves, and the movement can occur with no clear sense of the destination. Moreover, data often does not travel alone, and the paths of different data can be threaded together to forms knots or tapestries, or fray and split apart. White also binds the material and the discursive nature of data together, arguing that data travel as simultaneously socio-cultural and material entities that pass through discursively inscribed material processes, such as inscription and translation.

The discussion so far has largely conceptualised data movement as a singular passage that might detour and loop, but moves along a path, pipeline, stream or thread from origin to destination. The notion of *data circulation* denotes the mobility scaling effects caused by the non-rivalrous and non-excludable nature of data. Rather than consisting of a singular movement along a defined path, the concept of circulation recognises that a single source of data can rapidly multiply, bifurcate, and combine with other data, travelling along multiple paths to many destinations (Beer 2016). This is especially the case once data are published and shared (Borgman 2015). Data, and its derived products (e.g., analytics, information), circulate through archives, data repositories and data markets to potentially thousands or more users. Key datasets on open data sites might be downloaded tens of thousands of times, with key facts or a derived data visualisations moving into news media that are viewed by potentially millions of people. Data circulation then is a multitude of data threads stretching out over time and space, which can be stream-like in their nature (i.e., constantly in motion). Here, the challenge of making sense of data mobility grows from tracking a single flow, journey or a handful of threads, to tracing tens of thousands of movements across a complex network stretching from the local to the global, with this circulation always unfolding and in flux.

All of the data movement and circulation concepts discussed so far are united by an assumption that digital data travels in the same fashion as other entities, good and services. That is, data leaves an origin point and moves along a path between a series of sites and systems to a destination (or many destinations), and it can transform along this route as it is processed. However, we contend that such a conceptualisation is flawed. Data do not move as a flow, journey, or thread. Rather data *replicates*. As evident in the notion of data circulation, data most often travels through copying, with the source data remaining at the origin (e.g., on an open data site) and a copy transferred to a new site (e.g., on the computer of the person who has downloaded a dataset) (Figure 1a). At the new site the data might be transformed through data wrangling or data fusion or enrichment, or be converted into derived data, with these transformed data then transferred to the next site (with or without the original, copied source data). In some cases, the transformation can be significant, with downstream versions being derived tertiary data or the products of analysis (e.g., visualisations). The replication of data can give the impression of a journey if the data on the origin site is deleted after being copied to the new site. Nonetheless, unlike a journey, the data are briefly duplicated. Through replication digital data *proliferates* and *diffuses* across many sites, with local transformations creating multiple versions (Figure 1b). Data mobilities then predominately consist of processes of copying, transformation and erasure, as data are produced, processed, reworked, fused, used, re-used and re-purposed across sites, systems and actors. A consequence of proliferation and transformation is that some digital data are not easily deleted given their multiple copies across organisations and servers. As Thylstrup (2022) notes, the original source data and its residues and derivatives persist even when the data in the original host system are erased, though they might be shorn of supporting materials and context. This is why revenge porn, and other abusive and derogatory content, continues to haunt and circulate even after the original material has been removed (Thylstrup 2022).

Figure 1a and 1b: Data journey, data replication, data proliferation and data transformation



J1, J2, J3 = each separate leg of an overall journey; R1, R2, R3 = each additional step of replication.

Data mobilities then, we contend, are quite different in nature to the movement of water (a flow) or transportation (a journey). Nonetheless, our conception of data mobilities shares some characteristics with data journeys, threads and circulation. In particular, data mobilities are contingent, lively and relational in nature (Bates et al. 2016; Leonelli 2020). How data replicate and proliferate is never fixed or predetermined; rather, it is emergent and evolves in indeterminate ways depending on context, chance, and uncertainty. As Metzler et al. (2023: 3) notes, 'data mobility is always an effect of the relations in which data are entangled.' The pipeline of data production can be planned in detail through processes of articulation and scaffolding – identifying, assembling, scheduling, coordinating and monitoring all the tasks necessary to complete a job (Nadim 2016; Halfmann 2020). However, planning, and its implementation and operation is contingent, shaped by choices and decisions that are informed by personal experiences, negotiations, institutional politics, capacities, policies, and regulations, among other factors, along with the varying application of data practices, glitches, unanticipated interventions, and contextual matters (Kitchin and Lauriault 2014; Loukissas 2019).

The contingency of data mobilities is evident in the effects of data frictions and seams, and the work of maintenance and repair. Data frictions are impediments or blockages that hinder the replication of data across sites, systems and actors (Edwards 2010; Bates 2018). Frictions can be caused by a number of human and technical issues, such as: incompatible data formats, standards, and systems that limit interoperability; mistakes, glitches, and disruptions; resistance and refusal of actors to cooperate and share data; cost, resource and skills capacities; and regulatory and legal limitations. A consequence of data frictions is inefficient or partial replications, error and noise (Lindsay 2017), or an inability to replicate creating stranded data (Kitchin 2022) or broken data (Pink et al. 2018). While data frictions are mostly framed as a hindrance, some exist for good reason; for example, to protect privacy or proprietary knowledge and to ensure data security. Data seams, in contrast, are the points of contact between component parts of a data system, or between data assemblages that enables them to be conjoined or communicate (Vertesi 2014; Inman and Ribes 2018). The replication of data across seams is facilitated by data management and governance mechanisms, such as metadata, data dictionaries, data standards, and transfer protocols that enable interoperability (Millerand and Bowker 2009; Gal and Rubinfeld 2019). In addition to creating seams to smooth data frictions, processes of data maintenance and repair are employed, such as firmware patching and software updates, as well as repairs and replacement of physical system components (Mattern 2018). Without this maintenance and repair the systems underpinning data mobilities will atrophy and eventually stop working altogether.

In addition, our conception of data mobilities similarly recognises that digital data can be replicated in different forms (e.g., lists and tables using a variety of data formats, or assembled as hierarchical or relational databases), using different media, infrastructure and processes (e.g., CDs, DVDs, pen drives, external hard drives, email, file transfer, the cloud, and the use of APIs, ETL [extract, transform, load], scaping, and ingestion processes). Each of these work in varying ways, have different characteristics, and possess and create different data affordances. Likewise, our conceptualisation also acknowledges that data can be assembled into new forms and larger datasets, and data often replicates and diffuses with metadata, other datasets, and other entities (e.g., paradata, documentation, derived information, narratives) (Edwards 2010; Morgan 2011). As such, not only do data replicate, but so do their digital companions. As data are converted into information (e.g., visualisations, narratives) their forms and means of replication alters (e.g., via news and social media, and websites). In the remainder of the paper, we illustrate the constitution and operations of data mobilities through an examination of the Irish planning data ecosystem.

Methodology

To make sense of the planning data ecosystem and its attendant mobilities we undertook an empirical study using a number of related methods. The fieldwork was conducted between June and August 2023. Interviews were undertaken with 29 public sector officials involved in data work at local, regional and national scale within the planning system. For each IT system, at least one interviewee had familiarity with its operation. A number of the interviews were of a walk-through nature, with the participant demonstrating the workflow – data entry, data fusion, data management and data-informed decision-making – related to the use of an IT system. This was supplemented with a close reading of the user manuals for these systems. In addition, a full data audit was undertaken of the core planning application management systems used by local authorities (LAs) (iPlan, APAS and Odyssey), the Building Control Management System (BCMS), and planning.localgov.ie. In addition, interviews were conducted with 4 academics familiar with the Irish planning system. We also examined 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). The walkthrough interviews and data audits were used to construct data dictionaries for each IT system detailing all the variables generated, and to create a chart of the data ecosystem as a whole (see Figure 1). Next, focusing on the development and control aspects of the data ecosystem, we constructed charts of data mobilities for each stage of the process. Each chart, organised along a timeline, detailed the replication processes employed and key data-informed decision points (see Figure 2 for an example). Collectively, this set of charts provided a detailed mapping of data mobilities in the planning and development system from application to completion of properties.

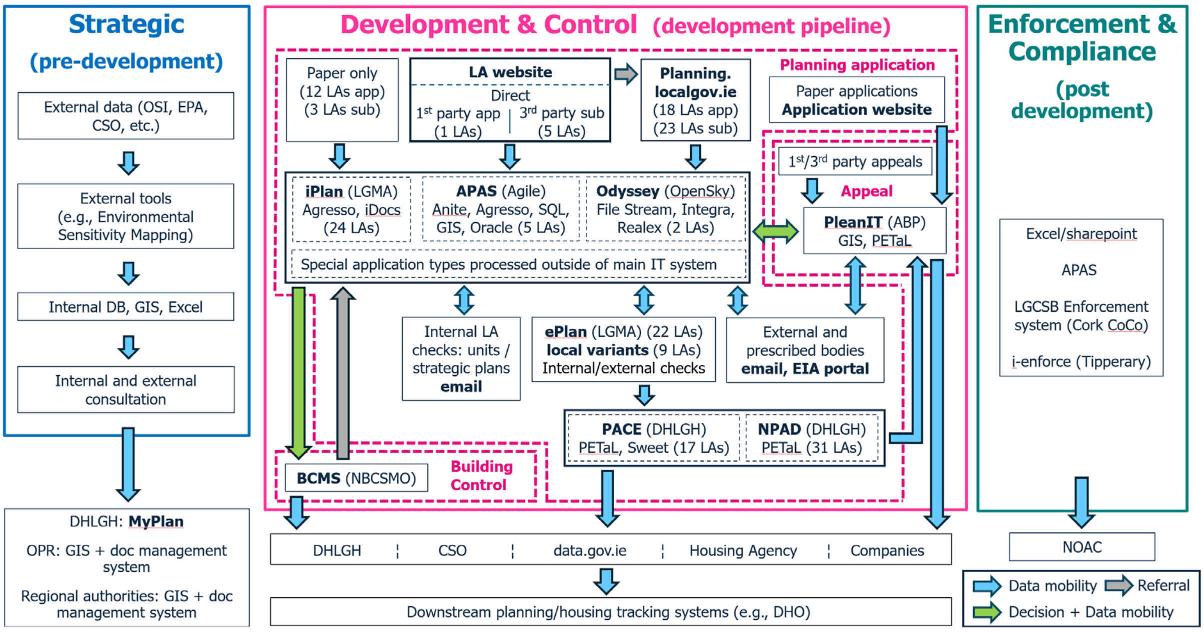
The Irish planning system, data ecosystem and data mobilities

The terrestrial planning system in Ireland, as in other countries, consists of three broad blocks of planning work: (1) strategic planning; (2) development and control; and (3) enforcement and compliance. Strategic planning is composed of forward and pre-development planning work, determining how an area is to be developed in the coming years and ensuring these plans meet policy and regulatory requirements. Development and control consists of assessing planning applications and appeals for new development and, once permission is secured, monitoring compliance with building control through the construction phase. Enforcement and compliance occurs post-development and involves checking whether a development complies with the terms and conditions of a planning permission and any subsequent legal process if these have been breached.

These three blocks of planning work are organised and overseen by a multi-level, tiered system of governance, with planning practice and policy delivered by 31 LAs operating at the county scale, three regional authorities, and a handful of state agencies and government departments at the national scale. LAs are responsible for producing and delivering county development plans and local area plans, and managing development and control. Regional authorities are small and have a limited role in the planning system, constrained to preparing regional spatial and economic strategies and metropolitan area strategic plans, and guiding and coordinating the strategic planning work of LAs. At the national scale: the Local Government Management Agency (LGMA) supports and oversees the work of LAs and provides some core IT systems; the Office of the Planning Regulator (OPR) vets and approves all local, county and regional spatial plans and provides planning advice and guidance; An Bord Pleanála (ABP) is the national planning appeals body and adjudicates on all planning appeals, as well acting as the planning authority for specific types of large-scale development; the National Building Control and Market Surveillance Office (NBCMSO) oversees the construction phase of development and compliance with various building control standards; and the Department of Housing, Local Government and Heritage (DHLGH) formulates planning policy and legislation, and oversees the strategic direction of the sector.

In all three blocks of planning work, all the key stakeholders in the system make extensive use of IT systems to undertake their planning function, and they generate, handle, process, analyse and share substantial volumes of data. Importantly, they do not work independently, but are bound together into a relatively complex data ecosystem that is dependent on the sharing of data between assemblages to function (see Figure 2). These planning IT systems are also interlinked with other IT systems, such as financial systems (related to payment of fees; e.g., Agresso, Integra), file management systems (e.g., iDocs, Oracle), and analysis systems (e.g., GIS) that form key operational components. At points in the assessment process, additional data are sourced from internal and external sources, and data are shared with the public as open data. The critical role of data mobilities in the functioning of data ecosystem is well illustrated by detailing how the development and control role of planning is organised and operates.

Figure 2: The planning data ecosystem in Ireland (as of summer 2023)



Bolded text = Planning IT system or means of capturing data. Each system is discussed in the next section.

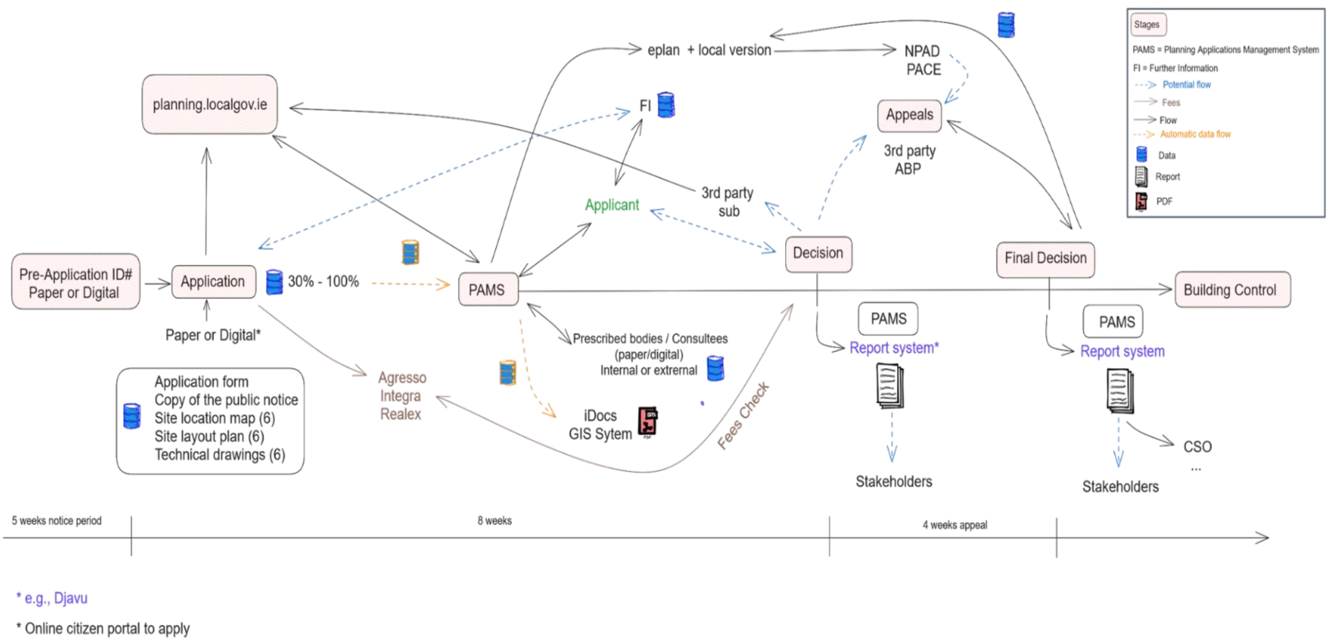
Development and control

The development and control function of the Irish planning system (the pink box in Figure 2) consists of a sequential and time-ordered set of phases and tasks for assessing planning applications, appealing decisions, and monitoring the construction process. Each phase has an associated workflow of tasks and at least one IT system that is used to undertake and manage these tasks. Just as an application passes through phases and systems, so too are data shared between them (as indicated by the blue and green arrows), in order to inform operations and decision-making. As Figure 2 illustrates, this system of development and control is reasonably complex but, in brief, it is organised and operates in the following way.

Planning application

After an initial stage of pre-planning consultation, applicants are required to post a notification about the proposed development in a local newspaper and on a site notice at least two weeks prior to submission, thus sharing some outline data and information with the local community. After a five-week notice period, they then submit a planning application that consists of a copy of the site notice, the site location map, site layout plan, technical drawings, and details of the site development (materials, infrastructure, services). If submitting by paper and post (required for 12 LAs), six sets of documents must be included. On receipt, a LA worker will scan and file the documents and manually enter key data into a planning application management system (PAMS). 19 LAs allow an application to be made digitally, with 18 using a shared services web portal – planning.localgov.ie (administered by the LGMA) – which provides a data entry and upload facility. The uploaded documents are filed in a document management system and the data in the online application form imported into a PAMS. There are three main variants of PAMS used across the 31 local authorities: iPlan (produced by the LGMA) used by 24 LAs; Odyssey (produced by Open Sky) used by 2 LAs; and APAS (produced by Agile) used by 5 LAs. In the case of the five LAs using APAS, each LA has locally configured the system to its own needs, with each instance quite different in terms of design and workflow. These systems are used to manage and track the progress of assessment, including sourcing relevant information, monitoring fee payment, tracking all communications with the applicant and third parties, noting observations and decisions, and ensuring that tasks are undertaken on the proscribed timeline (see Figure 3).

Figure 3: Data mobilities in the planning application phase



After initial processing and assessment, a case manager will seek feedback and an assessment of the application. Requests for expert views and additional information are sent by email to internal LA units (e.g., transportation, environment, and archaeology and heritage departments) and selected external bodies (e.g., OPR, Land Development Agency) and prescribed bodies (e.g., DHLGH, Transport Infrastructure Ireland, Health Service Executive, Environment Protection Agency). Prescribed bodies have to be consulted and are defined by legislation. Checks are also made against the registry of protected structures using an Environmental Impact Assessment (EIA) portal, and with respect to strategic plans (e.g., local area plans, land zoning). The responses to these queries are recorded as data fields in the PAMS.

To enable the general public to assess and make submissions on proposed developments, selected details of the application are replicated on a publicly accessible website and are also made available to walk-in visitors to the LA planning department. 22 LAs use the ePlan system (eplanning.ie, produced by LGMA), with the four Dublin LAs using an Agile platform (planning.agileapplications.ie), the two Galway LAs using a shared site, and the remainder using their own dedicated sites. These systems are purely used for communication, with third party submissions on applications made in three ways depending on LA: 23 LAs using planning.localgov.ie, 5 LAs using their own dedicated website, and 3 LAs via paper submission only. Key information from third party submissions are added to the PAMS and all materials stored in the document file system. In addition to each LA replicating the data on an eplanning website, a selection of data for all LAs are collated within a single site, the NPAD (National Planning Application Database). NPAD is a national, publicly accessible, online GIS mapping tool that displays the location of all planning applications from all 31 LAs, along with summary planning application information (25 fields), and a link to the planning files in ePlan or its variants. The replication of data into NPAD is automated, using a custom ETL process called PETaL (Planning, Extract, Transform and Load) developed by the DHLGH. In addition to NPAD, Tailte Éireann and DHLGH provide LAs (used by 17) with a standardised tool for digitally capturing the site boundaries of planning applications, pulling all these data into a centralised system, PACE (Planning Application Capture Environment).

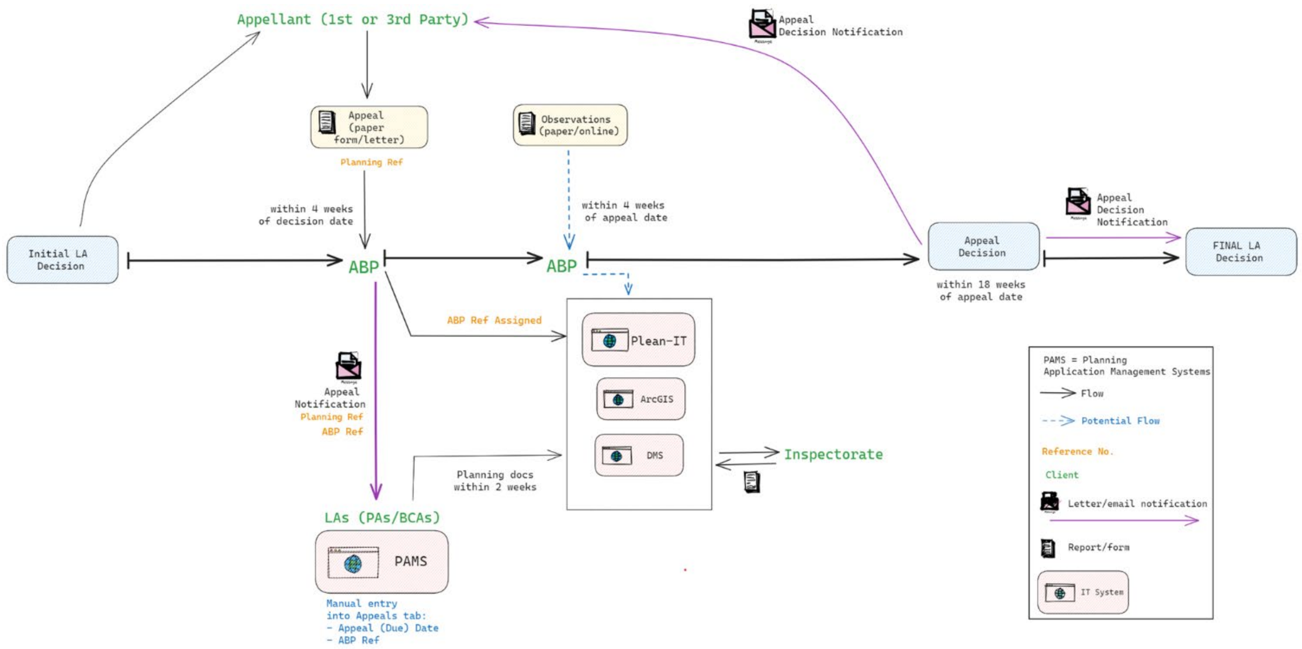
LAs are mandated to process and make a decision on applications within eight weeks of receiving them. If at any point further information is sought from the applicant by the case manager, this can reset the clock and extend the time for making a decision. As a consequence, decision-making can often extend well beyond eight weeks. Once a decision has been made, there are two routes forward to the next stage. Either to appeal the decision (the first party objecting to the denial or by a third party objecting to approval or denial), or to proceed to development and building commencement.

Appeals

An Bord Pleanála (ABP) is the national planning appeals body and has two principal remits. First, to investigate and adjudicate on first or third party appeals to planning permission decisions. Second, to process specific types of planning applications that are sent directly to ABP (e.g., Strategic Infrastructure Developments (SIDs) such as motorways and hospitals, and specific cases regarding the Development (Emergency Electricity Generation) Act 2022). ABP remains a paper-based organisation for legal and statutory reasons (each page is stamped to prove it has been read and an amendment to existing legislation is required to enable the adoption of a digital approach). While it uses IT systems, notably PleanIT (its own planning application management system), a document management system, and a GIS, appeals and applications are made through a paper submission and it has paper-based versions of all files and it prints out all emails to add to its paper-based document filing system. For large applications and appeals the applicant is expected to create a website that contains all pertinent information that ABP can then access and replicate.

On receipt of an appeal details are entered into PleanIT, a unique ID assigned to track progress, and a case inspector assigned to assess the appeal. ABP will at this point also correspond with the responsible LA to notify them of an appeal and request relevant planning documents, replicate information from its ePlan or variant system, and from NPAD via PETaL. Basic data on the planning application, along with technical drawings, maps, and documents, are replicated automatically in PleanIT, while other details are entered manually. A similar process to LA assessments then takes place in which prescribed bodies and the public are invited to provide observations about an application/appeal, with the provided information being recorded in PleanIT. ABP has 18 weeks, organised into a number of timed phases, in which to investigate and make a decision on an appeal (see Figure 4). When ABP makes a decision, it notifies the applicant and appellant, as well as the relevant LA. The appeal decision will be recorded in the LA PAMS, along with the ABP ID number, the date of appeal, appeal decision, and other relevant information.

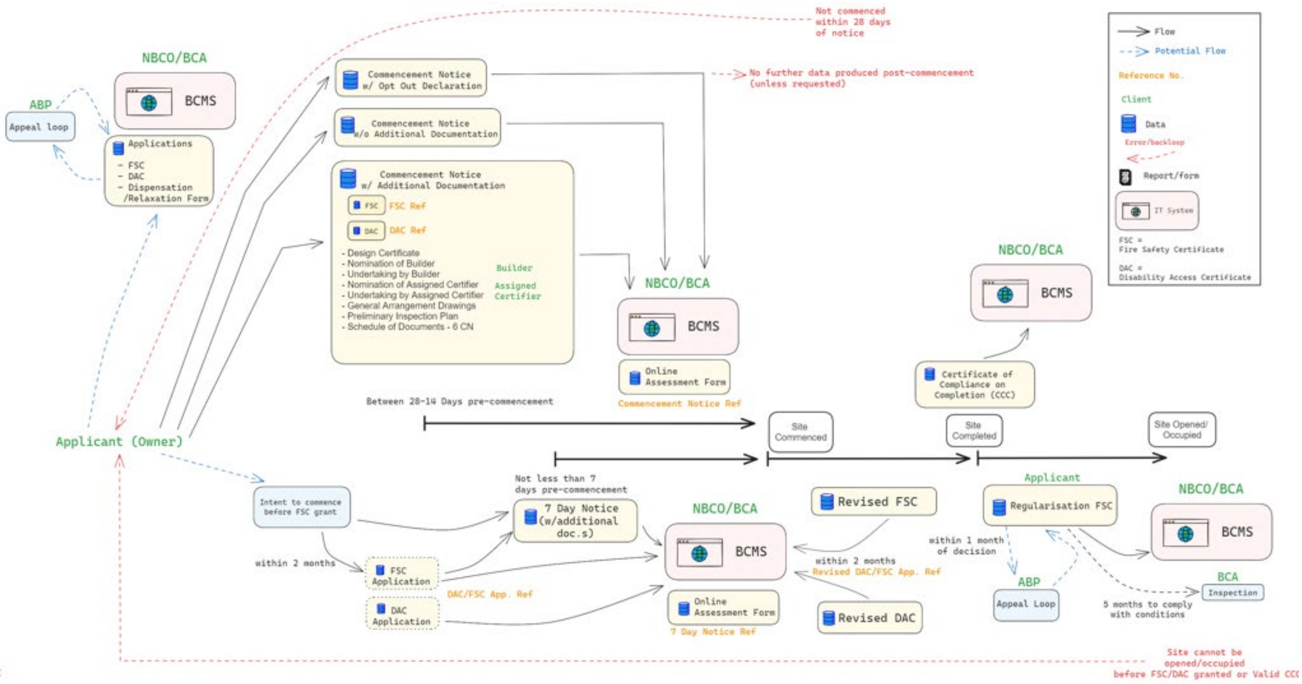
Figure 4: Data mobilities in the appeal process



Building Control

Once planning permission has been granted the applicant is able to move to the development phase. All development in Ireland must be compliant with the regulations and standards set out in Building Control Act 1990, with this being monitored through the BCMS, first introduced in March 2014 as a shared services for all LAs. The data and documentation to be submitted at commencement stage differs depending on the nature of the works. One-off houses and extensions are exempt from tracking using BCMS, though the intent to proceed with an exemption needs to be registered via the submission of an opt-out declaration. For these works, no other documentation is required at this or any subsequent stage of development. All other development must register and upload relevant data at set points along the construction process. These data enter BCMS in the form of a series of notices, certificates, and statutory documents filled out by building owners and assigned certifiers, and validated by the relevant Building Control Authority (i.e., LA). In the simplest case, this occurs at two main points in the development timeline. First, between 28 and 14 days prior to works being commenced (Commencement Notice); second, upon completion of the works (Certificate of Compliance on Completion). While online submissions are encouraged for all types of notice, paper submissions are accepted, with relevant information then entered into the system. For works which meet the criteria as set out in legislation, additional compliance documentation is required, including the nomination of a designer and assigned certifier, a granted Fire Safety Certificate (FSC), a certificate of design, and technical drawings/maps which demonstrate the proposed development's compliance with the regulations (see Figure 5).

Figure 5: Data mobilities in building control



Beyond the planning system

Some of the data generated within the planning application and assessment process, as well as the building control process, are made available as open data and are published as official statistics in an aggregated form. The CSO are responsible for producing official planning permissions statistics and receive details on 14 variables from LAs/ABP monthly which are compiled into aggregated data at LA, regional and national scale. These data are also published on the national open data portal, data.gov.ie. The DHLGH receive regular cuts of development and control data from LAs, and can also request LAs to compile planning data as needed for the monitoring of planning and policy. A limited set of compiled data are shared via the DHLGH open data portal and through NPAD, and also by the Housing Agency. ABP publish a set of data relating to planning appeals and applications in their annual report. The data made available through open data sites, or through specially prepared cuts of the data, are also accessible to sector stakeholders and the public through a number of data visualisation sites that are designed to enable the monitoring and tracking of planning and development. 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/)
- Dublin Housing Observatory (https://airomaps.geohive.ie/dho/)
- Dublin Housing Task Force mapper (https://housinggovie.maps.arcgis.com/apps/View/index.html?appid=3fa56a71ee774f9487d14a9e5336b00c)

In addition to the official, mandated sites for accessing data, unofficial mobilities exist through data scraping by private companies. Construction Information Services Ireland (CIS Ireland) and Building Information Ireland (BII) are two companies that specialise in harvesting, validating, wrangling, and enriching such data to create data products and tools. Each day (CIS) or week (BII) they scrape data and documentation relating to new applications from the ePlan and variants systems, and new registrations in BCMS. They extract relevant data from the documents, link data with other relevant datasets, such as procurement data sourced through e-tender websites, and compile it with respect to each proposed development. The data are also used to compile a number of high-level planning and construction statistics at LA, regional and national level, and across different sectors of activity (e.g., residential, commercial, industrial, etc.).

The nature of data mobilities

As made clear in the previous section, the development and control process in Ireland involves extensive data mobilities to sustain its organisation and operation. Indeed, it is reliant on such mobilities to function as a constituent whole given that data needs to be sourced at various points in the application assessment and data generated or assembled at one phase is required for use in other phases. Here, we consider the nature of these data mobilities, how they work in practice through replication and proliferation, the various data frictions at play, and how data mobilities transition in organisation and form.

Forms and entanglements of replication

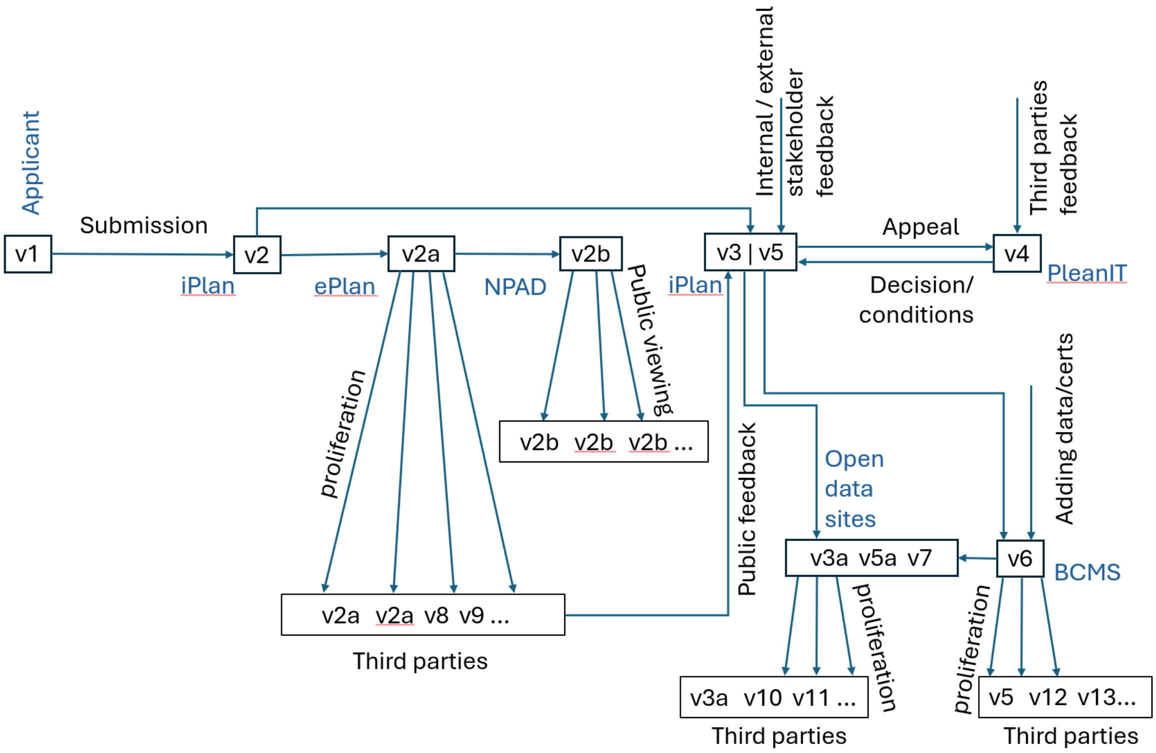
Throughout the development and control process each actor and IT system, at each stage, retains the data shared with other actors and IT systems. For example, the applicant retains a copy of their application at submission (regardless of whether it is a paper or digital submission), and a PAMS retains a copy of the material shared with the ePlan systems. In other words, data are replicated, not moved. As documented, how replication takes place is multiple in form, including post, email, manual data entry through typing and cut-and-paste, manually-directed file transfer, data scraping, and automated APIs and ETL processes. These replication processes are not fixed and invariant in nature, but contingent and contextual. For example, manual data entry practices vary across operatives, with some staff entering only the required fields and others entering as many fields as possible, with most somewhere between; automated ETL processes can be glitchy, failing to work if there are changes to firewall settings, server configurations, or database design, and need constant maintenance and repair. Replication processes are stream-like in that they are continually taking place – every day thousands of replications are occurring between the various systems. As such, Figures 2 to 5 would ideally be animations rather than fixed images to illustrate the on-going dynamism of their data mobilities. In addition, replication is rarely a singular occurrence, but involves companions, such as correspondence and instructions (e.g., email, forms), other data (e.g., metadata, derived data, additional datasets), and information (e.g., documentation, narrative, visualisations), and data are often assembled into, and circulate as, larger datasets (e.g., databases) (Edwards 2010).

The replication of data within the planning system has a relative degree of path dependency (a defined and self-reinforcing sequence of tasks), due to the rule-set for evaluation, scripted practices, established protocols, embedded institutional workflows, and technical configuration of systems (Payne 2014; Poirier 2022). Nonetheless, as Figures 2-5 make clear, the mobility of data within the planning data ecosystem does not consist of a neat, linear sequence of replication, transformation and proliferation. Rather, how development and control works is through entangled sets of replications used to assemble data, populate IT systems, and make decisions. Some of these replications are unidirectional, data replicated from one system into another (e.g., data from PAMS into ePlan). In other cases, replications are bidirectional, with data from one system replicated into another, and other data replicated in response (e.g., application data from PAMS is shared with internal and external units, along with a request for evaluation, with an assessment and associated relevant data returned, which is added to the PAMS). In some cases, the data pass through a number of replications and transformations before looping back to an origin point (e.g., from PAMS to ePlan to NPAD to PleanIT to PAMS). At certain points, there can be loop backs, with a process re-set and re-performed (e.g., at any stage of the planning application phase (Figure 3), a case officer can ask the applicant to revise their application to address identified issues or to supply additional data and documentation, then re-run the tasks already performed). The data mobilities for each planning application are contingent and variable: despite being evaluated against the same criteria, using the same workflows, no one application consists of the same set of replications and processes.

Transformation and proliferation

Once replicated, new data in a system are often transformed in a number of ways to make them amenable to how they are to be managed, analysed and used. For example, data might be subjected to data cleaning to remove personal and sensitive data to comply with data protection regulations and to improve data quality, data wrangling to restructure data and produce derived data, data fusion to merge and assemble more extensive datasets, and data analysis to create information (e.g., visualisations and data narratives) (Kitchin 2022). As a result, multiple new versions of the data are produced that differ from the data retained at the source. In turn, these transformed datasets are themselves replicated. These process of mutation and transformation are illustrated in Figure 6, with v1 representing the original source data. This data is replicated in the iPlan system with personal data removed (v2). A selection of these data are then replicated in ePlan (v2a), and a further refined selection are replicated in NPAD (v2b). The addition of new data replicated via internal/external stakeholder feedback and via third party submissions in response to the ePlan data and documents produce an extended set of data (v3). If a decision is appealed, a selection of the v3 data is shared with PleanIT, which is subsequently extended by third party feedback (v4), with some data fed back to iPlan (v5). A selection of the v3/v5 data is also shared with the open data sites (v3a/v5a) and with the BCMS system (v6). Data mobilities then leads to multiple incarnations of an original source dataset (v1-v6), each of which has different data affordances (what processes and outcomes they enable) (Fjortoft and Lai, 2021), though not all of these incarnations exist simultaneously (e.g., the v1 dataset has been progressively extended to become v3). These incarnations can potentially multiply radically with proliferation.

Figure 6: Data replication, transformation, proliferation



v = data version. a and b indicate a selection of data rather than a transformation.

For much of the development and control phases, replication occurs in defined and stable forms that involve specific actors, limiting proliferation. Proliferation occurs, however, at three points that enable the data to be downloaded and used by anybody globally with an internet connection for re-use and repurposing. When replicated onto ePlan, key data relating to the planning application are open to viewing and download by the general public and to scraping by private companies, who can subsequently transform the data (v8, v9). The same is the case for the data recorded in the BCMS (v12, v13), which is openly accessible. Likewise, proliferation can occur when selected, aggregated data are published on open data sites (v10, v11), from which the data might replicated hundreds or thousands of times. These data, and their transformed incarnations, might then be hosted on other sites, and be used in publications and news media, thus spreading like the diffusion of a mutating disease (in which the host retains the disease and while passing it on, with mutation occurring in the new host before the disease is passed on again). In contrast, the data on NPAD (v2b) is viewable by the general public but not available to download, meaning they cannot easily be replicated and hence are less likely to proliferate or transform. A consequence of transformation and proliferation is that: two sets of analysis supposedly of the same source data might be undertaken on data sets that vary, producing discrepancies in outcomes and interpretation (Kitchin, 2022); there is no end to the data lifecycle as data incarnations persist across many systems and continue to be mobile outside the control of data processor/controller (Thylstrup, 2022).

Data seams and frictions

A number of mechanisms, such as internal and shared data management and governance, standardised forms and templates, standardised workflow within organisations, and the use of APIs and ETL processes, have been put in place throughout the data ecosystem to try to ensure that replication and proliferation occurs as intended and to minimise data frictions. How these processes work in practice to facilitate replication across data seams varies depending on personnel and institutional cultures and priorities, and technical specifications. In the case example, the seams linking data systems can be unstable and break; for example, changes to system or firewall configuration and block the functioning of ETL processes. There are several active data frictions that hamper the smooth operation of replication and proliferation. For example, 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. The lack of standardisation of data form and availability across PAMS hinders the production of harmonious national-scale datasets and official statistics by making it difficult to conjoin data (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). It presently takes a planning officer approximately a week every month to extract the required data from PAMS for submission to the CSO for the compiling of official planning statistics. In turn, the CSO then spends a considerable amount of time cleaning and wrangling these data into the required standardised measurement units, types and classes. There are a number of specialised planning applications handled by local authorities that are not processed through the standard development and control process, and in some LAs these are not administered by IT systems identified in Figure 2. These include applications under Sections 5, 35, 42, 44, 44, 57, 247 and Parts V, VII and XI of the Planning Act. A consequence of their separate administration is that special procedures are required in order for the data recorded to be passed into downstream systems. ETL processes require on-going maintenance and repair. Data frictions expose the contingent and relational nature of data mobilities revealing how they unfold varies; that they are always in the process of taking place.

Data mobility transitions

How the data ecosystem is assembled and how its data mobilities operate is not static, but is subject to alteration as new versions of software, new work processes and new IT systems are put in place, and new regulations, policies and legislation alter the rules and processes of assessing planning applications. In general, changes occur through slight, incremental shifts in workflows, system configurations, and data practices over time, though occasionally a more radical alteration can occur through a critical juncture (e.g., regime change, a crisis, or a 'game-changing' new innovation) (Rast 2012). At the time we were undertaking our fieldwork parts of the development and control process were undergoing digitalisation. Twelve local authorities still only accepted planning applications in a paper form and one LA used its own portal. In the 12 months afterwards, 9 local authorities have transitioned to using the planning.localgov.ie portal and the remaining 4 are due to phase out paper applications by the end of 2024. At the start of 2024 the two Cork LAs transitioned from using Odyssey to APAS as its PAMS. During the first half of 2024 four LAs transitioned to using ePlan for sharing planning application documentation with the public, raising the number of participant LAs to 26. The remaining five LAs are due to transition in the coming months. In addition, four more LAs (all iPlan users) have adopted the use of PACE.

Since we undertook our fieldwork, the LGMA have undertaken a review of the IT support and processes for development and control and are working on a business case for a new national planning system designed to significantly reduce data frictions and improve data harmonisation and quality. The new system, if adopted, would replace iPlan and the various iterations of APAS with a single planning application management system that has a single workflow and data schema. This would mean all 31 LAs using in a consistent manner planning.localgov.ie, the new PAMS, ePlan, PACE and NPAD. Further, the Department of Taoiseach (Prime Minister's office) and the Housing for All inter-departmental group have been undertaking a review of BCMS with a view to significantly improving its data capture processes to ensure more accurate and consistent data and to improve being able to trace a proposed development from planning application to turn-key. These reconfigurations would mean the associated data mobilities would be somewhat reorganised, streamlined and harmonised, with consistent, higher quality data becoming available as official statistics and open data.

Conclusion

The mobility of data is vital to the work of data assemblages and data infrastructures, linking them together into functioning data ecosystems. Data mobilities enables data enrichment and data fusion, new forms of data analysis, enhanced data-informed decision-making, and the creation of data products and services. In the case of planning, data mobilities are fundamental to the operation and delivery of the state's planning function, binding a set of mutually constitutive data assemblages together into a data ecosystem. Data mobilities are the means of sourcing, assembling and sharing evidence in the form of plans, facts, figures, views and opinions, assessing these evidence, and enacting evidence-based decision-making, to ensure that built structures in an environment comply with regulations and policy at all stages of the development pipeline from pre-planning to turn-key.

Up until now, how data are shared and circulate has been theorised as a movement from here to there, where the data flows or journeys along a path, pipeline, stream or thread from one system or place to another. In contrast, we have argued that data does not flow (like water) or journey (like other goods and services), rather it replicates and proliferates. Data are copied, with the source retaining the data and the destination receiving a replica. As illustrated through our case study, at all phases of development and control data are replicated across sites, systems and actors using a variety of processes (post, email, manual data entry through typing and cut-and-paste, manually-directed, file transfer, data scraping, and automated APIs and ETL processes). At selected points the data are made openly available, enabling them to proliferate by being replicated (downloaded), stored and shared by the other actors. Once replicated, the data often transformed through processes of cleaning, wrangling, fusion and analysis. This produces various incarnations of an original source dataset, with multiple versions existing simultaneously. The availability of many data incarnations raises some interesting questions regarding the integrity and validity of data-driven work and the repeatability, replicability and reproducibility of science.

Our analysis suggests three lines of future work is required in relation to data mobilities. First, we need more attention paid to data mobilities as a constitutive feature of data work, recognising that the creation and use of data are never static but inherently consists of mobile practices. In particular, there is a need to conduct further empirical analysis of data mobilities in action and their characteristics and nature through case studies relating to different contexts. Second, we need to more thoroughly ground an analysis of data mobilities within a mobilities perspective. To date, consideration of how data replicates and proliferates has largely developed independently of the mobilities literature, though it is occasionally cited, and mobilities scholars have rarely made data the central focus of their analysis despite significant attention being paid to the mobilities enabled and enacted through digital technologies and platforms (Sheller and Urry, 2006; Stehlin et al. 2020). Third, we need to continue to develop mobilities theory with respect to data. We have sought to advance the conceptualisation of data mobilities, arguing that it is more productive to consider the mobility of data as replication and proliferation, rather than as a flow, journey, chain, thread, or circulation. This conceptualisation needs to be applied, tested and refined with respect to other cases, and no doubt its veracity and utility will be challenged by new ideas and concepts that will further advance our understanding of data mobilities. Given the critical role of data mobilities to the management and governance of society and the functioning of the global economy, such theoretic and empirical work, we believe, is vital.

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 for feedback on an initial draft.

References

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

Bates, J., Goodale, P., Lin, Y. and Andrews, P. (2019) 'Assembling an infrastructure for historic climate data recovery: Data friction in practice', *Journal of Documentation* 75(4): 791-806.

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.

Beer, D. (2016) *Metric Power*. Palgrave, London.

Borgman, C.L. (2015) *Big Data, Little Data, No Data: Scholarship in the Networked World*. MIT Press, Cambridge, MA.

Central Statistics Office (2023) CSO Statement on New Dwelling Completions. https://www.cso.ie/en/csolatestnews/csostatementonnewdwellingcompletions/

Davies, T., Walker, S., Rubinstein, M. and Perini, F. (eds) (2019) *The State of Open Data: Histories and Horizons*. African Minds and International Development Research Centre Cape Town and Ottawa.

Dourish, P. and Gomez Cruz, E. (2018) 'Datafication and data fiction: Narrating data and narrating with data', *Big Data & Society*, 6(2): 1–10.

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

Fjortoft, H. and Lai, M.K. (2021) 'Affordances of narrative and numerical data: A social-semiotic approach to data use', *Studies in Educational Evaluation* 69: 100846.

Flensburg, S. and Lai, S.S. (2023) Follow the data! A strategy for tracing infrastructural power', *Media and Communication* 11(2): 319-329.

Floridi, L. (2010) *Information: A Very Short Guide*. Oxford University Press, Oxford.

Gal, M. and Rubinfeld, D.L. (2019) 'Data standardization', *NYU Law Review*, 94(4): 737–770.

Glouftsios, G. and Leese, M. (2023) Epistemic fusion: Passenger Information Units and the making of international security. *Review of International Studies* (2023), 49(1): 125–142.

Gabrys, J., Pritchard, H. and Barratt, B. (2016) 'Just good enough data: Figuring data citizenships through air pollution sensing and data stories', *Big Data & Society*, 3(2): 1–14.

Halfmann, G. (2020) Material origins of a data journey in Ocean Sciences: How sampling and scaffolding shape data practices. In Leonelli, S. and Tempini, N. (eds) (2020) *Data Journeys in the Sciences*. Springer Open, Cham, Switzerland. pp. 29-44.

Hilgartner, S. and Brandt-Rauf, S.I. (1994) 'Data access, ownership, and control: Toward empirical studies of access practices', *Knowledge: Creation, Diffusion, Utilization* 15(4): 355-372.

Howlett, P. and Morgan, M.S. (2011) *How Well Do Facts Travel?* Cambridge University Press, Cambridge.

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/

Kitchin, R. (2022) *The Data Revolution: A Critical Approach to Big Data, Open Data, and Data Infrastructures*. Second edition. Sage, London.

Leonelli, S. (2020) 'Learning from data journeys', In Leonelli, S. and Tempini, N. (eds) (2020) *Data Journeys in the Sciences*. Springer Open, Cham, Switzerland. pp. 1-26.

Leonelli, S. and Tempini, N. (eds) (2020) *Data Journeys in the Sciences*. Springer Open, Cham, Switzerland.

Lindsay, J.R. (2017) 'Target practice: Counterterrorism and the amplification of data friction', *Science, Technology & Human Values* 42(6): 1061-1099.

Loukissas, Y. (2019) *All Data Are Local: Thinking Critically in a Data-Driven Society*. MIT Press, Cambridge, MA.

Mattern, S. (2018) 'Maintenance and care', *Places Journal*. https://doi.org/10.22269/181120

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

McNally, R., Mackenzie, A., Hui, A. and Tomomitsu, J. (2012) 'Understanding the 'intensive' in 'data intensive research': Data flows in next generation sequencing and environmental networked sensors', *The International Journal of Digital Curation* 7(1): 81-94.

Metzler, I., Ferent, L-M. and Felt, U. (2023) ‘On samples, data, and their mobility in biobanking: How imagined travels help to relate samples and data’, *Big Data & Society* 9(1): 1-13.

Milan, S. (2018) ‘Political agency, digital traces and bottom-up data practices’, *International Journal of Communication*, 12: 507–527.

Millerand, F. and Bowker, G.C. (2009) ‘Metadata standards: Trajectories and enactment in the life of an ontology’, in Star, S.L. and Lampland, M. (eds), Standards and Their Stories: How Quantifying, Classifying and Formalizing Practices Shape Everyday Life. Cornell University Press, Ithaca, NY, pp. 149–167.

Morgan, M.S. (2011) ‘Travelling facts’, in Howlett, P. and Morgan, M.S. (eds) *How Well Do Facts Travel? The Dissemination of Reliable Knowledge*. Cambridge University Press, Cambridge. pp. 3-39.

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

Payne, R. (2014) ‘Frictionless sharing and digital promiscuity’, *Communication and Critical/Cultural Studies*, 11(2): 85-102.

Pelizza, A. (2016) ‘Disciplining change, displacing frictions: Two structural dimensions of digital circulation across land registry database integration’, *Tecnoscienza: Italian Journal of Science and Technology Studies*, 7(2): 35-60.

Pink, S., Ruckenstein, M., Willim, R. and Duque, M. (2018) ‘Broken data: Conceptualising data in an emerging world’, *Big Data & Society*, 5(2): 1–13.

Plantin, J-C. (2019) ‘Data cleaners for pristine datasets: Visibility and invisibility of data processors in social science. *Science, Technology & Human Values* 44(1): 52-73.

Poirier, L. (2022) ‘Accountable data: The politics and pragmatics of disclosure datasets’, *FAccT '22: Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. ACM, New York. pp. 1446–56.

Rast, J. (2012) ‘Why history (still) matters: Time and temporality in urban political analysis’, *Urban Affairs Review*, 48(1): 3–36.

Ruppert, E. and Scheel, S. (2019) 'Introduction: The politics of making up a European people', in Ruppert, E. and Scheel, S. (eds) *Data Practices: Making Up a European People*. Goldsmiths Press, London. pp. 1-28.

Sadowski, J. (2019) 'When data is capital: Datafication, accumulation, and extraction', *Big Data & Society*, 5(1): 1–12.

Sheller, M. and Urry, J. (eds) (2006) *Mobile Technologies of the City*. Routledge, London.

Stehlin, J., Hodson, M. and McMeekin, A. (2020). Platform mobilities and the production of urban space: Toward a typology of platformization trajectories. *Environment and Planning A*, 52(7), 1250-1268.

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

Thatcher, J., O'Sullivan, D. and Mahmoudi, D. (2016) 'Data colonialism through accumulation by dispossession: New metaphors for daily data', *Environment and Planning D*, 34(6): 990–1006.

Thylstrup, N.B. (2022) The ethics and politics of data sets in the age of machine learning: deleting traces and encountering remains. *Media, Culture & Society*, 44(4): 655-671.

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.

Vertesi, J. (2014) 'Seamful spaces: Heterogeneous infrastructures in interaction', *Science, Technology & Human Values*, 39(2): 264-284.

Weltevrede, E. and Jansen, F. (2019) 'Infrastructures of intimate data: Mapping the inbound and outbound data flows of dating apps', *Computational Culture*, 7. http://computationalculture.net/infrastructures-of-intimate-data-mapping-the-inbound-andoutbound-data-flows-of-dating-apps/

White, J.M. (2017) 'Following data threads', in Kitchin, R., Lauriault, T.P. and McArdle, G. (eds), *Data and the City*. Routledge, London, pp. 85–97.