Laying the foundations for a new approach to Census taking in Ireland

John Dunne\textsuperscript{a,}\textsuperscript{*} and Steve MacFeely\textsuperscript{b,c}
\textsuperscript{a}Central Statistics Office, Cork, Ireland
\textsuperscript{b}Centre for Policy Studies, University College Cork, Cork, Ireland
\textsuperscript{c}United Nations Conference on Trade and Development, Palace of Nations, Geneva, Switzerland

1. Introduction

Increased use of shared and linked information can play a significant role in improving public services and government performance by increasing administrative efficiency and improving policy formulation and assessment. This is a worldwide opportunity, not just in Ireland. However, the potential of ‘seamless’, ‘networked’ or ‘joined-up’ government can only be realised if the supporting data are also ‘joined-up’ [28]. Too often these issues are misdiagnosed as information technology problems. They are related of course but separate. If underlying data infrastructures are substandard, then IT solutions will not deliver on their promise. Technology systems can only communicate if databases are properly organised and populated with data codified with universal identifiers.

The current drive nearly everywhere towards ‘Open Data’ and ‘e-Govt’ allied with the ambitions outlined in the 2011 Public Service Reform Plan [10,11] require a National Data Infrastructure. Happily several initiatives have created a unique opportunity to consider the data needs for a modern democratic State, like Ireland, and the impetus to implement such a data infrastructure.

The introduction of property taxes and plans to introduce water charges, post codes and person cards, offer real and exceptional opportunities to make progress towards developing a data infrastructure for Ireland. Conversely, the cost of not taking this opportunity could be large and enduring. A logical, organised and joined-up data infrastructure that is capable of improving public services, supporting an efficient administration, e-government initiatives and evidence informed policy will not happen by accident. It will only come from a deliberate, top-down, system wide initiative that encompasses the full breath of the public sector.

The elegant and simple model outlined by Norbotten in 1960 and subsequently implemented in Norway and other northern European states offers a successful and workable blueprint for planning and organising public sector information. The organisation of public sector information in this manner laid the foundation for modern Census taking as described by Lars Thygesen in a recent interview [39].

In the UK, the Shakespeare Report has recommended a top-down ‘National Data Strategy’ for public service information, noting the requirement for ‘a bold strategy of investment in an infrastructure of data in order to make the UK the world leader in this field, thereby gaining the greatest advantage in this new wave of the digital revolution while also increasing the availability of data to external users’ (2013: p. 27). Similar infrastructure is required in Ireland.

A comprehensive National Data Infrastructure that explicitly facilitates data-sharing and linking naturally make some uncomfortable or nervous of inadvertent disclosure or deliberate misuse. Without question the infrastructure proposed in this paper brings risks but with proper governance it will also realise significant benefits. An open debate as to where Ireland should position itself along the privacy-efficiency/public ser-
vice spectrum, to determine the appropriate balance between personal privacy and administrative efficiency and the needs of a modern state, should be encouraged so that the public fully understand the trade-offs and implications.

This paper is presented in five sections. The first outlines what is meant by ‘public sector’ or ‘administrative’ data and why they are so valuable. The second section proposes how a National Data Infrastructure in Ireland might be organised. Section three details some of the benefits that such a data infrastructure would bring. Section four raises the issue of a privacy-efficiency trade-off. The fifth and final section discusses how we might begin to put a National Data Infrastructure in place.

2. What are administrative or public sector data?

In 1985, Blackwell defined administrative or public sector data as ‘information which is collected as a matter of routine in the day-to-day management or supervision of a scheme or service or revenue collecting system’ [25, p. 78]. Across the civil and wider public service, a huge volume of administrative records are collected, maintained and updated on a regular basis. These data pertain to the wide range of administrative functions in which the State is involved, ranging from individual and enterprise tax payments to social welfare claims or education or farming grants. Typically these administrative records are collected and maintained at the lowest level of aggregation i.e. transaction or interactions by individual taxpayer/applicant/recipient with the state, making these data very rich from an analytical perspective.

While considerable resources are expended by the public service in maintaining these records across the state to ensure they are accurate and up-to-date, with some additional effort, these records could become exponentially more powerful, not only as a tool in helping to design and appraise policy but also as an instrument to assist in implementing policy itself. In effect, administrative data should be viewed, not as an unfortunate burden or cost to the state but as a valuable asset. Perhaps because most official statistics and disseminated administrative data are viewed as a public good, their proper value has not been understood or fully appreciated by compilers or users. Nevertheless, well organised and open public sector information can contribute to democratic transparency, administrative efficiency and economic value [4,28,33,37].

With better organisation and coordination, the potential of public sector information in Ireland can be unlocked. This is the logic of section 31(2)1 of the 1993 Statistics Act, which recognises that statisticians can assist, if involved at a sufficiently early stage, in helping to design efficient information databases from administrative data sources. Unfortunately such consultation has not always taken place with the result that over the past three decades a proliferation of uncoordinated, independent administrative datasets have built-up across the public service, all using different classification, identifiers, definitions and codes. The ‘Statistical Potential of Administrative Records’ studies conducted by CSO between 2003 and 2009 illustrate the lack of uniformity in approach across Government departments and agencies [1–3]. The National Statistics Board has drawn attention to this matter on a number of occasions [2,11] where they have highlighted both the costs of an uncoordinated administrative system for downstream official statistics and for wider efficiencies across the public service.

3. A national data infrastructure

In 1960, Svein Norbotten presented a seminal paper on administrative based statistical systems2 that was to transform not only the administrative and statistical systems in Norway but those of several Nordic and northern European countries [38]. The ideas outlined in this paper are heavily influenced by his thinking and the subsequent transformation of the Nordic public and private information systems. Norbotten viewed all public service data as one coherent system and thus recognised that unique and permanent official identifiers were central to more effective and efficient public administration systems [30]. Implicit to his paradigm was the obvious but often forgotten canon, that data sets are valuable assets. The system that Norbotten envisaged and that Norway implemented has three simple dimensions: object identification; time specification; and attribute observation (see Fig. 1).

The Norwegian example, demonstrates how a logical information system, built on unique, permanent of-

---

1Section 31(2) of the 1993 Statistics Act states: ‘If any public authority proposes to introduce, revise or extend any system for the storage and retrieval of information …it shall consult with the Director General and accept any recommendations that he may reasonably make in relation to the proposal’.

2An abridged version of this paper was published in English in 1966.
Attributes

Fig. 1. Data model dimensions.

Objects

Time

ficial and commonly used identifiers, can permit public sector data to be analysed in a way that facilitates the identification of longitudinal, latitudinal, spatial and relational linkages. These linkages allow movements in time and space to be properly understood. Thus an ‘object’ or unit (e.g. individuals, enterprises or buildings) can be tracked over time as can their ‘attributes’ or characteristics (e.g. spatial location) and their relations to other units (e.g. family, employer, school, car). The importance of permanent or ‘persistent’ official identifiers is central to this approach. This was also the conclusion of the Finch Report [18] which was commissioned by the UK government to address the question of how to improve the transparency and openness of public data to improve public policy and research.

To develop a comparable National Data Infrastructure for Ireland a number of key databases must be developed. Specifically, three comprehensive databases or ‘lists’ are required: (i) a list of all persons in the state (with a unique ID), (ii) a list of all businesses in the state (with a unique ID) and (iii) a list of all locations/buildings in the state (with a unique ID and location co-ordinate). Furthermore, the inter-linkages between these lists are required, so that the various interactions between them can be measured and understood e.g. where does a person live and work (see Fig. 2).

If the Irish Public Sector is to provide efficient administration systems for businesses, persons and property within the state then it must maintain some basic information about those same businesses, persons and properties. To get maximum benefit from such an information system, the infrastructural design is crucial, and must involve the relevant permanent, official unique identifiers associated with each database or list. For those interacting with the state in any service or activity, use of these official identifiers must be mandatory. A move to such a universal design would de-silo the existing plethora of systems. Only with such a system can the interactions and inter-relationships between citizens/business and the state be measured and understood.

4. Why do we need a national data infrastructure?

The benefits from joined-up public service information are many and varied. The Shakespeare Report for the UK lists these as ‘transparency, accountability, improved efficiency, increased data quality, creation of social value, increased participation, increased economic value, improved communication, open innovation, and data linkage’ [35, p.7]. Several government policies in Ireland (noted above) have also highlighted these benefits. The issue that is not addressed typically in these reports is how to join-up data so that efficient and accurate linkage is possible. As noted above, there is a risk that deficits in data infrastructure are misunderstood or incorrectly identified as information technology problems. Technology can only provide solutions if the underlying data are properly structured and organised and populated with universal codes and classifications.

It is worth noting, that the term ‘National Data Infrastructure’, does not in any way limit the scope to national issues and policies. On the contrary, the approach outlined in this paper implicitly incorporates sub-national or regional aspects of public service data such as those held by local authorities. The importance
of properly harnessing administrative data to support regional policy and spatial planning has been clearly highlighted by the Regional Assemblies [35, p. 16] where they note ‘What is needed to facilitate [a national and regional data infrastructure] is not just joined-up thinking, but a common and homogeneous manner for the collection, storage and harmonisation of data. Without this, a very significant opportunity to enhance the evidence-base within Ireland, through the development of a comprehensive data resource, will not be achieved’. A coordinated approach to organising locally and regionally held public service will be essential if key strategic objectives are to be achieved, such as developing and maintaining regional economic models [8].

The recent difficulties surrounding the household charge have clearly illustrated the importance of data infrastructure to implementing policy. The absence of a comprehensive household register meant that compliance could not be targeted and non-compliance could not be followed-up. Several other initiatives outlined within Public Service Reform will also require improved data infrastructure if they are to be properly implemented. If services are to be streamlined and if duplication is to be reduced, then a harmonised, logical data management system across the full breadth of the public service must be put in place. The ambitions of Shared Services, the Public Service Card, the Revenue Business Register, IrelandStat and the recognition that information must be shared to reduce transaction costs all require that data are collected, stored and codified within a harmonised basis to facilitate data matching. The plan itself notes that support structures will be required to deliver the plan – a National Data Infrastructure is one such structure.

Across Public Sector Reform there are explicit implications for data organisation and infrastructure. In particular, the sections dealing with implementation, eGovernment, information sharing, shared services and evaluation – see Section 1.3 Section 2.4 Section 3.5 and Section 9.6. For example, it is not clear how the objective of developing a consistent approach to identification of residents across the public service (see Section 2.3) could be achieved without introducing mandatory use of the PPSN across all public service data registers or lists. Furthermore, it is also clear from the intention of several other initiatives contained within the plan (see Sections 7, 8, 10, 11 and 13) that a National Data Infrastructure would greatly support these plans, even if there is no explicit reference to information or data. For example, Sections 8.3 and 8.4 outline structural and rationalisation initiatives for both Local Government and the VEC. These plans will clearly need to address a range of data infrastructure issues if they are to be successfully implemented.

There are also important implications for public sector data in Public Service Reform and in both the Croke Park [9] and Haddington Road [21] agreements where greater emphasis is now placed on outsourcing service provision. It is vital that the underlying data generated or associated with these services are organised in a coordinated way using the permanent public service identifiers, classifications and codes. Furthermore, it is critical that these data remain in the ownership and control of the State i.e. remain public sector information.

5. The privacy – efficiency trade-off

Both at home and abroad concerns over privacy and protection of individual information are live and increasing. Concerns ranging from the activities of UK Government Communications Headquarters (GCHQ), the National Security Agency (NSA) in the US and ‘information rich’ multinational enterprises have all prompted reaction, ranging from the establishment of the UK civil liberties group ‘Big Brother Watch’ to the Data Protection Commissioner in Ireland conducting a ‘privacy’ audit of Facebook Ireland Ltd. [6]. In Ireland, where privacy is highly valued, this is a particularly culturally sensitive issue [32, p. 211]. Yet, at the same time there is an appetite for increased public sector efficiency and a growing intolerance towards the administrative burdens imposed by Government departments and State agencies on individuals and enterprises. The necessary trade-off between these two positions does not however appear to be well understood. If administrative burdens on businesses and on individuals are to be significantly reduced and if efficiencies across the public service (such as shared services, reduced fraud, improved e-government and targeted probabilistic audits) are to be fully realised, then the implicit, if all too often unstated, consequence is that unit level or micro-data must be exchanged and linked by government

---

3Section 1.6.  
4Sections 2.1, 2.2, 2.3, 2.4, 2.5, 2.6 and 2.10.  
5Sections 3.2, 3.3.  
6Section 9.7.  
7Section 7 – External Service Delivery.
departments and agencies. The OECD review of the Irish public service in 2008 highlighted this issue: ‘Significant future efficiencies are likely through greater sharing of data within and between governments. The sharing of individuals’ personal information, however, does raise privacy protection issues, and the potential trade-offs between increased efficiency and privacy protection, need to be carefully assessed’ [32, p. 199].

There is always the potential for conflict between data protection and public sector reform, particularly when so much of the reform hangs on sharing public sector data. Nevertheless, a practical balance must be struck between privacy and efficiency. Safeguards are needed but an appropriate balance must also be struck between risk-management and risk-avoidance. Three inviolable legislative pillars exist to help strike this balance: the Freedom of Information Act 1997; the Data Protection Acts 1988 and 2003; and the Statistics Act 1993. Any considerations must be cognisant of these important pieces of legislation and should centre on the benefits that can and will be realised versus the risks that data may be abused or mishandled. Public trust is at the heart of this debate; can government be trusted to protect personal information and only use it sensibly? If the answer is yes and Government is to be trusted then the implementation of a National Data Infrastructure will facilitate good governance and allow appropriate audit controls and access logs to be put in place. If the answer is no, then many of the ambitions outlined in current government strategies must be reassessed.

6. A possible way forward

When contemplating major structural reform, it is not always easy to determine where to begin. In Ireland however, a number of initiatives have begun or have been recently announced that create natural start points. The recent announcement by Government that legislation to improve data-sharing and data-governance in the public service [13] is to be drafted is one such start point. In particular, the review of the PPSN and the establishment of a health Identifier (IHI) with a one-to-one link to the PPSN is an ideal opportunity to further enhance and formally recognise a single public-service ‘persons’ register or list; one of the three key pillars envisaged in a National Data Infrastructure.\(^8\)

European legislation [5] now requires that every member state has a ‘point of single contact’ for service providers. The intention here is to simplify the administrative formalities of establishing service activities, so that ‘undertakings or individuals’ can carry out all necessary formalities on-line using a single point of contact. Again, this legislation offers an opportunity to transform the way businesses interact with the State in Ireland. To really achieve the objectives intended by this legislation, a unique business identifier (UBI) is required. Developing a UBI should not be a herculean or costly task; several countries in Europe have a UBI. In Ireland, the Revenue Commissioner ‘customer number’ could be transformed into or used as the basis to create a UBI, making a single point of contact possible, making it easier for businesses to interact with the state and facilitating the data-sharing envisaged by the Department of Public Expenditure and Reform. This approach would allow the development of the second pillar of the National Data Infrastructure.

The public service in Ireland uses a plethora of different spatial and regional boundaries and identifiers [22]. A common spatial identifier is required. While successive Governments have discussed and promised to implement a system of post-codes in Ireland, it has never happened. However, legislation to ‘provide for the establishment, operation and maintenance of a system of postcodes’ was enacted in 2011 [31]. Importantly, this legislation recognises post-codes as ‘infrastructure in non-physical form’ (Part 2, Section 34.1 (b)) and defines a post-code as a geospatial identifier ‘...that identifies the locality of an address and, where appropriate, the geographic location of an address’ (Part 3, Section 6.1). The recent Government announcement [7] that a ‘next generation’ postcode system, where there will be a unique code for each individual address, will be introduced in 2015 is of huge significance. This initiative will play a very important role in the development of household and building or location registers, the third pillar of a National Data Infrastructure.

As outlined, a number of very important initiatives are underway or may soon be underway, that could, contribute towards the implementation of a National Data Infrastructure. Individually these initiatives represent real progress and form vital pieces of infrastructure but on their own they are not enough. An overall architecture is necessary, where other important legal, organisation and governance issues must also be addressed; for example, who should own and maintain public sector registers or lists? Who should be allowed

---

\(^8\)Given the necessity to have a one-to-one concordance between the PPSN and the new Health Identifier, it is not immediately clear, why two separate identifiers are required.
to link public sector information, under what conditions and for what purposes? Furthermore, it is vital that the use of key permanent official identifiers becomes mandatory when interacting with the State.

7. A background afterword on the new Irish Census opportunity

Only very occasionally does an opportunity present itself when, if taken, will reap significant benefits. Such an opportunity, to implement efficiencies in the Irish Public Service, emerged in late 2011 with the publication of ‘Public Service Reform’ [10]. This paper argues that, in order to realise the objectives of ‘seamless’ or ‘joined-up’ government outlined in that plan and supporting strategies, such as e-Govt [11] and open government [12], administrative data in the Irish public sector need to be organised in a planned, rational and coordinated way. Several of the actions outlined in Public Service Reform, such as, the introduction of a public service card, property taxes, water charges and the recent Government commitments to improve data-sharing [13] and to introduce postcodes [7] provide a unique opportunity to reap major benefits from restructuring administrative information systems and developing a coordinated National Data Infrastructure. Significant improvements could be achieved by taking a number of relatively straightforward but fundamental decisions regarding the implementation of new public administration systems or infrastructure. A National Data Infrastructure would provide a platform to transform the way public administration in Ireland operates, facilitating, not only improved public services and administrative efficiency but greatly improving the management information systems necessary to monitor and evaluate programmes.

The architectural design for a National Data Infrastructure must take a whole-of-system perspective to ensure that various strands of reform are integrated so that departmental data systems can easily talk to one another. If designed properly, the resulting data infrastructure would not only contribute to public sector efficiency but would also better support public policy design, implementation and evaluation by allowing public sector data to be shared and linked across Government departments and agencies.

Although it has long been understood that good quality information is necessary to inform policy, the rhetoric and the practice have sometimes diverged [32]. What perhaps has been less well understood is that robust data are also often required to simply implement policy. In Ireland, the difficulties associated with implementing the household charge in 2012 provide a good illustration. Many of the complex, cross-cutting reforms articulated in Public Service Reform [10] require the integration of data across public service organisations (e.g. Shared Services, the Public Service Card, the Revenue Business Register, IrelandStat) if those ambitions are to be realised. In order to share and link data in an efficient manner a greater degree of standardisation is required. A National Data Infrastructure will also require the universal adoption of permanent official identifiers on public administration systems. The OECD alluded to this in its 2008 report ‘Ireland – Towards an Integrated Public Service’ where it noted ‘Mechanisms to improve or streamline systems so that the appropriate data can be better shared within and across the Public Service is something that should also be examined further’ (p. 84).

Of course the OECD was not the first organisation to highlight deficiencies in Ireland’s data infrastructure. The National Economic and Social Council (NESC) have raised concerns about the sufficiency of adequate information for policy making since the 1970’s [23–25]. The importance of being able to re-use and match public sector information has also been highlighted in several government strategies and reviews [11,14,15,19]. These reports have all, in one way or another, highlighted the potential role of public sector data in reducing administrative burden, promoting openness and transparency, supporting better policy information and advice and improving downstream official statistics [17]. These reports however make no reference to how public sector information should be organised in order to achieve these objectives. There is also a risk in these reports that technological and data infrastructure issues have been confused; so more work is needed.

Finally, while the primary driver for a National Data Infrastructure is a more efficient and effective public administration [28], such a National Data Infrastructure lays the foundation for a new approach Census taking based on administrative data sources. First steps towards an administrative enumeration in Ireland using this National Data Infrastructure are described by Dunne, 2014.

References