

CONFERENCE OF EUROPEAN STATISTICIANS

Expert Forum for Producers and Users of Climate Change-Related Statistics

29-30 September 2022, Geneva

IRISH APPROACH TO DATA STEWARDSHIP IN THE ENVIRONMENT DOMAIN

Prepared by Paul Morrin, Central Statistics Office, Ireland

Abstract

NSOs are actively considering how they can increase their influence in their own national contexts in an increasingly competitive data ecosystem. This increased influence should achieve improvements in administrative data collections (indirectly, as improvements for statistical purposes also benefit administrative purposes) and production of more policy relevant statistical products over time. 'Data Stewardship' is the strategic description of these activities. UNECE are currently leading discussions on 'what it is' for our region. Part of this new role for NSOs involves modern data management practices, such as metadata, quality information, data acquisition procedures and modern data science methods, which are common to all data organisations. This is particularly relevant for NSOs as we shift to more secondary data usage, since we no longer have control over the data collection. The second part is more externally focussed based on services to other public service bodies. This includes partnerships, secondments, researcher access, quality/methodology advice, accreditation and bespoke multidimensional analysis in the Irish context. These services are particularly relevant to the environment domain, which has a number of strong, well funded scientific and regulatory bodies (who are also producers of statistics) that CSO can collaborate with based on agreed programmes of work. This paper describes the Irish approach to data stewardship in the environment domain.

1 DATA STEWARDSHIP – WHAT IS IT?

1. Data stewardship is a set of roles and standards for ethical data use and reuse within and across organisations. Data stewardship activities are essential in all organisations with data holdings. Data stewardship can be organised across a range of business areas with defined roles, or can also be centralised into a particular function. An example of the list of activities that can be considered in this function from a draft UNECE report on data stewardship is given in Table 1 below.

Table 1. **Examples of data stewardship activities**

External/internal networking	Data management/technical	Ethics and legal
Supporting acquisition of new data and using new data sources	Maintaining metadata and classifications codes	Link to legal services
Liaising with external providers for metadata and life cycle	Data life cycle management	Management/compliance of Data Governance Model

Liaising internally for quality	Monitoring ‘once only’ principle	Developing rules for acquisition of data
Making internal and external data available based on FAIR ¹ principles	Overseeing quality and security processes	

Source: UNECE Task Force on data stewardship

2. In addition to these organisation level tasks, there also needs to be a National ‘Data Steward’ who co-ordinates data skills, services and standards in the public sector so that all Public Sector Bodies (PSBs) can benefit from enhancements in the ‘data ecosystem’. The Central Statistics Office (CSO) is willing to take on this role, in collaboration with others, and our ‘catalogue of services’ for PSBs varies depending on the context in each policy domain. CSO prepared a paper for the 2020 Conference of European Statisticians setting out the data stewardship services provided during the early stages of the COVID crisis². This paper similarly sets out the services we can potentially provide in the environment domain in the coming years.

2 THE IRISH CONTEXT FOR THE ENVIRONMENT DOMAIN

3. Ireland has a highly centralised statistical system, with CSO responsible for 82% of European statistics. The environment domain is different, since statistics for many European legislative requirements are produced by the scientific agencies who have regulatory responsibility. Of the 14 ‘Other National Authorities’ reporting European statistics, 3 are environmental and 3 more are in related fields such as sustainable agriculture, food production and fisheries. In addition, while many PSB’s produce statistics as a by-product of their administrative functions, statistical production methodologies are much more complex for environmental statistics, such as the European Energy Balance Sheets. These more complex statistics require multi-agency collaboration, including CSO involvement. Data sharing between public agencies is quite limited in Ireland and is usually for defined operational purposes rather than e.g. for multidimensional policy analysis. Finally, much of the data required for environmental statistics is held within the public sector in Ireland, such as climate sensor and utility meter data.

3 SERVICES THAT ARE PARTICULARLY RELEVANT IN THE ENVIRONMENT DOMAIN

3.1 PARTNERSHIPS

4. Partnerships will be essential for National Statistical Offices (NSOs) in future if we are to meet the ever-growing demand for insight. Finding the right partner of course is not always easy and our focus tends to be on those who share our core values based around our reputation for professionalism, integrity and objectivity. CSO has engaged in a number of strategic partnerships in recent times with agencies that have complementary technical skills. A successful partnership has been with the national mapping authority in Ireland, Ordnance Survey Ireland (OSi). We are now deepening our partnerships with the two key environmental agencies, the Environmental Protection Agency (EPA) and the Sustainable Energy Authority of Ireland (SEAI). We have long standing liaison groups with the agencies. We also have mutual respect for our roles as producers of Official statistics, while acknowledging that CSO has responsibility in the European Statistical System for co-ordinating all European statistics (described under Section 3.5). The areas where we feel CSO can particularly add value in the environment domain are described in this paper.

¹ Guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets.

² https://unece.org/fileadmin/DAM/stats/documents/ece/ces/2020/ECE_CES_2020_29-2007757E.pdf

3.2 SECONDMENTS

5. CSO professional statistical staff have been seconded to PSBs on request for more than 30 years with the seconding PSB paying for the cost of the service. The seconded statistical staff provide a broad range of statistical services and are also a liaison for services being provided by CSO within the seconding organisation. It is not sufficient to pass down guidelines, classifications, standards etc. from the CSO within the framework of the Statistics Act, 1993 (sections 10 and 31 refer) and expect that they will be adopted seamlessly without direct support from the CSO. Many PSBs do not have access to the skill sets to the scale necessary to engage with such standards or indeed to interrogate data holdings in a meaningful way for their own business purposes. It is now apparent to PSBs that the CSO are seconding out these skill sets or services and the main challenge we have in rolling out this service has been our ability to meet the demand. We have a number of secondments in environmental agencies, who complement the scientific, statistical and analytical skillsets already present and link the agencies to CSO stewardship supports. The key difference to other sectors is that some of the resources are organised as 'programmes of work' funded by the agencies. Staff working on the programmes are based in both organisations with loan arrangements. This reflects the interlinked nature of the work involved in topics such as energy balances and the circular economy, where CSO performs key tasks in the work programmes without overall responsibility for delivery. There is also much more potential for secondments in to CSO than for other domains given the complementary skillsets mentioned earlier.

3.3 BESPOKE AND REGULAR MULTIDIMENSIONAL ANALYSIS BASED ON DATA LINKAGE

6. Gerry Brady from CSO has prepared a paper for the Forum in Session 3 which covers this service in some detail. CSO has long standing access to identifiable data from across the administrative system which facilitates matching based on identifiers for people and properties where available. Where identifiers are not available, matching based on deterministic methods is possible with high rates where names, sex and dates of birth in particular are recorded accurately. The service CSO provides across all domains, in addition to our Official statistics, includes 'pathfinder projects' and short term project resources via the 'Data Science and Statistical Support Unit'. Pathfinder projects are policy-relevant research projects that CSO develops in collaboration with policy makers on specific policy questions. The aim of these projects is to deliver insight and value to the policy makers while at the same time highlighting the value and critical importance of the National Data Infrastructure (i.e. the collection and storage of identifiers related to persons, locations and businesses by PSBs in their administrative systems). These projects have increased the profile of CSO as a provider of analytical services through the provision of aggregated data to policy makers and demonstrated an agility and flexibility in how we do our business. The 'Data Science and Statistical Support Unit' can provide additional short-term analytical capacity in Departments who are either facing 'crisis' situations or are developing complex multi-dimensional policy proposals.

3.4 RESEARCHER ACCESS

7. Recently there have been demands from the policy side, health and education sectors primarily, for CSO to link and integrate relevant data sources, and to subsequently provide research access to these newly created Research Microdata Files (RMFs) within the framework of the Statistics Act, 1993. This would effectively involve CSO creating a 'Research Data Hub' for policy and research purposes in some key domains. Enhanced governance around these RMFs involving our data providers has been implemented in the context of COVID data access for researchers, which will inform approaches in other domains. We are already providing this service for research funded by environmental agencies for a small number of projects. There is potential for expansion of the service, especially in the context of European Data Spaces for the Green Deal and the rapid expansion of third level research into environmental issues.

The potential for 'Big data' in the environment domain (such as smart meters, climate sensors, etc) will require upgraded CSO researcher infrastructure. Also, synthetic data based on these 'real' datasets may be sufficient for some researchers which will require more data science skillsets in CSO via new hires and training.

3.5 ACCREDITATION

8. As producers of Official statistics, the Irish Statistical System Code of Practice (ISSCoP) accreditation is particularly valued by environmental agencies. The EPA has already attained this standard for some of its statistical compilations, and SEAI are well advanced in the process. There are obligations under the Statistics Act, 1993 for CSO to coordinate Official statistics; under the amended 223/2009 CSO will be the port of call by the European Commission where issues arise irrespective of whether the data is compiled by CSO or an Other National Authority (ONA) such as EPA or SEAI. The accreditation process has been a positive development for all participants including CSO staff, and has been useful for informing PSB's of our stewardship services as well as for bedding in seconded staff into their new organisations.

3.6 QUALITY AND METHODOLOGY ADVICE

9. The CSO has a formal Quality Management Framework (QMF) consistent with international standards which covers a broad range of activities:

- Establishment of standards, based on international best practice, for all relevant parts of the data life cycle;
- Quality Review System;
- Data and process governance initiatives so that our data assets are stored, governed and managed in a consistent, standardised manner across all business areas;
- Data Management and governance support tools;
- Classifications Management;
- Metadata Management;
- Survey documentation;
- Process Mapping;
- Process maintenance;
- Process Metrics and Indicators;

10. The CSO Quality Management and Methodology Divisions develop the standards and support their implementations across the business areas of the CSO.

11. While much of the work to date has been focussed within the CSO, Government Departments and their agencies are now approaching the CSO to support them on quality assurance related issues. We will use the QMF as the basis for these engagements and over time the vision will be to ensure a consistent approach across the system based on the QMF. As with accreditation, the two main environmental agencies have been particularly keen to investigate and adopt the QMF given their role as producers of Official statistics. We have also supplied methodological advice to a number of PSBs on an informal basis, though many statistics compiled in the public sector are directly reported from administrative collections. There appears to be much more scope for this service in the environmental domain as there are extremely complex methodologies underpinning e.g. energy balances.

* * * * *