# An Approach to Support Data Stewardship through the Implementation of Data Standards across the Irish Data System

**Laura Reddin**

*Central Statistics Office, Ireland*

## Abstract

Data stewardship ensures the ethical and responsible creation, collection, management and use of data, which will improve the quality of data. In addition, public trust will be strengthened across the Irish State. Robust, well-structured data puts decision makers in the best position to make well informed decisions, thereby serving our society. The development of data standards supports data being collected and published in a consistent, harmonised manner. Harmonised data standards provide guidance on best practices around collecting and producing quality data. It also ensures the availability of quality information across the Irish civil and public service.

In order to advance its strategic priority of advancing data stewardship, the Central Statistics Office (CSO), Ireland has begun work on developing and implementing a broad suite of data standards.

This paper will describe the steps taken by the CSO in order to successfully develop and implement the use of data standards across the Irish Statistical System. This involves the research around individual concepts and consultation with key stakeholders in the early development stages of each standard concept being developed. The paper will continue to outline the promotion of established data standards, along with the governance and monitoring of the use of the standards.

It will also outline the approach undertaken by the CSO in the implementation of data standards and it will conclude with detailing the obstacles encountered and the proposed plan to overcome expected barriers.

**Keywords:** Harmonisation, Interoperability, Comparability

## 1. Introduction

Data stewardship is embedded within the core values of the Central Statistics Office. Improvements to data quality continue to be a corporate priority by providing standards on how to implement consistent and coherent data classifications. One of the strategic aims outlined in the CSO Statement of Strategy is to “Increase the use of secondary data sources”. (Central Statistics Office, 2022). Further analysis of the strategic aim highlights the need to improve the quality of administrative data sources by developing standards, rules and guidelines to govern the classifications, storage and management of data across the Irish Statistical System. The

Irish data ecosystem comprises of data holdings which offer significant insights for decision makers and the people of Ireland if they are structured and organised according to agreed standards and classifications. Without structure, data is locked in silos resulting in lost opportunities for efficiencies and insight. The data steward role, as set out in sections 10(2), 10(3) and 31 of the Statistics Act, 1993, will influence the improvement of the quality of administrative data sources to allow for more efficient management of data for the creation of official statistics. (Houses of the Oireachtas, 1993).

Additionally, the Central Statistics Office's long-term strategy, CSO Vision 2030 envisages that the CSO will provide statistical leadership nationally and advocate for the management of data as a strategic asset, driving the development of the Irish Statistical System and championing the expansion of the National Data Infrastructure (NDI). (CSO Vision 2030, 2020). A strong data ecosystem will be crucial to reap the benefits of establishing the CSO's data steward role consistent with its legal mandate.

## 2. Maturing Data Ecosystem

The Irish System of official statistics and the Central Statistics Office have been going through a period of transformational change. This transformational change leads to our evolving Irish data environment. Our society is changing rapidly with policy makers requiring increasingly complex insights to be drawn from data-based evidence. To deal with these demands, a higher reliance has been placed on the acquisition of administrative data through improved collaboration with government departments and other agencies. This has resulted in an organisational shift to a "secondary data first" approach supplemented by existing primary data collection. The pronounced shift towards reliance on administrative data was made possible by a coordinated approach to the use of unique identifiers, and much greater coordination within the system has left a stronger data ecosystem in place and improved the foundations for future developments of statistics and data-driven policy analysis.

As a result of sustained efforts led by the Central Statistics Office and involving holders of administrative data throughout the public sector, a National Data Infrastructure is progressively being embedded across the Irish Statistical System. Administrative datasets are increasingly being equipped with unique identifiers for persons, enterprises and locations, along with standards and classifications being harmonised. Subsequently the potential for data linkage and concomitant improvements in statistics and data-driven policy analysis is continually heightened. Increasingly, the Central Statistics Office is seen as the leader in the efforts to improve our public data ecosystem, thereby effectively becoming a data steward.

The rapidly expanding and evolving landscape of data production and sharing is forcing National Statistical Offices (NSOs) to reconsider their position and role in the national data ecosystem. The increasingly popular term used to denote this changing role is Data Stewardship. As the transformation is multifaceted, there are different interpretations of how the NSO role should change to meet its data stewardship responsibilities. Data stewardship encompasses functions and competencies to enable access to, and re-use of, data for public benefit in a systematic, sustainable, and responsible way. Data stewardship is operating *in service of* rather than *in control of* the data ecosystem. (CES Task Force, 2023)

This rapid change is set to continue, with the desirable delivery empowering the Central Statistics Office to deal with future challenges.

## 3. Challenges

Change often involves risk, and unfortunately the establishment of data stewardship takes no exception to risk holding or risk management. One challenge in the current data environment is that many of our data holdings exist in isolation. With the power of hindsight, it is evident that the lack of structure has guided a siloed approach. National governments are taking steps to reduce policy siloes and make public sector data strategies, projects and initiatives more coherent. (OECD, 2019). There are multiple ways of representing the same information, even within a department. Therefore, it is not difficult to determine that across the various government departments and agencies of the Irish Statistical System the problem of multiplicity multiplies. National Statistical Offices face not only the challenge, but the opportunity, to place themselves as key players in the data ecosystem given their well-grounded expertise in data management, access and sharing practices in the public sector, as well as their key role as producers of statistics and indicators.

Data stewardship is an extension of the role that the Central Statistics Office has played in coordinating the Irish Statistical System, where the key objective is to ensure good coordination between statistical agencies. The growing role has the potential to go deeper into the statistical system to ensure interoperability, and to go wider, to the whole data ecosystem.

All data can be powerful but only if it is planned and structured from the outset. Thus, the need for data standards to drive a better data infrastructure allowing more consistency to data. (Eolas Magazine, 2023)

## 4. Data Standards Solution

Through consideration of the challenges being encountered, the solution reached involves providing information around general standards while also providing people with the tools to align to the recommended standards. Data standards provide a practical approach to tackling the challenges resulting from our maturing data environment. The use of standards across the Irish Statistical System ensures that common definitions and processes are used within and between organisations, helping to remove the barriers to collaboration, thereby making data more comparable, transparent, and insightful.

The creation of data standards ensures better quality information across the Irish Statistical System. Access to more consistent and comparable data is important for policymakers and service providers in allowing them to better understand how policies and services are meeting the needs of users. Examining your own data provides a certain level of insight but this is greatly enhanced if you can compare your data with other data sources that have applied the same data standards.

Developing alternative data sources, such as public sector administrative data holdings, is critical to addressing the growing demand for data along with a coordinated approach to data linkage and integration. To facilitate the linking of data across administrative systems, the National Data Infrastructure (NDI) was established in 2017 which involves the collection and storage of three key identifiers on public sector data holdings. The identifiers include a person identifier being the Personal Public Service Number (PPSN), a location identifier being EIRCODE, and an enterprise identifier being the Unique Business Identifier (UBI). (NDI Champions, 2023)

Building on from the three established identifiers of the NDI brings us to the development of common data standards to improve the usability and quality of data on administrative data across government organisations, which is referred to as the NDI+. The successful rollout of common data standards across the Irish Statistical System will allow data to be structured to maximise comparability, leading to more reliable outcomes. The current situation lends to Public Sector Bodies measuring and defining concepts inconsistently, making it difficult to gain a complete understanding of the data across the Irish Statistical System, thus limiting potential comparability.

Each data standard is developed with the mindset that full standardisation may not be feasible initially, but a harmonised approach embraced. Harmonisation offers a realistic approach to improving data comparability as it recognises that there is a balance to be struck between standardisation and operational demands. Statistics do not always have to be fully

standardised to be harmonised, but development and use of harmonised standards allows data producers to develop outputs that align reasonably with information from other data producers thereby increasing the usefulness of their statistics for research, analysis, decisions, and policymaking.

## 5. Elements of a CSO Data Standard

A consistent layout has been adopted for each individual data standard consisting of the following basic elements:

- General description of concept being measured
- Concept definition
- Guidance for data collection (inputs)
    - Standard questions
    - Standard response options with coding structure
- Standard reference classification, displaying different levels of detail whereby level 1, summary level, is required at a minimum
- Descriptors of response options
- Visual example of harmonising to the standard
- Links to related standard concepts
- Examples of usage
- Link to relevant API, allowing easy re-use of standards

Look up files and coders will be attached to standards, where appropriate, acting as a shared coding facility to support standard implementation. Coders are currently being developed by the CSO for relevant standard concepts, such as Religion and Occupation.

## 6. Example of Harmonising to the Standard

Diving into an example, firstly, the challenges of integrating data from across the system will be apparent, but more importantly the CSO harmonised data standards solution will be evident. Looking at a common variable collected by many departments, such as marital status, the current barriers to data comparability are immediately visible. Each department works independently to specify what they need to measure, how it is defined, the internal coding structure, and the publishing of results. The siloed approach creates difficulty bringing together data from the different sources. To delve into the collection of marital status, each department ask the question at data collection using slightly different wording and phrasing. Moreover, different format of response options are provided, all with inconsistent definitions. As a result,

the analyse of marital status across the Irish Statistical System becomes very limited. The limitation can be overcome quite easily through the implementation of common data standards. Standardising the question and the response options ensures that the data is comparable at micro-data level, while harmonising the level at which outputs are produced allows a degree of comparability, even where the question and response options have not been fully standardised.

## 7. Development Process

The process for drafting each individual data standard follows a journey consisting of several steps. The necessary components include engagement, research, consultation, promotion and prioritisation. The initial development stages involved presenting at the formal statistical conference, to firstly deliver the need for the change, followed by addressing any concerns or doubts. Another important element within the engagement step included the partaking in both the Census Advisory Group consultations and the Equality Data Strategy meetings. In addition, the NDI Champions Group meetings granted a further support in engaging with external departments. Internally, initial engagement took shape via the hosting of an online presentation and answer session casted to all internal stakeholders. An important element of the engagement step was the presence and feedback of key decision makers in the early stages of development.

Continuing the journey to the next step of the process, research, it was very important to carry out research of international offices and their approach to the successful implementation of standards. The research involved exploring the published work around the available standards from numerous National Statistical Offices such as, Office of National Statistics, Australian Bureau of Statistics, Statistics New Zealand and Statistics Canada. This is an invaluable part of the process to ensure successful implementation.

Moving onto the next step, consultation, it was necessary to consult with key stakeholders in the early stages in order to be able to drive the change for effective implementation of data standards. Consultation included working with the Questionnaire Design Unit, who are the experts around the initial social standard concepts.

Progressing onto the promotion step, this will involve the promotion of approved standards both internally in the CSO and externally across the Irish Statistical System. The implementing and sustaining change stages involve working with government departments and offices in order to empower them to support the change by utilising the data standards in their data

collections. At this stage the establishment and use of data standards will be anchored into the culture of the data ecosystem.

An important component of the overall process is prioritisation. Following consideration and analysis the CSO decided to focus on demographic variables initially, with a plan to expand further while also being responsive to feedback or requests. All standards are built around current regulations in place, mainly being EuroStat's IESS regulation which aims to ensure that social statistics are produced in a more harmonised and coordinated manner across Europe.

While a process is in place, it is also important to be mindful that challenges exist preventing the effortless implementation of data standards. However, the benefits of harmonisation continue to outweigh the obstacles by offering more consistent and comparable data across the Irish data ecosystem.

## 8. Initial Implementation

Promotion of data standards has commenced, including a recent data standards workshop at the Administrative Data Centre seminar in November 2023. This was followed by a presentation to the NDI Champions Group at the December 2023 meeting. All feedback was gathered and taken on board prior to the standard sign off.

A decision was made to initially focus on standardising the demographic block of variables in order to tackle the inconsistencies across the social concepts. In addition, the focus around social variables aligned favourably with the ongoing work of developing a national equality data strategy.

Currently two standards are at publishing stages, being:

- Marital Status
- Full-time or Part-time Employment Status

Both standards will be launched in the coming weeks through the CSO website.

Several more standards have been drafted and are awaiting approval, including:

- Household Relationship
- Main Activity Status
- Religion
- Country
- Sex
- Gender
- Sexual Orientation

Finally, standards next in line to be drafted include:

- County
- Local Authority
- NUTS regions
- NACE
- Ethnicity

## 9. Future Steps

The main focus going forward is around the promotion of the approved data standards. All approved data standards will be published and made available to all through a CSO portal. The portal is currently under construction with the hope of being accessible in Q2 2024. Intermediately, data standards will be made available via the CSO website.

Forecasting the successful promotion and implementation of the data standards, monitoring the use of the data standards will become crucial. The NDI Champions Group developed a dashboard to summarise the coverage of record-level unique personal identifiers and unique location identifiers for key data holdings of government departments and agencies under their aegis. It was developed to highlight datasets where coverage is low and is used to communicate the key data holdings and coverage metrics to stakeholders. It is envisaged through an expansion of the NDI Dashboard that it will be possible to monitor the use of standards across the Irish Statistical System.

Future steps will also involve engagement with external departments to support standard implementation. Support and guidance will be offered to assist in the adaptation of harmonised data standards. Training sessions and workshops will be provided to share research and best practices to develop knowledge and expertise around harmonising to the approved data standards.

## 10. Conclusion

The use of the data standards will improve transparency and comparability of data across the Irish Statistical System. Data standards also support data linkage and integration which increases the ability to provide evidence and insight to support policy. Not only will this lead to advanced innovation, while saving time and money. Implementation of data standards allows national data assets to be structured to realise their full potential for the benefit of serving the Irish society. Standardisation of questions and response options for concepts ensures that maximum value can be extracted from the data, inevitably producing better outcomes.

At present, the early stages of the rollout of data standards have commenced, transitioning from the development stages to the implementing and sustaining change stages.

## References (12pt. bold)


Central Statistics Office. (2022). Statement of Strategy Annual Report 2022.

CES Task Force. (2023) Data Stewardship and the Role of National Statistical Offices in the New Data Ecosystem.

CSO Vision 2030. https://www.cso.ie/en/media/csoie/releasespublications/documents/corporatepublications/CSO_VISION_2030.pdf

Eolas Magazine. (2023). National Data Infrastructure: Data as a strategic asset to the public service. Eolas, Issue 61.

Houses of the Oireachtas (1993), "The Statistics Act, 1993". Statistics Act, 1993 (irishstatutebook.ie)

NDI Champions. (2023). National Data Infrastructure Annual Report.

OECD (2019), The Path to Becoming a Data-Driven Public Sector, OECD Digital Government Studies, OECD Publishing, Paris, https://doi.org/10.1787/059814a7-en.