

Revenue Statistics

Quality Statement

September 2017

Statistics & Economic Research Branch

Table of Contents

Table of Contents	1
Overview	2
Introduction	2
Other Documents.....	2
General Information	2
Organisational Unit Responsible	2
Periodicity.....	2
Users	2
Statistical Concepts and Methods	3
Units	3
Data Source	3
Methods.....	4
Classifications.....	4
Quality Principles	5
Relevance	5
Accuracy and Reliability	5
Timeliness and Punctuality	5
Coherence and Comparability	6
Accessibility and Clarity	6
Production of Statistics	7
Overview	7
Specify Needs.....	7
Design Phase.....	7
Build Phase	8
Collect, Process & Analyse Phase	8
Disseminate Phase	9
Evaluation Phase.....	10
Confidentiality	11
Confidentiality	11
Appendix 1 – Levels 1 & 2 of the Generic Statistical Business Process Model ..	12

Overview

Introduction

This document provides information on how Revenue addresses key principles, as identified in the Irish Statistical Code of Practice, in relation to Revenue's published statistics, including:

- Commitment to quality
- Timeliness and punctuality
- Accessibility and clarity

Additionally, this statement outlines the procedures in place to provide assurance on the quality of Revenue statistical processes and outputs. Revenue statistics refer to statistical products which are published on the Revenue Statistics webpage and which are referenced in the 'Revenue Statistics – Publication Calendar'.

Other Documents

It should be noted that this document details the quality activities and procedures in place at an organisational level. The 'Method' statements for individual Revenue statistical products detail what specific measures, process and activities are in place in each case.

Aside from the individual 'Method' statements, the following documents are available on Revenue's Statistical webpage:

- Publication and Dissemination Policy
- Statistical Disclosure Control
- Publication Calendar
- Timeliness Monitor

General Information

For the most part Revenue Statistics are based on administrative data extracted from Revenue Systems. Revenue Statistics which are not based on Revenue administrative data are identified in the relevant individual 'Method' statement.

Organisational Unit Responsible

Revenue Statistics are prepared and published by the Statistics & Economic Research Branch, Corporate Affairs & Customs Division, Dublin Castle.

Queries should be addressed to: statistics@revenue.ie

Periodicity

Revenue Statistics will be released in accordance with the frequency and publication dates outlined in the 'Revenue Statistics – Publication Calendar'.

Users

Revenue Statistics are used by a diverse range of stakeholders, including:

- Members of the public
- Media
- Political parties, public representatives and the Houses of the Oireachtas
- Public bodies including the Department of Finance and Department of Public Expenditure & Reform
- Researchers & research bodies

Statistical Concepts and Methods

Units

The units presented in each statistical product are dependent on the measure or attribute being published. Unless otherwise stated all monetary values will be shown in Euros or € million Euros.

Income Tax Statistics are generally presented on the basis of taxpayers/tax units. It should be noted that married persons or civil partners who have elected or who have been deemed to have elected for joint assessment are counted as one tax unit.

Data Source

Revenue Statistics are based on data extracted from Revenue systems. These data will typically have been collated on foot of taxpayer registration, taxpayer/agent statutory returns/payments or returns to Revenue from 3rd parties (e.g., employers).

A number of Revenue data sources are used in Revenue's statistical products and the 'Method' statement of the individual statistical products details the particular data sources for that product.

Examples of Revenue data sources include:

Taxpayer Registrations Data: These data are based on administrative data gathered when a taxpayer or business registers for tax. This registration data may be subsequently updated.

Taxpayer Returns Data: These data are based on the returns taxpayers (or 3rd parties) are obliged to submit to Revenue on a regular basis (e.g., Form 11/12 for self-employed/PAYE workers for Income Tax, CT1 from businesses for Corporation Tax, P35s from employers for Income Tax). These data sources may be accessed through a range of methods including dashboards and dedicated database query tools.

Income Tax IDS Data: An 'Income Distribution Statistics' (IDS) dataset is constructed each year integrating a number of different tax record sources including taxpayer registration data, self-assessed taxpayer returns and returns by employers on behalf of employees (P35, P60, P45 and Form 11). The dataset is only extracted when specific filing thresholds have been achieved (96% of PAYE returns and 83% of self-assessment returns). Aside from taxpayer and returns data, additional fields are included which contain the combination of other fields that are used for calculating a taxpayers liability.

The production of the IDS dataset is dependent upon validation of the income tax returns which takes place once the returns have been filed for that year. As the returns are not due to be filed up to 11 months after the end of the tax year, there is generally a 15 to 18 month offset between the tax year end and the availability of the IDS dataset for that year. Thus the 2015 IDS dataset only became available in mid 2017.

The IDS dataset is a comprehensive source of information on income distribution in Ireland. It represents the taxpayer population which comprises a large part of the total population. Other cohorts of the population, who are not captured in the IDS dataset,

excluding non-compliance, would include persons in receipt of welfare payments (who are not obliged to file a return) and full-time students.

Corporation Tax IDS Data: The Corporation Tax 'Income Distribution Statistics' (IDS) dataset provides the data for reporting and statistical analyses on Corporation Tax returns for Revenue. The Corporation Tax (CT) assessing system in the Integrated Taxation Processing (ITP) system is responsible for processing the Form CT1 from CT customers.

Payment Data: These data are based on administrative data gathered in respect of payments made to Revenue by taxpayers in statutory tax returns. Payments are credited by Revenue's Collector General's Division to the taxpayer and is associated with the appropriate period and taxhead, on the basis of the payment returns information. This process is reviewed by the Revenue Accountant General.

Methods

The 'Method' statement of the individual Revenue statistical products detail how each set of Statistics are developed.

Classifications

Revenue Statistical releases utilise a number of classifications, each of these are cited in the 'Method' statement of the individual Revenue statistical products. There are a number of classifications which are employed across a range of products, with a common classification being the NACE Code, which is a basis for statistical classifications of economic activities. Revenue uses NACE Rev 2.

Quality Principles

The following section outlines Revenue's approach to delivering on its commitment to ISSCoP quality principles in the production of statistics. In the case of the production of individual statistical products, the approach detailed below may be further refined in the individual 'Method' statements.

Relevance

Most of Revenue's statistical output is well established and has been in place to meet a range of public interest needs, including:

- To meet information requirements from Revenue data identified on foot of Parliamentary Questions and Freedom of Information requests.
- To service a range of information requests from within Revenue, from other government bodies, and from researchers/academics.
- To provide information to the media and other public requests for information as received through our dedicated statistics email address – statistics@revenue.ie.
- To disseminate information on important and significant Revenue activities such as tax collection.
- To provide oversight and transparency of the cost or yield to the Exchequer of various schemes/reliefs/allowances and credits.

Revenue's statistical outputs were designed to achieve the best fit between available resources, user needs, data availability and data confidentiality.

Accuracy and Reliability

There is a high reputational interest for Revenue in assuring the accuracy and reliability of its statistics. Due to the nature and scope of our role, our statistical outputs form the basis of much budgetary and policy analysis. As almost all Revenue statistical outputs are based on administrative data the errors usually associated with sampling are not present. However, Revenue undertakes a number of validations on both source material and final outputs to ensure the delivery and dissemination of accurate statistical outputs, for example there is an extensive validation process embedded within the production of both IT and CT IDS datasets.

Revenue's approach to revisions is outlined in our 'Revenue Statistics – Publication and Dissemination Policy'. In brief where the format, content or methodology of a Revenue statistical dataset is altered or revised, notice of such revisions will be communicated to customers on release. Where statistical figures are revised, the updated data will be uploaded to the relevant platform at the earliest opportunity. Notice of any revisions will be communicated through Revenue's statistics bulletin which is distributed via email through the statistics@revenue.ie email account.

Timeliness and Punctuality

Revenue Statistics will be released in accordance with the frequency and publication dates outlined in the 'Revenue Statistics – Publication Calendar' in accordance with our 'Publication and Dissemination Policy'. Our performance against these targets will be tracked in our 'Timeliness Monitor'.

Coherence and Comparability

To the greatest extent possible, Revenue statistics will be consistent by using the same unit granularity, time intervals and unit of measure – for example pre-Euro currency values have all been converted to Euro. Where statistics are produced on an annual basis, previous years figures will be published side-by-side to facilitate comparison. Where it is necessary to deviate from this approach, or where there is a change in methodology, the revision protocol detailed in Revenue’s ‘Publication and Dissemination Policy’ will apply.

Revenue’s published datasets typically contain data from previous years and the data for each year are formatted and broken-down in the same units and at the same level to permit year-on-year comparison allowing for annual comparability to identify trends etc.

Accessibility and Clarity

Most Revenue datasets are Open Data compliant. As such, in compliance with the Open Data Technical Framework, the dataset is published in both PDF and CSV formats, and under a CC-BY creative commons attribution license. The data are published in the statistical area of the Revenue website, and on the Open Data portal of the Department of Public Expenditure & Reform, using clear English descriptions and detailing the metadata tags and other key information in plain English.

This approach provides clarity regarding the nature of the dataset and ensures that the statistics are easily found through keyword searches.

Production of Statistics

Overview

The following section describes and defines Revenue's processes for the production of statistics. These are outlined using the Generic Statistical Business Process Model (GSBPM) – see Appendix 1 for Model overview. As envisaged within the framework, the GSBPM has been applied flexibly to describe the statistical production processes in the context of the overarching Revenue environment. The production model described below is a generic one describing in broad terms the processes involved, each product may have a separate production path. Variations from the generic model along with product specific comments will be set out in each product's 'Method' statement.

Specify Needs

The requirement to deliver new statistical products or to modify existing outputs may be triggered by a by a number of events. These include:

- Introduction of a new scheme/ tax relief/ tax credit in a Budget or Finance Act, for example the 'Help to Buy' (HTB) scheme in Budget 2017.
- Direct requests received either from within Revenue or from outside (e.g., general queries, PQs, FOI requests), such as other government bodies, research academic organisations and/or the public in respect of Revenue activity or the administration of schemes/relief/credits etc.
- Internal evaluation of the existing statistical offerings in the context of the operating environment and changes to available data.

This list is an indication of the drivers that trigger Revenue to undertake a review of the published data and the need to expand on this offering. Revenue will establish in more detail how this need might be met, which may include a process of consultation with the stakeholders. There is significant overlap between this phase and the Design phase. The requirement, for example, may require the tracking of the complete lifecycle of a scheme. If there are comparable national or international statistics being produced, the delivery of these solutions may also be considered at this stage.

A fundamental aspect of this process is establishing whether current data sources could meet user requirements, and the conditions under which they would be available, including any restrictions on their use. If the data are available to meet the identified need two further issues must be considered before proceeding further:

- i. Whether there is any organisational, regulatory or disclosure impediment to developing the new/ amended output, for example issues of taxpayer confidentiality.
- ii. Whether the work involved is justified in terms of the resources required to publish this material versus the benefit of doing so.

Before proceeding further, approval to production of the new/amended statistics will be sought from the Head of the Statistics & Economic Research Branch or senior management, as appropriate.

Design Phase

In brief, the purpose of this phase is to determine how Revenue staff will interact with the necessary data retrieval systems and who will be responsible for the production and assimilation of such material. Building on the work of the previous phase, the detailed

design of the statistical outputs will be completed. This may involve designing a prototype to both illustrate the format of the product and to provide stakeholders with a sample of what may be delivered. Outputs should be designed to follow existing standards wherever possible. Where appropriate, Open Data metadata will be prepared at the Design phase.

There are a number of potential data sources available to Revenue and the most appropriate collection method(s) and instrument(s) will be determined at this stage. This decision will be addressed for each product in their specific 'Method' statement and will usually refer to one of the data sources detailed previously in this document. Responsibility for the production of the statistics should also be finalised. The mechanism for data extraction and compilation will be dependant on the product itself, but may include the use of dashboard reports, specific database query tools for use on our IDS datasets or other databases. In other cases the data collection will be detailed in the 'Method' statement.

At this stage any statistical processing or analysis should be considered and finalised, for example if there is imputation for any under coverage or use of an Average Marginal Rate (AMR) to calculate tax costs. Validation and disclosure controls will also be designed at this stage. Finally the dissemination and process should be developed. Issues such as the frequency of issue, platform(s), new product vs extension of existing product and whether the product will be publish in Open Data format will be determined.

Build Phase

Many of the Build phase activities may be undertaken in parallel with Design phase activities. In the course of this phase the various collection components are built and tested. These will include queries, reports, report templates. Where appropriate the template report should also be translated at this stage. All documentation should also be drafted at this stage including:

- i. The Method statement
- ii. Documentation about the process components, including technical documentation and user manuals
- iii. Any explanatory or training materials for business area users on how to operate the process

Collect, Process & Analyse Phase

There is a degree of overlap between these three phases. Where issues are identified at the Analysis phase the production process may revisit the Collection stage. It is therefore worthwhile considering the three elements together.

The Collection phase for most Revenue statistics involves the extraction of data from a range of administrative databases and loading them into the appropriate environment for further processing. In the case of some datasets (e.g., where Revenue's Collector General's furnish Revenue payments data), the data provider will be contacted seeking extraction and delivery of the data from their systems. The collection phase is dependant on the source of the data and will be referenced in each Method statement. In some cases this involves validation, for example the extraction and collation of both the Income Tax and Corporation Tax IDS datasets involves extensive validation processes.

Where Revenue statistical outputs are produced from a number of sources, the data are then integrated. For example, the 'Cost of Tax Expenditures' dataset contains the cost to the Exchequer of various credits, reliefs, allowance and schemes for various taxheads and

as such data which are derived from multiple sources is integrated at this phase. At this stage aggregate data and totals are also produced. In the case of a number of Revenue statistics it is necessary at this stage to 'grow' the data to take account of incomplete or late returns. While the appropriate tax rate is used in many cases to create new data units to display, for example, the cost of a measure/relief, in a number of instances it may be necessary to use a derived Average Marginal Rate (AMR) that calculates the average tax rate paid by relevant taxpayer constituencies. Each dataset's 'Method' statement will specify where any of these or other processes are applied.

In the analyses phase statistical outputs are produced, examined in detail and made ready for dissemination. It includes preparing statistical content (including commentary, technical notes, etc.), and ensuring outputs are 'fit for purpose' prior to dissemination to stakeholders. First, initial draft outputs are produced. These outputs are then validated through a range of activities including:

- Checking that the population coverage and response rates are in line with expectations.
- Comparing the statistics with previous cycles.
- Checking that the associated metadata is in line with expectations.
- Cross-checking the statistics against other relevant data (both internal and external).
- Investigating inconsistencies in the data.
- Performing macro editing.
- Validating the statistics against expectations and business intelligence.

At this phase confidentiality and disclosure concerns are addressed. Confidentiality is central to Revenue's relationship with taxpayers. However, we also recognise the value of providing a broad range of statistical insights into tax administration. [Statistical Disclosure Control](#) protocols have been developed and this document outlines our approach to ensuring the balance between the need to inform society and safeguarding taxpayer confidentiality are achieved in the production of Revenue Statistics. The draft statistical output will be tested against these controls.

Disseminate Phase

This phase manages the release of Revenue Statistics to customers. It includes all activities associated with assembling and releasing a range of static and dynamic products via a range of channels. These activities support customers to access and use the statistical data released by Revenue.

Revenue statistics are scheduled for release based on the availability of tax returns data. However, for some products the releases are synchronised with calendar events such as a Budget, Annual Report publication, etc. The production and release schedule of statistical products is detailed in the 'Revenue Statistics – Publication Calendar'. The format and dissemination process for each product will be detailed in the 'Method' statements.

Once a proposed publication of data has reached this phase and has been deemed 'fit for purpose' the data are now formatted and the metadata finalised. The format of the data will have been finalised during the design stage but can range from one or all of '.pdf', '.csv', '.doc' and '.xls'. Where distribution statistics are hosted on the Central Statistics Office (CSO) Statbank the relevant files are prepared. In accordance with Revenue calendar, the datasets are then released and metadata updated on the Revenue.ie

website, OpenData Portal and/or CSO Statbank platforms. Dataset releases are publicised through a Revenue Statistics Bulletin.

Evaluation Phase

Newly produced statistical outputs are evaluated at each production cycle to establish whether they should continue to be produced, whether they should continue to be produced in this manner, or whether they should be changed on the basis of changing user needs or the availability of the data. In the case of well-established outputs this evaluation may not be formally implemented at each cycle – evaluation is dependent on a number of factors including feedback from users, system metrics (e.g., the number of downloads/accesses of that data on the Revenue website) and suggestions from within Revenue. The evaluation is undertaken by the relevant Business area within Revenue along with the producers of the statistics for publication.

Confidentiality

Confidentiality

Revenue amasses a significant amount of data, much of it of a very sensitive nature. As in other public bodies, the collection, management, use, access to and sharing of Revenue data is currently guided and regulated by a body of existing laws and policies on the Re-use of Public Sector Information, Government Copyright and Freedom of Information. The standard governance measures, controls and roles are in place in Revenue to ensure adherence with these obligations.

In relation to the technical and physical provisions to protect taxpayers, Revenue is subject to audit from the Data Protection Commissioner. All associated potential threats are identified managed locally by specialist IT and Corporate Services areas. In some cases these risks are on the Corporate Risk Register and are subject to ongoing assessment and evaluation through Revenue's Risk Management Committee.

At the heart of Revenue's relationship with taxpayers is our commitment to protecting the confidentiality of the information shared with us. Section 851A of the Taxes Consolidation Act 1997 formalises taxpayer confidentiality and provides a specific tax-related provision to reassure taxpayers that personal and commercial information disclosed to Revenue is protected against unauthorised disclosure to other persons. Staff obligations regarding access to official data and regarding Revenue's data security and confidentiality policies are set out in a broad range of documents and extend to include:

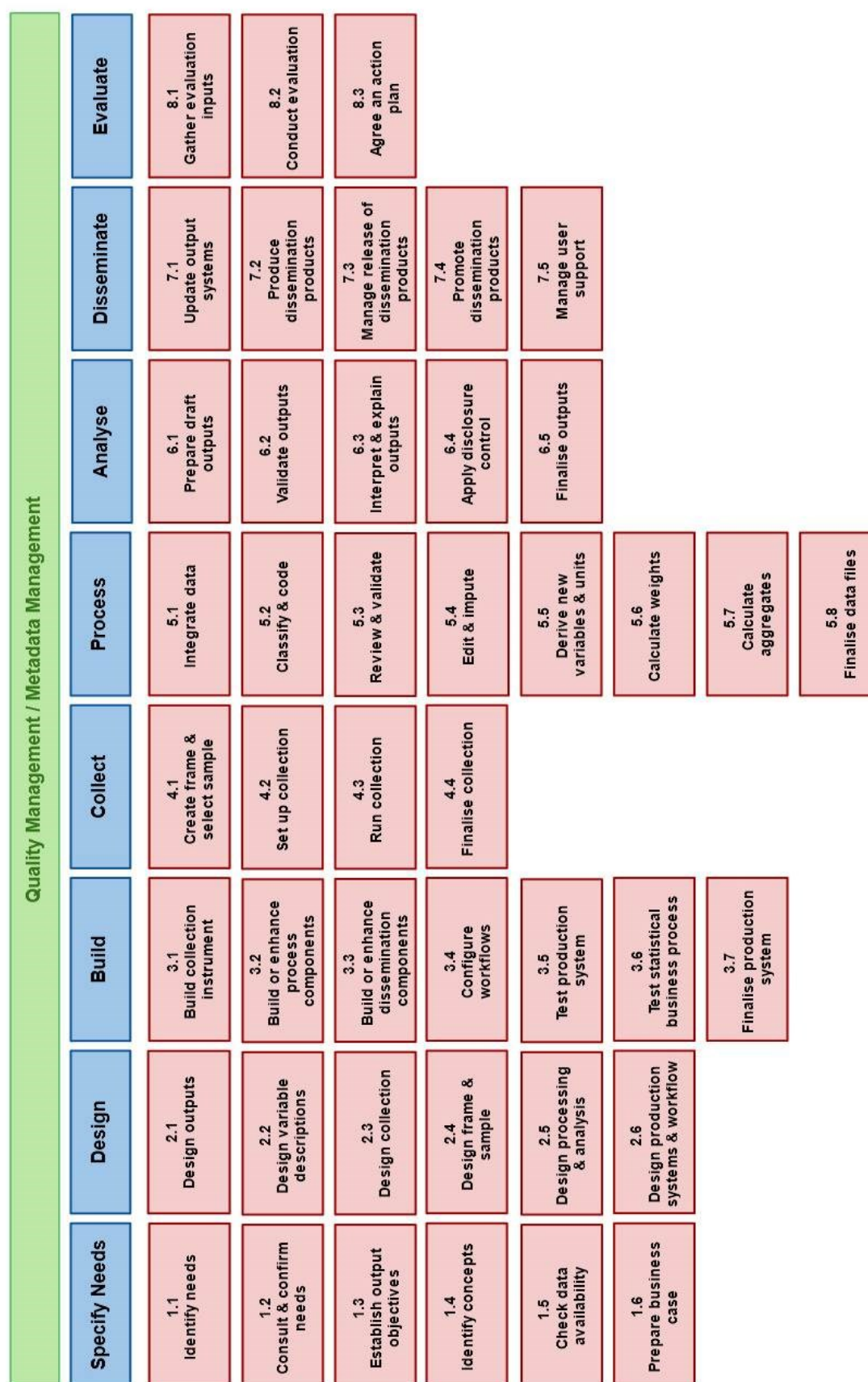
- Revenue Code of Ethics
- Civil Service Code of Standards and Behaviour
- Revenue security and confidentiality policy
- Data security and confidentiality policy
- Unauthorised access to Revenue data
- Inappropriate data access and disclosure

These guidelines and regulations are conveyed to all new staff members and reminders are issued to all staff from time to time via circular or other means. All Revenue staff sign the Official Secrets Act. Non-compliance with data confidentiality obligations are treated very seriously and breaches are subject to appropriate action under the disciplinary code.

External parties who require access to Revenue data are required to make similar confidentiality undertakings. Revenue has extensive data exchanges in place with the CSO. These are mostly transfers from Revenue to the CSO. Approved researchers can apply for access, under controlled conditions to work on Revenue's data through the CSO's Research Microdata File (RMF) process.

Within Revenue [Statistical Disclosure Control](#) protocols have been developed. This document outlines Revenue's approach to ensuring the balance between the need to inform society and safeguarding taxpayer confidentiality are achieved in the production of Revenue Statistics. It outlines the various rules and checks that are applied to statistical outputs to evaluate whether a release is appropriate in the context of a disclosure risk. It also details the various anonymization techniques that may be applied in cases where such a risk presents.

Appendix 1 – Levels 1 & 2 of the Generic Statistical Business Process Model



Source : <http://www1.unece.org/stat/platform/display/GSBPM/GSBPM+v5.0>