



**Oifig Luachála**  
Valuation Office

# Data Audit

## Valuation Office

Prepared by  Derilinx



# Table of Contents

3	Executive Summary
4	Introduction
4	The Valuation Office & Data
5	The Need for a Data Audit
6	Data Audit Process & Results
7	Audit Process
8	Identifying Datasets
8	Data Demand, Impact & Change Readiness
8	Data Publishing
9	Data Quality Assessment
11	Summary
12	Analysis of Results
15	Recommendations



# Executive Summary

Derilinx carried out a data audit with the Valuation Office in September and October 2022, which consisted of a full eight departments.

These eight departments were: Accounts, Corporate Services, Procurement, Public Office and Archives, Human Resources, Management Support, ICT and the Valuation Administration Department.

For the data audit, representatives from each of the divisions filled out a data audit template. That is, within a spreadsheet, they listed each data asset and details on the data assets including their type, publication status and data quality. These templates were then returned to Derilinx for analysis.

The aim of this analysis was to answer what is the current state of data within the Valuation Office, and what should be done next to improve the state of this data.

## The main findings of the data audits were:

- + The Valuation Office holds 48 data assets
- + The Valuation Office currently publish (or publish partial versions of) 12 of their 48 data assets externally
- + Overall, Valuation Office data tended to be around mid-level of quality
- + Most divisions marked little demand for the data assets they hold
- + 14 Additional datasets have been identified as being suitable for publishing open data

## From these results and the Valuation Office's strategic aim of better use and access of data, the main recommendations to improve the data in the Valuation Office are:

1. Create a well organised, accessible data catalogue
2. Publish data directly on to [data.gov.ie](https://data.gov.ie)
3. Standardized metadata
4. Investigate if more currently closed data can be openly published using anonymisation
5. Promotion of Valuation Office Data to potential users
6. Publish to the [Public Service Data Catalogue](#)
7. Anonymous data



# Introduction

## The Valuation Office & Data

The Valuation Office understand that data is core to its mission and strategy, which is to become an organisation that will deliver an impartial and trusted property valuation service in an efficient manner to support the government of Ireland. The development of Project AXIA has generated further momentum in its focus to ensure its data are of the highest possible quality.

The Office has been a pioneer in Open Data and has been publishing data including via API for many years. It recognised the opportunity to complete a Data Audit across the organisation that will enable to identify which datasets can be published as open data. The Valuation Office needs to be able to collect, store, share and analyse data – both internally and externally – in order to deliver its functions, provide policy insights and drive performance, while complying with applicable data protection and governance laws and best practices.

### **The Valuation Office has recognised the opportunity to become insight rich as a whole**

- + Leveraging more external data
- + Addressing known data gaps internally
- + Processing and leveraging internal data more efficiently
- + Building an overall corporate Data & Analytics capability that addresses current business needs
- + Enables the organisation to adapt in an agile fashion to changing business needs/opportunities (future proofing)
- + Putting in place a Data Governance framework that will positively impact current and future data practices (collection, access, integration, quality, curation etc)
- + Progressively building business insights across the analytics spectrum (business intelligence and advanced analytics)
- + Building a Data Infrastructure that supports the above goals.

The Valuation Office is already a publisher on Ireland's Open Data Portal ([data.gov.ie](https://data.gov.ie)) directly (with 5 statistical datasets and an open data API).



## The Need for a Data Audit

The transcription into law of the Open Data and Public Sector Information Directive in July 2021 creates a requirement/opportunity to complete a Data Audit across the organisation that will enable it to identify which datasets can be opened. The completion of the data audit will not only help the Valuation Office comply with the Open Data and Public Sector Information Directive but will also enhance the way the Valuation Office processes data.

The existence of a robust evidence-base to support decision making and performance monitoring is a fundamental element of a Data Governance Initiative i.e., having processes and procedures to ensure that data is consistent and trustworthy throughout an organisation.

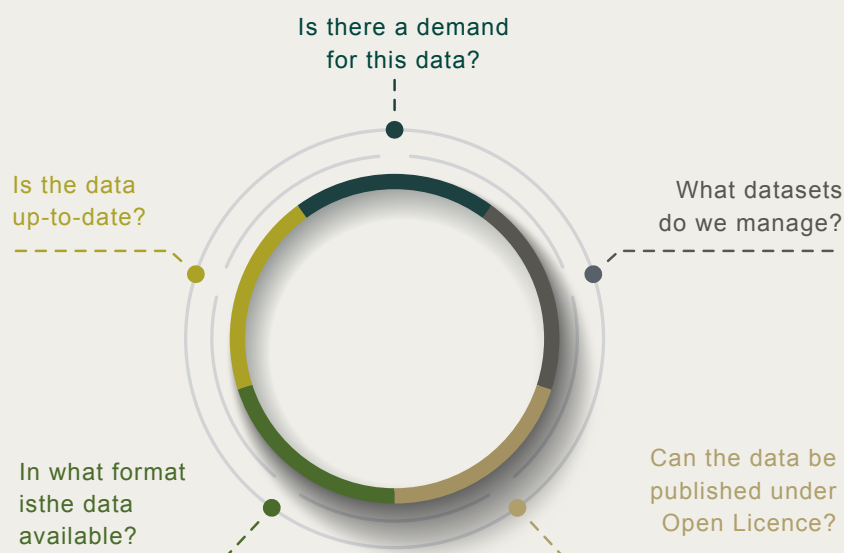
**However, there are a number of challenges in obtaining and maintaining a robust evidence base, including:**

- + There is often no clear and unambiguous picture of the data held by the organisation.
- + The absence of a common overview of the data makes it difficult to get value from the data i.e., there is no common data model.
- + Data is held in a variety of systems/formats, and it can be difficult to access.

**There are sometimes multiple and/or partial ‘versions of the truth’ which means that data is not always trusted.**

- + Duplication of data is common, where multiple business areas are collecting and storing the same or overlapping datasets.

The Derilinx Data Audit Framework directly addresses these challenges and provides a proven method for the identification of datasets, the compilation of a comprehensive inventory, and the definition of clear recommendations for managing this information and sharing it in a controlled environment. The Framework will enable the Valuation Office (during and beyond the pilot project) to find out what data they hold, where it is located and who is responsible for it.





### **The purpose of the audit:**

1. Find out what data assets are being created and held within organisations
2. Explore how those data are stored, managed, shared and reused
3. Identify any risks e.g., misuse, data loss or irretrievability
4. Learn about stakeholders' attitudes towards data creation and sharing
5. Suggest ways to improve ongoing data management.

Using strict data governance processes, each data asset can be reviewed for suitability for data sharing. The data asset can remain as private data, or approved for sharing with certain stakeholders, or approved for publication as Open Data.

# Data Audit Process & Results

This section contains the steps taken to complete a data audit are outlined as well as the findings which were reached after completion of the audit.

## Audit Process

The audit was completed with the following steps:

1. Identify and list all datasets
2. Assess the quality of each dataset
3. Assess each dataset for publishing criteria, rating the potential for opening or sharing.

Reviews of the spreadsheets were held over multiple calls between Derilinx and a nominated data champion (i.e., someone whose responsibility it was to gather the names of the datasets and fill out the template) from each department within the Valuation Office in September and October 2022, which included advice on further improving the findings of the audit.

The tables detailed in the next number of sections are counts of how many data assets have been labelled a certain way. The full individualised results of the data audit can be found in the accompanying spreadsheet.

## Identifying Datasets

During step 1, a data champion from each department was chosen for the task of identifying any datasets within their respective departments. A total of 48 different datasets were identified.



## Data Demand, Impact & Change Readiness

Each dataset was reviewed in terms of its use by others and its impact along three categories.

### 1. Demand

Is there a demand for this data, e.g., from the public, from the wider public sector, from the private sector?

### 2. Impact

What is the level of impact linked to publishing this data or linking this data to other data points?

### 3. Change readiness

Do people in the organisation have the time, skills or resources to act upon this data?

All of these were graded as high, medium or low by marking them as green, amber or red.

	Green	Amber	Red
<b>Demand</b>	8	10	35
<b>Impact</b>	7	15	26
<b>Change Readiness</b>	5	20	23

Table 1 Demand Impact & Change readiness counts

## Data Publishing

These results focus on what data is currently published as open data, what should be published as open or shared data, and the priority of publishing.

	Yes	No
<b>Is the data openly published?</b>	12	36

Table 2 Openly published counts

Following from this we want to know if this reflects the intended future availability of the data.

	Open	Internal	Do not publish
<b>Decision to publish</b>	17	12	10

Table 3 Decision to publish counts



## Data Quality Assessment

The quality of the datasets was assessed using the themes: Accessibility, Usability, Completeness, Granularity, Consistency, Interoperability, Metadata and Freshness/Timeliness.

The users were then asked questions using these themes and had to give either straight “Yes/No” answers or tiered answers in the form of “Green/Amber/Red”.

Based on the answers, each dataset was assigned a high-level score of “Gold/Silver/Bronze”.

For step 2, the Valuation Office then had to assess the quality of each of the datasets identified using the spreadsheet tool. The results can be seen in the tables below. Each subsection has a brief definition of the data topic.

### Accessibility

Data accessibility is the extent to which data users in an organisation can access and use data to achieve organisational goals.

High Level Definition	Yes	No
Is everyone in the organisation aware that this data is available?	31	19
Is the data well-managed and presented (within the organisation)?	41	9
Is the data readily accessible (within the organisation)?	37	13

Table 5 Accessibility counts

### Usability

Being able to access data is just one step. Data must be well-structured, well-maintained and publishable such that data users can easily get the information and insights they need from the data.

High Level Definition	High	Medium	Low
How easy is it to work with the data?	23	15	12

Table 6 Ease of working counts

High Level Definition	Yes	Now
Is the dataset ready to publish?	28	22

Table 7 Ready to publish counts



### Completeness

Data should capture everything that is needed by users of the data. That the data has comprehensive information on what it aims to measure.

High Level Definition	High	Medium	Low
How complete is the data?	34	8	8

Table 8 Completeness counts

### Granularity

Some data captures information at a broad or overview level. Other data captures it at a very exact and detailed level. Atomic data is the lowest level of detail possible. This question helps understand if low or high-level data is held.

High Level Definition	High	Medium	Low
How atomic is the data?	26	11	13

Table 9 Granularity counts

### Consistency

Data should stay aligned with anywhere else it exists. Every user who views the data in whatever place, should see the same information.

High Level Definition	High	Medium	Low
The degree to which data has attributes that are free from contradiction and are coherent with other data in a specific context of use	26	18	2

Table 10 Consistency counts

### Interpretability

Can all potential users get the information they need without more background knowledge? If data needs context or explanation to fully use, then that should be provided with the data.

High Level Definition	High	Medium	Low
How easy is it to understand the data (for a non-specialist)?	23	8	3

Table 11 Interpretability counts

## Metadata

Metadata is structured information provided along with data that helps users understand what a dataset contains. A number of popular metadata standards exist.

High Level Definition	High	Medium	Low
Is the data well described, does it have a standard metadata?	4	1	29

Table 12 Metadata counts

## Freshness / Timeliness

Ideally, the time between the data reference period and date of data publication is as short as possible. For example, data capturing a previous year would be available in January and not, say, August.

High Level Definition	High	Medium	Low
How close to creation is publication?	5	5	24

Table 13 Freshness/Timeliness counts

## Composite score

The composite score is an overall score to each dataset, summarising the results from the other data quality questions. There are three levels:

- + **Gold:** Scores high on all components of Data Quality
- + **Silver:** Scores medium to high on most components of Data Quality (no more than two Low)
- + **Bronze:** Does not meet 'Silver' criteria (no High and more than two Low)

Composite score	High	Medium	Low
Composite Score	2	28	4

Table 14 Composite score counts

## Summary

As an exercise involving data champions from departments across the Valuation Office, the process was positive. 34 datasets have been identified by this organisation and fully reviewed according to this audit template. This has allowed the Valuation Office to better understand the landscape of what data it holds and its potential users. It also allows the Valuation Office to ask questions of why the data is in its current state. We answer some of these questions in the next chapter.



# Analysis of Results

This section takes a more high-level, qualitative, look at the results of the data audit in the previous chapter. Summarising the main findings and highlighting areas that may need attention in the final recommendations.

## Which data assets are openly published?

Just 6 of the 42 data assets are currently published as open data. All these datasets appear to be statistical which is one of the 6 categories of high value datasets in the open data directive

These datasets are:

Department	Dataset name	Dataset description
Management Support	Data Protection	Number of Data protection queries and breaches received each year
Management Support	Parliamentary Questions	Number of PQs issued per year
Management Support	Political Representations	Number of Political Reps issued per year
Management Support	Press Queries	Number of Press queries received per year
Management Support	Website Analytics	Number of Sessions and Users
Procurement, Training & Development	Training Statistics	Various types of training courses from health & safety one day courses, third level college degrees etc.

Table 15 Openly published datasets

## Which data assets are to be published internally?

The datasets mentioned here are also statistical datasets which are some of the high value datasets mentioned in the open data directive. Majority of datasets come from the accounts department, which may help other departments with budgeting and increase transparency within the organisation.

These data assets are:

Department	Dataset name	Dataset description
Accounts	Consultancy expenses	Total Consultancy expenses
Accounts	Incidental Expenses	Total incidental expenses 2022
Accounts	Legal expenses	Total Legal expenses
Accounts	Office Machinery & other supplies	Total Office Machinery & other supplies
Accounts	Office Premises expenses	Total Office Premises expenses
Accounts	Post & Telecommunications expenses	Total Post & Telecommunications expenses
Accounts	Total accumulated spending	Total accumulated spending
Corporate Services Unit	Statistics regarding Records transferred to National Archives	Circuit Court Appeal files and 1st Appeals to the Commissioner files have been transferred to the National Archives this year. It is intended to transfer further records. Historically other records have been transferred but details on these may not be readily available.
Corporate Services Unit	Amounts remitted to the climate fund	Co2 usage and distance travelled by staff
Corporate Services Unit	Corporate services stock take 2021	'Stock-take' of all stationary items within the store rooms
Management Support	Technical Review Group decisions	Statistics on Cases discussed and outcomes

Table 16 Datasets to be published internally

### For the shared or closed data assets, why aren't they openly published?

Not all data is suitable for sharing due to certain constraints like GDPR on personal data, agreements on datasets shared from other public service bodies, commercial sensitivity or operational data that can't be shared outside the organisation.

Only 12 of the 54 datasets listed by the Valuation Office in the data audit are currently available as open data. During the data audit and additional 14 datasets that were identified as being suitable for open data. The change readiness of these datasets has been identified as 'Red' indicating that there is a lot of work needed to get these datasets to a state where they can be published.

### Is there a demand for the data assets?

The demand breakdown between the 53 datasets was, 8 green, 10 amber and 35 red. Indicating that there is not a strong demand for data.

4 of the published datasets scored highly for demand, but not all can be shared. 2 of the datasets can't be shared due to the Valuation Office not owning the datasets. The other 2 datasets are "Training statistics" from the "Procurement, Training and Development" Department & the "Valuation Certificates to ratepayers/agents" from the Valuation Administration Unit. For these high-demand datasets, the Valuation Office could anonymise or de-risk them such that they could be published openly.

An example of this would be the API which provides real time data in relation to post-Revaluation Rating Authority area.

### Which data assets are in high demand but aren't openly published?

There are four of these data assets. Helpfully all but one, are available internally. These data assets are:

Name	Publish Decision	Classification Comments
Minimising waste generation and Maximising Recycling	Amber	It shows how we are trying to reduce our waste and how we can work as a team within the office to maximise our recycling
Flooding information	Red	Sourced through data sharing agreement, not available for open data
Public Office & Archive Section	Amber	A joint effort is in line with the merger of three organizations VO, PRA and OSI to form Tailte Éireann, discussions, and working groups are in place and more data regarding archives and other public-facing activities will be consolidated and published as open data once the formation of Tailte Éireann is finalised.
Valuation reports to all Local Authorities	Amber	This will be streamlined as part of AXIA roll out and gradually becomes open data.
Details of Properties from Pre-Revaluation Counties" that are not available on the API	Green	This will be streamlined as part of AXIA roll out and gradually becomes open data.

Table 17 High demand datasets that aren't openly published

While Valuation Office has decided not to openly publish these, versions of some of the data assets could be open. With the Valuation reports to all local authorities' dataset, the Valuation Office is already publishing something similar with the REST API, this could be expanded to include the reports to local authorities.

### Is the data deemed to be of high quality?

Looking at the composite scores it seems more data assets fall at a middling level of data quality. Different data assets do better in certain categories, but some patterns emerge.

- + Most people within their departments are familiar with what data they hold.
- + Use of metadata standards seems to be reasonably high with 25 datasets across the organisations scoring high.
- + Overall, the data seems to be well-defined, consistent, and interpretable.
- + There also seems to be a gap between when data is collected and when it is published. The majority of data seems to be published close to the collection data, but not a lot is published very close to the collection date. This should be addressed with the completion of Project AXIA which aims to improve data collection processing and sharing within the organisation.



# Recommendations

From these results from the audit, the main insight is that the Valuation Office maintains mostly good quality data, but finding out about, accessing and using this data is currently a challenge.

The exercise was illuminating to show what data exists and it's potential for sharing and openness. Participants commented that it was very positive to have a more complete view of the data landscape which can provide a foundation for building and maintaining a data infrastructure in future.

## We recommend that the Valuation Office:

1. Create a well-organised, accessible data catalogue. This catalogue should list not just the datasets but also well-describe the data they hold. Even if people in departments can't access certain datasets, at least they can know information about them.
2. Publish further identified data to [data.gov.ie](https://data.gov.ie). A quick path to publishing is to publish data directly on <https://data.gov.ie/>. PDFs are valid for now but in the future aim for CSVs and other open, machine-readable formats.
3. Standardise Metadata. Use of metadata standards. Choosing which is most appropriate and then applying that. In combination with this, reference guides for staff on how to add good metadata to data assets and what metadata is expected. Metadata is important as it enables a potential user to find datasets.
4. Anonymous data. For datasets which can't be openly published but which there is some demand for, try and see if there are versions of the data that could be published. This may involve some discussion with users of the data to see what level of information removal, or anonymisation, would be acceptable to them in the published data.
5. Promotion of available data. The high awareness in each department of how well the data is known within each department is a positive finding. After the data catalogue is created, and publishing is improved, the next step is to tell other departments. Valuation Office should promote that data is available within other departments, how people can access it, an overview of what the data holds, and that each department can help them with any use cases. This will help to encourage a culture of data sharing within the organisation
6. Publish to the public Service Data Catalogue. Any datasets that will be published internally for other departments to use can be published on the [Public Service Data Catalogue](#). The PSDC publishes the metadata of these datasets and so will not give away any sensitive data but will alert the public and other Public Service Bodies (PSB) to the existence of such datasets, which will help increase overall transparency and help promote data sharing with other PSBs

Overall, these steps will significantly mature the value the Valuation Office generates from its data.



**Date** 12/01/2023

**Web** [www.derilinx.com](http://www.derilinx.com)

**Email** [info@derilinx.com](mailto:info@derilinx.com)

**Address** 38-39 Fitzwilliam Square, Dublin 2, D02 NX53

**Tel** +353 (0)1 254 4316

**Mob** +353 (0)87 798 1808

**Twitter** @derilinx