Data Sharing for National Statistical Offices

A step-by-step handbook for sharing data safely, legally, and usefully in the Caribbean

Selvi Jeyaseelan, Ian R Hambleton

2026-01-26

Table of contents

A Introduction

A.1 Who this guide is for

This guide is for staff working in National Statistical Offices (NSOs) across the Caribbean who are involved in preparing, approving, releasing, or managing access to data.

It is written for people who are comfortable working with data, but who may not have formal training in data sharing frameworks, licensing, or data governance.

This is a practical operational guide. It does not replace national legislation, institutional policy, or legal advice.

For a more in-depth discussion of the data sharing landscape in 2026, please read our companion guide: *"Data Sharing for Data Producers in Small Islands"*. For help with deciding *whether* to share, please read our: *CaribData Decision Guide*.

A.2 Why data sharing matters for NSOs

National Statistical Offices collect and manage data using public resources. This creates a responsibility not only to protect data, but also to ensure that data are used to generate public value.

Well-managed data sharing can:

- improve evidence-informed policymaking,
- strengthen regional and international reporting,
- increase trust in official statistics,
- reduce duplication of data collection,
- and enable new insights through secondary analysis.

At the same time, poorly managed data sharing can undermine public trust and expose institutions to legal and reputational risk. This guide focuses on how to share data well, not simply whether to share.

A.3 How to use this guide

This guide is structured as a *data sharing playbook*.

It follows the typical lifecycle of a data sharing decision, from early assessment through to post-release review. Each section explains why a step matters and what to do in practice. You do not need to apply every step in every situation. Judgement is required.

A.4 What we mean by data sharing

Data sharing means making data available for use by others in a controlled and purposeful way.

This may include:

- publishing summary statistics or datasets openly,
- providing access to approved users under defined conditions,
- sharing data with regional or international partners,
- or enabling access within secure environments.

Data sharing does not mean unrestricted release, and it does not mean giving up institutional control.

B The data sharing workflow at a glance

Most data sharing decisions follow the same broad pathway, outlined here as *13 steps*:

```
(1) Assess value → (2) Define purpose → (3) Identify audience → (4/5) Assess risk
      ↓
(6) Prepare data → (7) Describe data → (8) Choose data format → (9/10) access method
      ↓
(11) Record decision → (12) Release data → (13) Review use
```

Each step builds on the previous one. Skipping early steps usually increases risk later.

C Step 1: Is this dataset worth sharing?

C.1 Why this step matters

Preparing data for sharing takes time—occassionally considerable time. Not every dataset needs to be shared, and not every dataset needs the same level of investment. This step helps NSOs prioritise effort and avoid unnecessary work.

C.2 Questions to ask

A dataset is often a good candidate for sharing if one or more of the following apply:

- it supports national or regional policy priorities,
- it contributes to international reporting obligations,
- it represents a unique or hard-to-repeat data collection,
- it has potential for secondary analysis,
- it is frequently requested by external users.

If none of these apply, it may still be appropriate to share the data later, but it may not be a priority.

> ℹ Example
>
> A labour force survey conducted annually is a strong candidate for sharing because it supports policy, reporting, and trend analysis. A one-off internal pilot dataset may not be.

D Step 2: Clarifying the purpose of sharing

D.1 Why purpose comes first

Every data sharing decision should begin with a clear purpose. Purpose determines acceptable risk, appropriate access, and suitable safeguards. If the purpose is unclear, the safest option is to pause.

D.2 What to do in practice

Write a short purpose statement that explains why the data are being shared and how they are expected to be used. One or two sentences is enough, and the statement should be clear to someone who is not familiar with the dataset. A good purpose statement helps guide decisions about access, safeguards, and acceptable use, and provides a clear reference point if questions arise later.

Typical purposes include:

- informing public debate,
- supporting research,
- enabling regional comparison,
- improving service planning,
- or meeting reporting requirements.

> ℹ Example
>
> Purpose: *"To support regional analysis of youth unemployment trends for a CARICOM labour market report."*

If you cannot clearly state the purpose, the data should not be shared yet.

E Step 3: Identifying who should access the data

E.1 Why access decisions matter

Different users require different levels of access. Not everyone needs the same data or the same level of detail. Access should always be proportionate to risk—low-risk data can usually be shared more widely, while data that are more detailed, sensitive, or potentially identifiable require tighter controls and clearer conditions of use. This ensures that the level of access granted reflects the potential harm that could arise from misuse, misinterpretation, or unintended disclosure.

E.2 Common access types

Table 1: Common Data Access Types

Access type	Typical users	Risk level
Open access	General public	Low
Registered access	Analysts, journalists	Low–medium
Restricted access	Approved researchers	Medium
Secure access	Sensitive analysis	High

F Step 4: Open versus controlled data

Not all shared data needs to be open, and openness is not the only way data can deliver public value. Some datasets are well suited to open access, particularly when the risk of disclosure is low and the information supports transparency or public understanding. Other datasets can still generate significant public benefit when shared under controlled conditions, for example by enabling research, policy analysis, or regional collaboration while protecting confidentiality and public trust.

The table below summarises some of the key differences between open-access and controlled-access data, and highlights how each approach balances accessibility, risk, and public value. Importantly, both approaches can support good data-sharing practice, including alignment with FAIR principles, which are discussed later in this guide.

Table 2: Open vs. Controlled Data

Feature	Open-access data	Controlled-access data
Who can access	Anyone	Approved users only
Typical use	Transparency, reporting	Research, policy analysis
Disclosure risk	Low	Medium to high
Suitable for small populations	Sometimes	Often
Can be FAIR	Yes	Yes

G Step 5: Legal and ethical checks

G.1 Why this step cannot be skipped

NSOs have legal obligations to protect confidentiality, safeguard personal information, and maintain public trust in official statistics. Data sharing decisions must align with national statistics legislation, data protection laws, and widely accepted ethical norms, particularly where data relate to individuals, households, or small groups.

Even when data are de-identified or aggregated, legal and ethical considerations still apply, especially in small populations where re-identification risks may be higher.

G.2 Minimum checks

Before sharing, confirm that:

- the data can legally be shared under national statistics and data protection legislation,
- confidentiality obligations are met, including any limits on disclosure of identifiable or sensitive information,
- consent conditions are respected, where applicable, or that data sharing is otherwise permitted,
- contractual or institutional restrictions on use or onward sharing are understood.

These checks do not need to be complex, but they should be explicit and documented.

G.3 Practical guidance

In many cases, legal and ethical clearance can be handled through established internal processes, such as review by a senior officer, data governance committee, or designated data custodian. Formal legal review is usually only required for higher-risk datasets or unusual sharing arrangements.

Documenting who approved the decision, on what basis, and with what conditions helps protect both the institution and individual staff members if questions arise later.

H FAIR data principles

Many organisations refer to the FAIR principles, which describe good practice for making data **Findable, Accessible, Interoperable, and Reusable**. These principles are increasingly referenced by funders, regional bodies, and international partners.

FAIR does not mean open. Data can meet FAIR principles even when access is restricted, provided that potential users can discover the data, understand how access works, and use the data appropriately once permission is granted.

The table below summarises what FAIR typically looks like in practice for NSOs.

Table 3: FAIR Data Principles

Principle	What it means for NSOs
Findable	The dataset has a clear title, description, and a stable link or reference
Accessible	Users can easily understand how to request or obtain access
Interoperable	Data use standard formats, definitions, and classifications where possible
Reusable	Conditions of use, limitations, and provenance are clearly described

H.1 Applying FAIR in practice

Applying FAIR principles does not require new systems or specialist tools. In most cases, small improvements make a meaningful difference, such as writing clearer dataset descriptions, using common file formats, or documenting variables consistently.

FAIR is best used as a checklist rather than a compliance exercise. NSOs can improve findability and reusability even for sensitive datasets by publishing metadata, describing access conditions clearly, and documenting how data were created and processed.

Used this way, FAIR provides a helpful structure for improving data sharing and data quality without increasing disclosure risk or reducing institutional control.

I Step 6: Preparing data for sharing

I.1 Why preparation matters

Raw data is rarely suitable for sharing. Preparation reduces disclosure risk, improves usability, and ensures that shared datasets are consistent with legal, ethical, and quality expectations.

Just as importantly, preparation should be *transparent and reproducible*. NSOs should be able to demonstrate how a shared dataset was derived from the original (canonical) dataset, and to repeat the process if the data need to be updated or corrected.

I.2 Canonical datasets and derived versions

NSOs should treat the original, authoritative dataset as the *canonical version*. Any dataset prepared for sharing should be a *derived version*, created through a documented and repeatable process.

The canonical dataset should never be edited directly for sharing purposes.

I.3 Typical preparation actions

Preparation often includes:

- removing direct identifiers such as names, identification numbers, or addresses,
- reviewing indirect identifiers that could allow re-identification when combined,
- grouping or suppressing small categories,
- checking for obvious errors or inconsistencies,
- creating new variables (for example, age bands) specifically for sharing,
- documenting all changes made.

I.4 Creating an auditable preparation process

Wherever possible, preparation should be carried out using *scripted code* rather than manual edits. This can be done using whatever statistical software is routinely used within the NSO (for example, Stata, R, SPSS, Python).

Using code allows the NSO to:

- clearly show how the shared dataset was produced,
- reproduce the dataset if updates are required,
- reduce the risk of accidental changes,
- and support internal review or audit if needed.

The preparation code should:

- start from the canonical dataset,
- apply all transformations explicitly,
- generate a clearly named output dataset for sharing.

Both the code and the derived dataset should be retained.

I.5 Version control and naming

At all times, datasets prepared for sharing should be clearly versioned. File names or metadata should reflect:

- the source dataset,
- the purpose of the shared version,
- the date or version number.

This helps prevent confusion and ensures that users, internally and externally, are working with the correct version of the data.

ℹ Example

A health dataset has names removed, age grouped into five-year bands, and rare conditions combined into broader categories using a scripted data preparation file. The original dataset remains unchanged, while the shared version is saved with a new version name and accompanied by the preparation code.

J Anonymisation in practice

Anonymisation is the process of reducing the likelihood that individuals, households, or organisations can be identified in a shared dataset. This is a critical step in data sharing, particularly for NSOs working with population, health, education, or administrative data.

In practice, anonymisation usually involves a *combination of techniques*, rather than a single action. These may include:

- removing direct identifiers such as names, identification numbers, or exact addresses,
- modifying indirect identifiers such as age, occupation, or geography so they are less precise,
- grouping or top-coding variables where extreme or rare values increase identification risk,
- suppressing or aggregating small counts that could reveal individuals or small groups.

Top-coding is a common disclosure control technique where values above a chosen threshold are grouped together and reported as a single maximum category. For example, exact ages above 80 might all be recorded as "80+", or very high income values might be capped at a single upper value. This reduces the risk that individuals with rare or extreme characteristics can be identified, while still preserving analytical usefulness.

Anonymisation should always be assessed in context. Data that appear anonymised in isolation may become identifiable when combined with other datasets or local knowledge, especially in small populations.

NSOs should apply anonymisation as part of a documented, repeatable preparation process, and should consider whether the remaining level of detail is necessary for the stated purpose of sharing. If detail does not add clear analytical value, it is usually safer to remove or aggregate it.

Anonymisation reduces risk, but it does not eliminate risk entirely. For this reason, anonymisation decisions should be considered alongside access controls, conditions of use, and documentation.

K Step 7: Describing the data (metadata)

K.1 Why metadata is essential

Data that cannot be understood cannot be used safely or effectively. Metadata provides the context that allows users to interpret data correctly and reduces the risk of misuse, misinterpretation, or repeated follow-up queries to the NSO.

Metadata should be treated as a *core product of data sharing*, not as an optional add-on. Well-prepared metadata protects the NSO by clearly stating what the data can and cannot be used for, and by documenting limitations up front.

K.2 What metadata looks like in practice

In practice, metadata may be provided as:

- a short description on a data portal or repository page,
- a separate "readme" or documentation file,
- a data dictionary describing variables and values,
- or structured metadata entered into a repository system.

The format matters less than clarity and completeness.

K.3 Minimum metadata to include

At a minimum, every shared dataset should include the following information.

Table 4: Common Metadata Inclusions

Metadata area	What to include
Purpose	Why the data was collected and why it is being shared
Contents	What the dataset contains and what it does not contain
Coverage	Time period covered and geographic scope
Methods	How the data was collected or generated
Processing	Key transformations, anonymisation, or derivations
Quality notes	Known limitations, caveats, or data quality issues
Access	Conditions of use and access restrictions
Contact	Who to contact for questions or clarification

This level of metadata is usually sufficient for most users and most sharing scenarios.

K.4 Using metadata standards

Many NSOs already follow recognised metadata standards, either formally or informally. Commonly used standards include:

- **Dublin Core** for general dataset description,
- **DDI (Data Documentation Initiative)** for survey and statistical data,
- **DataCite metadata** when datasets are assigned persistent identifiers such as DOIs.

NSOs do not need to implement these standards in full to benefit from them. Using their basic structure as a guide can help ensure that important information is not overlooked and that metadata is compatible with regional or international platforms.

K.5 Keeping metadata aligned with the data

Metadata should be reviewed and updated whenever a dataset is updated or a new version is released. Metadata that no longer reflects the data can be misleading and may create risk.

As a practical rule, metadata should always describe:

- the *specific version* of the dataset being shared, and
- the *conditions under which that version was prepared and released.*

Example: Metadata for a shared NSO dataset

Dataset title
Youth Labour Force Indicators, 2022–2024
Purpose
This dataset was collected to monitor labour force participation, employment, and unemployment among persons aged 15–29, and is shared to support policy analysis and regional reporting.
Contents
Aggregated quarterly indicators including labour force participation rate, employment rate, and unemployment rate, disaggregated by sex and age group.
Coverage
Time period: Q1 2022 to Q4 2024
Geographic coverage: National level only
Methods
Data were collected through the national Labour Force Survey using a stratified household sampling design. Estimates are weighted to reflect the national population.
Processing and anonymisation
Direct identifiers were removed. Age was grouped into five-year bands.

Small cell counts were suppressed. Derived indicators were calculated using standard ILO definitions.

Quality notes

Estimates are subject to sampling error and should not be used for subnational analysis. Changes in questionnaire design in 2023 may affect comparability over time.

Access and conditions of use

This dataset is available under controlled access for non-commercial research and policy analysis. Users must acknowledge the National Statistical Office as the data source.

Contact

Statistics Division, National Statistical Office
Email: statistics@nso.gov

L Step 8: Choosing data formats

The format in which data are shared has a direct impact on how easily they can be used, combined with other datasets, and preserved over time. Choosing appropriate formats also reduces the risk of data being misunderstood or becoming inaccessible.

As a general rule, NSOs should prefer formats that are:

- widely used and well documented,
- machine-readable,
- non-proprietary or openly specified where possible,
- suitable for long-term preservation.

L.1 Practical guidance on format selection

When preparing data for sharing:

- export data directly from statistical software rather than copying and pasting values,
- avoid formats that lock data into a single software product,
- check that character encoding (for example, UTF-8) is handled correctly,
- ensure that variable names and value labels are preserved or clearly documented.

Where possible, include a short description of the file format in the metadata so users know how to open and use the data.

L.2 Common data types and formats

Table 5: Common Data Types

Data type	Preferred formats	Acceptable formats
Tabular data	CSV	Excel (XLS/XLSX), Stata (DTA)
Text documents	TXT, RTF, MD	DOCX, PDF
Images	TIFF	JPEG, PNG
Audio	WAV, FLAC	MP3
Video	MP4 (H.264)	MOV
Geospatial data	GeoJSON, Shapefile	KML

L.3 International standards and specifications

Many data formats are underpinned by international standards that support interoperability and long-term use. For example:

- *ISO standards* define common approaches for dates, country codes, languages, and geographic identifiers,
- formats such as *CSV*, *GeoJSON*, and *TIFF* are open specifications widely supported across platforms,
- geospatial data formats often align with standards from organisations such as the Open Geospatial Consortium (OGC).

NSOs do not need to implement standards formally in every case, but aligning with commonly used standards makes data easier to combine with national, regional, and international datasets.

L.4 New and emerging data types

NSOs are increasingly working with data beyond traditional surveys and administrative records. These may include:

- sensor or environmental data,
- mobility or transport data,
- satellite or remote-sensing imagery,
- digital transaction or platform data,
- text or social media data.

For newer data types, best practice may still be evolving. In these cases:

- document the format clearly in metadata,
- describe any specialised software required,
- consider sharing derived or aggregated versions alongside raw data.

L.5 Format decision guide in practice

Use the guide below when deciding how to share a dataset.

- If the data are mainly numeric and tabular, start with *CSV* for maximum compatibility.
- If users need formulas, multiple sheets, or familiar workflows, *Excel* may be acceptable, but avoid using it as the only format. An open source alternative (such as OpenDocument, suffix *ODS* is preferable).
- If the data will be analysed statistically, consider providing *CSV plus a statistical format* (for example, R, Stata or SPSS).
- If the data include geography, use **GeoJSON or Shapefile** rather than static maps.
- If the data are primarily for reading rather than analysis, provide a *PDF in addition to*, not instead of, the underlying data.

L.6 Preserving value labels and codebooks

When exporting data from statistical software, value labels, category definitions, and variable descriptions are often lost in simple formats such as CSV.

To avoid this:

- include a separate *codebook or data dictionary* describing variables and values,
- export labels to a documentation file where possible,
- reference the codebook clearly in the metadata.

As a practical rule, users should be able to understand every variable in the dataset without referring back to the original data producer.

L.7 A practical rule of thumb

Choose the simplest format that preserves the structure and meaning of the data, aligns with common standards, and can be used by the widest possible audience without specialised tools.

M Step 9: Choosing a licence or conditions of use

Data is protected by copyright by default. A licence (or clear conditions of use) explains what others are allowed to do with the data and what obligations they have when using it. Clear licensing reduces misuse, prevents misunderstandings, and lowers the need for case-by-case clarification by the NSO.

M.1 Creative Commons in practice

Creative Commons (CC) licences are widely used, internationally recognised, and well suited to many types of statistical data. They provide a simple, standardised way to communicate permissions and restrictions, and they are easily understood by researchers, policymakers, and data platforms.

Creative Commons licences range from very open (placing minimal restrictions on reuse) to more restrictive (limiting commercial use or adaptation). Importantly, using a Creative Commons licence does **not** mean giving up ownership of the data — it simply defines how others may use it.

M.2 Other licence types

While Creative Commons licences are common for data, other licence types also exist and may be appropriate in some contexts. For example:

- **MIT** or similar open licences are sometimes used for code or software shared alongside data,
- bespoke or institutional licences may be used for controlled-access datasets,
- data sharing agreements may function as licences for restricted data.

The key requirement is that conditions of use are *clear, explicit, and documented*.

M.3 Common licence options

The table below summarises commonly used licences and what they allow.

Table 6: Useful Creative Commons Licensing

Licence	Attribution required	Commercial use allowed	Adaptation allowed	Typical use
CC0	No	Yes	Yes	Maximum reuse, open data
CC BY	Yes	Yes	Yes	Open data with attribution
CC BY-SA	Yes	Yes	Yes (share alike)	Data intended to remain open
CC BY-NC	Yes	No	Yes	Non-commercial research
CC BY-ND	Yes	Yes	No	Data where modification is discouraged
MIT	Yes	Yes	Yes	Code or scripts shared with data

M.4 Practical guidance for NSOs

When choosing a licence or conditions of use:

- match the licence to the *purpose of sharing* and *level of risk*,
- avoid restrictive licences unless there is a clear reason,

- ensure the licence is stated clearly in metadata and documentation,
- apply the licence consistently across versions of the same dataset.

For controlled-access data, a formal licence may be replaced or supplemented by a data sharing agreement that sets out permitted uses and responsibilities.

M.5 A guiding principle

Choose the least restrictive licence consistent with the purpose of sharing, legal obligations, and the level of risk, while ensuring that expectations for attribution and use are unambiguous.

N Step 10: Choosing a sharing platform

Where data is shared matters as much as how it is shared. The platform you choose affects who can find the data, how it can be accessed, how it is cited, and how long it remains available.

A good platform choice supports visibility and reuse while maintaining appropriate control and reducing administrative burden for the NSO.

N.1 A practical decision process

When selecting a sharing platform, work through the following questions in order:

- **Who needs to access the data?**
  Public users, approved researchers, government partners, or internal staff may require different platforms.
- **What level of control is required?**
  Fully open data can be hosted on public portals, while sensitive or controlled-access data require platforms with access management and user authentication.
- **Is long-term access important?**
  For datasets that will be cited or reused over time, platforms that support persistence and versioning are preferable.
- **Does the platform support good documentation?**
  The ability to attach metadata, codebooks, and licences is essential.
- **Does the platform create unnecessary lock-in?**
  Avoid platforms that restrict export, reuse, or migration of data in the future.

The simplest platform that meets these needs is usually the best choice.

N.2 General principles for platform selection

When possible:

- use established and widely recognised repositories,
- prefer platforms that assign persistent identifiers (such as DOIs),
- ensure metadata and access rules are visible and clear,
- avoid platforms that depend on proprietary formats or software.

Institutional and regional repositories often provide a good balance between control, credibility, and visibility.

N.3 Common platform types: features, benefits, and drawbacks

The table below summarises common types of data-sharing platforms and their typical strengths and limitations.

Table 7: Common Data Sharing Platform Types

Platform type	Features	Benefits	Drawbacks
National open data portal	Public access, metadata pages	High visibility, transparency	Limited support for sensitive data
Institutional repository	Metadata, persistent identifiers, versioning	Long-term preservation, credibility	May have limited access controls
Regional repository	Shared standards, regional visibility	Supports comparison and collaboration	Governance may be more complex
General-purpose repository (e.g. Zenodo)	DOIs, metadata, broad access	Easy to use, widely recognised	Limited custom access controls
Secure research platform	User approval, controlled environments	Strong confidentiality protection	Higher administrative overhead
Ad hoc file sharing (e.g. email, cloud folders)	Simple, quick	Low setup effort	Poor documentation, high risk

N.4 Practical guidance for NSOs

As a general rule:

- use *open data portals or general repositories* for low-risk, open datasets,
- use *institutional or regional repositories* for curated, citable datasets,

- use *secure or controlled platforms* for sensitive or detailed data,
- avoid relying on informal file sharing except for short-term internal use.

Where possible, document why a particular platform was chosen. This helps maintain consistency and supports future reviews as platforms and needs evolve.

O Step 11: Recording the decision

O.1 Why documentation matters

Clear records protect both staff and institutions. They provide evidence that appropriate checks were carried out, support consistency over time, and make it easier to respond to questions or challenges about how and why data were shared.

Good documentation also reduces reliance on individual memory and helps ensure continuity when staff roles change.

O.2 What to record

At a minimum, record the following information for each data-sharing decision:

- the dataset that was shared (including version),
- the stated purpose of sharing,
- access conditions or licence applied,
- the date of approval or release,
- the responsible officer or approving authority.

This information can be captured in a simple log, spreadsheet, or internal register. The format matters less than ensuring that records are complete, consistent, and easy to retrieve.

Where preparation code, agreements, or correspondence exist, note where these are stored.

P Step 12: Releasing and promoting the data

Once data are shared, they should be easy for intended users to find, access, and cite correctly. Release is not only a technical step, but also a communication step.

Good practice includes:

- linking shared datasets to reports, dashboards, or publications that use them,
- providing a recommended citation for the dataset,
- ensuring metadata and documentation are visible at the point of access,
- using shared datasets in internal training, briefings, or analytical work.

Where platforms provide basic metrics (such as downloads or access requests), these can be monitored to understand demand and demonstrate the value of sharing.

Promotion should always be proportionate to risk. Sensitive or controlled-access data may require targeted communication rather than public promotion.

Q Step 13: Reviewing downstream use

Q.1 Why review matters

Data sharing does not end at release. Reviewing how data are used helps NSOs identify problems early, improve future releases, and maintain trust with data users and the public.

Review also provides feedback on whether the original purpose of sharing is being met.

Q.2 What to do in practice

Where appropriate:

- review publications, reports, or outputs that use the data,
- check that data are interpreted and cited correctly,
- note any patterns of misunderstanding or misuse,
- refine guidance, metadata, or preparation steps for future releases.

Focus review effort on higher-risk or high-profile uses. Not all shared data require active monitoring.

R Final considerations

R.1 Governance models in brief

NSOs organise data sharing in different ways depending on size, mandate, and capacity. Common governance approaches include:

- institution-led control, where the NSO manages all decisions centrally,
- cooperative or network-based arrangements, where responsibilities are shared,
- formal data trusts or stewardship models for specific data types.

Most organisations begin with strong internal control and adapt governance arrangements as experience, partnerships, and confidence grow.

R.2 A guiding principle

Start conservatively. Expand access gradually as safeguards, skills, and institutional capacity develop. Decisions that feel cautious early on often enable more confident sharing later.

R.3 Common pitfalls to avoid

Experience across many settings highlights recurring challenges:

- sharing data without a clearly stated purpose,
- overestimating the protection provided by anonymisation alone,
- failing to document decisions and approvals,
- adopting models designed for large countries without adaptation to small populations.

Being aware of these pitfalls helps reduce risk and improves consistency.

R.4 Final readiness check

Before releasing data, confirm that:

- the dataset is an appropriate candidate for sharing,
- the purpose is clear and documented,
- access conditions match the level of risk,
- legal and ethical checks are complete,
- data and metadata are prepared and aligned,
- small population risks have been considered,
- decisions and approvals are recorded.

If these conditions are met, sharing can proceed with confidence.

R.5 Closing note

Effective data sharing balances openness with care. It relies on proportionate judgement, clear documentation, and trust between data producers and users.

Start small. Learn from experience. Improve with each release.