

## Reviewer Template

**Instructions:** This template aims to provide guidance to the reviewers of case studies submitted for inclusion in the Handbook. As a reminder, the Handbook aims to provide practical, actionable guidance to researchers, administrators, policy makers, and other practitioners faced with the issue of implementing a new data access mechanism, or improving an existing mechanism. They should be able to draw on concrete examples here. Case studies should be comparable amongst each other, which is why we ask that they follow a certain template. In the following, we provide general and detailed guidance on what a reviewer should assess when creating feedback for the authors.

Case studies should describe the process of deciding about and ultimately providing researcher access to administrative data. This can include a description of options considered but not ultimately implemented. For solutions that were implemented, the authors should explain why this solution was best for the context at hand and what considerations informed the final decision. The case study author should also indicate in which areas they think they innovated or have particularly deep knowledge to share. Case study submissions can include tools/materials where appropriate.

### General Notes for Reviewers

- Imagine the case study as a sort of instruction manual for doing something similar in another location or with another partner - what information would be useful?
- The readership is global - authors should not assume that readers know US or other country's institutions.
- Authors should properly expand acronyms upon first use. No knowledge of acronyms should be assumed.
- Authors should support assertions with citations where possible. For example, if a paper mentions a policy decision that was influenced by the research, it would be great to provide a reference.
- While we are currently not imposing a citation style, citation content should follow Chicago Manual of Style ([https://www.chicagomanualofstyle.org/tools\\_citationguide/citation-guide-2.html](https://www.chicagomanualofstyle.org/tools_citationguide/citation-guide-2.html)). In particular,
  - URLs pointing to online documents should cite the document with all the usual information (author, publication date, etc.).
  - If citing a generic webpage, use a footnote citation indicating the date when last consulted.
  - Legal code should expand on acronyms ("Common Rule", "USC"), and ideally point to online repositories of legal code so interested parties can easily find the text of the legal code. Example:

5 United States Code (U.S.C.) § 552b. Accessed at <https://www.law.cornell.edu/uscode/text/5/552b> on February 28, 2020.

- Where appropriate: What were the software solutions & systems used?
- Who provided legal advice, if any? On the researcher side? On the data provider side?
- What share of the cost or funding went towards the different project aspects, e.g. hardware vs. staff?
- What was the staffing, what skills were they hiring for or using?

### Summary

**Instructions:** The summary should provide an overview of the key elements of the case study: role of the authors, the data provider, data content, data collectors, data curators, and interesting uses of the data. It

might also highlight certain elements not covered.

## Motivation

**Instructions:** The motivation section aims to identify what drove the original project to make the data available. Was the initiator a researcher or the data provider (business, non-profit, government)? What were they after - a one-off research project that turned out to be useful, a legal mandate to make data more accessible? What outcomes were the originators of the collaboration after? Note that this section should not be talking about the legal context of the *how*, only the *why*. Other sections address the *how* of communicating outcomes, of defining who can access and who cannot.

As a reviewer, can you relate to this motivation? How could the authors better express their motivation?

## Data use examples

**Instructions:** This section should give a clear idea of the *initial* use of the data. Graphics can be used to support the description. Note that this should be concrete - it should provide examples of how the data was actually used, not prospective or speculative uses. Examples are listed in the [Template].

## Legal context

### Access

**Instructions:** In this section, the authors should describe in what context the data provider is allowed to, able to, or mandated to provide access to the data. The legal framework for access should be described in a way that is intelligible to a non-expert or a non-native (references are encouraged). The legal framework presumably mandates the use of some form of contract - a non-disclosure agreement, a protocol, a data use agreement, a framework, etc. How much leeway did the team have to craft those, and what drove those decisions? What kind of sanctions can the data provider impose (financial, reputational, penal) for unauthorized uses of the data? If yes, why were those chosen? If possible, the authors should provide a concrete example - a DUA template, etc. - pruned of any identifying information.

As reviewer, can you relate to the description? Can you translate some of the lessons from this experience into your own context?

## Intellectual property

**Instructions:** Does the data provider assert Intellectual Property (IP) on the original data, or on any derivative data or products created by researchers with access to the data, such as tables, research papers, source code, etc.? How is IP enforced? Does the data provider require co-authorship on researcher output? If there is a review process, does it focus on IP (if not, it should be mentioned later under "[Safe Outputs]"). Does the data provider have a veto right over outputs? Is there a specific copyright on the data and code generated by the project, or a specific license imposed by the data provider, or chosen by all?

## Making Data Usable

**Instructions:** The description should focus on the many steps to create usable data - details are encouraged. Note that usable is meant from a statistical/ quality/ processing point of view. Modifications to the data to be able to release the data should be described under "Safe Outputs", modifications to allow for access (de-identification, etc.) should be described under "Safe Data". Who made the data usable, how did the process work? How repeatable is it?

## Five Safes framework

**Instructions:** The "Five Safes framework" [DesaiFiveSafesdesigning2016] provides structure. The authors should discuss safe projects, safe people, safe settings, safe data, safe outputs. The discussion should allow the reviewer and the reader to identify useful interpretations, and generalizability. How are "safe people" defined or identified? How did data become "safe"? If describing "safe settings", identify what that meant both from the computer equipment side, as well as from the physical rooms (refer to the chapter on "Physical security" if necessary).

**Instructions:** We would also like to know how onerous each dimension was to implement (cost in terms of time, personnel, and money).

## Revenue

**Instructions:** The revenue discussion identifies the counter part to the various cost components. How were incremental costs covered, was there full cost recovery? What was the pricing model? Again, where possible, precise numbers should be used. At the end of the authors' initial project, how sustainable do they think it is? Are there recurring revenue streams, did the data provider turn it into a revenue source?

## Metadata

**Instructions:** The authors should describe how metadata (file formats, variable codebooks, etc.) were generated, how they were published (references should be provided). Again, examples should be provided explicitly or by reference. Did the provision of metadata play a role from the start? (One reference might be the FAIR Principles)

## Outreach

**Instructions:**

- how do (additional) [Safe People](#) become aware of your access point and data offerings? Examples of newsletters, reference to particularly well-documented replication archives or such...
- How do/did you interact with the data providers/ other stakeholders: what metrics on outcomes do/did you provide: measures of cost savings, effectiveness, custom presentations, etc.

## Data Life-cycle and Replicability

### Preservation of Researcher-accessible Files

**Instructions:** Are researcher-accessible files preserved? Are master files preserved? If yes, over what time period (5, 10 years, forever)? Is there active curation, or only preservation of current bitstreams? Does the

data provider perform the archival/curation functions itself, is another entity within the organization responsible for this activity, or is a third-party (national, regional, university archive) responsible for this? Are persistent identifiers generated for these?

### Reproducibility of Researcher-accessible Files

*Instructions:* Can researcher-accessible files be consistently regenerated, or are they snapshots of a dynamic database or data pipeline? Are the query mechanism/program code reproducible/archived (can similar files be recreated, can older files be retrieved), or are they manual queries (with risk of inconsistency, variation across personnel)?

### Preservation of Researcher-generated Files

*Instructions:* Researchers may generate intermediate and final data files and tables. Researchers generate processing code/programs. Are these preserved? If yes, for what time? Can other researchers access these files, with or without the permission of the original researchers (and subject to [Safe projects], [Safe settings], and [Safe people] conditions)? Does the data provider generate persistent identifiers for any of these?

### Disposal of data

*Instructions:* does the data provider delete data files or purge records within the data due to legal requirements or limited resources to maintain? Are records purged upon request or at certain time intervals (eg juvenile justice, deleted tweets)?

### Metrics

*Instructions:* How is the success of the data access project evaluated?

- Do you, as a reviewer, see the particular measures as being well explained? What is being counted: Number of people involved, number of people affected, number of datasets provided, number of papers, number of citations, number of mentions?
- Who is the audience for metrics - funders, data provider, legislature? Provide some sample statistics.