

# Beta: Agile Open Data Group Playbook

Inspired by the [Digital Service Playbook](#).

<b>Draft: Agile Data Review Group Playbook</b>	<b>1</b>
Introduction	2
Rationale	2
The Plays	3
Get buy-in	3
Talk to 9	4
Meet regularly	6
Respond to real asks	7
Structure as open data	8
Appendix - Data Embargo Agreement	9

## Introduction

This playbook outlines a new collaboration approach to improving the transparency and quality of government-maintained data. Using lessons learned during the release of data during the COVID-19 pandemic, this playbook is an agile framework for government employees responsible for data, and the journalists, researchers, and citizens who have an interest in the data.

We want the data quality acceleration and data transparency that happened in response to COVID to become the new normal for government operations.

## Rationale

There is a disconnect between outside users of open government data, government employees who maintain the data, and government policy makers who use the data. This disconnect makes it difficult for these groups to solve hard problems related to government data access and data quality. Small agile feedback loops improve data, but they are not possible using Federal committees or a full-blown regulatory process.

During the COVID crisis, the Trump Administration formed an ad hoc data feedback and testing group to address these issues. The group included healthcare data journalists, patient advocates, outside privacy experts, and federal employees.

We quickly released COVID data and documentation that were usable by journalists and policy makers at the state, city, and county levels. This made it possible to triangulate data from multiple sources, which increased the credibility of Federal experts and inoculated against misinformation. The data powered articles, research papers, and online dashboard tools to better inform the public. This helped to establish trust in the validity of COVID reporting, and, by implication, the COVID response.

As a result of this success, we've captured best practices in this playbook. This guide is geared towards public servants at all levels who want to foster open dialog with open data enthusiasts. We see Chief Data Officers, heads of government agencies, data scientists, journalists, policy experts, and members of the public interested in open government all being able to benefit from the lessons we've learned.

Our aim is to create a data-centered dialog between the government and the people it serves, drive greater transparency, formally facilitate trust, and improve the quality of work across government.

# The Plays

## Get buy-in

This agile data group will need endorsement and participation from the leadership of the agency. There should be a consensus that data outputs will be as high-quality and transparent as possible, and that a collaboration to achieve these goals is worthwhile. The agency leadership will need to agree that data uncertainties will be addressed through the group's expertise, and if any uncertainties can't be resolved, that the individuals in the group will properly weigh them against the value of releasing the data.

### Checklist:

- Write a proposal to follow this playbook for your agency sponsor. Include:
  - The value proposition.
  - The single accountable person as per the [USDS playbook](#).
  - Which data systems will be proposed to the group for review.
  - A read-out schedule to review progress.
- Engage individuals who will join the group from the government side. Include the following people, some who will be able to fill multiple roles:
  - Leaders who
    - Understand policy aims.
    - Have authority and willingness to release data.
    - Have clout to bring participation of System of Record owners.
    - Can get approval for controversial data releases.
  - Data engineers who have working knowledge of and access to the underlying databases.
  - Someone familiar with FOIA and privacy practices.
- Bring different policy makers and data system owners to the group as needed.

### Key Questions:

- What value will the group create?
- Who is your agency sponsor?
- When is the check-in with the agency sponsor?
- How will you demonstrate that the group is addressing risks in releasing data?
- Who in the group will release the data?
- Who will field technical questions about the underlying databases?
- Who has the ability to conduct queries against the database in question?
- Who is responsible for protecting the security and privacy of US citizens and businesses by using appropriate aggregation methods?

## Talk to 9

While choosing the non-government data testers, find people who are providing constructive feedback about the data. They should already be committed to making a positive impact on society by leveraging the data or related data.

Federal paper-work reduction act rules require that the group will consist of no more than 9 people from outside the government. While participants may change over time, that limit should be kept for any given meeting or feedback request. Further, the agency will gather input from the individuals, not from the group as a whole for consensus. This obviates the need for managing the group under the terms of the Federal Advisory Committee Act.

Individuals in the Open Data Group will need to speak freely and review data that is not ready for the public. Conversely, the public needs transparent access to the group interactions. The group will need a shared agreement which provides for temporary confidentiality.

### Checklist:

- Get up to 9 people as outside data testers. One person may fill multiple roles:
  - **Data consumers** working with data from the agency.
  - **Someone who is included in or impacted by the data.** For instance, patient voices for patient-related datasets.
  - **Data Journalists**, especially those who make dashboards or release raw data.
  - **Commercial data users** who make the data useful in commerce.
  - **FOIA requesters** who make requests on relevant data systems.
  - **Academic researchers**, whose studies or publications rely on the data.
  - All data testers should:
    - Be capable of reviewing complex technical details of data releases.
    - Have a working knowledge of data released by an agency, and its limitations.
    - Have a history of providing value to the public using currently available data, for example through a service or product.
    - Have a deep understanding of data ethics and data privacy.
- Understand and follow the Federal Rules and actions that apply to Open Data processes:
  - [Paperwork Reduction Act](#)
  - [Federal Advisory Committee Act](#)
  - [Exceptions to the need to form a FACA committee](#)
  - [Talk to 9](#) project
  - The [Freedom of Information Act exemptions](#)
  - The [OPEN Government Data Act](#) and the larger [Foundations for Evidence-Based Policymaking Act](#)
  - [Privacy Act](#)
  - [OMB Open Data Policy](#)

- The [requirements to use Open Data Formats for Open Data](#).
- The [Federal Data Strategy](#)
- The [Data \(Information\) Quality Act](#)
- Establish confidentiality ground rules that may vary agency-to-agency. Use the Data Embargo Agreement in the Appendix as a starting point.
- Review applicable regulations that create the tension between the need for confidentiality and transparency, for example, open meetings laws and HIPAA.

### **Key Questions:**

- Which data testers have working knowledge of the data that has already been released?
- What use-cases do the data testers represent?
- Which data testers have submitted FOIA requests on the databases in question?
- Which data-tester is comfortable with de-identification and aggregation methods, and the implications of [FOIA](#)?
- Which data testers have a track record of exercising data for the public good?
- Are all team members familiar with FOIA, FACA and the Paperwork Reduction Act?
- How is the group tracking how many people from outside the government are giving feedback?
- Is it clear to the group that the goal is not to reach consensus on any topic, but to provide individual feedback?
- Are there any agency-specific confidentiality requirements that must be met?

## Meet regularly

Weekly meetings are recommended to allow for a week's worth of independent work to be reviewed. Members should expect to be committed to this schedule for at least 1-2 years. Members should be rotated as the data needs of previous members are addressed and retired.

### **Checklist:**

- Choose the meeting schedule.
- Determine if and when Agile Data Group members should be changed.

### **Key Questions:**

- Do you have time to discuss all of the important issues and progress that have happened since the last meeting?

## Respond to real asks

Best practices established by prior agile data groups will be the foundation for the current group's data approach. But the group will continuously need to act on the requests and feedback from data users and the public regarding what makes data useful.

### Checklist:

- Create a list of datasets to release and/or improve using feedback from the data testers.
- Prioritize datasets that can be released quickly, ideally within a few weeks.
- Then focus on highly valuable datasets that may have been deemed too difficult or controversial to release.
- Attempt to identify and address any consequences of data uncertainty.
- Datasets that have significant privacy, ethics, or safety concerns that cannot be addressed by standard de-identification and quality assurance practices should take the FOIA and/or regulatory path to release.
- Determine how often data should be updated, with a preference to more frequent updates.
- Create and implement plans to apply data correction or improvement strategies.
- Focus on data ingestion processes to ensure that data is vetted when it is collected.
- Relay policy questions and feedback to other government leaders when appropriate.

### Key Questions:

- How will your group measure progress?
  - Has the group released new datasets?
  - Has the group improved existing datasets?
  - Has the group fulfilled its value proposition?
- What maximizes the data's downstream usefulness?
- What data limitations are preventing the agency from meeting its mandated goals?
- What data problems are slowing down data analysis and decision making?
- What data structure errors are stalling use-cases?
- What uncertainties exist around releasing the data? How will you address them?
- Are there any risks that cannot be resolved? How are they weighed against the value of releasing the data?
- Are policy issues raised by the team ever dismissed as "that's not about data"?
- Have the beta datasets been deleted at the appropriate time by the data testers?

## Structure as open data

Datasets should default to open. There will be privacy and commercial concerns that need to be addressed, but the “government as a data platform” idea requires data generosity.

### Checklist:

- Adhere data releases to [Open Data Standards](#).
- Structure data to make it friendly to the ETL processes and pipelines that are maintained by downstream consumers.
- Ensure that data releases will be backward compatible as much as possible.
- Create private resources to help each other use and understand the data. Once the data is released, these resources should be made public and potentially expanded. Such as:
  - Wiki's and/or ReadMes for working with released data.
  - Blog posts discussing how best to use the dataset and/or discussing the relevance of the dataset.
  - Webinars to discuss the implications and impact of the data release.
- Understand the regulations dealing with citizen access to data:
  - [Government in the Sunshine Act](#)
  - [Foundations for Evidence Based Policy Making Act](#)
  - [The Data/Information Quality Act](#)
  - [The Freedom of Information Act](#)

### Key Questions:

- Is it simple for outside users to compare current data releases with previous data releases?
- Is the data file in CSV, JSON, or some other open file format?
- Can this data file be made easier to ETL? Is the data normalized? Is the aggregation level clear?
- Does the data file have all the relevant fields?



## Appendix - Data Embargo Agreement

Do you agree to respect an embargo for the [insert filetype] file of [insert data description] data that will be shared with you via [insert method of delivery]? The embargo will lift at the time this file is publicly posted on [insert publication site] or after [4] weeks of you receiving this data, whichever is first. We will notify you when it has been posted. This data set is being shared with you for evaluation of the usefulness of some of the changes so this file may not ever be shared publicly.

By accepting the terms of the embargo you agree to not share this file with anyone or publicly discuss it, until the embargo lifts.

We further expect that discussion related to this dataset will be confidential until the embargo is lifted, and after data is released discussion of technical issues will be treated under Chatham House rules: i.e. not attributed to any specific person.

Please respond back individually (do not reply all) with your acceptance or denial of this embargo request. Additionally, if you intend to share the file with others in your organization, please include their names as well.