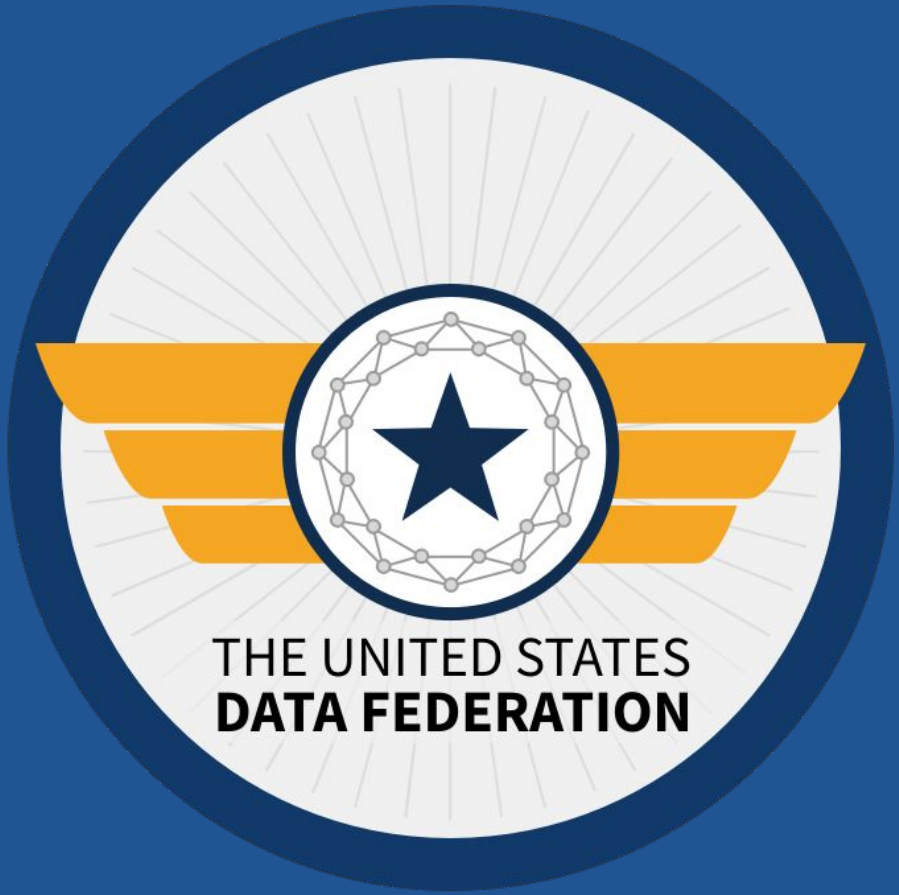

THE UNITED STATES
DATA FEDERATION

Hello!

Please introduce yourself in the chat and let us know what prompted you to join us today

Phil Ashlock
Director, Data & Analytics Portfolio
GSA Technology Transformation Services

Julia Lindpaintner
Project Lead & 18F UX Designer
GSA Technology Transformation Services

Amy Mok
18F Engineering Lead
GSA Technology Transformation Services

Overview
Our work
Our ask

Overview

Our work

Our ask

THE PROBLEM

Getting good, reliable data is hard.

ANOTHER PROBLEM

Getting data from different sources makes this even harder.

- Aligning incentives
- Translating between data formats
- Specifying data standards
- Driving adoption of standards
- Making data collection user-friendly
- Implementing reliable validation

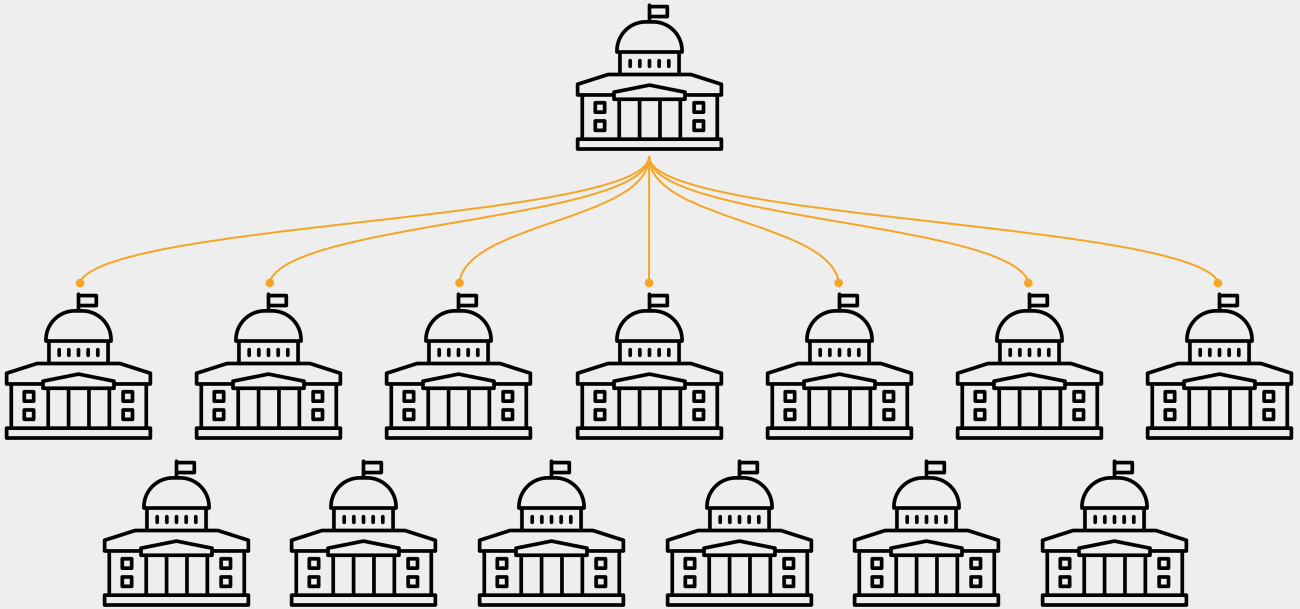
federated data effort

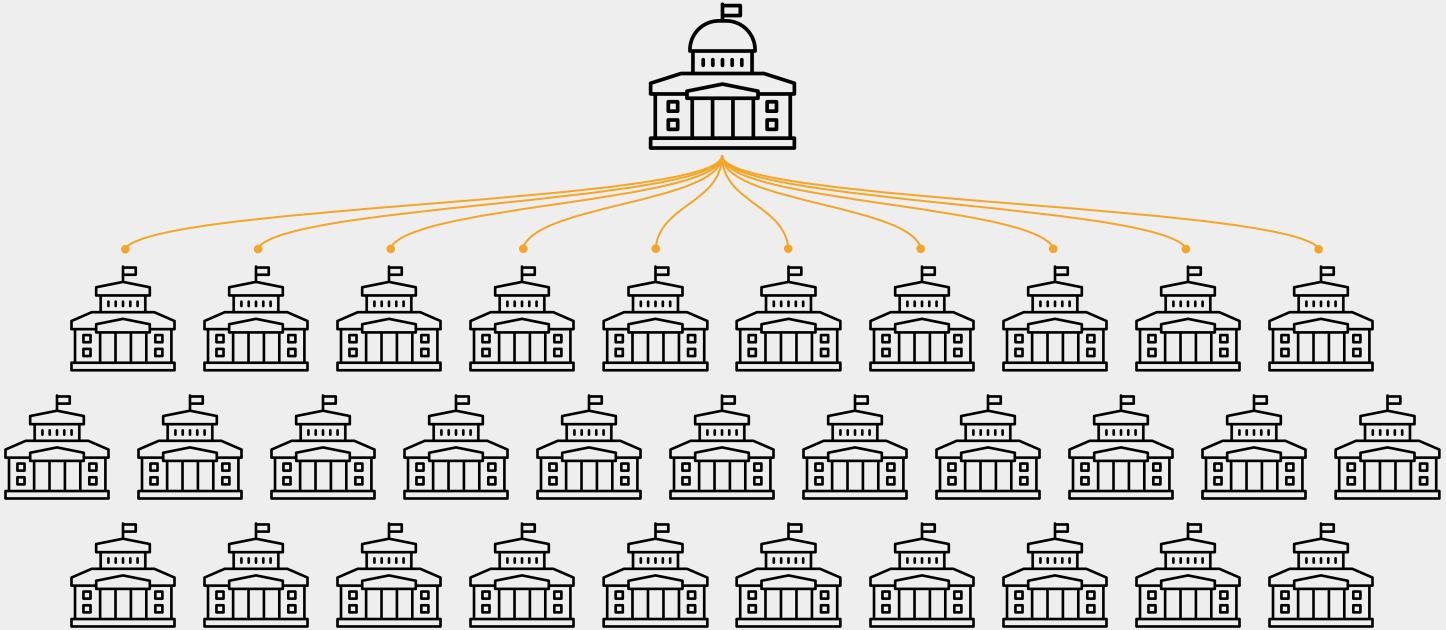
federated data effort

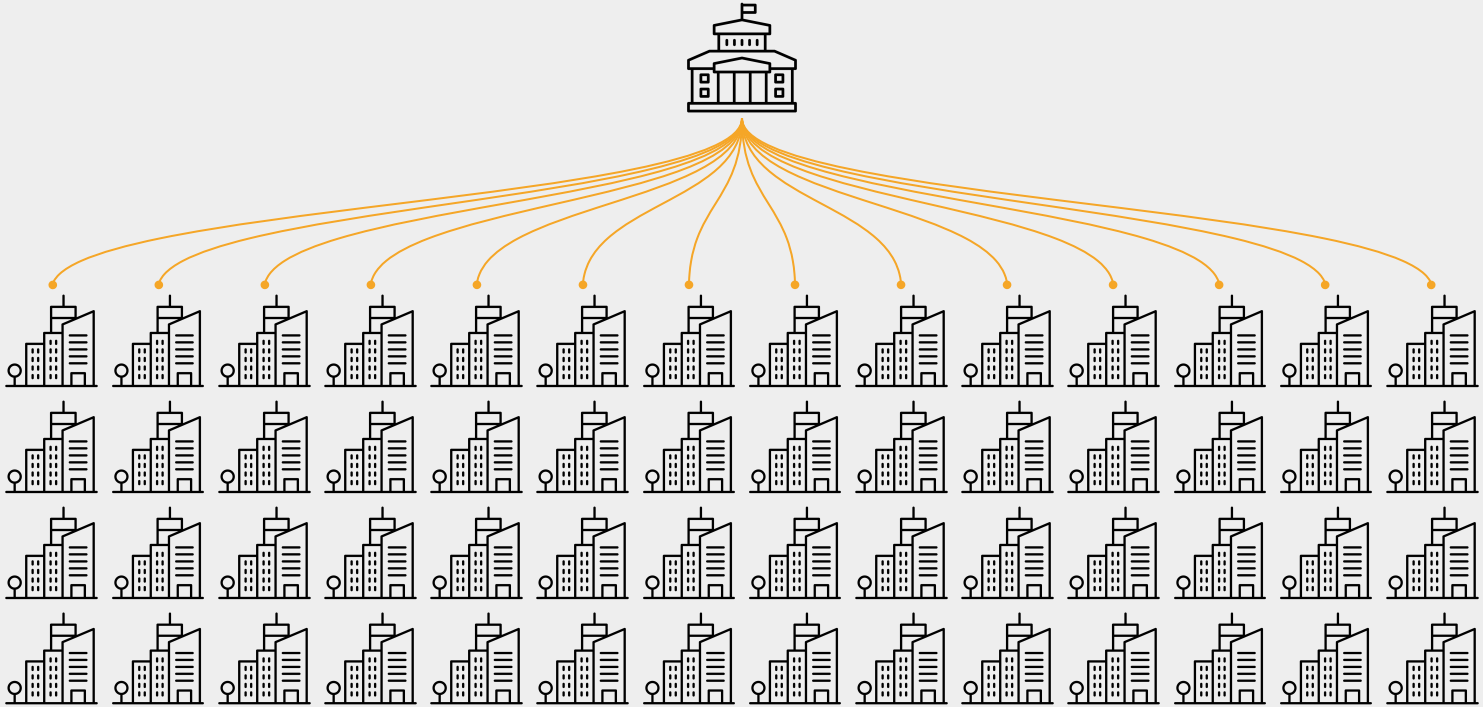
=

effort in which a **common type of data** is collected or exchanged across **disparate organizational boundaries**

In our distributed style of government, this happens **all the time.**





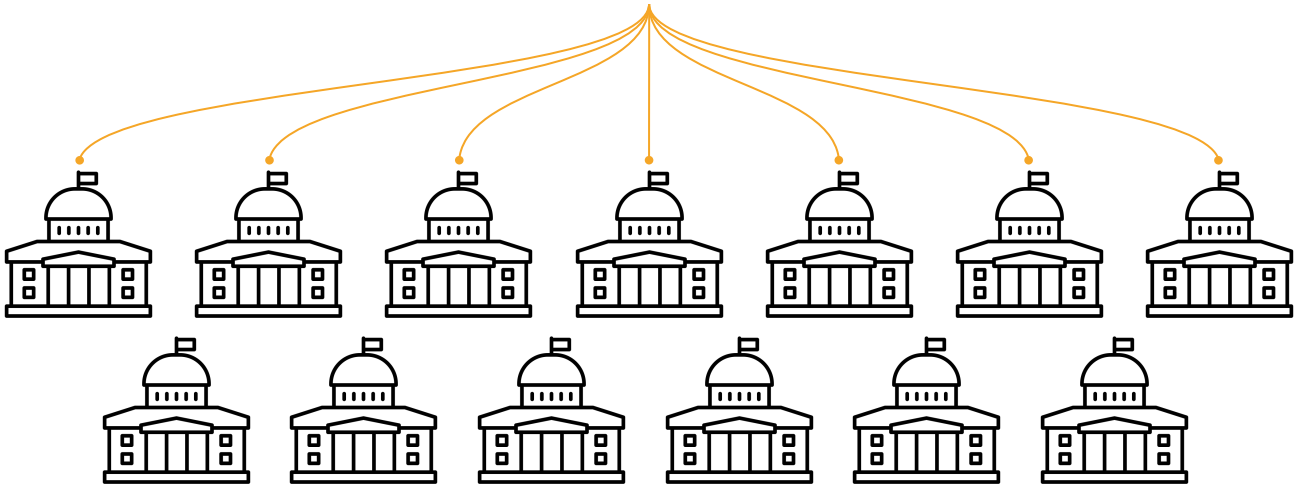


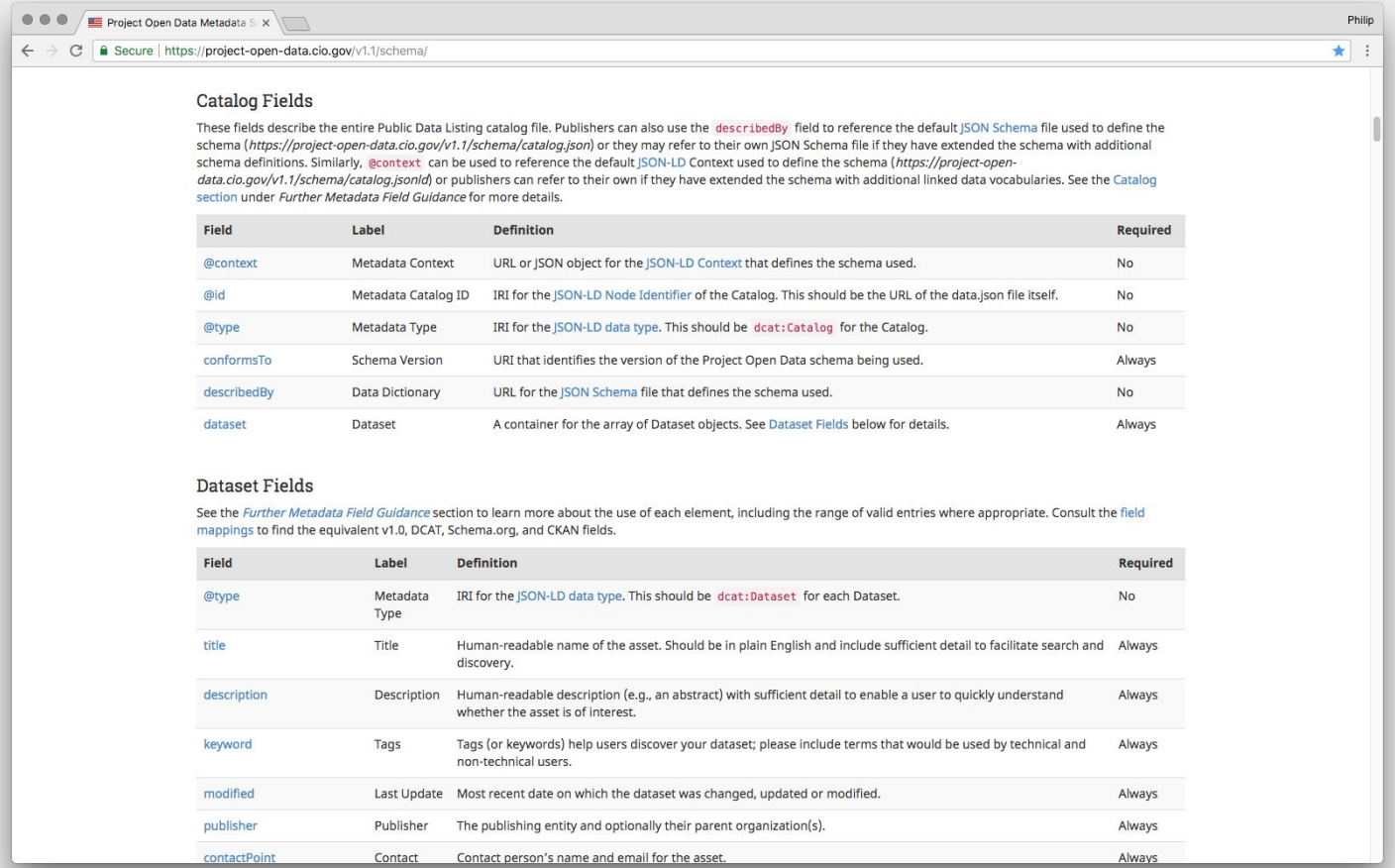
- Shape policy
- Inform budget decisions
- Track program performance
- Support operational efficiencies
- Make available to 3rd parties

Google Dataset Search

data.world

DATA.GOV





Catalog Fields

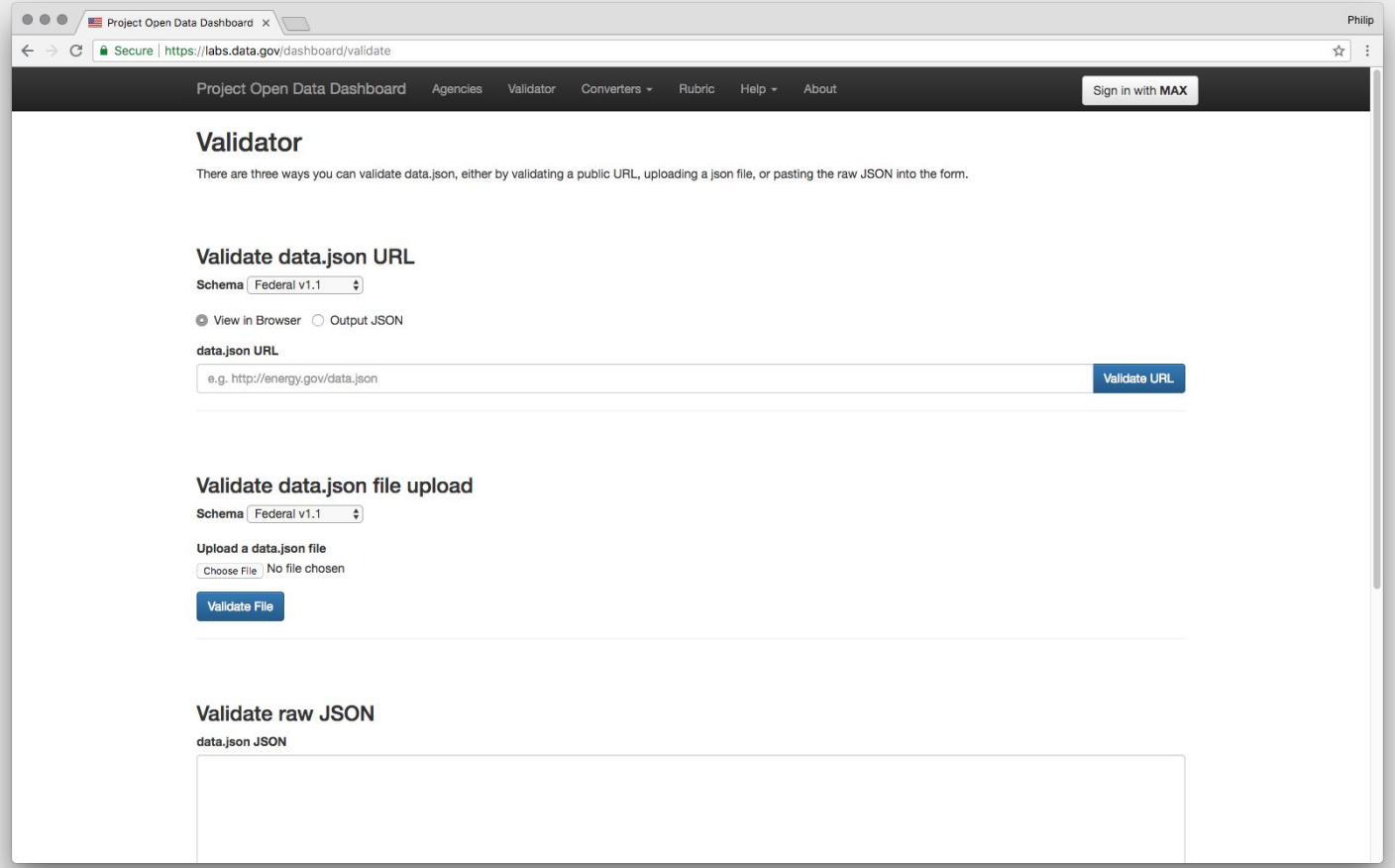
These fields describe the entire Public Data Listing catalog file. Publishers can also use the `describedBy` field to reference the default JSON Schema file used to define the schema (*https://project-open-data.cio.gov/v1.1/schema/catalog.json*) or they may refer to their own JSON Schema file if they have extended the schema with additional schema definitions. Similarly, `@context` can be used to reference the default JSON-LD Context used to define the schema (*https://project-open-data.cio.gov/v1.1/schema/catalog.jsonld*) or publishers can refer to their own if they have extended the schema with additional linked data vocabularies. See the Catalog section under *Further Metadata Field Guidance* for more details.

Field	Label	Definition	Required
@context	Metadata Context	URL or JSON object for the JSON-LD Context that defines the schema used.	No
@id	Metadata Catalog ID	IRI for the JSON-LD Node Identifier of the Catalog. This should be the URL of the data.json file itself.	No
@type	Metadata Type	IRI for the JSON-LD data type. This should be `dcat:Catalog` for the Catalog.	No
conformsTo	Schema Version	URI that identifies the version of the Project Open Data schema being used.	Always
describedBy	Data Dictionary	URL for the JSON Schema file that defines the schema used.	No
dataset	Dataset	A container for the array of Dataset objects. See Dataset Fields below for details.	Always

Dataset Fields

See the *Further Metadata Field Guidance* section to learn more about the use of each element, including the range of valid entries where appropriate. Consult the field mappings to find the equivalent v1.0, DCAT, Schema.org, and CKAN fields.

Field	Label	Definition	Required
@type	Metadata Type	IRI for the JSON-LD data type. This should be `dcat:Dataset` for each Dataset.	No
title	Title	Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery.	Always
description	Description	Human-readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether the asset is of interest.	Always
keyword	Tags	Tags (or keywords) help users discover your dataset; please include terms that would be used by technical and non-technical users.	Always
modified	Last Update	Most recent date on which the dataset was changed, updated or modified.	Always
publisher	Publisher	The publishing entity and optionally their parent organization(s).	Always
contactPoint	Contact	Contact person's name and email for the asset.	Always


Project Open Data Dashboard
Philip
Secure | https://labs.data.gov/dashboard/validate
Project Open Data Dashboard
Agencies
Validator
Converters
Rubric
Help
About
Sign in with MAX
Validator
There are three ways you can validate data.json, either by validating a public URL, uploading a json file, or pasting the raw JSON into the form.
Validate data.json URL
Schema
Federal v1.1
View in Browser
Output JSON
data.json URL
e.g. http://energy.gov/data.json
Validate URL
Validate data.json file upload
Schema
Federal v1.1
Upload a data.json file
Choose File
No file chosen
Validate File
Validate raw JSON
data.json JSON

Data integration models in other governments

Enterprise Service Bus
(centralized, not very scaleable)

National Interoperability Framework
(distributed, more scaleable)

THE CORE PROBLEM

Federated data efforts haven’t had adequate infrastructural support.

THE GOAL

Make managing federated data efforts in government easier

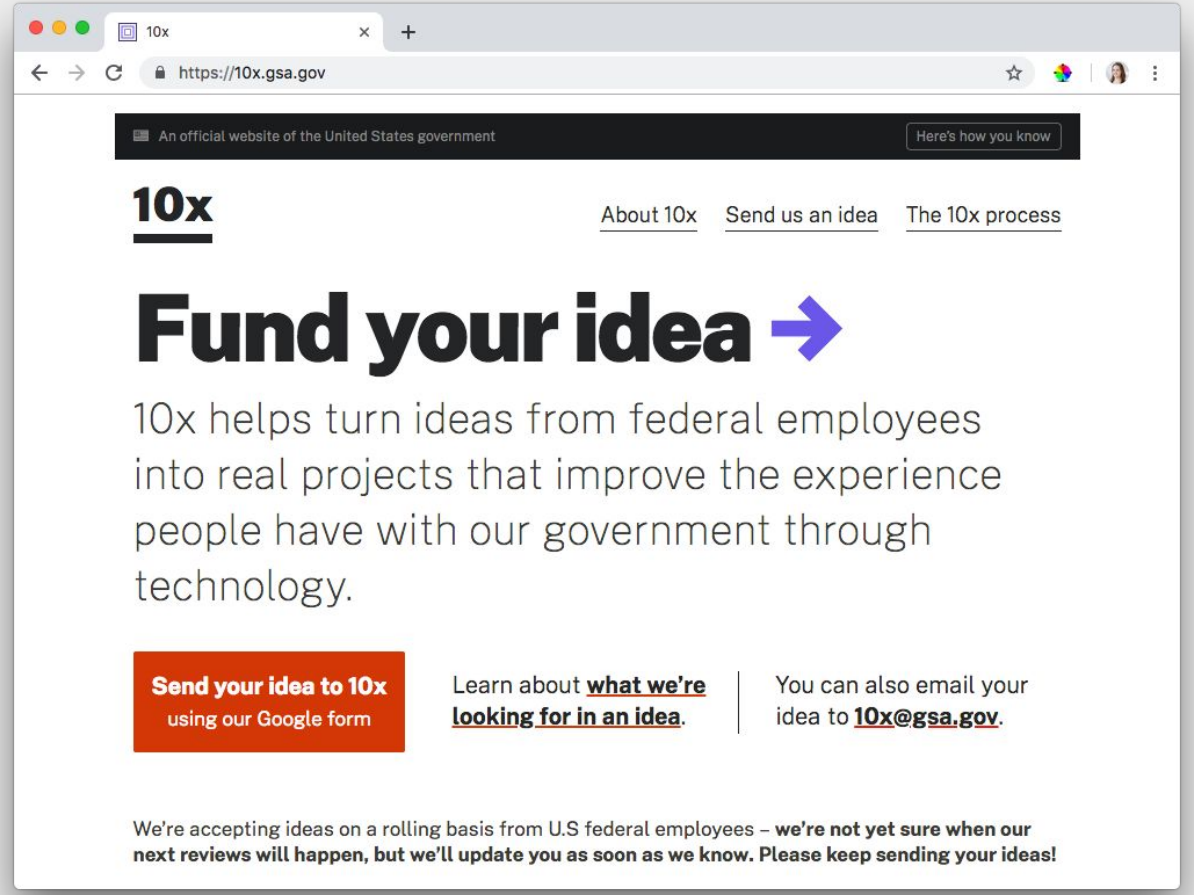
THE APPROACH

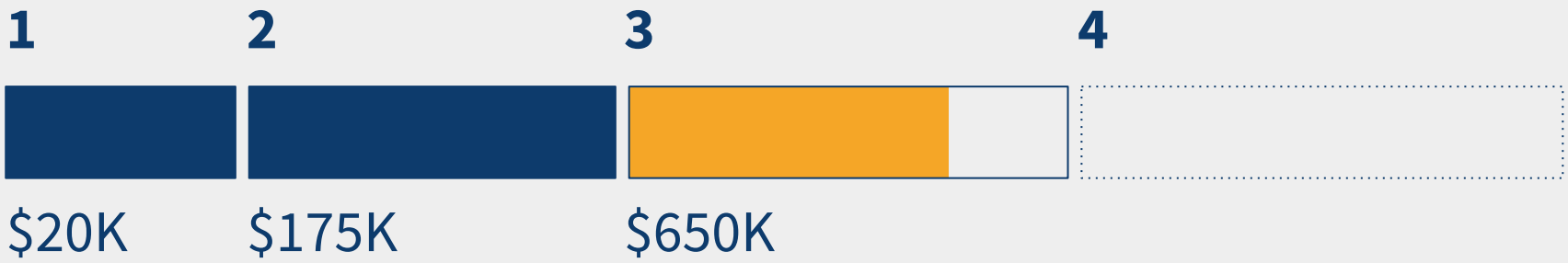
- Develop reusable tools
- Document and share repeatable processes

Overview

Our work

Our ask


10x
https://10x.gsa.gov
An official website of the United States government
Here's how you know
10x
About 10x
Send us an idea
The 10x process
Fund your idea
10x helps turn ideas from federal employees into real projects that improve the experience people have with our government through technology.
Send your idea to 10x using our Google form
Learn about what we're looking for in an idea.
You can also email your idea to 10x@gsa.gov.
We're accepting ideas on a rolling basis from U.S federal employees – we're not yet sure when our next reviews will happen, but we'll update you as soon as we know. Please keep sending your ideas!


1
2
3
4
$20K
$175K
$650K

PHASE ONE: October 2017–January 2018

Understand the problem space

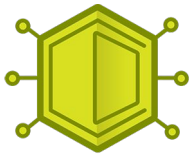
We interviewed examples of successful federated data projects

DATA Act

BETTER DATA. BETTER DECISIONS. BETTER GOVERNMENT.

DATA.GOV

code_



the Open Civic Data project

Open Data Philly

NIEM

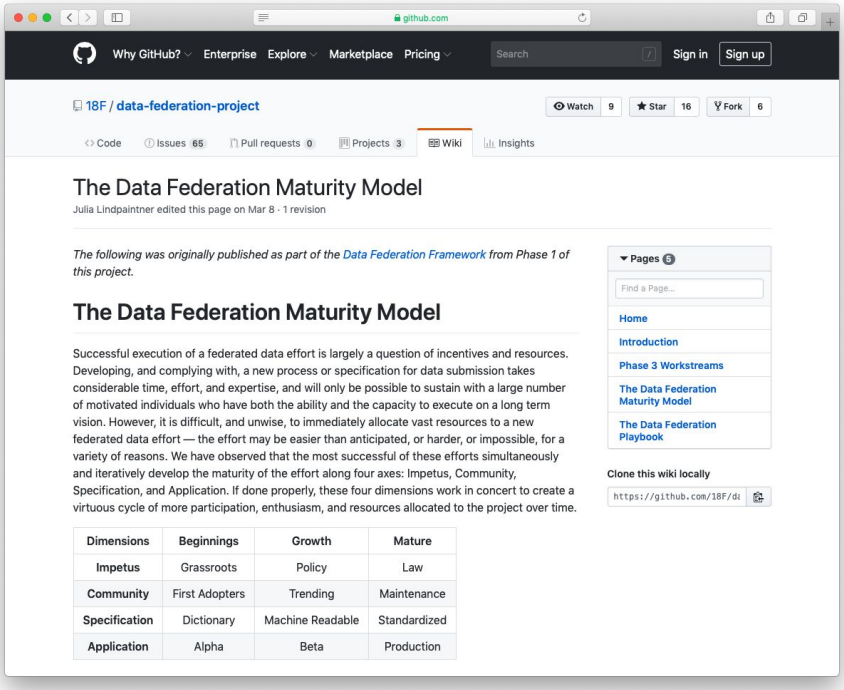
OPEN REFERRAL

FGDC

FEDERAL GEOGRAPHIC DATA COMMITTEE

Open311

We drafted a maturity model and playbook



github.com

Why GitHub? Enterprise Explore Marketplace Pricing Search Sign in Sign up

18F / data-federation-project

Watch 9 Star 16 Fork 6

Code Issues 65 Pull requests 0 Projects 3 Wiki Insights

The Data Federation Maturity Model

Julia Lindpaintner edited this page on Mar 8 · 1 revision

The following was originally published as part of the Data Federation Framework from Phase 1 of this project.

The Data Federation Maturity Model

Successful execution of a federated data effort is largely a question of incentives and resources. Developing, and complying with, a new process or specification for data submission takes considerable time, effort, and expertise, and will only be possible to sustain with a large number of motivated individuals who have both the ability and the capacity to execute on a long term vision. However, it is difficult, and unwise, to immediately allocate vast resources to a new federated data effort — the effort may be easier than anticipated, or harder, or impossible, for a variety of reasons. We have observed that the most successful of these efforts simultaneously and iteratively develop the maturity of the effort along four axes: Impetus, Community, Specification, and Application. If done properly, these four dimensions work in concert to create a virtuous cycle of more participation, enthusiasm, and resources allocated to the project over time.

Dimensions	Beginnings	Growth	Mature
Impetus	Grassroots	Policy	Law
Community	First Adopters	Trending	Maintenance
Specification	Dictionary	Machine Readable	Standardized
Application	Alpha	Beta	Production

Pages 5

Find a Page...

Home

Introduction

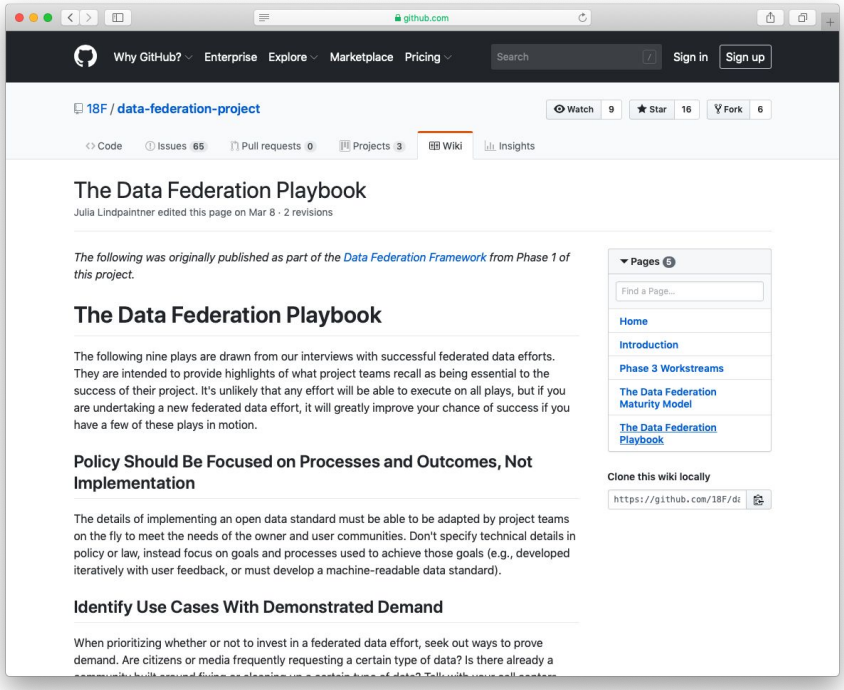
Phase 3 Workstreams

The Data Federation Maturity Model

The Data Federation Playbook

Clone this wiki locally

https://github.com/18F/da



github.com

Why GitHub? Enterprise Explore Marketplace Pricing Search Sign in Sign up

18F / data-federation-project

Watch 9 Star 16 Fork 6

Code Issues 65 Pull requests 0 Projects 3 Wiki Insights

The Data Federation Playbook

Julia Lindpaintner edited this page on Mar 8 · 2 revisions

The following was originally published as part of the Data Federation Framework from Phase 1 of this project.

The Data Federation Playbook

The following nine plays are drawn from our interviews with successful federated data efforts. They are intended to provide highlights of what project teams recall as being essential to the success of their project. It's unlikely that any effort will be able to execute on all plays, but if you are undertaking a new federated data effort, it will greatly improve your chance of success if you have a few of these plays in motion.

Policy Should Be Focused on Processes and Outcomes, Not Implementation

The details of implementing an open data standard must be able to be adapted by project teams on the fly to meet the needs of the owner and user communities. Don't specify technical details in policy or law, instead focus on goals and processes used to achieve those goals (e.g., developed iteratively with user feedback, or must develop a machine-readable data standard).

Identify Use Cases With Demonstrated Demand

When prioritizing whether or not to invest in a federated data effort, seek out ways to prove demand. Are citizens or media frequently requesting a certain type of data? Is there already a

Pages 5

Find a Page...

Home

Introduction

Phase 3 Workstreams

The Data Federation Maturity Model

The Data Federation Playbook

Clone this wiki locally

https://github.com/18F/da

PHASE TWO: April–June 2018

Pick a use case & build a prototype

We partnered with the Food & Nutrition Service (FNS)

OMB APPROVED NO. 0584-0026
Expiration Date: 04/30/2016

Department of Agriculture, Food and Nutrition Service

School Food Authority (SFA) Verification Collection Report

State agencies must report the information on this form ANNUALLY for each SFA with schools operating the National School Lunch Program (NSLP) and/or the School Breakfast Program (SBP).
All SFAs, including SFAs with all schools exempt from verification requirements, must complete applicable sections.

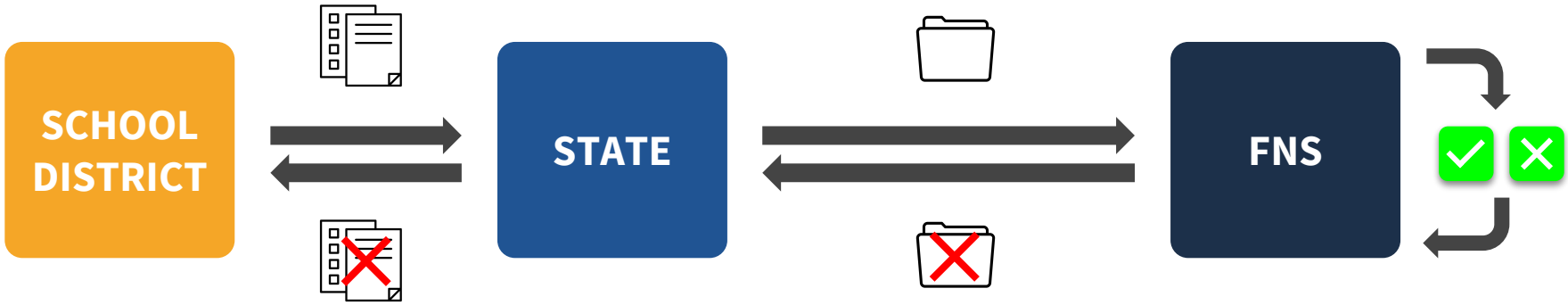
According to the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number. The valid OMB number for this collection is 0584-0026. The time required to complete this information collection is 45 minutes per response, including the time to review instructions, search existing data resources, gather the data needed and complete and review the information collection.

State Agency Name:	SFA ID#:	Type of SFA: ☐ Public ☐ Nonprofit/Private	School Year: From: 20 To: 20
SFA Name:		SFA City:	SFA Zip code: ☐☐☐☐☐

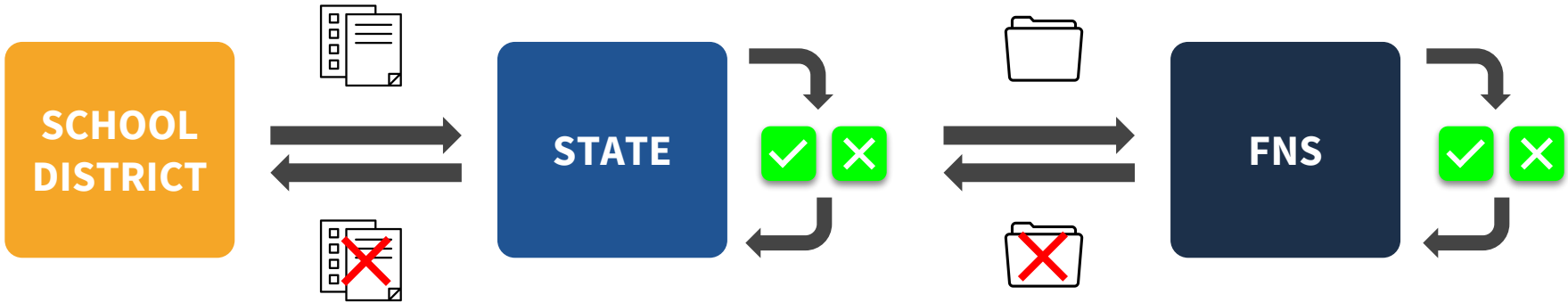
		All SFAs must report Section 1	A. Number of Schools OR Institutions	B. Number of Students
Section 1	Total Schools, Residential Child Care Institutions (RCCIs) and Enrolled Students	1-1: Total schools *(Do not include RCCIs)*:	1-1A	1-1B
		1-2: Total RCCIs *(Do not include schools counted in 1-1)*:	1-2A	1-2B
		1-2a: RCCIs with day students *(Report ONLY day students in 1-2aB)*:	1-2aA	1-2aB
		1-2b: RCCIs with NO day students:	1-2bA	1-2bB

		ONLY SFAs with alternate provisions must report Section 2	A. Number of Schools AND Institutions	B. Number of Students
Section 2	SFAs with schools operating alternate provisions	2-1: Operating Provision 2/3 in a BASE year for NSLP and SBP:	2-1A	2-1B
		2-2: Operating Provision 2/3 in a NON BASE year for NSLP and SBP:	2-2A	2-2B
		2-2a: Provision 2/3 students reported as FREE in a NON BASE year:		2-2aB
		2-2b: Provision 2/3 students reported as REDUCED PRICE in a NON BASE year:		2-2bB
		2-3: Operating the Community Eligibility Option:	2-3A	2-3B

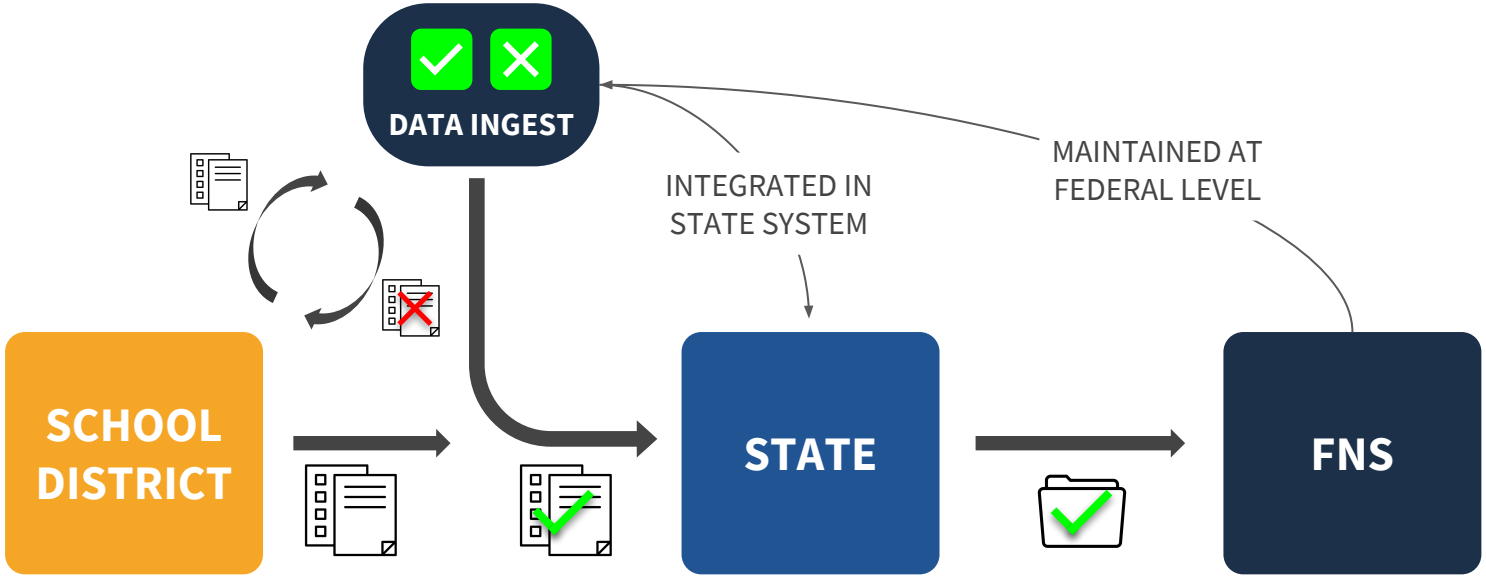
PREVIOUS STATE



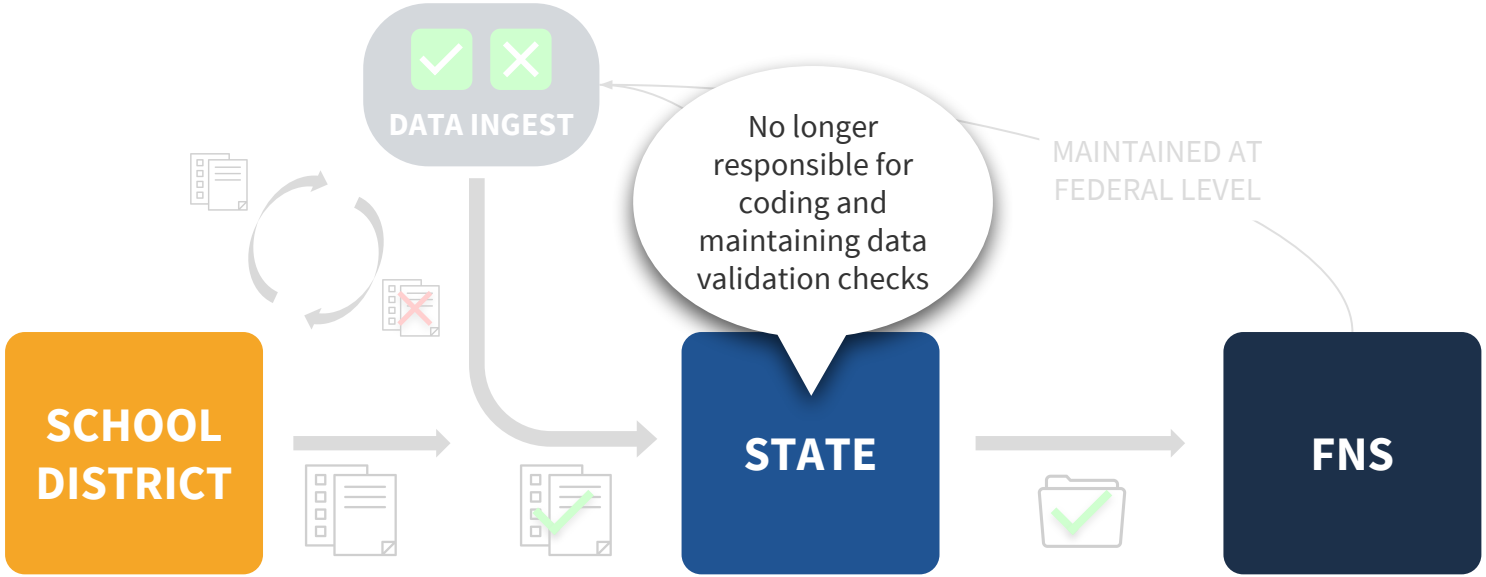
CURRENT STATE



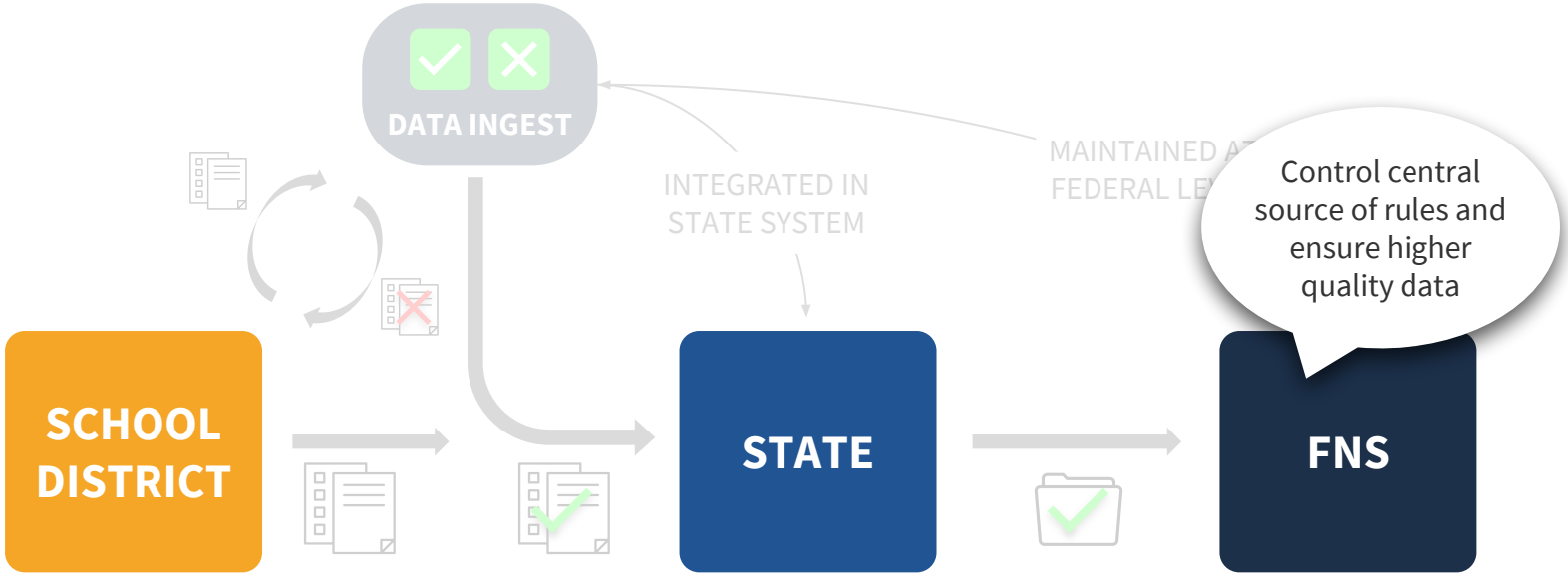
FUTURE STATE



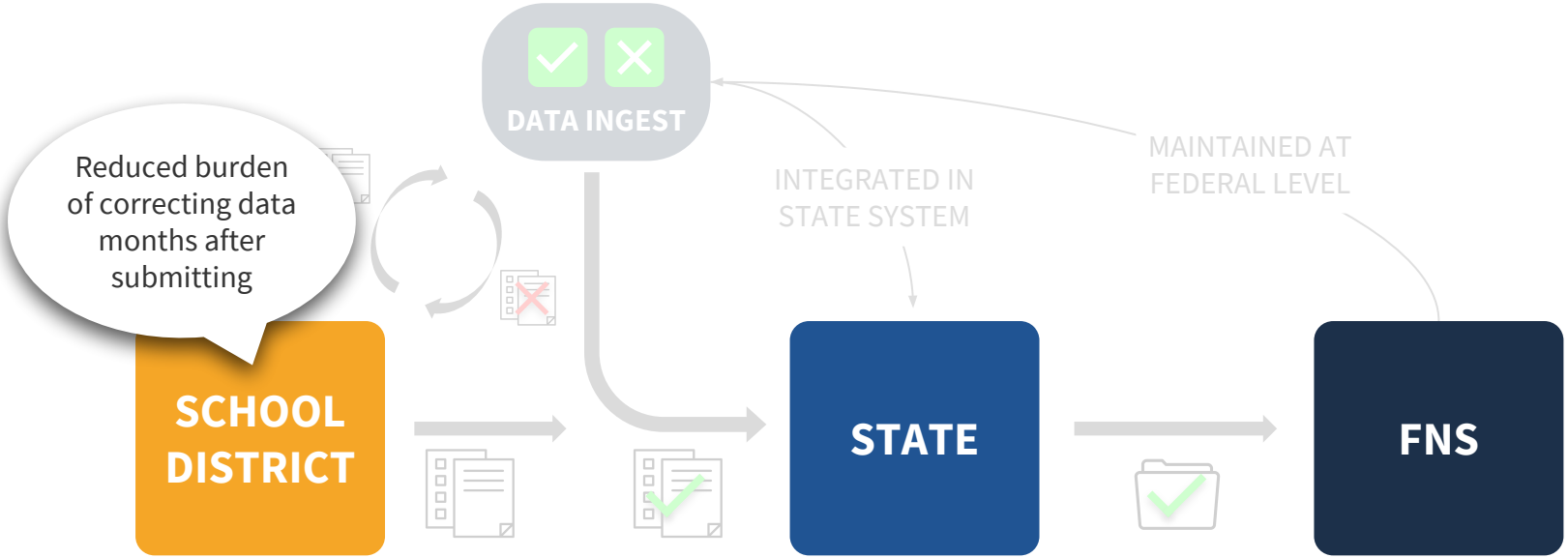
BENEFITS ALL LEVELS



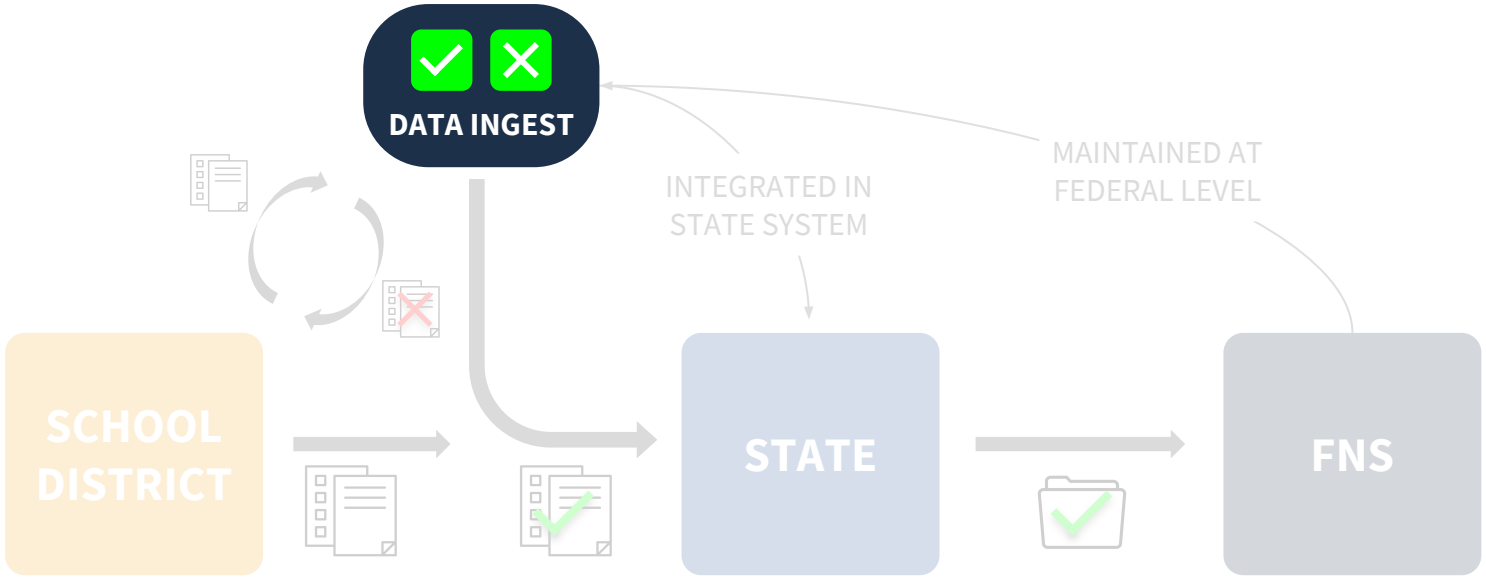
BENEFITS ALL LEVELS



BENEFITS ALL LEVELS



OUR TOOL



See the code on GitHub.

Demo!

PHASE THREE: December 2018–May 2019

Make it real & make it last

In Phase 3, we are going both deep and broad

- Set FNS up for success
- Demonstrate reusability of the tool with other use cases and data sets

- Identify other opportunities for reusable tooling or repeatable processes
- Develop a plan to sustain this project beyond 10x funding

Overview

Our work

Our ask

THE APPROACH

- Develop reusable tools
- Document and share repeatable processes

WE STILL NEED TO LEARN MORE ABOUT

Challenges facing federated data efforts

—

Opportunities for tools & resources

- Aligning incentives
- Translating between data formats
- Specifying data standards
- Driving adoption of standards
- Making data collection user-friendly
- Implementing reliable validation

YOU CAN HELP

Share data- and rulesets with us

Let us know about other one-off validation tools that exist

CONTACT THE TEAM

data-federation@gsa.gov

FOLLOW OUR WORK ON GITHUB

github.com/18F/data-federation-project

VISIT THE WEBSITE

federation.data.gov

WE STILL NEED TO LEARN MORE ABOUT

Challenges facing federated data efforts

—

Opportunities for tools & resources

Thank you!

Noun Project icons by Made x Made, ATOM & DewDrops