

U.S. Data Federation

10x Phase 4 Funding Request

13 JUNE 2019

“Data is a valuable national resource and a strategic asset to the U.S. Government, its partners, and the public.”

— PROJECT OPEN DATA







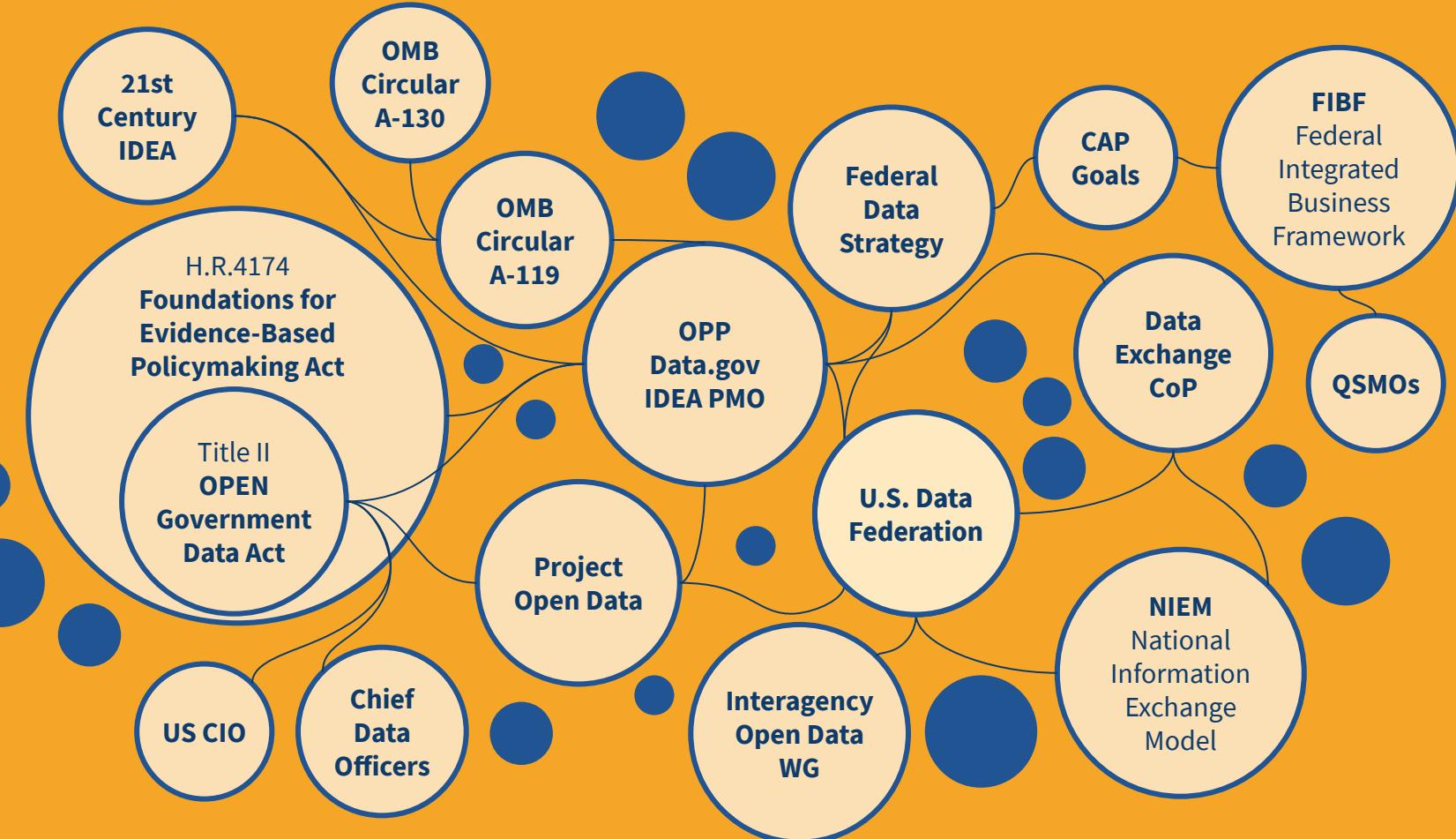
Vineland Pl

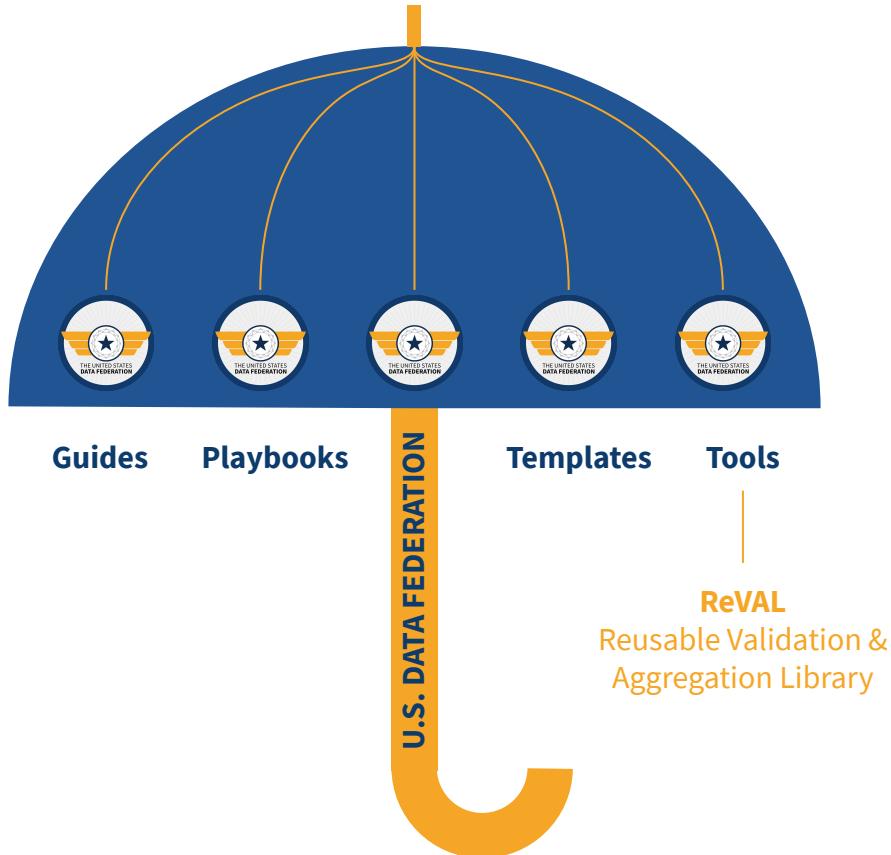
TO
EXITS
35 & 94

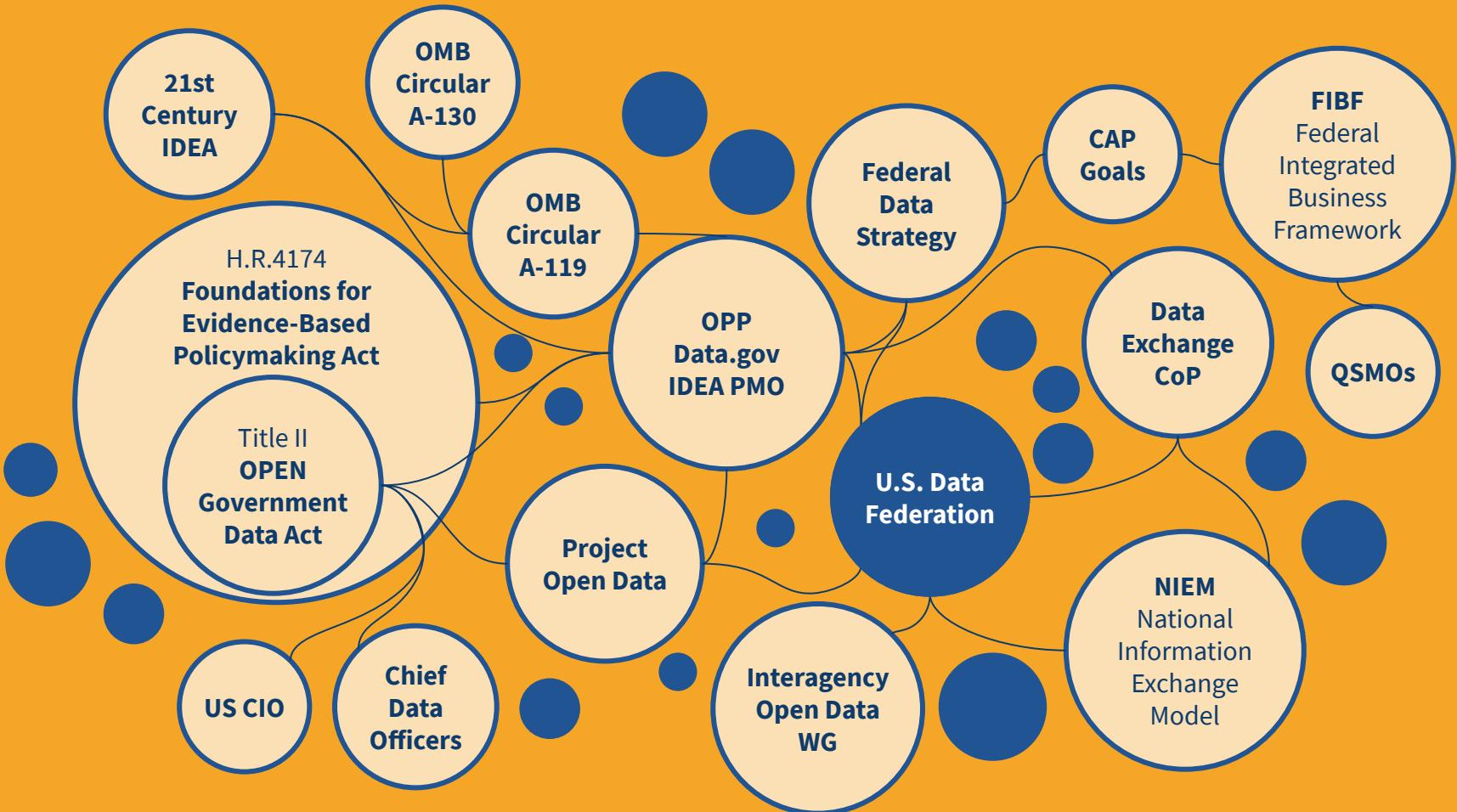
FEBAR ONLY

“The Federal Government needs a robust, integrated approach to using data to deliver on mission, serve customers, and steward resources.”

— CROSS-AGENCY PRIORITY GOAL ON
DATA, ACCOUNTABILITY & TRANSPARENCY







U.S. DATA FEDERATION MISSION

We promote government-wide systemic investment and capacity building to support distributed data management challenges, data interoperability, and broader data standards activities.

U.S. DATA FEDERATION APPROACH

We curate, create, and connect
reusable tools and repeatable processes
to support federated data efforts.

U.S. DATA FEDERATION APPROACH

We curate, create, and connect
reusable tools and repeatable processes
to support federated data efforts.

This is bigger than a single 10x project.

WHAT WE'VE ACCOMPLISHED SO FAR

1

Developed, scaled, and demonstrated demand for
ReVAL (Reusable Validation & Aggregation Library)

2

Validated a model for the identifying common needs and
creating reusable resources for data efforts

3

Began building a community around a shared need for
knowledge-sharing across data efforts in government

U.S. Data Federation



ReVAL
Reusable Validation &
Aggregation Library

WHAT WE WANT TO DO NEXT

1

Ensure ReVAL is maximally reusable by completing known technical enhancements

2

See ReVAL into production by continuing to collaborate with existing pilot partners

3

Ensure the discoverability of ReVAL by connecting it to other resources for data practitioners

WHAT WE WANT TO DO NEXT

1

Ensure ReVAL is maximally reusable by completing known technical enhancements

2

See ReVAL into production by continuing to collaborate with existing pilot partners

3

Ensure the discoverability of ReVAL by connecting it to other resources for data practitioners

“The Office of Management and Budget must develop and maintain an **online repository** of **tools, best practices, and schema standards** to facilitate the adoption of open data practices.”

— EVIDENCE ACT; OPEN GOVERNMENT DATA ACT

“The repository may include more detailed practice descriptions, case studies that demonstrate the practices ‘in action,’ and tool kits for implementing individual or groups of practices.”

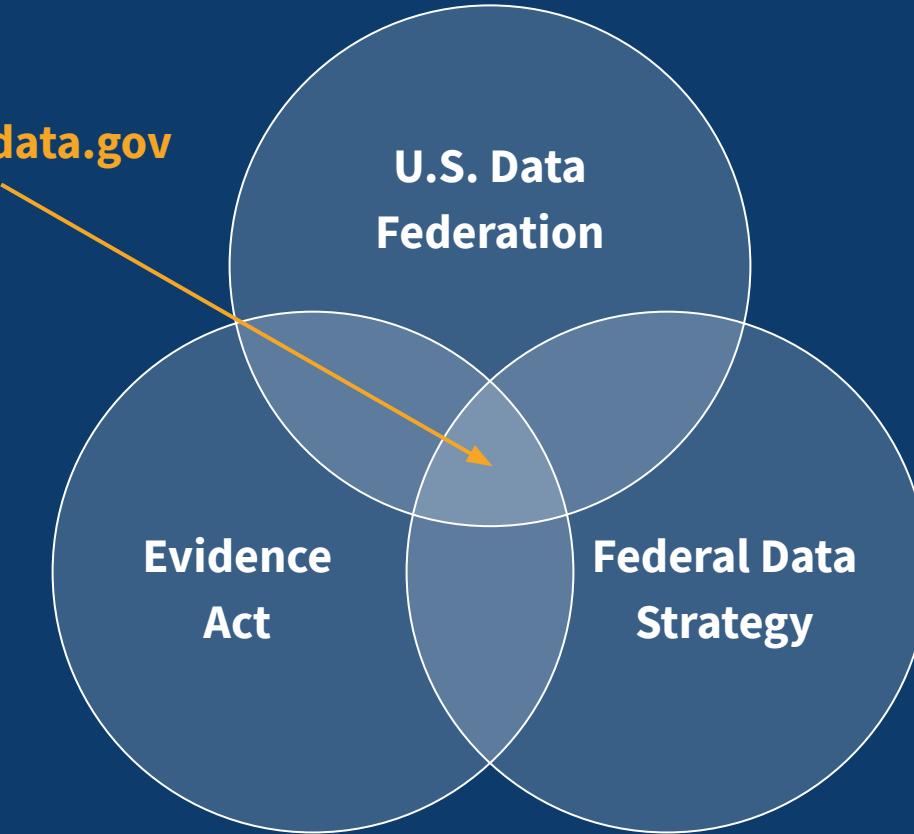
— FEDERAL DATA STRATEGY YEAR 1 ACTION PLAN

resources.data.gov

**U.S. Data
Federation**

**Evidence
Act**

**Federal Data
Strategy**



A UNIQUE OPPORTUNITY

We are seeking Phase 4 funding to leverage the completion of ReVAL and the momentum of the U.S. Data Federation work to support a long-term vision and strategic plan for a user-centered, maximally-effective resources.data.gov.

Overview of the U.S. Data Federation

BACKGROUND

Policies, initiatives, and citizen-facing products and services depend on aggregating and harmonizing data from disparate government sources.

OPPORTUNITY

Many of these federated data efforts face common requirements and common challenges, but lack common resources.

federated data effort



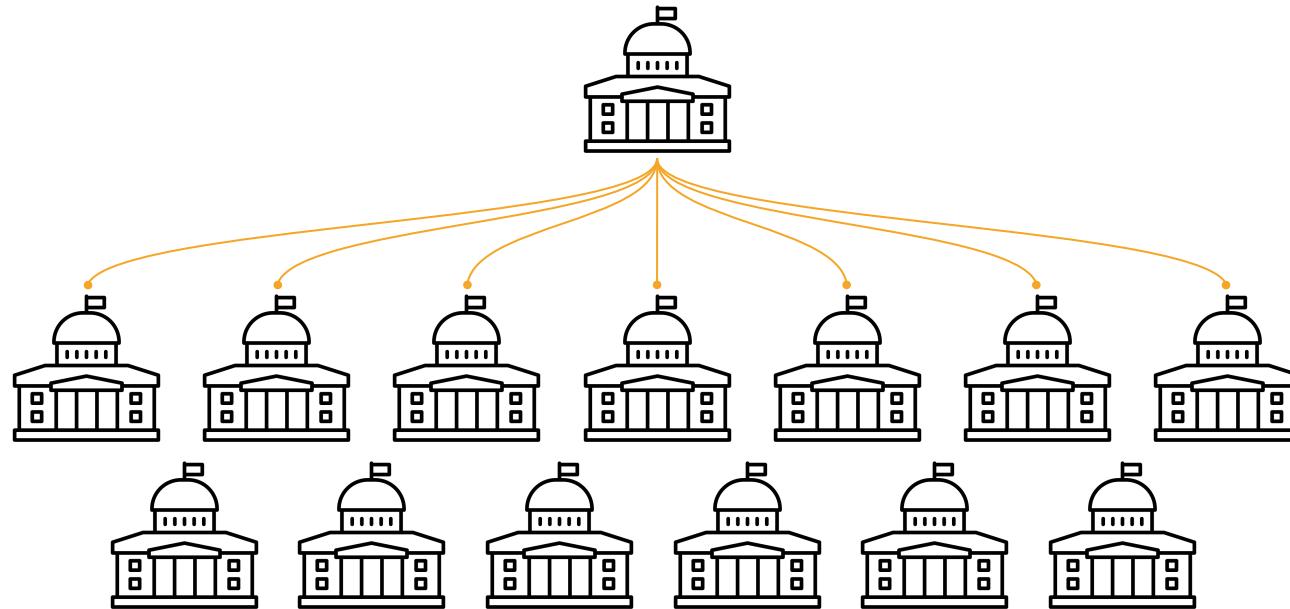
effort in which a **common type of data** is
collected or exchanged across
disparate organizational boundaries

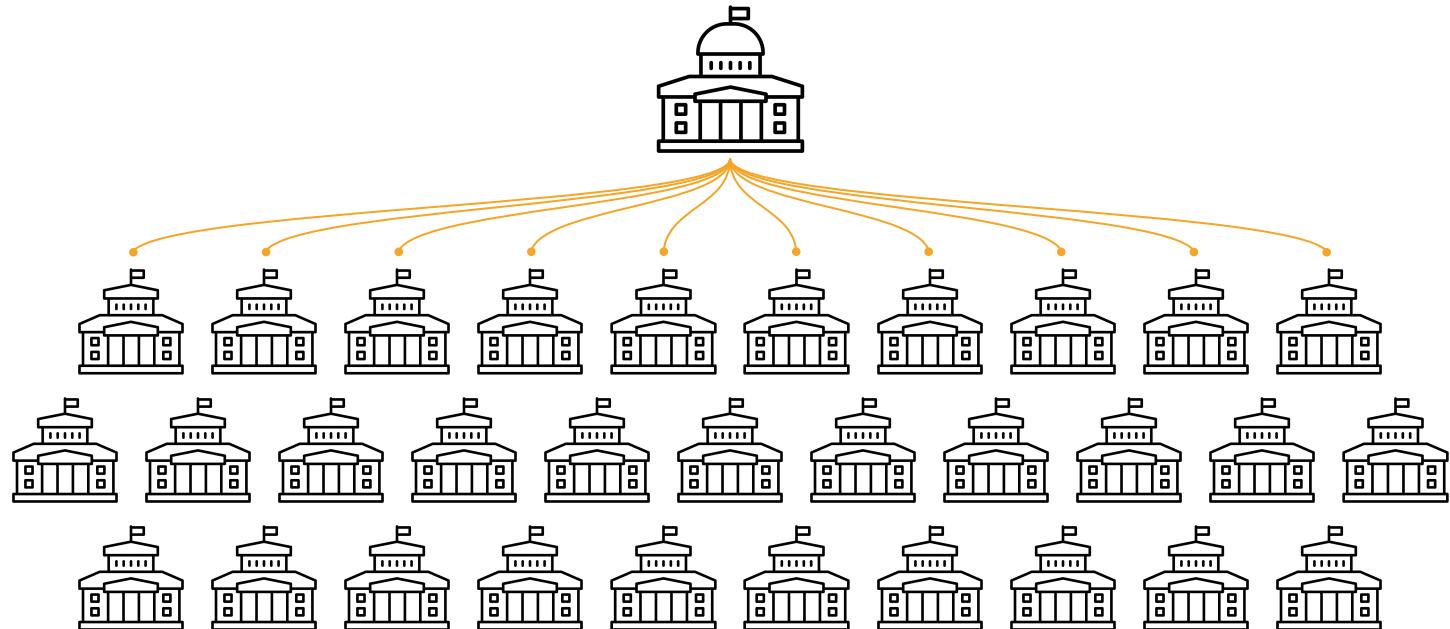
Efforts involving multiple data sources face organizational and technical challenges:

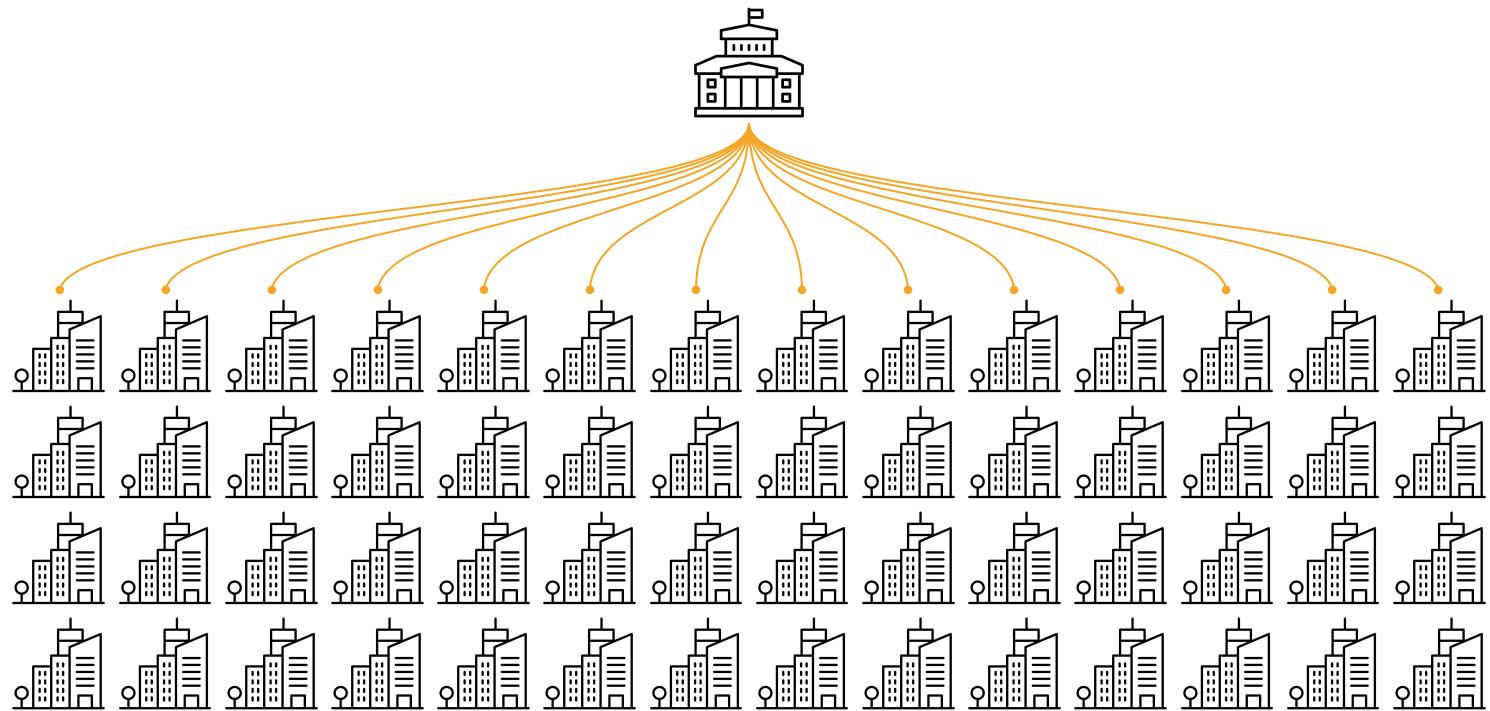
- Aligning incentives across entities
- Translating between data formats
- Specifying data standards
- Driving adoption of standards
- Making data collection user-friendly
- Implementing reliable validation



**Federated data efforts are critical to
the success of programs and services
throughout our distributed government.**







PROBLEM

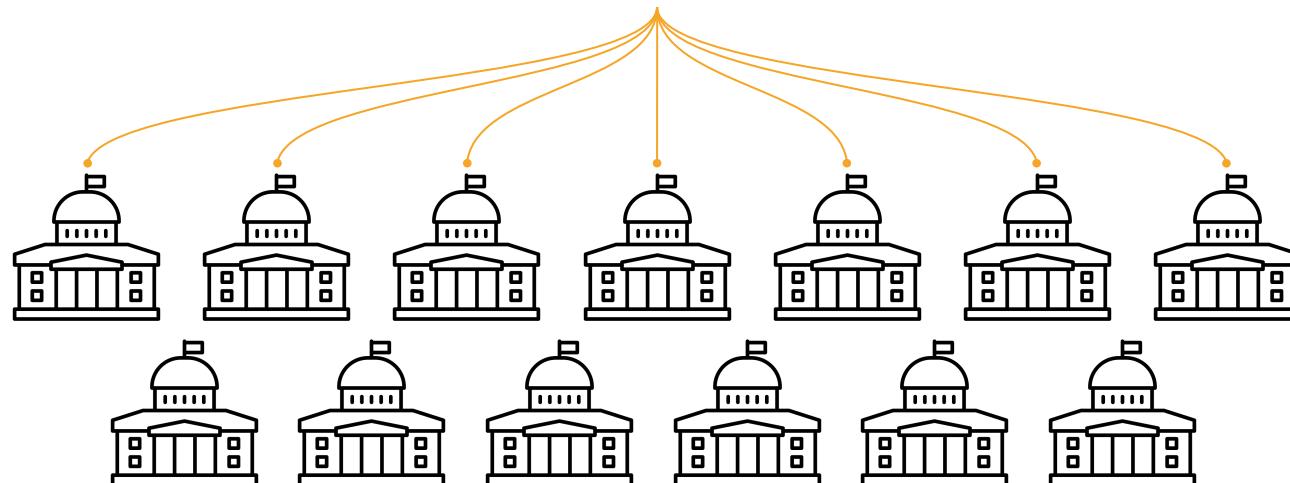
Federated data efforts haven't received the systemic investigation nor the infrastructural support they deserve.

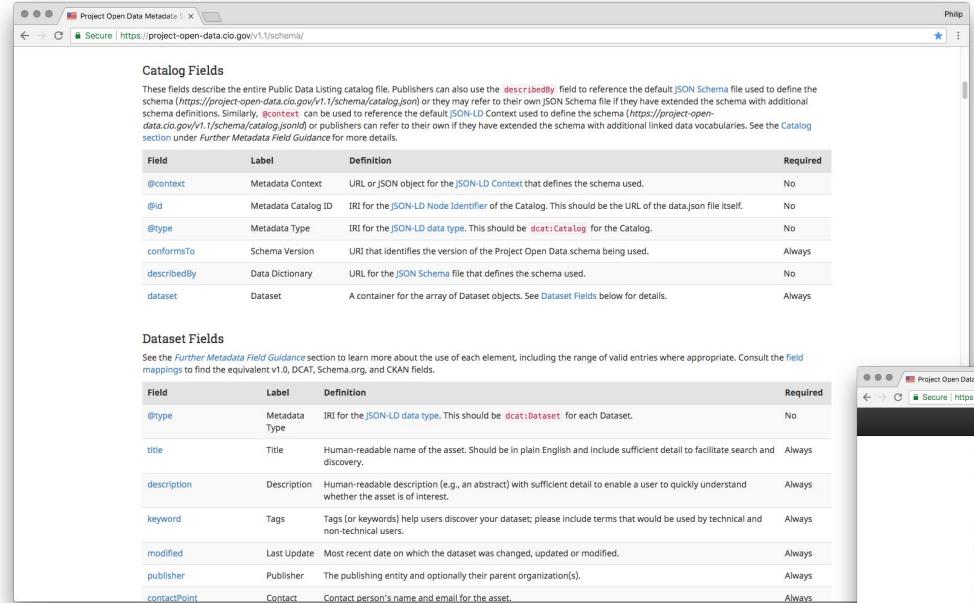
Google Dataset Search



data.world

 **DATA.GOV**



Project Open Data Metadata

Catalog Fields

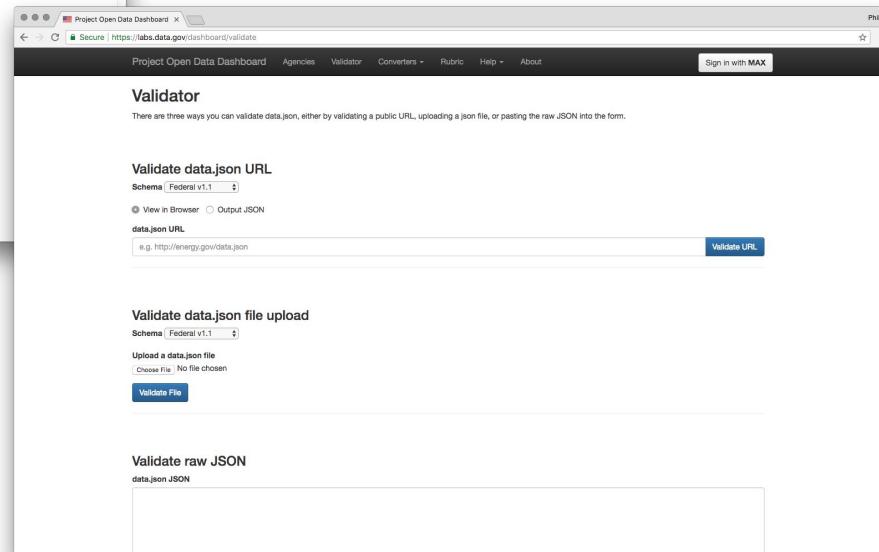
These fields describe the entire Public Data Listing catalog file. Publishers can also use the `describedBy` field to reference the default JSON Schema file used to define the schema (<https://project-open-data.cio.gov/v1.1/schema/catalog.json>) or they may refer to their own JSON Schema file if they 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
<code>@context</code>	Metadata Context	URL or JSON object for the JSON-LD Context that defines the schema used.	No
<code>@id</code>	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
<code>@type</code>	Metadata Type	IRI for the JSON-LD data type. This should be <code>dcat:Catalog</code> for the Catalog.	No
<code>conformsTo</code>	Schema Version	URI that identifies the version of the Project Open Data schema being used.	Always
<code>describedBy</code>	Data Dictionary	URL for the JSON Schema file that defines the schema used.	No
<code>dataset</code>	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
<code>@type</code>	Metadata Type	IRI for the JSON-LD data type. This should be <code>dcat:Dataset</code> for each Dataset.	No
<code>title</code>	Title	Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery.	Always
<code>description</code>	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
<code>keyword</code>	Tags	Tags (or keywords) help users discover your dataset; please include terms that would be used by technical and non-technical users.	Always
<code>modified</code>	Last Update	Most recent date on which the dataset was changed, updated or modified.	Always
<code>publisher</code>	Publisher	The publishing entity and optionally their parent organization(s).	Always
<code>contactPoint</code>	Contact	Contact person's name and email for the asset.	Always

Project Open Data Dashboard

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 data.json file upload

Schema: [Federal v1.1](#)

Upload a data.json file

No file chosen

Validate raw JSON

data.json JSON

AMBITION

The U.S. Data Federation seeks to provide an integrated resource for data practitioners that curates and creates reusable tools and documents and shares repeatable processes.

Phase 1: We conducted interviews, built a website to share their stories, and drafted a maturity model and playbook.



The screenshot shows two views of the Data Federation framework:

- GitHub Project Page:** A screenshot of a GitHub repository titled "The Data Federation Playbook". It shows basic project statistics: 18F / data-federation-project, 9 Watchers, 16 Stars, 6 Forks, and 6 Issues.
- Website:** A screenshot of the official website at federation.data.gov. The header includes the United States Data Federation logo. The main content area features a blue banner with the text: "The United States Data Federation is dedicated to making it easier to collect, combine, and exchange data across government through reusable tools and repeatable processes." Below this, there's a section titled "Getting good, reliable data is hard. Getting data from different sources makes this even harder." followed by a detailed paragraph about the challenges of data integration.

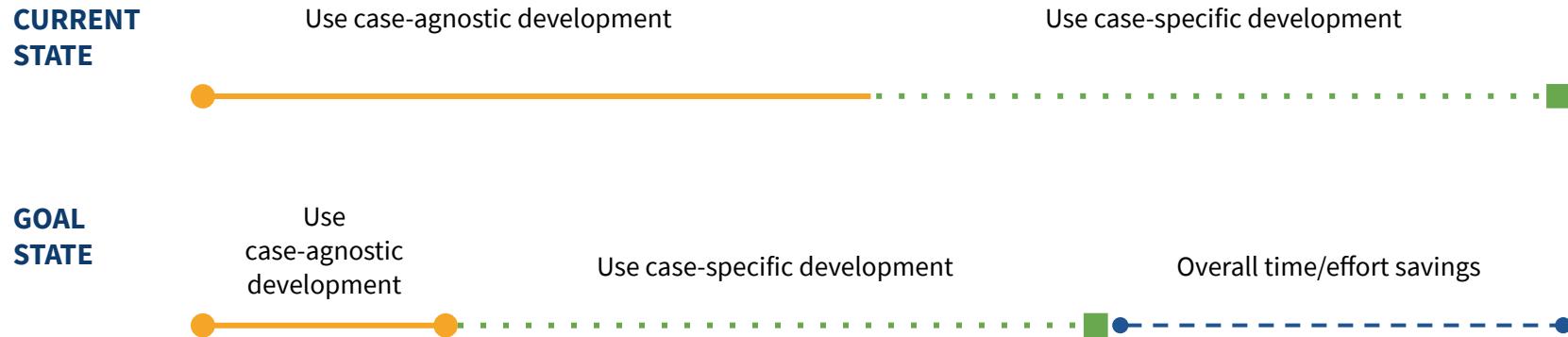
The more customized a tool is, the less reusable it is.

To identify the boundaries between use case-agnostic and use case-specific development, we would need to work with multiple partners.



Focus on use case-agnostic tools could accelerate progress towards custom solutions.

It also ensures that the tool is maximally reusable.



Overview of ReVAL

Reusable Validation & Aggregation Library





Improve data quality by streamlining FNS-742 data validation

Why the FNS-742 matters

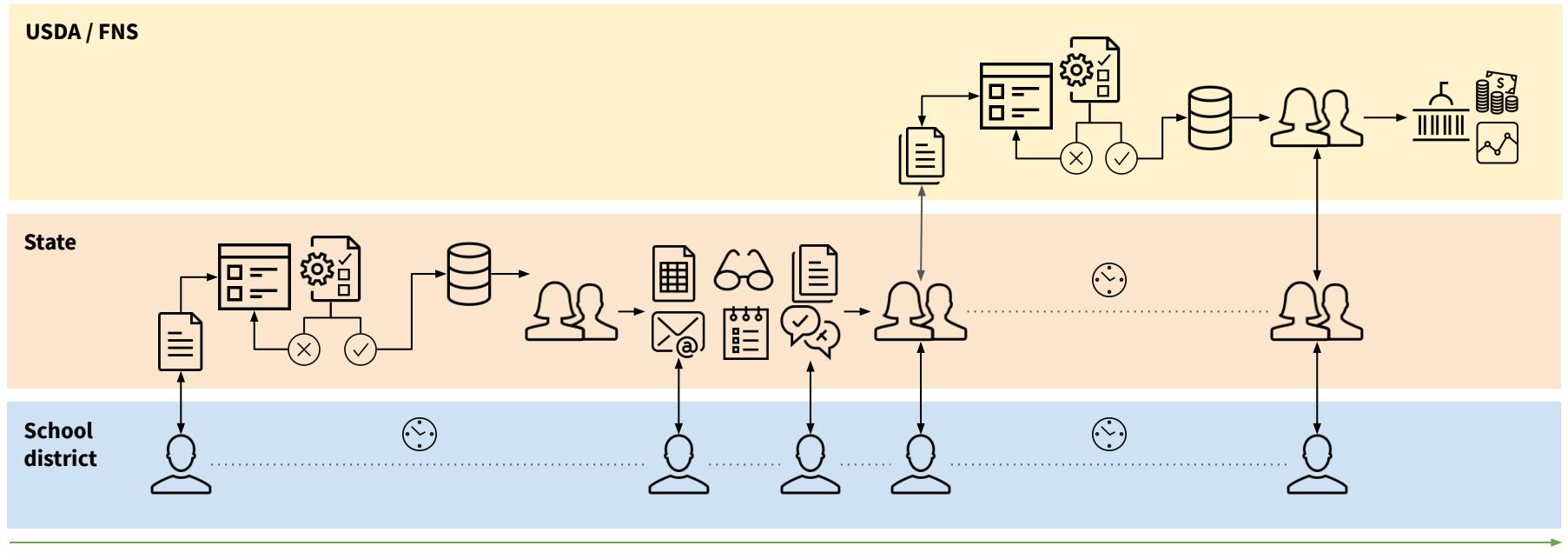
- Reports on an eligibility verification process that a percentage of households are asked to go through
- Collects data from approx. 20,000 school districts across the country and is the most important data they currently receive from the school district level
- Dataset is a critical source of administrative data used for research, Congressional advice, policy, etc.

OMB APPROVED NO. 0584-0026
Expiration Date: 04/30/2019

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: <input type="checkbox"/> Public <input type="checkbox"/> Nonprofit/Private	School Year: From: 20 To: 20	
SFA Name:		SFA CIVY:	SFA Zip code: <input type="text"/>		
Section 1					
All SFAs must report Section 1					
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					
Section 2					
ONLY SFAs with alternate provisions must report Section 2					
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					
2-4: Operating other alternatives for NSLP and SBP:					
2-4A 2-4B					
2-5: Operating an alternate provision(s) for only SBP or only NSLP:					
2-5A 2-5B					
Section 3					
ALL SFAs must report Section 3 or check box 3-4 if applicable					
3-1: <input checked="" type="checkbox"/> Check this box only if all schools and/or RCCIs in the SFA were not required to perform direct certification with SNAP through the Supplemental Nutrition Assistance Program (SNAP). Do not include students certified with SNAP through the letter method.					
3-2B					
3-2: Students directly certified through the Supplemental Nutrition Assistance Program (SNAP). Do not include students certified with SNAP through the letter method.					
3-2A					
3-3: Students directly certified through other programs: Include those directly certified through Temporary Assistance for Needy Families (TANF), Food Distribution Program on Indian Reservations (FDPIR), or Medicaid (if applicable); those documented as homeless, migrant, runaway, foster, Head Start, Pre-K Even Start, or non-applicant but approved by local offices. DO NOT include SNAP students already reported in 3-2.					
3-3B					
3-4: Students certified categorically FREE eligible through SNAP letter method: Include students certified for free meals through the family providing a letter from the SNAP agency.					
3-4B					
Section 4					
ALL SFAs collecting applications must report Section 4					
4-1: Approved as categorically FREE Eligible: Based on those providing documentation (e.g. a case number for SNAP, TANF, FDPIR or an application)					
4-1A 4-1B					
4-2: Approved as FREE eligible: Based on household size and income information					
4-2A 4-2B					
4-3: Approved as REDUCED PRICE eligible: Based on household size and income information					
4-3A 4-3B					
T-1: Total FREE Eligible Students Reported:			T-1	T-2: Total REDUCED PRICE Eligible Students Reported:	T-2

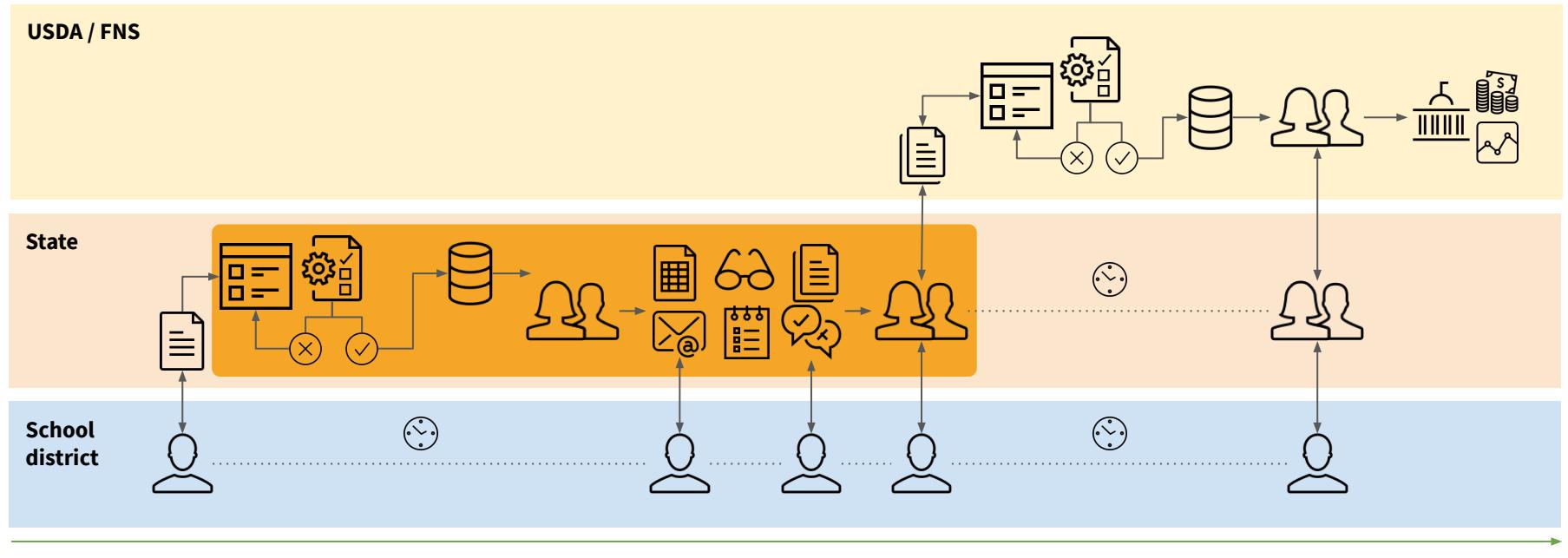
CURRENT STATE

The current process is characterized by a multitude of data validation processes and a high burden on states.



CURRENT STATE

FNS-742 data is validated using a combination of manual and automated edit checks.



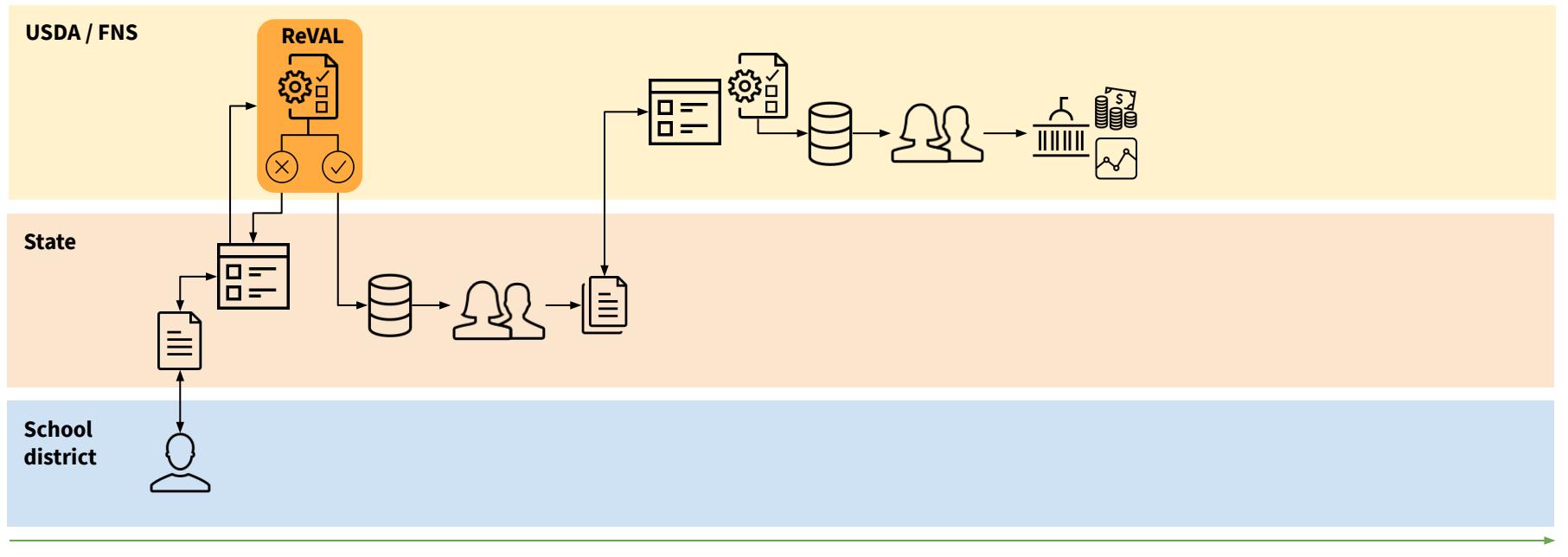
CURRENT STATE

Processes and systems vary from state to state.



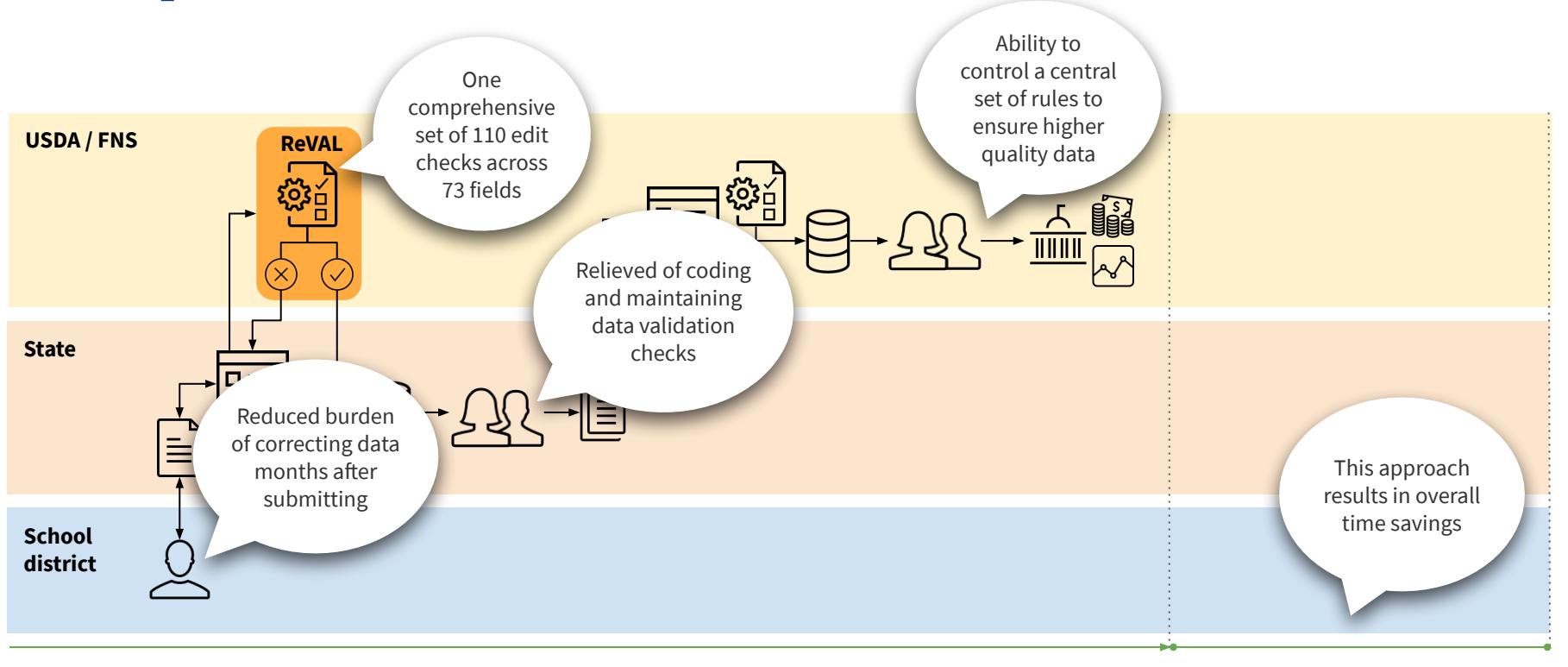
FUTURE STATE

ReVAL allows FNS to maintain validation rules centrally and states to access those rules from their systems via API.



FUTURE STATE

This provides benefits at all levels.





ReVAL is an open source, reusable Django app for data validation and aggregation. Accessible via API or web interface, it is a robust starting point for custom tooling.

Technical spec

- Accessible via GitHub; Django app; written in Python
- Accepts data submitted as CSV or JSON
- Utilizes various validators to perform complex validation (e.g., Goodtables, JSON Schema, SQL)
- Uses a standard JSON response

Implementation

1. Write rules
2. Inject rules into validator
3. Accept data submissions via API or as uploads through web interface
4. Return results via API or UI
5. [Save data in a designated location]

Django REST framework

Upload List / Validate

Validate

Apply all validators in settings to incoming data

:param request: HTTP request
:return: JSON describing validation results

Accepts Content-Types: "text/csv" OR "application/json"

CSV data is handled identically to as-if submitted via file upload.

JSON data must an array of JSON objects whose keys are the expected data columns.

Received JSON objects are converted to tabular format wherein all observed keys are considered headers/columns.

GET /data_ingest/api/validate/

HTTP 405 Method Not Allowed
Allow: POST, OPTIONS
Content-Type: application/json
Vary: Accept

```
{  
    "detail": "Method \"GET\" not allowed."  
}
```

Media type: application/json text/csv

Content:

POST

OPTIONS

Sample JSON of one school's data

```
[
  {
    "state_id": "VA",
    "sfa_id": "123",
    "type_sfa_public": "1",
    "type_sfa_private": 0,
    "school_year_from": 2017,
    "school_year_to": 2018,
    "sfa_name": "test 1",
    "sfa_city": "Alex",
    "sfa_zip_code": 22302,
    "S11A_non_rcci_schl": 30,
    "S11B_non_rcci_stud": 5,
    "S12A_rcci_schl": 10,
    "S12B_rcci_stud": 3,
    "S12aA_rcci_day_schl": 12,
    "S12aB_rcci_day_stud": 4,
    "S12bA_rcci_non_day_schl": 12,
    "S12bB_rcci_non_day_stud": 5,
    "S21A_base_23_schl": 8,
    "S21B_base_23_stud": 1,
    "S22A_non_base_23_schl": 14,
    "S22B_non_base_23_stud": 1,
    "S22aB_non_base_23_free_stud": 1,
    "S22bB_non_base_23_rp_stud": 1,
    "S23A_cep_schl": 17,
    "S23B_cep_stud": 1,
    "S24A_alt_both_schl": 16,
    "S24B_alt_both_stud": 1,
    "S25A_alt_one_only_schl": 57,
    "S25B_alt_one_only_stud": 1,
    "S31_dc_not_required": 1,
    "S32B_dc_snap_stud": 26,
    "S33B_dc_other_stud": 73,
    "S34B_free_snap_letter_stud": 45,
    "S41A_free_doc_apps": 1,
    "S41B_free_doc_stud": 1,
    "S42A_free_inc_apps": 1,
    "S42B_free_inc_stud": 1,
    "S43A_rp_inc_apps": 7,
    "S43B_rp_inc_stud": 100,
    "T1_tot_free": 20,
    "T2_tot_rp": 1,
    "S51_exempt": 0,
    "S52_yes_by_verif": 1,
    "S52_yes_after_verif": 1,
    "S52_no_verif": 1,
    "S531_verif_type_standard": 1,
    "S532_verif_type_alt_1": 1,
    "S533_verif_type_alt_2": 1,
    "S54_error": 10,
    "S55_selected_apps": 10,
    "S56_no_verif_sfa": 10,
    "S57A_confirm_verf_apps": 144,
    "S57B_confirm_verf_stud": 144,
    "S58A1a_free_cat_r_nc_apps": 10,
    "S58A1b_free_cat_r_nc_stud": 10,
    "S58A2a_free_cat_r_rp_apps": 10,
    "S58A2b_free_cat_r_rp_stud": 10,
    "S58A3a_free_cat_r_paid_apps": 10,
    "S58A3b_free_cat_r_paid_stud": 10,
    "S58A4a_free_cat_n_paid_apps": 10,
    "S58A4b_free_cat_n_paid_stud": 10,
    "S58B1a_free_inc_r_nc_apps": 10,
    "S58B1b_free_inc_r_nc_stud": 10,
    "S58B2a_free_inc_r_rp_apps": 10,
    "S58B2b_free_inc_r_rp_stud": 10,
    "S58B3a_free_inc_r_paid_apps": 10,
    "S58B3b_free_inc_r_paid_stud": 10,
    "S58B4a_free_inc_n_paid_apps": 10,
    "S58B4b_free_inc_n_paid_stud": 10,
    "S58C1a_rp_r_nc_apps": 10,
    "S58C1b_rp_r_nc_stud": 10,
    "S58C2a_rp_r_free_apps": 10,
    "S58C2b_rp_r_free_stud": 10,
    "S58C3a_rp_r_paid_apps": 10,
    "S58C3b_rp_r_paid_stud": 10,
    "S58C4a_rp_n_paid_apps": 10,
    "S58C4b_rp_n_paid_stud": 10,
    "VC1_for_cause": 10
  }
]
```

Django REST framework

Upload List / Validate

Validate

Apply all validators in settings to incoming data

:param request: HTTP request
:return: JSON describing validation results

Accepts Content-Types: "text/csv" OR "application/json"

CSV data is handled identically to as-if submitted via file upload.

JSON data must an array of JSON objects whose keys are the expected data columns.

Received JSON objects are converted to tabular format wherein all observed keys are considered headers/columns.

GET /data_ingest/api/validate/

```
HTTP 405 Method Not Allowed
Allow: POST, OPTIONS
Content-Type: application/json
Vary: Accept

{
  "detail": "Method \"GET\" not allowed."
}
```

Media type: application/json

Content:

```
[
  {
    "state_id": "VA",
    "sfa_id": "123",
    "type_sfa_public": "1",
    "type_sfa_private": 0,
    "school_year_from": 2017,
    "school_year_to": 2018,
    "sfa_name": "test 1",
    "sfa_city": "Alex",
    "sfa_zip_code": 22302,
    "S11A_non_rcci_schl": 30,
    "S11B_non_rcci_stud": 5,
    "S12A_rcci_schl": 10,
    "S12B_rcci_stud": 3,
    "S12aA_rcci_day_schl": 12,
    "S12aB_rcci_day_stud": 4,
    "S12bA_rcci_non_day_schl": 12,
    "S12bB_rcci_non_day_stud": 5,
    "S21A_base_23_schl": 8,
    "S21B_base_23_stud": 1,
    "S22A_non_base_23_schl": 14,
    "S22B_non_base_23_stud": 1,
    "S22aB_non_base_23_free_stud": 1,
    "S22bB_non_base_23_rp_stud": 1,
    "S23A_cep_schl": 17,
    "S23B_cep_stud": 1,
    "S24A_alt_both_schl": 16,
    "S24B_alt_both_stud": 1,
    "S25A_alt_one_only_schl": 57,
    "S25B_alt_one_only_stud": 1,
    "S31_dc_not_required": 1,
    "S32B_dc_snap_stud": 26,
    "S33B_dc_other_stud": 73,
    "S34B_free_snap_letter_stud": 45,
    "S41A_free_doc_apps": 1,
    "S41B_free_doc_stud": 1,
    "S42A_free_inc_apps": 1,
    "S42B_free_inc_stud": 1,
    "S43A_rp_inc_apps": 7,
    "S43B_rp_inc_stud": 100,
    "T1_tot_free": 20,
    "T2_tot_rp": 1,
    "S51_exempt": 0,
    "S52_yes_by_verif": 1,
    "S52_yes_after_verif": 1,
    "S52_no_verif": 1,
    "S531_verif_type_standard": 1,
    "S532_verif_type_alt_1": 1,
    "S533_verif_type_alt_2": 1,
    "S54_error": 10,
    "S55_selected_apps": 10,
    "S56_no_verif_sfa": 10,
    "S57A_confirm_verf_apps": 144,
    "S57B_confirm_verf_stud": 144,
    "S58A1a_free_cat_r_nc_apps": 10,
    "S58A1b_free_cat_r_nc_stud": 10,
    "S58A2a_free_cat_r_rp_apps": 10,
    "S58A2b_free_cat_r_rp_stud": 10,
    "S58A3a_free_cat_r_paid_apps": 10,
    "S58A3b_free_cat_r_paid_stud": 10,
    "S58A4a_free_cat_n_paid_apps": 10,
    "S58A4b_free_cat_n_paid_stud": 10,
    "S58B1a_free_inc_r_nc_apps": 10,
    "S58B1b_free_inc_r_nc_stud": 10,
    "S58B2a_free_inc_r_rp_apps": 10,
    "S58B2b_free_inc_r_rp_stud": 10,
    "S58B3a_free_inc_r_paid_apps": 10,
    "S58B3b_free_inc_r_paid_stud": 10,
    "S58B4a_free_inc_n_paid_apps": 10,
    "S58B4b_free_inc_n_paid_stud": 10,
    "S58C1a_rp_r_nc_apps": 10,
    "S58C1b_rp_r_nc_stud": 10,
    "S58C2a_rp_r_free_apps": 10,
    "S58C2b_rp_r_free_stud": 10,
    "S58C3a_rp_r_paid_apps": 10,
    "S58C3b_rp_r_paid_stud": 10,
    "S58C4a_rp_n_paid_apps": 10,
    "S58C4b_rp_n_paid_stud": 10,
    "VC1_for_cause": 10
  }
]
```

POST

Django REST framework

Upload List / Validate

Validate

Apply all validators in settings to incoming data

```
:param request: HTTP request
:return: JSON describing validation results

Accepts Content-Types: "text/csv" OR "application/json"

CSV data is handled identically to as-if submitted via file upload.

JSON data must an array of JSON objects whose keys are the expected
data columns.

Received JSON objects are converted to tabular format wherein all
observed keys are considered headers/columns.
```

POST /data_ingest/api/validate/

```
HTTP 200 OK
Allow: POST, OPTIONS
Content-Type: application/json
Vary: Accept

{
    "tables": [
        {
            "headers": [
                "state_id",
                "type_sfa_private",
                "school_year_from",
                "school_year_to",
                "S11A_non_rcci_schl",
                "S11B_non_rcci_stud",
                "S12A_rcci_schl",
                "S12B_rcci_stud",
                "S12A_rcci_day_schl",
                "S12B_rcci_day_stud",
                "S12B_rcci_non_day_schl",
                "S12B_rcci_non_day_stud",
                "S21A_base_23_schl",
                "S21B_base_23_stud",
                "S22A_base_23_schl",
                "S22B_base_23_stud",
                "S22B_non_base_23_free_stud",
                "S22B_non_base_23_rp_stud",
                "S22A_cpt_schl",
                "S23B_cpt_stud",
                "S24A_alt_both_schl",
                "S24B_alt_both_stud",
                "S25A_alt_one_only_schl",
                "S25B_alt_one_only_stud",
                "S31_dc_not_required",
                "S32_dc_not_required"
            ],
            "whole_table_errors": [
                {
                    "severity": "Error",
                    "code": "extra-header",
                    "message": "There is an extra header in column 74 (sfa_id)",
                    "error_columns": [
                        "sta_id"
                    ]
                }
            ],
            "rows": [
                {
                    "row_number": 2,
                    "errors": [
                        {
                            "severity": "Error",
                            "code": "1",
                            "message": "Total schools (not including RCCIs) cannot be greater than total students in school.",
                            "error_columns": [
                                "S11A_non_rcci_schl",
                                "S11B_non_rcci_stud"
                            ]
                        }
                    ]
                }
            ]
        }
    ]
}
```



Using ReVAL provides benefits in the form of time savings and reduced effort for both developers and end users involved in federated data efforts.

For developers and data practitioners

- Builds on a known processes and existing open source code, making it easy for users to spin up instances
- Very customizable and adaptable to unique needs and formats
- Flexible to be implemented as API on the backend or with a user interface

For program staff and end users

- Reduced manual data validation
- Minimize duplication of effort across federated data effort
- Centralized data validation checks, facilitating updates and version control
- Human-readable error messages make corrections easier





The Census Commodity Flow Survey (CFS) is a survey of approximately 100,000 establishments that ship commodities.

Applications of ReVAL

- Use the API-based version to streamline process of data validation from the CFS and create a central source of truth for edit checks within the Census Bureau
- Use the validation and aggregation features in the interface-based version to provide a better user experience for businesses reporting the CFS

Proof points

- Able to engage with ReVAL directly through via GitHub documentation
- Accelerated timeline of building a prototype by several weeks
- Has identified additional use cases
- Sees value in the creation of a reusable front-end interface for ReVAL



TO
EX 35
94
I-10
I-10

Vineland Pl

FEBAR ONLY

DO NOT ENTER



The Work Zone Data Exchange (WZDx) specification enables infrastructure owners and operators (IOOs) to make harmonized work zone data available for third party use.

Application of ReVAL

- ReVAL could provide schema validation for data providers adjusting their data feeds to conform with the new WZDx data specification
- Would benefit from lightweight user interface for data providers

Proof points

- ReVAL able to be upgraded to do more complex validation with nested data
- Engaging with this partner revealed opportunities to make ReVAL compatible with a wider set of data and rules formats

Review of work to date

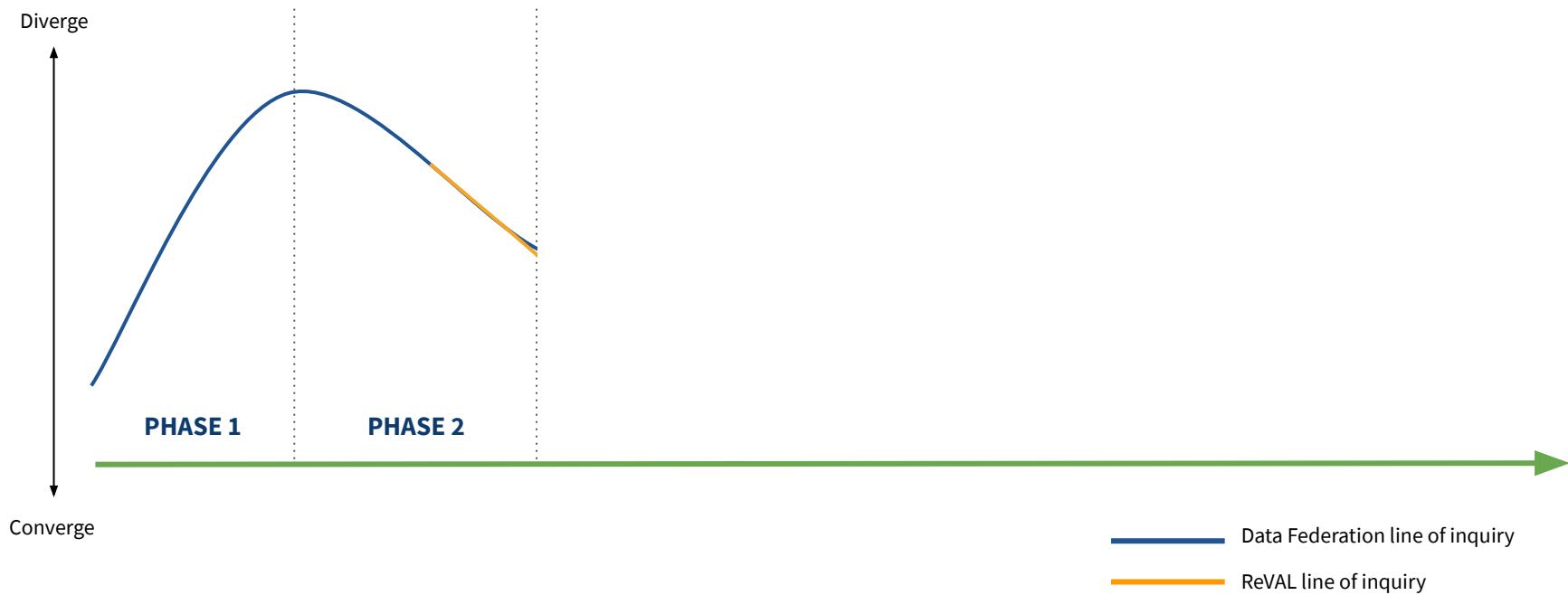
Phase 1 explored the breadth of the problem space.

Our findings suggested that reusable tooling and processes would benefit future federated data efforts.



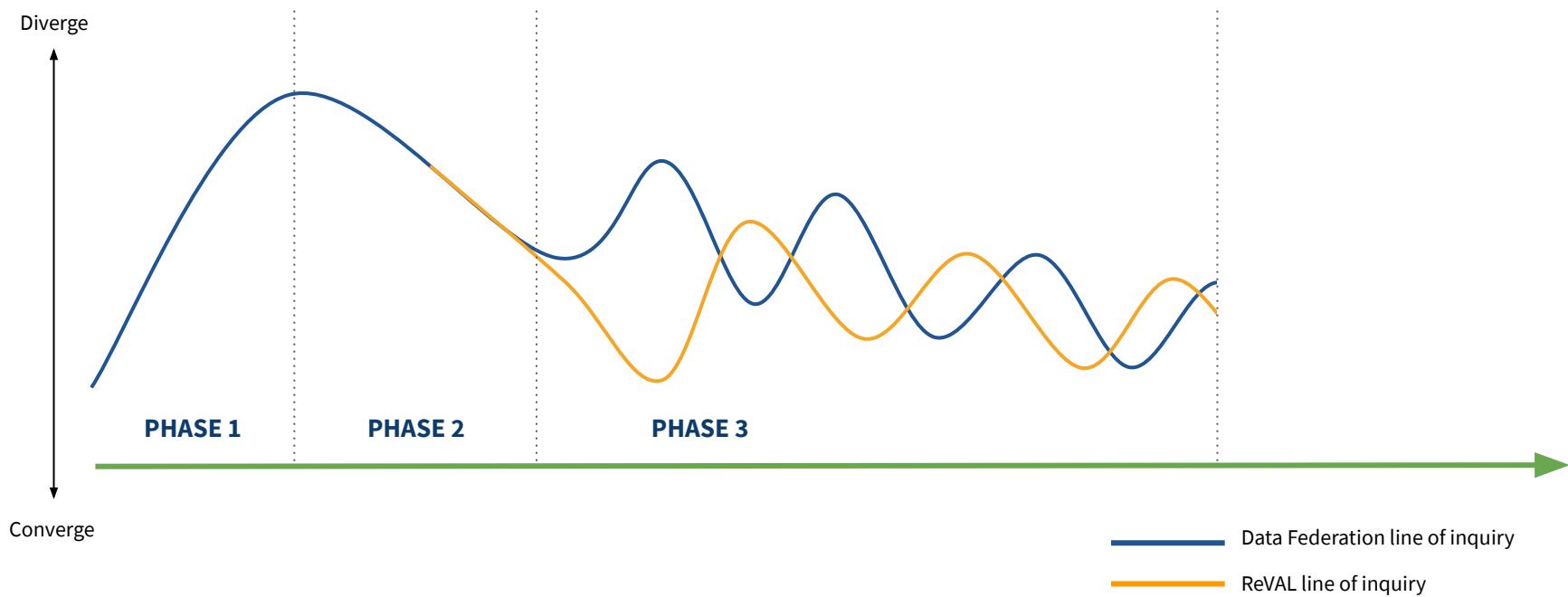
In Phase 2, our scope converged on an initial tool.

We created a prototype of ReVAL and found a partner in the USDA Food & Nutrition Service Office of Program Integrity.



In Phase 3, our lines of inquiry converged and diverged.

We continued research and outreach activities, built out ReVAL for FNS, and began to scale the tool with other partners.



Data Federation line of inquiry + ReVAL-specific line of inquiry

Tool development relied on research and outreach to find potential users and align with other similar efforts.



Federal Data Strategy
Leveraging Data as a Strategic Asset



DIGITALGOV



DATA.GOV



OMB
Office of Management
and Budget



FEMA



CENTERS FOR DISEASE
CONTROL AND PREVENTION



CENTERS FOR MEDICARE & MEDICAID SERVICES

VA



U.S. Department
of Veterans Affairs



Performance.gov



< N I E M >

FGDC.GOV
FEDERAL GEOGRAPHIC DATA COMMITTEE

DATA
COALITION



CIO
.GOV

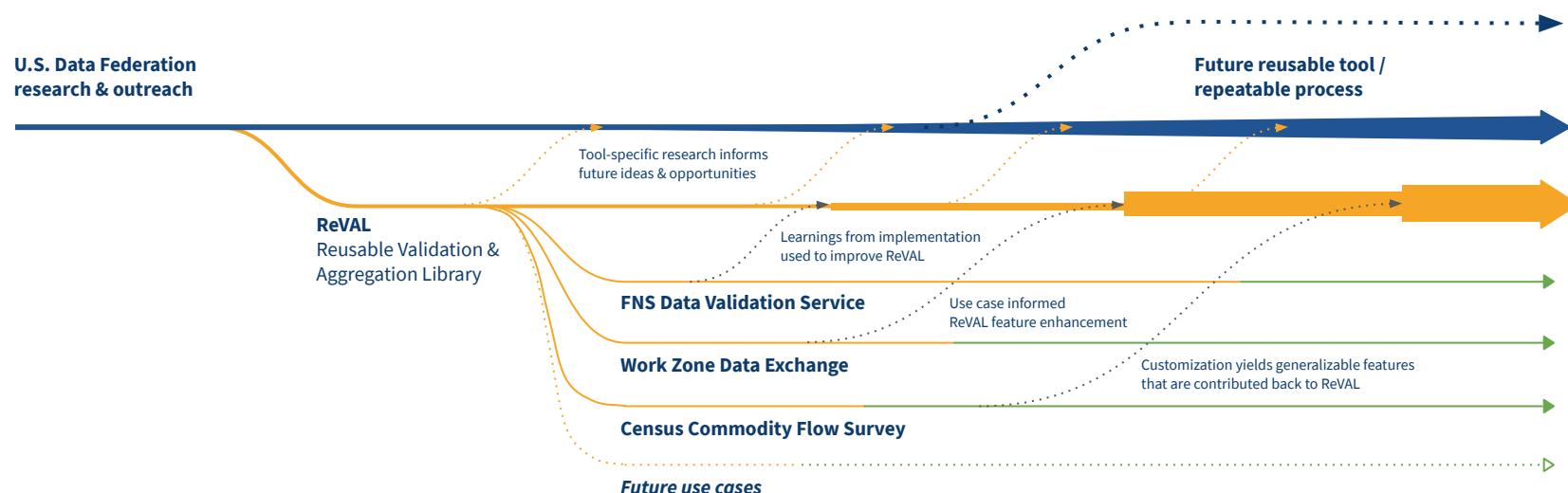
United States
Census
Bureau



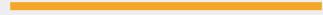
Data Federation line of inquiry + ReVAL-specific line of inquiry

Our diverging and converging lines of inquiry informed and supported each other.

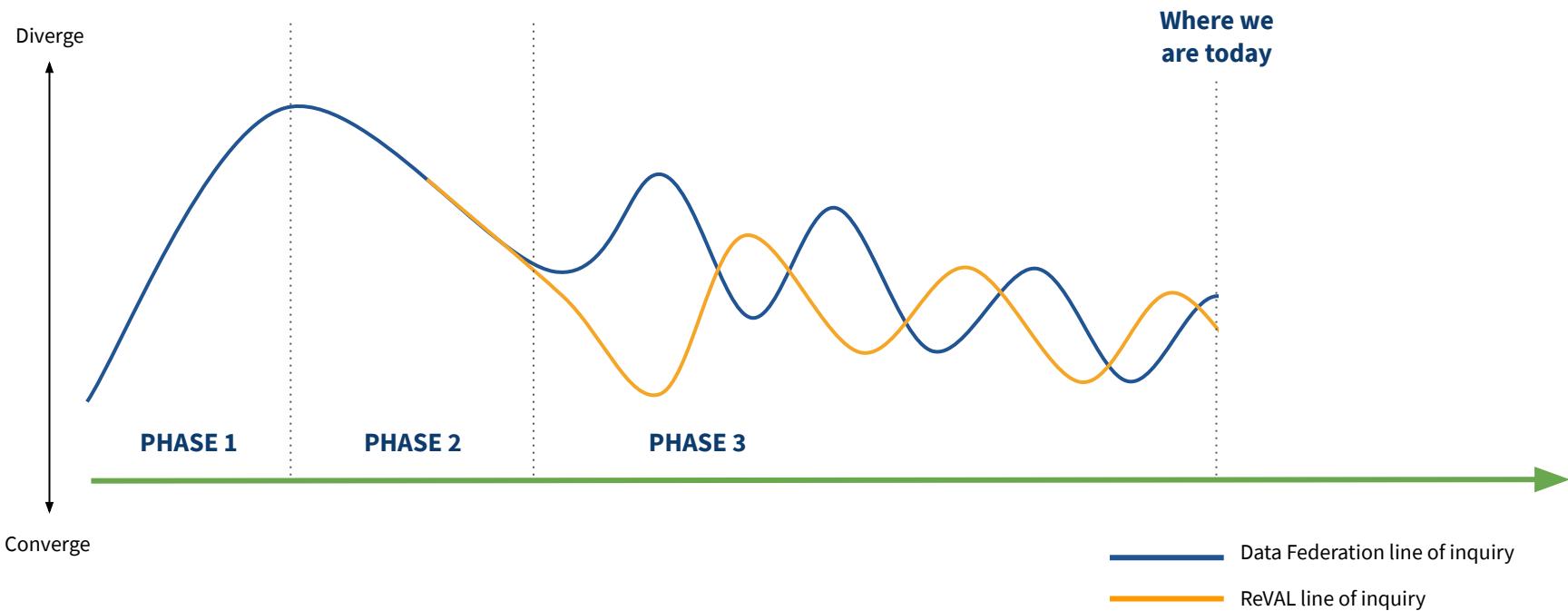
This suggests a model for future resource development.



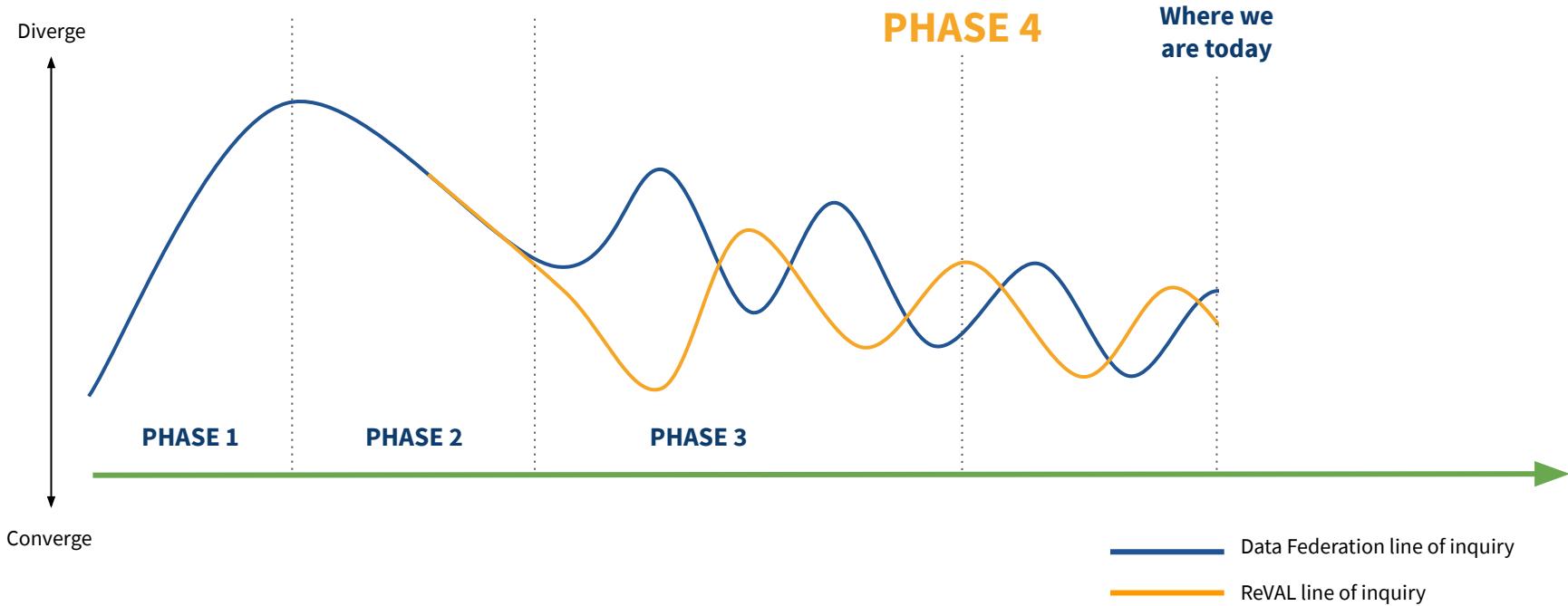
Vision for Phase 4



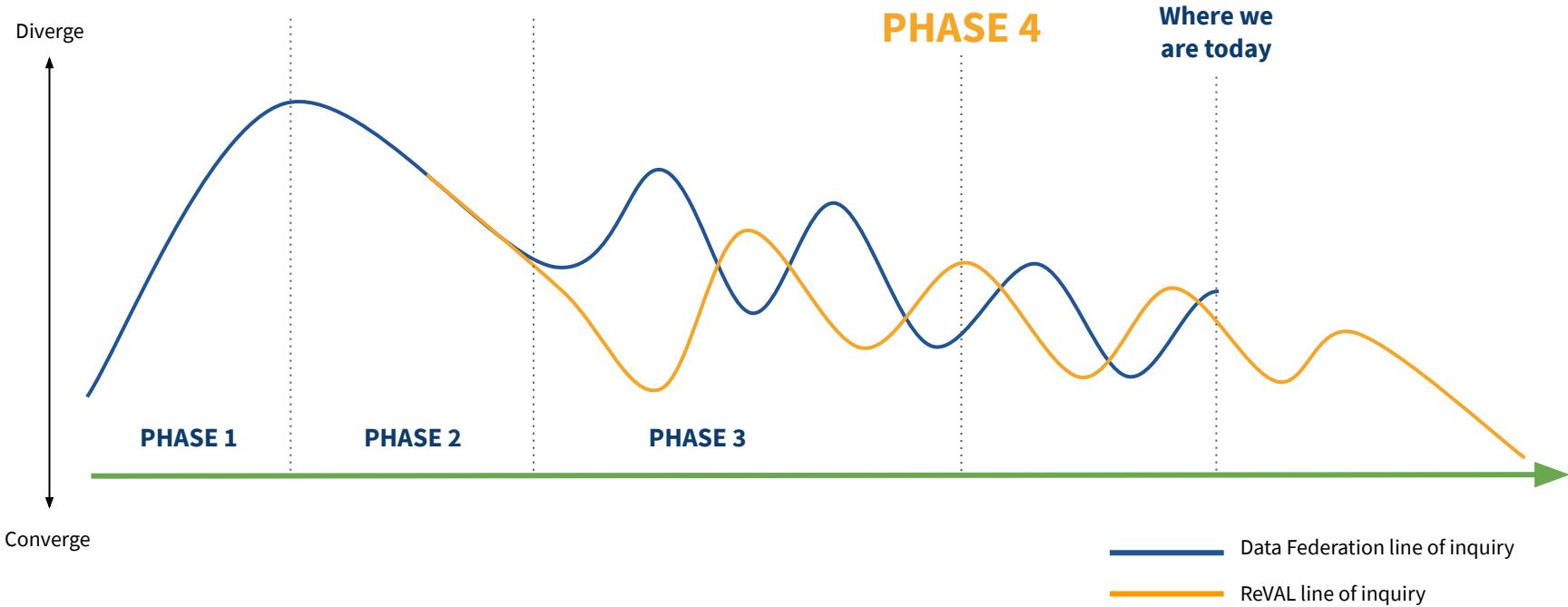
As we reach the end of Phase 3, we realize that we have already started doing the work we need to be doing.



We're already in Phase 4: Our work with partners across government has allowed us to start scaling ReVAL.



While we *could* limit our scope for Phase 4 to bringing the development of ReVAL to its logical conclusion...



Limiting Phase 4 to the completion
of ReVAL would be **short-sighted**
and antithetical to the goals of the
U.S. Data Federation.

MODEL FOR RESOURCE DEVELOPMENT

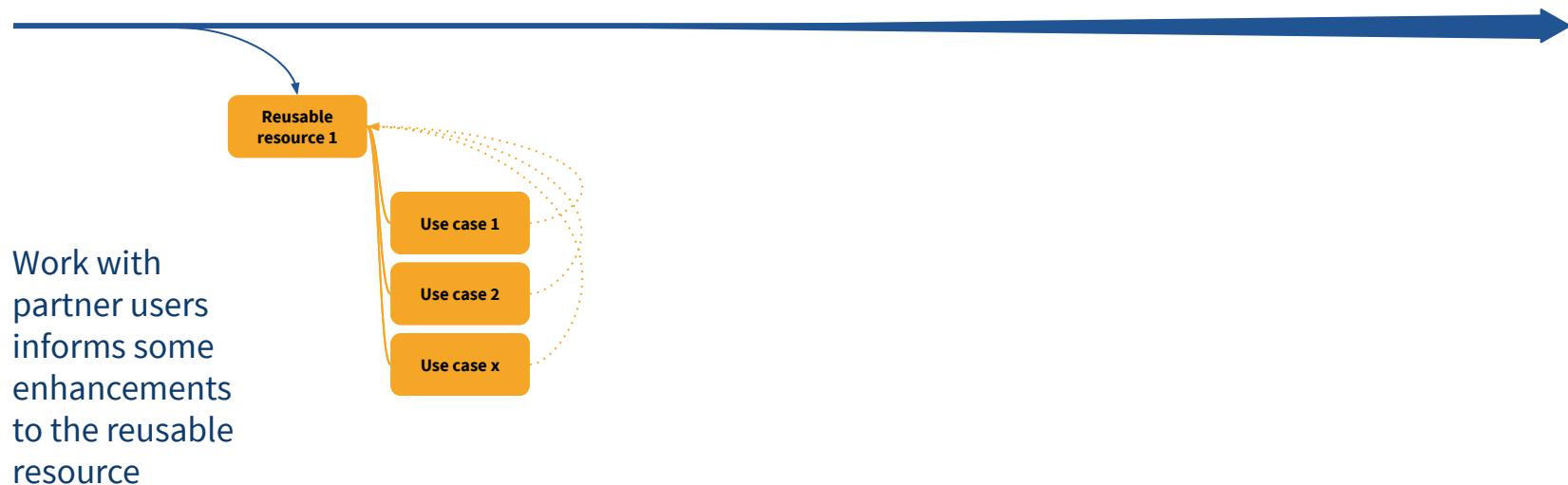
U.S. Data Federation

Reusable
resource 1

Research and outreach allow us to identify common needs, which suggest solutions to pursue

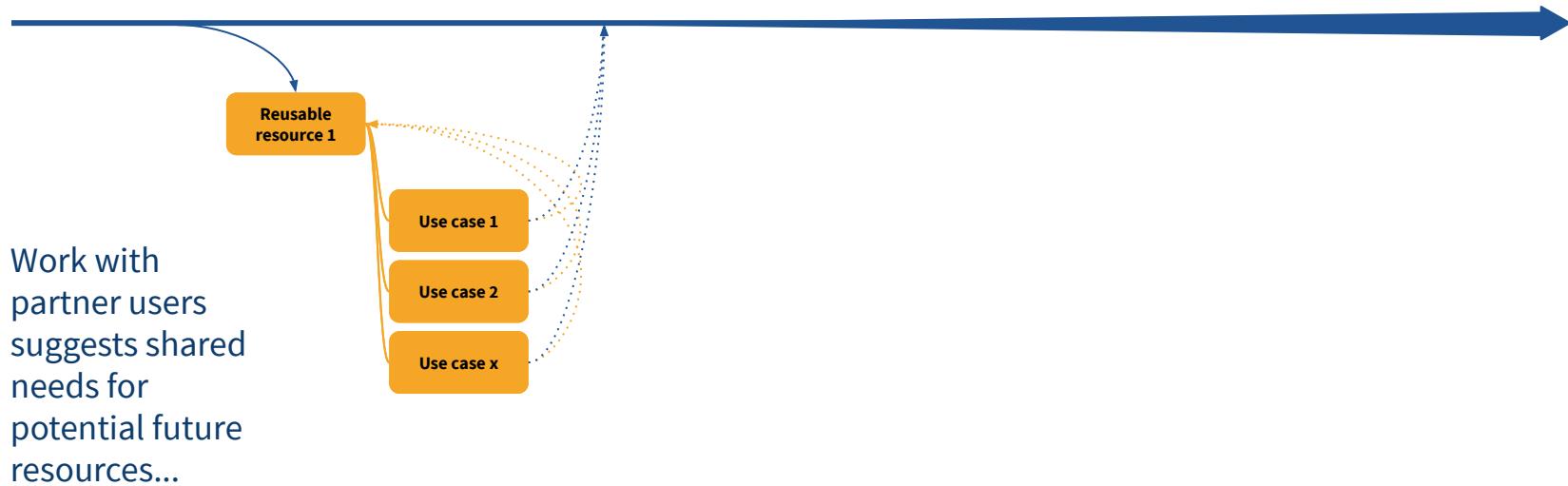
MODEL FOR RESOURCE DEVELOPMENT

U.S. Data Federation



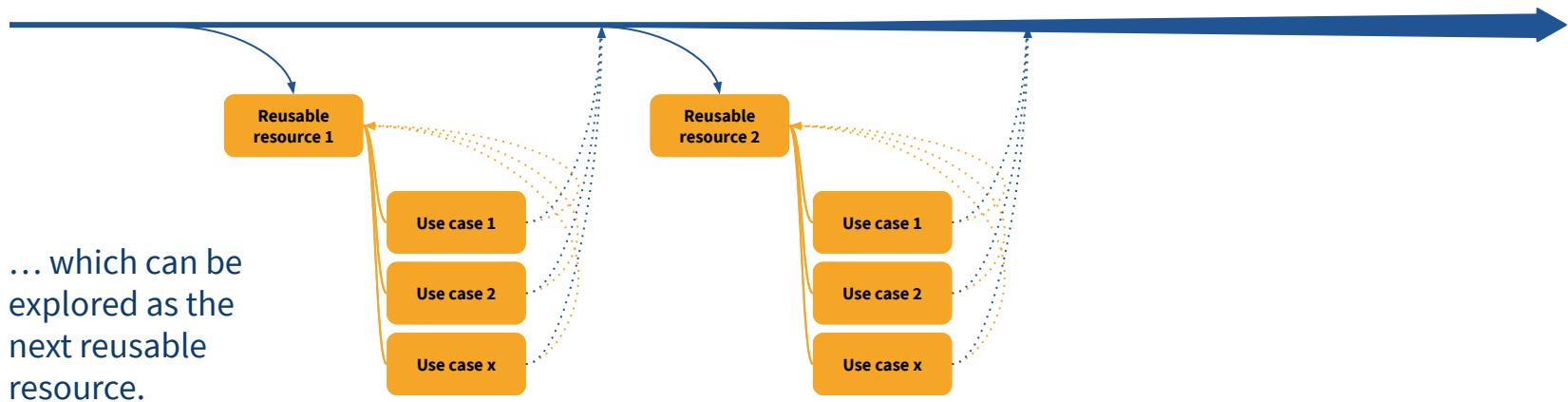
MODEL FOR RESOURCE DEVELOPMENT

U.S. Data Federation



MODEL FOR RESOURCE DEVELOPMENT

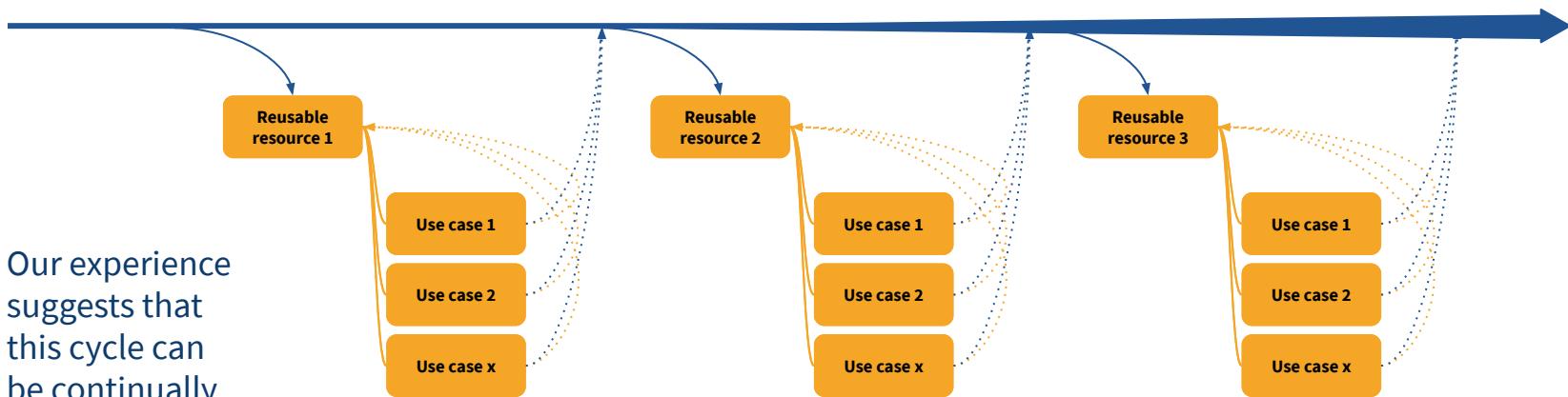
U.S. Data Federation



MODEL FOR RESOURCE DEVELOPMENT

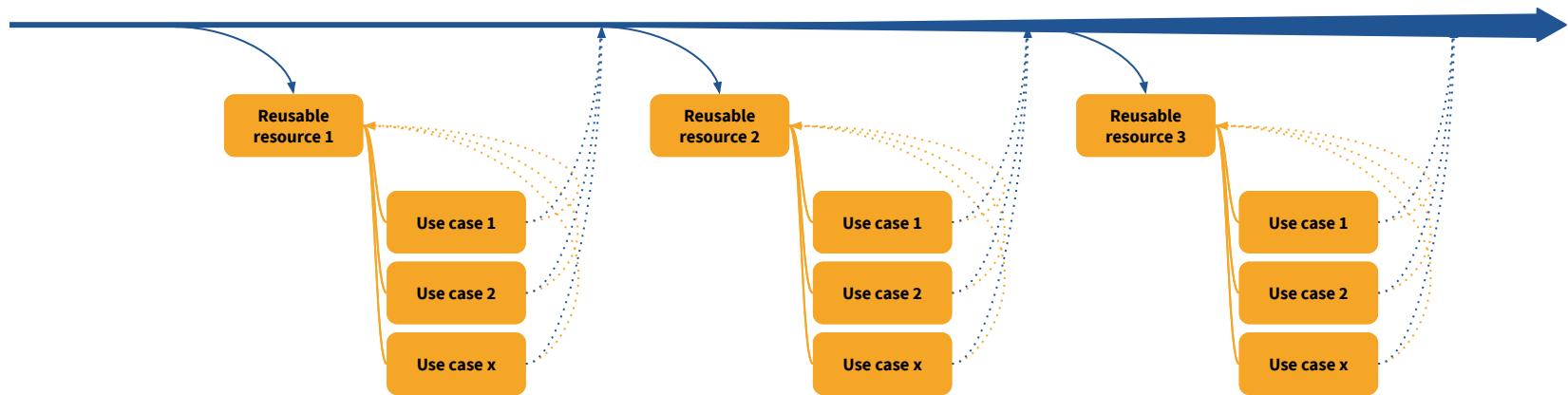
U.S. Data Federation

Our experience suggests that this cycle can be continually repeated



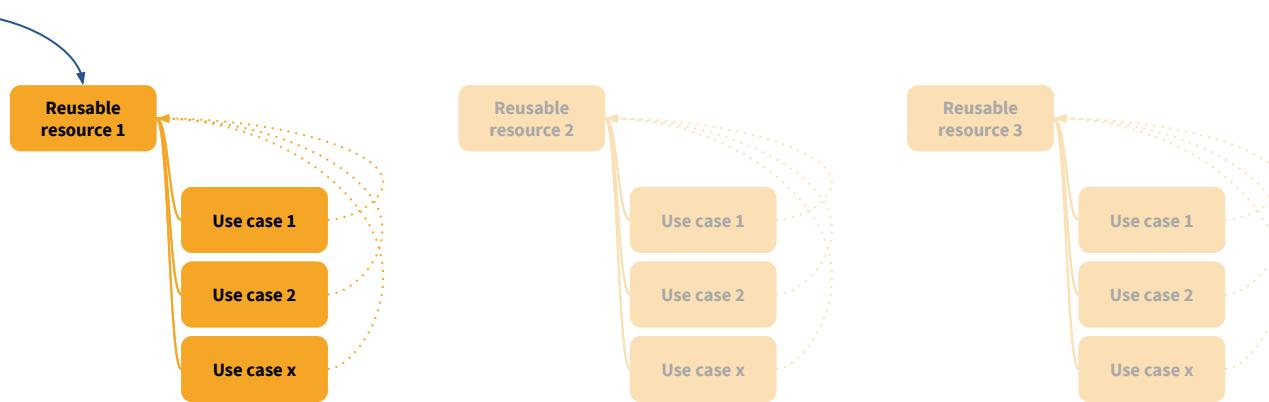
MODEL FOR RESOURCE DEVELOPMENT

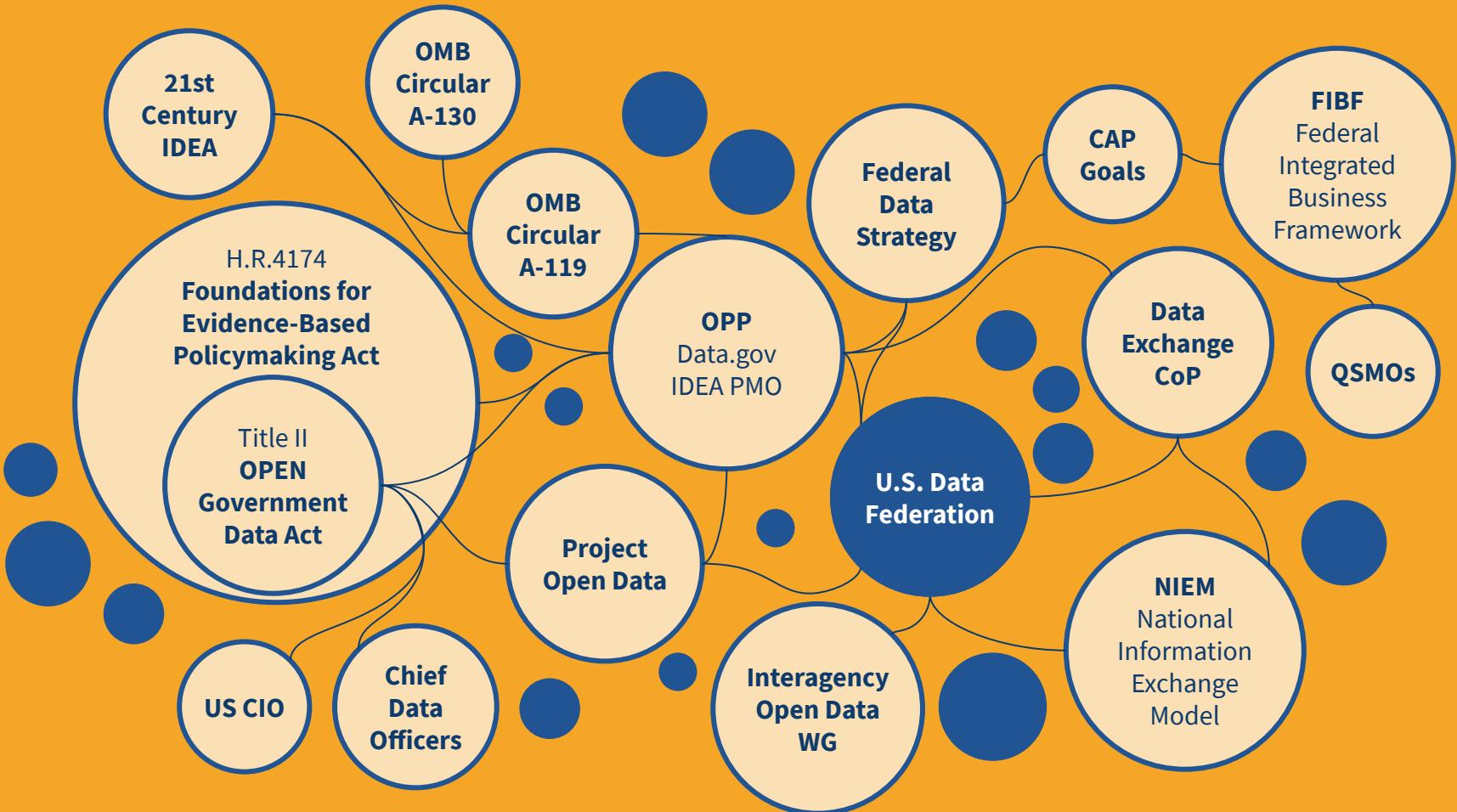
U.S. Data Federation



MODEL FOR RESOURCE DEVELOPMENT

U.S. Data Federation





The work of the U.S. Data Federation dovetails with both GSA and government-wide initiatives and priorities

Legislative mandates	<ul style="list-style-type: none">Foundations for Evidence-Based Policymaking Act (Title II: OPEN Government Data Act)21st Century IDEAOMB Circular A-119OMB Circular A-130	Government-wide initiatives	<ul style="list-style-type: none">Federal Data StrategyCross-Agency Priority Goals
OPP Data & Analytics efforts / Data.gov priorities	<ul style="list-style-type: none">Relaunch Project Open Data by November 2019 to meet mandate to create a shared repositoryCollaborate within OPP to support mandates from 21st Century IDEA	TTS Strategic Plan	<ul style="list-style-type: none">Measurable acceleration of project delivery and valuable resource creationCollaboration between 10x, OPP, and 18FSynergies with data.gov, leveraging of existing resources, and working in the openLeveraging CoPs and focusing on reusable resources
Communities of Practice & councils	<ul style="list-style-type: none">Data Exchange CoPOpen Data Working GroupChief Data Officers CouncilInteragency Council on Standards PolicyAgency-specific CoPs for Open Data	Related 10x projects	<ul style="list-style-type: none">10x 21st Century IDEA Bundle10x Federal Grantee Reporting10x Eligibility Rules Service10x Canonical Federal Identifiers10x PRA project

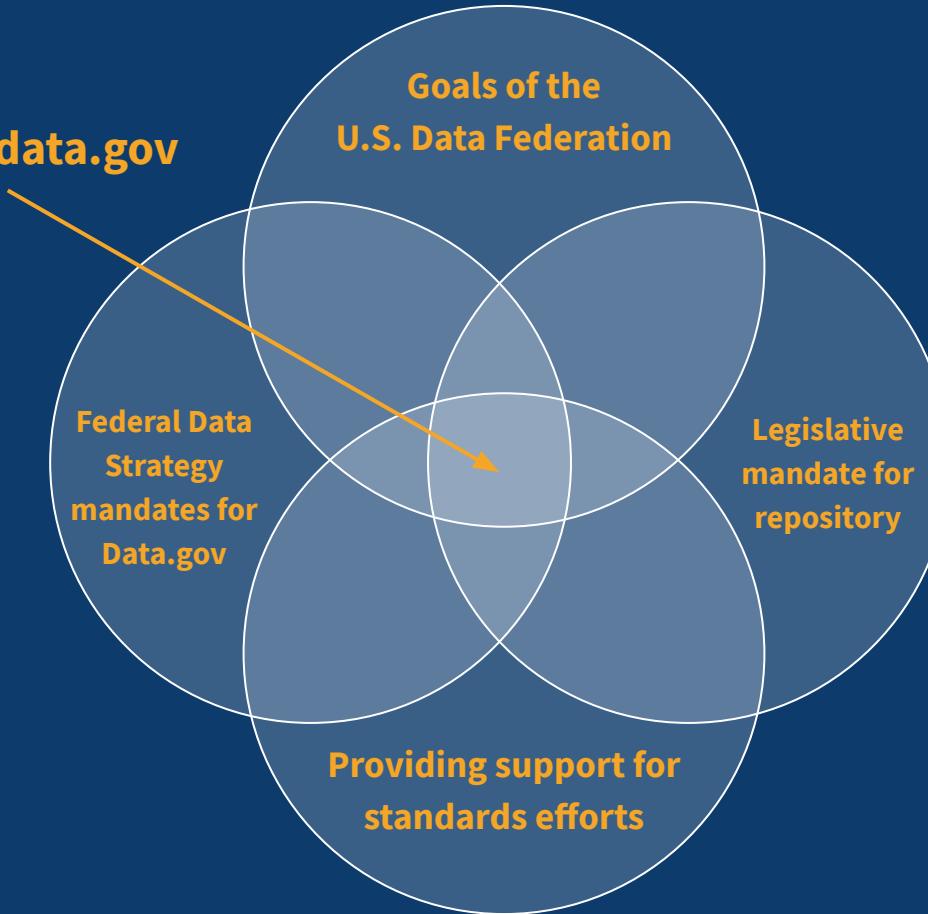
resources.data.gov

**Goals of the
U.S. Data Federation**

**Federal Data
Strategy
mandates for
Data.gov**

**Legislative
mandate for
repository**

**Providing support for
standards efforts**



The U.S. Data Federation and [resources.data.gov](#) have a lot of work and goals in common.

U.S. Data Federation

- Ensure ReVAL is maximally reusable by completing known enhancements
- Ensure ReVAL is documented and presented in a way that serves users
- Ensure that users discover ReVAL and understand how to integrate it into larger data efforts
- Raise awareness of ReVAL

[resources.data.gov](#)

- In 3 months, relaunch Project Open Data as [resources.data.gov](#)
- Ensure resources are documented and presented in a way that serves users
- Create a new information architecture that helps users navigate and discover appropriate resources
- Raise awareness of this new repository

The U.S. Data Federation and resources.data.gov have a lot of work and goals in common.

U.S. Data Federation

- Ensure ReVAL is maximally reusable by completing known enhancements
- Ensure ReVAL is documented and presented in a way that serves users
- Ensure that users discover ReVAL and understand how to integrate it into larger data efforts
- Raise awareness of ReVAL

resources.data.gov

- In 3 months, relaunch Project Open Data as resources.data.gov
- Ensure resources are documented and presented in a way that serves users
- Create a new information architecture that helps users navigate and discover appropriate resources
- Raise awareness of this new repository

resources.data.gov



U.S. Data Federation



Complete development of ReVAL



resources.data.gov

U.S. Data Federation

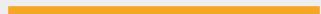
Complete development of ReVAL



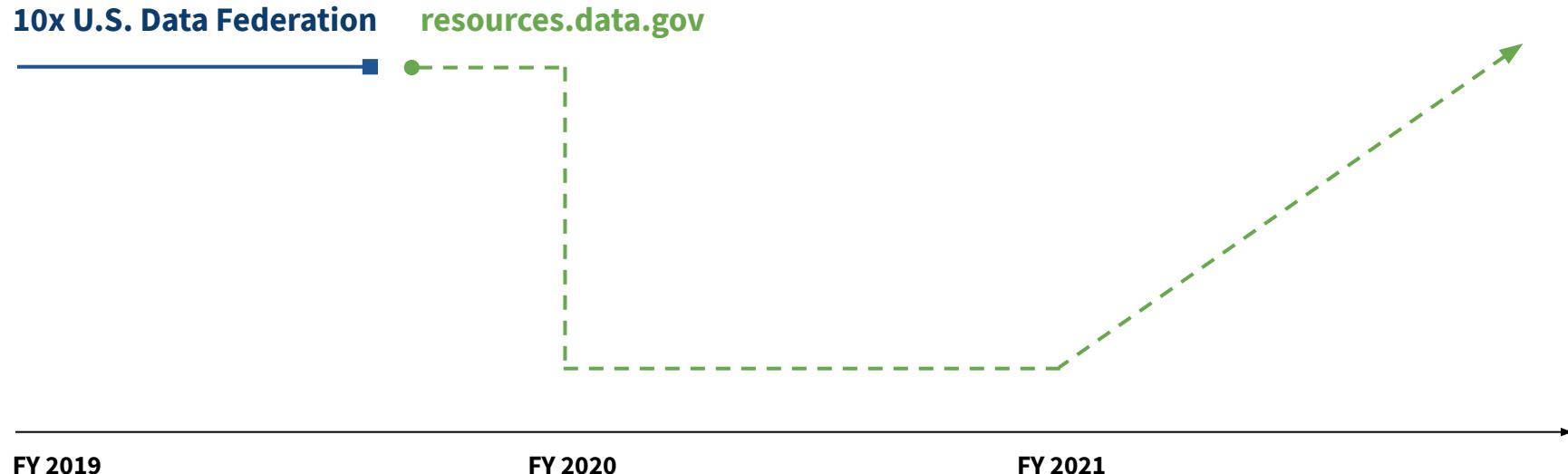
PHASE 4 VISION

Use the learnings, resource development model, momentum, and network of the U.S. Data Federation to build a user-centered resources.data.gov that helps data practitioners navigate the landscape of data standards, tools, and other resources.

Making Phase 4 a reality



Without additional funding, the work of the U.S. Data Federation will be paused. When resources.data.gov is funded in FY 21, work will have to be ramped up again.



We have the opportunity to build on the momentum of the U.S. Data Federation. Phase 4 funding will allow us to shape and accelerate the creation of a useful resources.data.gov.

10x U.S. Data Federation

Use the remaining Phase 3 budget (approx. 300 hours) to fund 15 weeks of limited ongoing support to partners and scope definition with resources.data.gov team

Phase 4 funding

Re-staff some of the Phase 3 team becoming available again at the end of FY 19 to carry the work forward into Phase 4

resources.data.gov

Transition to resources.data.gov funding and support

FY 2019

FY 2020

FY 2021

Funding need for Phase 4 as currently envisioned: \$875,000

STAFFING REQUEST

First 3 months:

- 1 engineer at 100%
- 1 product manager at 100%
- 1 strategist at 100%
- 1 UX researcher/designer at 100%

Following 6 months:

- 1 engineer at 25%
 - 1 product manager 25%
 - 1 strategist at 100%
 - 1 UX researcher/designer at 50%
-

PROPOSED ACTIVITIES

Complete development of ReVAL

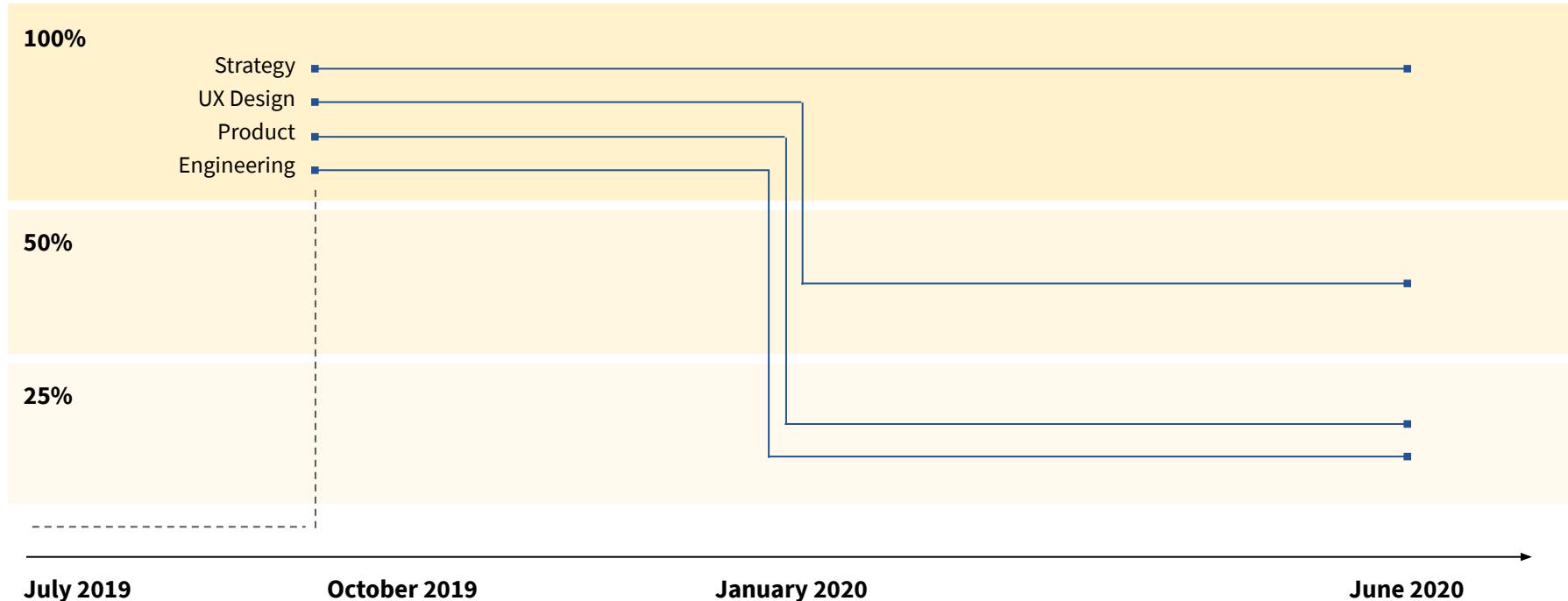
- Build out the front-end aggregation/validation interface
- Integrate API and interface sides of the tool
- Clean up code, finish documentation, and set up testing
- Enhance the JSON complexity ReVAL accepts
- Enhance the SQL validator and add other validators
- Finish documentation for reusable core ReVAL
- Have regular check-ins with partners

Collaborate with the team working on resources.data.gov

- Collaborate on a strategic roadmap for the Resource Center
- Conduct an inventory of resources on Project Open Data
- Conduct user research to understand the needs of data practitioners accessing the Resource Center
- Design the information architecture for the Resource Center
- Use ReVAL as a pilot and content model for future resources
- Leverage the U.S. Data Federation contacts and network to raise awareness of the effort and gather feedback

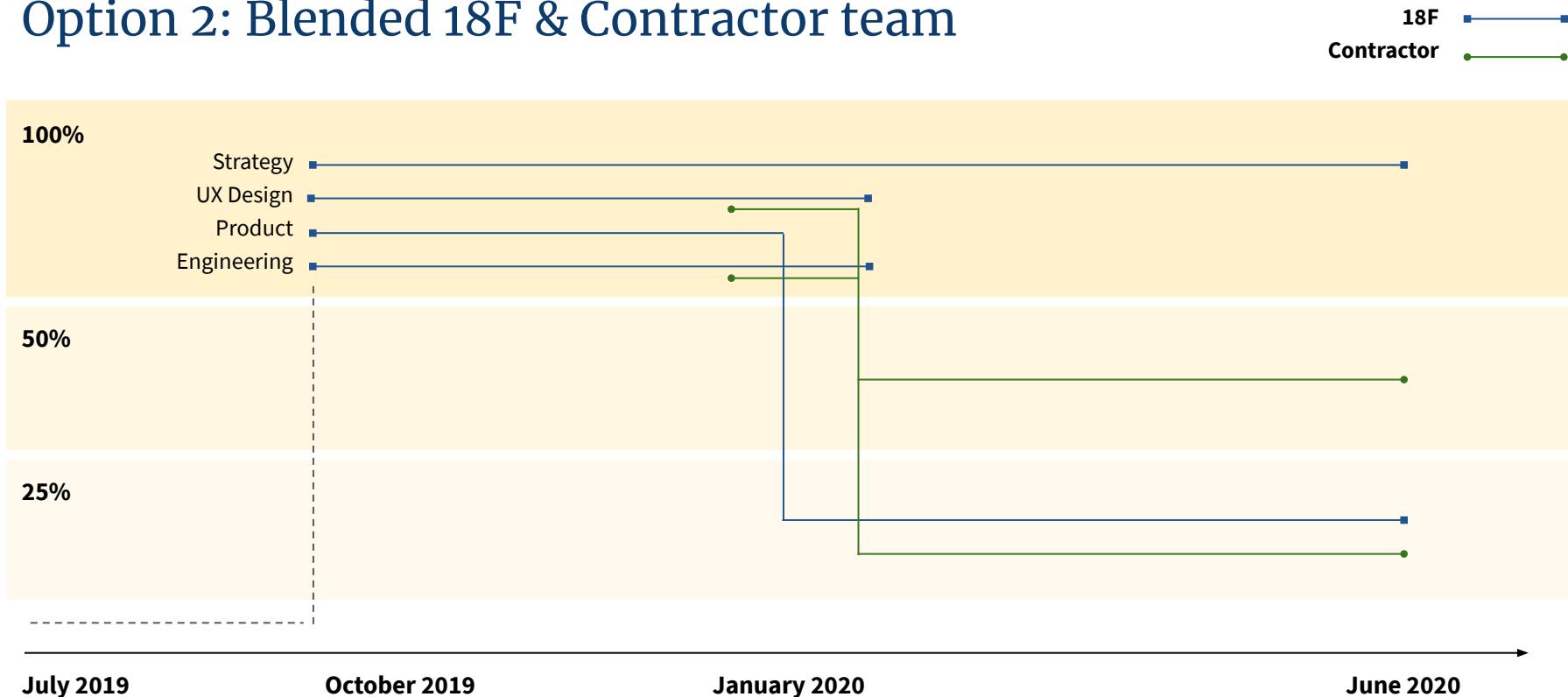
Timeline of staffing needs

Option 1: 18F team



Timeline of staffing needs

Option 2: Blended 18F & Contractor team



WHAT SUCCESS LOOKS LIKE

In 6 months

- ReVAL is in production with one or more of our pilot partners
- The API and UI versions of ReVAL are fully integrated
- ReVAL's documentation is templatized as a content model for resources in the new repository
- A model for continuous enhancement to ReVAL with low O&M is in place

In 12 months

- ReVAL has been adopted by 5 or more agency partners
- resources.data.gov provides a way to guide data efforts through different stages of maturity, integrating “people-focused” resources with “technical” resources
- The resource development model of the U.S. Data Federation guides the development of new resources for resources.data.gov

HOW TO MEASURE SUCCESS

Number of applications

- Number of instances of ReVAL and other resources being forked and applied to different projects

Reach of the resource

- Number of agencies and programs accessing ReVAL and other resources

Burden reduction

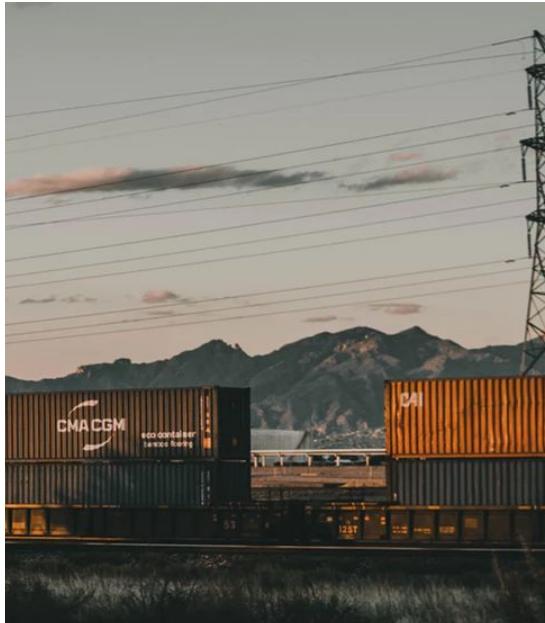
- Reduced PRA burden hours (incidentally a great source of potential use cases)

Time savings & cost avoidance

- Implementing ReVAL and other tools accelerates progress towards use-case specific development

The potential impact of this work is great.

Data problems are not abstract, they have real, significant impact on the country's people and economy.



Efforts towards data standards
and interoperability will build the
infrastructure of the 21st century.
That's worth investing in.

Discussion





Thank you!

Prepared by

Philip Ashlock
Mike Gintz
Mark Headd
Ethan Heppner
Julia Lindpaintner
Amy Mok

Appendix



Alternate funding request: Limit scope to completing development of ReVAL

Funding request: \$272,640

STAFFING REQUEST

- 1 engineer at 100% for 3 months
- 1 product manager at 100% for 3 months
- 1 UX researcher/designer at 50% for 3 months

PROPOSED ACTIVITIES

Finish known production-readiness enhancements to ReVAL

- Build out the front-end aggregation/validation interface
- Integrate API and interface sides of the tool
- Clean up code, finish documentation, and set up testing
- Enhance the JSON complexity ReVAL accepts; add ability to accept tabular data packages; consider other formats
- Enhance the SQL validator and add other validators
- Finish documentation for reusable core ReVAL

Continue to collaborate with pilot partners

- Have regular check-ins with partners as they move towards production, supporting them when needed
- Integrate learnings from their process into the documentation

Promote discoverability and adoption of ReVAL

- Continue marketing via presentations & mailing list
- Lobby for inclusion on the resources.data.gov