Demand-Driven Open Data

An introduction for data owners

HHS IDEA LAB

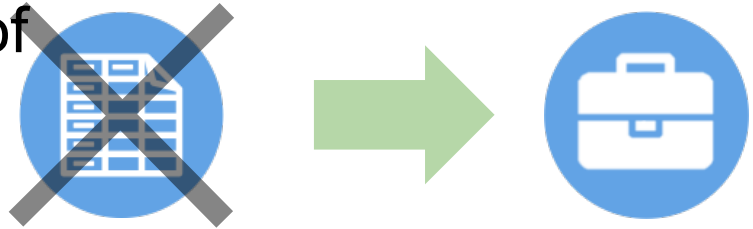
More info: http://demand-driven-open-data.github.io
Contact: David.Portnoy@HHS.gov

Aka, “Innovative Linkages” Project

The Opportunity

HHS can create additional economic, health and social value by changing the way it measures progress on Open Data efforts...

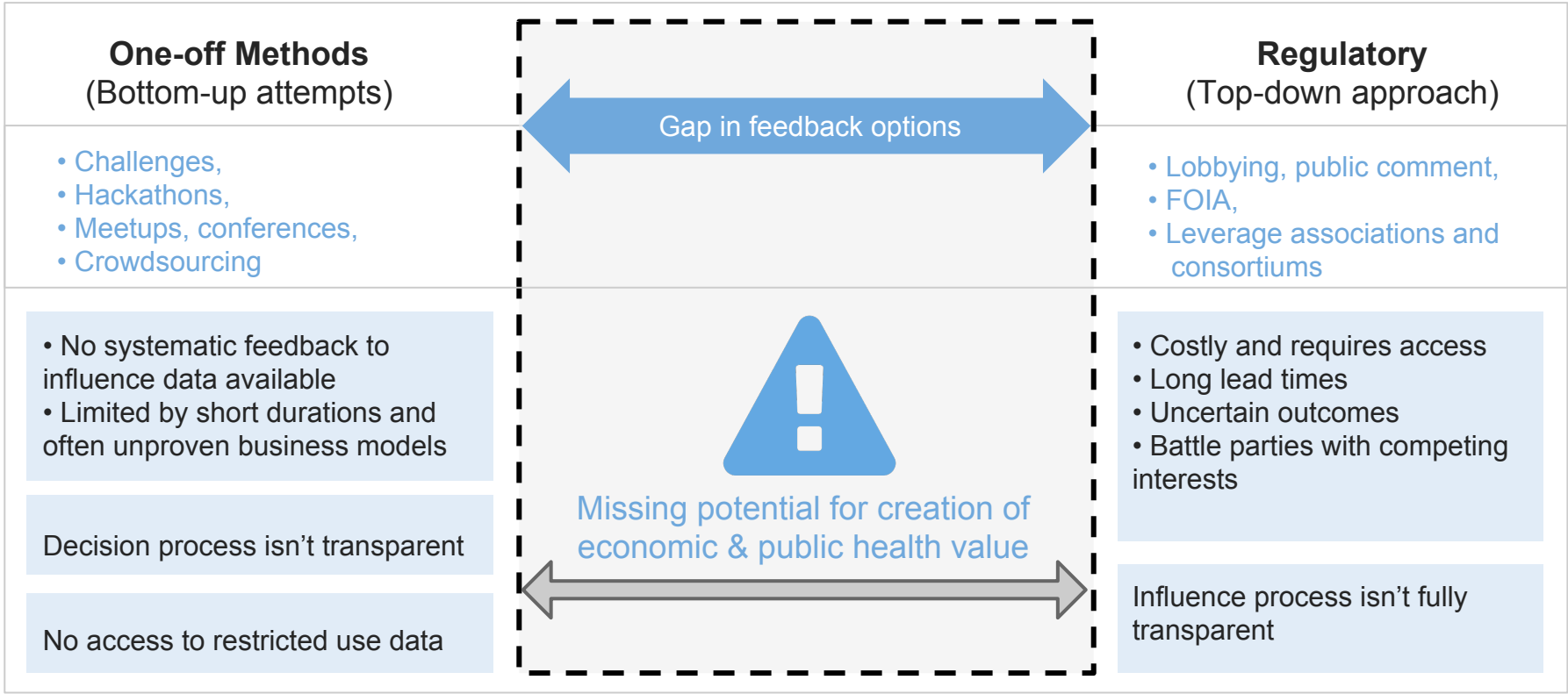
from number of datasets released



to value in terms of use cases enabled

Prior to DDOD, if you wanted to influence the data HHS provides there were primarily two extremes: participate in one-off events or attempt the regulatory path

...But each had significant limitations

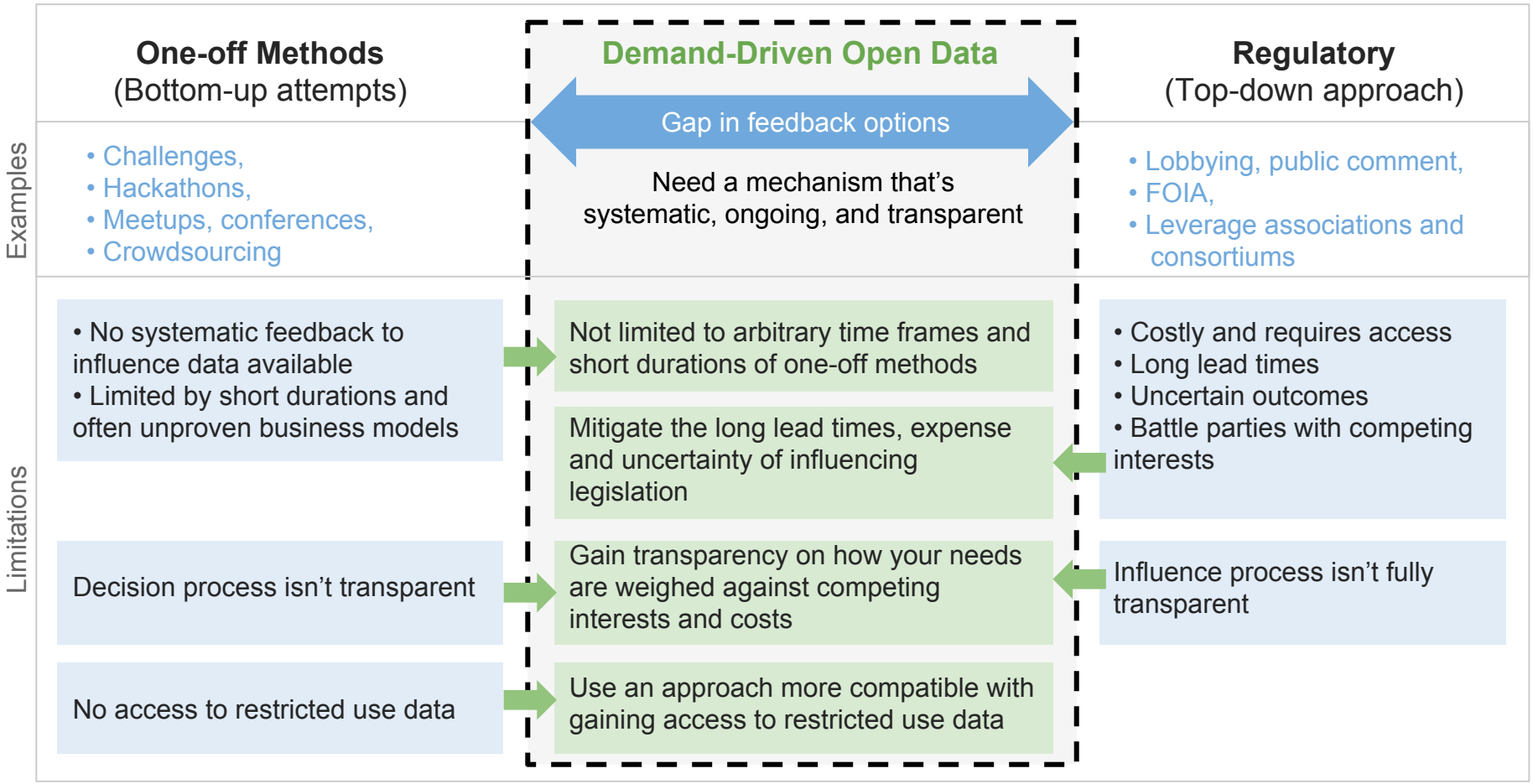


So we need a mechanism that…

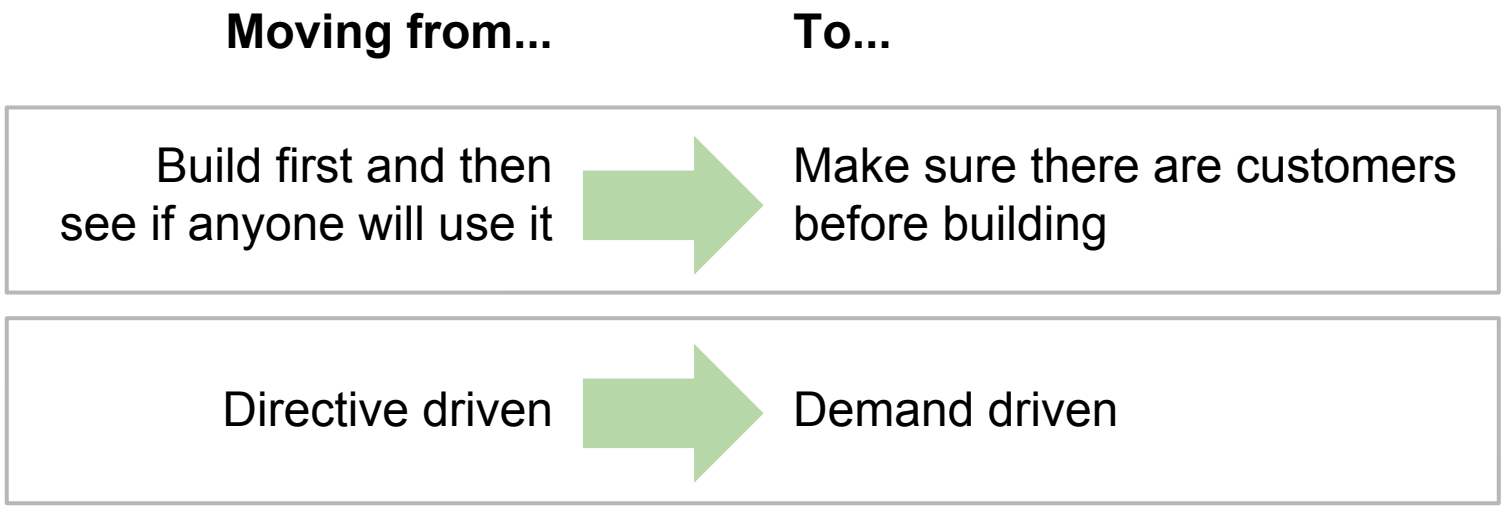
Enables systematic, ongoing and transparent signaling of relative value of data in a way that's inclusive of all types of participants

That's "Demand Driven Open Data" (DDOD)

DDOD fills the gap and addresses many of the limitations



By understanding the “market” for its data, HHS can better allocate resources by migrating to a “Lean Startup” methodology



DDOD can be broken into functional areas to execute in parallel

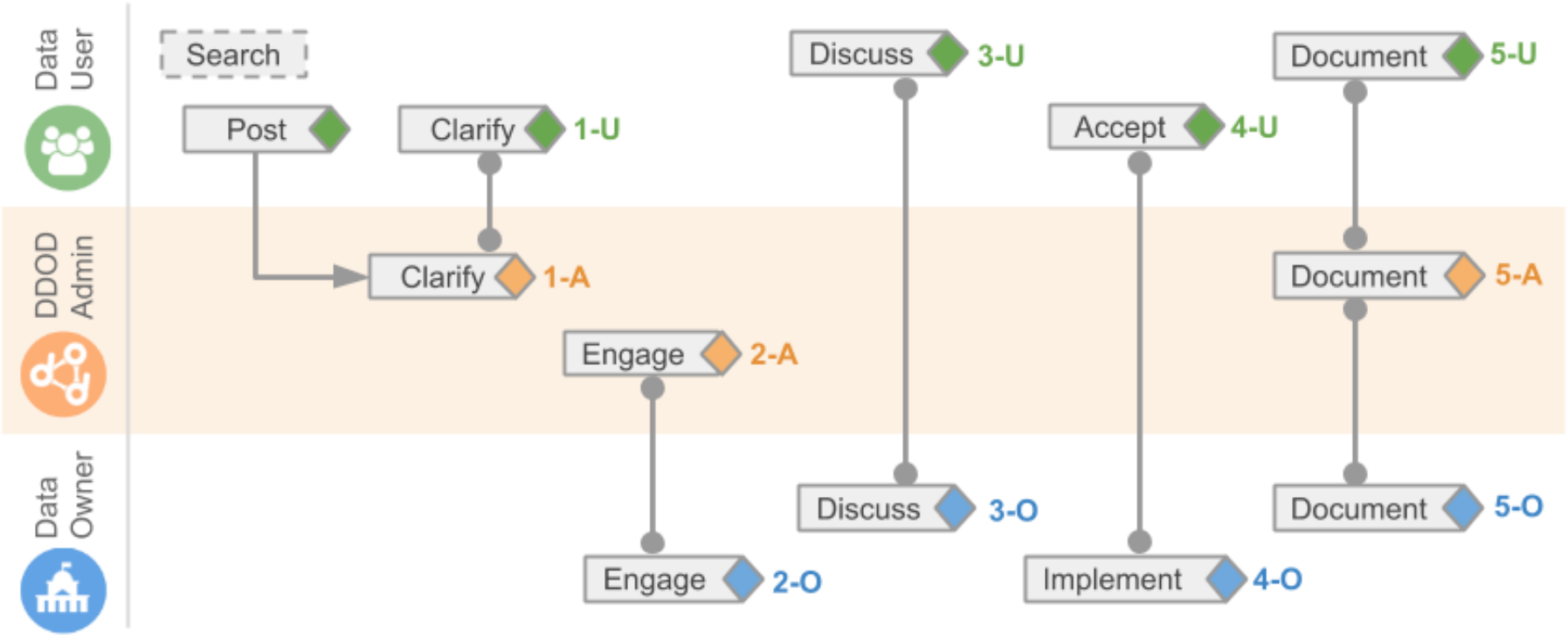
Note that promotion and incentive are a big part of the effort

Promotion	Ensuring there's enough liquidity and uptake	• Program-wide: Vision, publications, partners, events • By use case
Prioritization	Process and methodology for selecting	• Distributed requirements management • Calculation, announcement, decision transparency • Data sources & transparency: EDI, legislated, FOIA, etc.
Technology	Enabling gathering of input, selection and feedback loop	• Qualitative feedback • Quantitative non-monetary • Quantitative monetary
Execution	End-to-end pilot for improving 1 dataset and adding 1 dataset	• End-to-end practice of all components coming together • Allow execution of demand-driven Linkages
Incentive	Providing reason for implementing highest value work first	• Feedback mechanism, data value report card • Promotion of success stories • Advocate for policy

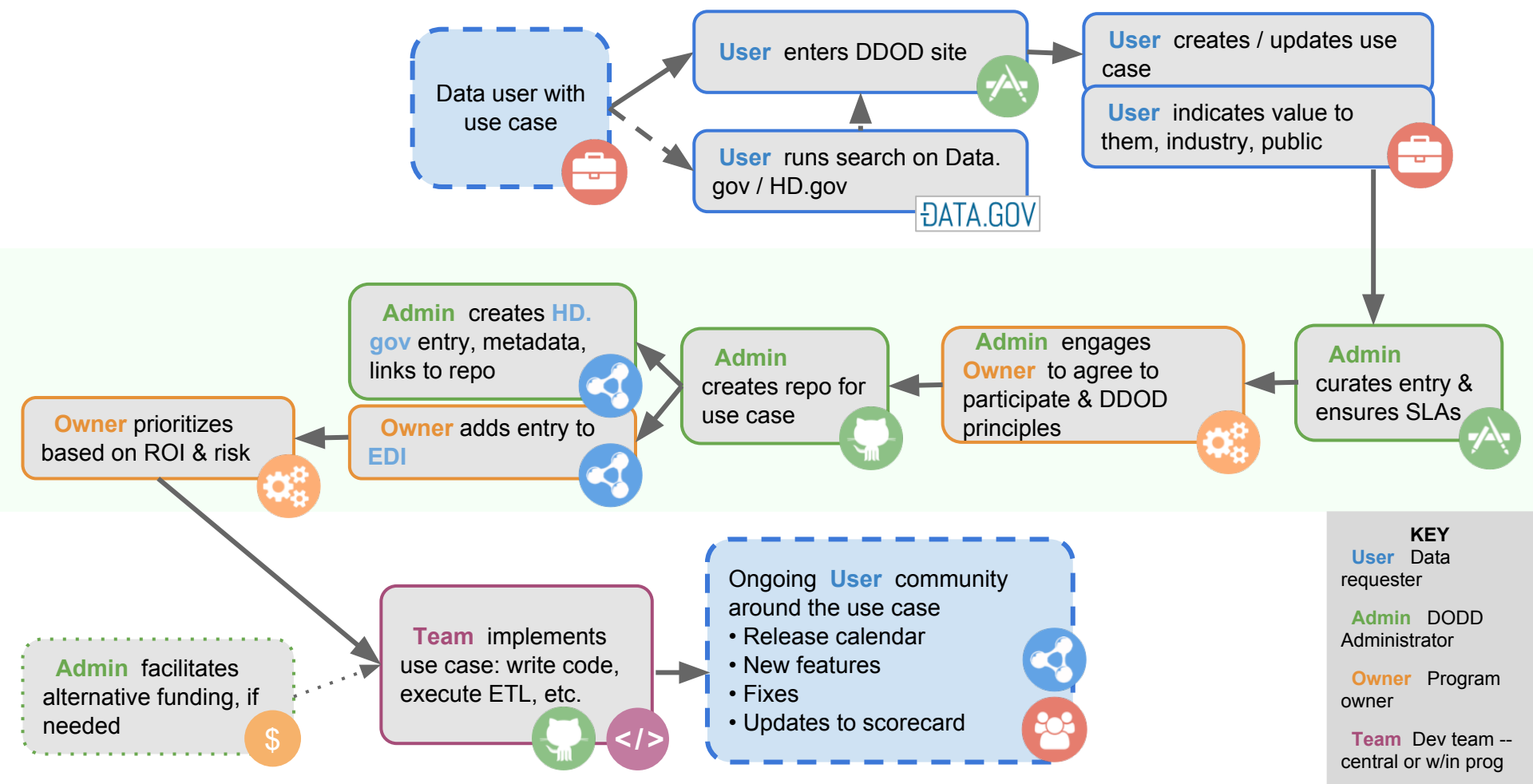
Proposed deliverables for the 12-month effort

Promotion	Use cases & Community
Prioritization	Process & org structure
Technology	Tools & Source code
Execution	Datasets & Success stories
Incentive	Observations & recommendations

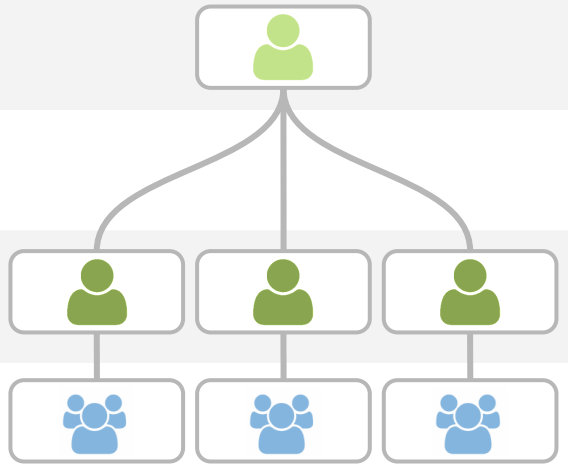
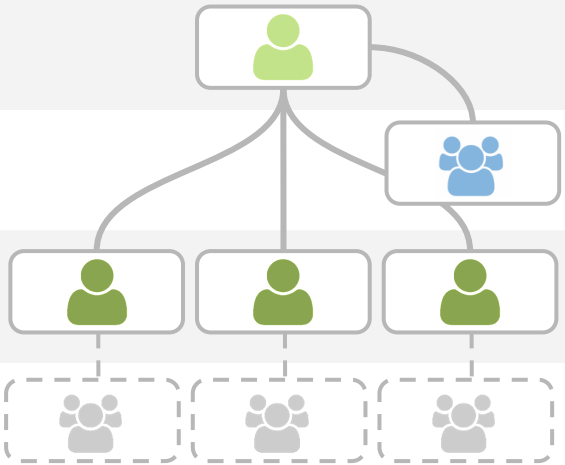
The process involves 3 participants: Data User (that's you), Data Owner, and DDOD Admin. Each is responsible for enabling a specific set of milestones



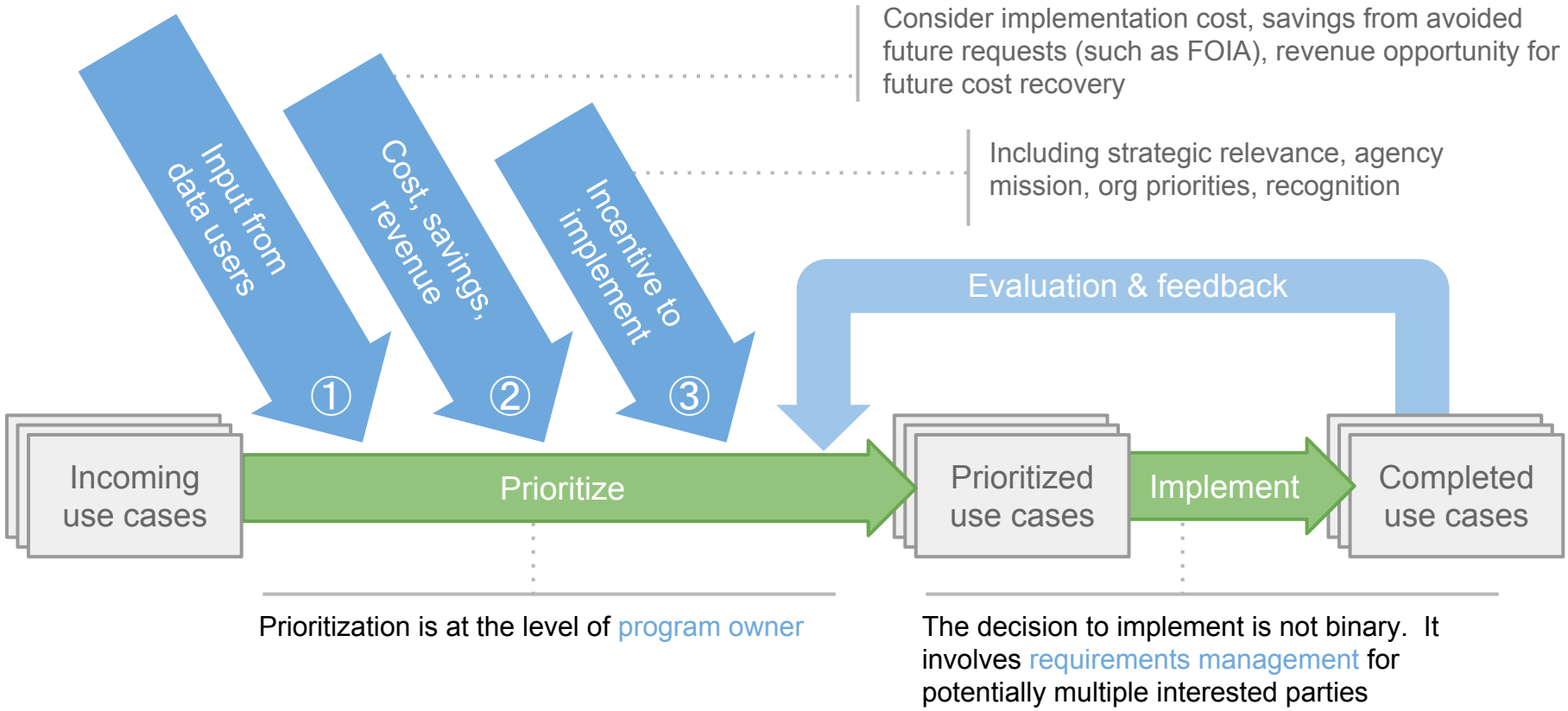
End-to-end workflow from use case to live project



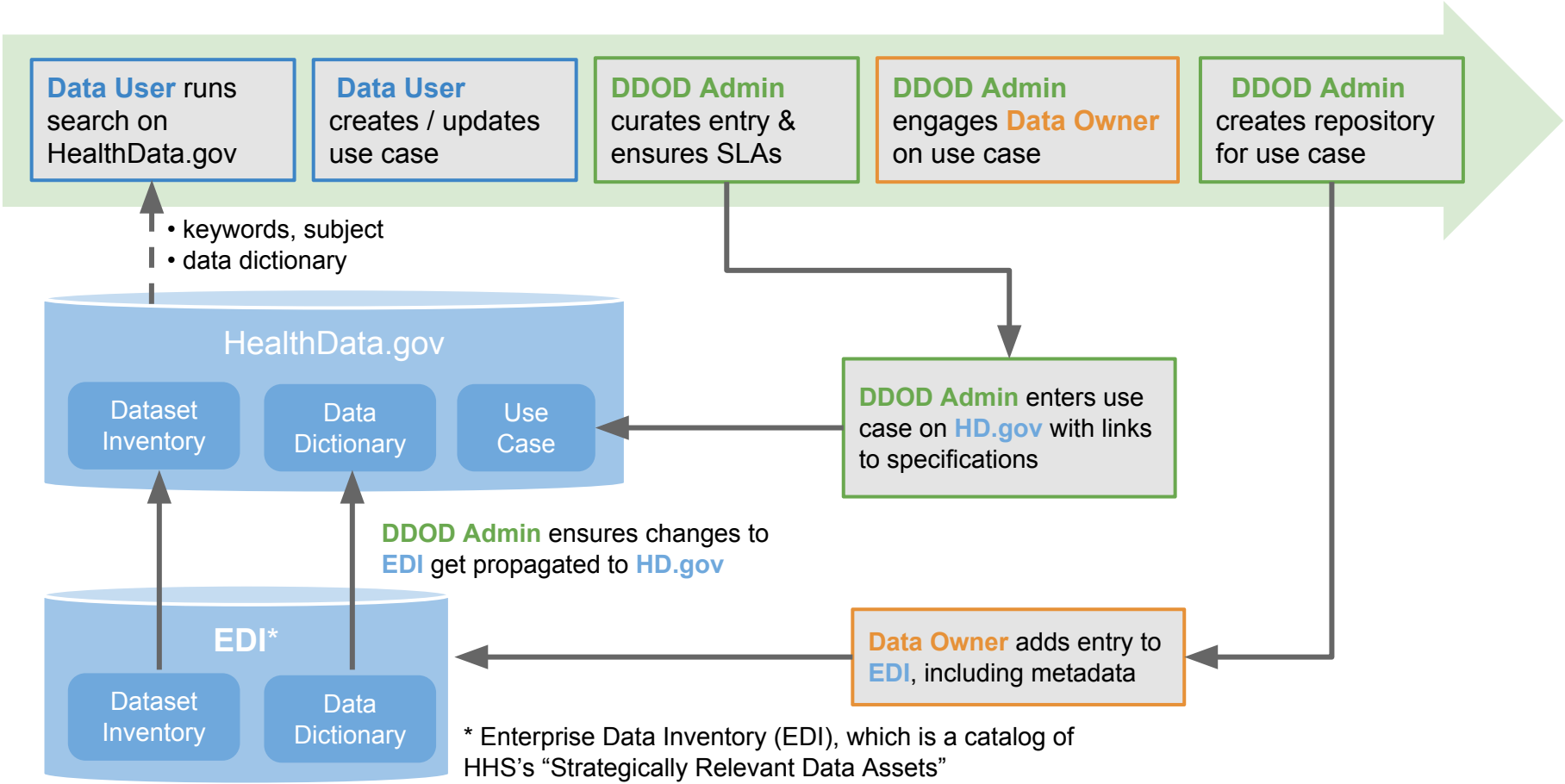
There are several options for organizational structures that enable DDOD, but ultimately the organizational dynamics are unique to each use case and program owner

	Decentralized implementation, with allocated team in each organization	Centralized implementation team with department-wide buy-in
Centralized DDOD manager		
Centralized dev team		
Health Data Leads, Program owners		
Dev team at level of program or data owner		

Prioritization of use cases is determined by 3 drivers



The DDOD process also serves to enhance content and discoverability for HealthData.gov, as well as ensuring key systems of record are in the EDI



Technology

The tools available would evolve

	Now	Short Term	Medium Term	Long Term
DDOD	(None)	• StackExchange, Github, Google Forms • Seed with pilot use cases	• Scripting of workflow among existing tools • Integration of tools into HD.gov	• Purpose built application • Methods to calculate quant value
Data maturity scorecard	(None, NCVHS is working on scorecard)	• Evaluate existing HD.gov datasets • Experiment with visualization & metrics calcs	• Improve tagging & search based on maturity factors	• Integrate features into relaunched HD.gov • Automation to spot drastic shifts in quality
Data activity scorecard	(No tracking of traffic or downloads for indexed sites)	• Get Google Analytics access to indexed sites. • Get CKAN / Socrata stats access	• Fortify GA with in-page analytics and custom dimensions for CKAN • Require registration	• Automation to spot drastic shifts in usage
Metadata Inventory	(Only data that's published has metadata)	• Merge EDI PDL into DH.gov -- regardless of availability or restrictions	• Add metadata from FOIA requests, procured studies and research	• Extend common core metadata with domain specific fields • CISP-like linkage visualization

Technology

Possible initial tool implementation

Github: Central DDOD Repo

Issues	Use Case, associated discussion about value, prioritization
Pages / Wiki	Specifications by use case (until moved to dedicated repo)
Tags	Unique Use Case name, program subject area

Move to dedicated repo after owner buy-in

Github: Dedicated repo for project / use case

Code	ETL, API
Issues	Features, bugs
Wiki	Reqs, docs

Add metadata to HD.gov, even before implemented

HealthData.gov

- Metadata
- Link to dedicated Github pages

Anonymous input
Calculated values

Common tag naming

Google forms

Form	Safe & anonymous data collection
Sheets	Used for calculation Publicly visible but with protected fields

Ops & support discussions moved to Stack Exchange

Stack Exchange

Users	Reputation based voting & support
Q & A	Discuss how to implement, fix, interpret
Tags	Unique Use Case name, program subject area

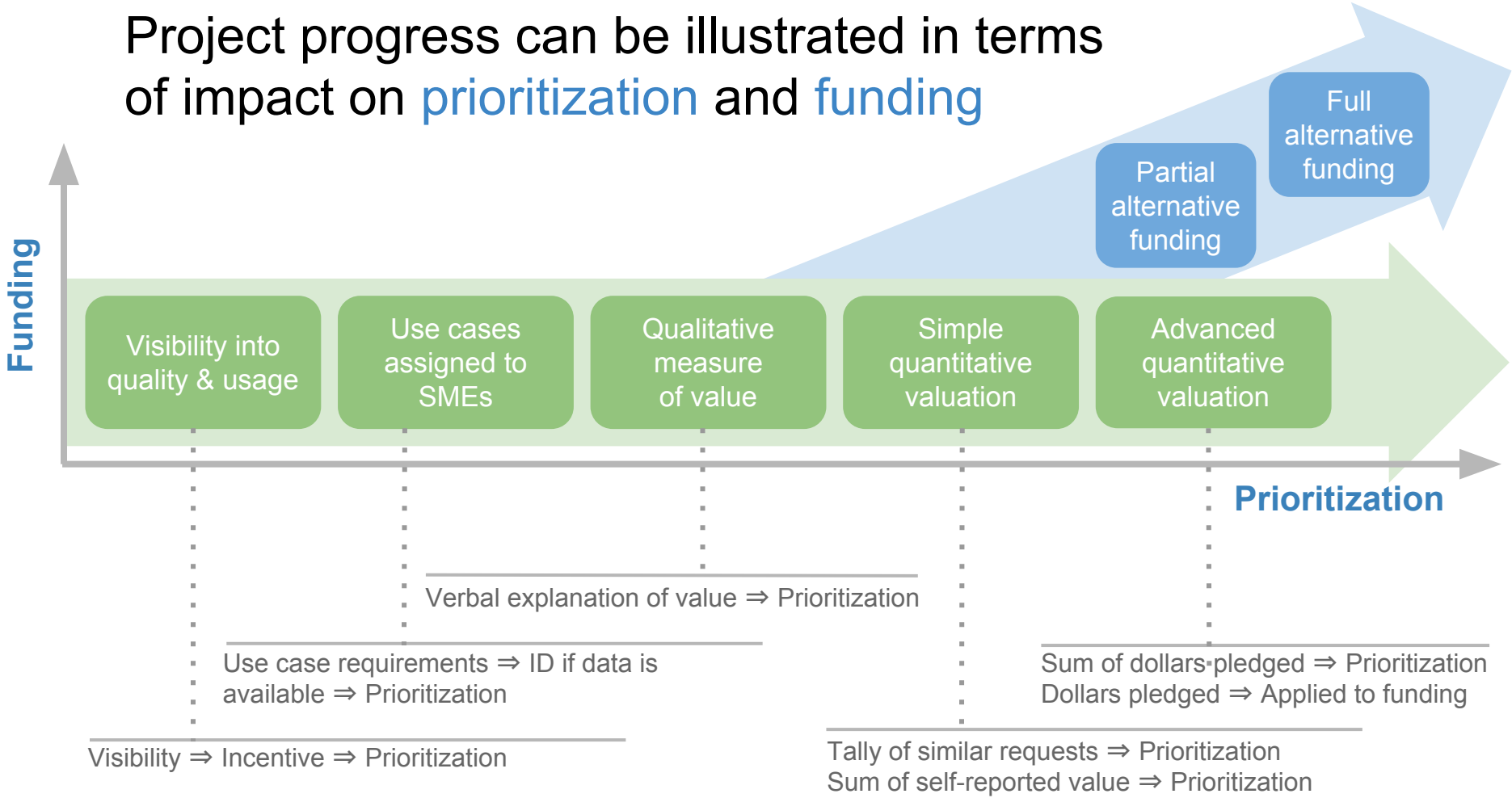
Pivotal Tracker

- Sprint / release planning

Workflow scripting

- Automate calcs
- Consistency / sync

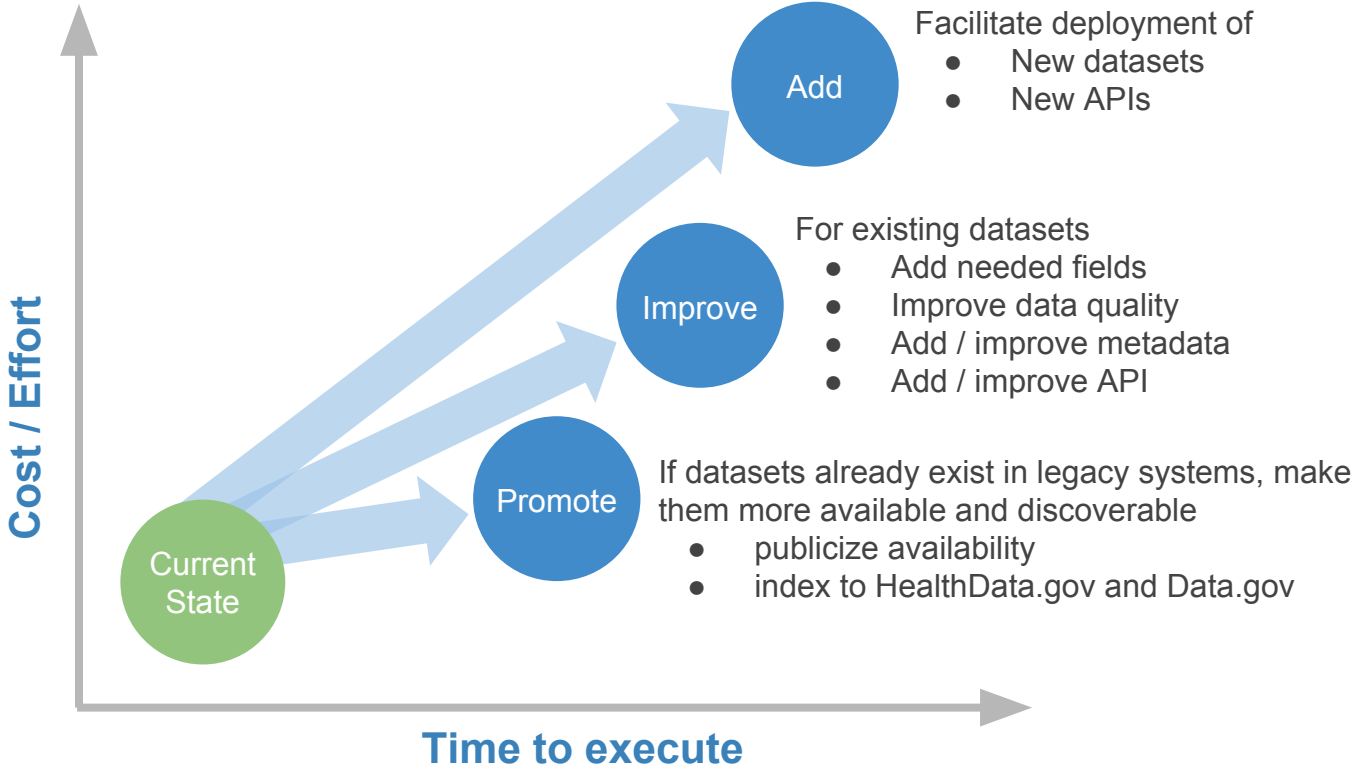
Project progress can be illustrated in terms of impact on prioritization and funding



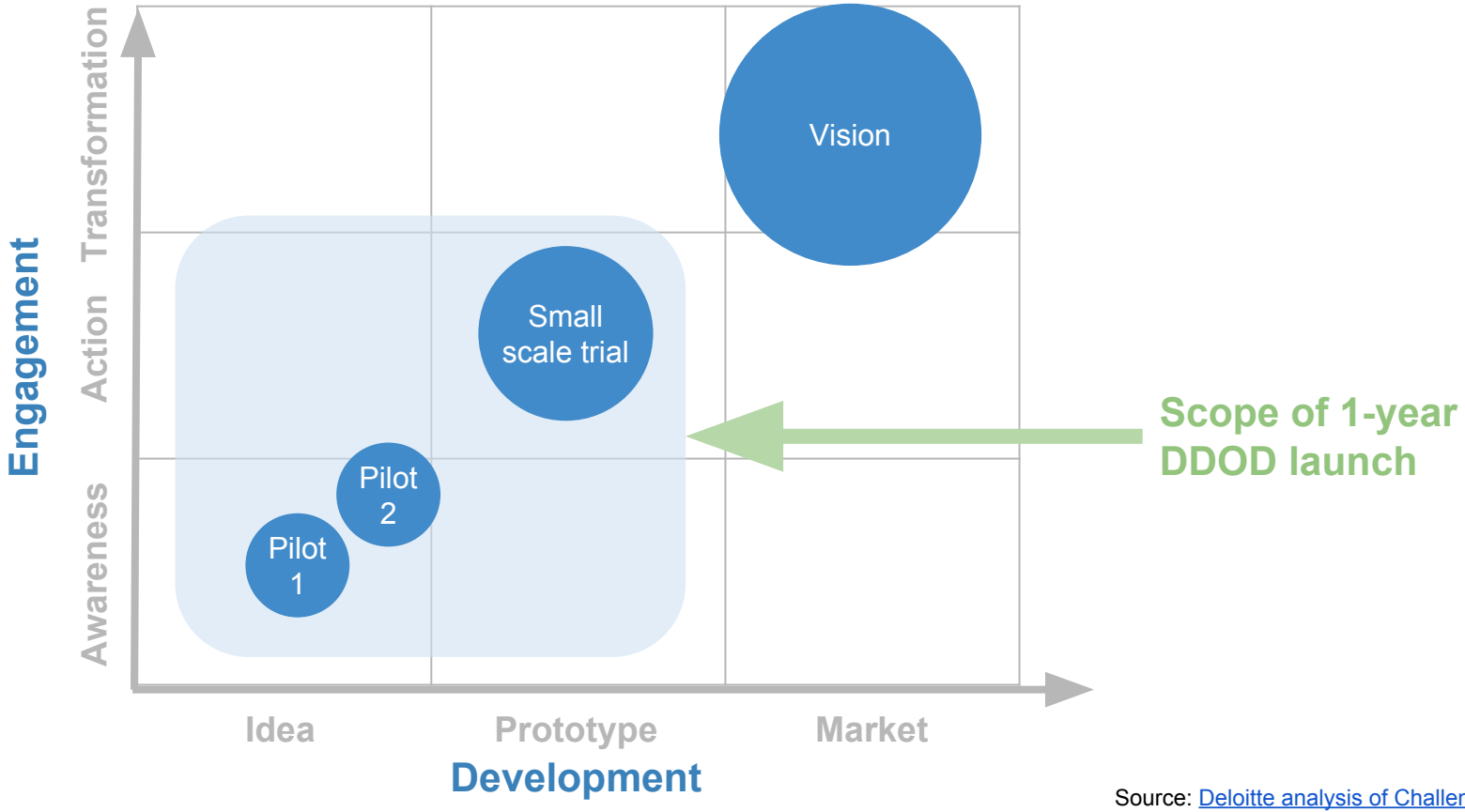
There are both qualitative and quantitative methods for prioritizing use cases

Qualitative	Quantitative
Prioritization based on self-reported descriptions from questions provided on a form with each new use case or feature	While absolute quantitative valuations might be difficult, it's possible to use known objective factors to assess relative value
1. What's the value to your organization? 2. What's the value to industry? 3. What's the value to public health? 4. How consistent is it to the mission of the agency? 5. How time sensitive is the request? What are the impacting factors?	1. Cost already spent on a procured study or survey requested 2. Revenue from cost recovery programs 3. Avoided costs from FOIA requests, manual periodic releases, etc. 4. Crowdfunding-style cummulative pledges from multiple parties as a proxy for value

Implementation of a use case could fall into one of 3 categories



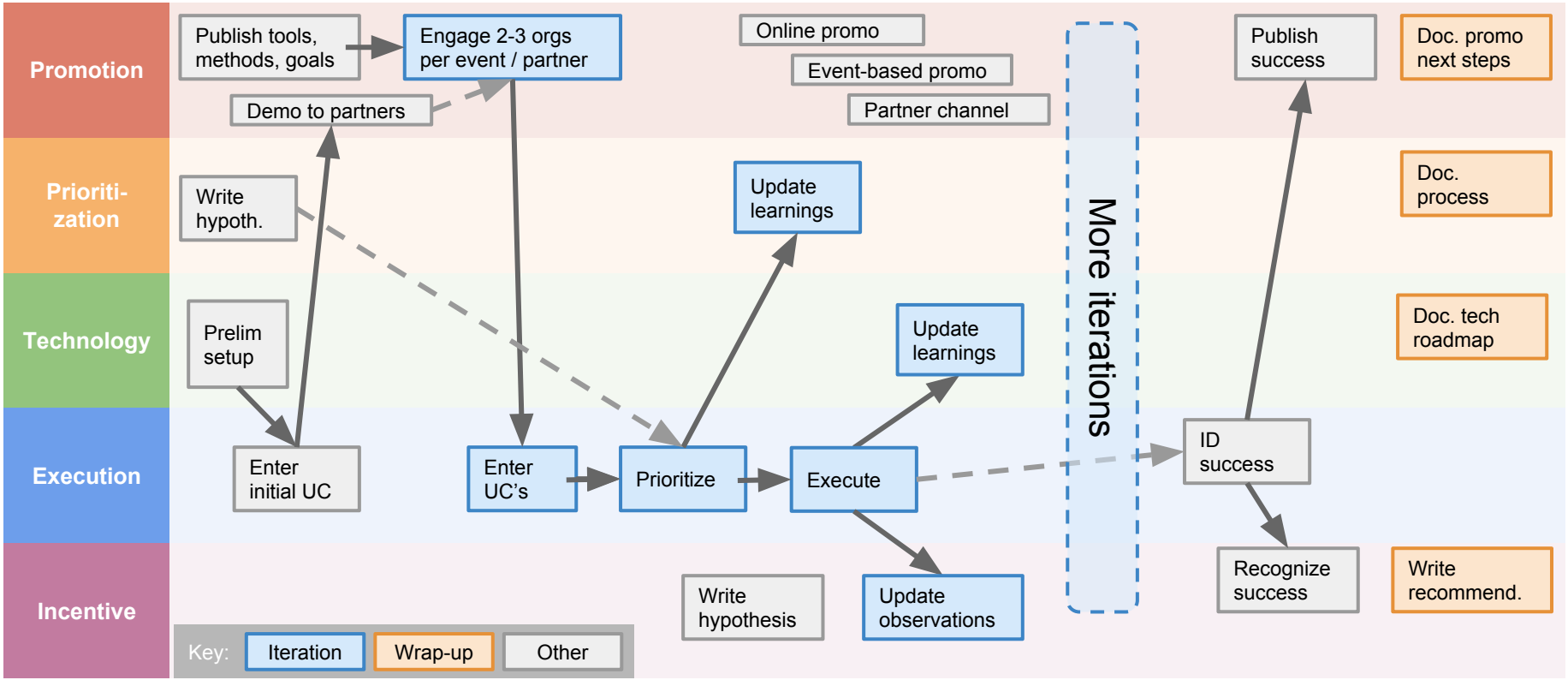
DDOD initiative can be categorized by phases along the Development and Engagement dimensions



Source: Deloitte analysis of Challenge.gov

Sequence of steps for 1-year DDOD launch scope by functional area

Setup. → Iterate, iterate, iterate! → Recommend sustainable model.



There are related initiatives that would make DDOD more effective...

One measures the usefulness of existing data, while the other enables users to discover new data

DDOD ✓	Signaling of **demand**	Enables ① *systematic* and ② *ongoing* and ③ *transparent* signaling of relative value of data for the ④ *full range* of market participants
Data maturity scorecard, Data activity scorecard	Usefulness of **supply**	Has a feedback loop on the *usefulness* of existing and future data
Full metadata inventory	Discovery of **supply**	Enables users to discover the *possible applications* for data, regardless of its privacy classification or availability

DDOD relies on signing up data users to advocate for their use cases, participate using DDOD tools and provide effectiveness feedback

Indicate your interest by sending us an email or submitting the participant form

We'll get you going, starting with a discussion that covers:

- Requirements for your use cases
- Criteria you use for prioritization

As we go about working on your use cases, you'll leverage the DDOD tools and processes for requirements management, voting and community engagement

You submit verbal and written evaluations of the DDOD tools and processes