Building a Data Science Innovation Lab at GAO

Digital Government Institute | 930Gov Conference

September 6, 2023

Andrew Kurtzman

Assistant Director for Data Science

Overview of Content for Today

- Portfolio Management Process

- Contributing to GAO's Mission
    - Examples
    - Alignment with GAO's Strategic Plan

- Project Management Process

1. Overview of GAO and the Innovation Lab

What is GAO's Innovation Lab?

- The Government Accountability Office (GAO) provides Congress, the heads of executive agencies, and the public with timely, fact-based, non-partisan information that can be used to improve government and save taxpayers billions of dollars. This is usually at the request of congressional committees or subcommittees or is statutorily required by public laws or committee reports

- The Government Accountability Office (GAO) is perhaps best known for evaluating the effectiveness of Executive Branch programs and publishing audit reports based on these evaluations.

- GAO's Innovation Lab uses data science and technology to enhance GAO's oversight and internal operations capabilities.

The Lab has grown rapidly but scalably

- We grew from 3 personnel in 2020 to about 20 in 2023: mathematical statistician (OPM 1529), general engineer (OPM 801).
  - GAO has about 3,200 employees.
- We developed data-rich case studies and presentations to evaluate candidates; select based on communication and team skills as much as technical skills
- We emphasize adjacent skillsets (deep research, data engineering) and train extensively
- We reward staff for challenging and improving our process, documenting new processes and create open fora for doing so
- We established contracting vehicles to onboard supplementary skillsets

Maintaining innovation is a challenge and requires managing major risks.

- Major Risk 1: rejected as "foreign entity"

- Major Risk 2: conform to normal operations and produce normal agency outputs

- Solution:
  - define and preserve distinct culture
  - establish portfolio-level and project-level processes to link to broader agency
  - hire staff with the correct primary and adjacent skills to succeed

1. Portfolio Management Process

The Lab maintains a distinct culture within GAO

- GAO culture closest to "order": "respect, structure, and shared norms…. Employees are united by cooperation; leaders emphasize shared procedures and time-honored customs."

- Lab culture closest to "learning": "exploration, expansiveness, and creativity…. Employees are united by curiosity; leaders emphasize innovation, knowledge, and adventure."

Source: The Leader's Guide to Corporate Culture, by Groysberg, Lee, Price, and Cheng, Harvard Business Review, January 2018

The Lab's portfolio management process lets us experiment safely and orderly

The Lab's portfolio management process is:

- Intended to align resource investment with level of certainty and size of potential impact

- A "funnel." Most ideas that start development will never enter production.

- Intended to align the Lab's work with GAO's strategic goals.
    - Work will only be accepted if the type of work is within the Lab's capabilities;
    - and if the likely project will likely have a positive impact related to how our client meets GAO strategic goals.

Portfolio Management Process Overview
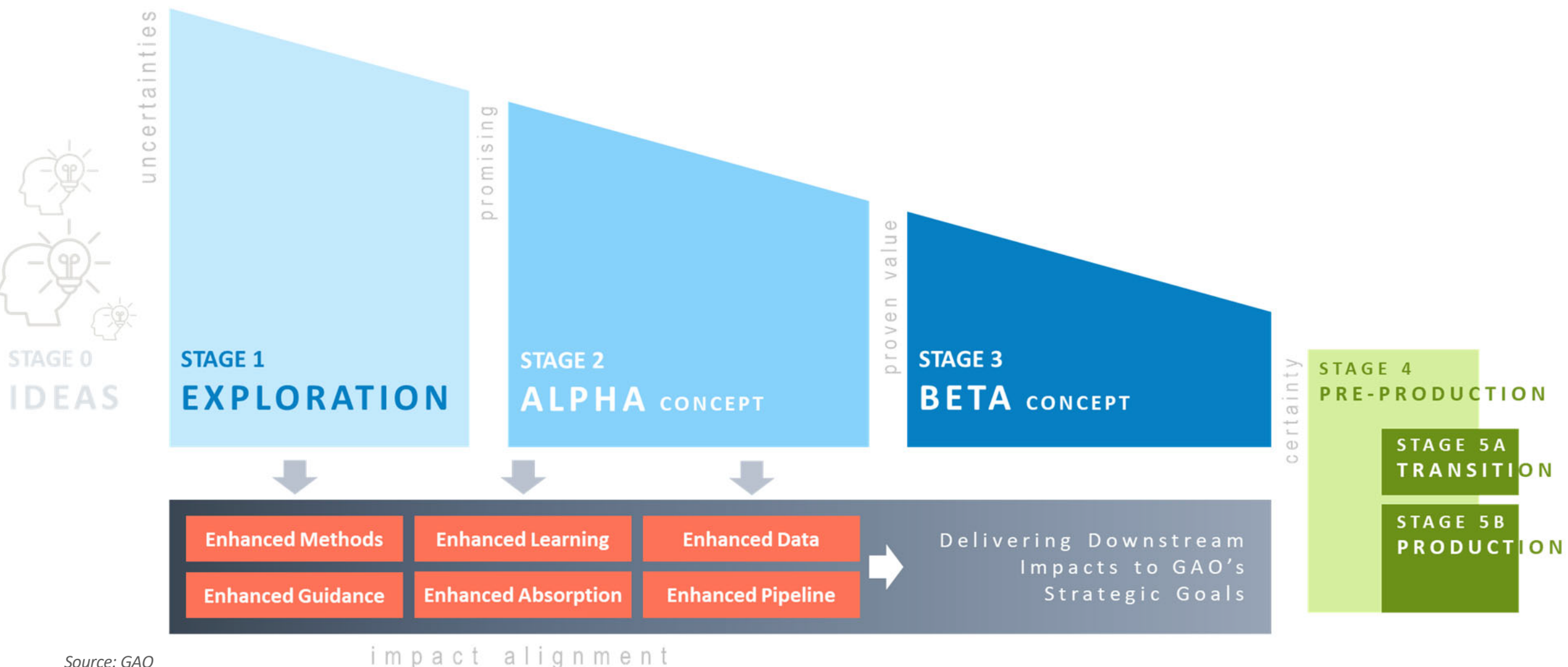
Source: GAO

Each Stage in the Process has specific goals and revisits the "Go / No Go" investment decision

- **Exploration** ends with agreement (or not) on stakeholders and a concept

- **Alpha** ends with agreement (or not) that key components of the concept have been proven and we should invest in a functional system

- **Beta** ends with agreement (or not) that it is worth investing in the system to make it secure, scalable, designed to match end-user needs

- **Preproduction** ends with agreement (or not) on specific roles and responsibilities and technical details

2. Contributing to GAO's Mission

Impacts of Lab Work on GAO Mission

- **Enhanced Efficiency**: replacing manual audit work with scripted programs to ensure speed, consistency, and quality of audit output.
- **Enhanced Analysis**: introducing new methods, data, and guidance suitable for use by future audits.
  - *Methods* that may enhance audit findings and recommendations.
  - *Data* that may help inform active audits or the decision to start a new audit.
  - *Guidance* that may serve as criteria or to define scope for audits.
- **Enhanced Absorption**: supplementing delivery of audit guidance and other GAGAS products through interactive tools that illustrate key concepts and conclusions.
- **Enhanced Pipeline**: identifying promising new areas to audit, including by developing tools to persistently monitor for audit opportunities.

Enhanced Efficiency

- Automated financial audits (Current Stage: Delivered)
- Automated Congressional monitoring and alerts (Current Stage: Beta)

Enhanced Analysis

- Methods
  - Machine Learning Precedence Paper: Topic Models (Current Stage: Beta)
- Data
  - Mortgage Database (Current Stage: Alpha)
- Guidance
  - AI Oversight Framework (GAO-21-519sp) (Current Stage: Delivered)
  - Identity Verification Practices to Prevent Improper Payments (JFMIP-22-1) (Current Stage: Delivered)

Enhanced Absorption

- Identity Verification Controls Simulator (Current Stage: Delivered)
  - Takes concepts from associated report (JFMIP-22-1) and guides users through building a complete control system for a hypothetical program.
  - gaoinnovations.gov/id_verification

Assessing Project Fit with ILX Project Impact Alignment Matrix

		Client Strategic Goal Impacted by Lab Work				
		Goal A	Goal B	Goal C	Goal D	Goal E
Type of Impact Achieved	Efficiency	Project 1				
	Methods			Project 4		
	Data		Project 2			
	Guidance					Project 6
	Absorption		Project 3		Project 5	
	Pipeline					Project 7

3. Project Management

Projects are managed with "Agile Data Science"

- The Innovation Lab uses Scrum methodology for Agile data science
- Agile data science is financed on technical debt
- Data scientist staff are generalists (research, data science, data engineering)
- All Lab data science projects follow a biweekly sprint schedule that includes development and customer interaction
- Customers may schedule meetings with Lab staff at their convenience and / or join us for our standard meetings.
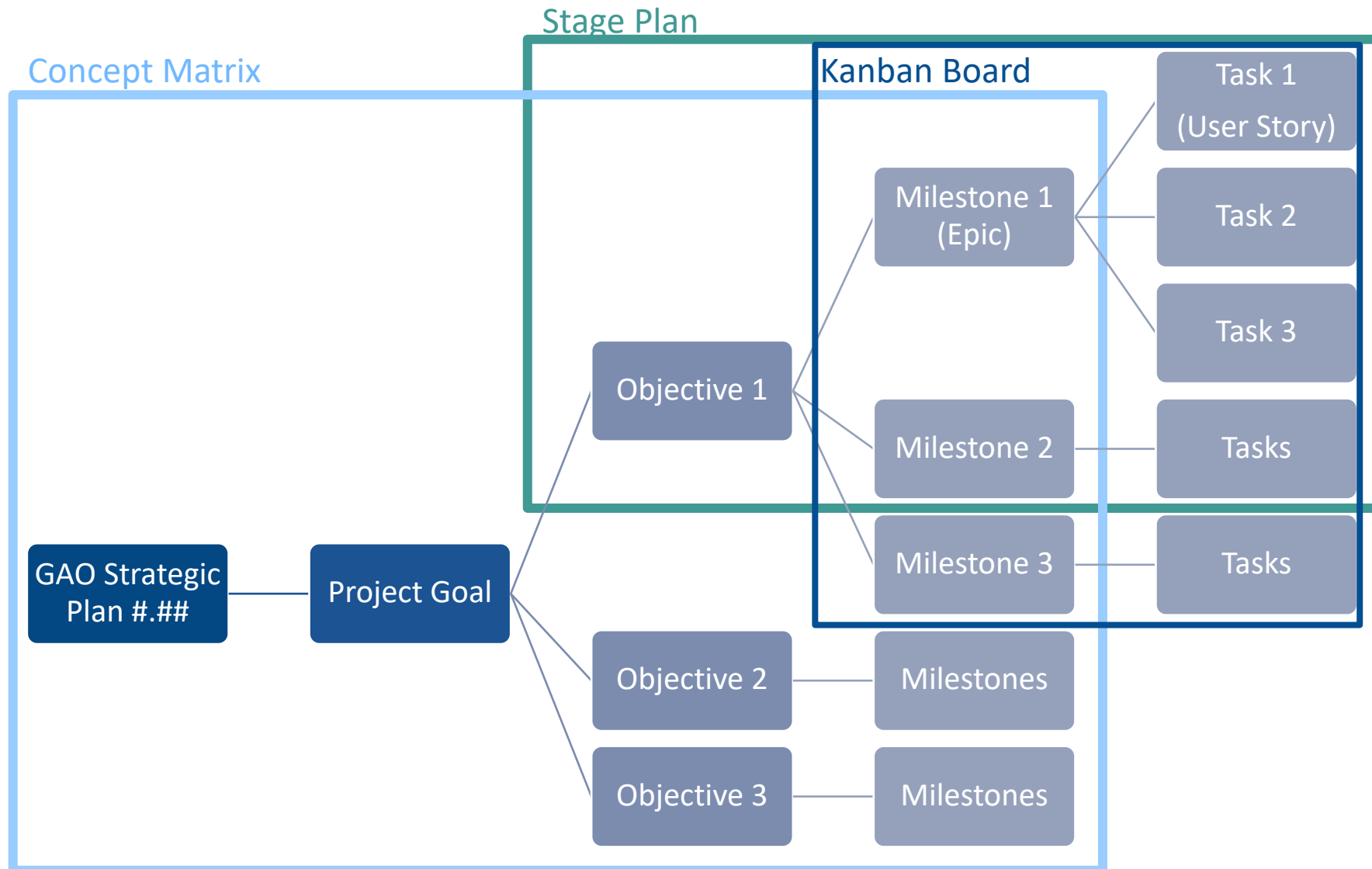
Goals of Project and Portfolio Logic Model

- Scrum tasks on a project should have a logical relationship to objectives and goals

- Project plans must have flexibility to change but must remain logically coherent

- Phases of a project should be expressible in terms of the logic model

Project and Portfolio Logic Model

Project and Portfolio Logic Model: Artifacts

Biweekly Sprint Schedule

	Week 1					Week 2					Monthly
	Mon	**Tue**	**Wed**	**Thu**	**Fri**	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	
Planning		x									
Standup			x	x	x	x	x	x	x	x	
Scrum of Scrums				x		x		x		x	
Backlog Refinement	x										
Client Meeting	x										
Lab-Wide Retrospective											x

Goal: By 10:20am every day, all staff on all projects should know what they are doing for the day and management should understand what institutional resources and permissions are needed.

4. Management of Internal Controls

The Lab needs a wide range of internal controls unique to the Lab's processes

Examples include:

- Every project is tracked within the Lab's portfolio management process, including customer and strategic plan alignment

- Required sign-offs are visibly requested and documented

- Projects are making reasonable progress

- Staff and management are visibly associated with each project

- Staff work is balanced and reasonable

- Agency-wide requirements (procurement, hiring, and many more)

The Lab built its controls to allow automatic, continuous self-audit

- Document database containing required forms as generic workflow
- All projects, staff, sign-offs tracked within systems
- Staff, management, stakeholders associated with projects by stage
- Role-based controls for sign-off of documents
- Required forms used to auto-populate project summary pages; internal reports updated in real time
- Project "showcase" for visibility
- Soon: technology dependency and ATO cross-tracking

Questions?

Taka Ariga (Lab Director and Chief Data Scientist): ArigaT@gao.gov

Andrew Kurtzman (Assistant Director for Data Science): KurtzmanA@gao.gov