U.S. Department of Labor (DOL)'s Analytics Platform: Driving Cultural Change By Leveraging Data as a Strategic Asset

Christina Yancey, *Chief Evaluation Officer, DOL*
Scott Gibbons, *Chief Data Officer, DOL*
David R. Judkins, *Principal Scientist, Abt Associates*

FCSM- September 2020

Chief Evaluation Office
U.S. DEPARTMENT OF LABOR

Presentation Agenda

- Christina, DOL Chief Evaluation Office
  - **Introduction And Context**
- Scott, DOL Chief Data Officer
  - **Challenges and Solutions**
- David, Principal Scientist-Abt Associates
  - **External user perspective**

Introduction: Evidence Act and DOL

- Chief Evaluation Office established in 2010 to coordinate, manage, and implement the DOL evaluation program, with 2 operating units:
  - **Evaluation**
    - Plan and oversee research studies (3rd party contractors)
    - Disseminate/publicly post findings and work with stakeholders to incorporate evidence
  - **Data Analytics**
    - Directly conduct analysis of extant administrative data
- Evidence Act builds on existing momentum

DOL's Co-Location of Analytics and Evaluation

- **Culture of collaboration and innovation**
  - Learning agendas, projects, capacity building
  - Evaluation perspectives inform analytics
    - Analytics driven by research questions
  - Analytics perspectives inform and benefit evaluation
    - QA/QC analytic work informs thinking on evaluation suitability
- **Not just intersection of interests, co-evolution**

DOL Data Board OPEN DATA DATA QUALITY DATA CULTURE STRATEGIC VALUE

Chief Evaluation Office
U.S. DEPARTMENT OF LABOR

Case Study for Using Administrative Data at DOL

Analytics platform as tool for-

1. Accessing and combining federal data
   - Repeatable secure data transfers, storage, analysis
   - Generalizable risks and requirements (statutory provisions, security protocols, MOUs)
   - Culture change to build capacity for leveraging data for multiple purposes
2. Evaluator analysis
   - Nimble external user access
   - Varied requirements for tools

Challenges In Leveraging Data As A Strategic Asset

- Resistance to data sharing, rigorous evaluation
- Data are collected as a by product of programs
- We have had little IT consolidation, no governance
- No enterprise analytic framework, tools are ad hoc
- No enterprise emphasis on data-informed decisions
- DOL has trouble retaining Data Scientists
- Staff are often not trained in analysis

Challenges In Leveraging Data As A Strategic Asset

- Resistance to data sharing, rigorous evaluation
- Data are collected as a by product of programs
- We have had little IT consolidation, no governance
- No enterprise analytic framework, tools are ad hoc
- No enterprise emphasis on data-informed decisions
- **DOL has trouble retaining Data Scientists**
- Staff are often not trained in analysis

DOL Data Board | OPEN DATA | DATA QUALITY | DATA CULTURE | STRATEGIC VALUE

Challenges In Leveraging Data As A Strategic Asset

Decision

- Co-develop a dynamic analytical sandbox
- Focus on practical evaluation, analytics use cases
- Select technology consistent with mission, vision, goals, and methods
- Development driven by stakeholders, users
- Feedback loops between collaborative work with agencies and architecture, tools

Challenges In Leveraging Data As A Strategic Asset

Solution:

- **An internal analytical hub that co-locates data and tools**
- **Containerization to rapidly prototype new capabilities**
- **Iterative development of platform components**
- **DevSecOps, Registries maintain variation in tooling**
- **Open source tools to keep costs low**
- **Leverage benefits of user communities**

Addressing The Cause, Not The Symptoms

Q: What do the Evidence Act, FDS asks us to do?

A: Build culture, capacity to leverage strategic value in data

- **Addressing symptoms is easy but addressing the root cause is more complicated.**
- **We need to be honest about limiters, and appropriately design and build services and tools**
- **Federal IT culture makes it challenging to innovate**

We aim to build capacity that:

- **Rather than limiting staff, enables innovation, creativity, and testing feasibility of new ideas**
- **Generates products that resonate with our staff and leaders**
- **Disrupts in a "good" way: supports staff, maintains trust relationships with leadership**
- **Consistent with the change and evolution we seek to create**

Addressing The Cause, Not The Symptoms

Concern # 1: Resistance to data sharing / MOU issues

Approach: Technical challenges < legal, admin issues

- All data are now local, directly controlled
- Bringing researchers in rather than sending data out
- Less time with legal, parochial data mgmt. issues
- Develop comfort, trust with the process
- Cultural Change -> Common Enterprise process
- Example: CEO manages outcome data from NDNH

Addressing The Cause, Not The Symptoms

Concern # 2: No enterprise analytic framework, tools

Approach: Leverage analytic, evaluation work to inform effort

- **Fill that need in ways customers are asking for.**
  - **Embrace CD/CI and varying tool sets, containerization, high frequency deployment, open source analytics tools**
- **Concurrent provisioning of proprietary software for more users**
- **Cultural Change -> Increase in experimentation; less attrition;**
- **Benefit-> Better analytics, science, cost effectiveness, efficiency**
- **Example: Use DevSecOps, Registry to host variations on one tool**

Addressing The Cause, Not The Symptoms

Concern # 3: Limited Staff Skills

Approach: Leverage tools with amazing COPs

- Abundant training templates for open source tools
- Have software champions provide template code
- Training sessions with template code in all platforms
- PUDF repos with code to ingest, weight, benchmark
- Cultural Change -> Why reinvent what works well?
- Faster prototyping; easier experimentation; more trust
- Many of our new services come from ideas on blogs

Addressing The Cause, Not The Symptoms

Concern # 4: Limited use of data to inform programs, planning

Approach: Collaborative work is key to building capacity

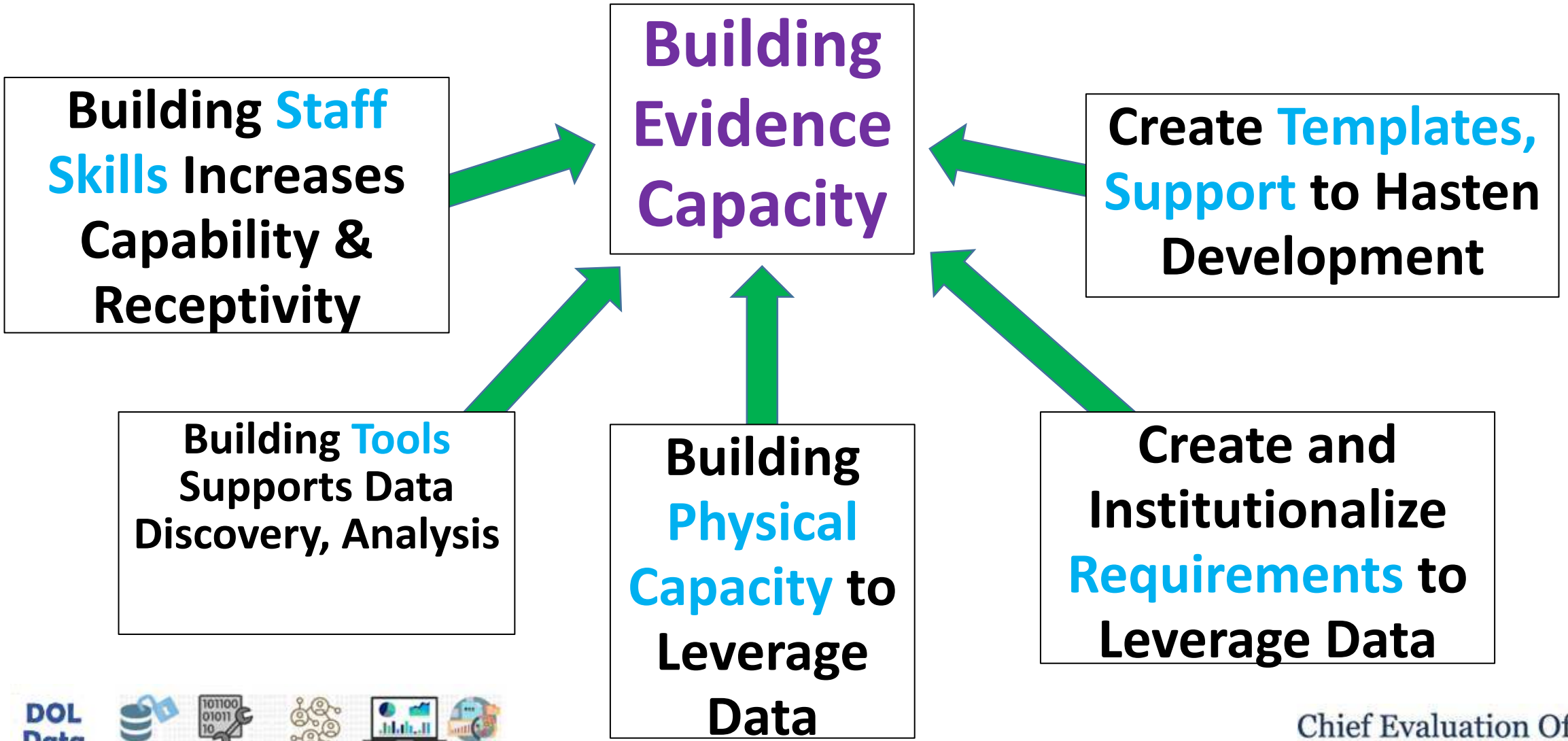
- **Leverage sandbox to host capacity building efforts**
- **Bring program staff into process through research questions**
- **Ensure analysts understand constraints of data product users**
- **Ensure that program staff understand what is possible**
- **Develop mutual understanding of goals, methods, constraints**
- **Exposure to iterative approach builds trust and comfort**
- **Cultural Change -> Successful elimination of real barriers**

Addressing The Cause, Not The Symptoms

Concern # 5: Transitioning to Data Science

- **Advocating person-autonomous, repeatable, consistent**
- **Integrating tools like git, ETL, governance**
- **Training tools is also communicating expectations**
- **Cultural Change -> Transitioning staff to better science, better workflows, more rigor, more transparency**

What Is It That The Evidence Act Asks Us To Do?



DOL Data Board — OPEN DATA · DATA QUALITY · DATA CULTURE · STRATEGIC VALUE

Chief Evaluation Office
U.S. DEPARTMENT OF LABOR

What Is It That The Evidence Act Asks Us To Do?

Analytics capacity is supporting research and evidence

- **Leading culture change; building trust & receptivity**
- **Using favorable experiences with analytics to push towards more rigorous efforts**
- **Bringing value to the enterprise:**
  - **Using analytics to test data for evaluation suitability**
  - **Familiarizing users with the methods**
  - **Proceed up the cascades from descriptive > QED > Causal?**
- **As analytics integrate data into decisions, it lays the groundwork for greater use of evidence in planning, policy**

Abt ASSOCIATES

BOLD THINKERS DRIVING REAL-WORLD IMPACT

DEAP: The User Experience

FCSM 2020

Using Data in New Ways: Leveraging the Evidence Act to Coordinate Evaluation, Statistics and Policy

Research needs

- Both SAS and RStan
- SAS for frequentist analyses
- RStan for Bayesian analyses
- Highly secure processing environment to tabulate data from employer UI tax forms

Why Bayesian?

- Reporting training outcomes for each of 34 programs
- Sample sizes too small at many of these to serve as a useful guide for likely performance of future trainee cohorts
- Bayesian methods specifically designed for this task, including variance estimation
- Similar to small-area estimation in federal surveys

Why RStan?

- Very flexible priors, very flexible models, and post-model processing (e.g., aggregation of individual predictions into program-level means)
- Blistering speed thanks to Hamiltonian Monte Carlo (no U-turn sampling)
- 10-20 times faster than Stata despite use of less congenial priors (most advanced method is blocked Metropolis-Hastings sampling)
- Much easier to program than Bayesian procedures in Stata (at least for my star collaborator, Stas Kolenikov)

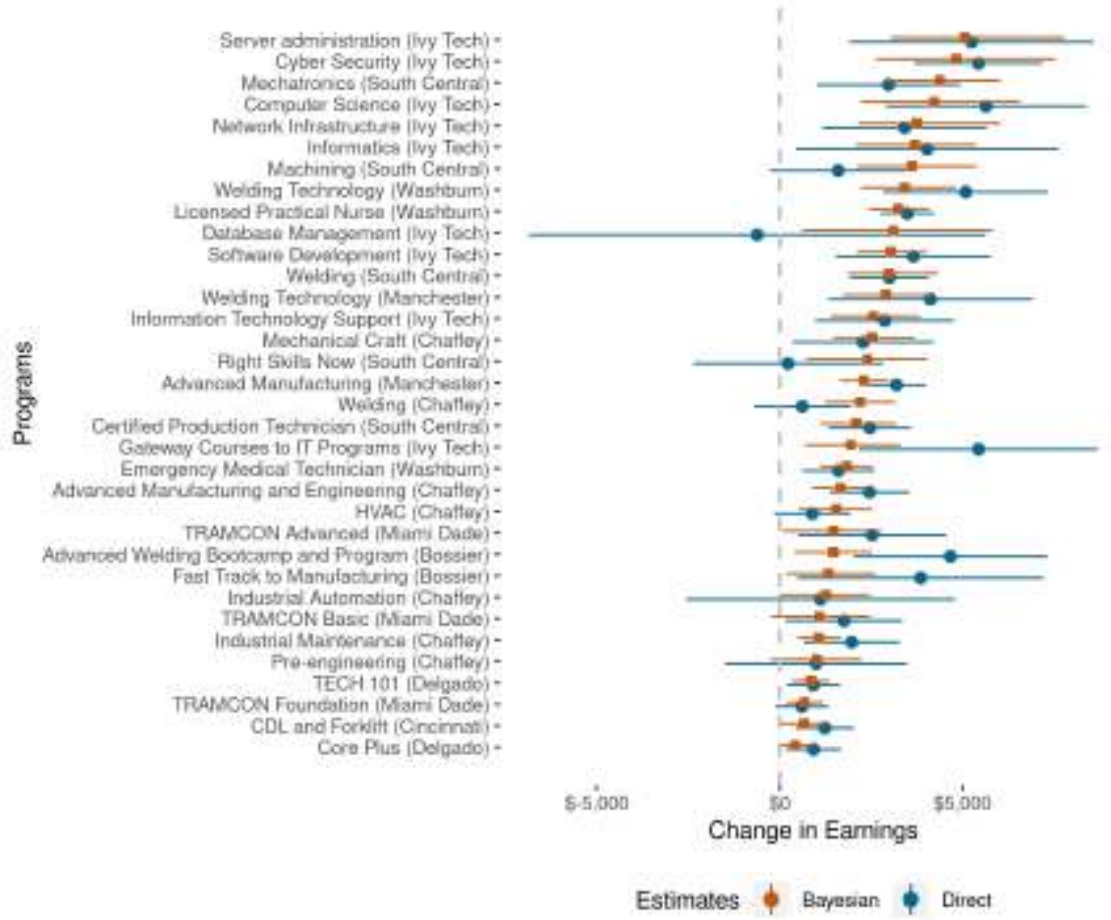
But…

- RStan achieves it speed and flexibility thanks to run-time compilation with a C++ compiler
- This compiler triggers anti-malware software on most systems that prevents successful compilation
- Scott and his collaborators developed a great safe environment with containerization. C++ compilers are dangerous to system security, but with the container approach, we cannot break out and compromise DOL server system

Smooth flexible operations

- With RSA security, workers with proper clearance can use DOL laptops from home
- No need for visits to a research data center
- No need even for locked rooms on contractor premises
- Vetted users are responsible for ensuring that downloaded tabulations and models do not compromise data confidentiality

Beautiful results



Contact

David Judkins

Principal Scientist

David_Judkins@abtassoc.com

BOLD THINKERS DRIVING REAL-WORLD IMPACT

abtassociates.com