

**Introductory Comments:
Opportunities for Accelerating Scientific
Discovery: Realizing the Potential of
Advanced and Automated Workflows**

March 16-17, 2020

Distributed Event Across 8 Time Zones

Daniel E. Atkins

University of Michigan

atkins@umich.edu

Study Committee Chair

Sponsored by [Schmidt Futures](#)

The Study Team

- **Daniel Atkins**, University of Michigan
- **Ilkay Altintas**, San Diego Supercomputer Center
- **Shreyas Cholia**, Lawrence Berkeley National Laboratory
- **Mercé Crosas**, Harvard University
- **Alfred Hero**, University of Michigan
- **Rebecca Lawrence**, F1000 Group, UK
- **Bradley Malin**, Vanderbilt University
- **Lara Mangravite**, Sage Bionetworks
- **Brian Nosek**, Center for Open Science
- **Tapio Schneider**, Caltech
- **Tom Arrison**, National Academies
- **Emi Kameyama**, National Academies

Preliminaries

- **Thank you for your participation. Your expertise is critical to the success of our study.**
- **Thank you for your rapid accommodation of the distributed meeting. (And please be tolerant of glitches.)**
- **The session is being recorded, but only for the Study Team to use in preparing the consensus report.**
- **We would like to keep copies of all your slides to inform the report. We will contact you for permission to use figures from your slides in the report should that arise.**

This Workshop is the Centerpiece of a NASEM Study

- **Given Study Goals: Examine current efforts** to develop advanced and automated workflows for scientific research. **Identify promising research** approaches to promote progress by the community in creating and using more productive research workflow systems.
- Use this **workshop as a primary way to learn** about:
 - **Status of workflow in selected research fields**
 - **Key barriers and enablers** for advanced workflow systems
 - **Emerging opportunities** resulting from their development and use
 - **Other advice from the community they believe relevant to the study**
- Based on this workshop, literature review, and committee expertise **produce a NASEM consensus report including recommendations for strategic actions to be taken by multiple stakeholders to seize the perceived opportunities.**

Continuing the Evolution

- Computers as a **programmed calculator** for science and engineering calculations.
- Centralized, **center-based computers** for general purpose computing including science and engineering.
- Specialized research **supercomputers plus network**.
- Distributed, **ubiquitous computing** and the rise of the all-digital world.
- **Cyberinfrastructure ecologies** of computing, data, information, acquisition, activation and distance-independent collaboration. (collaboratory, e-science)
- **Above systems augmented by artificial intelligence and automation/robotics to make the workflow of research more efficient and productive. (accelerate discovery)**

“Research Workflow” is a Broad Sense, e.g.

- Computation pipelines
- Steering simulation sweeps
- Guide data explorations and analysis
- Supporting human in the loop
- Producing metadata
- Establishing provenance
- Controlling wet and dry lab automation
- Process manage distributed facilities and business process
- *Includes organizational, policy, regulatory, incentive, economic, educational and workforce issues.*

Acceleration includes

- accelerate generation of workflows
- accelerate exchange/reuse of workflows
- accelerate the workflows themselves.

From email with Carole Goble

Motivation for the Study

Why Now?

- Growing **general awareness** of the transformative potential of AI, e.g. [HR 6216, National Artificial Intelligence Initiative Act of 2020](#)
- Growing **activity in science communities** to apply artificial intelligence approaches to a range of research tasks.
- **Significant progress in open science** that can now be leveraged in workflow systems to help make data more findable, accessible, interoperable, and reusable (FAIR).
- The growing **use of contemporary digital lab books, workflow system**, especially in science education, can be built upon.
- **Challenges to reliable and reproducible** research that workflow systems can address
- Hope that the acceleration of scientific discovery will help **meet the many grand challenges** in our world

HR Findings 6216

- Artificial intelligence is a tool that has the **potential to change and possibly transform every sector** of the United States economy and society.
- **Requires a major role** by the Federal government.
- **Lack of clear understanding** of capabilities and potential
- **Researchers have too limited access** to datasets, computing resources, real world test environments
- **Lack of standards and benchmarking** for AI systems
- Requires **interdisciplinary collaboration** and still too many barriers.
- Current Federal **investments are insufficient** to incentivize and support what is required.
- **US education pipeline for AI faces significant challenges.**
- Federal government **must do more.**

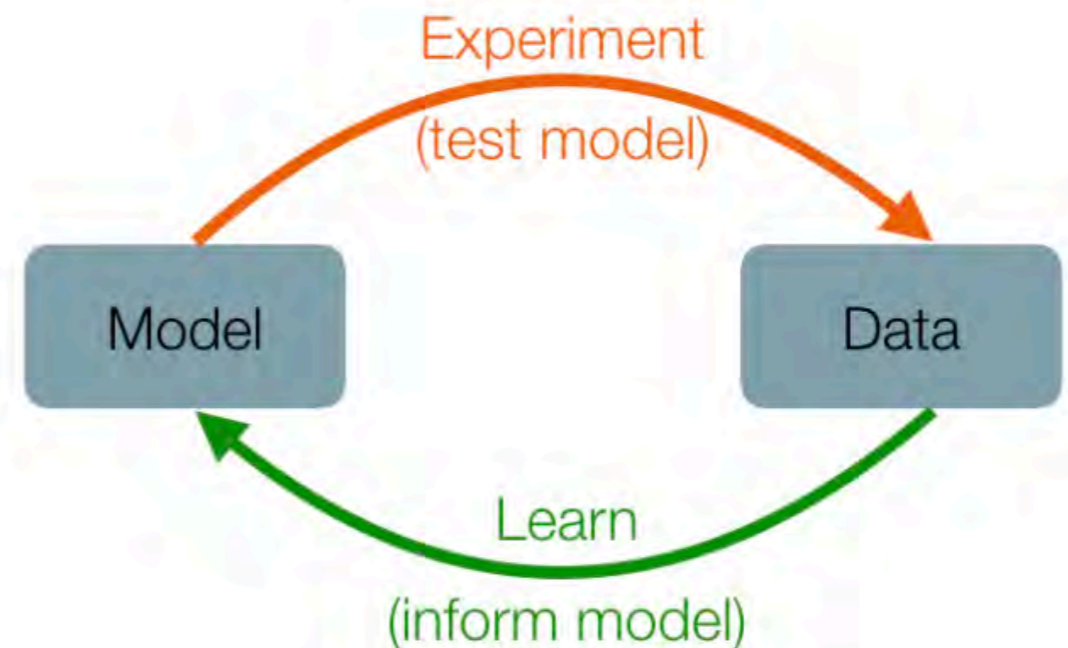
Motivation for the Study

Why Now?

- Growing **general awareness** of the transformative potential of AI, e.g. [HR 6216, National Artificial Intelligence Initiative Act of 2020](#)
- Growing **activity in science communities** to apply artificial intelligence approaches to a range of research tasks.
- **Significant progress in open science** that can now be leveraged in workflow systems to help make data more findable, accessible, interoperable, and reusable (FAIR).
- The growing **use of contemporary digital lab books, workflow system**, especially in science education, can be built upon.
- **Challenges to reliable and reproducible** research that workflow systems can address
- Hope that the acceleration of scientific discovery will help **meet the many grand challenges** in our world

The opportunity more specifically..

- Computerized methods for experimental design and learning about models have **been used for decades**,
- Now in some fields it is beginning to be possible to enclose the entire scientific knowledge generation **loop in an automated system**.
- This has the potential to transform scientific fields because it can accelerate iteration of the knowledge generation loop by **orders of magnitude**.



per Tapio Schneider

- The opportunities for accelerated progress here **arise because** increasingly in the sciences, models are expressed as **computer models**, the acquisition and processing of **data involves computation**, and experimental apparatuses (e.g., gene sequencers) are automated and **controllable by computers**.
- Given a model, not only one but **thousands or more experiments** can be designed automatically and can be optimized to be maximally informative about the model if they can be carried out by automated systems.
- The **learning step** can likewise be **automated**.
- Algorithms for designing experiments to be maximally informative and for learning about a model from data can **become broad purpose and transcend** individual scientific fields, and be delivered by automated platforms.
- Models and data, and the means of acquiring data, will **remain domain specific**

per Tapio Schneider

Synthesis: What have we learned to inform our report?



atkins@umich.edu

Meeting Protocols

- Panel Chairs will be moderating and calling call upon panel members to speak. Panelist, *please watch your own time and do not go over designated speaking time*. During discussion the panel leaders will moderate and steer the discussion but they may also decide to use the Zoom “hand raising” feature.
- You will find this feature by clicking on the *Participant* tab on your screen, then clicking *raise hand*. Panel chairs will need to monitor the raised hand panel on their screen. Click again to lower hand after speaking.
- We are asking those other than panelist to NOT use the hand raise feature but rather the Zoom Chat feature to send questions which we will moderate and try to address as time permits.

This is a Study By and For the Research Community

- We Need Your Help
- Please suggest a few key references.
- Please send written comments after the workshop: atkins@umich.edu
- Thank you

Stuart Feldman, Chief Scientist

