



AI and workflows for accelerated science

Beth Plale

Indiana University

NASEM Opportunities for Accelerating Scientific Discovery

March 16-17, 2020

Topics

- Openness as foundation
- Reference workflow-driven lifecycle; illustrative
 - AI distributed, not all workflows equal
- Upstream
 - Reducing barriers
- Downstream
 - MVP for reuse
 - Rights to scientific result
- Recommendations



Open Science



Increased science *rigor* and *transparency*

Open access to publications from federally funded research

Brings greater visibility to failed results



Increased return to society

Increased *impact* and visibility of research

Opportunities for citizen engagement

Provides credit to the researcher



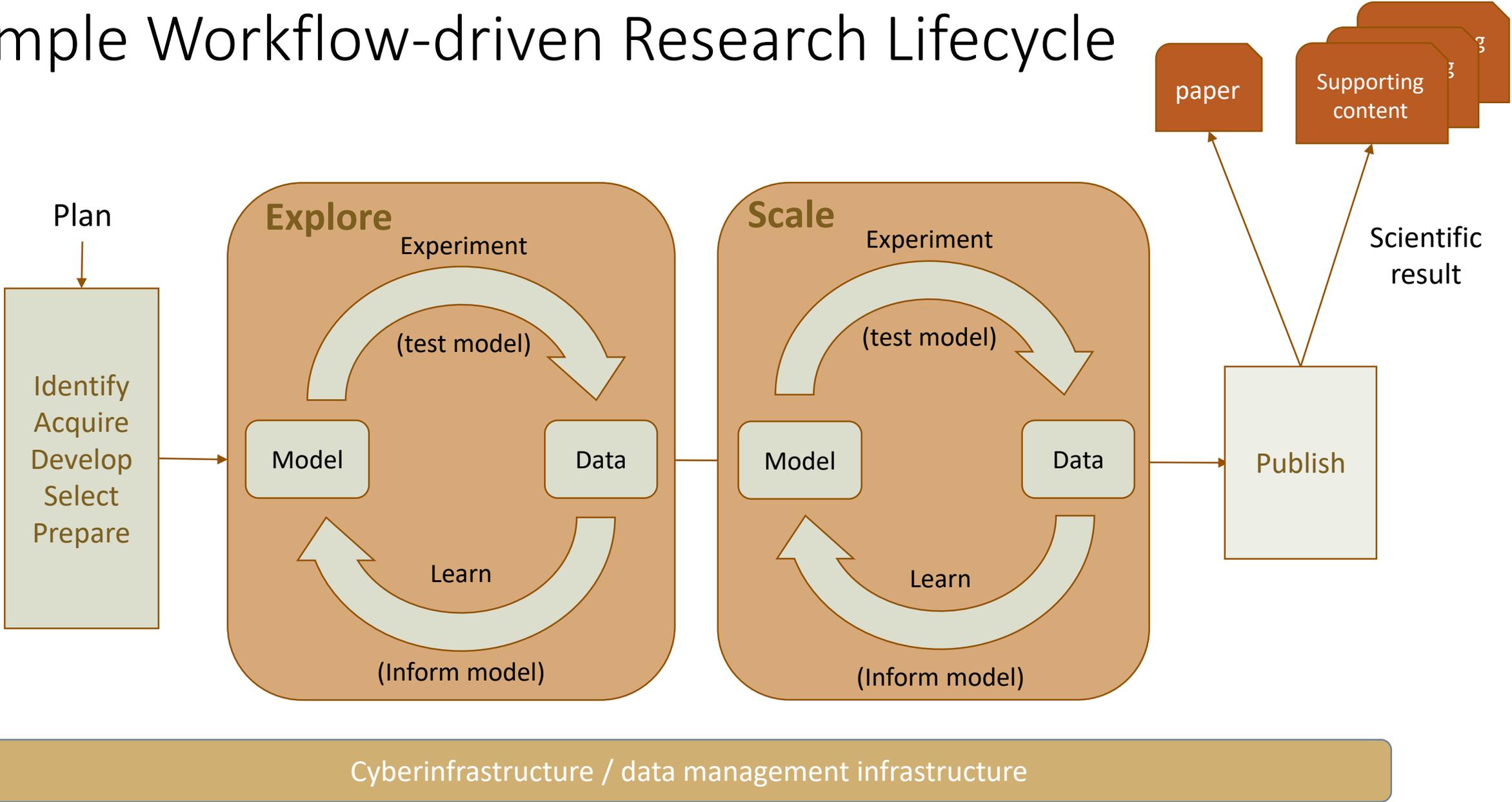
Increased return to science

Reduces cost of duplicating data collection

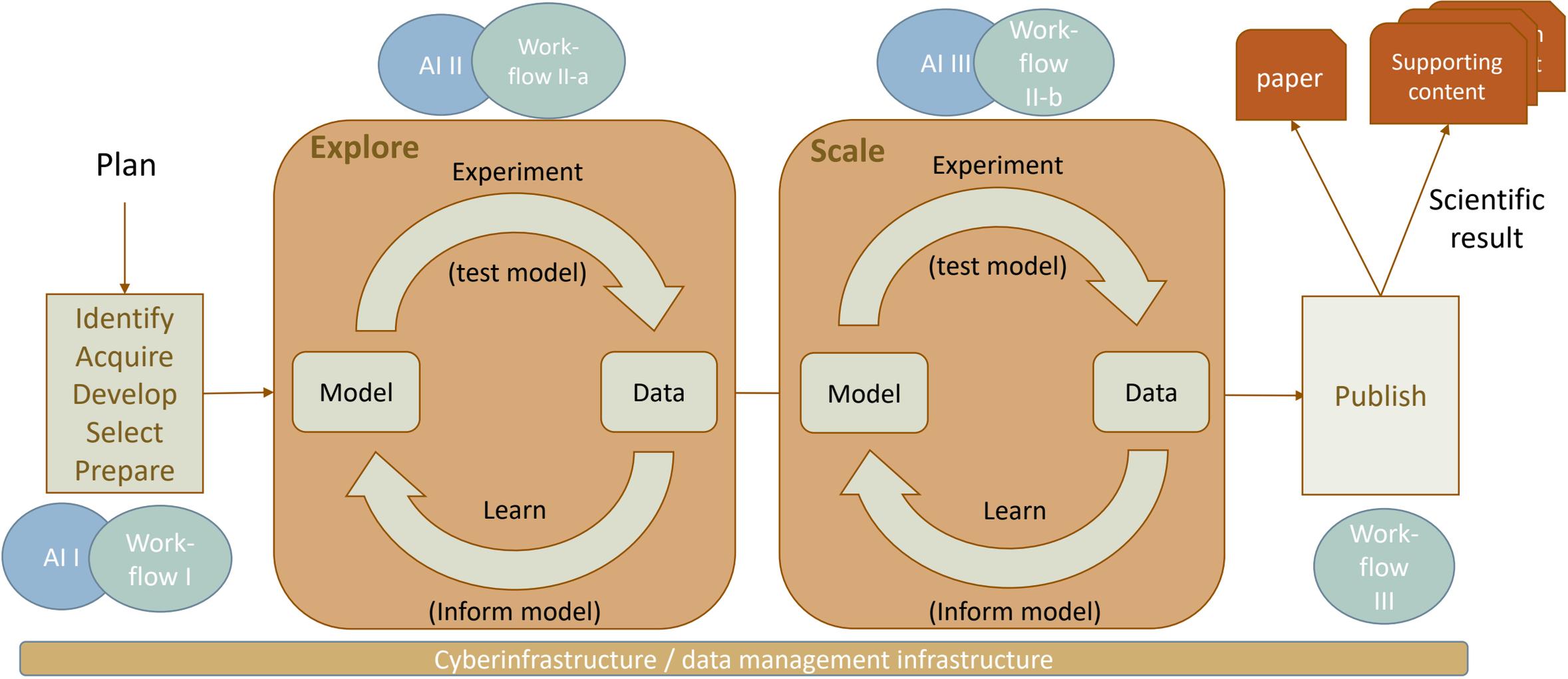
New scientific discoveries not achievable otherwise



Simple Workflow-driven Research Lifecycle



AI and Workflows: Independent and Distributed



i) AI takes different forms in research lifecycle, ii) workflows take different forms (more/less automated), iii) final outcome is “scientific result”

Reducing upstream barriers to research

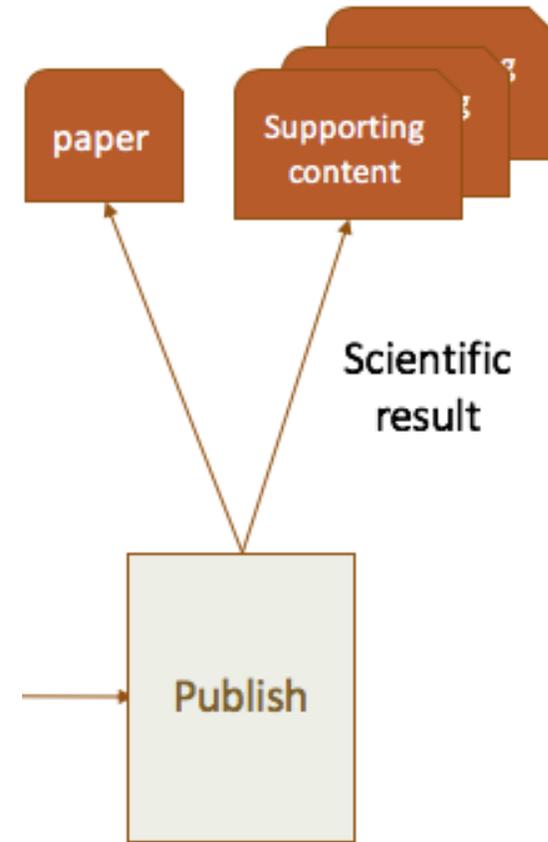


- Data Use Agreements (DUA's) negotiated between data supplier and researcher's institution
 - Templates are being developed to speed up negotiation*
 - DUAs good that specify an open or standard format for data
- Format standardization across instrument types ready for standardization

* Federal Demonstration Pilot (FDP) is example

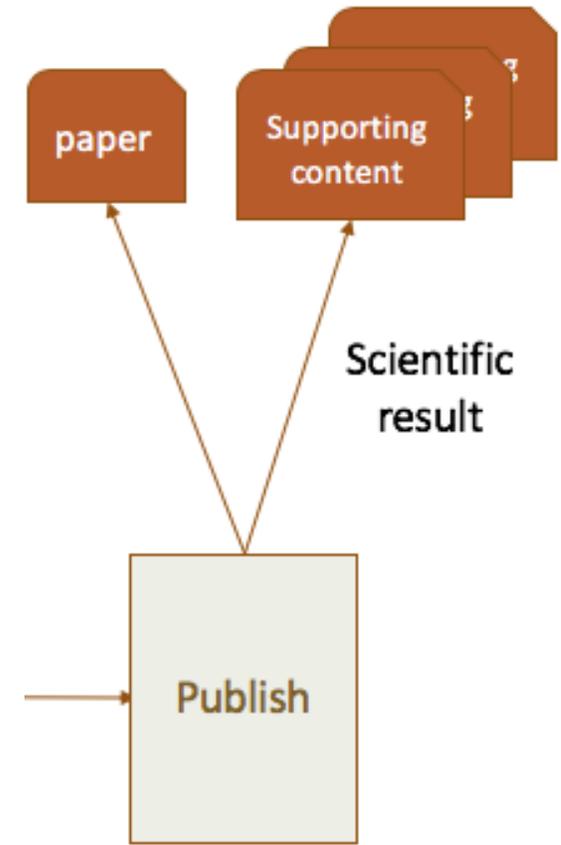
Scientific result is the primary outcome of federally funded research

For a *scientific result* to be reusable, information needs to be transparent



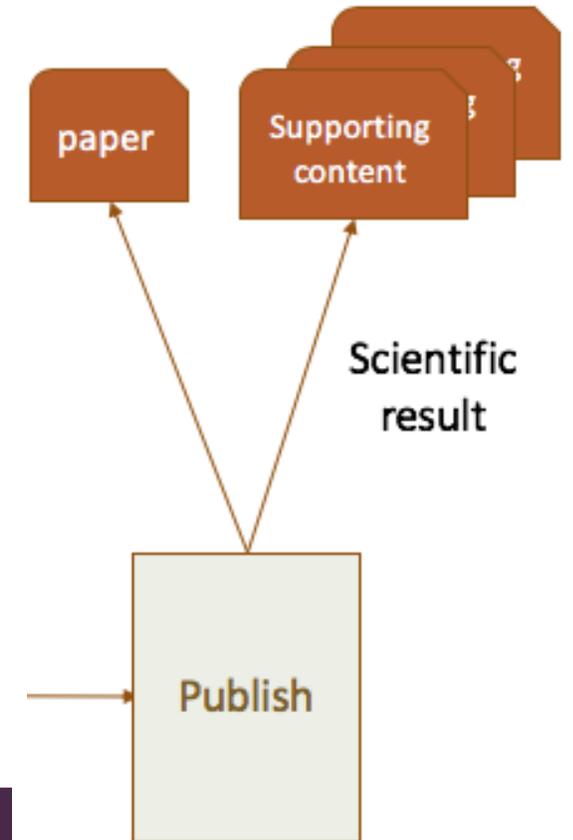
Reuse (from FAIR principles) is important manifestation of Open Science

- What information is needed to enable reuse?
 - i. Narrative (methodology, outcomes)
 - ii. Data, software
 - iii. Workflows, provenance
- From where is information available? Options:
 - Methodology/experiment section of manuscript
 - Appendix of manuscript
 - Supporting content (datasets, software, workflows)
- AI raises new questions
 - Cultural change needed to represent in manuscript



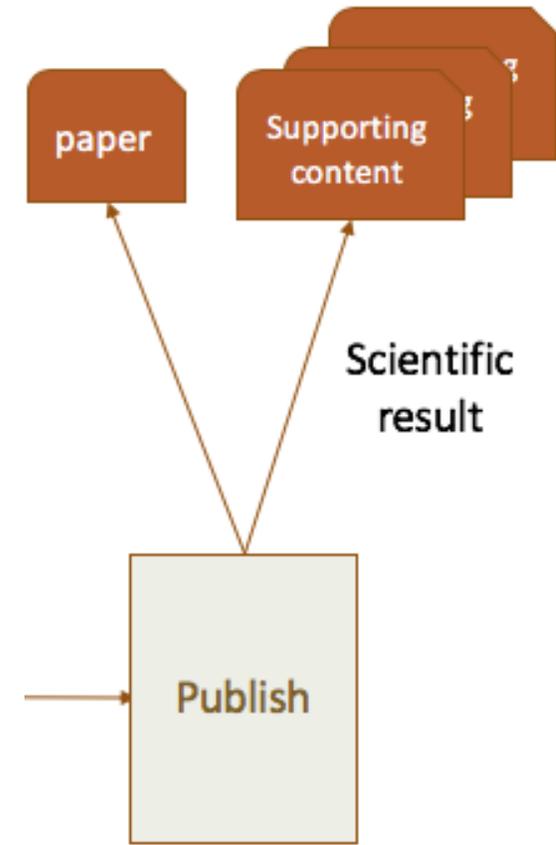
Differing AI needs for transparency due to accountability (e.g., accountable, responsible AI)

Approach used in SC20 to increase accountability: Transparency of decisions of use around AI addressed in methodology / experiment part of paper with additional linked information in appendix if needed



<https://sc20.qldclient.com/submit/transparency-reproducibility-initiative/>

Rights and the scientific result



Funder: NSF Data Management on sharing

Investigators are *expected to share with other researchers*, at no more than incremental cost and within a reasonable time, *the primary data, samples, physical collections and other supporting materials* created or gathered in the course of work under NSF grants. ...

Investigators and grantees are *encouraged to share software and inventions* created under the grant or otherwise make them or their products widely available and usable.

Proposal & Award Policies & Procedures Guide (PAPPG): XI D4



Funder: NSF Data Management on ownership

NSF normally *allows grantees to retain principal legal rights to intellectual property* developed under NSF grants to provide incentives for development and dissemination of inventions, software and publications that can enhance their usefulness, accessibility and upkeep. Such incentives do not, however, reduce the *responsibility that investigators and organizations have as members of the scientific and engineering community, to make results, data and collections available to other researchers.*

Proposal & Award Policies & Procedures Guide (PAPPG): XI D4



Conclusions and Recommendations

- AI use is independent and distributed, and its use can't be captured through workflow provenance
- Not all workflows (and their provenance) are equal in the value of the information that they contribute to assessments of reuse
 - E.g., publishing workflow is largely operational
- Format standardization within instrument categories likely to increase with prevalence of (smartly argued) Data Use Agreements (DUAs)
- Open science balances sharing and ownership; for common good
- AI puts new demands on transparency of scientific result for reuse
- Standardization of scientific result? Some early successes with research objects

Beth Plale
Indiana University

