Accelerating Scientific Discovery: The Need for Platforms

Ian Foster

Workshop on Opportunities for Accelerating Scientific Discovery: Realizing the Potential of Advanced and Automated Workflows
March 16-17, 2020

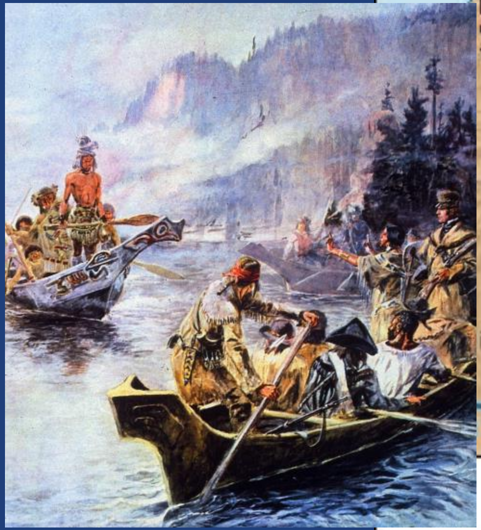
Argonne NATIONAL LABORATORY

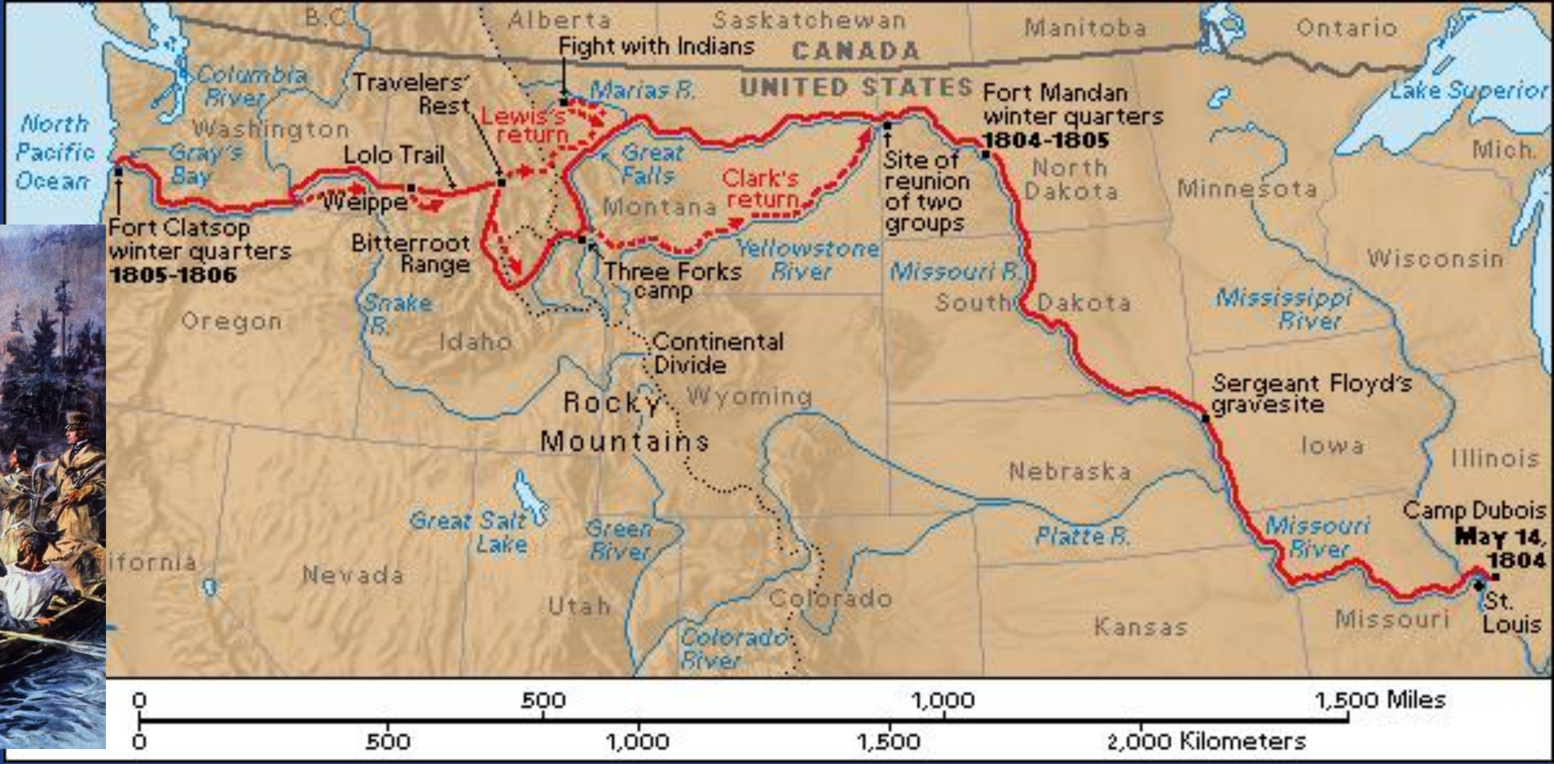
THE UNIVERSITY OF CHICAGO

globus labs

Lewis and Clark Expedition, 1804

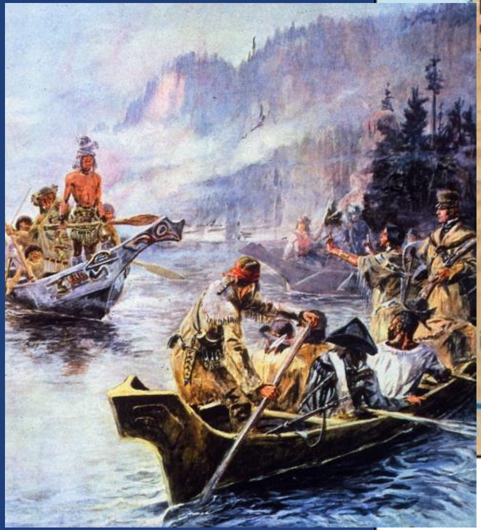
Months of preparation
18 months one way
Expert guides
Much discovery
Mortality, bugs, …

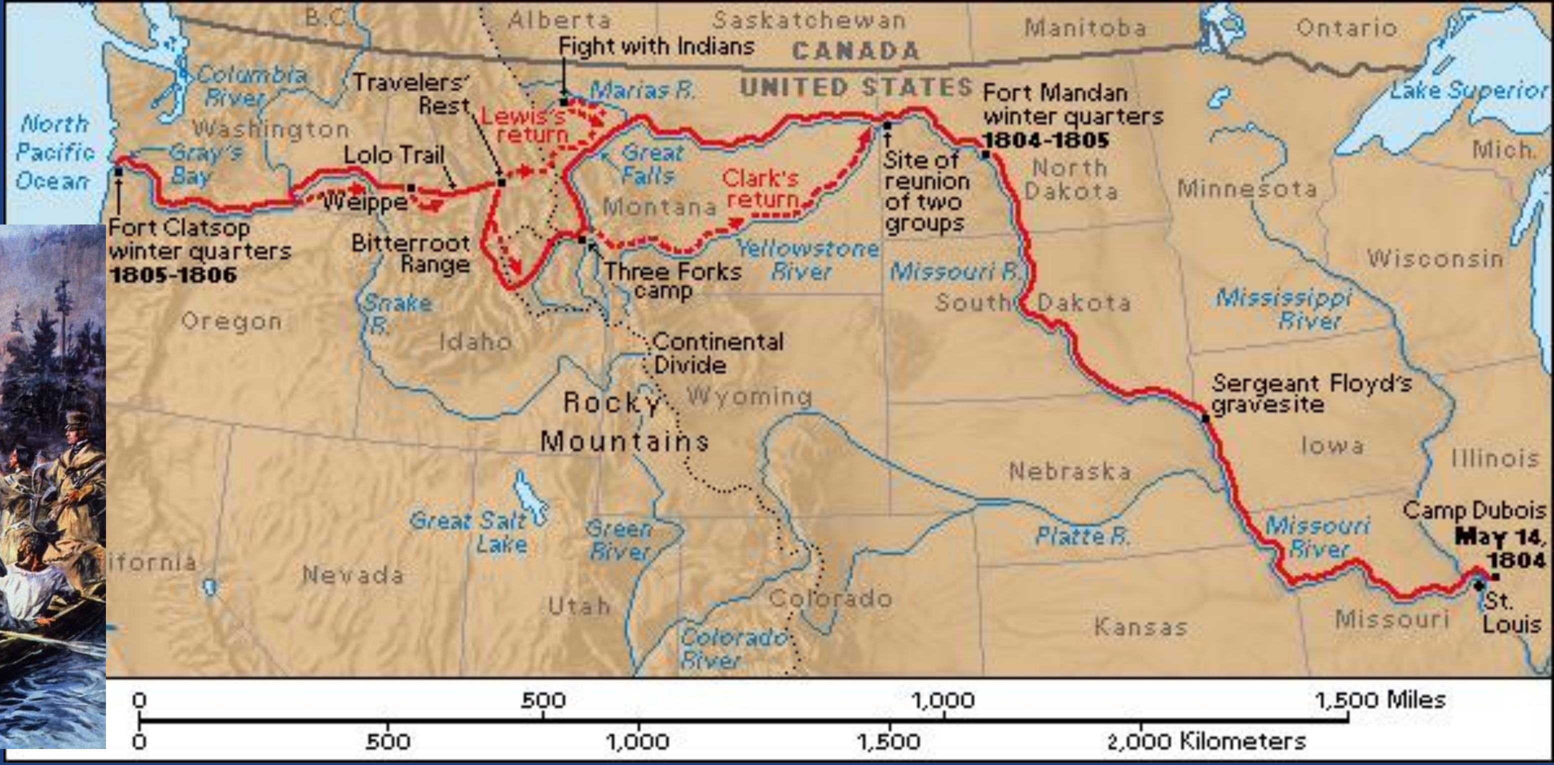




Lewis and Clark Expedition, 1804

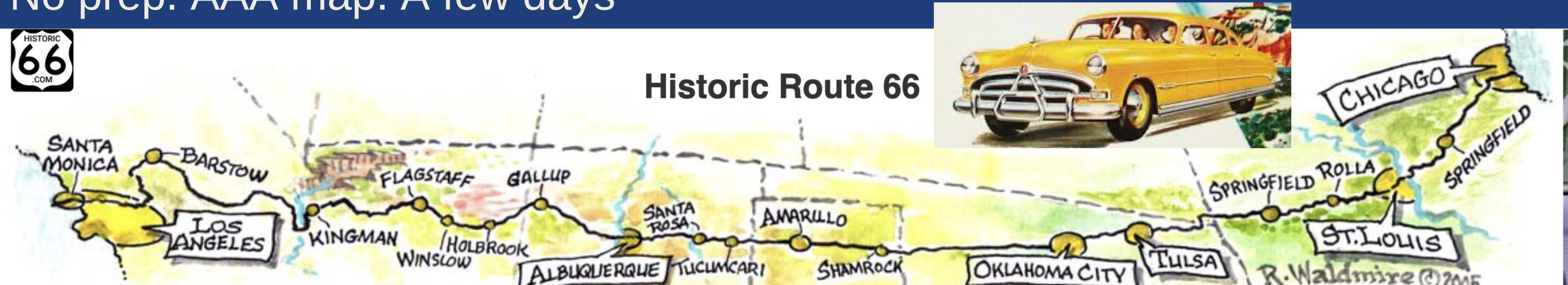
- Months of preparation
- 18 months one way
- Expert guides
- Much discovery
- Mortality, bugs, …





Route 66, 20th Century

No prep. AAA map. A few days

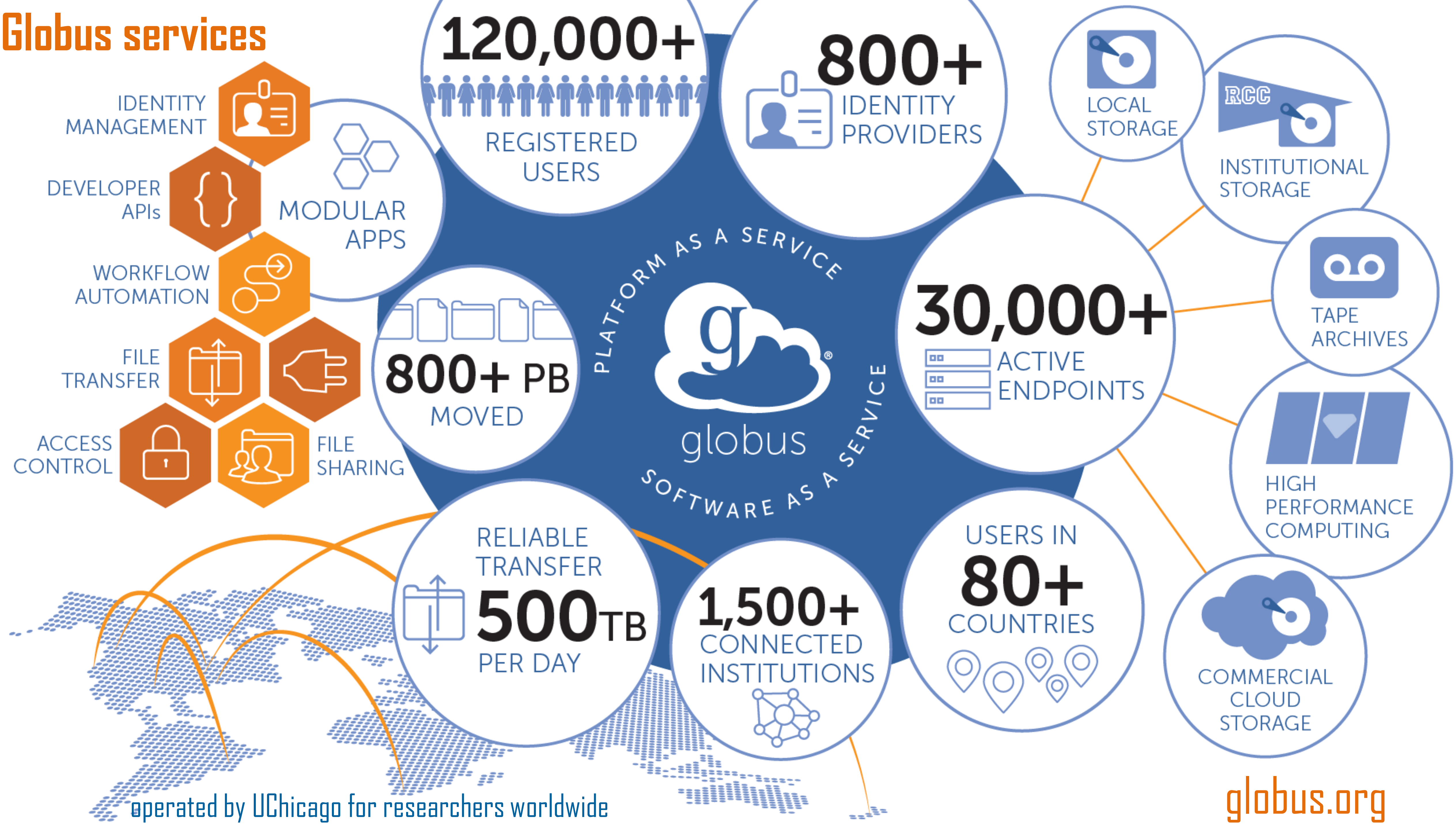


Route 66 crosses IL, MO, KS, OK, TX, NM, AZ and CA

https://www.historic66.com

Higgs discovery “only possible because of the **extraordinary achievements** of … **grid computing**”—Rolf Heuer, CERN DG

Globus services



operated by UChicago for researchers worldwide

globus.org

Takeaway lessons

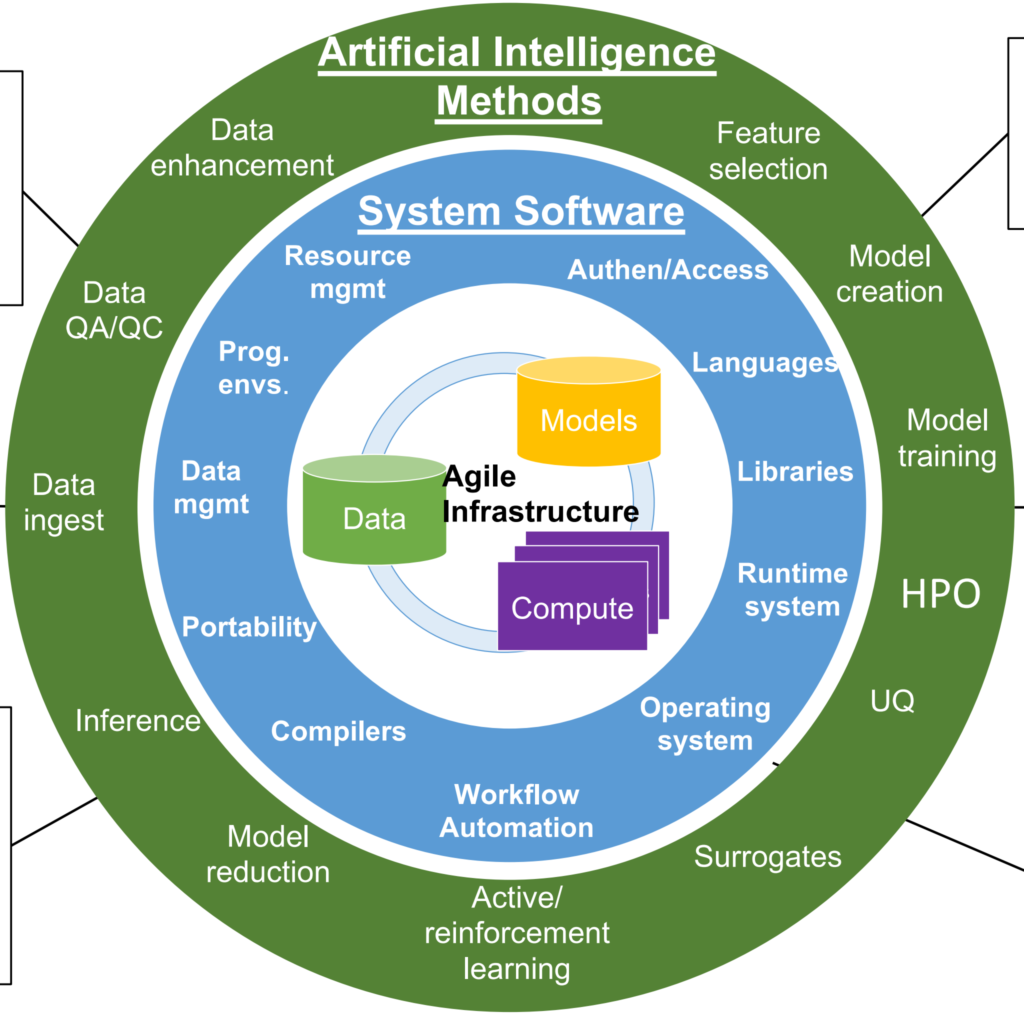
- **The power of platforms**
- **The power (but also the cost) of quality**
- **The power of combining public cloud & science resources**
- **The need for sustainability so that projects can build with confidence**
- **Subscriptions as a sustainability model that works**
- **The need to address the economies of scale**

New instrumentation and AI methods require new platform services

<u>Scientific instruments</u>
Major user facilities
Laboratories
Automated labs
…

<u>Sensors</u>
Environmental
Laboratories
Mobile
…

<u>Databases</u>
Reference data
Experimental data
Computed properties
Scientific literature
…



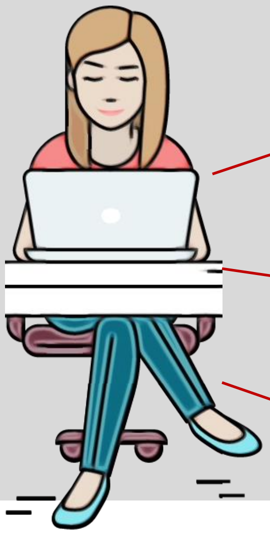
<u>Simulation codes</u>
Computational results
Function memoization
…

<u>Industry, academia</u>
New methods
Open source codes
AI accelerators
…

<u>Scientists, engineers</u>
Expert input
Goal setting
…

Model registry	DLHub
Flows	Automate
Write programs	Parsl
Cost map	SCRIMP
Function fabric	funcX
Data fabric	Data services
Trust fabric	Auth

Next steps: Automated discovery engines. E.g., for materials



Give me a **polymer dielectric** with dielectric constant >10, energy density > 30 J cm^{-3}, breakdown strength > 700 MV m^{-1}, efficiency > 95%

Give me a **solid-state electrolyte** with current density >10 mA cm^{-2}, high stability, low toxicity, moderate cost

Give me a **semiconductor** with bandgap suitable for photovoltaics

Composition, properties, synthesis pathway

Materials Analytics and Design Engine

Organism Designer

Physics Comprehension Sys.

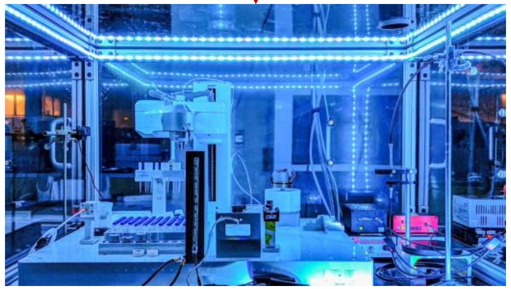
. . .

Self-driving laboratory

AI for scient. program.

Generalized inverter

. . .



- Incorporate state-of-the-art methods for simulating complex materials systems
- Learn continuously as new data are obtained from simulations, experiments, and the literature, transferring knowledge across materials systems

Our goal is to make all research data reliably, rapidly, and securely accessible, discoverable, and usable.

https://labs.globus.org

Relevant references

- Blaiszik, Ben, Kyle Chard, Jim Pruyne, Rachana Ananthakrishnan, Steven Tuecke, and Ian Foster. "The Materials Data Facility: Data services to advance materials science research." *JOM* 68, no. 8 (2016): 2045-2052.
- Ananthakrishnan, Rachana, Ben Blaiszik, Kyle Chard, Ryan Chard, Brendan McCollam, Jim Pruyne, Stephen Rosen, Steven Tuecke, and Ian Foster. "Globus platform services for data publication." In *Proceedings of the Practice and Experience on Advanced Research Computing*, pp. 1-7. 2018.
- Babuji, Yadu, Anna Woodard, Zhuozhao Li, Daniel S. Katz, Ben Clifford, Rohan Kumar, Lukasz Lacinski et al. "Parsl: Pervasive parallel programming in Python." In *28th International Symposium on High-Performance Parallel and Distributed Computing*, pp. 25-36. 2019.
- Chard, Ryan, Zhuozhao Li, Kyle Chard, Logan Ward, Yadu Babuji, Anna Woodard, Steven Tuecke, Ben Blaiszik, Michael Franklin, and Ian Foster. "DLHub: Model and data serving for science." In *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pp. 283-292. IEEE, 2019.
- Madduri, Ravi, Kyle Chard, Mike D’Arcy, Segun C. Jung, Alexis Rodriguez, Dinanath Sulakhe, Eric Deutsch et al. "Reproducible big data science: a case study in continuous FAIRness." *PloS one* 14, no. 4 (2019).