

Networking & Information Technology Research and Development Program

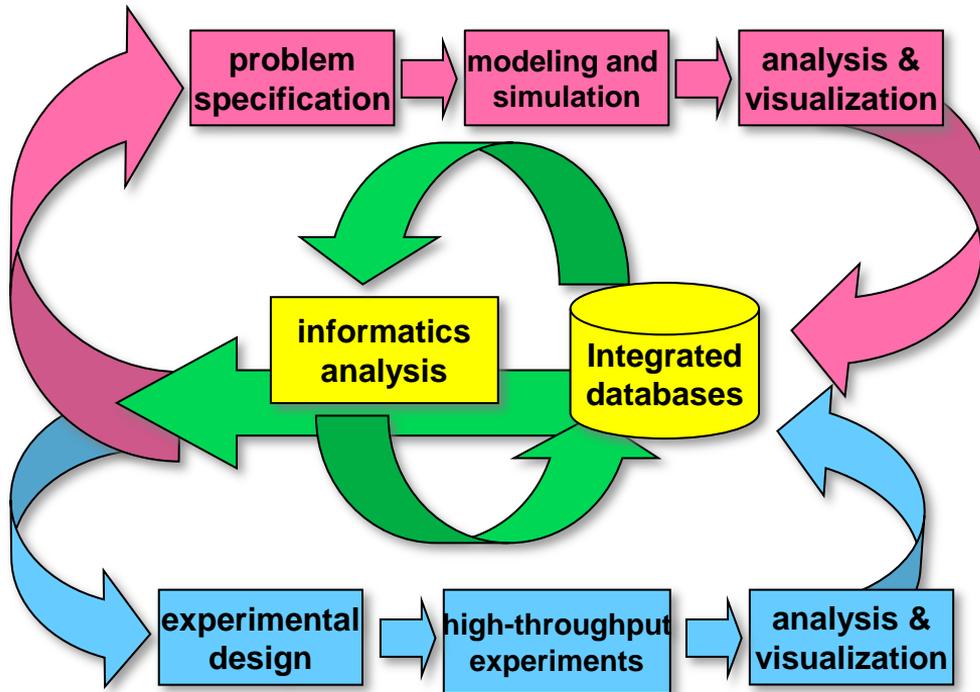
Ian Foster (foster@anl.gov)

Department of Computer Science, University of Chicago
Math and Computer Science, Argonne National Laboratory

Discovery engines for 21st Century Science



Data-driven discovery requires discovery engines



Discovery engine

- Knowledge base & computing engine for a disciplinary research program
- Tight coupling and automation for high-throughput discovery
- Centralized for economies of scale, knowledge sharing, and collaboration

Such systems exist, but are specialized and expensive

Ad pricing

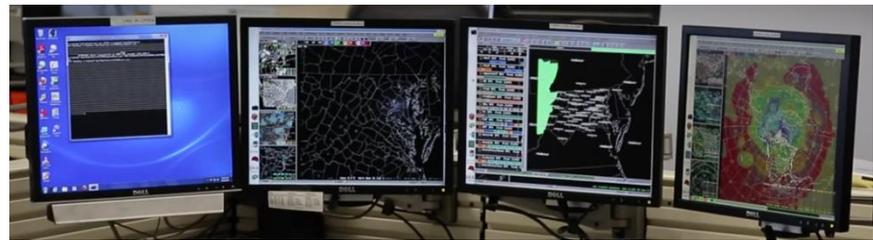
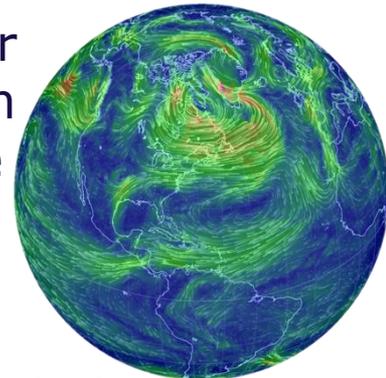
- Ingest stream of web pages, searches, click stream; compute model offline; run experiments (?)



- Using a carefully architected and tuned pipeline on large and specialized computer system

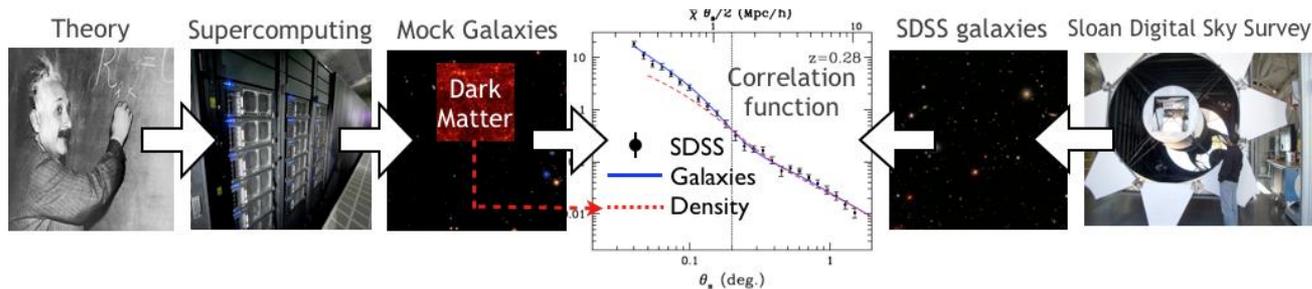
Weather forecasting

- Ingest small number of data streams; run ensembles; propose expts; prepare standard products
- ... using a carefully architected and tuned pipeline on a large, specialized computer system



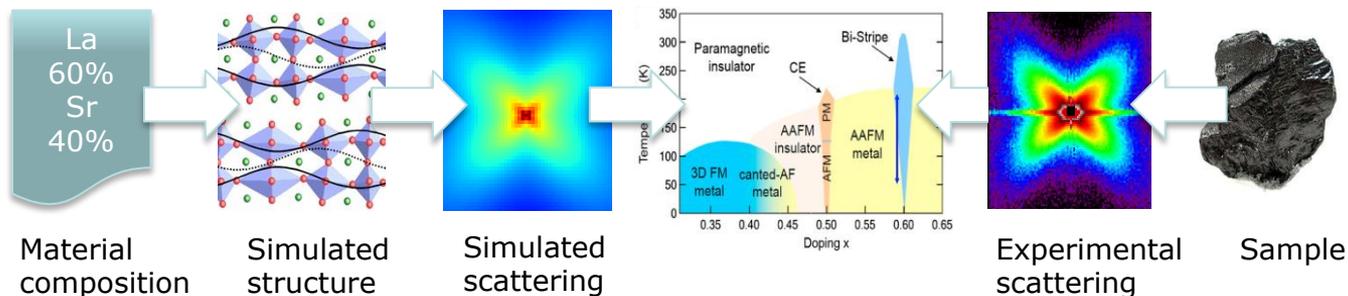
Science needs many discovery engines

Understanding dark matter and dark energy

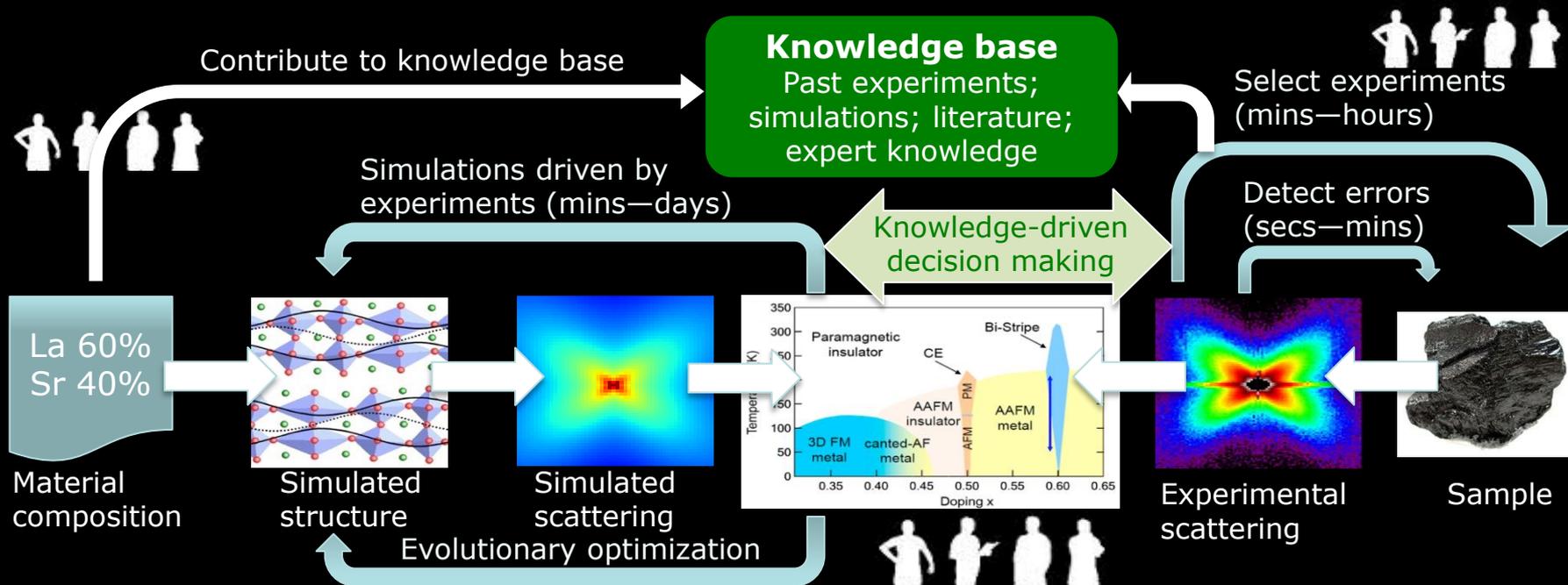


- And more more:
- Health care
 - Climate change
 - Environment
 - Ecosystems
 - ...
 - ...

Understanding & designing disordered structures



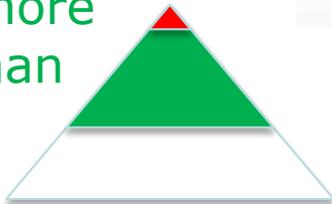
Example: A discovery engine for disordered structures



Diffuse scattering images from Ray Osborn et al., Argonne

Discovery engines and extreme-scale computing

Opportunity:
Reach many more
researchers than
extreme-scale
simulation



Exciting research agenda

- End-to-end automation to slash costs
- Massive knowledge management and fusion
- Rapid inference and knowledge-based response



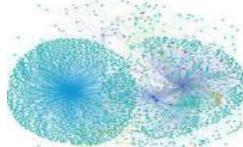
Challenges for exascale technologies



- Reliable, secure, high-speed system integration beyond the machine room



- On-demand scheduling to align with human decision taking timelines



- New computational problems that stress computer architectures in new ways