Cybertools for Materials Research, Education and Collaboration

Creating distributed learning & research environments

Sharon C. Glotzer & Christopher R. Iacovella
Departments of Chemical Engineering and Materials Science and Engineering, University of Michigan, Ann Arbor

Cybertools abound

- Advances in IT have dramatically changed aspects of the way we work.
  - email, ichat, web, ...
  - now podcasts, blogs, wikis, ...
- Has the materials community taken full advantage of what IT has to offer?
- What other types of tools do we need?

How do we currently work?

- Students run simulations and generate data. Lots of data, plus analysis.
- They send us data & analysis as graphs, images, etc. via email, ichat, link to public folder on desktop, or bring hardcopy to meetings.
- We discuss, make suggestions, student goes off to generate more data, analysis, etc.
- Repeat.

How do we currently work?

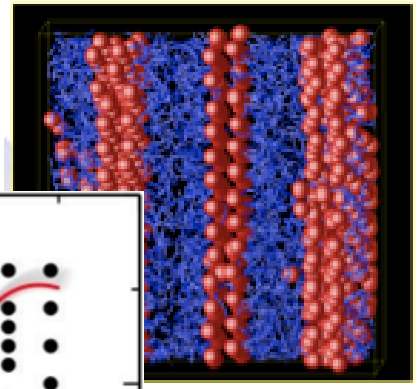
- If collaborators involved, they receive info via email or file uploaded to website, or real-time video conferencing via ichat
- Eventually, paper is written, selected data & analysis is presented and published, raw data & analysis is stored on student's computer, and we move on to new project.

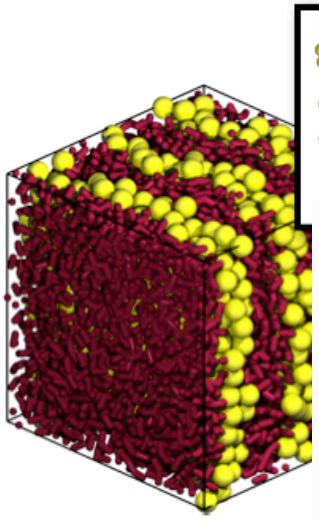
- What's wrong with this workflow?

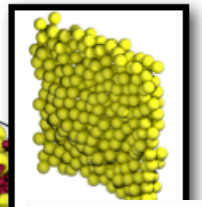
How do we currently work?

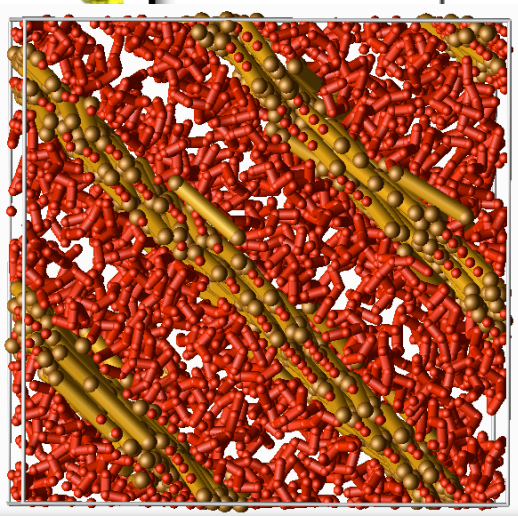
- Limited sharing of data and data products
  - Provided only when prompted
  - Peer-to-peer sharing, learning difficult
  - No meaningful relationships between files containing related data and data products
  - Files difficult to find by anyone other than student who generated them
  - Data lost over time
  - Lots of useful information "locked away"

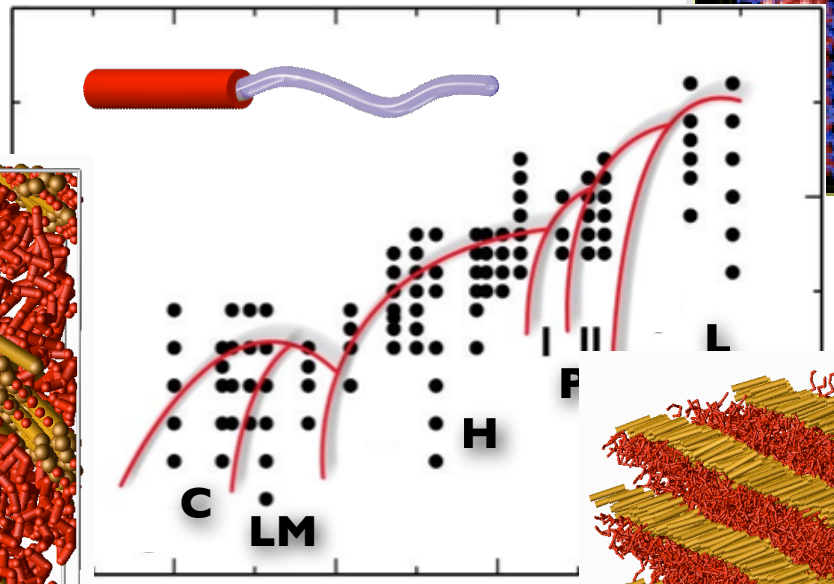
Example: Lamellar phases

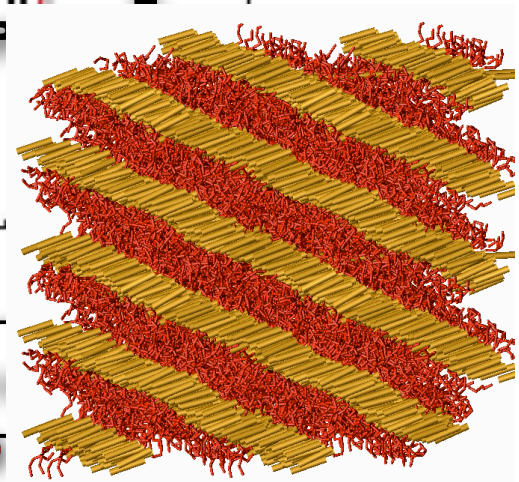


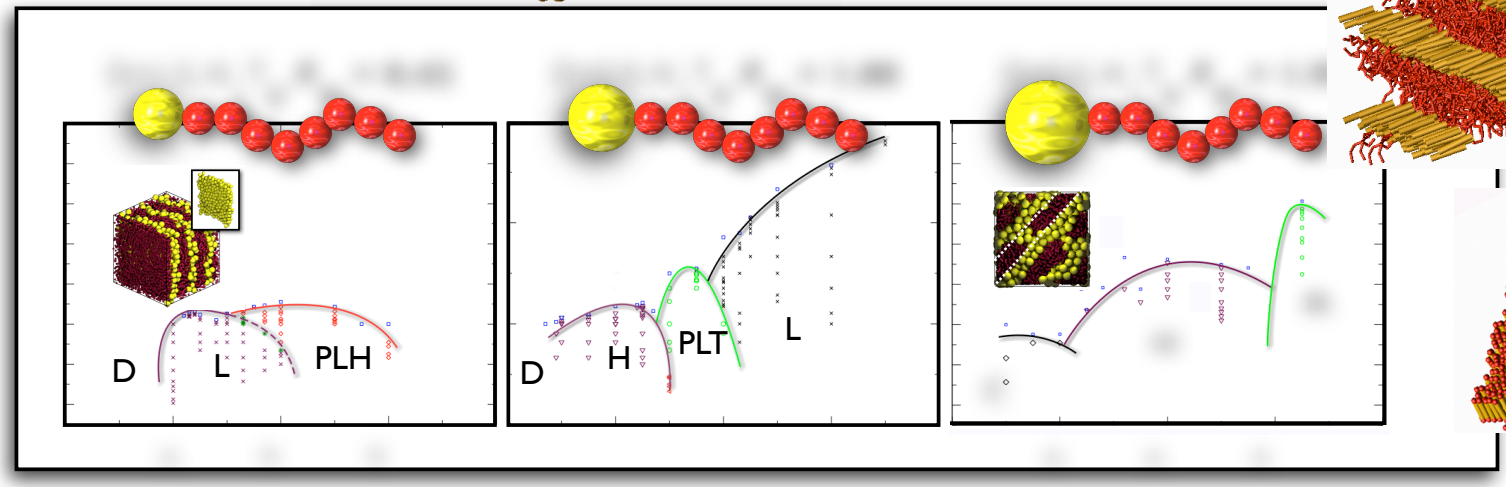


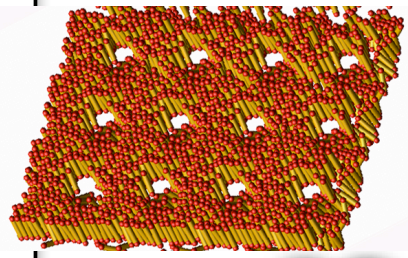












Examples of research output from materials simulation

- Here: materials structures, thermo. phase diagrams
- Not shown: assembly kinetics, materials properties, background info, definitions, simulation method, theoretical equations, related publications and web material, experimental results - *all inter-related*

Want a cyberinfrastructure that brings all that information together in a sensible, dynamic, substantive way.

...a living, community-developed materials cyberenvironment

(> database+portal+shared notebook+mining tools+open source....)

The Vision: How do we want to work?

- Want easy, searchable access to full research product, anytime, from anywhere.
  - Access should reflect **inter-relationships** between data, analysis, related work, commentary, etc. with links to appropriate related information
  - Should be able to query based on any property, behavior
  - Transformative for both research and teaching
- Want collaborations with seamless and protected sharing of data and other files.

What else do we want?

- Perpetual analytics: the process of performing real-time analytics on data streams (Jeff Jonas, IBM Entity Analytics)
  - Provides real-time enterprise awareness: streaming recognition of related data and reconciliation.
  - e.g. if a user changes a data entry, the system "knows" and relevance updated accordingly
  - e.g. link new microstructures with similar descriptors before the question is asked

Vital for real-time situational awareness, useful for research

If we're really greedy...

- Sequence neutrality: regardless of the order in which data or queries occur, the end-state, once all data points are known, is the same. (Jeff Jonas, IBM Entity Analytics)
  - e.g. if data added to system two months after initial query that changes outcome of query, you are instantly notified

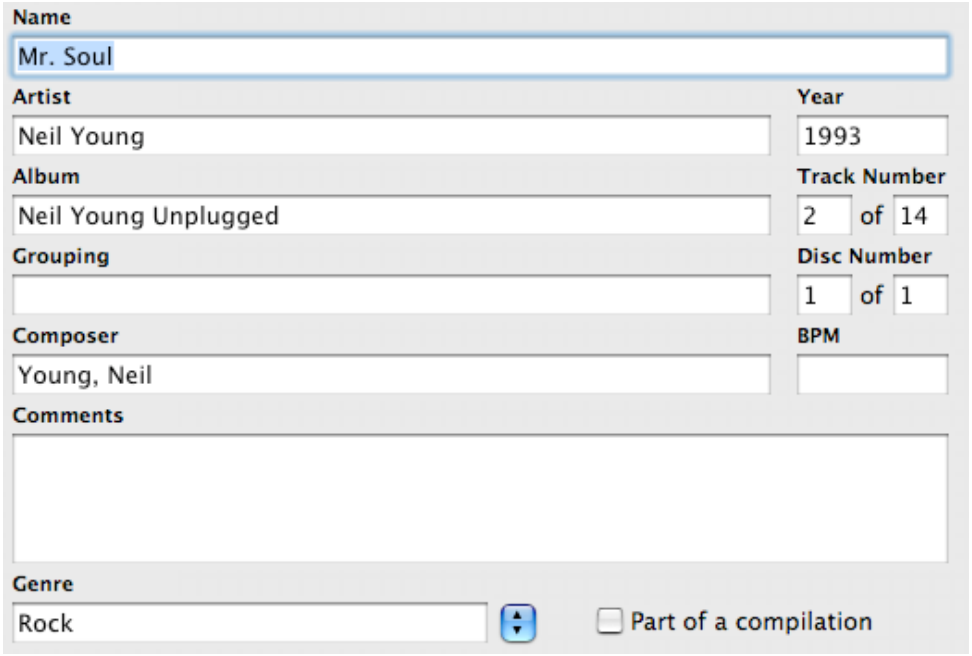
Changing the way we work

- We don't just want faster existing tools
- We want *transformative tools, that fundamentally change the way we work* on our own, with students and with other researchers
  - *to see widespread use, must fit easily into workflow*
- Much of the underlying cyberinfrastructure we need will come from other industries/ communities, even other science communities; we need to adapt it for our purposes
  - search engines, entertainment, intelligence, etc.

Example from entertainment

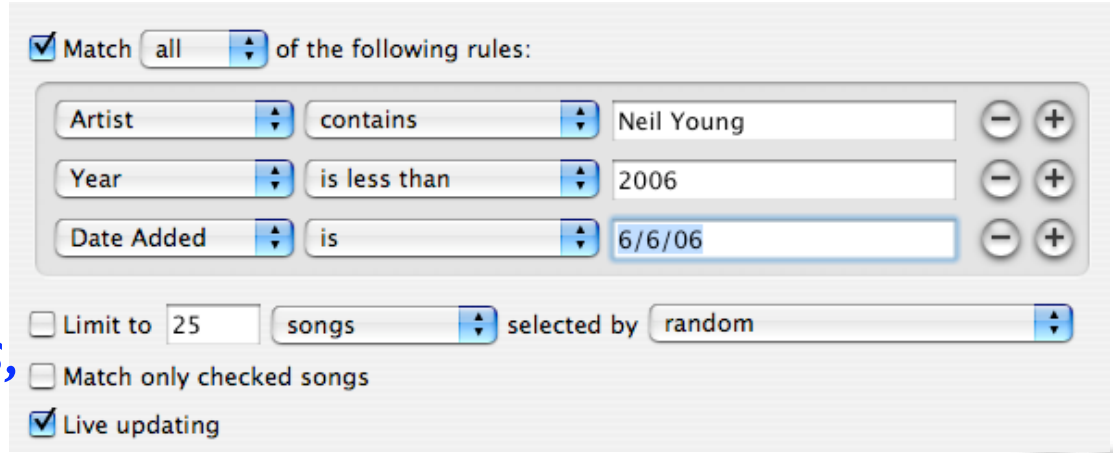
Consider iTunes:

- A data-driven program...each song tagged with a variety of "meta-data" that enables inter-relationships, etc.
  - Song Name, Album Art, Artist, Album, Year, Bitrate, Comments, Genre, Grouping, etc.



iTunes

- • Each song lives in the library, but can be associated with different playlists...multiple playlists can point to the same song
  - Each field is easily searchable
  - Search by artist, album, year, song title, etc...
- • Can create "smart-playlists" driven by simple rules, which update automatically as new fields matching rules are added to library



iTunes for Materials:
e.g. song = state point, etc.
smart playlists for structures, densities, etc.

First steps: MatDL Pathway

Collaboration w/L. Bartolo (PI, Kent), J. Warren (NIST), D. Sadoway (MIT), A. Powell (MIT), K. Rajan (ISU) under NSF DUE NSDL grant

- Build materials-centric web portal based upon current industry-standard digital library protocols like Fedora and Fez.
- Develop metadata protocols for files to allow searching, inter-relationships, etc. *Through metadata, cyber becomes part of the research.*
- Establish MatForge, an open-source initiative for materials simulation code community development & download
- Establish wiki sites for materials subdomains with links to data, relevant literature, analysis and simulation tools, etc.

First steps: MatDL Pathway

Collaboration w/L. Bartolo (PI, Kent), J. Warren (NIST), D. Sadoway (MIT), A. Powell (MIT), K. Rajan (ISU) under NSF DUE NSDL grant

- Enable bridging of research, classroom teaching, and independent UG/grad learning in a simple way by integrating several CI platforms.
  - Students share data, codes
  - Hot-off-the-press research data, codes used in virtual labs for undergraduate and graduate courses (U-M, MIT)
- New cybertools for collaborating
  - within research group, between research groups at one or multiple universities (NIRT, MRSECs)
  - peer-to-peer collaborations in particular