HTR Informatics

DOW

Issues and Challenges in the Management of Data and Knowledge from High Throughput Research in Catalysis and Material Science

Bruce E. Wilson, Technical Leader
The Dow Chemical Company
Midland, MI USA

HTR at Dow Chemical

- High Throughput Research is a ***HOW*** not a ***WHAT***
- High Throughput Research can be applied to any of Dow's product lines and processes
  - Homogeneous catalysis processes
  - Heterogeneous catalysis processes
  - Polymers (thermoforms & thermosets)
  - Formulated products
  - Agrochemicals
- Focus is often on the process, not the product, except that the process often defines the product
- High project turnover

Discovery vs. Process Research

Discovery Research

- Screening
- Yes or No tests
- Look for an event
- # samples large
- Sample size small
- # condition sets small

Process Research

- Optimization (multi-valued)
- Mixtures
- Measurement precision
- # samples small
- Data complex
- Sample size larger
- # condition sets larger
- Data integration
  - With other reactors
  - Across scales

Example Project: Catalyst Modifier

- **Problem:** Polymer product; small amounts of a side product coat reactor over time
- **Task**: Modify the catalyst package and process conditions to eliminate the side product
- Resulting data
  - Structures & amounts of catalyst modifiers (~300 candidates)
  - Reactor conditions (Temperatures, pressures, flows versus time)
  - Analytical results (GPC, DSC, FTIR)
  - 3 months work; ~50 boxes of paper (if printed)
- Solution was a *combination* of process changes and a catalyst modifier

Informatics

- Conducting experiments is no longer the bottleneck
- It can often take longer to collate the pertinent synthesis, reaction and screening data that it does to run the experiments
- Multiple screens required for multivariable optimizations
- Still have people who prefer a week in the laboratory to an hour in the library

Challenges for Informatics

- Complex experiments
- Complex data
- Need data about how the experiments were executed (i.e. process variables)
- Dynamic workflow
- High project turnover (~100/year)
- Multi-variable optimization problems

Big Challenge #1

Can We Afford the High-Tech Solution?

Big Challenge #2

Can’t just cut open the shrinkwrap

Addressing the Challenges

- Leverage what we can from pharmaceutical industry experience
- Define a flexible architecture
- Use an agile methodology
- Communicate, communicate, communicate

Informatics Architecture Principles

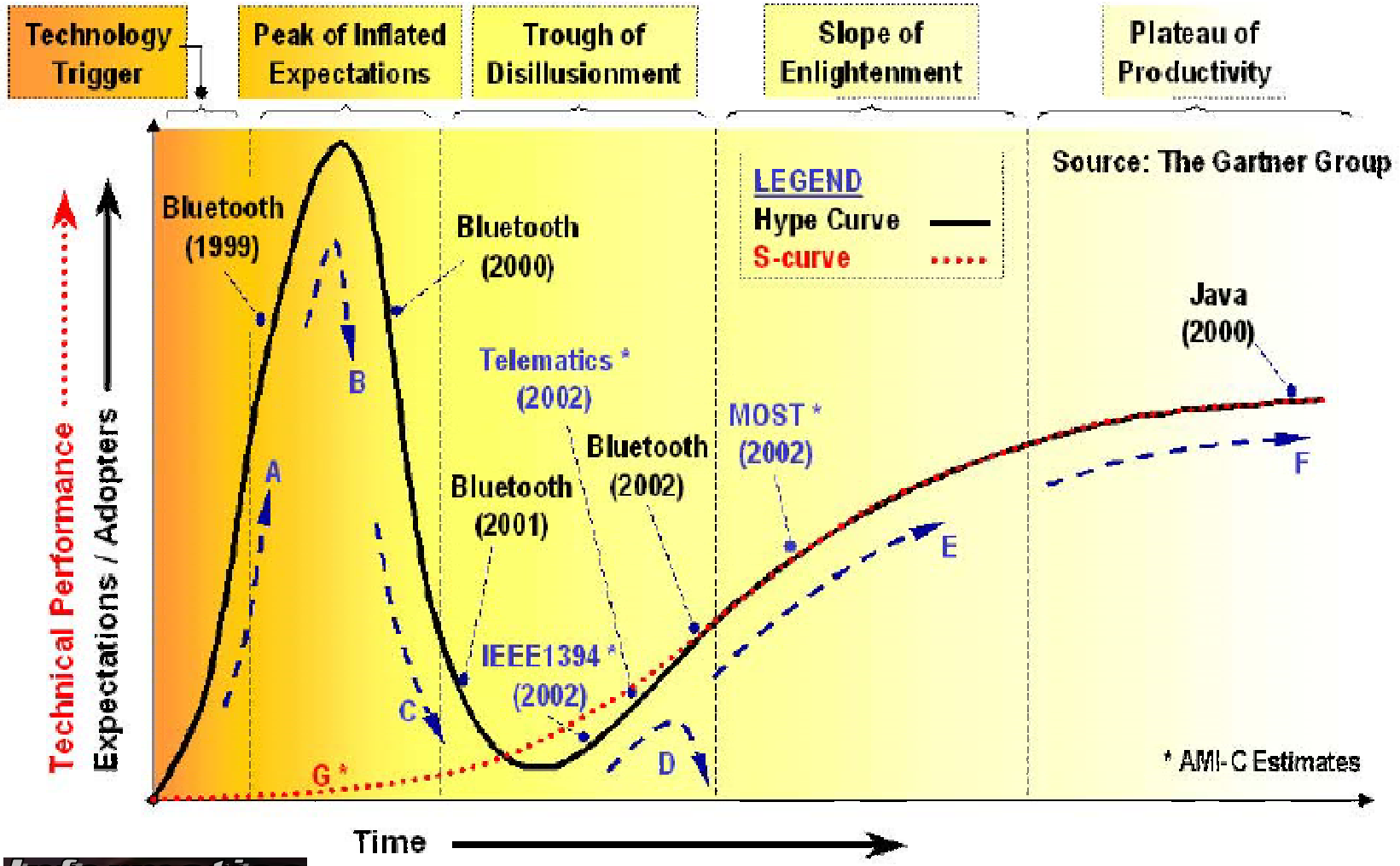
- Use purchased software where practical
  - Understand want versus need
  - When can the workflow adapt to the solution
  - When must solution adapt to the workflow
- Design for the long term (possible 75 year retention)
  - Use published data formats
  - Keep data migration paths in mind at all times
  - Underlying applications will change
  - Maintenance dominates total cost of ownership
- Applications which must be “in charge” are bad
  - Must be able to talk with other applications
  - Use Web Services as application glue

Agile Development

- Lots of Agile Methods: XP, Orange, Crystal, Scrum, …
- Studied available methodologies, chose something of a hybrid
- Short inspect and adapt cycles
- Focus on delivering something usable by the customer
- Understand the cost of changes
  - Build now, refactor later
  - Getting something useful now helps drive out requirements

Communications

Stages of the Hype Cycle



Summary

- Wetware is at least as important as hardware and software
- Architecture and implementation methodology are at least as important as the tools
- “Center of the universe” applications and tools are bad