

A man in a dark sweater stands with his arms raised, looking upwards. The background is a blurred molecular model with various colored spheres (red, purple, green, blue) connected by thin lines, set against a light blue background.

Chemical information integration: A changing perspective

Mitch Miller, American Chemical Society, 10 September 2003



Contents

- Integration: what are we talking about?
- An (anecdotal) historical perspective
- The current need
- What follows

Chemical Data Integration

- What does it really mean?

Chemical Data Integration in 1989

- Quantities were smaller
- Types of data were simpler
- Information systems were proprietary
- Expectations were lower
 - Incorporate very different types of data in a single document
 - Using disparate systems in tandem was a kind of integration

1989 Example: Incorporating structures in a document

- Remember, cut and paste have been around for a while...



1989 Example: Disparate systems in tandem

- 3D database -> Molecular Dynamics
- Steps
 - Performed database search using one proprietary system
 - Exported molfile for promising molecule
 - Run molecular dynamics in second proprietary system
 - Export trajectory file
 - Ran secondary program to convert to SD format
 - Imported into database
 - Ran 3D search
- (Oh!)

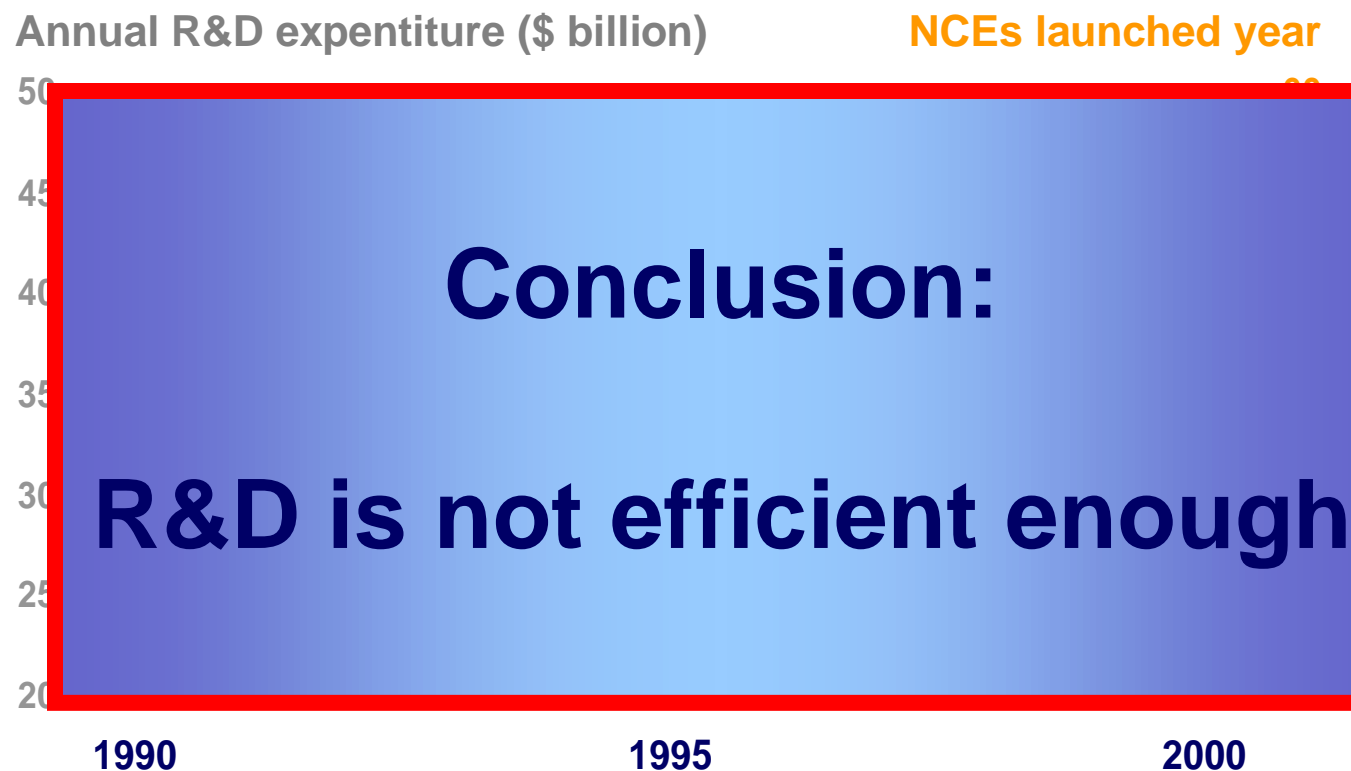
Chemical Data Integration in 2003

- Large increase in quantity of data
- diversity of data types
 - high-throughput data is different from lower-throughput data
 - image data
 - spectroscopic data
 - genomic data
- New enabling technologies
 - .Net, SOAP, XML, J2EE
- Buzzwords
 - ROI, CRM, ERP enterprise resource planning, SCM supply chain management, RTI Real-Time Integration, analytics, e-business, data warehouses, 360-degree view, end-to-end

But....

- Buzzword compliance is not enough

The R&D Investment Paradox



- R&D spend more than doubled in 10 years
- Number of compounds entering development is static
- Number of NCE launched is steadily declining

Scientific Data Map and Domains

Generating data is one
thing
sharing data
meaningfully is another

Prediction

Results

Compound
Libraries

Some 'numbers'

- Google search for 'chemical data integration'
 - on 03 September 2003
 - First page only
- Results
 - 6 commercial products + one interview with interview with CTO of software company
 - 3 conferences on integration
 - Exclusive of 9 additional sponsored links
- Of 12 speakers here today, 6 have something to sell you
- **Conclusion: data integration is a highly commercial proposition**

What follows...

- Variety of perspectives on our topic
- Speakers were asked to provide a brief definition of 'integration'
- Half from software vendors; half from implementers

Finally, Some Meeting Business

- One cancellation – two talks
- At 1:30, a panel discussion
- Implementers: what's missing in commercial data integration software?