Integrating High Throughput Workflows: Value and Build vs. Buy Analyses (Plus A Bit More)

ACS Fall Meeting
CINF Emerging Trends in Discovery Data Integration

Peter E. Cohan
The Second Derivative
September 10, 2003
Agenda

- Some Definitions of “Integration”
- Workflow Value Analysis
- Workflow Software Build vs. Buy Analysis
- Impact of Demonstrations and Ideas to Improve Results
- Summary
Some Definitions of “Integration”
Definitions of “Integration”

1. **Customer’s Perspective:**
   
   Data or information that we **need** to bring together to solve a problem we face today.

2. **Vendor’s Perspective:**
   
   Data or information that we **can** bring together to solve a problem we hope our customers face today.

3. **Visionary’s Perspective:**
   
   Data or information that it would be **really cool** to bring together to enable some really cool things…
Definitions of “Integration”

Workflow Perspective:

- “Integration” is the enabling of a workflow and its results from the perspective of a workflow user or beneficiary.
- A “workflow” can include 1 or more linked, nested, or concurrent other workflows
  - (e.g. the drug discovery and development pipeline)
- A workflow can include hardware, software, and know-how
  - The definition of a workflow is in the eye of the beholder…
High Throughput Workflows – Value Analysis
Pre-Summary

- If you know what you want to accomplish, and the problem is tractable, then high-throughput methods pay-off big time…
Where Can We Apply High Throughput Methods?

- **Pharma Discovery**
  - Combinatorial and Parallel Synthesis
  - Synthetic Methods Scoping and Development
  - High Throughput Screening
  - Toxicology, ADME

- **Pharma Chemical Development**
  - Process Development
  - Polymorphs and Salt Selection
  - Formulation

- **Materials Science**
  - Homogeneous Catalysis (e.g. Polyolefins)
  - Heterogeneous Catalysis
  - Pigments
  - Polymers, Coatings and Formulations
  - Electronic Materials (e.g. Phosphors, Fuel-Cells)
Key Business Objectives

There needs to be an economic driving force:

- Dramatically increase the rate of innovation in the discovery and optimization of…
- Implement high throughput core competency…
- Achieve cultural change from traditional to high throughput methods…

For Example:

- “We need a critical commercializable discovery in a key business area”
- “We just had a critical commercializable discovery in a key business area
  □ And we need to exploit and protect it rapidly”
Workflow Value Analysis

- Understand the value chain
- Identify the critical, measurable values
- Estimate the value of high throughput research and/or optimization

Critical Factors → Increasing Rate of Innovation → New products
More Products
More Profit
Critical Factors

- Improve the probability of success of projects
- Perform more projects with the same resources  
  (Measurable)
- Reduce the time to market
- Make critical business/market decisions earlier  
  (Valuable, but Intangible)
- Increase visibility into projects’ success or failure ("fail early")
- Present the best candidate to the next phase
- Improve IP protection and management of competitive IP
- Capitalize rapidly on opportunities
- Uncover unexpected results and discoveries
Measurable Value: 
Increase Probability of Success

Long-term Research Projects

Optimization Projects

Data Sources:
• Dow, ExxonMobil, W.R. Grace, Symyx Technologies
# Types of R&D Projects

<table>
<thead>
<tr>
<th></th>
<th>“Research”</th>
<th>“Optimization”</th>
<th>“Firefighting”</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of R&amp;D Work</td>
<td>17%</td>
<td>57%</td>
<td>26%</td>
</tr>
<tr>
<td>FTE Rate / Pair (1 Pair = 1 Scientist + 1 Technician)</td>
<td>$560K / year</td>
<td>$560K / year</td>
<td>$560K / year</td>
</tr>
<tr>
<td>Experiments per Year (per Pair)</td>
<td>435</td>
<td>610</td>
<td>610</td>
</tr>
</tbody>
</table>

Data Sources:
- Dow, ExxonMobil, W.R. Grace, Symyx Technologies
Throughput Assumptions

- **Workflow includes:**
  - Specification
  - Synthesis
  - Sample Preparation
  - Screening
  - Data Management and Analysis

- **All steps are equally matched in throughput with one another**
  - [No bottlenecks]

- **Each “Cell” can be run once per day (200 working days per year) - sustainable**
  - 24 cells = 4,800 experiments per year
  - 48 cells = 9,600 experiments per year
  - 96 cells = 19,200 experiments per year

- **Metrics:**

<table>
<thead>
<tr>
<th></th>
<th>“Research”</th>
<th>“Optimization”</th>
<th>“Firefighting”</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of R&amp;D Work</td>
<td>17%</td>
<td>57%</td>
<td>26%</td>
</tr>
<tr>
<td>FTE Rate / Pair</td>
<td>$560K / year</td>
<td>$560K / year</td>
<td>$560K / year</td>
</tr>
<tr>
<td>(1 Pair = 1 Scientist + 1 Technician)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiments per Year (per Pair)</td>
<td>435</td>
<td>610</td>
<td>610</td>
</tr>
</tbody>
</table>
Value Calculation Examples

- **Case I: Optimization**
  - Ia: 24 cell system used by a team of 1 pair (2 people)
  - Ib: 48 cell system used by a team of 1 pair (2 people)
  - Ic: 96 cell system used by a team of 2 pairs (4 people)
  - Id: 48 cell system used by a team of 1 pair (2 people) with increased POS

- **Case II: Research**
  - IIa: 24 cell system used by a team of 2 pairs (4 people)
  - IIb: 48 cell system used by a team of 2 pairs (4 people)
  - IIc: 96 cell system used by a team of 2 pairs (4 people)
  - IIId: 48 cell system used by a team of 2 pairs (4 people) with increased POS

- **Case III: Combined Research and Optimization**
  - IIIa: 96 cell system used by a team of 2 pairs (4 people) – 1 research success
  - IIIb: 96 cell system used by a team of 2 pairs (4 people) – 2 research successes
## Summary – Comparison of Traditional vs. High Throughput

<table>
<thead>
<tr>
<th>Case</th>
<th>Traditional Time to First Success</th>
<th>HT Time to First Success</th>
<th>Annual Value of HT System</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimization 24</td>
<td>2 years</td>
<td>3 months</td>
<td>$3.9M</td>
</tr>
<tr>
<td>Optimization 48</td>
<td>2 years</td>
<td>1.5 months</td>
<td>$8.4M</td>
</tr>
<tr>
<td>Opt. 96 (2 pairs)</td>
<td>2 years</td>
<td>0.8 months</td>
<td>$15.7M</td>
</tr>
<tr>
<td>Opt. 48 ↑ POS</td>
<td>2 years</td>
<td>4 months</td>
<td>$2.1M</td>
</tr>
<tr>
<td>Research 24</td>
<td>50 years</td>
<td>9 years</td>
<td>$5.1M</td>
</tr>
<tr>
<td>Research 48</td>
<td>50 years</td>
<td>4.5 years</td>
<td>$11.3M</td>
</tr>
<tr>
<td>Research 96</td>
<td>50 years</td>
<td>2.25 years</td>
<td>$23.8M</td>
</tr>
<tr>
<td>Res. 96 ↑ POS</td>
<td>50 years</td>
<td>2.82 years</td>
<td>$18.7M</td>
</tr>
<tr>
<td>Combined 1</td>
<td>50/2 years</td>
<td>2.25yr/0.8 month</td>
<td>$19.3M</td>
</tr>
<tr>
<td>Combined 2</td>
<td>50/2 years</td>
<td>2.25yr/0.8 month</td>
<td>$19.3M</td>
</tr>
</tbody>
</table>
Summary – Sensitivity Analysis
Summary – Cost per Experiment and Pro-Forma Analysis

Cumulative Earnings
"Generic" Plastics Development Program

<table>
<thead>
<tr>
<th></th>
<th>Cost Per Experiment Over 3 Years</th>
<th>Cost Per Experiment Over 5 Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry Today</td>
<td>$1075</td>
<td>$1075</td>
</tr>
<tr>
<td>Commercial System at 12,000 Experiments / Year</td>
<td>194</td>
<td>97</td>
</tr>
<tr>
<td>Commercial System at 24,000 Experiments / Year</td>
<td>135</td>
<td>68</td>
</tr>
</tbody>
</table>
Workflow Software
Build vs. Buy Analysis
Pre-Summary: “Core” or “Context”?

- **Core**: Stuff that makes your business what it is
- **Context**: Stuff that has to be done but does not really differentiate

**Recommendation:**

Invest in “Core”, Outsource “Context”  
- Geoffrey Moore
Implementation Effort vs. Productivity

Zone of Highest Risk
Workflow and Implementation Requirements

- Design
- Synthesis / Modification
- Sample Preparation
- Properties Screening
- Data Analysis

DOE - Simulation - Modeling - Statistical Analysis - Analytics - Data Mining - Visualization

Database Infrastructure
## Workflow and Implementation Requirements

<table>
<thead>
<tr>
<th></th>
<th>DOE</th>
<th>Simulation</th>
<th>Modeling</th>
<th>Statistical Analysis</th>
<th>Analytics</th>
<th>Data Mining</th>
<th>Visualization</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Design</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Synthesis / Modification</strong></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sample Preparation</strong></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Properties Screening</strong></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data Analysis</strong></td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Hardware
- N/A

### Instrument Control SW
- N/A

### End User SW
- ✓

### Data Collection / Workflow Mgt. SW
- N/A

### Database
- ✓

### Know-How
- ✓
Generalized Informatics Blocks
Implemented at Example Company

- DOE
- Simulation
- Modeling
- Statistical Analysis
- Analytics
- Data Mining
- Visualization
- Synthesis / Modification
- Sample Preparation
- Properties Screening
- Design
- Data Analysis
- Library Design Tool
- Instrument Control, Data Collection and Management Tools/Toolkit
- Searching / Reporting
- Database and Database Application Server
First Workflow – Informatics Blocks

Library Design Tool → Preparation Station → Reactor System → Analytical Station #1

Analytical Station #1 → Analytical Station #2

Analytical Station #2 → Analytical Station #3

Library Design Tool → Inst. Control, Data Collection and Management Toolkit

Inst. Control, Data Collection and Management Toolkit → Searching / Reporting

Searching / Reporting → Database and Database Application Server

Copyright 2003 The Second Derivative - All Rights Reserved
Second Workflow – Informatics Blocks

Library Design Tool → Preparation Station → Reactor System

Analytical Station #1 → Analytical Station #2 → Analytical Station #3 → Analytical Station #4 → Analytical Station #5

Library Design Tool → Inst. Control, Data Collection and Management Toolkit → Searching / Reporting

Database and Database Application Server → Searching / Reporting
New Workflow at Example Company–Module Blocks

Assumption: Target throughput is significantly accelerated over traditional throughput

Library Design Tool → Liquid-handling Station → Proprietary Synthesis Station

Commercial Analytical Station
- Inst. Toolkit

Proprietary Analytical Station
- Inst. Toolkit

Database and Database Application Server

Proprietary Analysis Software
- New Software

Searching / Reporting
- Inst. Toolkit

Searching / Reporting
- Inst. Toolkit

Copyright 2003 The Second Derivative - All Rights Reserved
New Workflow – Informatics Implementation Based on Existing Foundation

Software Implementation Effort at Company (based on Existing Foundation):
67 months (~5.5 person-years), over a period of 16 months

Staffing and Disciplines (SW Only):
9 individual staff members, in addition to the core foundation team
- Oracle / SQL
- Object-Relational Mapping
- Automation
- Database: Reporting, querying, chromatograms/xy datasets, images, numbers
- Client Applications / Web Applications
- Developer kit
- C++ / ATL / COM / XML / XSLT / JAVA / HTTP...

Existing Foundation:
100 person-years invested
New Workflow – Informatics Blocks

Assumption: Target throughput is significantly accelerated over traditional throughput

Database: Track & Correlate data - #’s, images, xy data, chromatograms, etc.
New Workflow – Informatics Blocks

Design
- 1 Month
  - Excel
  - Fixed Designs

Liquid-handling Station
- 1 Month
  - Commercial Robot
  - Fixed Designs

Proprietary Synthesis Station
- 48 Months
  > 2 years to execute

Commercial Analytical Station
- 3 Months Each
  - Data Loaders

Proprietary Analysis Software
- 24 Months

Proprietary Analytical Station
- Web Reports / Project Reports
- 6 Months Each
  - Web Reports
  - Project Reports

Database: Track & Correlate data - #’s, images, xy data, chromatograms, etc.

- 18 Months
- 3 Months Each
  - Data Capture
  - Data Analysis
  - Data Loaders

- 3 Months Each
  - Data Loaders

- 24 Months
- 48 Months
- 1 Month
- 3 Months Each
- 6 Months Each
- 18 Months
- 1 Month
New Workflow – Informatics Implementation From Scratch

Software Implementation From Scratch:
116 months (~10 person-years), over a period of 2-4 years

Staffing and Disciplines (SW Only):
? individual staff members
- Oracle / SQL
- Automation
- Database: Reporting, Querying, Chromatograms/xy datasets, images, numbers
- Client Applications / Web Applications
- C++ / ATL / COM / XML / XSLT / JAVA / HTTP…

Support ~20% of Total
~1-2 people on an ongoing basis
New Workflow – Informatics Implementation From Scratch – Implications and Risks

Issues and Risks:

- Investment of people in core-vs.-context activities
  - 10 person-years to begin
  - 1-2 person-years ongoing

- Time-line to completion

- Risk of failure

- Initial cost estimates vs. true costs (“Are you in denial?”)

- Complexity beyond expectations

- Experience (or lack of) with large projects

- Uncertainly of workflow and methods

- Risk of data integrity

- Risk of data errors / erroneous data handling

- Lack of inter-project leverage

- Risk of ability to support on an ongoing basis
# New Workflow – Pre-Summary Build vs. Buy

## Costs to Build

<table>
<thead>
<tr>
<th>Category</th>
<th>Cost (M)</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hardware</td>
<td>~$1.0M</td>
<td>~15-20%</td>
</tr>
<tr>
<td>Software</td>
<td>~$2.5 - 3.0M</td>
<td>~40-50%</td>
</tr>
<tr>
<td>Engineering</td>
<td>~$1.0M</td>
<td>~15-20%</td>
</tr>
<tr>
<td>Chemistry</td>
<td>~$1.2M</td>
<td>~15-25%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>~$5.7 – 6.2M</td>
<td></td>
</tr>
</tbody>
</table>
Summary

- Evaluate business objectives:
  - Single project vs. many projects (Core or Context?)
  - Time to implement
- Evaluate the value of the workflow contemplated
  - Throughput, TTD, FTE’s, etc.
- Evaluate Costs to Build and Support Over Lifetime
  - Hardware: 15-20% $?
  - Software: 40-50% $?
  - Engineering: 15-20% $?
  - Chemistry: 15-20% $?
  - Ongoing Support: ~20% of (HW, SW, ENG, CHEM)
  - Total: $?
“Core” or “Context”?

- **Core:** Stuff that makes your business what it is
- **Context:** Stuff that has to be done but does not really differentiate

**Moral:**
Invest in “Core”, Outsource “Context”
Impact of Software Demonstrations
Demos…

THE TECHNOLOGY DEMO

THE SOFTWARE ISN’T 100% COMPLETE.

IF IT HAD A USER INTERFACE YOU WOULD SEE SOMETHING HERE… HERE…and sometimes here.

AND THEN YOU’D BE SAYING, “I GOTTA GET ME SOME OF THAT.”

ANY QUESTIONS?

DILBERT reprinted by permission of United Feature Syndicate, Inc.
A Great Demo

“Do the Last Thing First!”
A Great Demo

(Introduction)
1. Provide the Illustration - Summarize
2. Do it - Summarize
3. Do it again - Summarize
4. Questions & Answers
5. Summarize
How can demos go wrong?

A feature failed – software bugs/crashes
Failure to identify Critical Business Issues
Demonstrator didn’t know the product
Unknown or unqualified audience needs
Can’t drive the message
No story
Confusing story
Too long
Too boring
Too many features
Didn’t stop in time
Unclear story
Got lost in the story

No point to the story
No conclusion or poor conclusion
Broad range of audience needs
Disconnect between Sales and Technical
Capabilities didn’t match needs
Lack of demo skills
Lack of clear objectives for the demo
Too little time
Too much time
Equipment failure
Equipment unavailable
Questions interrupted the flow
People interrupted the flow
What bad things happen when demos fail?

**Development**
- Product capabilities are implemented poorly
- Wrong product built
- Wasted product development iterations
- Missed release dates

**Deployment:**
- User adoption is slowed or stalled
- Feedback cycle to vendor is attenuated – missed opportunities
- Training costs increase
- Professional services (consulting) costs increase
- Adoption is limited – “shelfware”
- Benefits delayed (ROI)

**Sales**
- Cost of sale increases
- Lost opportunities
- Value of sale is reduced
- Sales cycle is extended
- Sale is lost
- Fewer products/services sold
- Company misses quarterly or annual goals
- Salesperson misses quota
- Commission is lost or reduced
- People leave
Why do you (or your customers) build or buy software?

Existing Problems or Business Issues:
- Existing Problem 1 → Solution
- Existing Problem 2 → Solution
- Existing Problem 3 → Solution

Anticipated Problems or Business Opportunities
- Anticipated Problem → Solution
- New Opportunity → Realized
Solutions: What are the key capabilities provided by the software?

- Capability #1
  - Illustration
- Capability #2
  - Illustration
- Capability #3
  - Illustration
- Capability #4
  - Illustration

Solutions... that address a problem, an objective, or Critical Business Issue
Developing a Great Demo....
What do you need to create a Great Demo?

1. **Objective**
   - Proof of capabilities
   - Vision generation

2. **Organization Chart**
   - Audience members and roles

3. **Critical Business Issues**
   - CBI(s) and Specific Capabilities needed for each audience member

4. **Infrastructure**
   - When, where, how, who, with what…
Executing a Great Demo

(Introduction)
1. Provide the Illustration - Summarize
2. “Do it” - Summarize
3. “Do it again” - Summarize
4. Questions & Answers
5. Summarize
Some Comments About Collaborative Tools (WebEx and Similar…)

- Good alternative way to provide a demo
  - Reduces cost - travel and lost time on-the-road
- Most effective when a representative from your firm is present at the customer site (the demonstrator can be in California, via WebEx)
  - See the reaction of the audience
  - First-hand observations are important!
- Resist the urge to point at your own screen with your finger…!
  - I've SEEN this…! Shockingly, the customer can't see your hand…
- *Interactivity is the name of the game*
Some Definitions of “Integration”

Beauty, Workflows, and Integration are in the eye of the beholder…

HT Workflow Value Analysis

ROI is terrific

Workflow Software Build vs. Buy Analysis

Invest in “Core”, Outsource “Context”

Improving Demonstrations

Do the Last Thing First!
The Second Derivative helps software organizations improve their sales and marketing results by helping folks with demonstration and presentation skills, value analyses, pricing, and related thorny problems.

Contact Information
Peter E. Cohan
Principal
The Second Derivative
1538 Winding Way
Belmont, CA 94002
Telephone: +1 650 631 3694
Email: P.Cohan@comcast.net