Simplifiers and Complicators:
A Tale of Two Cultures

Dan Port
University of Southern California
CSE Technology Week
February 1999

"It was the best of times, it was the worst of times..."

Customers
Are often are not fully aware of relative cost and difficulty in ...

... implementing a requirement:
- Detailed features
- "Research" CS
- Unfeasible approaches
- Labor and/or machine intensive activities
- Not easily automated processes
- People-Technology clashes

Developers
... deploying, using, and maintaining a system:
- overly general features
- changes to workflow
- assumed maintenance skills
- "gizmos" and "gadgets"
- overly optimistic robustness
- accounting for evolution of the organization
- additional administration costs

A tale of great expectations...
Some Examples

1. A commercial customer specified a natural language interface for an otherwise simple query system. The project was cancelled after the natural language interface caused a factor-of-5 overrun in project budget and schedule.

2. A government customer specified a 1-second response time for an extremely large transaction processing system. Meeting this requirement involved a custom architecture and a $100M project. The customer authorized a prototyping activity, which determined that 90% of the transactions needed only 4-second response time. With this performance requirement, a commercial-technology-based solution was achieved for $30M.

3. A commercial developer delivers (late) a reservations system to a major resort organization that allows reservation operators to visually customize their interfaces. A large, full-time support team was hired to help operators configure and use the system. The system was scrapped after operator productivity dropped 40% and the developers are inundated with bug reports from the support team.

Significant Indicator of Potential For Unmet Expectations

LCO success condition:
- Describes at least one feasible architecture relative to cost/schedule/resource constraints
- Viable cost-effective business case
- Stakeholder concurrence on key system parameters

CS577 Projects That Failed LCO Criteria

1996: 4 out of 16 (25%)
1997: 4 out of 15 (27%)

why?
The Two Culture Problem Within CS577

- Difficult Win-Win issues did not get resolved by LCO
- Domain specific top-10 risk items and requirements taxonomy for Win-Win not detailed enough to manage expectations
- Hard fought compromises and resolutions in LCA causing late breakage and less than satisfactory results

1996 CS577 Expectation Problems

- A large photographic archive project focused on the issues of automating the archive’s operation, and seriously underestimated the amount of effort required to digitize the photos.

- A student film archive project focused on cataloguing and querying issues, and underestimated the performance problems involved in digital film distribution across a campus network.
1997 CS577 Expectation Problems

• A Virtual Reference Librarian project underestimated the complexity of handling natural language queries, and of building AI learning capabilities to infer frequently asked questions from a set of natural language queries.

• A project to convert legacy periodical records into a new set of common formats underestimated the complexity of creating a general set of record translators across a wide variety of periodical characteristics (variable publication frequency, volume numbering discontinuities, etc).

An Experiment in 1998 CS577: “Pre-position” Expectations

1. For each of the major digital library application sub-domains, we developed a characterization of the sub-domain, including lists of simplifiers and complicators (S&C’s) which would make the applications easier or harder to implement.

2. We provided the S&C’s to Library clients, with explanations in librarian terms by the Library project coordinators.

3. We highlighted the S&C’s in class lectures on risk management. Provided teams with Client-side S&C’s

4. We made the S&C’s the subject of an early student homework exercise: to pick two candidate 1998 projects, to identify their sub-domains, and the analyze their S&C’s by providing a specialized block-diagram and specialized S&C’s for those projects.
Example S&C’s

<table>
<thead>
<tr>
<th>Type of Application</th>
<th>Simple Block Diagram</th>
<th>Examples</th>
<th>Simplifiers</th>
<th>Complicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multimedia Archive</td>
<td><img src="image" alt="Simple Block Diagram" /></td>
<td>1, 2, 3, 4, 7, 12, 13, 14, 15, 16, 17, 18, 21, 24, 25, 26</td>
<td>• Use standard query languages  • Use standard or COTS search engine  • Uniform media formats</td>
<td>• Natural language processing  • Automated cataloging or indexing  • Digitizing large archives  • Digitizing complex or fragile artifacts  • Automated annotation/description or meanings to digital assets  • Integration of legacy systems</td>
</tr>
</tbody>
</table>

Example Specialized Block Diagram

Asian Film Database

Find Film by:
- Film Title
- Language
- Genre
- Keywords

Update Film Archive
- Add New Film Clip
- Add New Film Still
- Update Film Clip
- Update Film Still
- Delete Film Clip
- Delete Film Still

Film Information:
- Film Title
- Language(s)
- Genre
- Film Clip(s)
- Film Still(s)
- Synopsis
- Detailed Information
- Contact Information

Film Catalog
- Japanese
- Korean
- Chinese
- English

Asian Film Archive
- Film Clips
- Film Stills

Update Catalog

Asian Film Archive
- Display Film Clip
- Display Film Still
The first approach is a complex, error-prone, expensive natural language processing issue.
The second approach will require more storage space, in addition to acquiring the translations.

Dedicating film clips from the entire collection of films (which grows at a very fast rate of 800 films per year for Indian films alone).

If each film clip requires around 10 MB, then the size of growth of the database will be of 8 GB a year (exclusive of catalog information).

If each film clip requires around 10 MB, then the size of growth of the database will be of 8 GB a year (exclusive of catalog information).

Integration of Legacy'Systems

We cannot use more effective multi-media formats, which are becoming standard technologies.

Uniform Media Formats

This means that we may have to convert existing digital assets or digitize the original media which may be costly.

An unique file format limits the user base to those familiar with or using that particular file format.
The chosen file format may not be the most efficient for the various types of media (e.g. in terms of compression rates, quality, etc.)

We cannot use more effective multi-media formats, which are becoming standard technologies.

Uniform Media Formats

This means that we may have to convert existing digital assets or digitize the original media which may be costly.

An unique file format limits the user base to those familiar with or using that particular file format.
The chosen file format may not be the most efficient for the various types of media (e.g. in terms of compression rates, quality, etc.)

May not be as effective for "discovering" assets in the archive, users must know what they're looking for, in order to search for it.

For a standard Relational Database Management System (RDBMS) that supports storing multi-media assets.

A Relational Database Management System may not be most suited for archival of multi-media assets.

A Relational Database Management System may be a high initial cost, high implementation, set-up and training requirement (requires specialized knowledge)
### More Example S&C's

#### Selective Dissemination of Information

<table>
<thead>
<tr>
<th>Data Analysis</th>
<th>New library material notification</th>
</tr>
</thead>
</table>
| **Data Assimilation** | **Use of existing or
standard information base** |
| | **Well defined distribution points** |
| | **COTS notification and event processing** |
| | **WWW/Internet based** |
| | **Restricted interest vocabulary and filtering structures** |
| | **Single information base** |

#### Data Analysis

- Use of data analysis packages (e.g., PERL)
- Implementation using interpreted or script languages (e.g., PERL)
- Data stored in an RDBMS
- COTS reporting packages (for graphics, etc.)
- Simple, consistent data formats

**11.23, USC web-page stats**

- Use of natural language processing
- Highly unstructured data
- High degree of formalization or conversion
- Computationally intensive reporting
- Spatial data analysis

### More Example S&C's

#### Activity Monitoring and Control

<table>
<thead>
<tr>
<th>Room scheduling, Reserve materials management</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standards based agent interfaces</strong></td>
</tr>
<tr>
<td><strong>Simple, well-defined policies</strong></td>
</tr>
</tbody>
</table>

- Real-time or embedded monitoring
- Synchronization of monitoring activities
- Concurrency management of activities
- Distributed monitoring or activities
- Natural language processing of policy
More Example S&C's

<table>
<thead>
<tr>
<th>Distributed Borrowing</th>
<th>Asset Source</th>
<th>9, 19</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interactive Communication</th>
<th>10, 19</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Homogeneous asset sources
- Simple asset source interfaces
- Few asset sources
- Complex borrowing policies and requests
- Organization politics and economics
- Internet/WWW interface
- CONIX communication services
- High degree of integration
- Real-time
- Synchronous
- Concurrency
- Rich media (video, voice, NLP, etc.)

List of Examples

1. Architecture & Fine Arts Databases
2. Bella Lewitsky Archives
4. CNTV Moving Image Archive
5. EDGAR Corporate Data
6. General Library FAQ's
7. Hancock Library Photographic Archives
8. Hancock Museum Virtual Tour
9. Interlibrary Loan Data & Interface
10. International (French) Cross-Cultural Teaching Model
11. ISLA Forms Development for Data Ingest
12. ITV Course Materials
13. Korean-American Museum
14. Latin American Pamphlets
15. Lion Fauchranger Archive
16. Los Angeles Regional History Photographic Archives
17. Maps for Instruction
18. Medieval Manuscripts
19. Networked Consultation Support
20. Online Catalog Search Strategies
21. Planning Documents
22. Serial Control Records for the Integrated Library System
23. Statistical Charts
24. Stereoscopic Slides
25. Technical Reports Archive
26. TV Show Files
27. Virtual Business Reference Assistant
28. Virtual Education Reference Assistant
### Client-side S&C’s

<table>
<thead>
<tr>
<th>Simplifiers</th>
<th>Complicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Project scope fits within client’s authority scope</td>
<td>• Scope crosses organizational boundaries</td>
</tr>
<tr>
<td>• Solution reduces job tedium, reduces procedural delays</td>
<td>• Solution creates more user work, dehumanizes personal interactions</td>
</tr>
<tr>
<td>• Solution reduces organizational friction, infrastructure clashes</td>
<td>• Solution shifts power, confuses lines of authority, puts outside parties on critical path</td>
</tr>
<tr>
<td>• Task-tailorable user interface</td>
<td>• Mismatches between user interface and user tasks, capabilities</td>
</tr>
<tr>
<td>• COTS product features anticipate direction of growth</td>
<td>• COTS product features evolving toward different marketplace</td>
</tr>
<tr>
<td>• Hidden costs: licenses, data entry, conversion</td>
<td>• Mismatches with existing legacy-system constraints</td>
</tr>
<tr>
<td>• Creeping (baroque) elegance</td>
<td>• Single-criterion optimization: speed, correctness</td>
</tr>
<tr>
<td>• Creeping (baroque) elegance</td>
<td>• Creeping (baroque) elegance</td>
</tr>
</tbody>
</table>

### The Results

"It is a far, far better thing that I do, than I have ever done…"

**CS577 Projects That Failed LCO Criteria**

1996: 4 out of 16 (25%)
1997: 4 out of 15 (27%)
1998: 1 out of 20 (5%)
The Skewed Tao of S&C’s

- Every complicator has a complementary simplifier.
  - Example:
    - multiple media formats
    - uniform media formats

- Not every simplifier has a (non-trivial) complementary complicator
  - Example: Use of data analysis packages