



100 Year Archive Dilemma

SNIA-DMF

100 Year Archive Task Force



Introduction

- **Peter Mojica**
- **VP, Product Strategy, Management**
- **AXS-One Inc.**
www.axsone.com
- **Member 100 Year Archive Task Force (SNIA-DMF)**

- CreekPath Systems (Boulder, CO.)
 - Chief Solutions Architect
- Selectica (San Jose, CA.)
 - Worldwide Director, Channel Sales
- First Union National Bank (Charlotte, NC)
currently Wachovia Securities
 - VP Research & Development
- Wall Street Financial Services
 - Nomura Securities, CSFB, Chemical Bank
- British Airways
 - Systems Development Management
- Legal Industry
 - Herzfeld & Rubin (Wall Street)
 - Automotive Product Liability
- Sargoy Stein & Hanft (New York City)
 - Motion Picture Anti-trust



100 Year Archive Task Force

- Charter
 - Drive problem awareness
 - Provide education to end users and executive managements
 - Deliver pertinent and actionable information that is usable today
 - Create a long-term solution with wide adoptability across all of the moving pieces



What's the problem?

- A simple example:
 - The validity of my patent is being challenged - the stakes are high
 - I have my original work-product contained in word processing documents and spreadsheets
 - I saved these on a floppy disk, the 360K variety, and I need to access the data
 - Today's date is 2006
 - The data was created and saved in 1986
 - Just 20 short years ago



Not that tough?

- **I need a media reader?**
 - Solution: eBay
- **I need additional hardware accessories, PC, cables, etc.**
 - Solution: eBay
- **I need an Operating System, Dos 2.0**
 - Solution: Microsoft Museum
- **I need the original applications or something to convert the data to a common readable format?**
 - find correct versions of Multi-mate, Lotus 123 and Harvard Graphics

- I found a media reader – 7 days
- I found additional hardware that I need – 14 days
- I located an Operating System – 7 days
- I have not located the original applications but I have a converter program which translates Multi-mate and Lotus 1-2-3 to text
- Harvard Graphics is a problem, I've hired a developer to work on that problem
- The judge has provided an extension of time to produce my evidence
- I'm in good shape ?



Now let's take a quick look at the importance of "context" in relation to long-term data archival

- I successfully converted the Lotus 123 documents to text
 - In the conversion I lost the spreadsheet formulas
 - The resulting numbers without the formulas are useless
- Lesson:
 - Content and Context are Critical

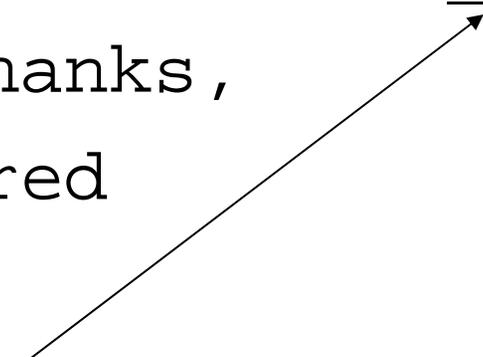


Another look at the importance of “context”

John,
Meet me in the
boardroom now.
Thanks,
Fred

John,
Meet me in the
boardroom now.
Thanks,
Fred

Add underline mark





Another look at “context”

John,
Meet me in the
boardroom now.
Thanks,
Fred

John,
Meet me in the
boardroom **now**.
Thanks,
Fred



What difference does “context” make

John ,
Meet me in the
boardroom now.
Thanks ,
Fred

John ,
Meet me in the
boardroom **now**.
Thanks ,
Fred



Possibly all the difference in the world

John ,
Meet me in the
boardroom now.
Thanks ,
Fred

John ,
Meet me in the
boardroom **now**.
Thanks ,
Fred



Simple examples, now add scope and different perspectives

- Exasperate the simple example from a single person to 100,000 end users participating in complex business transactions, using 100's of complex business systems and applications across the globe for a single global organization
- Do you get a sense of the enormity of the problem
- Now add:
 - hardware obsolescence
 - Application and data migrations
 - Conversions to future versions of hardware and software
 - Electronic Information and Paper



The post S.O.X, Enron business driver

- Business Contracts
- Medical Records
- Product Liability
- Financial Transactions
- Electronic Communications
 - E-mail
 - Instant Messages

- **Litigation:** *“were getting sued, and we need to prove “beyond a shadow of a doubt” what we did last year, last decade, last”*

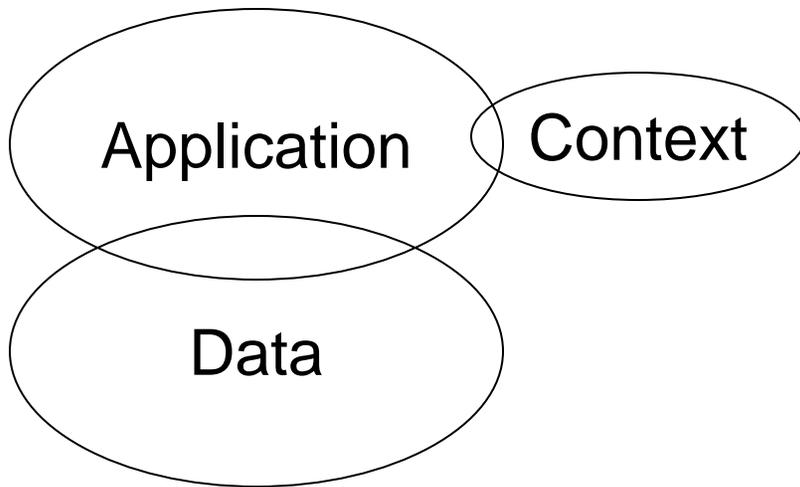


Evolving Business requirements are driving change

- **A simple analogy of doing more to keep up with the new business requirements.**
- First off: What was the business requirement?
 - One way
- What is the new business requirement?
 - Two-way +
 - Multiple systems technologies and business processes have to be more closely aligned
- Lesson: New business drivers are adding to the complexity of the 100 year archive dilemma.
- Context remains critical
- Cooperation across vendors and technologies is key

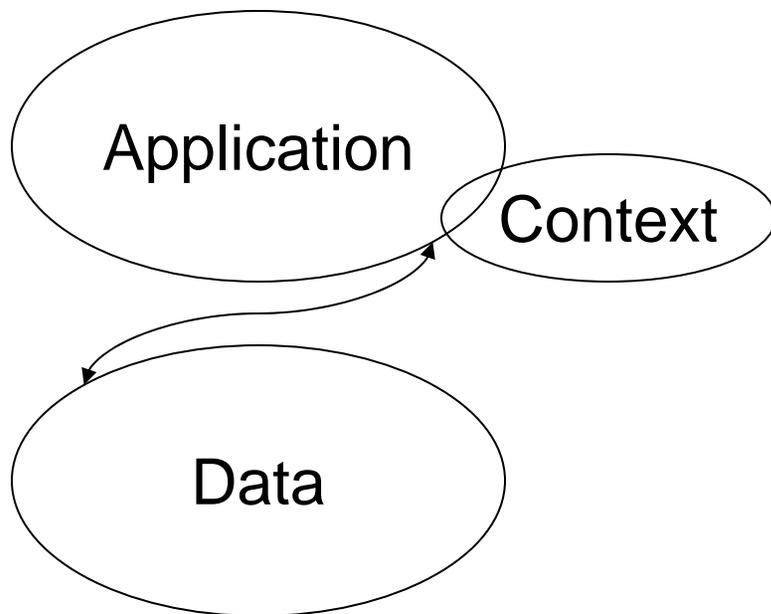


Applications, Data and Backup



- **First generation applications**
- **Tightly coupled**
- **Context contained within application**
- **Backups preserve application and data for disaster recovery**
- **Require application to restore usability**
- **Does not protect over the ravage of time**

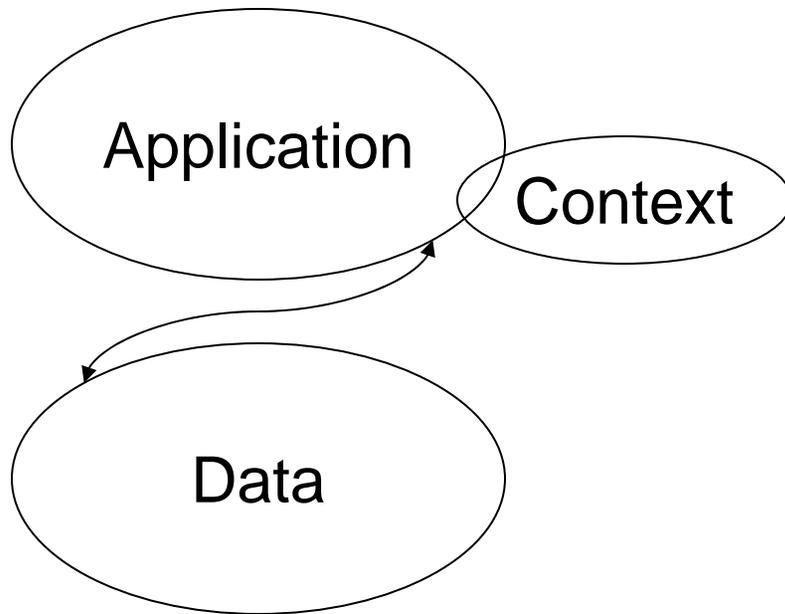
Applications, Data and Backup



- Second and third generation applications
- Loosely coupled
- Context still contained within application
- Backups still require applications to restore usability and maintain context
- Other important issues:
Timed expiration of data



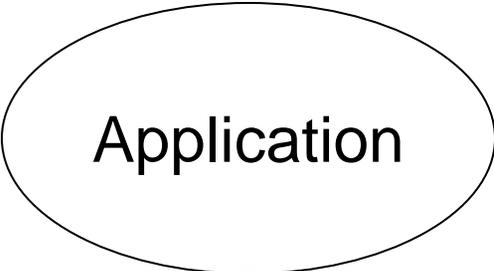
Applications, Data and Context



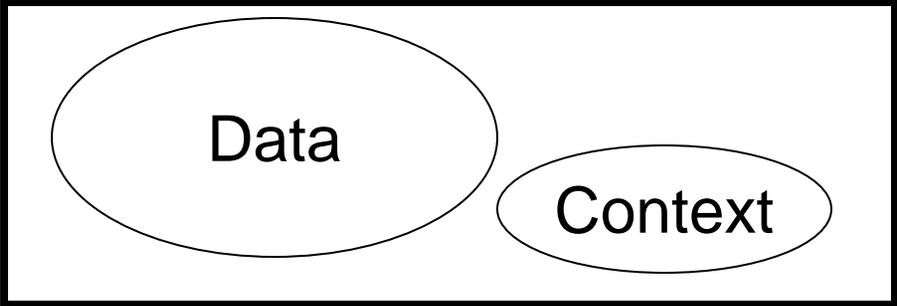
- Second and third generation applications
- Loosely coupled
- Context still contained within application



Applications, Data and Archival



Archive



- Archival stores the data
- The context
- Stores efficiently
- Reproduces irrefutability
- Maintains rich text
- Provides Audit
- Full-text indexing for search
- It solves business problems which span over a > 100 year life
- Key: archive data without regard to hardware platform, OS, or application logic.



Archiving Standards





The solution

- What is the 100 Year Archive Task force doing ?
 - 1) Talking to the business users, fortune 500, government agencies, across verticals and geography
 - Information and proper requirements gathering for a deeper understanding of the new business requirements is critical.
 - 2) Developing best practices which can be utilized immediately.
 - 3) Longer-term standards for solving meta data taxonomies that describes context, solves format, delivers independence from hardware, operating systems and applications.
 - 4) Working on cooperation and understanding that people, process and business requirements are key, underlying technologies require transparency.
- Thank you.
 - Many moving pieces working together to form a cohesive understandable archive that solves real world business problems in an immediate fashion.



More questions

- www.snia-dmf.org/100year



Introduction

- **The 100 Year Archive task force**
- The challenges of retaining online archives for extended periods of time in the face of variable media life, equipment obsolescence, and changing operating environments is daunting. This concept of a "100 Year Archive" is the guiding principal for LTACSI's long-term archive projects. Two major challenges include maintaining physical readability, the ability to get back the bits written, and maintaining logical readability, the ability to understand and manipulate those bits. The 100 Year Archive task force is a multi-organizational team preparing educational materials for end users on how to tackle these problems.
- Among other projects, the task force will create a white paper to describe best practices and recommendations for extended archive and compliance lifetimes, including:
 - Decision factors and best practices for long-term retention
 - Technology choices, media life and technology refresh considerations
 - Discovery, search, legal and regulatory considerations