Digital Libraries and Repositories: Issues and Trends

March 2009

Art Pasquinelli
Education Market Strategist
Global Education & Research
Sun Microsystems, Inc.

Agenda

- Sun and the Library Market
- Today's Trends and Challenges
- Addressing Today's Challenges
- Sun Directions and Summary

Sun's Digital Library Value Proposition

We Help Digital Libraries, Archives, and Repositories Develop Open, Scalable, Secure Environments for Knowledge Development, Discovery, Management, and Sharing

Sun and the Library Community

- **Types of Sun Customers:**
  - National and State Libraries & Archives
  - Academic and Research Libraries
  - Public Libraries
  - Large Museums
  - Library Consortia
  - Large Primary and Secondary Organizations

Sun Partner Library Solutions

Ex Libris

SirsiDynix

EXANET
BREAK THE STORAGE LIMITS

VTLS

webfeat

INNOVATIVE INTERFACES

grokker

CONTENTdm®
Digital Collection Management Software by DiMeMa, Inc.

DSPACE

LUNA

Access Innovations

OCLC PICA

intrallect

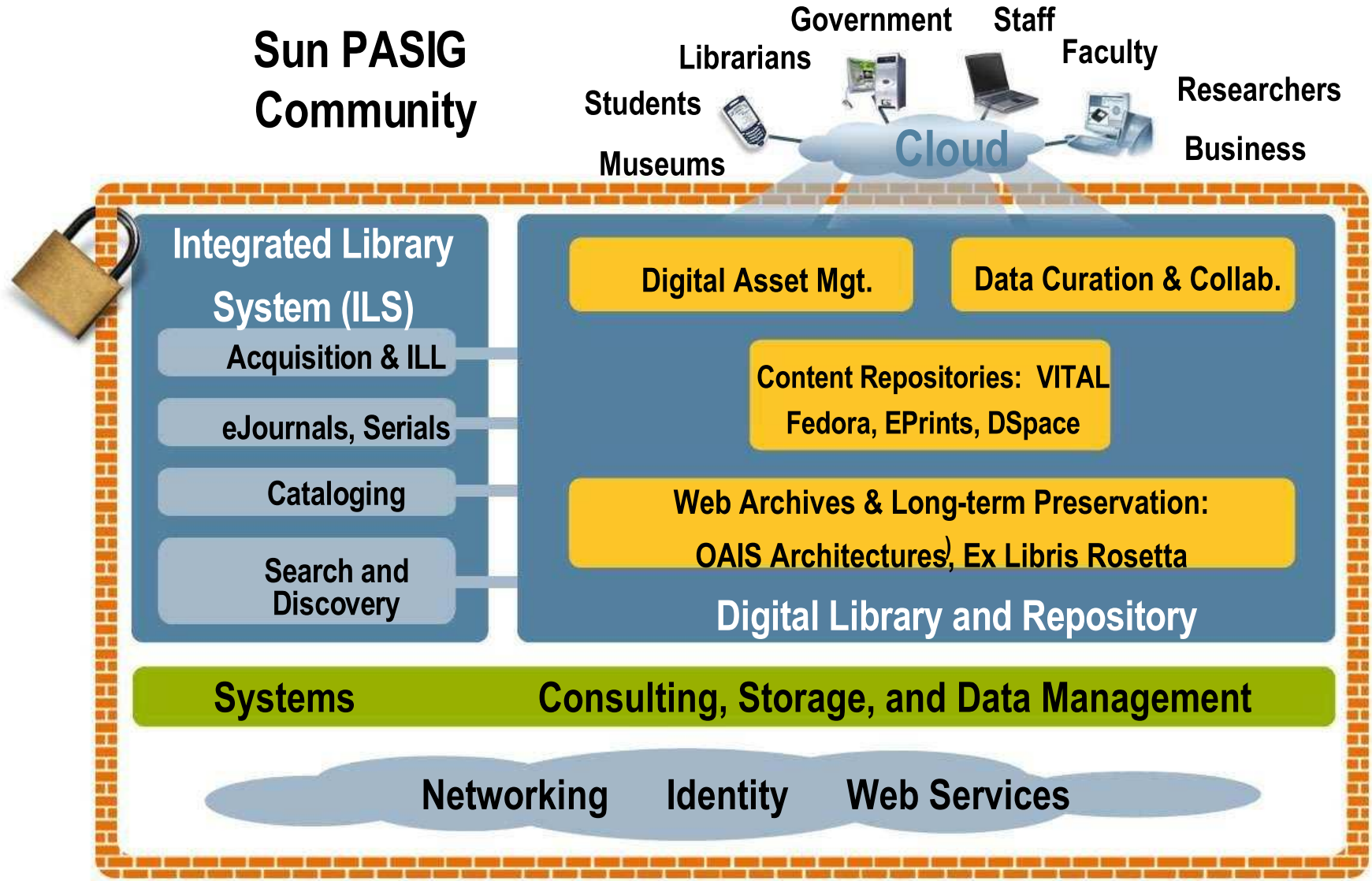
AXIELL bibliotekernes IT-partner

fedora

artesia
The Open Text Digital Media Group

Sun's Vision for Tomorrow: The Knowledge Marketplace

The Digital Library "Map"



Scalability, Security, Sustainability, Sharability

Some Challenges Libraries Face

- Impact of the Social Web: New Services and Demands
- Staff Training and Technology Developments
- Traditional Content Growth
- Security and Longevity of Materials
- Re-inventing the Physical Library as a Social Space

Some Challenges Digital Libraries and Knowledge Institutions Face

- Data Management
  - Long-term Sustainability and Permanent Access
  - Content Growth – Francine Berman's “Got Data” Article...Zettabytes!
  - Types, Value, New Uses of Content - Long-Tail Data, Curation Impact
  - Disaster Recovery and Business Continuity
  - Predicting Future IT Economic Models – Scalability, New Technology, Power Cost

- Growth of Repositories and Cloud Computing
  - Federation and Connectivity – Trucks vs Datagrid?
  - Service Level Agreements – Public and Private Clouds?
  - Broader Access to New Services – Duraspace, Mellon's AVAN, Datanet Projects

Some Challenges Digital Libraries and Knowledge Institutions Face

- Content and Data Creation and Sharing Models
  - Industry Collaboration and Standards Co-Development – XAM, SNIA, iRODS
  - New Funding Models and Data Use Requirements - Datanet
  - Breaking Out of the University 'IT Grasp' into the Knowledge Marketplace: How much does your IT department charge you to host a TB of data?
  - Documents Tied to Data – Drug and Pharmchem Companies

- Functional Collaboration
  - Communicating Around Datasets – Semantic Web Opportunity?
  - User Rights – Authentication, Access, Authorization
  - Immersive Education Technologies

Challenges of Repository Projects

- Objects Must be Always Retrievable (Access)
- Cost and Complexity as Systems Scale (Economic Sustainability)
- Finding Data through Sophisticated Metadata Handling (Discovery)
- Data Integrity Must be Assured (Trust)
- Seamless Scaling Must be Provided (Scale & Extensibility)

Carl Grant, President, President Ex Libris N. America

The Challenges of Digital Preservation

- Bit Rot
- Obsolescence
  - Format
  - Technology
- Distribution and Dissipation
- Migrations and Transitions
  - People (2 – 20 years)
  - Software (5 – 10 years)
  - Hardware (3 – 5 years)

Benign neglect doesn't work for digital objects. Preservation requires active, managed care.

Tom Cramer, Stanford U. Library

Addressing the Challenges: Mutual, Enlightened Self-Interest

- Government, Business, and Education Cooperation
  - Funding Optimization
  - Standards Development
  - Collaboration
- Goals: Lower Risk, Increase Innovation, Broaden Participation
  - Industry Associations
  - Open Computing and Storage
  - Sharing Best Practices via Communities

Industry: Storage Networking Industry Association (SNIA) 100 Year Archive Task Force

- Over 80% report a need to retain information over 50 years, and 68% report a need of over 100 years
- Long-term generally means longer than 10 to 15 years
- Over 40% of respondents are keeping email records over 10 years
- Database information was considered most at risk of loss
- 70% of respondents say they are 'highly dissatisfied' with their ability to read their retained information in 50 years
- Current practices are too manual, too prone to error and too costly
- Collaboration is recognized as necessary in order to define information retention requirements

Key Findings

Logical and Physical Migrations Do Not Scale Cost-Effectively

Only operating standard today is to migrate information physically (to new media) every three to five years and logically (to new formats) before the applications and readers die and become obsolete (every 5-10 years).

> **A never-ending, costly cycle of migration**

Practitioners are struggling to keep up with migration requirements. Only 30% claimed to be doing physical migration correctly on disk & none on tape or optical. Only 20% claimed they were confident in their ability to logically migrate some of the data.

> **Information is at risk long-term**

Raymond Clarke, Enterprise Storage Architect, Sun, SNIA Technical Board Member

Open Computing and Storage

Open Architectures are Essential for Long-term Sustainability

The Sun PASIG (www.sun-pasig.org)
Started in 2007 by Stanford U. and Sun Microsystems

- Comparison of High-level OAIS Architectures, Workflows, and Use Cases
- Sharing of Best Practices and Community-developed Solutions and Technologies
- Cooperation on Standard, Open, Solutions and Replicable Reference Architectures
- Review of Storage Architectures and Trends and their Relation to Preservation and Archiving and Research Data Set Management
- Exposition of Relevant Commercial Third Party Expertise and Solutions

Focus of the Sun PASIG June 24-26, 2009 Meeting in Malta

1) Storage and Data Management Architectures
2) Preservation and Long-term Sustainability
3) Repository Directions
4) Research Data Curation
5) Cloud Computing
6) Managing Large eResearch Datasets
7) Digital Asset Management

Upcoming Sun Community Events

- IS&T Archiving 2009 – May 4-7
- Ex Libris N. America User Group – May 5-8
- Open Repositories 2009 – May 18-21
- Sun PASIG Europe – June 23-26
- I-Pres – October 2009
- Sun PASIG N. America – October 2009
- Others:
  - Sun Immersion Special Interest Group
  - Sun HPC Consortium

Sun Reference Architectures

Develop Collaborative, Replicable Reference Architectures for and With the Community and Partners

- Fedora
- Fedora/Drupal
- DSpace
- iRods
- EPrints
- Duraspace
- Ex Libris Rosetta
- VTLS VITAL
- Internet Archive in a Sun Modular Datacenter (March 25)

Sun References by Topic

Slovakian National Library	Broad Digitization, Sun Rays
Stanford U.	OAIS Digital Repository
Johns Hopkins U.	eResearch/Data Curation (Fedora)
CNRS/CINES	eResearch Data
Oxford University	Fedora Repository, Sun Rays
National Library of New Zealand	Digital Preservation (Ex Libris)
California Digital Library	Large Scale Digitization
NYU	Digital Asset Management
US Library of Congress	Broad Digitization, SAM-QFS
French National Library	Broad Digitization
Texas Digital Library	DSpace Repository
Norwegian National Library	Broad Digitization, SAM-QFS
Alberta Digital Library	Long-Term Sustainability

Sun's Future: Focus on Your Architectures

- **Focus on Tiered, Open Architectures – Not All Content has the Same Value**
  - SAM, ZFS, Infinite Archive Solution (IAS), MySQL, Tape
  - Open Storage
  - Focus on Policies and Procedures

- **Distributed Architecture and Business Continuity**
  - Sun Modular Datacenter “Black Box”
  - Cloud Computing
- **Rich Media and Digital Asset Mgt.**
  - SAM, Tape, Low Cost Disk, JBODS
- **Identity Management**
  - Authentication and Authorization for Repositories

Thank you!

art.pasquinelli@sun.com

http://blogs.sun.com/pal/