VTLS Update on VITAL
PASIG 2009 Fall Meeting
San Francisco, October 7-9, 2009

Vinod Chachra, President & CEO
VTLS Inc. Blacksburg, VA, USA
About VTLS Inc.

VTLS is the first spin-off corporation from Virginia Tech.

Vinod Chachra served as VP, Information Technology at Virginia Tech before starting VTLS in 1985.

VTLS does business in 40 countries.

VTLS has offices/wholly owned subsidiaries in 7 countries.

VTLS has four major products

- **Virtua**: Integrated Library System
- **VITAL**: Institutional Repository Software
- **Visualizer**: Discovery tool for distributed content
- **VTRAX**: RFID based tracking & security systems for libraries
More About VTLS Inc.

Employee loyalty at VTLS is extremely high. With more than 20 Years of service are:
- Vinod Chachra, President & CEO
- George Nichols, VP Finance
- Jack Bazuzi, Executive VP
- Rebecca Shaver, Product Manager, Virtua
- John Espley, Chief Librarian
- Ginny Tillage, Director of Development
- Todd Perry, Chief Technologist

VTLS World Headquarters are located in Blacksburg, VA in a beautiful 32,500 square foot facility located at the Corporate Research Center (CRC) at Virginia Tech.

VTLS was the first tenant at CRC

VTLS employs
- 80 FTE in Blacksburg;
- 120 FTE worldwide.

VTLS is privately owned and privately financed.
VTLS Information Technology
for Advanced Learning
It is a FEDORA based institutional repository
VITAL / Fedora Relationship

Application Developers  Public Access  Staff

Access Control

Web Services  Web Interface  Vital Manager

Open Source Fedora™ Repository

Other Digital Collections

Dublin Core/EAD/TEI/XML

Digital Content

OAI
How does VITAL work with Fedora™?

Workflow tools that simplify the creation, modification, submission and dissemination of digital objects

Web Service Interfaces (API’s)

Management Service (API-M)
- Ingest – XML-encoded object submission
- Create – interactive object creation via API request
- Maintain – interactive object modification via API requests
- Validate – application of integrity rules to objects
- Identify – generate unique object identifiers
- Security – authentication and access control
- Preserve – automatic content versioning and audit trail
- Export – XML-encoded object formats
VITAL architecture overview

- Access Portal
- Batch Loading Tool
- Vital Manager
- Valet
- Indexes
- Web services
- FEDORA
Enter ARROW - What is ARROW?

Australian Research Repositories Online to the World

Project funded by the Australian Commonwealth Department of Education, Science and Training (DEST), under the Research Information Infrastructure Framework for Australian Higher Education. (Funding ended December 31, 2008)

“The ARROW project will identify and test software or solutions to support best practice institutional digital repositories comprising e-prints, digital theses and electronic publishing.”

[Reproduced with permission from ARROW]
ARROW/VTLS – Funding Model

- ARROW started as a “demonstration project” to prove feasibility of IRs
- ARROW partnered with VTLS-VITAL and ended as a “operational IR”
- VTLS developed software in close co-operation with Fedora & ARROW.
- The funding model:
  - ARROW provided funding to VTLS against some deliverables
  - The VTLS development project was required to coordinate and synchronize with Fedora developments.
  - As a part of the overall plan, VTLS would place certain modules in “open source” for all Fedora users whether they were VITAL users or not.
  - The initial ARROW consortium would get the VITAL/Fedora software; but would be required to pay for installation, training and support services
  - Additional members could “buy in” at advantageous prices for a predefined period
Some key benefits of VITAL include:

- Storage and management of any content format, including rich-media, due to VITAL's repository object architecture
- Integration with existing systems through open, standards-based protocols
- Search full-text content of PDF, DOC, RTF and other document formats
- Display high resolution imagery, multi-page documents and specialized data formats (MARC, EAD, TEI, etc.)
- Automatically capture preservation metadata and create long-term, citable DOIs
- Storage of content in native XML provides support/validation against any metadata schema
- Support for aggregation of like-content in collections
- Automatic metadata extraction and validation (facilitates preservation)
Additional benefits of VITAL include:

- Web-crawler indexing and exposure (Google, etc.)
- Language support and UNICODE compliance
- Support for custom index creation for resource discovery
- SRW/SRU Interface for exposure of repository content
- Tracking content changes through versioning
- Robust facilities for batch ingest of content
- Support for link resolving against an institution’s OpenURL server
- Support for dynamic creation/export of content citations (EndNote, text)
- Facilities for web-based self-submission of diverse content types (VALET)
- Integration with Fedora™ repository architecture allows for extensibility
VITAL 3.x - What’s New

- Web-Based Content Administration
- Highlighting:
  - Author Content
  - Works within Collections
- Automatic Notification of New Content (RSS Feeds)
- Quick Search ‘cart’ for selecting and exporting content citations
- Filter search results using categorized metadata ‘facets’
- Enhanced browsing of image content
- User Authentication via LDAP and Shibboleth
VITAL 3.x - What’s New

- Multilingual Support (more on this ...)
- Performance Testing and Tuning
- ARROW performance tests and acceptance (more on this ...)
- Quick Edit Improvement
  - Add tags as well as edit existing tags
- Report Results Filtering by Date
- New OAI-PMH Provider
VITAL 4.x - What’s New

- Incorporates all 3.4.2 bug-fixes
- Includes 3.4.2 performance enhancements done for ARROW
- New 4.0 features include support for
  - Consortia environments
  - Vocabulary Lists
- Access Control and Authentication (more on this ...)
- Additional Bug-fixing
Future Releases – Coming Soon

- Agile Development Process Implemented
  - 2-week fast cycle iterations
  - Starting with all releases following 4.X
- Enhancements under development include:
  - VITAL serving as a DRUPAL module
  - Embargo
  - RDF GUI
  - Support of additional metadata formats
In Chinese
and other languages
寻找机构知识库解决方案？VITAL是最佳答案！

VITAL是为大学、研究机构、博物馆、档案馆以及信息中心等设计的知识库解决方案。基于Fedora®开发平台，VITAL的设计目标是提供简化知识库的部署。它为系统管理员、联系文献管理学者和研究用户提供一套通用工具和使用管理资源的机制。VITAL为各种类型的机构广泛提供有价值的资源提供了方便的途径。在VITAL系统创建完成后，那些有价值的资源将被整合于统一的搜索系统和管理工具中。

VITAL仍然是知识库解决方案的领先者。我们继续改进我们的系统，以适应研究人员的需求。继续访问我们的资源。
显示文献 1 - 15 of 24.
鸟巢
<table>
<thead>
<tr>
<th>题名</th>
<th>创作者</th>
<th>日期</th>
<th>全本文</th>
</tr>
</thead>
<tbody>
<tr>
<td>无线电爱好者登黄岩岛</td>
<td>高, 如实, 张</td>
<td>2006</td>
<td>—</td>
</tr>
<tr>
<td>墨尔本风光</td>
<td>高, 如实, 张</td>
<td>2007</td>
<td>—</td>
</tr>
<tr>
<td>葛本海姆博物馆</td>
<td>高, 如实, 张</td>
<td>2004</td>
<td>—</td>
</tr>
</tbody>
</table>

保存/发送引文

- 引文格式：HTML引文
- 主题：VTLS VITAL引文
- 发送引文：发送引文到硬盘

- 马, 建霞, 张, 晓林, ..; 2007. 机构知识库问卷。
- 张, 晓林, 无名氏, ..; 2007. 墨尔本风光。
- 倪, 宁宁, 戎, 华, ..; 2006. 无线电爱好者登黄岩岛。
- 张, 晓林, ..; 2005. 关于语义架构的一些思考。
- 高, 如实, 张, 晓林, ..; 2004. 葛本海姆博物馆。
Load Test Results
VITAL Performance Testing for ARROW

Goal

Establish that the VTLS VITAL/FEDORA solution had reasonable performance on an average hardware configuration with 120 simultaneous users.

*The ARROW system acceptance was based on satisfactory results from the performance tests.*

Functional tests were previously completed. The production software was VITAL Release 3.4.2 running on a Linux x86 platform.
Hardware Configurations

Performance tuning was initially done on VTLS machines and performance testing was done by a third party on the customer machines.

This arrangement was not satisfactory due to the

   a) delays in the process of transferring software from the tuning machine to the testing machine

   b) inability to monitor the Australian systems (while the tests were running) from USA.

Subsequently the hardware configuration in Australia was replicated at VTLS.
Hardware Configurations

Customer hardware configuration:
4 Intel(R) Xeon CPU 2.80GHz each with 512KB cache and 2.5 GB of RAM.

VTLS test system hardware configuration:
4 Intel(R) Xeon CPU 3.00GHz each with 2048KB cache and 6 GB of RAM.

For both systems, the Java Virtual Machine was allowed a maximum heap size of 1GB, so the difference in RAM should not be relevant. In any case, the performance tests had to pass on the customer machine.
VITAL Performance Testing for ARROW

Software Configurations and Scripts

Both systems were running the same version of the repository software – VITAL Release 3.4.2

FEDORA 2.2.4 Version was used on both systems

HP LoadRunner 9.50 (testing tool used by ARROW’s independent consultant)

Test Scripts were developed by ARROW.

Testing on the VTLS system was done by VTLS.

Testing on the ARROW system was done by the independent consultant.
Load Profile and Transaction Mix

Searches using single term: 20%
Searches using string of up to 5 terms: 15%
Searches using two terms and single boolean operator: 15%
Display of simple objects: 30%
Display of complex objects: 10%
Downloading of file: 10%

Data streams in the test database were limited to a maximum size of 10MB.
Simultaneous User Load Profile

Number of users: 120

Each user was running through scripts containing the transaction mix.

Average hits per second: 41

Average throughput: 800,000 bytes/second
**Length of Performance Tests**

VTLS ran several 72 hour full load tests with tuning and software fixes between runs.

ARROW ran several tests in the performance acceptance process.

a) An initial full load test was run for 22 hours.

b) After the test above VITAL was left running for 5 full days,

c) A second second 48 hour full load test on the same instance of VITAL without a restart.
LoadRunner Average Transaction Response Time
Graph and chart from customer's 48 hour 120-user load test follows

"Description: Displays the average time taken to perform transactions during each second of the load test. This graph helps you determine whether the performance of the server is within acceptable minimum and maximum transaction performance time ranges defined for your system."
# LoadRunner Performance Results Chart

<table>
<thead>
<tr>
<th>Color</th>
<th>Scale</th>
<th>Measurement</th>
<th>Graph's Minimum</th>
<th>Graph's Average</th>
<th>Graph's Maximum</th>
<th>Graph's Median</th>
<th>Graph's Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>Portal_AdvancedSearchPage</td>
<td>0.535</td>
<td>0.697</td>
<td>1.223</td>
<td>0.655</td>
<td>0.157</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Portal_AdvancedSearchSubmit</td>
<td>1.323</td>
<td>1.603</td>
<td>2.408</td>
<td>1.548</td>
<td>0.254</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Portal_AdvancedSearchSubmitAndOpen</td>
<td>3.416</td>
<td>4.483</td>
<td>6.116</td>
<td>4.372</td>
<td>0.641</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Portal_DownloadFile</td>
<td>1.234</td>
<td>1.465</td>
<td>2.105</td>
<td>1.419</td>
<td>0.184</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Portal_Home</td>
<td>0.61</td>
<td>0.71</td>
<td>1.106</td>
<td>0.673</td>
<td>0.115</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Portal_OpenObject</td>
<td>1.9</td>
<td>2.396</td>
<td>3.898</td>
<td>2.254</td>
<td>0.463</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Portal_PhraseSearch</td>
<td>1.392</td>
<td>1.679</td>
<td>2.473</td>
<td>1.63</td>
<td>0.245</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Portal_PhraseSearchAndOpen</td>
<td>3.296</td>
<td>4.577</td>
<td>6.982</td>
<td>4.463</td>
<td>0.692</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Portal_SingleSearch</td>
<td>1.418</td>
<td>1.75</td>
<td>2.524</td>
<td>1.711</td>
<td>0.252</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>Portal_SingleSearchAndOpen</td>
<td>3.342</td>
<td>4.438</td>
<td>6.286</td>
<td>4.312</td>
<td>0.665</td>
</tr>
</tbody>
</table>
Graph of Data from Performance Test
Displays the average time taken to perform transactions during each second of the load test.
Graph of Data from Performance Test
These spikes were due to some external backup process.
Results

In all cases the average response times were much lower than the required response times.

System Accepted !!
VITAL users in Australian Universities

Monash
UNSW
Swinburne
South Australia
Central Queensland
LaTrobe
Murdoch
Macquarie
UNE
Sunshine Coast
Newcastle
New England
Ballarat
Western Sydney
Who else is using VITAL?

Australia  16 Institutions including Monash University
USA  Duke Medical Library
Virginia College of Osteopathic Medicine
Mary Washington University;
Columbia University (CIESIN project),
Virginia Tech
Europe
UK:  National Library of Wales
Greece:  National Theatre and Athens Archaeological Society
Belgium:  UCL and its consortium
Slovakia:  National Library
Middle East
Kuwait Institute of Scientific Research
Asia
Pakistan Medical Research Council
Singapore National Library

Others are testing the system.
Highlighted VITAL Users

- National Library of Wales
  - Digitizing over 1 million historic wills from Welsh collection
  - Integrated repository with ILS
  - Metadata publicly available, Images can only be viewed within library

- National Library Board of Singapore (internal access only)
  - Over 1.5 million objects
  - Base for all data in National Discovery Service

- Virginia College Of Osteopathic Medicine (VCOM)
  - OSTMED-DR – digital library of research publications in OM.
  - Streaming videos of recorded classes using clouds.
  - Access control system described later used by VCOM.
ARROW’s view… Why partner with VTLS?

Partnering for success, support and survivability

ARROW needed to partner with a developer who could not only produce the software but could provide ongoing user support and development after December 31, 2008

Why VTLS Inc.?

- VTLS wanted to be a development partner
- Had begun work on a repository solution already
- Familiar with library sector
- Willing to produce a combination of a proprietary solution, Fedora and other Open Source software
More VITAL Performance Testing

Moving Forward

Similar tests to be conducted on SPARC platform.

Some VITAL users have large repositories running on SPARC.

VTLS is also running some performance tests on streaming videos. These videos are recorded lectures used as refresher classes by students and for self paced offsite learning.

Apache’s Jmeter will be used in future performance tests
Authentication and Access Control
Authentication

The normal authentication mechanisms are supported.

1. Local userid and password
2. LDAP
3. Shibboleth
Access Control

There is interaction among

a. Object permissions
b. Tasks
c. Users permissions
Access Control – Objects & Permissions

a. An object consists of one or more data streams.

b. An object can belong to one or more “sites”

c. A repository consists of one or more sites.

d. Permissions can be assigned at the repository level (global permissions), at the site level, at the object level or at the data stream level.

e. Unassigned permissions are inherited from the higher level.
Access Control – Tasks

a. The system has more than 60 “tasks” that users can perform.

b. Examples of tasks are –
   a. View records or data streams
   b. Enter records or data streams
   c. Edit records or data streams
   d. Give permissions
   e. Run reports etc. etc.
Access Control – User Permissions

a. Users belong to one or more groups.

Groups can be created for functions like cataloging or acquisitions; or by enrollment in classes like English 101; or by habits or hobbies like Jogging.

b. Members of the group are given permissions or denied permissions to perform “tasks” from the task list.

c. Members of the group are given access to the whole repository, to one or more sites, to one or more objects or to one or more data streams.

d. Permissions are inherited (as previously mentioned)

e. Over-rides of inherited “permission denied” are supported.
VITAL 4.1 Authentication & Access Control

Access Control -- Summary

Permissions And Overrides

- Users
- Groups
- Tasks
- Repository
- Sites
- Objects
- Data Streams

And
Overrides
After today’s talk the student will be able to:

- Describe various benign lung conditions and their management.
- Recognize and manage patients with lung cancer.
- Make rational decisions faced with an incidental radiographic finding of a pulmonary nodule.
- Recognize various diagnostic measures used to differentiate pulmonary nodules.
Closing Remarks
Often asked: Why not use only open source?

- Open source is not really “free”.
- If you include the cost of personnel then the total cost to the institution for using “open source” is much higher.
- Open source solutions without commercial support are not reliable enough for dependable production needs.
- Commercial vendors have greater experience in offering services like:
  - implementation
  - support
  - documentation
  - quality assurance testing
  - training & additional required development
- Sustainability is compromised unless the institution has lots internal resources and time.

VTLS has adopted a solution based on “mixed source” getting (we believe) the best of both worlds.
Closing Thought

IF PEOPLE ARE WILLING TO GO AS FAR AS THEY CAN SEE,

THEY WILL BE ABLE TO SEE FARTHER WHEN THEY GET THERE.

Research and development is a never ending continuous process