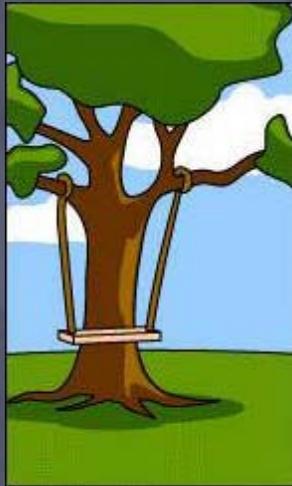


National Digital Newspaper Project Architecture

Presentation by George Schlukbier
NDNP Technical Project Manager
Office of Strategic Initiatives
Library of Congress



How the customer explained it



How the Project Leader understood it



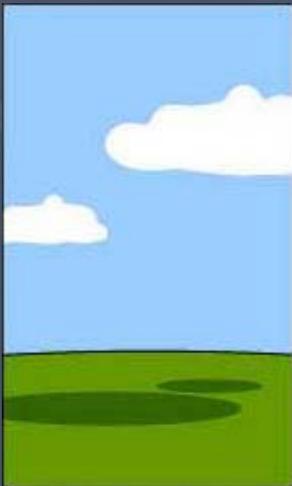
How the Analyst designed it



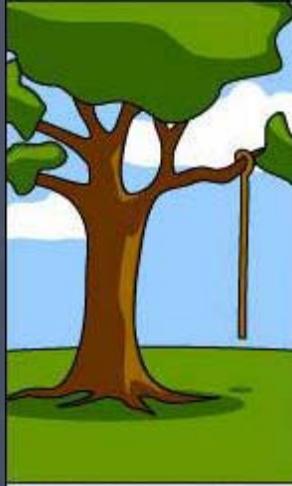
How the Programmer wrote it



How the Business Consultant described it



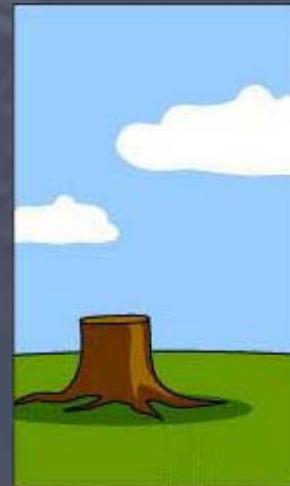
How the project was documented



What operations installed



How the customer was billed



How it was supported



What the customer really needed

Organization

NDNP Committee - Guidelines - Management - Use Cases

Development Team - PM George Schlukbier

Architecture/Design - Fedora condensed and extensively modified
(Corey Keith and Brian Vargas)

Module Development - Pre-ingest and Validation/Ingest /Repository
/Search/ Disseminators
(Justin Littman, Dave Woodward, Corey Keith, Brian
Vargas, Ray Murray, Stephen Engratt, Eric Ku, Waldo Ford)

Facade Disseminators
(Dave Woodward, Brian Vargas, Waldo Ford)

UI team - Use Cases/ Wire frames/ Test pages/ UI testing
(Web Services Team)

QA Team - Code testing, Script testing, UI testing
(George Sadusk, Jeremy Ossa)

Agile Development Process

Weekly meetings as a team. Follow the principles of Agile Development, Modular Design, Open Source, Standards and Best Practices, Project Management (PMI)

Modules

Pre-ingest- Validation-Ingest

Facade - Fedora with an abstraction facade used with cocoon to extend the usefulness of the current system.
- Black Box

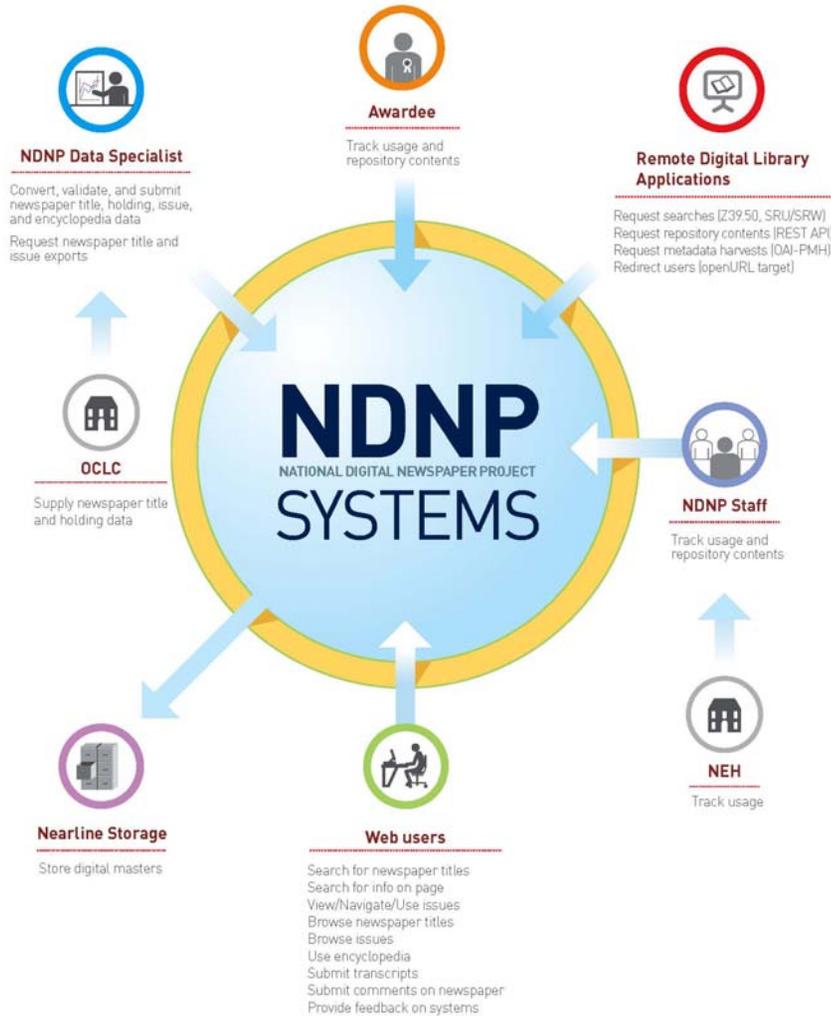
Search/Index - Lucene but plans to also experiment with K2 by Verity.

Dissemination Services- User Interface, Remote digital library applications
- Expose and develop API's.

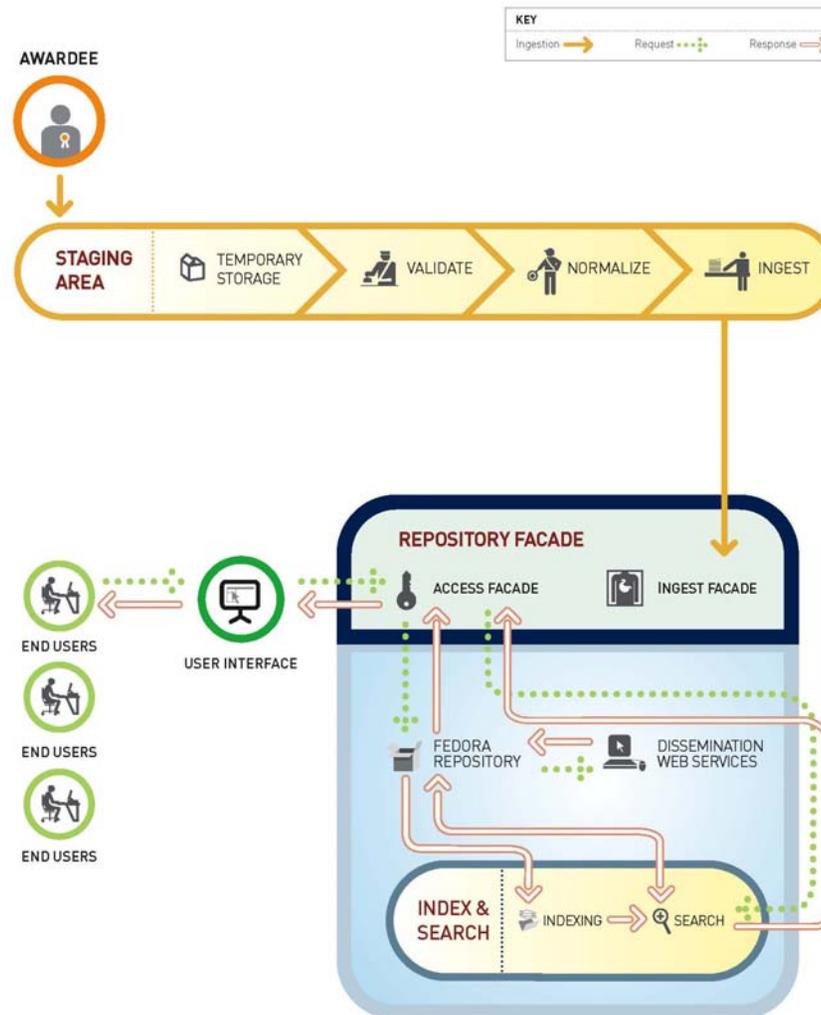
Quality Assurance

-Automated testing, scripts and use case testing
-UI testing

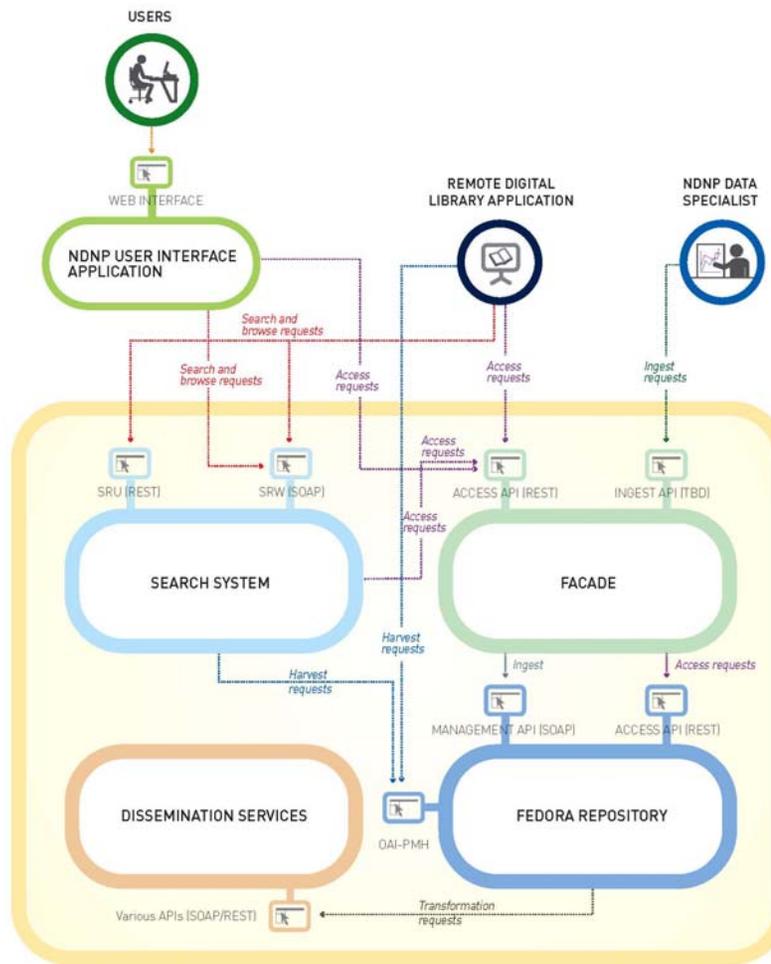
NDNP CONTEXT DIAGRAM



NDNP DATA FLOW DESIGN



NDNP INTERACTION DIAGRAM



Hardware

Build Environment - Location is a top secret so the server was named

“Peach” is a dual-processor Intel PIII Xeon 1.0 GHz machine with 4GB of RAM and two 28GB SCSI Ultra-160 drives.

Pilot Environment - Crystal City, VA

Six to ten servers - each with dual AMD processors (a combination of 32-bit and 64-bit processors) and 4GB of RAM.

Each machine will have 80GB of RAID0 SATA hard drive space for the Gentoo Linux OS, software packages, and fast temporary working space.

The total storage is targeted at 55 TB managed by a single NAS head.

Production Environment - Washington, DC.

TBD based on system analysis after Prototype is built.

Software

Peach is running the following software:



Gentoo Linux (<http://gentoo.org>), kept up-to-date with the latest package updates on an as-needed basis (<http://packages.gentoo.org>).

- Fedora repository 1.5 maybe 2.1
- Tomcat
- Cocoon
- Shell access
- MySQL database
- Apache web server
- Sun JDK 1.4
- Sun JDK 1.5

Development Software

Our Development Project Structure uses Subversion for version control

Build Tool is Maven
Cocoon 2.1.7
Java 1.5 or 5.0

Communications Software

Wiki
Open Office
Email/ IM.

Development Team

George Schlukbier - Technical Project Manager for NDNP

Justin Littman - Information Technology Specialist, Integration Management

Dave Woodward - Computer Specialist, Information Technology Services

Ray Murray - Newspaper Digital Conversion Team Leader, Collections and Services, Library Services

Corey Keith - Digital Project Coordinator, Network Development and Marc Standards Office, Library Services

Brian Vargas - Repository Architect, Technical Development

Steve Engratt - Information Technology Specialist