NIH Enterprise Architecture

Collaboration Architecture

Version 1.0

22 June 2004

Table of Contents

Table of Contents

(Continued)

List of Figures

List of Tables

1.0 Introduction

Collaboration is vital to achieving the NIH mission. Collaboration includes the technologies and tools that enable NIH users to access vital information resources, share information, and work and communicate effectively and efficiently with peers, customers, and the public. These technologies and tools range from web browsing and searching capabilities to large scale document sharing and workflow management.

In 2003, a Collaboration domain team, Collaboration I, focused on enterprise messaging, email, and Active Directory guidelines and standards. This document establishes collaboration architecture standards and guidelines across NIH and specifies common components that have been developed and agreed upon by two other separate cross-Institute and Center (IC) domain teams and can be implemented enterprise wide. The first domain team, Collaboration II, concentrated on search engines, web browsers, and enterprise reporting. The second domain team, Collaboration III, concentrated on web content management, document management, and workflow tools.

The Collaboration domain teams clarified the technologies in use at NIH, the current state or baseline architecture, and developed the future state direction. Additionally, the domain teams developed “patterns” to show how these products should work together within the enterprise. Throughout the domain team meetings, NIH technologists considered how to leverage current technologies while planning how to expand NIH’s Collaboration capabilities. This allows NIH to keep its existing investments where needed and plan for innovations, trends, and anticipated business needs to provide the maximum business benefits to its user base.

1.1 Collaboration Domain Teams

This report comprises the compilation of findings and recommendations derived from the joint NIH-Gartner Enterprise Architecture project team. Two teams with representation from various Institutes and Centers (ICs) and the Center for Information Technology (CIT) worked together for four weeks each to develop the Collaboration architecture patterns and bricks that are presented in this report. The Collaboration II team, which focused on search engines, web browsers, and enterprise reporting, was made up of 16 of subject matter experts. The Collaboration III team, which focused on web content management, workflow tools, and document management, was made up of 17 subject matter experts. The IC representatives that contributed to this effort are shown in Table 1:

Table 1. Collaboration Domain Teams

Collaboration II	Collaboration III
Robert Bosworth, NIAID	James Blagaich, CSR
Toni Calzone, NIAAA	Toni Calzone, NIAAA
Peter Carothers, NIDCD	Eric Charles, NIAID
Blaise Czekalski, NCI	Megan Columbus, OD
Vernita Dawkins, NIGMS	James Del Priore, CIT
James Del Priore, CIT	Sandy Desautels, CIT
Susan Ensley, NCRR	Starr Kline, OD
Laura Hall, NHLBI	Marguerite Lewis, NIDA
Matthew Lightner, OD	Jose Lopez, NIGMS
Terry Luedtke, NLM	Terry Luedtke, NLM
Judy Mahaffey, CIT	Mary Ellen O'Meara, NCI
John Price, CIT	Ed Ost, NIDCD
Sheila Taylor, NIDCR	Mark Perry, OD
Jackie Sanders, NIMH	Brenda Sonneveldt, NHLBI
Doug Sur, CC	Robert Steele, NIMH
Virginia Vinton, CIT	Doug Sur, CC
	Rayseen Woodland, NCRR

1.2 Scope

The two Collaboration domain teams focused on establishing principles, patterns, and bricks for all of NIH. When appropriate, the suggested architecture components were divided by their relative scale of features and costs (i.e., lightweight vs. heavyweight solutions).

The Collaboration II Domain Team looked at web browsers, search engines, and enterprise reporting tools. Web browsers are accessible by all NIH users and work on all platforms supported by the NIH ICs. Search engines can search and index intranet and Internet web sites for HTML (hypertext markup language) files and native documents. Enterprise reporting (business intelligence) solutions, at a minimum, provide the ability to produce standard reports and can scale to include advanced features such as ad hoc querying and analysis.

The Collaboration III Domain Team looked at web content management, document management, and workflow tools. Web content management is used for web page creation and publication throughout NIH. Document management technologies provide simple file structure management of documents within the ICs or across the enterprise. Workflow tools provide design and automation of work processes within ICs and across the enterprise. During team discussions, the domain team also addressed digital signatures and imaging as they relate to the other collaborative toolsets discussed in

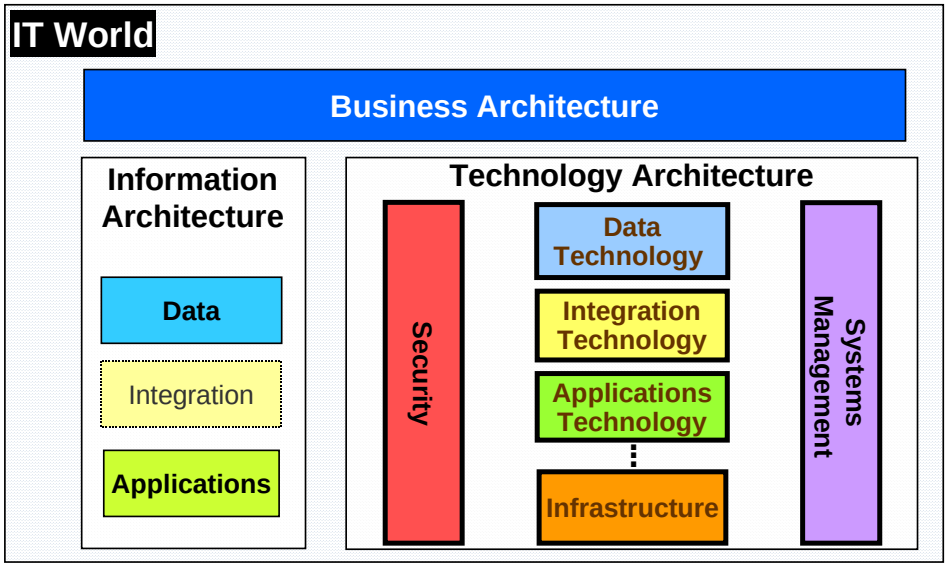
Collaboration III. Throughout the Collaboration III discussions, it was apparent that several technologies categorized had functionality across multiple collaborative domain areas. When technologies had functionality across multiple areas, the team categorized them by primary function.

1.3 Collaboration Domain in the NIH Enterprise Architecture Framework

The *NIH Enterprise Architecture Framework* and *NIH Enterprise Architecture Matrix* are based on the Federal Enterprise Architecture Framework (FEAF) and the FEAF Matrix.[1]

The NIH EA Framework recognizes three distinct component architectures: the Business Architecture, Information Architecture and Technical Architecture. The NIH EA Framework is illustrated in Figure 1.

Figure 1. NIH Enterprise Architecture Framework



The Collaboration Domain is part of the Technology Architecture within the NIH EA Framework.

The NIH EA Matrix provides five potential *perspectives* or views of the architecture, at increasing levels of detail. The NIH EA Matrix is shown in Table 2.

[1] Level IV of the FEAF, derived from the Zachman Framework

Table 2. NIH Enterprise Architecture Matrix

	Data Architecture	Application Architecture	Technology Architecture
Planner Perspective	List of Enterprise Business Objects	List of Business Processes, + multi-enterprise processes.	List of Business Locations and Business Partners
Owner Perspective	Semantic Model	Business Process Models (including multi-enterprise)	Business Logistics System + multi-enterprise logistics
Designer Perspective	Logical design patterns; Use enterprise business objects	Logical design patterns, by style	Integration technology for enterprise systems
Builder Perspective	Physical design patterns; Use shared database if applicable	Logical design patterns, by style	Physical design patterns; Use bricks from TRM or request a waiver. TRM includes security, NIH network, and other infrastructure
Subcontractor Perspective	Project scope	Use common services or APIs, if defined	

This architecture report focuses on the Planner, Owner, and Designer views (reference Table 3).

Table 3. Collaboration Alignment With the NIH Enterprise Architecture Matrix

	Data	Applications	Technology
Planner View			■ Scope (Section 1.2) dictates the type of Collaboration tools in use ■ Principles (Section 1.4) provide high-level guidance and the fundamentals of Collaboration technologies at NIH
Owner View			■ Patterns (Section 2.0) depict the interconnectivity of Collaboration and other components
Designer View			■ Bricks (Section 3.0) depict the building blocks that are referenced by the Collaboration patterns

1.4 Principles

The Collaboration domain teams were chartered with establishing principles to guide the development and usage of the architecture standards within the Collaboration area of focus. The Collaboration domain teams started with the principles established in the previous Collaboration effort from May 2003. Throughout the four meeting efforts, the domain teams affirmed the principles defined by the Collaboration I domain team and added two new ones. The principles identified here are to be universally accepted by all of NIH and should withstand changes and trends in Collaboration markets and technologies. In addition, they should remain relevant with changes in NIH and Federal Government policies, programs, and procedures. Table 3 depicts the principles and the rationales that explain how NIH and the public benefit from the principle.

Table 4. Collaboration Architecture Principles

Principle	Rationale
Level of Security: Security is fundamental to collaboration systems. The level of security will be based upon the sensitivity of data generated and maintained and will meet or exceed the directives of departments or agencies with oversight.	With the introduction of stricter government mandates, such as HIPAA, and the increasing digitization of sensitive information, any Collaboration system must be grounded in an appropriate level of security, balancing the sensitivity of the data with ease of use.
Investment Leverage: Collaboration systems will leverage existing and future enterprise software, management systems, infrastructure and standards.	As a practical matter, NIH has invested heavily in a number of technologies, infrastructures and standards. Therefore, any system that is expected to be widely used will need to leverage as much investment as possible.
Integration: Collaboration systems will integrate with enterprise application and directory structures and support the import and export of information, facilitating sharing at NIH and external organizations as needed to accomplish the NIH mission.	Any Collaboration system used will need to interface with the appropriate existing systems at NIH and to external organizations to better facilitate sharing and communication.
Common Access: Collaboration systems will provide common access for authorized personnel and offer full access from standard platforms. Common access for authorized personnel includes: ❑ Access from supported platforms and user interfaces ❑ Access within NIH ❑ Access via remote connectivity ❑ Access by other entities to NIH-wide collaboration systems	Because of NIH's global character, any successful enterprise collaboration system must address the requirements of NIH's user communities, especially in regard to remote access, access within NIH, access by other entities, and platform support.
Best Practices and Open Standards: Collaborative systems will be based upon industry best practices and open standards.	In order to effectively manage risk, NIH requires that any Collaboration system employed be based upon open standards, when possible, and best practices.
Reliability and Availability: Collaborative systems will be highly reliable, striving for 24x7 service.	In order to provide maximum value to NIH, Collaboration systems must be available round the clock to its users.
Resources: Collaboration systems will support the effective use of NIH resources, enabling them to meet user needs while allowing minimal duplication of effort and technologies between ICs.	Collaboration systems need to keep in mind the efficient use of NIH resources and should eliminate redundancy of efforts across the organization.
User Rights: Collaboration systems will allow a user to control accessibility and define managerial privileges as appropriate to one's role.	NIH 's Collaboration systems must provide the capability for controlling accessibility based on user privileges.

1.5 Summary of Key Decisions

- New versions of Internet Explorer (IE) and Safari are considered strategic options for web browsers.
- Search engines considered in this architecture concentrate only on search of websites.
  - Separate tactical options for searching intranet and internet sites have been established
  - Strategic options to consider include Google Search Appliance and Verity Ultraseek
  - Enterprise search technology has been considered a technology to track in the future.
- Enterprise reporting includes the following strategic options:
  - Crystal Enterprise for lightweight web (reporting only),
  - Adobe Acrobat or Microsoft Office for lightweight non-web (reporting only), and
  - An emerging business intelligence product for heavyweight (reporting, querying, and analysis).
- Documentum (Application Xtender) is considered the strategic tool for document management.
- Microsoft CMS and Interwoven Teamsite are considered the strategic tools for web content management.
- Handysoft BizFlow, OpenText LiveLink, and NMS eFlow are considered strategic options for workflow tools.
- Digital signature and document imaging bricks have been defined in the Collaboration architecture to set a provisional direction.
  - Digital signature and document imaging baselines and tactical/strategic toolsets have yet to be fully determined.

1.6 Benefits of the Collaboration Architecture

In addition to the general benefits of Enterprise Architecture, stated elsewhere, this domain architecture provides the following specific benefits:

- Greater knowledge sharing across the enterprise achieved via Collaboration systems will improve productivity and efficiency within the NIH workplace.
- Collaboration technologies and the potential integration of these technologies can complete gaps within NIH processes.

- Efficient sharing of data and information across NIH will reduce duplication and redundancy of data
- Greater insight into information through document sharing, knowledge sharing, and data reporting can lead to better business decisions at NIH.
- Leveraging fewer Collaboration solutions across NIH can improve the productivity of IT support staffs.
- Potential cost savings can accrue if volume purchase agreements are deemed to be advantageous.

2.0 Collaboration Design Patterns

Design patterns may be *logical* or *physical.* Logical design patterns do not specify specific technology platforms, products or brand names. A logical design pattern may be implemented by one or more related physical design patterns. Patterns provide design guidance to implementation teams and can occur in one domain or span multiple domains. Patterns provide a reference model ("blueprint") for the technology elements that can be combined to solve a specific problem.[1]

2.1 Pattern 1: Web Information Access

2.1.1 Description

The Web Information Access pattern provides a logical representation of accessing NIH information resources available on the web. Specifically, the pattern shows the interconnectivity of browsing and searching components and capabilities for both intranet (including portal) and Internet NIH websites.

2.1.2 Web Information Access Solution

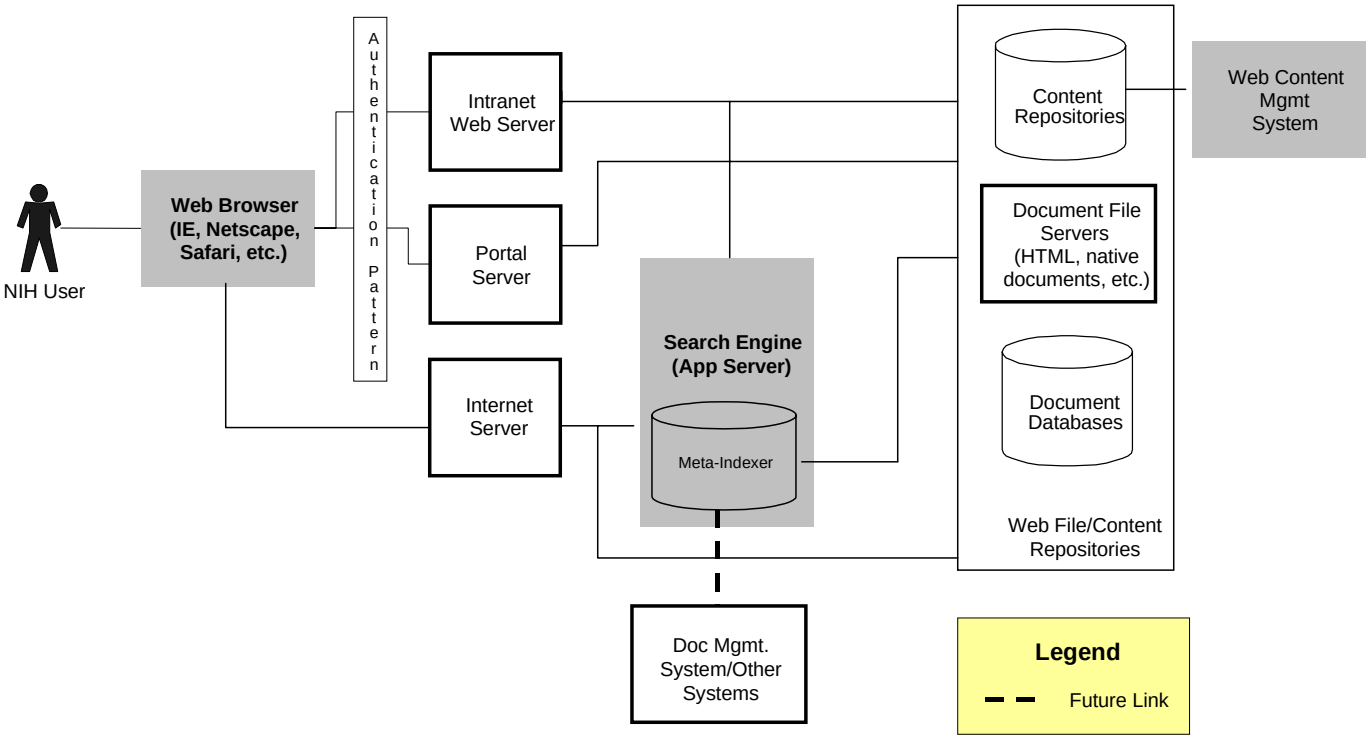
This solution, shown in Figure 3, provides access to web information resources via a desktop web browser. The applications highlighted in gray represent the Collaboration components or "bricks" addressed in the domain teams. An NIH user can access NIH intranet, portal, and Internet websites utilizing a web browser directly or can search these sites for key information. Intranet and Internet web servers, along with portal applications, can connect directly to:

- Document file servers storing HTML files and native file formats,
- Content repositories connected to a web content management system, and
- Other document databases to serve up web content to the end user.

For search-based information retrieval, the search engine application server can be accessed directly by an NIH user through the desktop web browser via web servers. The meta-indexer on the search engine platform indexes files located in these file servers, content repositories, and document databases. Access to the intranet web server and portal applications requires authorization as specified in the Security Architecture. Optionally, a robust enterprise-wide search engine application can index documents and other content (i.e., multimedia) found across the enterprise by connecting to other information systems.

[1] Technology represented by "bricks" or specific technologies inside a brick.

Figure 2. Logical Design Pattern for Web Information Access



2.1.3 Benefits

- This solution enables NIH to access web content and web-based applications via the desktop.
- This solution provides simple access to both web content and search results of web information required for NIH end users.
- This solution allows for connectivity from the search engine to other enterprise information resources including business information systems, document management systems, and other file repositories.
- This solution provides remote accessibility to NIH web sites and web-enabled applications.

2.1.4 Limitations

- This solution does not yet show the linkage to document management systems or access to external content services.

2.2 Pattern 2: Data Reporting Pattern

2.2.1 Description

The Data Reporting pattern provides a logical representation of the components necessary to access and create reports based on NIH data. Specifically, the pattern shows the interconnectivity of web browsers, non-web reporting tools, web-based enterprise reporting tools, and NIH data repositories.

2.2.2 Data Reporting Solution

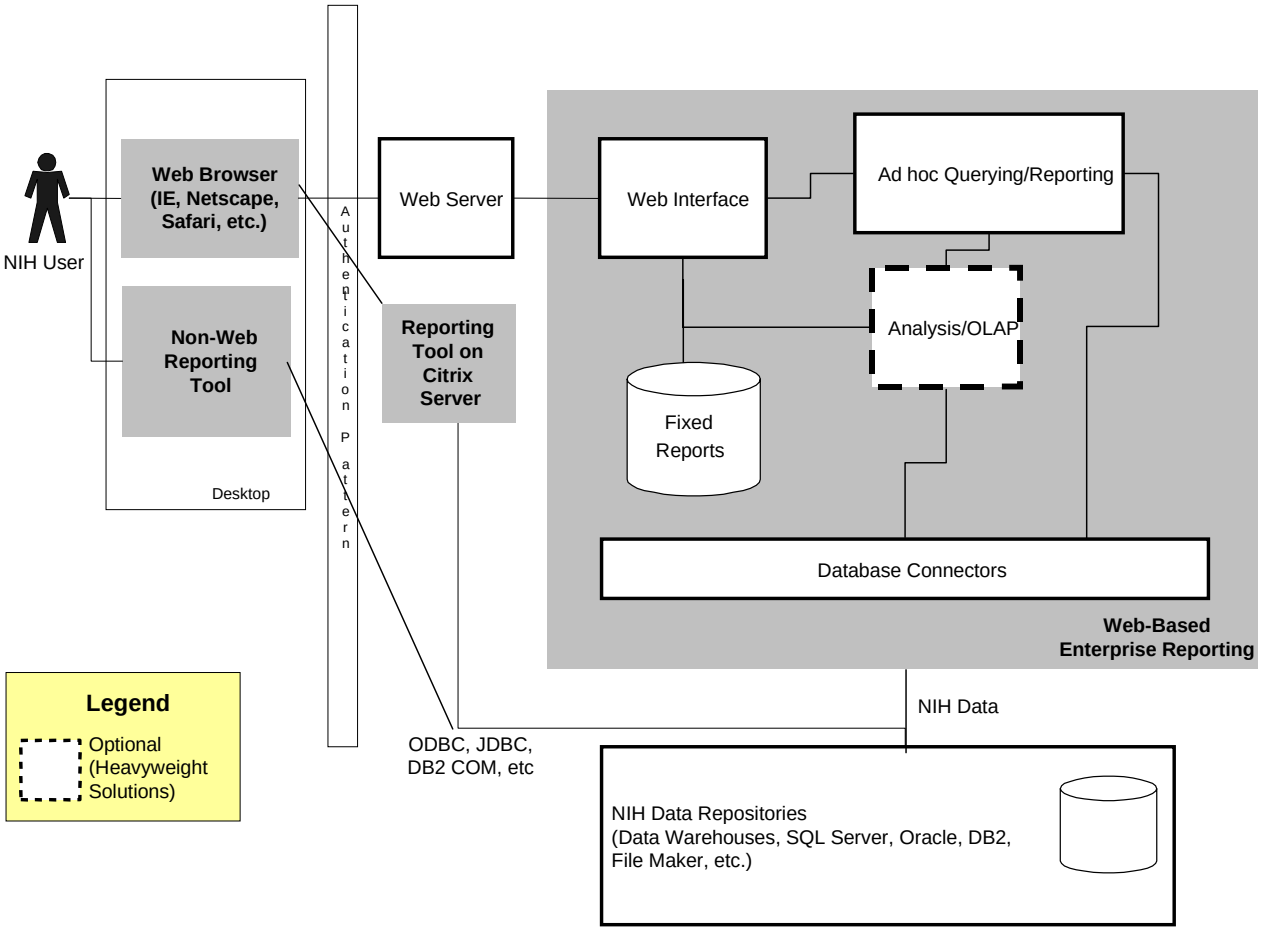
This solution, shown in Figure 4, presents multiple options for data reporting at NIH. An NIH user can access NIH data repositories including individual databases (Oracle, SQL Server, DB2, etc.) and data warehouses such as n-Vision and report from them through either a non-web reporting tool on a desktop, a reporting tool running on a Citrix server, or a web-based enterprise reporting tool (highlighted in gray). Access by NIH users to either the non-web reporting application, the reporting tool running on a Citrix server, or the web-based enterprise reporting tool requires authentication as specified within the Security Architecture.

The non web-based reporting tool uses a database connection mechanism such as ODBC, OLEDB, or DB2 COM to directly access NIH data and create or view reports. Reporting tools running on Citrix servers require access through a web browser and directly connect to the NIH data repositories using the same mechanisms.

A web-based enterprise reporting solution can be accessed through a standard web browser and web server. A typical web-based enterprise reporting platform includes a canned reports repository and a real-time reporting and ad-hoc querying component that accesses data directly from the NIH data repositories through database connectors.

Heavyweight enterprise reporting tools also include the ability to do multi-dimensional online analytical processing (OLAP) from data in the NIH data repositories. Database connection mechanisms used by the enterprise reporting tools include ODBC, OLEDB, DB2 COM, web services, etc.

Figure 3. Logical Design Pattern for Data Reporting



2.2.3 Benefits

- This solution provides multiple options for reporting and accessing data based on scale and platform.
- This solution depicts both lightweight reporting and optional advanced querying/analysis capabilities for the enterprise.

2.2.4 Limitations

- This solution does not address validation of the quality of data accessed and reported.
- This solution does not address the data aggregation needs or the structures required to support the reporting styles.

2.3 Pattern 3: Web Content Management Pattern

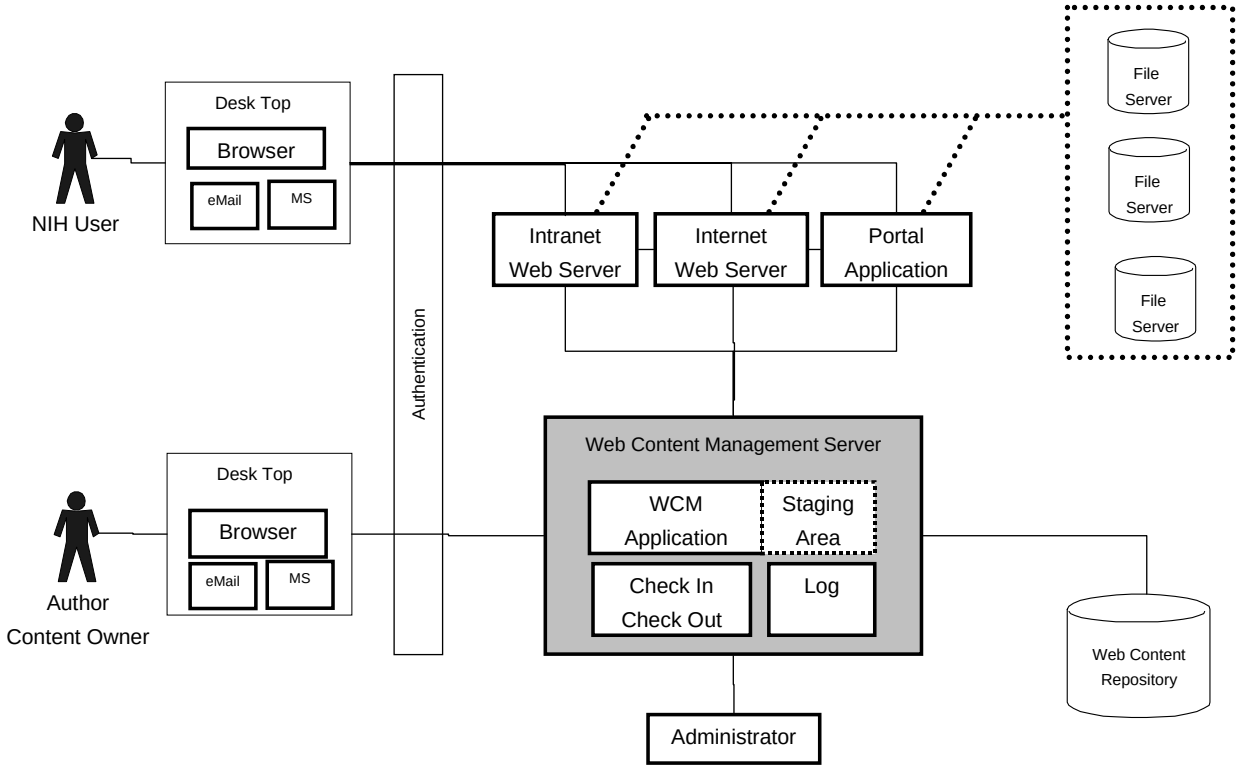
2.3.1 Description

Web Content Management applications provide the ability to author, edit, and publish information to be displayed via an established NIH inward or outward facing website. Information may include text, audio, graphics or other digitized information in file format.

2.3.2 Web Content Management Solution

The solution, shown in Figure 5, provides a logical depiction of how a web content management system can be accessed and utilized within NIH. There are two primary user groups for this solution: the authors and the readers. The primary user is the web author or content manager. This individual could be an individual within NIH simply publishing text to a static website or an individual within the graphics division building an outward interactive and collaborative. The secondary user of this solution is the reader. The reader can access the web site through an intranet, Internet, or web portal. The reader's access is typically limited to viewing, but in some cases the web content might include and offer add-on capabilities to establish collaboration rooms or the ability to comment on published web content. The web content author or content owner can access the web content application through their desktop or a dedicated terminal by authenticating their permission to add, edit, or remove content. Edit access can be restricted to specific areas of content to protect content from unauthorized modification. Within the web content management application, the author can stage content, check in and check out content, and publish as necessary. Each web content management application has varying levels of scalability and capability that vary from simple static publishing, changing Microsoft Word documents to HTML or Adobe format, to publishing XML (extensible markup language). One common goal that web content management applications strive toward is allowing individuals outside of the typical information technology organization to manage and publish their own data.

Figure 4. Logical Design Pattern for Web Content Management



2.3.3 Benefits

- Allows an organization to publish information on inward or outward facing websites to enable knowledge sharing and collaboration.
- Typically, the user interface for the author or content owner is simple and intuitive, which allows for non-technical personnel to manage and publish their own information.
- Applications allow automatic update or removal of data once the information value has expired.

2.3.4 Limitations

- Some web content management applications have limited capability to publish XML and other file formats. Add on capabilities are available, but will increase overall procurement and maintenance cost.

2.4 Pattern 4: Workflow Pattern

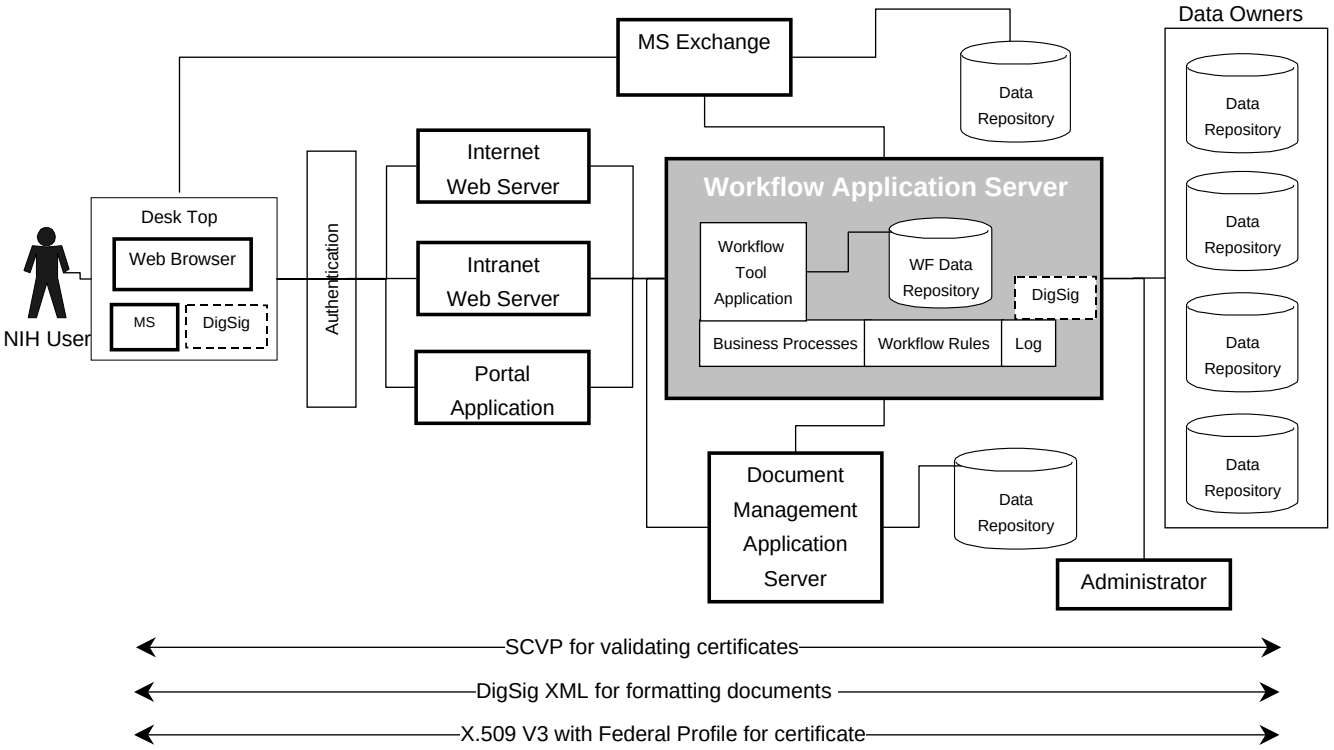
2.4.1 Description

Workflow tools allow NIH to exchange and to collaborate on documents through a defined, organized process. Applications typically allow the user to construct standard business process rules including approval, quality assurance, review, and/or information access.

2.4.2 Workflow Solution

The solution, shown in Figure 6, provides the user with the ability to access the workflow tool applications through a web-enabled interface or through the organization's email system. Documents needed by the workflow application can be created in Microsoft Office. Access is controlled through the mechanisms specified in the Security Architecture. This solution provides for data content to be stored on external repositories separate from the workflow application, allowing NIH organizations to house and to manage their own data. Digital signature capabilities are typically provided through add-ons and are represented in a general format. Additional detail on digital signature solutions is provided in Section 3.7.

Figure 5. Logical Design Pattern for Workflow



2.4.3 Benefits

- Allows data ownership and rights to be established by the data owners.
- Provides for an interface with document management and records management applications.
- Users can access the application through a web interface and/or receive notifications through their desktop email system.
- If appropriate authentication and security is established, users can access through a remote server.
- Digital signature can be achieved with add-on capabilities.

2.4.4 Limitations

- Integration of workflow tools with email systems, document management systems and records management are not represented in detail as the interface mechanisms depend on the products chosen.

2.5 Pattern 5: Document Management Pattern

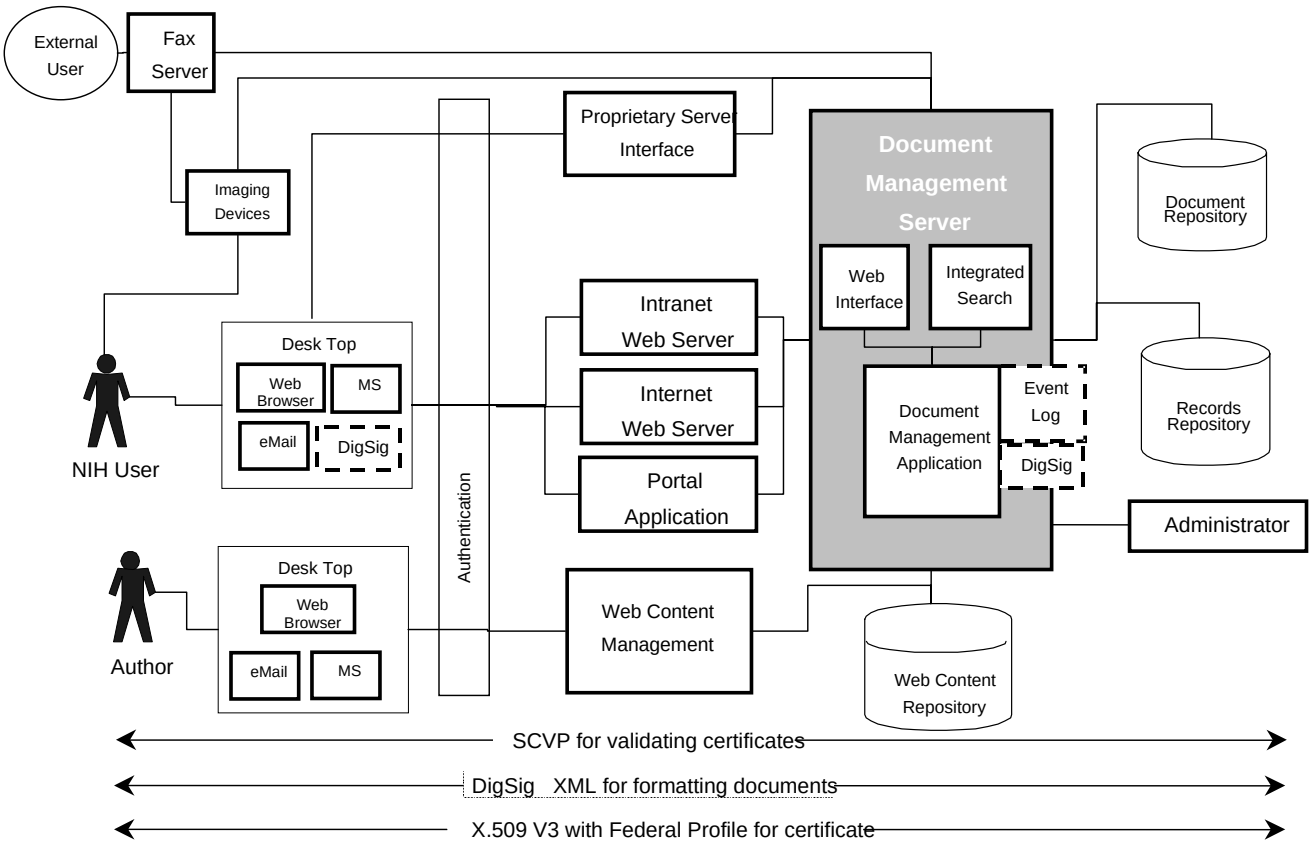
2.5.1 Description

Document management systems provide the capability for NIH to store documents logically within a standard taxonomy. Access and version control of documents can be managed effectively therefore enabling an organization to avoid multiple instances of duplicate data.

2.5.2 Document Management Solution

As shown in Figure 6, the document management solution allows the user to deposit documents through multiple interfaces. Most users will access the document management system through a typical desktop configuration via a web interface or an existing proprietary application. Access can also be obtained through imaging devices or through the organization's email system, which archives emails as historical artifacts. Search capabilities are typically built into the functionality of the document management system. Searches can be driven by a keyword search or through other designated parameters. The web content management interface depicts how documents within the document management system are published onto a website. Access through this web content management interface would be independent of the access directly to a document management system, but defined accessibility and authentication would have to be established. Digital signature capability is shown to provide a high level understanding of how a digital signature add-on might interface with a document management system. Digital signatures are addressed in more detail in Section 3.7.

Figure 6. Logical Design Pattern for Document Management



2.5.3 Benefits

- Enables establishment of a taxonomy to ensure appropriate document management.
- Reduces the likelihood of maintaining duplicate data.
- Allows for documents to be accessed through multiple channels within the organization.

2.5.4 Limitations

- Depicts a standalone tool for document management only. Other collaborative tools can provide basic and, in some cases, mature document management capability.

3.0 Collaboration Bricks

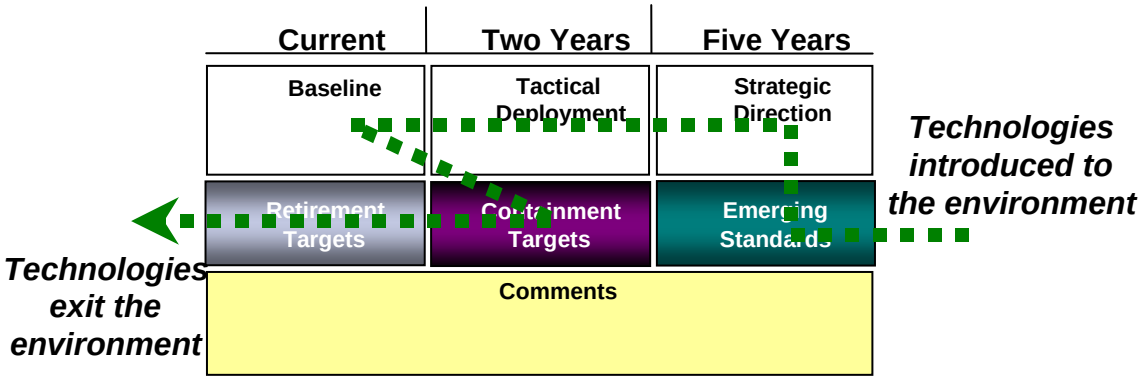
In the Technical Reference Model (TRM), baseline and planned technology choices for elements meet in a chart called a "brick." Bricks represent the physical building blocks of the enterprise IT systems – they identify specific technologies used to implement solutions. Bricks document both NIH's current ("as-is") environment and future ("to-be" or target) states. The planning horizon is five years.

Each brick captures:

- A description of the technology and its role,
- Specific implications, dependencies, and deployment and management strategies, and
- Technology elements, categorized.

A brick template is shown below:

Figure 7. Technology Planning "Brick"



The technology choices for architectural elements are categorized as follows:

- **Baseline** technologies include current technology and/or process element(s) in use.
- **Tactical** technologies are recommended for use in the near or tactical time frames (next two years). Currently available products needed to meet existing needs are identified here.
- **Strategic** technologies provide strategic advantage and might be used in the future. Usually, marketplace leaders are identified here, as they are likely to provide better benefits and meet the anticipated needs of the business.
- **Retirement** technology and/or process elements targeted for de-investment during the architecture planning horizon (five years).
- **Containment** includes technology and/or process elements targeted for limited (maintenance or current commitment) investment.

- **Emerging** technology and/or process elements are to be evaluated for future use based on technology availability and business need. These technologies may not be new to the marketplace, but are simply not yet in use at NIH. In this case, the products may be a fit for emerging needs at NIH.

The Collaboration architecture is comprised of the following bricks:

- Web Browsers
- Search Engines
- Enterprise Reporting Tools
- Web Content Management
- Document Management
- Workflow Tools
- Digital Signature
- Document Imaging

3.1 Brick 1: Web Browser

Web browsers are programs that "read" hypertext and display it as formatted text and images. Browsers allow users to view the contents of a site and navigate from one site to another.

The NIH has a variety of browsers in use today throughout its ICs, ranging from Microsoft Internet Explorer to Opera and Safari as indicated in Table 5. It is important to note that two key criteria were utilized when populating the bricks: application dependencies and Section 508 compliance to maintain accessibility including the usage of screen readers, such as Jaws, for browsing. Applications requiring certain versions of web browsers for access were considered and those browsers were kept in containment.

Table 5. Web Browser Brick

Baseline Environment (Today)	Tactical Deployment (0-2 years)	Strategic (2-5 year)
■ PC/Unix/MAC: Internet Explorer 4.0x, 4.5, 5.0x, 5.1x, 5.2x, 5.5, 6.0 ■ PC/Unix: Netscape 4.0x, 4.61, 4.7x, 5.0, 6.0, 6.1, 7.0, 7.1 ■ PC/Unix: Opera/Opera 7.1 ■ PC/Unix: Mozilla 1.31, 1.4, 1.6 ■ PC/MAC/Unix: Firefox ■ MAC: Safari/Safari 1.2/Safari 2.x	■ PC/Unix/MAC: > Internet Explorer 6.X ■ PC/Unix: > Mozilla 1.6 ■ MAC: > Safari 2.X ■ Unix: Netscape 7.X	■ PC: Internet Explorer (Longhorn Embedded) ■ MAC: Safari
Retirement (Technology to eliminate)	**Containment (No new deployments)**	**Emerging (Technology to track)**
■ PC/Unix/MAC: <= Internet Explorer 5.X ■ PC/Unix: <= Mozilla 1.4 ■ MAC: <= Safari 1.X ■ Unix: <= Netscape 6.X	■ PC/MAC: Internet Explorer 5.X ■ PC/Unix: Netscape 4.X ■ PC/Unix: Opera ■ PC/Unix/MAC: Firefox	■ New Browsers for the PC, MAC, and Unix platform with enhanced speed and capabilities, such as Internet Explorer (Longhorn), Mozilla, and Opera ■ MAC-based browsers that may be supported in the future by PCs, such as Safari

Comments
■ '<' or '<=' indicate versions less than or less than or equal to the one specified; '>' or '>=' indicate versions greater than or greater than or equal to the one specified. ■ The latest versions of browsers that are still heavily in use within NIH have been designated Tactical. ■ Older versions of web browsers that do not have enterprise application dependencies are in Retirement. ■ Older versions of web browsers that have enterprise application dependencies are in Containment. ■ Tactical and Strategic products were selected to leverage NIH's investment in products that are a proven fit for NIH's known future needs. Leveraging baseline products in the future will minimize the operations, maintenance, support and training costs of new products. ■ Some baseline products have been designated Retirement and Containment. These products are either not as widely or successfully deployed at NIH, or they do not provide as much functionality, value, or Total Cost of Ownership as the selected Tactical and Strategic products.

3.2 Brick 2: Search Engines

A search engine, in the context of NIH and this Collaboration domain team, includes a "robot" or "crawler" that goes to every page or representative pages on a Web site, or the whole Web, and creates an index; as well as the program that receives a search request, compares it to the entries in the index, and returns results to the end user.

It is important to note that the brick shown in Table 6 only addresses search capabilities for web pages. However, several vendors and technologies that can be used enterprise-wide (i.e., search information systems within an organization) will be tracked (in Emerging) moving forward.

Table 6. Search Engines Brick

Baseline Environment (Today)	Tactical Deployment (0-2 years)	Strategic (2-5 year)
■ Copernic ■ Google Search Appliance ■ HTDig 3.1.6 (Shareware) ■ Recommind Mindserver ■ MS Indexing Service ■ Verity Info Server 3.6/3.61 ■ Convera XX ■ Phantom ■ Swish-E ■ Webcrawler ■ Verity Ultraseek 5.04 ■ MS Search Engine ■ Convera Excalibur Retrievalware 6.9/8	■ Google Search Appliance (internet) ■ >= HTDig 3.1.6 (Shareware)* ■ Verity Ultraseek 5.04/5.1 (intranet)	■ Google Search Appliance ■ Verity Ultraseek >= 5.1

Retirement (Technology to eliminate)	Containment (No new deployments)	Emerging (Technology to track)
■ Verity Info Server 3.6/3.61	■ Convera XX ■ Phantom ■ Swish-E ■ Webcrawler ■ Recommind Mindserver (Research Heavy) ■ Convera Excalibur Retrievalware 6.9/8	■ Shareware search engines, such as HTDig ■ Multimedia search technology ■ Enterprise search products, such as Convera (A&V) , FAST, Autonomy (A&V), VerityK2, and Endeca ■ Embedded search capabilities, such as those that come with Netscape or ColdFusion ■ Open Source search engines, such as Apache Lucene ■ Additional search engine products such as new Verity products, ISIS, dtSearch, and Thunderstone
Comments		
■ '<' or '<=' indicate versions less than or less than or equal to the one specified; '>' or '>=' indicate versions greater than or greater than or equal to the one specified. ■ Due to security concerns, separate search engines for intranet and Internet sites were designated in Tactical. ■ Tactical and Strategic products were selected to leverage NIH's investment in products that are a proven fit for NIH's known future needs. Leveraging baseline products in the future will minimize the operations, maintenance, support and training costs of new products. ■ Some baseline products have been designated Retirement and Containment. These products are either not as widely or successfully deployed at NIH, or they do not provide as much functionality, value, or Total Cost of Ownership as the selected Tactical and Strategic products.		

3.3 Brick 3: Enterprise Reporting Tools

As defined by the NIH domain team, an enterprise reporting tool is one that allows organizations, including federal agencies, to gain a better understanding of their business by putting critical information in the hands of all those who need it – employees, managers, partners, and customers. An enterprise reporting tool must be configurable to meet the needs of its user base, capable of accessing information assets in the enterprise based on access rights, and must support wide-scale deployment.

Enterprise reporting tools as defined above generally contain several types of functionality, from generic reporting capabilities (including report design) to robust ad hoc querying and online analytical processing (OLAP) capabilities. This second layer of capabilities is considered business intelligence (BI). There is an overall convergence of

these two layers, robust reporting and business intelligence, within the marketplace today. This makes it possible to select a single vendor for the full suite of capabilities.

The business intelligence marketplace can further be delineated into enterprise BI suites (EBIS) and BI platforms. Enterprise business intelligence suites (EBISs) are the successors to the basic query and reporting tools. EBISs emerged from the previously disjointed query, reporting and OLAP offerings while addressing criteria such as product scalability, usability and manageability. BI platforms offer complete sets of tools for the creation, deployment, support, and maintenance of BI applications. They combine database access capabilities, Structured Query Language (SQL), OLAP data manipulation, modeling functions (what-if analysis), statistical analysis, and graphical presentation of results (charting) to create data-rich applications, with custom end-user interfaces, organized around specific business problems.

The NIH domain team considered “lightweight” web-based and non web-based reporting solutions for standard, fixed reporting. Non web-based reporting solutions consist of applications that exist on a user’s desktop and don’t require a web browser for accessibility. In addition, the team looked at “heavyweight” solutions for advanced ad hoc querying/reporting and analysis (BI). The brick in Table 7 also includes options for statistical reporting, as its use is widespread within the scientific research community.

A required capability for the solutions listed in the brick below is compliance with Section 508 requirements. Several vendors listed in the brick are not Section 508 compliant as of March 2004. It is important to note that Vendor solutions listed in the brick have to be Section 508 compliant by October 1, 2004, and must have completed a VPAF form verifying such. This brick assumes that all listed vendors will be Section 508 compliant by that date.

Table 7. Enterprise Reporting Tools Brick

Baseline Environment (Today)	Tactical Deployment (0-2 years)	Strategic (2-5 year)
■ Actuate Server 6/7i ■ Microstrategy 7I ■ InfoMaker (PowerBuilder) ■ Crystal Reports 8 ■ Crystal Enterprise 8.5, 9.5 ■ InetSoft Style Report ■ MS Office 2000, 2002, 2003 ■ Neon Shadow ■ Oracle 9I App Server ■ PDF.Lib ■ Cognos PowerPlay ■ SAS (Strong Stat Analysis) ■ SPSS (Stat Analysis) ■ Adobe Acrobat 6.0 ■ Cognos (Client Server) ■ Hummingbird BI Query (Report/Query Focused) ■ Filemaker (MAC)	Lightweight (Reporting) ■ MS Office 2002 / 2003 (Non-Web) ■ Adobe Acrobat 6.0 (Non-Web) ■ Crystal Enterprise 9.5 (Web) ■ Filemaker (MAC) Heavyweight (Reporting, Ad-Hoc Querying, Analysis - BI) ■ Business Objects or Microstrategy 7I Statistical Reporting ■ SAS	Lightweight (Reporting) ■ Crystal Enterprise 10+ (Web) ■ MS Office *OR* Adobe Acrobat (Non-Web) Heavyweight (Reporting, Ad-Hoc Querying, Analysis – BI) ■ TBD (Emerging)
Retirement (Technology to eliminate)	**Containment (No new deployments)**	**Emerging (Technology to track)**
■ None Identified	■ Actuate Server 6/7i ■ MS Office 2000 ■ Crystal Reports 8 ■ Neon Shadow ■ PDF.Lib ■ Hummingbird BI Query ■ Cognos (C/S) ■ SPSS	■ Merging vendor capabilities, such as Crystal Reports and Business Objects ■ Future reporting/business intelligence vendor products, such as Microstrategy Report Svcs/BII, Cognos ReportNet/Powerplay, and MS SQL Server Reporting Services ■ Future PDF-based products, such as Adobe ■ Lightweight tools with strong 508 capabilities, such as Infragistics Net Advantage ■ BI Platforms, such as ProClarity and SpotFire ■ Lightweight embedded reporting tools, such as InetSoft Style Report

Comments
■ Vendors listed above must be Section 508 compliant by October 1, 2004 and must have completed a valid VPAT (Voluntary Product Accessibility Template - http://www.itic.org/policy/508/Sec508.html) form. ■ Vendors with outstanding Section 508 compliance issues have been placed in Containment. ■ Products with industry leading BI capabilities have been included as heavyweight querying/analysis Tactical options. ■ Products with strong lightweight non-web and web reporting capabilities have been included in Tactical and Strategic. ■ Tactical and Strategic products were selected to leverage NIH's investment in products that are a proven fit for NIH's known future needs. Leveraging baseline products in the future will minimize the operations, maintenance, support and training costs of new products. ■ Some baseline products have been designated retirement and containment. These products are either not as widely or successfully deployed at NIH, or they do not provide as much functionality, value, or Total Cost of Ownership as the selected tactical and strategic products.

3.4 Brick 4: Web Content Management

The NIH collaboration domain team has defined web content management as a technology that automates the content creation, approval, and publication process of any digital items (e.g. video, audio, text, graphic, links to physical resources, etc.), thereby providing internal and external web accessibility, management, and search functionality based on user roles and access rights. Web content management systems are instrumental in publishing and editing inward and outward web sites. The web content management marketplace has grown stagnant and several vendors have moved out of this marketplace to focus on smart enterprise suite solutions that combine multiple collaboration capabilities into a single application. Although the market is saturated with vendors, a few vendors have established a significant presence in this marketplace and provide scalability as needed to support the users requirements.

The NIH enterprise has multiple web content management applications within their baseline today. Some of these applications provide a light solution for web content management including PaperThin CommonSpot, RedBridge Dynabase, Apache Cocoon and Zope (open source).

As with all applications used by the Federal Government, it is required for these applications to be section 508 compliant. Both Interwoven and Microsoft CMS advertise themselves as compliant, but prior to selection of the technology, a compliance certification should be sought. The detailed break out of web content management applications is represented in the brick shown in Table 8.

Table 8. Web Content Management Brick

Baseline Environment (Today)	Tactical Deployment (0-2 years)	Strategic (2-5 year)
■ PaperThin CommonSpot ■ RedBridgeDynabase ■ InfoSquare OpenShare ■ Interwoven Teamsite ■ Microsoft CMS ■ Percussion Rhythymx ■ Merant Collage ■ Zope (Open Source) ■ Novell X-tend ■ Apache Cocoon (MAC) ■ Macromedia Contribute ■ Miscellaneous Software Source Control Mgmt	■ Microsoft CMS (Robust) ■ Merant Collage ■ Zope (Open Source) ■ PaperThin CommonSpot (Light) ■ Interwoven Teamsite ■ Percussion Rhythymx ■ Apache Cocoon (MAC)	■ Microsoft CMS ■ Interwoven Teamsite
Retirement (Technology to eliminate)	**Containment (No new deployments)**	**Emerging (Technology to track)**
■ Miscellaneous Software Source Control Mgmt (for HTML content)	■ RedBridge Dynabase ■ InfoSquare OpenShare ■ Macromedia Contribute	■ Other vendors with leading WCM capabilities, such as Vignette ■ Vendors that have grown into the WCM space, such as Documentum, Stellent, and Novell X-tend ■ Tools with strong connection capabilities, such as Venetica VeniceBridge
Comments		
■ WCM solutions that are robust, light, open-source, or MAC-based were included as Tactical options. ■ Miscellaneous Software Source Control Management products used throughout NIH were included in Retirement as they do not provide true and robust WCM capabilities. ■ Tactical and strategic products were selected to leverage NIH's investment in products that are a proven fit for NIH's known future needs. Leveraging baseline products in the future will minimize the operations, maintenance, support and training costs of new products. ■ Some baseline products have been designated retirement and containment. These products are either not as widely or successfully deployed at NIH, or they do not provide as much functionality, value, or Total Cost of Ownership as the selected tactical and strategic products.		

3.5 Brick 5: Document Management

The collaboration domain team defined document management as software systems that allow enterprises to generate, produce, store, manage, retrieve, and distribute electronic files (e.g. text, image, audio, video) yielding greater efficiencies in the ability to reuse information and to establish workflow constructs. Document management

tools typically interface with record management and workflow tools to provide a seamless collaborative environment.

Document management applications have also been subsumed into the functionality of smart enterprise suites and other multiple use applications currently popular in the marketplace. Several applications that were historically considered document management tools also contain functionality for standalone workflow capabilities, depending on how the application is integrated into the information technology environment.

Table 9. Document Management Brick

Baseline Environment (Today)	Tactical Deployment (0-2 years)	Strategic (2-5 year)
■ Infosquare OpenShare ■ File Magic ■ Open Text LiveLink ■ Innopac ■ Percussion Rhythymx ■ Xerox Docushare ■ Documentum Application Xtender	■ MS Sharepoint Team Services (lightweight) ■ Open Text Livelink ■ Documentum Application Xtender	■ Documentum Application Xtender
Retirement (Technology to eliminate)	**Containment (No new deployments)**	**Emerging (Technology to track)**
■ File Magic	■ Infosquare OpenShare ■ Percussion Rhythymx ■ Xerox Docushare ■ Innopac	■ Smart Enterprise Suites (SESs)
Comments		
■ Tactical and strategic products were selected to leverage NIH's investment in products that are a proven fit for NIH's known future needs. Leveraging baseline products in the future will minimize the operations, maintenance, support and training costs of new products. ■ Some baseline products have been designated retirement and containment. These products are either not as widely or successfully deployed at NIH, or they do not provide as much functionality, value, or Total Cost of Ownership as the selected tactical and strategic products.		

3.6 Brick 6: Workflow Tools

The collaboration domain team defined workflow tools as tools that allow the design, support, and implementation of specific work processes. These tools typically augment workflow with collaborative functionality, including document management, instant messaging, chat, e-mail, white boarding and other tools that facilitate employee coordination and collaboration. These tools are common within administrative environments that require a high volume of collaboration across the organization.

Workflow tools have significant usefulness within the NIH environment and are currently being considered to assist in the grants review and approval process. As with the document management marketplace, workflow tools are moving toward the same integrated single application offering that might be found in a smart enterprise suite or a knowledge management application.

Table 10. Workflow Tools Brick

Baseline Environment (Today)	Tactical Deployment (0-2 years)	Strategic (2-5 year)
■ Handysoft BizFlow* ■ NMS Imaging E-Flow ■ Open Text LiveLink ■ Oracle Workflow ■ Metastorm eWork ■ Bamboo Solutions ■ Sitescape	■ Handysoft BizFlow ■ Open Text LiveLink ■ Oracle Workflow/9I+ ■ NMS Imaging E-Flow	■ Handysoft BizFlow ■ Open Text LiveLink ■ NMS Imaging E-Flow
Retirement (Technology to eliminate)	**Containment (No new deployments)**	**Emerging (Technology to track)**
■ None Identified	■ Sitescape ■ Bamboo Solutions	■ Vendors that have grown into the workflow space, such as Open Text (Open Image) and Oracle (10G) ■ Other leading robust workflow products, such as Metastorm eWork
Comments		
■ Tools that will no longer be supported within NIH have been placed in Retirement.. ■ * The NIH Application Integration domain team has identified Handysoft BizFlow as a tactical option for workflow-based application to application integration. ■ Tactical and Strategic products were selected to leverage NIH's investment in products that are a proven fit for NIH's known future needs. Leveraging baseline products in the future will minimize the operations, maintenance, support and training costs of new products. ■ Some baseline products have been designated Retirement and Containment. These products are either not as widely or successfully deployed at NIH, or they do not provide as much functionality, value, or Total Cost of Ownership as the selected Tactical and Strategic products.		

3.7 Brick 7: Digital Signature

Digital Signature systems are defined as applications that provide identification and storage of digital signatures along with validation against a centrally housed database. Applications for this technology are somewhat limited at this time, due to the fact that the market and technology is changing rapidly.

Digital signatures can be applied at the desktop, on the web, or through forms. The market is very volatile at this time and there are few vendors that are recognized by the Federal Government as viable providers of digital signature technology. Vendors must be able to provide an X.509 V3 certificate with a federal profile in order to be considered for implementation within the NIH enterprise. Specific functions for consideration with digital signature technology include sectional signing, supporting multiple signatures, getting and managing digital signatures, and incorporating digital signatures into forms.

Standards are evolving and the latest standard that should be sought when considering digital signature technology is SCVP. Mr. Mark Silverman serves as the NIH subject matter expert for digital signature technology and should be consulted during any digital signature implementation. The brick provided in Table 11, was created in collaboration with Mr. Silverman and provides choices for tactical deployment of digital signature technology.

Table 11. Digital Signature Brick

Baseline Environment (Today)	Tactical Deployment (0-2 years)	Strategic (2-5 year)
Web Toolkit: ■ Infomosaic SecureXML Web Toolkit Desktop Solution: ■ Adobe's SelfSign (only solution for Mac) ■ Silanis ApproveIt Desktop	Forms-based Package Solution: ■ PureEdge ■ Adobe Acrobat ■ Infomosaic ■ Silanis Web Toolkit: ■ Infomosaic SecureXML Web Toolkit ■ Silanis ApproveIt Web Server Desktop Solution: ■ Infomosaic SecureSign (Desktop Solution) ■ Silanis ApproveIt	■ TBD

Retirement (Technology to eliminate)	Containment (No new deployments)	Emerging (Technology to track)
■ TBD	■ TBD	■ Other leading solutions, such as: ❑ CIC Redshore ❑ Documentum Trusted Content Services ❑ AINS Inc. CATXpress ❑ NMS Imaging eFlow ❑ CSI CyberSign ❑ Thunderbird
Comments		
■ A full inventory of baseline tools has not been completed ■ The PKI architecture for NIH is in the process of being designed and deployed, and is therefore subject to some change in the near-term. ■ Strategically, the preference is for open standards for content and digital signature formats like XML and DigSig XML over proprietary standards like Adobe and Microsoft. ■ Tactical products may change because the marketplace is still emerging and products and companies may shakeout further through acquisitions. ■ All products must support X.509 V3 certificates according to the Federal profile standard. ■ Tactical and Strategic products were selected to leverage NIH's investment in products that are a proven fit for NIH's known future needs. Leveraging baseline products in the future will minimize the operations, maintenance, support and training costs of new products. ■ Some baseline products have been designated Retirement and Containment. These products are either not as widely or successfully deployed at NIH, or they do not provide as much functionality, value, or Total Cost of Ownership as the selected Tactical and Strategic products.		

3.8 Brick 8: Document Imaging

Document imaging devices can be standalone or embedded into the organization's existing, fax, copy or duplication systems. Imaging systems take existing documents and create a digitized standard format. Files created by document imaging systems might be in Adobe, Microsoft Word, or other recognizable formats. By digitizing documents, an organization can then utilize document management systems to file and manage documents as needed to support the organizational objectives.

A typical imaging system is built from a series of hardware and software subsystems, each designed for a particular function in the process, such as:

- Capturing and converting the images from paper to electronic code,
- Classifying the images for later retrieval (indexing),
- File searching (via keywords, text retrieval, etc.),
- Managing the images and index data (metadata),
- Storing/archiving the images, and

- Distributing or routing the images as part of a business transaction.

The core component of an imaging system is, of course, the software for processing and managing the scanned document images. Most imaging software today is developed for Microsoft Windows environments with server software running on Microsoft Windows NT or Windows 2000 and desktop clients supporting Windows 98, Windows NT, and Windows 2000. Typically, imaging software is developed using Microsoft's Distributed Component Object Model (DCOM) certified for Microsoft BackOffice

The NIH enterprise has not completed a baseline inventory of document imaging devices within the NIH organization. The conclusions represented within the brick at Table 12 were developed with domain team members' knowledge of existing systems and direct experience with the Kofax Ascent Capture software.

Table 12. Document Imaging Brick

Baseline Environment (Today)	Tactical Deployment (0-2 years)	Strategic (2-5 year)
■ Xerox Docushare ■ Canon E-Copy ■ Kofax Ascent Capture ■ Adobe Acrobat 6.0	■ Kofax Ascent Capture	■ TBD
Retirement (Technology to eliminate)	**Containment (No new deployments)**	**Emerging (Technology to track)**
■ TBD	■ TBD	■ Other leading solutions, such as: ❑ WMS eiStream ❑ FileNET ❑ Tower Technology ❑ Captiva
Comments		
■ A full baseline of the NIH tool set has not been conducted ■ Emerging Technologies were established through Gartner Research and proposed solutions to current procurement ■ Tactical and Strategic products were selected to leverage NIH's investment in products that are a proven fit for NIH's known future needs. Leveraging baseline products in the future will minimize the operations, maintenance, support and training costs of new products. ■ Some baseline products have been designated Retirement and Containment. These products are either not as widely or successfully deployed at NIH, or they do not provide as much functionality, value, or Total Cost of Ownership as the selected Tactical and Strategic products.		

4.0 Gap Analysis

The domain teams have defined the following gaps between NIH's current environment and the vision (as defined in the patterns and bricks):

- ICs and CIT have not addressed: ownership, sponsorship, politics, and funding related to the solutions presented in the bricks and patterns.
- The recommendations presented in the bricks and patterns may not address the unique requirements and process definitions of the ICs. ICs must do their own requirements analysis to see how the recommended solutions fit in their environment and seek an exception to the EA if they are found to be insufficient .
- The recommendations presented in the bricks and patterns provide solutions for multiple cost points, however, each IC is responsible for analyzing its own budget and requirements before selecting a solution.

5.0 Next Actions

A few next actions need to be performed or executed in order for NIH to begin to migrate to the future state as defined by the Collaboration patterns and bricks. At the conclusion of the domain team meetings, the team identified these next actions:

- Communicate and share information among domain team members and ICs. Work towards centralization of Collaboration efforts to avoid duplication.
- Monitor implementation and enforcement of EA standards (EA governance) to ensure the ICs can maintain flexibility they need. Keep the EA process soft, easy, painless, and streamlined.
- Synchronize application development standards with web browser standards as many web browsers defined in the brick have dependencies with custom applications developed within NIH.
- Ensure ICs and CIT create a working relationship of mutual benefit and can provide IT knowledge across the enterprise.
- Work with procurement and CIT to get site licenses for products defined in the bricks; specifically look at the Google search engine and Crystal Reports licensing.

6.0 Change History / Document Revisions

Date	Change Author	Change Authority	Change Event	Resulting Version
6/22/04	Raj Patel, Greg Early	Jack Jones	Original Production	1.0

Appendix A—Glossary of Terms

7.0 Appendix A—Glossary of Terms

Term	Definition
Brick	The physical building blocks of the architecture, representing the building blocks of the architecture -- both hardware and software. They describe both the baseline and target technologies for the components identified in design patterns. Bricks document the architecture's physical technology.
CIT	Center for Information Technology that provides IT support to the 27 NIH ICs.
Digital Signature	Applications that allow the identification and storage of digital signatures along with the ability of that system to validate the signature against a centrally housed database.
Document Imaging	System that takes existing documents and creates a digitized standard format. Files created by document imaging systems might be in Adobe, MS Word, or other recognizable formats. By digitizing documents, an organization can then utilize document management systems to file and manage documents as needed to support the organizational objectives.
Document Management	Software systems that allow enterprises to generate, produce, store, manage, retrieve and distribute electronic files (e.g. text, image, audio, video) yielding greater efficiencies in the ability to reuse information and to establish workflow constructs.
Enterprise Reporting Tool	A tool that allows organizations, including federal agencies, to gain a better understanding of their business by putting critical information in the hands of all those who need it – employees, managers, partners, and customers. An enterprise reporting tool must be configurable to meet the needs of its user base, capable of accessing information assets in the enterprise based on access rights, and must support wide-scale deployment.
HTML (Hypertext Markup Language)	The set of markup symbols or codes inserted in a file intended for display on a World Wide Web browser page.
IC	One of the 27Institutes and Centers that comprise NIH.
OLAP (Online Analytical Processing)	Computer processing that enables a user to easily and selectively extract and view data from different dimensions.
Pattern	A collection of logical or physical components arranged in a specific configuration. Patterns provide design guidance to implementation teams and can occur in any of the architectural areas, and may sometimes cross-areas.
PKI (Public Key Infrastructure)	Enables users of a basically unsecured public network such as the Internet to securely and privately exchange data and money through the use of a public and a private cryptographic key pair that is obtained and shared through a trusted authority.
Search Engine	A "robot" or "crawler" that goes to every page or representative pages on a Web site, or the whole Web, and creates an index; or, a program that receives a search request, compares it to the entries in the index, and returns results to the end user.
Web Browser	Programs that "read" hypertext and display it as formatted text and images. Browsers allow users to view the contents of a site and navigate from one site to another.
Web Content Management	A technology that automates the content creation, approval and publication process of any digital items (e.g. video, audio, text, graphic, links to physical resources, etc.), there by providing internal and external web accessibility,

Term	Definition
	management and search functionality based on user roles and access rights.
Workflow	Tools that allow the design of specific work processes, support of these processes, and improve these processes with collaborative functionality, including document management, instant messaging, chat, e-mail, white boarding and other tools that facilitate employee coordination and collaboration.
XML (Extensible Markup Language)	The set of markup symbols that provide a flexible way to create common information formats and share both the format and the data on the World Wide Web, intranets, and elsewhere.

Client Contact Information

John F. Jones, Jr.
Chief IT Architect
Telephone: +1 (301) 402-6759
E-mail: jonesjf@mail.nih.gov

Gartner (Contractor Support) Contact Information

Terry McKittrick
Gartner Consulting
Telephone: +1 (703) 226 4779
Facsimile: +1 (703) 226 4702
E-mail: Terry.McKittrick@gartner.com

Gartner

research consulting measurement community news