Building Solutions with EPA Enterprise Application Platforms Getting it Right

Magrogan, Philip

ITS-ESE CTO

Sponsored by
John Sullivan,
Chief Architect

What is Our Context?

EPA has recognized strengths in computer security and enterprise architecture

- Full applications lifecycle support
- Driven to Include Service Architecture by OMB FEA
- Over 300 Applications Across Multiple Contractors
- Need for Long Term Cost Efficiencies

What is our Vision?

Build reusable application platforms that a Franchisee can leverage

Cost of development	Develop / deploy cycle time	Quality of Applications
Lowered by 30 % or more	Lowered by 50 % or more	Significantly Improved by Reuse

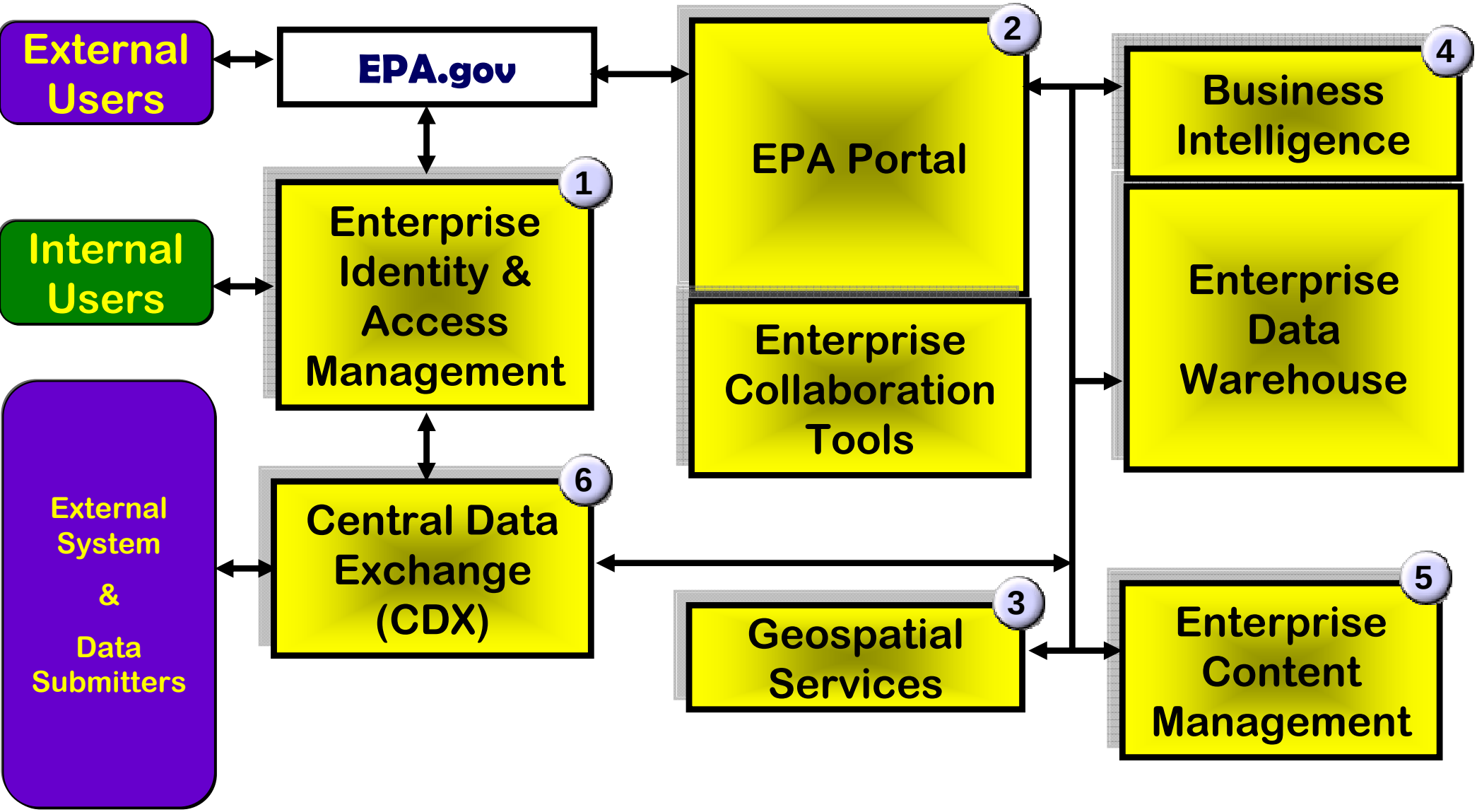
Platforms

Include reusable components, services, security plan, and infrastructure

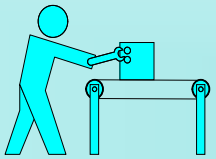
Franchise Developer Handbooks

Capture best practices and standards for interoperability

Target Architecture
Service Oriented Platforms



Franchise Development—An Open Business Model



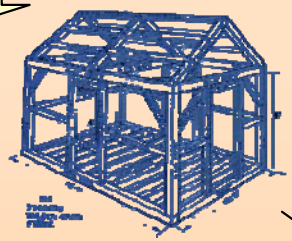
Preferred Contractors

Business Owner

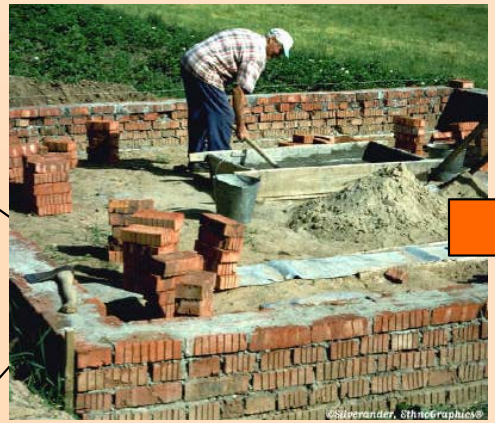
- Domain knowledge
- Skills
- Regulations

Franchisee

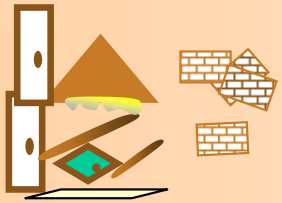
- Services
- Standards and conventions
- Deployment process



Franchise Handbook



Application Solution



Components

Consulting Training

Testing Env

Franchise

Franchise Development Methodology

An Open Business Model

Franchise Package includes:

Products:

Components of COTS products, published services, security plan, test cases etc.

Processes:

Handbooks of reusable design patterns, best practices for implementing applications, naming standards for interoperability, and deployment instructions

Knowledge Transfer:

Training, consulting, working group membership

Expected Benefits

Faster and more flexible reconfiguration of business processes

Decrease of IT operational costs

Secure and reliable services

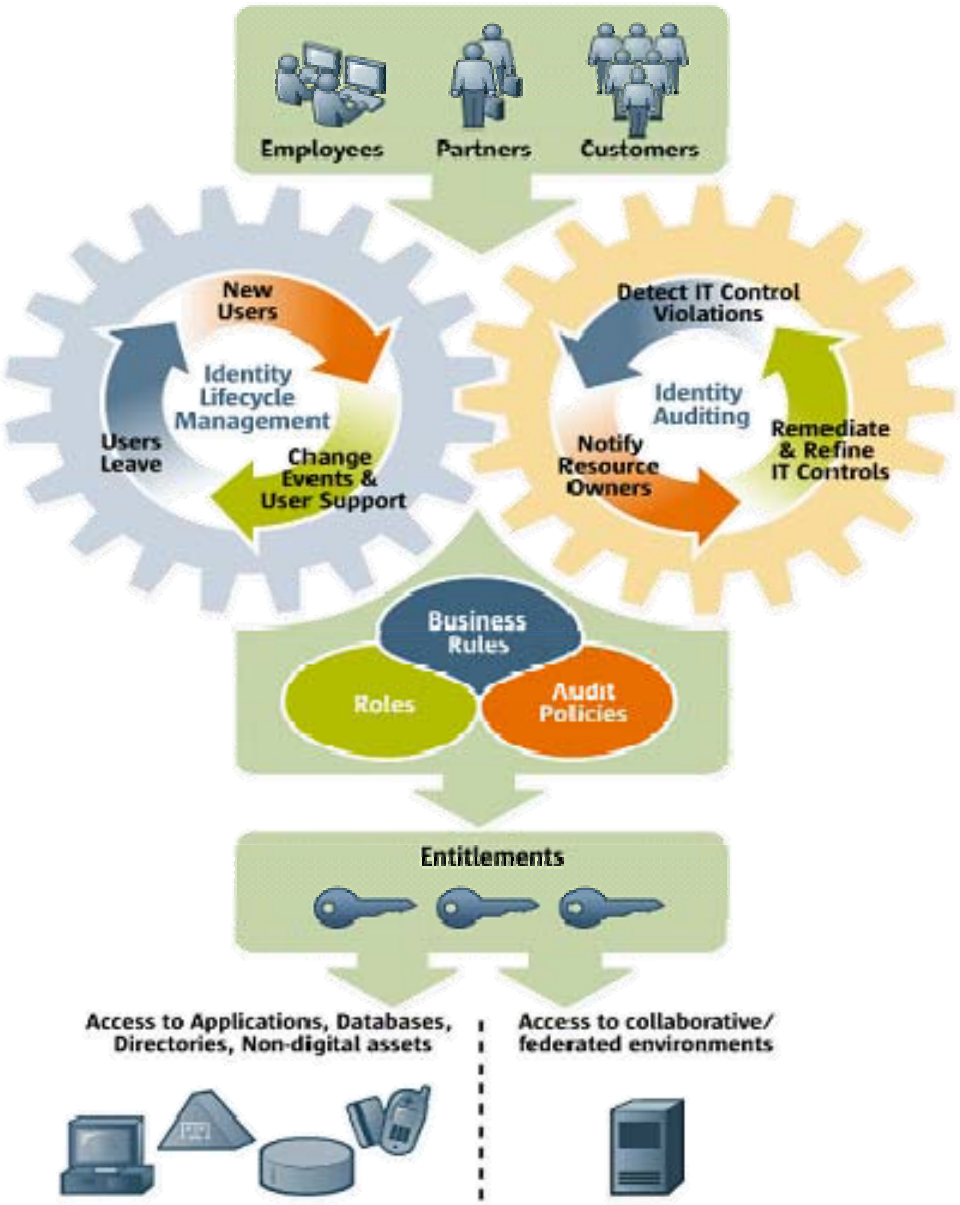
On the fly upgrades

After successful deployment of SOA enabled application platforms, it is expected to achieve 30% overall cost saving to develop applications with similar scope. This estimate is based on experience.

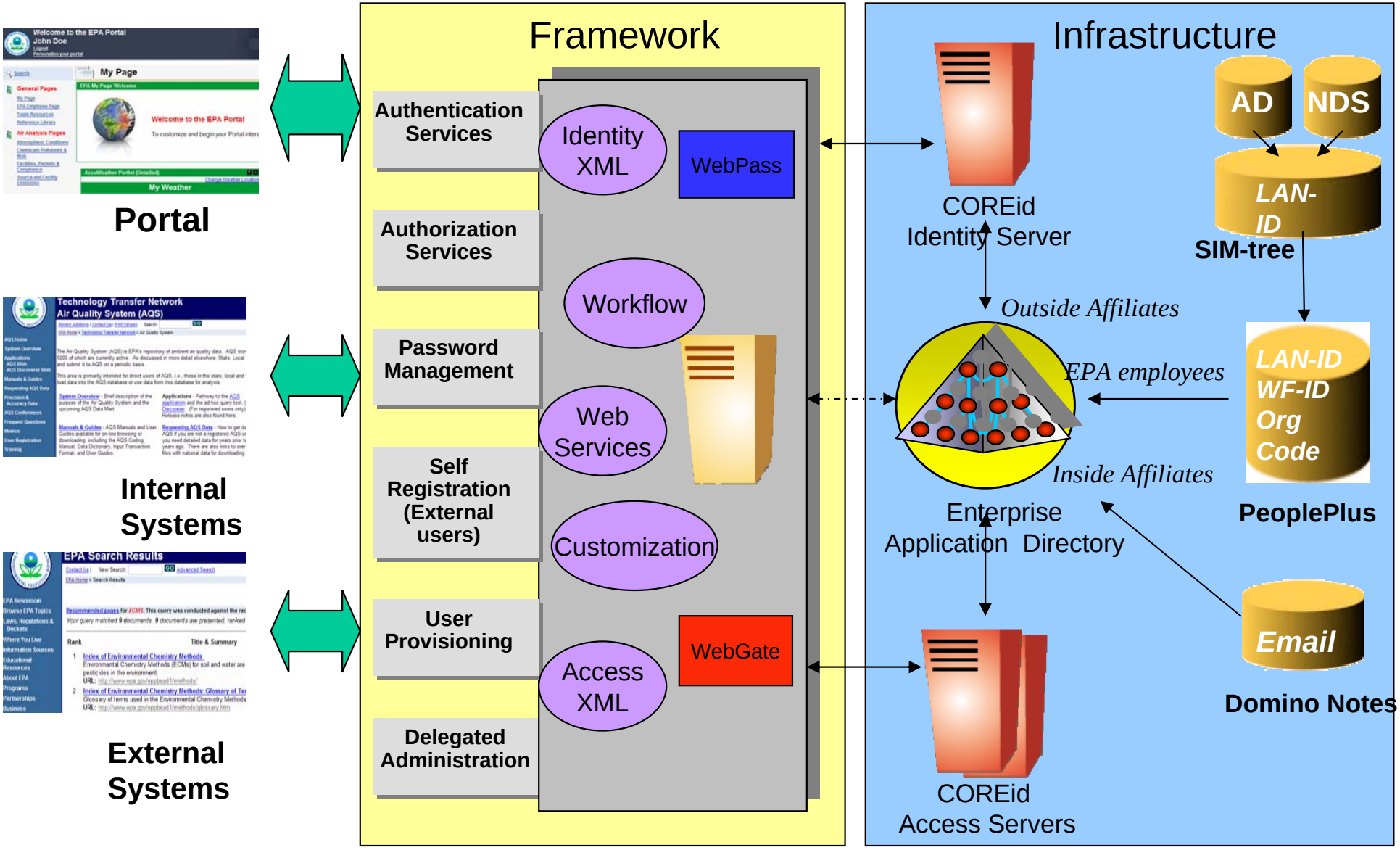
- Increase ROI
- Reduce TCO
- Improve Quality
- Shorten Development Cycle

Gartner Industry Survey Results

Identify & Access Mgt (IAM) Conceptual Architecture



EPA IAM Solution Architecture



Example: Office of Emergency Management Application

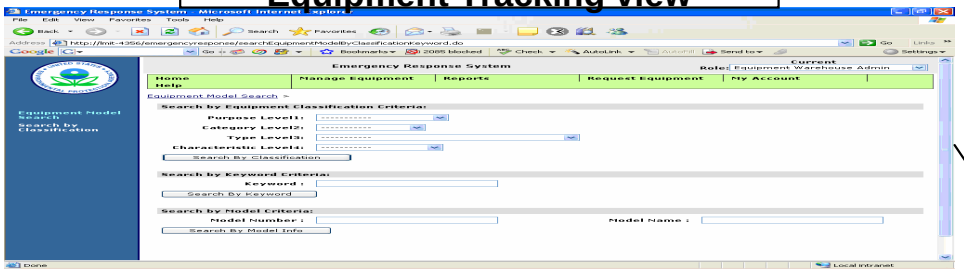
Goal was to deploy role-based OEM portlets that link to external applications.

OFFICE OF ENVIRONMENTAL INFORMATION

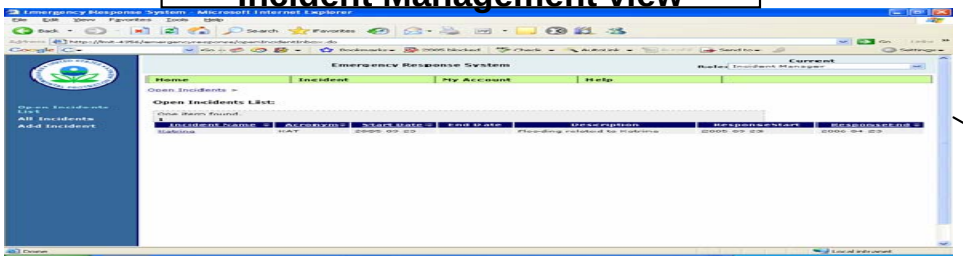
OEM Applications

IAM Web Services

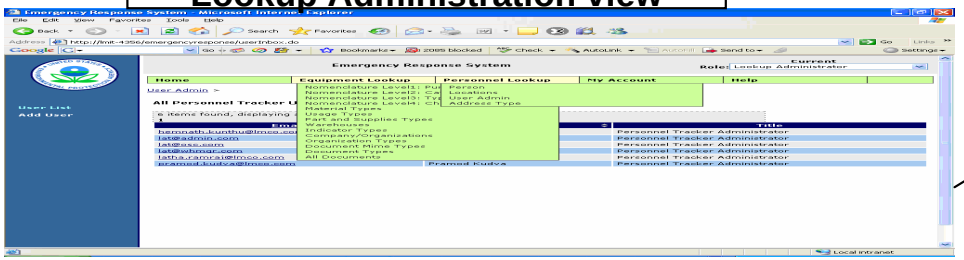
Equipment Tracking view



Incident Management view



Lookup Administration view



Auth
authenticate()
validate()
centralAuth()
centralValidate()

UserManager
getAttributes()
getUser()
getUsers()
addUser()
changeUserPassword()
deleteUser()
deactivateUser()

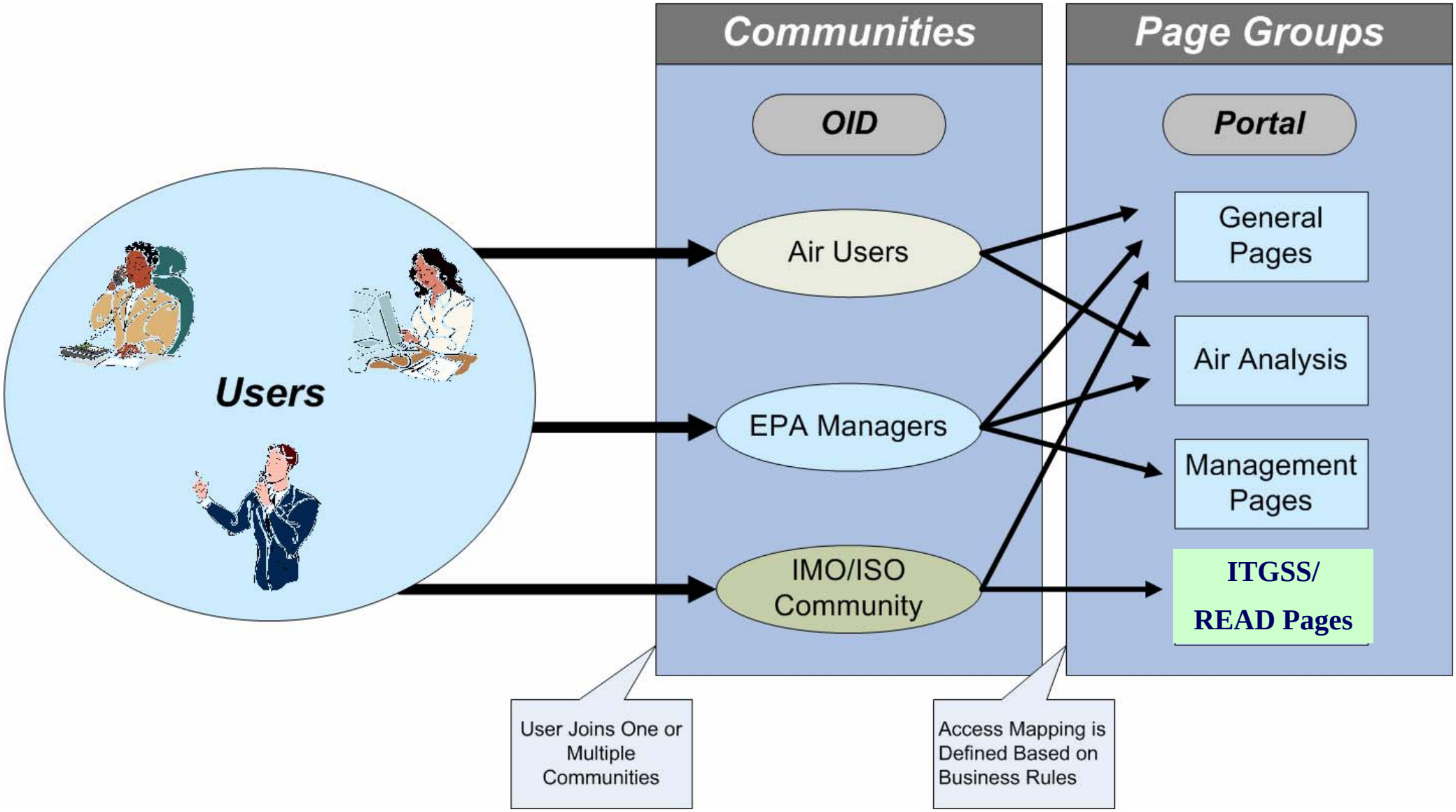
GroupManager
getGroupMembers()
getUserGroups()

IAM Platform Cost Savings

- **Reduction of dozens of passwords to single sign-on offers**
  - **Increased usability**
  - **Increased security**
  - **Significantly reduce the cost of duplicate password management.**

Cost of Duplication			
	Low	High	Mean
Cost per User	$ 80	$ 100	90
Number of Duplicates	18,000	72,000	45,000
Annual Cost of Duplicates	$1,440,000	$ 7,200,000	$ 4,320,000
5 year Opportunity	$7,200,000	$36,000,000	$ 21,600,000

Portal Concept of Operations - Communities Use Case View



Example: Portal Integration via Inter-Portlets Communication

OFFICE OF ENVIRONMENTAL INFORMATION

Geo Portlets

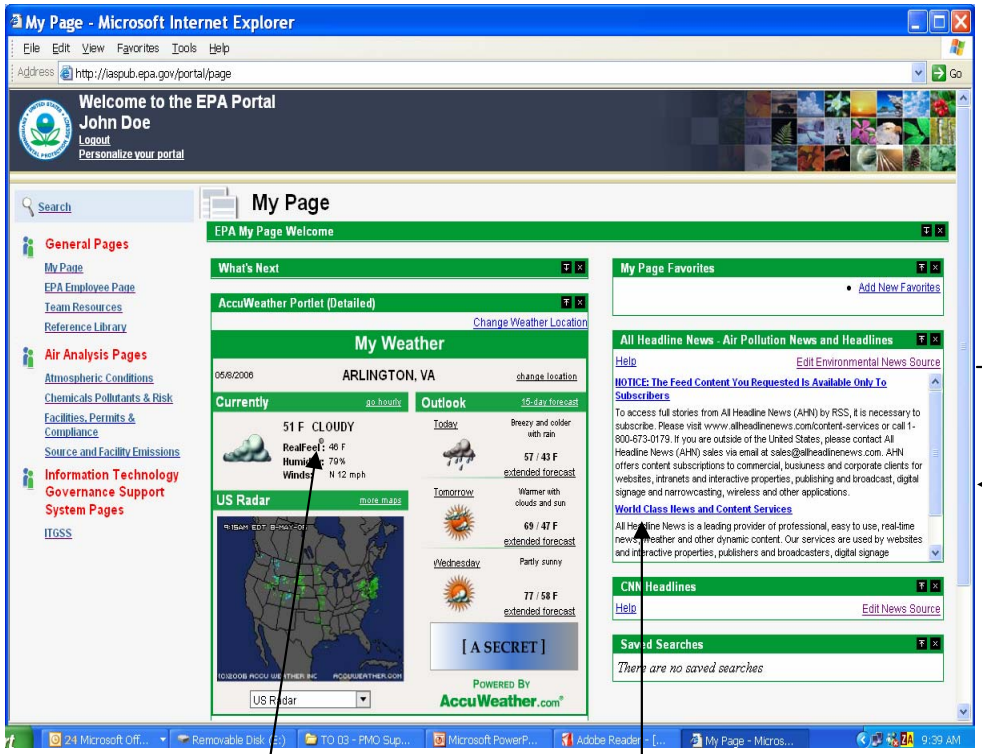
ECMS Portlets

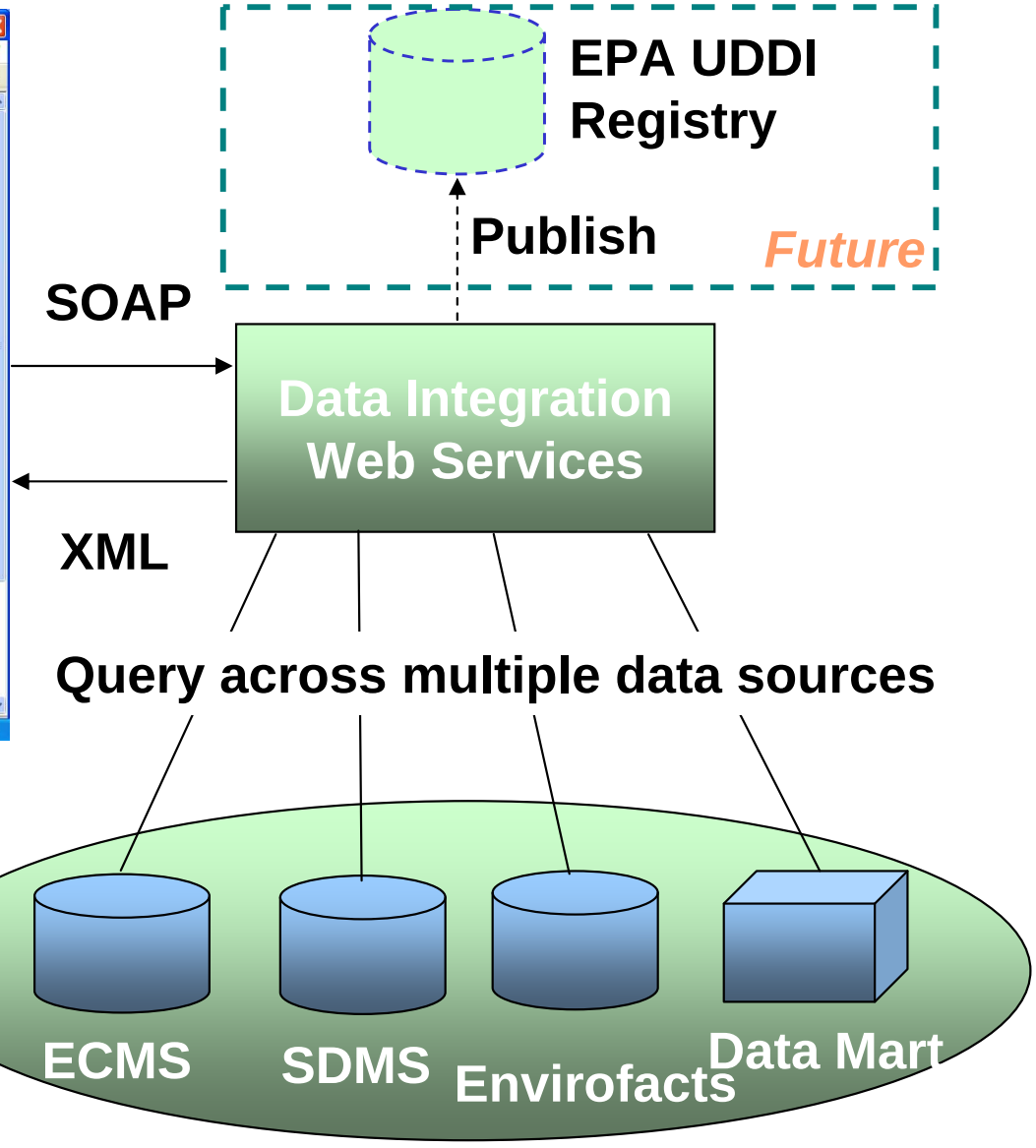
Search Portlets

Business Intelligence Portlets

Portal Integration is more than "the sum of its parts" – the components in a portal can add value to other components through inter-portlets communication.

Example: Portal Application Leveraging Other Integration Mechanisms



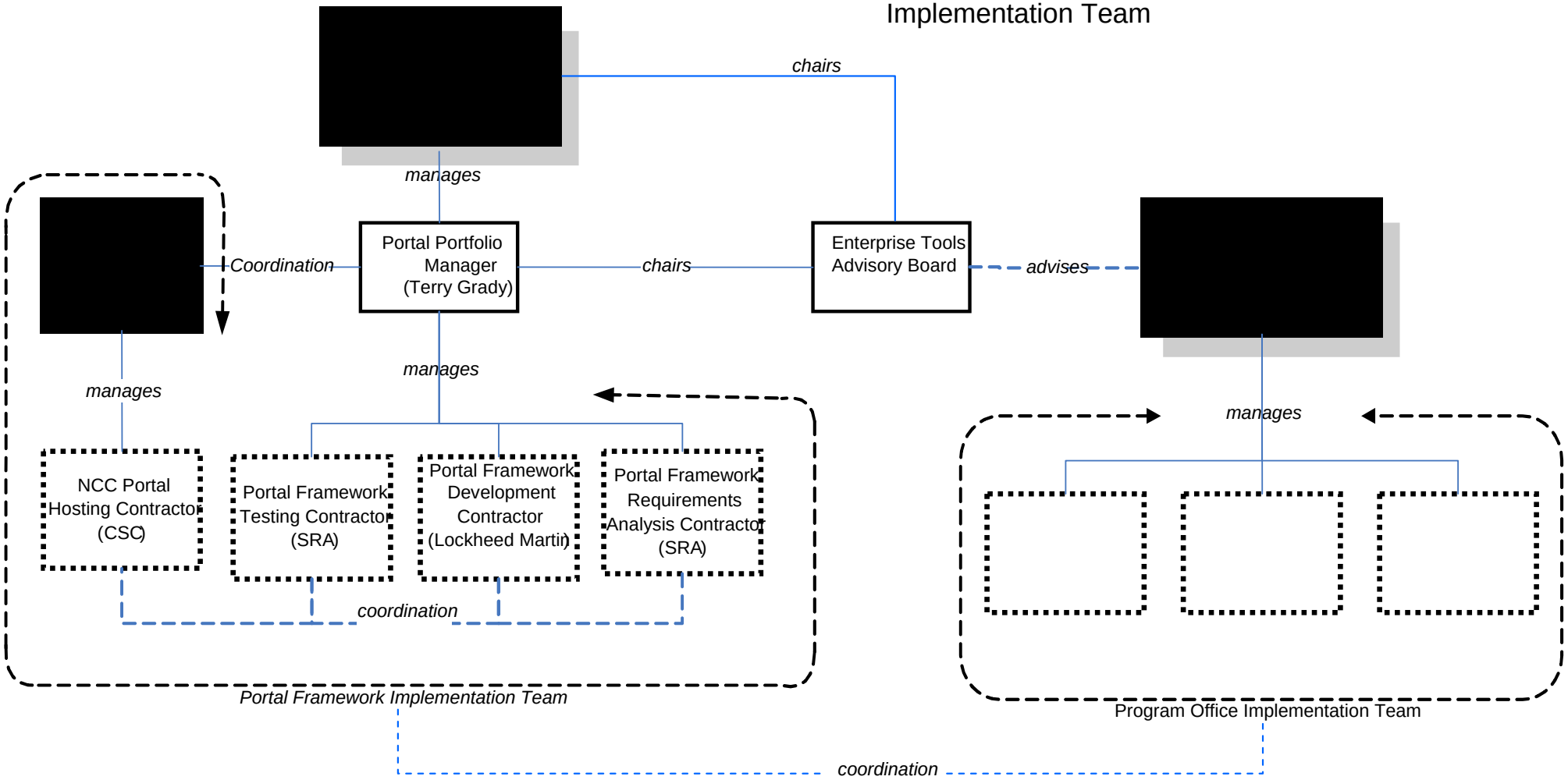


Create Portlets from web services, RSS news feed, etc

Portal Governance

EPA Groups and their Roles in Portal Application Development

- Portal Management Organization
- Portal Framework Implementation Team
- Program/Administrative/Regional Office Implementation Team



Portal Cost Savings

- **Having one community with consistent, dynamic content tailored to the individual is a huge positive cultural change**
  - **Portal knows who you are (MySpace, MyOrg, MyApps)**
  - **Portal knows where you are (Local Weather/traffic/news)**
- **Cost savings per application is averaging**
  - **30% in development**
  - **50% in deployment and infrastructure support**
  - **Many Portlets are reused with zero modification or simply a change in input stream**

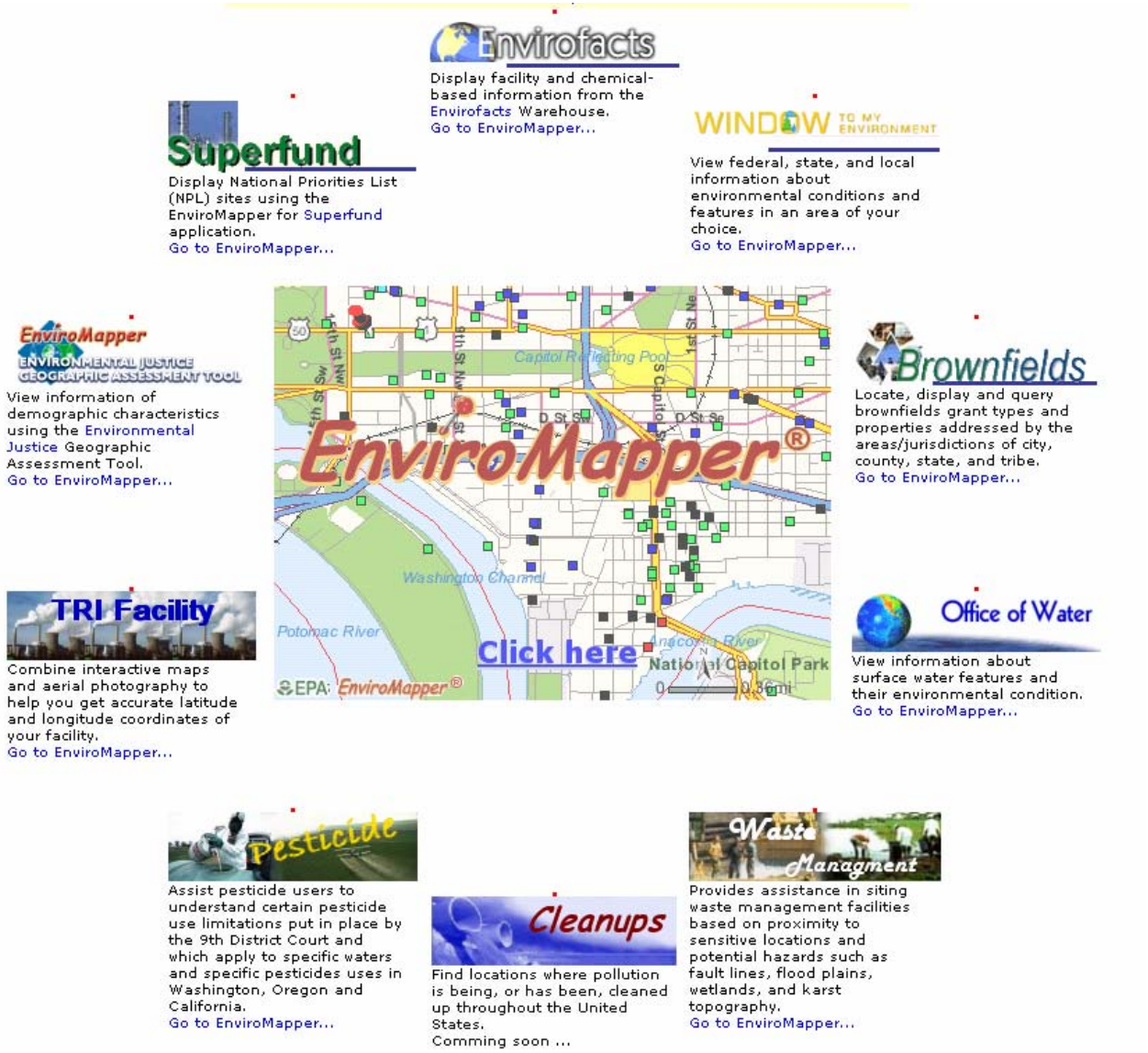
Portal Cost-Benefit Analysis: SDMS Business Case

- Option 1: Implement the application requirements without Portal
- Option 2: Adopt the Portal framework
- The estimate shows 30% less development time with Portal

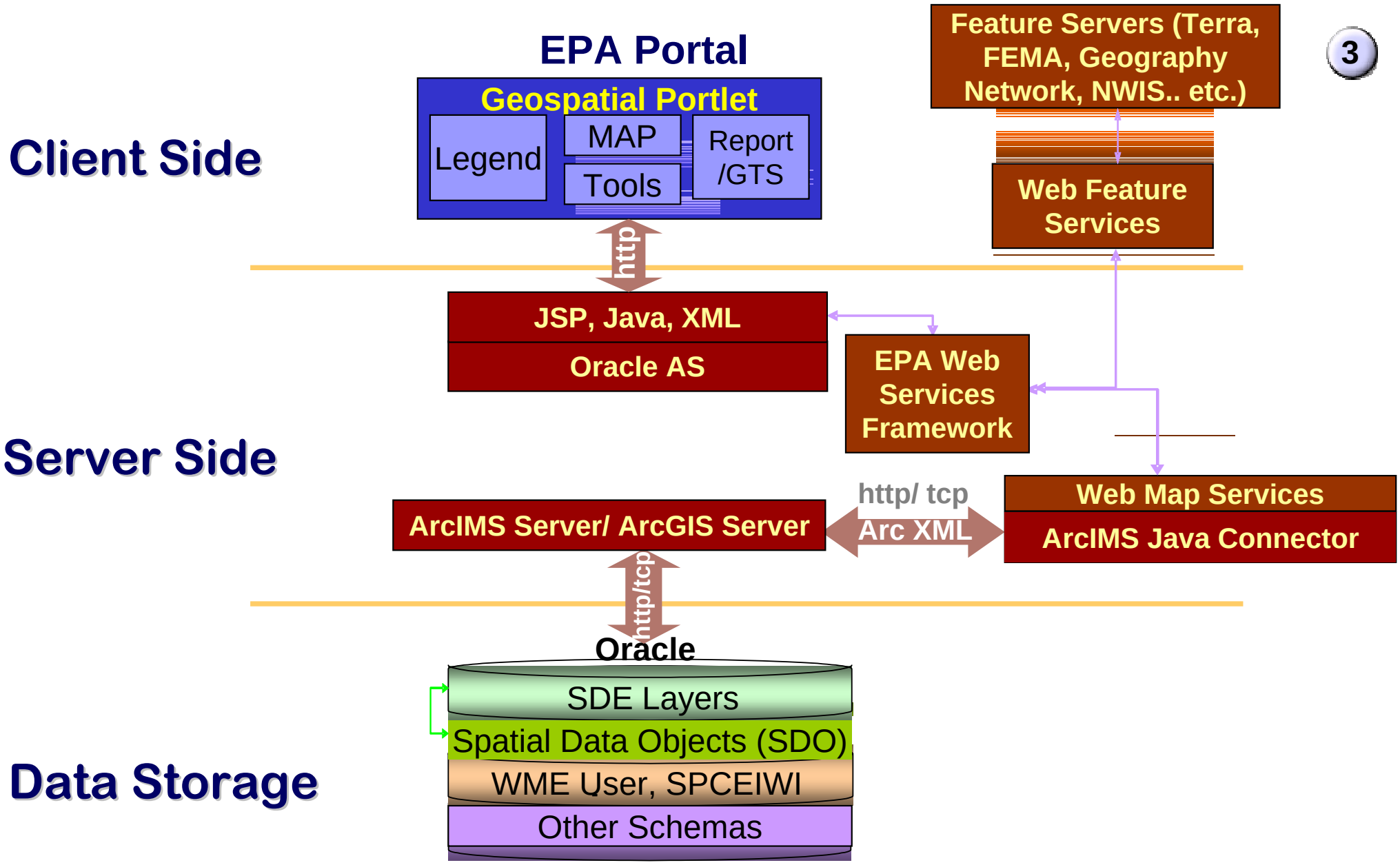
Modules	Option 1 Hrs	Option 2 Hrs
Home Page - News Bulletins and Announcements	110	95
SDMS	50	50
LOGIN/SECURITY MODULE	100	94
SEARCH	88	45
MESSAGE BOARDS	108	45
DOCUMENTATION	160	60
REGIONAL PERSPECTIVE	84	65
Contacts	91	60
Events Calendar	60	45
News Letter and News Letter Archive	90	55
Core File Structure	45	45
QuickPlace/Useful Links	9	9
Total	**993**	**668**

Current Geospatial EnviroMapper Product Family

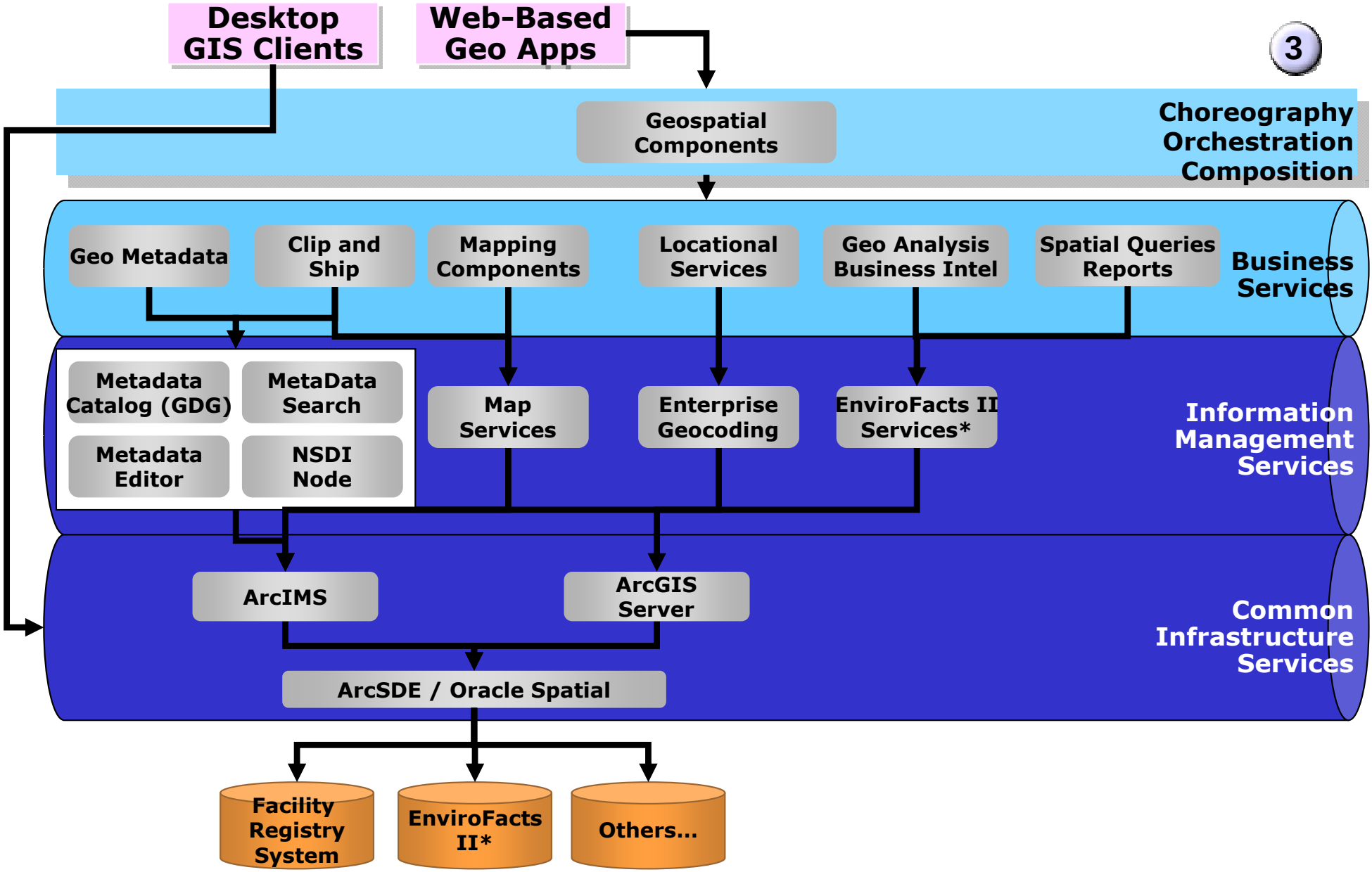
- Multiple applications with much overlap
- "Clone and own" development paradigm
- Multiple redundant code bases



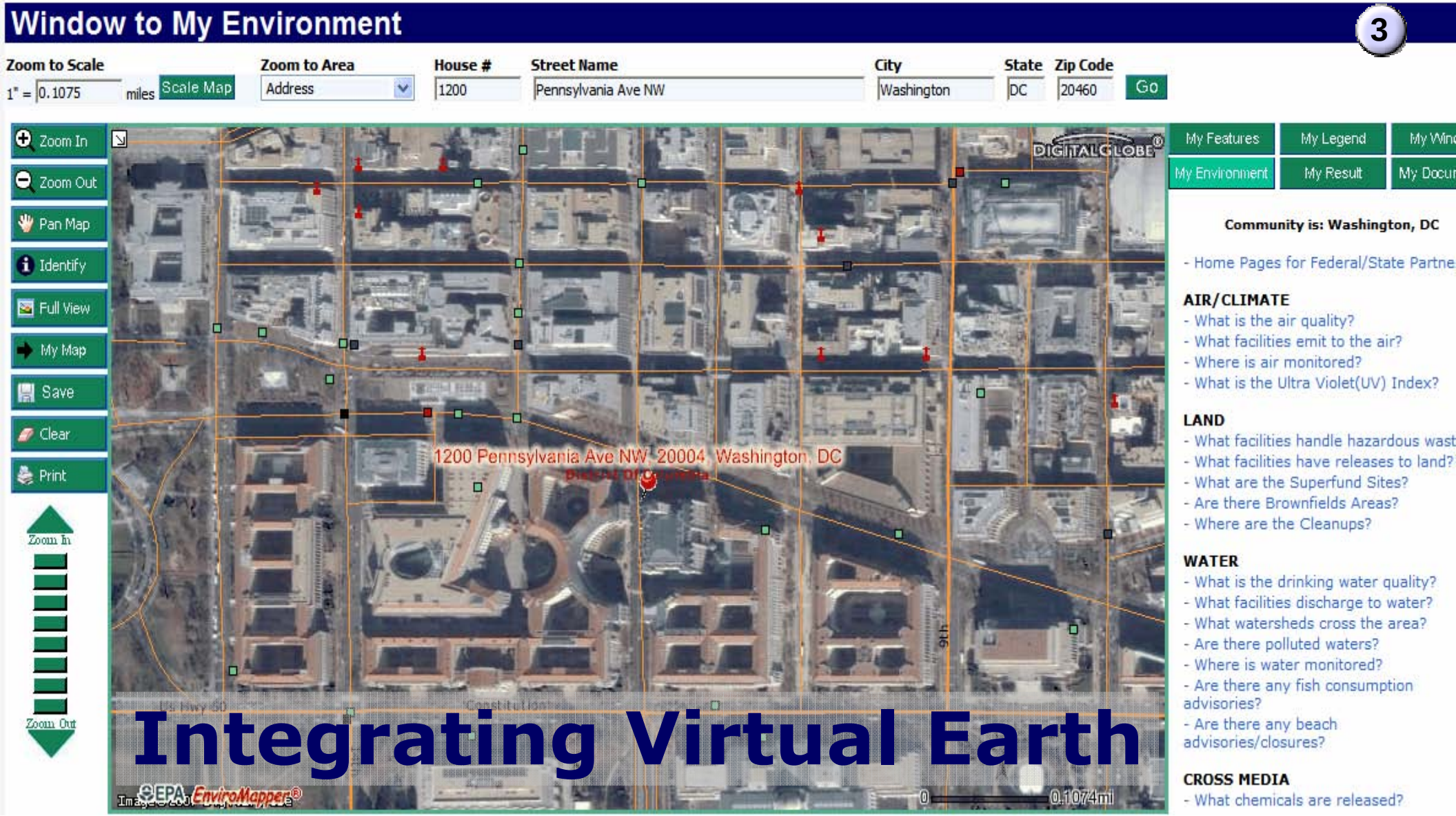
GeoSpatial Conceptual Architecture



Geospatial Services Platform



Adding New Technologies Seamlessly with the New EnviroMapper



Integrating Virtual Earth

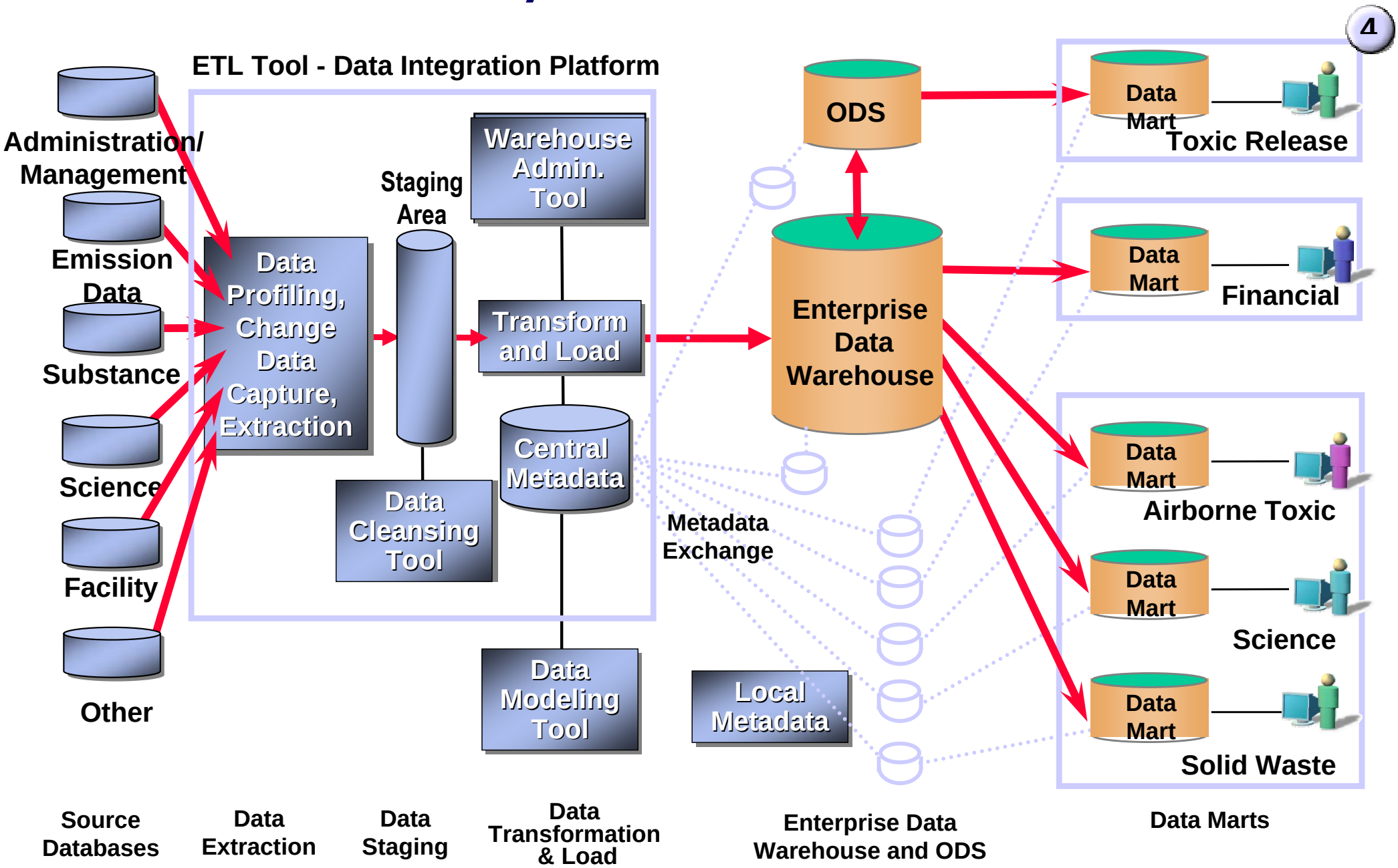
GIS Platform Cost Avoidance Estimate

Assumptions:

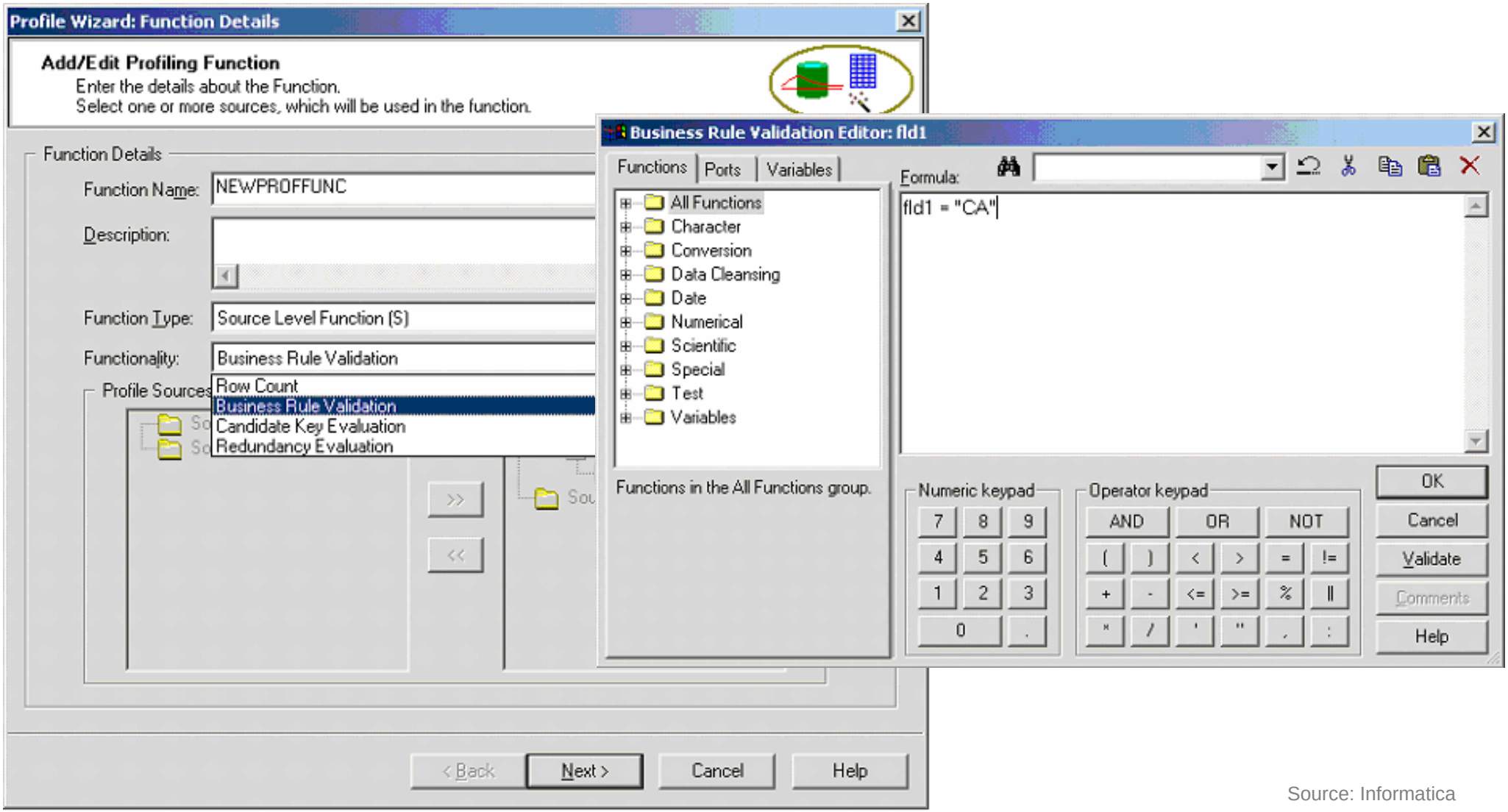
- 9 applications exist and must be maintained as separated code bases
- Cost for hosting and storage is estimated at $25K/yr
- Based on 2006 actual for Katrina, one Full Time Developer per base application is required
- 10% of base for maintenance on remaining 8 applications
- EPA Labor estimated at 5% of one person for 8 applications

Labor Rate EPA	$ 75	
Labor Rate Contractor	$ 90	
Number of Apps	9	
Hours/Yr HQ/App	100	$ 67,500
Hrs/Yr/NCC/App	100	$ 67,500
Hrs/Yr/Contractor/App	2000	$ 342,000
(Maintenance of 8 apps at 10%)		
WCF Charge/Yr/App	$ 25,000	$ 225,000
Total Annual Cost		**$ 702,000**

Data Warehouse / Mart Services Platform

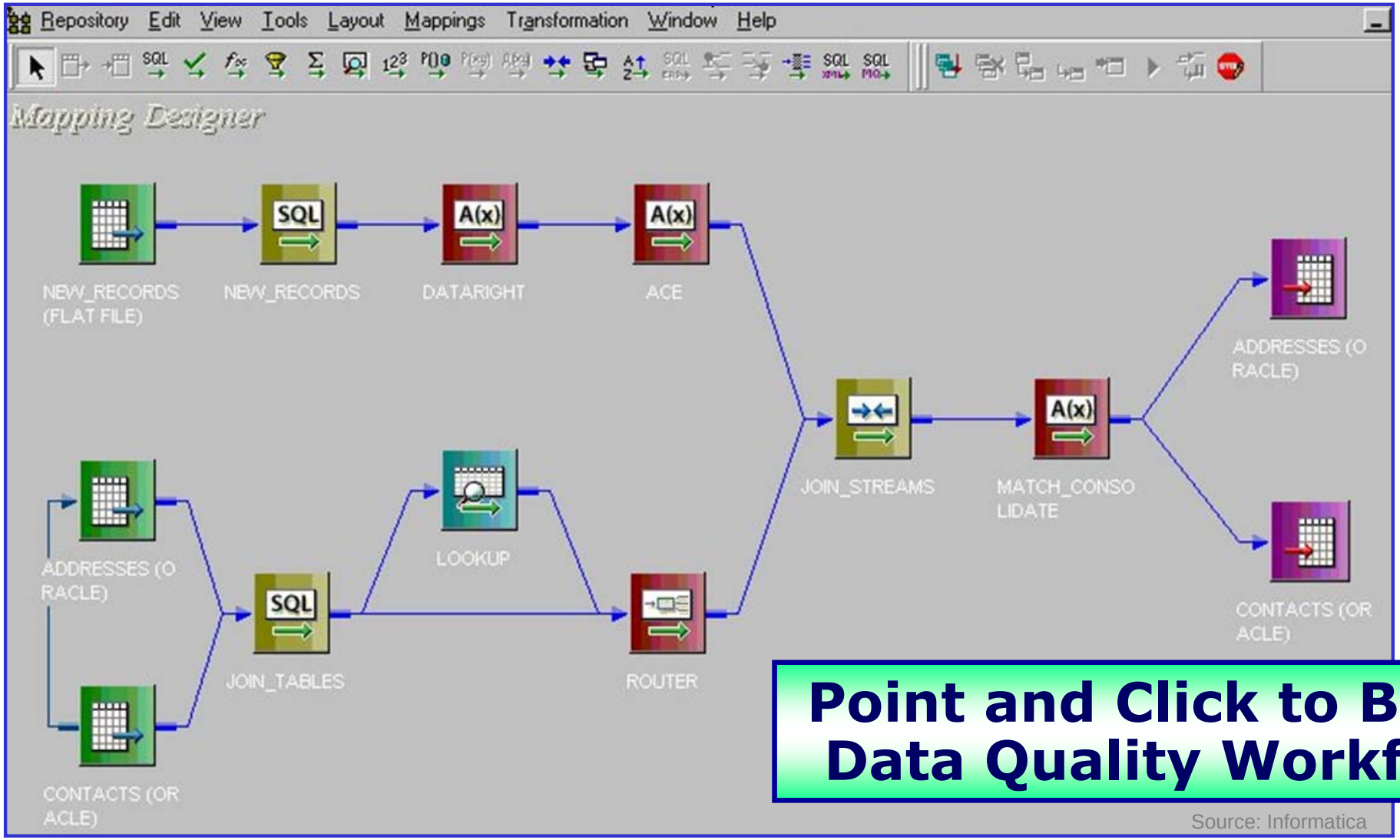


Automated Setup of Profiling or Correction Rule Data Quality Objects



Source: Informatica

Assemble Data Quality Objects into Automated Workflow



Enterprise Content Management System Concept of Operation

OFFICE OF ENVIRONMENTAL INFORMATION

5

Research & Development

- Design Doc Control
- Engineering Release Mgt
- Patent/IP Mgt
- Product Release Mgt
- Program Mgt

Operations

- Engineering Change Mgt
- Corrective Action Mgt
- Bulk Record Storage
- Standards Mgt
- Quality Assurance Mgt
- Reporting

Procurement Management

- Contract Mgt
- Supplier Score Carding
- RFP/RFQ Mgt
- Supplier Portal

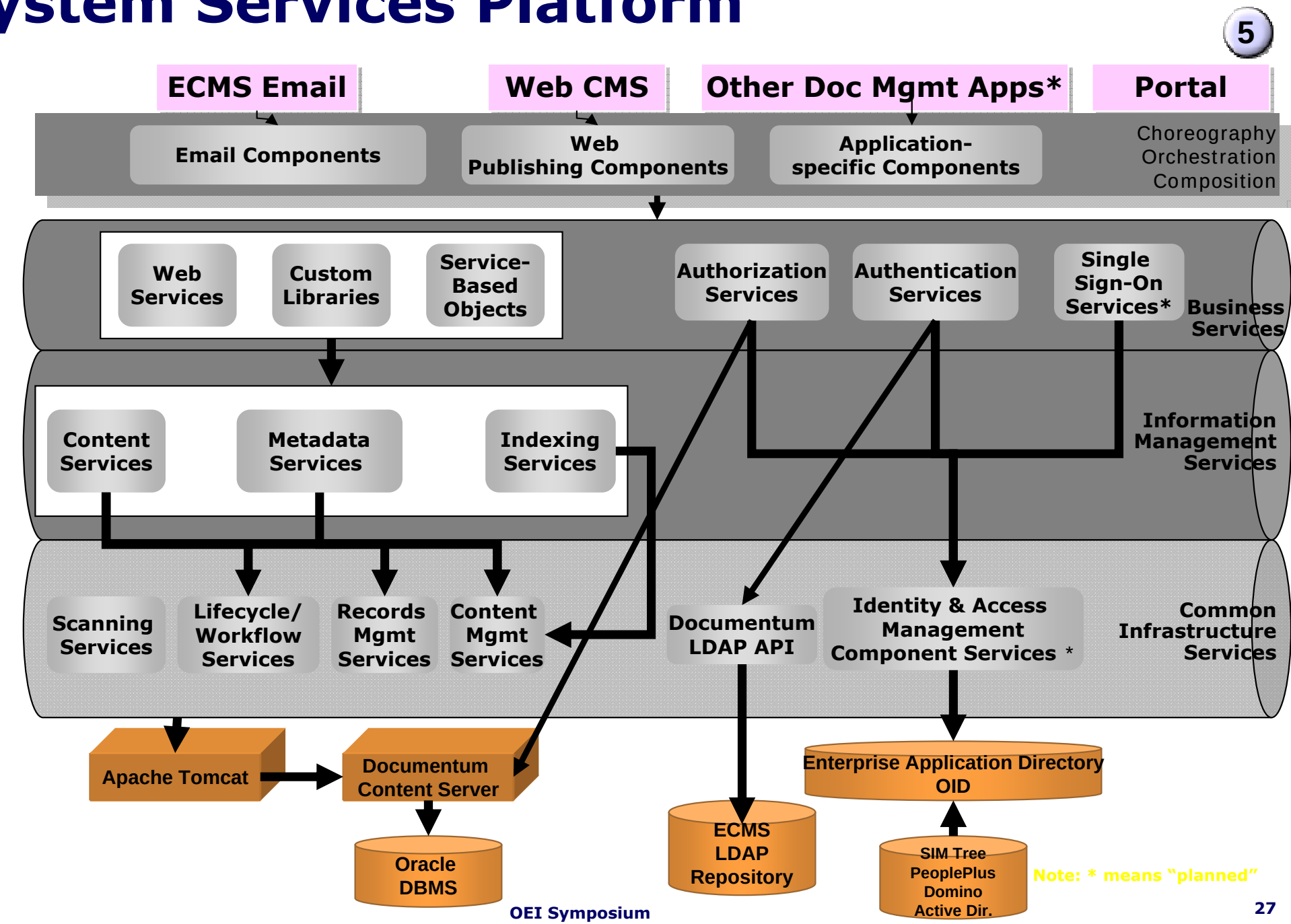
Administration

- Litigation Support
- Records Mgt
- Knowledge Mgt
- Compliance
- Corporate Portals

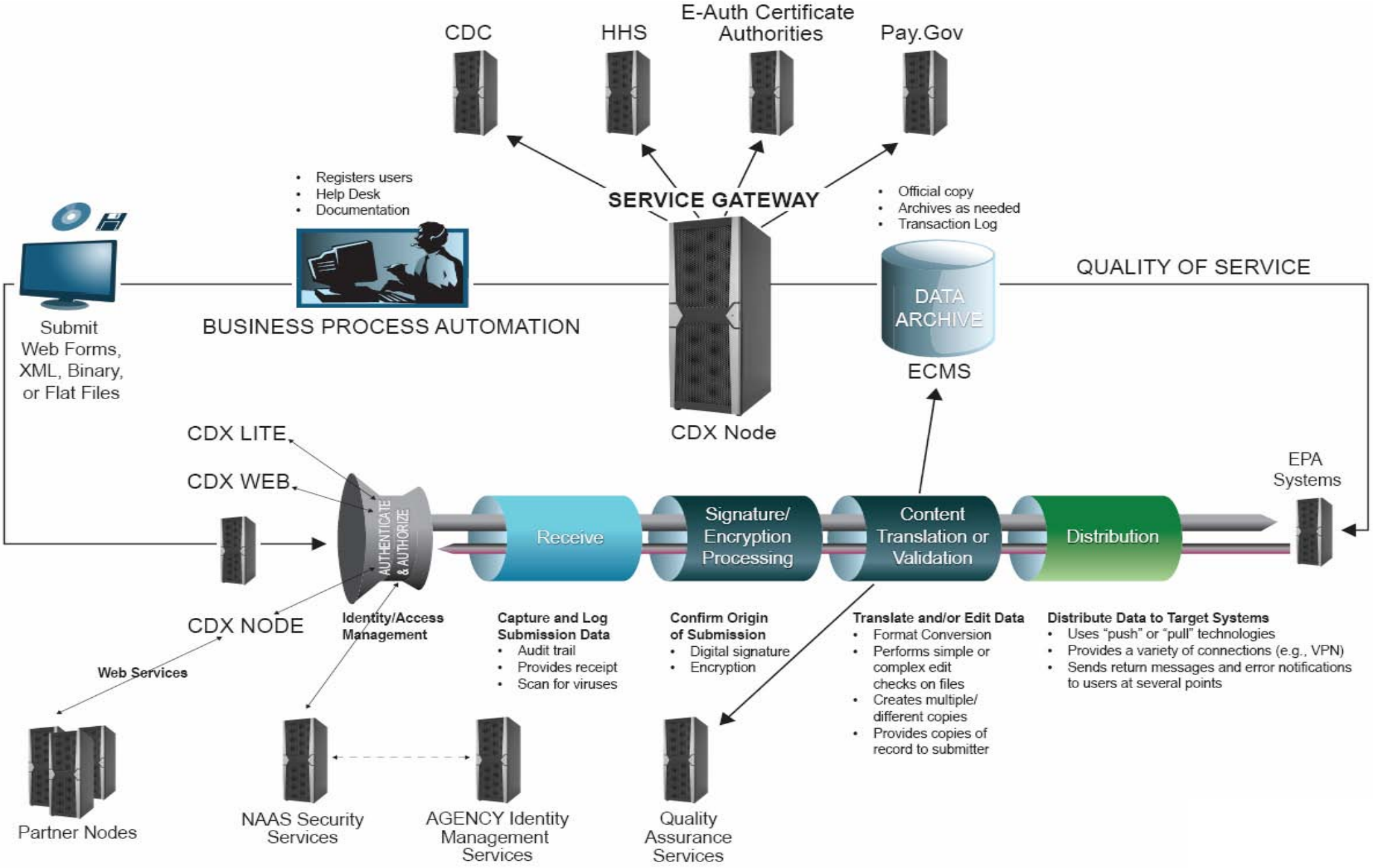
Communications

- Technical Publications
- Website Mgt
- Agency Collaboration
- Collateral Production
- FOIA Requests

Enterprise Content Management System Services Platform



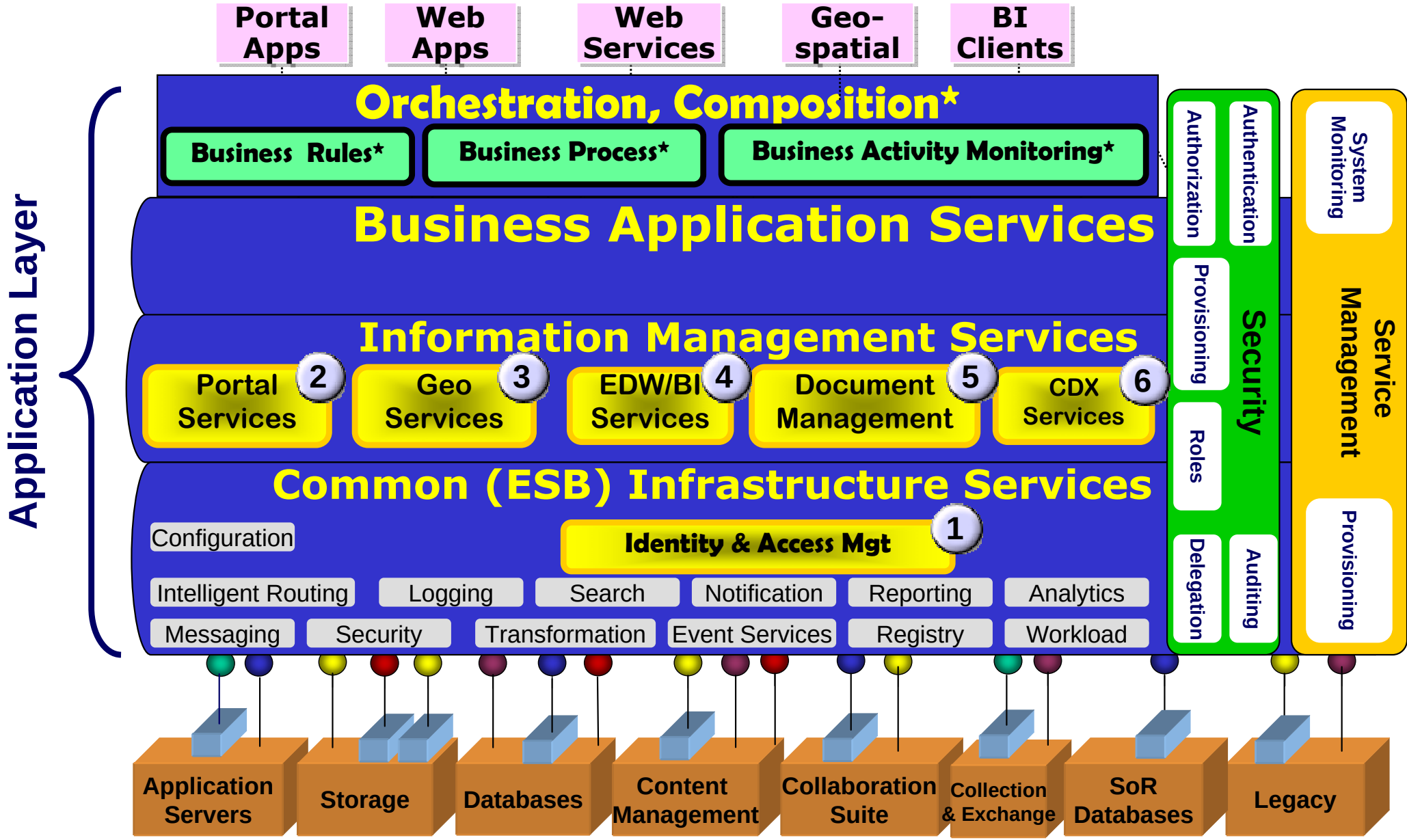
CDX Services



CDX Value Proposition

- **Cost Reduction**
  - Centralized administration of data exchange functions
    - Eliminates redundant infrastructure, applications and their associated maintenance costs
    - Automates flow development thru innovative, reusable components (CDX Lite and Exchange Network)
- **Enables faster, lower-cost implementation of new or modified data flows through re-use and modularization**
  - No more stove pipe applications
- **Improvements in Data Quality/Timeliness**
  - Decreases time to make information publicly accessible
    - Eliminates data entry tasks
  - Promotes higher quality data through enforcement of data specification or content standards

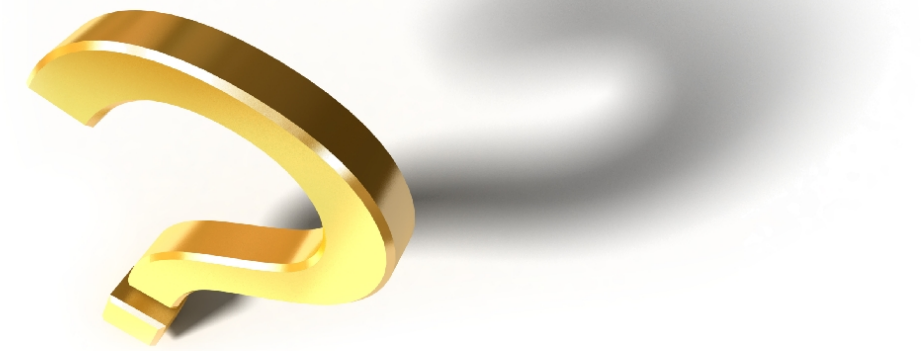
Vision Architecture: COTS-Based, Service Oriented, Reusable Components



Note: * means "planned"

Additional Information

- Appendix A: Finding and Documenting Solutions
  - How to find all the reusable “stuff” OEI has to offer
- Appendix B: Technology Roadmaps
  - Plans for new versions of COTS products
- Additional Information
  - Michael J. Cullen PMO, OEI
  - Phil Magrogan, CTO ITS-ESE



Phil Magrogan

CTO for Environmental Programs

(703) 731.5236

Phil.Magrogan@lmco.com

Appendix A

Finding and Documenting Solutions

Objectives

- Describe where to find reusable solutions or parts of solutions
  - Software applications and modules
  - Web services and other data services
  - XML schema
  - Code sets
- Describe where to find reusable designs
  - Software designs and templates (such as portal templates)
  - Data models
  - Data element names and definitions
  - Data standards
- Describe the future plans for reusable solutions and designs
- Describe how to document various types of solutions and designs

Finding Reusable Solutions: Software Applications and Modules

- **Registry of EPA Applications and Databases (READ)**
  - **Official agency inventory of systems and applications**
    - **Use full version found under ITGSS Portal**
    - **Designed to store information about modules as well as systems/applications**
    - **Contact information is included – at minimum call the system owner to see if they may have code that is in condition to be reused.**
    - **Some data dictionaries available (many more in future after new EDR operational)**
    - **Searchable by system name, architecture category, etc.**
  - **You might find a reusable application because:**
    - **Regions often have same needs**
    - **Some offices have similar functions**
  - **First effort has been to get complete and accurate information about systems – next step will be to enter information about modules**

Finding Reusable Solutions: Software Applications and Modules (continued)

- **GeoData Gateway (GDG) -**
  - EPA Technical Specification 1.0 for geospatial metadata - Agency-specific implementation of the FGDC Content Standard for Digital Geospatial Metadata (CSDGM)
  - EPA National Geospatial Data Policy - http://www.epa.gov/esd/gqc/pdf/epa_natl_geo_data_policy.pdf
  - GeoData Gateway: http://geogateway.epa.gov
    - Automate metadata contributions and harvesting
    - Manage centralized metadata resources
    - Marketplace for planned geospatial data acquisitions
    - Additional references on geospatial data resources and management
  - EPA Metadata Editor 2.1 is available for download at GDG
  - Using the EPA metadata solutions facilitates improved decision making and NSDI compliance
    - Consistent documentation - standardized taxonomies, better keyword searches
    - Fulfill all requirements at once (FGDC/GOS/EPA)
    - Use Pre-defined templates, EPA Metadata Editor Tool and EPA Specification
    - Sensitive/non-sensitive information is documented consistently for internal/external access
    - Rapid, painless development and publishing of geospatial metadata

Finding Reusable Solutions: Web Services and Other Data Services

➢ Exchange Network Services Discovery Tool (ENDS)

- Currently being adopted by Exchange Network partners to facilitate registration and discovery for data flows

➢ UDDI Registry

- UDDI (Universal Description, Discovery, and Integration) is an XML-based registry for service providers to list themselves. Its goal is to streamline online transactions by enabling service consumers to find one another and make their systems interoperable
- Secured with a CDX login
  - Future: Anyone with EPA login will have access
- In future, these registry will be part of Service Component Registry and Repository (SCRR)

Finding Reusable Solutions: XML Schema

➢ XML Registry

- Provide the capability to share information about
  - XML Data Exchange Template (DETs)
  - XML Schemas
  - Namespaces
  - WSDL files
  - Other supporting files needed to map data flows between Exchange Network partners or EPA organizations
- Has information about schemas approved for use on the Network and those under development
- In future, part of SCRR

Finding Reusable Solutions: Code Sets

- **Environmental Data Registry (EDR)**
  - http://iaspub.epa.gov/edr/codeset$.startup
  - **County and State Codes, Tribal Codes, etc**
  - **Has capability to hold non-hierarchical codesets and record additions and deletions over time**
  - **New code set management module will be completed in FY-2008 to allow cross mapping between code sets (promoting use of standard translation web services between systems)**

Finding Reusable Solutions: Code Sets (continued)

- **Environmental Terminology System and Services (ETSS)**
  - http://www.epa.gov/etss/
  - **An agency terminology repository of environmental terms, their relationships, and their definitions**
  - **Has capability to hold hierarchical (taxonomic) lists of terms and record additions and deletions over time**
  - **Has capability to manage keyword lists**
  - **Has official EPA Web Taxonomy**
  - **New front end under development to make it very easy to search for and download taxonomies and keywords for reuse (completion in early FY-2008) – until then contact DSB for assistance**

Finding Reusable Solutions: Code Sets (continued)

- **Substance Registry System (SRS)**
  - Chemicals and biologicals
  - Standard metadata and program specific metadata
- **Facility Registry System (FRS)**
  - Offers the most authoritative record about a given facility subject to environmental regulations and particular environmental interest
  - Locational Reference Tables (LRT): Locational (i.e., latitude and longitude) information

Finding Reusable Design: Designs and Templates

- In the future these will be in the Service Component Registry and Repository (SCRR)
- Until then check with members of contractor community, PMO working groups, and other EPA and developer networks

Finding Reusable Design: Data Models

- **Core Reference Model Version 2.0**
  - The Core Reference Model (CRM) is a high-level depiction of major groupings of environmental data and their relationships.
  - Provide federal, state, and tribal environmental agencies with guidance for consistently building and sharing environmental data on the Exchange Network
- **For substance data model – contact John Harman of OIC Data Standards Branch**
- **For facility data model**
  - http://www.epa.gov/enviro/html/frs_demo/new_docs.html
  - Contact Pat Garvey of OIC Information Services and Support Branch
  - In future, in SCRR

Finding Reusable Design: Data Element Names and Definitions

- **EDR Compare Tool**
  - http://iaspub.epa.gov/edr/compare_tool$.startup
  - **Search any data element list to compare data elements from various sources (e.g., data standards, data systems)**
- **Future EDR Compare Tool in FY-2008**
  - **Has Term search**
  - **Also has Concept based searches**
  - **New EPA data standards website coming with guidance, best practices, developer collaborative site**
    - **Data standards are best used as a package – data blocks within standards may also be used to assist in component design**

Finding Reusable Design: Data Standards

- **Data Standards Documentation**
  - **Data Standards page at EDR:** http://iaspub.epa.gov/edr/epastd$.startup
- **EPA Data Standards address**
  - **Semantics (meanings)**
  - **Formats**
  - **Value/codes sets**
- **New Data Standards Web Page coming soon**

Steps for Finding Reusable Solutions & Designs

Steps	Registry/Repository	Description
Step1. Software Applications and Modules	READ	Agency inventory of systems and applications and links to data dictionaries kept in EDR
	Geospatial Data Gateway	GIS
Step 2. Web Services and Other Data Services	ENDS (future)	Data services
	UDDI Registry	UDDI for web services
Step 3. XML Schema	XML Registry	XML schema for data flows
Step 4. Code Sets	EDR	Non-hierarchical environmental data
	ETSS	hierarchical (taxonomic) environmental data
	SRS	Substance Information
	FRS/LRT	Facility, Locational (i.e., latitude and longitude) information

Steps for Finding Reusable Solutions & Designs

Category	Registry/Repository	Description
Step 5. Designs and Templates	SCRR (future)	Reusable designs and templates including portal templates
Step 6. Data Models	SCRR (future)	Conceptual, logical and physical data models
	Exchange Network Web Page	Core Reference Model Version 2.0
Step 7. Data Element	EDR	Data dictionaries and data element concepts
Step 8. Data Standard	EDR	Data Standards Section Data Standards Web Page (future)

Documenting Solutions & Designs

- Update READ for Application Discovery
- Register Data Dictionary in EDR
- Update ENDS, UDDI Registry, and XML Registry for Web Services
- Register SRS for substances list
- Register FRS for facility data

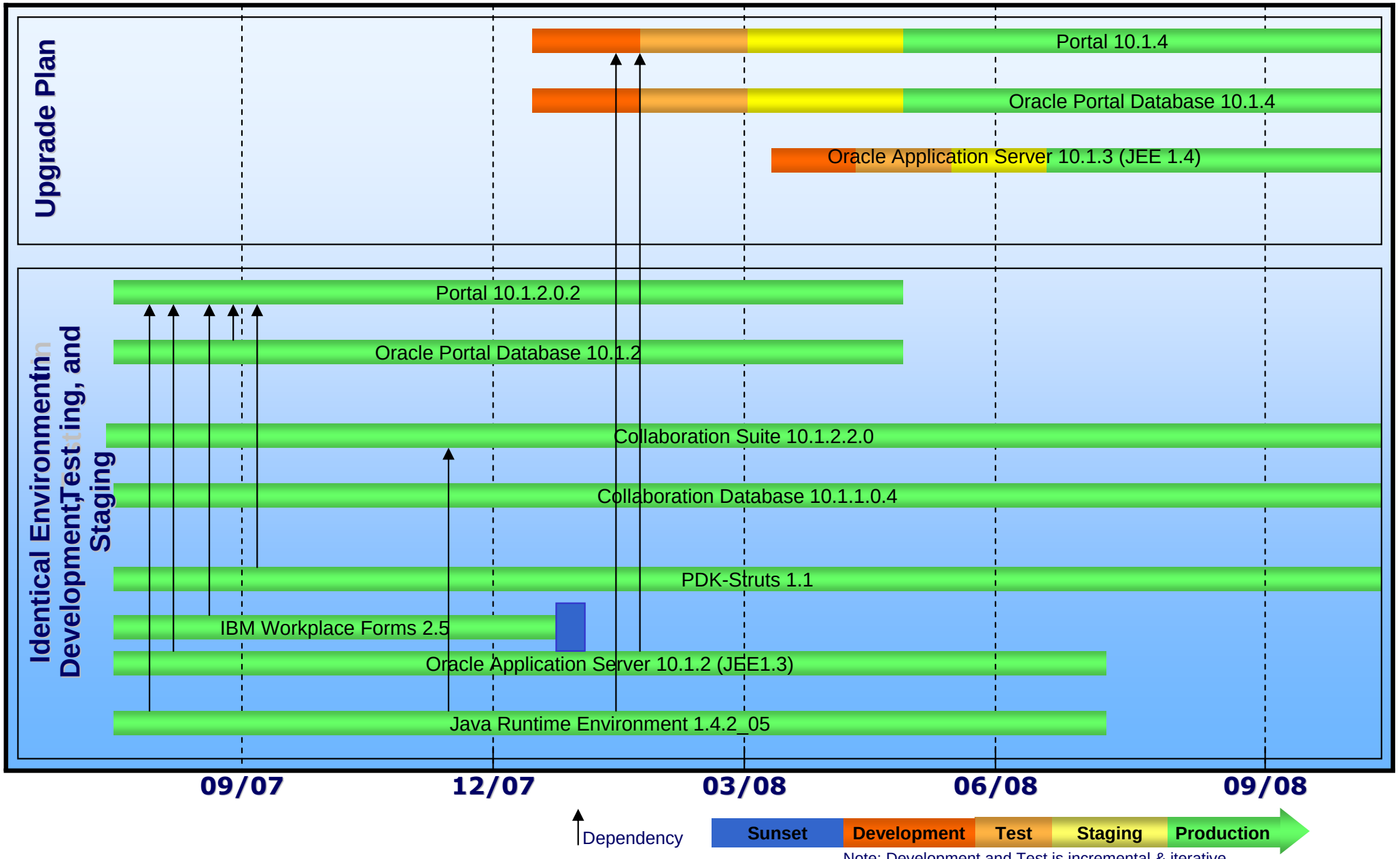
Appendix B:

Technology Roadmaps

Technology Roadmaps

- Portal
- Business Intelligence and Analytics (BIA)/ Extract, Transform, and Load (ETL)
- Identity and Access Management (IAM)
- Enterprise Content Management System (ECMS)
- Geospatial

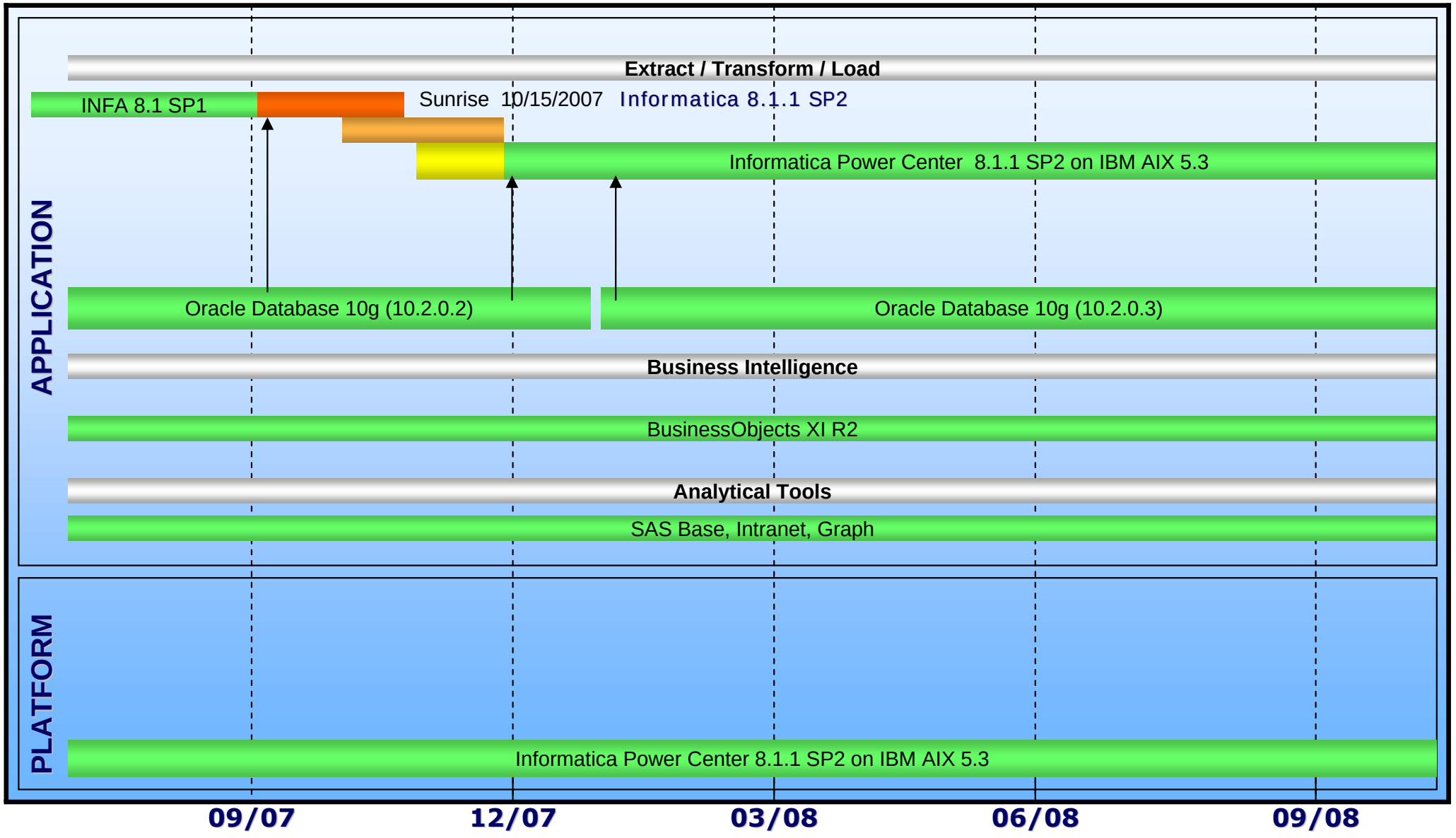
Technology Roadmap: Portal



Notes/Next Steps

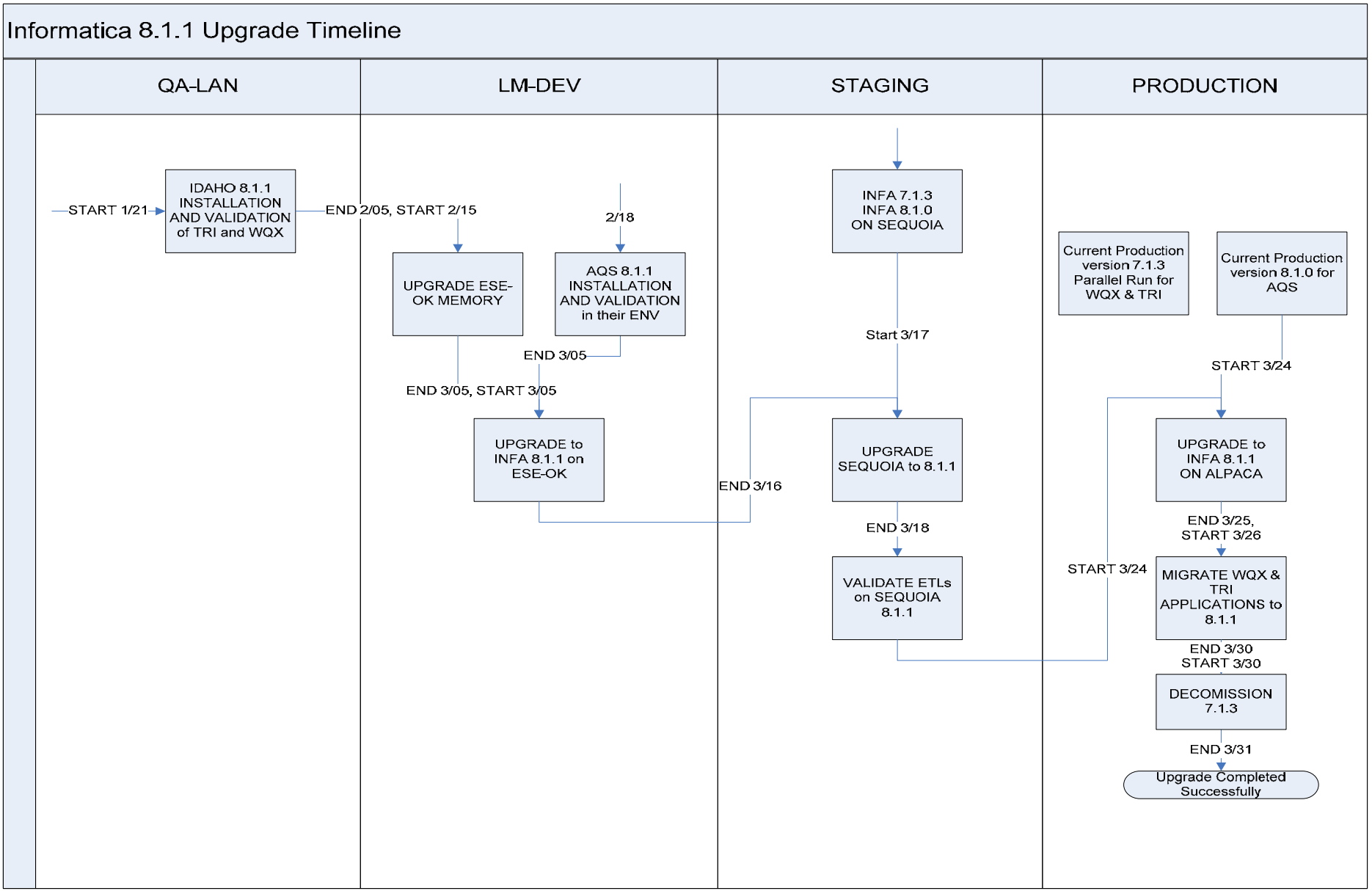
- The following Portal communities would like to leverage new features in Java Enterprise Edition (JEE) 1.4 and Portal 10.1.4 that help in reducing the development efforts
  - Clean Watershed Needs Survey (CWNS) (TO 32)
  - Information Technology Governance Support System (ITGSS) (TO 03)
  - Office of Emergency Management (OEM) (TO 31)
- Planning for next 12 months
  - Recommend an Integrated Project Team (IPT) with all stakeholders EPA National Computer Center (NCC), Lockheed Martin (LM), Computer Sciences Corporation (CSC), and Oracle (as needed)

Technology Roadmap: BIA/ETL



Note: Development and Test is incremental & iterative

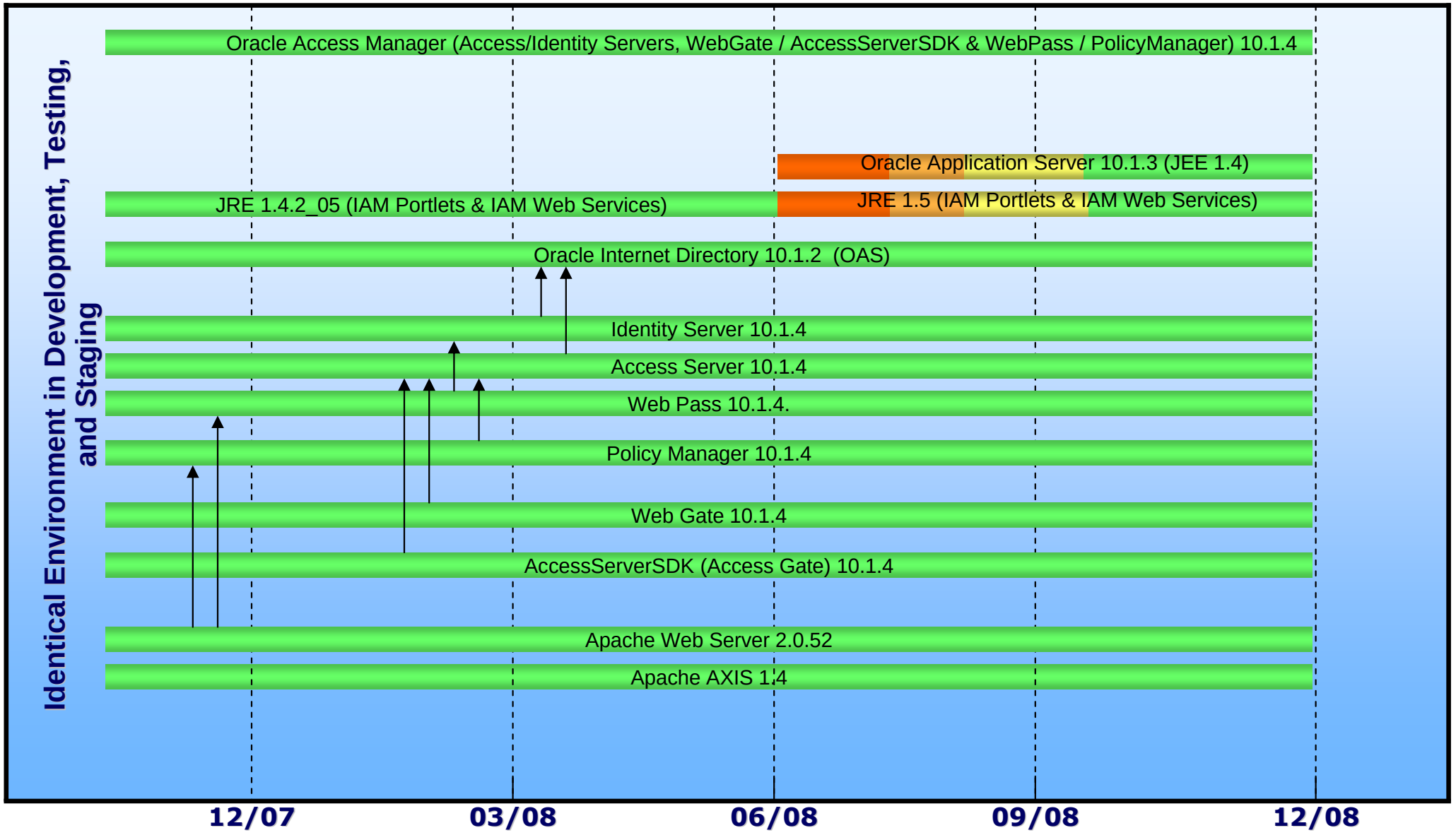
INFORMATICA MIGRATION SUMMARY



Notes/Next Steps

- **ACCOMPLISHED**
  - **Migration to v8.1 SP1**
- **NEXT STEPS**
  - **Plan & Install 8.1 SP4**
  - **Update NCC Roadmap to show SP4 Upgrade**

Technology Roadmap: IAM

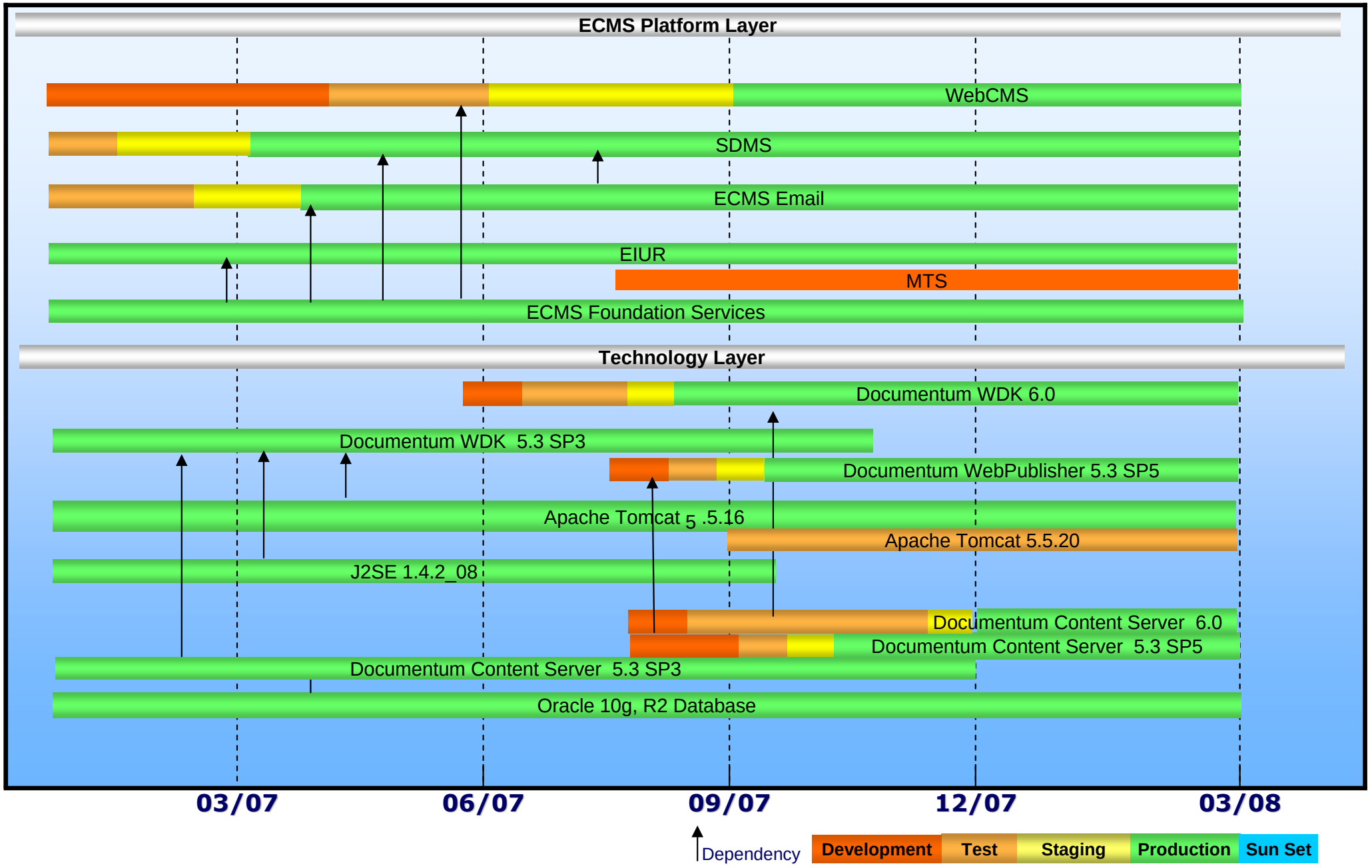


Note: Development and Test is incremental & iterative

Notes/Next Steps

- The following applications are either integrated or currently being integrated into IAM Solution
  - CWNS (TO 32) including External Document Management System (EDMS) Integration
  - OEM (TO 31)
  - Facility Registry System (FRS) Web (TO 51)
  - SDMS (TO 45)
- Planning for next 12 months
  - Recommend an IPT with all stakeholders EPA NCC, LM, CSC, and Oracle (as needed)

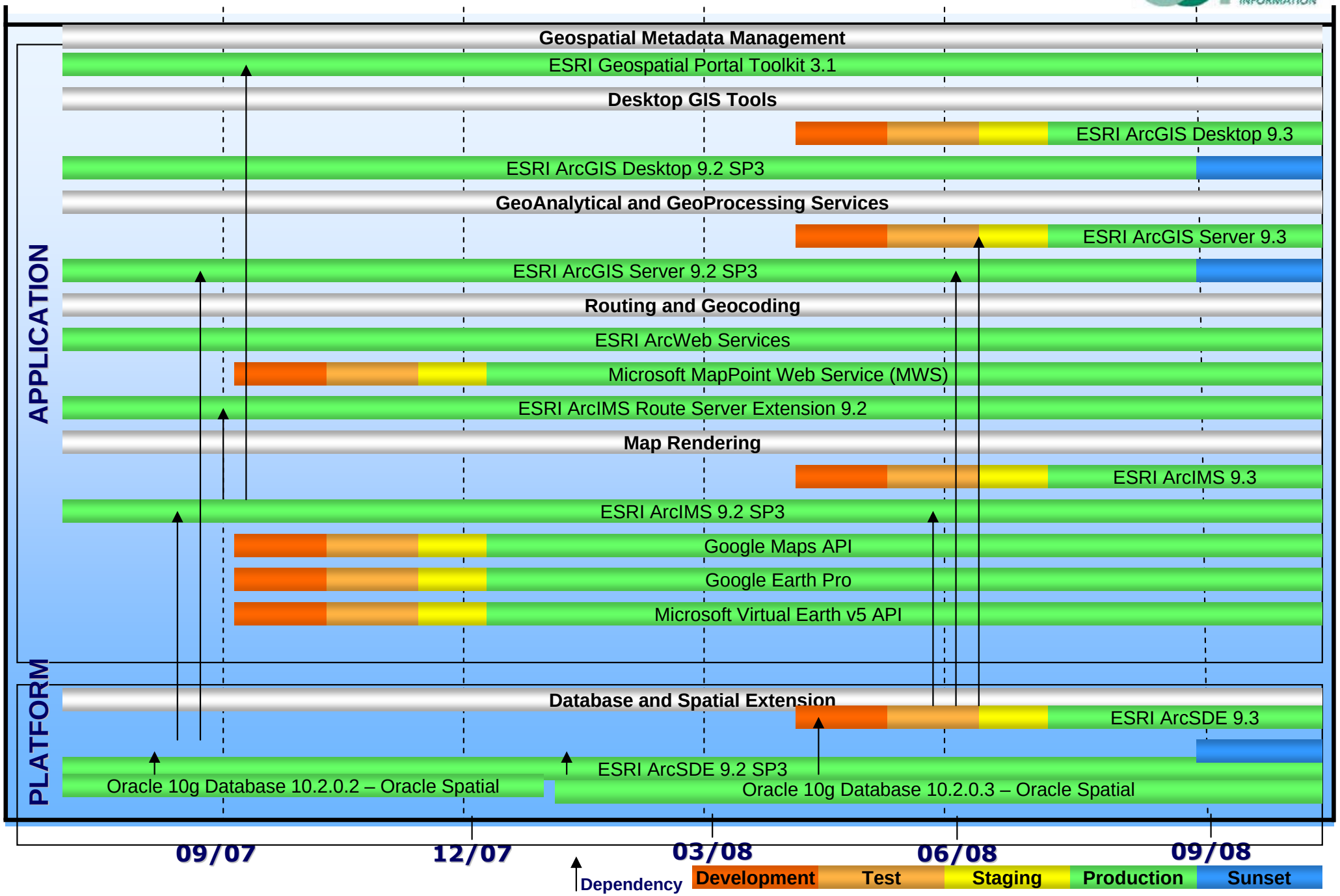
Technology Roadmap: ECMS



Notes/Next Steps

- ECMS is a platform of EMC/Documentum Commercial off the Shelf (COTS) products and applications that leverage functionality provided via the ECMS Foundation Services.
- COTS products installed include: Content Server, DA, WebTop, RMA, Records Manager, WebPublisher, Site Delivery Services, Web Development Kit (WDK), Content Intelligence Services, Documentum Transformation Services.
- Discrepancies with the NCC Technology Upgrade Schedule (TUS) need to be resolved.
  - NCC TUS lists 5.3 SP4. SEC is targeting 5.3 SP5 and D6 releases.

Technology Roadmap: Geospatial



Note: Development and Test is incremental & iterative

NOTES-Product Versions

- **ESRI ArcSDE/ArcIMS/ArcGIS Server/ArcGIS Desktop – EPA Roadmap lists:**
  - ArcSDE 8.1 as Standard: http://basin.rtpnc.epa.gov/ntsd/ITMatrix.nsf/ReportView/2BE1FB6B5976E3FB85256C7900625946?opendocument
  - ArcIMS 3.1 as Standard:
  - *ArcSDE, ArcIMS, ArcGIS 9.3 versions all scheduled for release 1Q08; available via ESRI ELA at no additional cost*
- **ESRI ArcIMS Route Server Extension – *Category B product, discounted within ESRI ELA, not anticipated for upgrade in the near future***
- **Oracle DB / Oracle Spatial – *EPA Roadmap lists:***
  - *Oracle 8i as Legacy: http://basin.rtpnc.epa.gov/ntsd/ITMatrix.nsf/ReportView/9177EC838901D2B785256C790062592D?opendocument*
  - *9i as Standard: http://basin.rtpnc.epa.gov/ntsd/ITMatrix.nsf/ReportView/947CF240200C7E0585256C790062592E?opendocument*
  - *Oracle Spatial as Standard: http://basin.rtpnc.epa.gov/ntsd/ITMatrix.nsf/ReportView/49160038D79F456085256C7900625947?opendocument*
  - *Oracle 10g as Target: http://basin.rtpnc.epa.gov/ntsd/ITMatrix.nsf/ReportView/55A478E38EA8B78685256E7F005605BD?opendocument*
  - *Versions predicated on individual licenses, upgrades entail additional cost; 11 not anticipated to be implemented within current horizon; Regions still maintaining older versions (7,8,9) and Microsoft SQL Server*
- **ESRI ArcWeb, Microsoft MapPoint, Virtual Earth, Google Earth Pro, Google**