

Securing
Communication with
the future



 **SHAMAN**
Sustaining Heritage Access through Multivalent ArchiviNg



Co-funded by the European Union

SHAMAN will provide a next-generation **Digital Preservation Framework**.

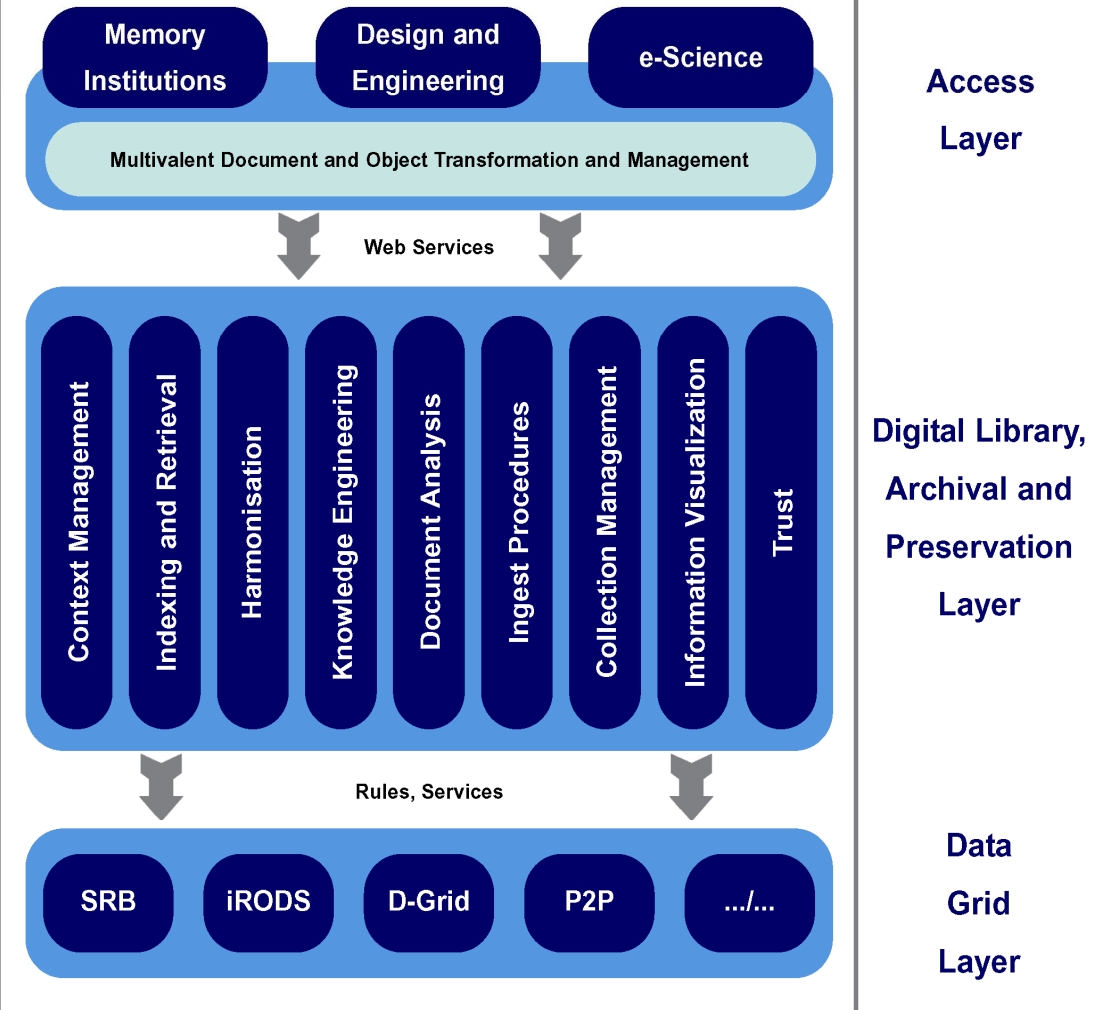
It develops exemplary **Application Prototypes** to investigate advantages and impacts of integrating SHAMAN's **New DP Component Technologies** as well as legacy into **New Application Solutions** along SHAMAN's **Reference Architecture** that **extends OAIS**.

Validation of the SHAMAN framework viability will be focused in **three Application Trial Domains**:

- scientific publishing in libraries and documents in governmental archives
- digital objects used in industrial design and engineering processes
- data resources used in e-Science applications



SHAMAN Digital Preservation Environment Framework



SHAMAN's PCAs supporting Interoperability

PCA	PCA NAME	PCA COORDINATION
1	Distributed Resource Management Infrastructure Framework and Gridbased Resource Integration	Jose Borbinha (INESC-ID)
2	Contextual and Multivalent Archival and Preservation Processes	Matthias Hemmje (Univ. of Hagen)
3	Semantic, Constraint-based Collection Management Systems	Jean-Pierre Chanod (XEROX RCE)
4	Managing Future Requirements	Adil Hasan (Univ. of Liverpool)

SHAMAN's WPs and PCAs supporting Component R&D

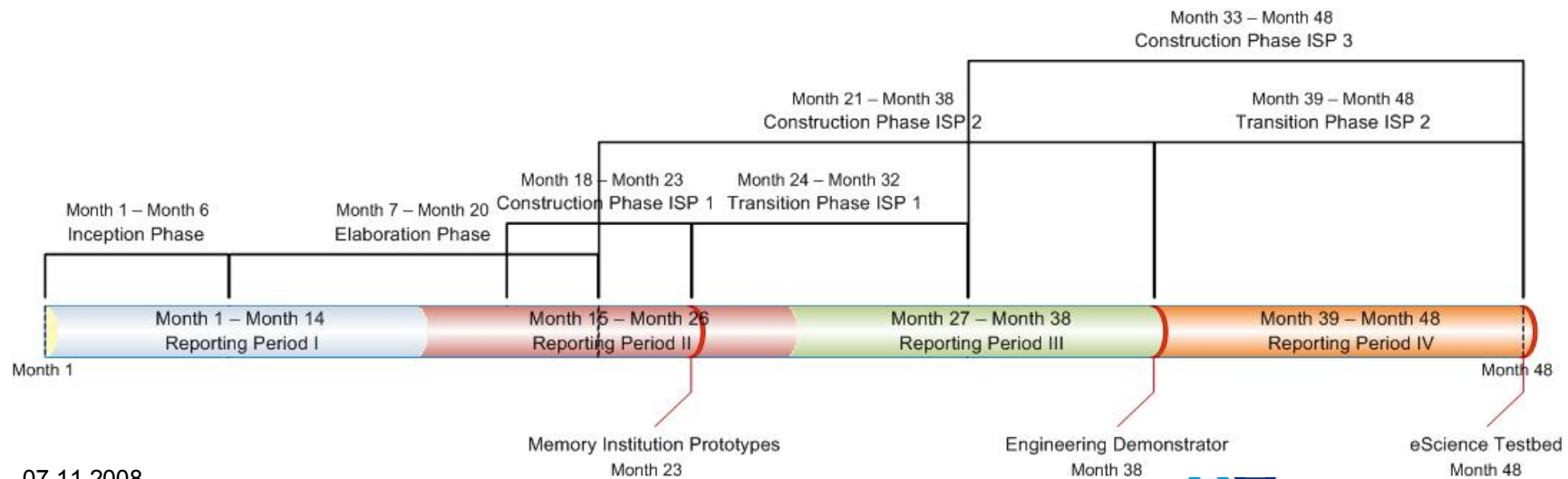
WP 1	Requirements Analysis and Identification of User Scenarios	Perla Innocenti (HATII, Univ. of Glasgow),	PCA1
WP 2	Design and Specification of the SHAMAN Digital Preservation Framework	Milena Dobрева (U. Strathclyde, Glasgow)	
WP 5	Data Grid Implementation	Martin Mois (Univ. of Hagen)	
WP 3	Context Capturing, Representation, and Management	Claus-Peter Klas (Univ. of Hagen),	PCA2
WP 4	Multivalent Preservation Interface and Media Engines	Paul Watry (Univ. of Liverpool),	
WP 6	Harmonisation, Basic Analysis and Ingest	Jean-Pierre Chanod (Xerox RCE)	
WP 7	Advanced Information Extraction and Knowledge Engineering	Jean-Pierre Chanod (Xerox RCE)	PCA3
WP 8	Managing Shared Collections	Jens Ludwig (SUB)	
WP 9	Interoperability with Future Environments	Paul Watry (Univ. of Liverpool)	PCA4
WP 10	Maintaining Essential Properties	Jana Dittmann (Univ. of Magdeburg)	

SHAMAN's ISPs – Supporting Cohesion, Integration, Evaluation, and Demonstration

WP 11	Document Production, Archival, Access and Reuse in the Context of Memory Institutions for Scientific and Governmental Collections	Alfred Krahnstedt (DNB)	ISP1
WP 12	Simple and Connected Object Production, Archival and Reuse in the Industrial Design and Engineering Domain	Andreas Hundsdörfer (InConTec)	ISP2
WP 13	eScience Data-Acquisition and Harmonisation Testbed	Jose Borbinha (INESC/ID)	ISP3

SHAMAN's Integration & Demonstration Subprojects (ISPs)

- Foster Systematic Evolution of Project Results...
 - **ISP 1** – Document Production, Archival, Access and Reuse in the Context of Memory Institutions for Scientific and Governmental Collections
 - **ISP 2** – Simple and Connected Object Production, Archival and Reuse in the Industrial Design and Engineering Domain
 - **ISP 3** – eScience Data-Acquisition and Harmonisation Testbed
- Horizontal Integration of RTD Contributions

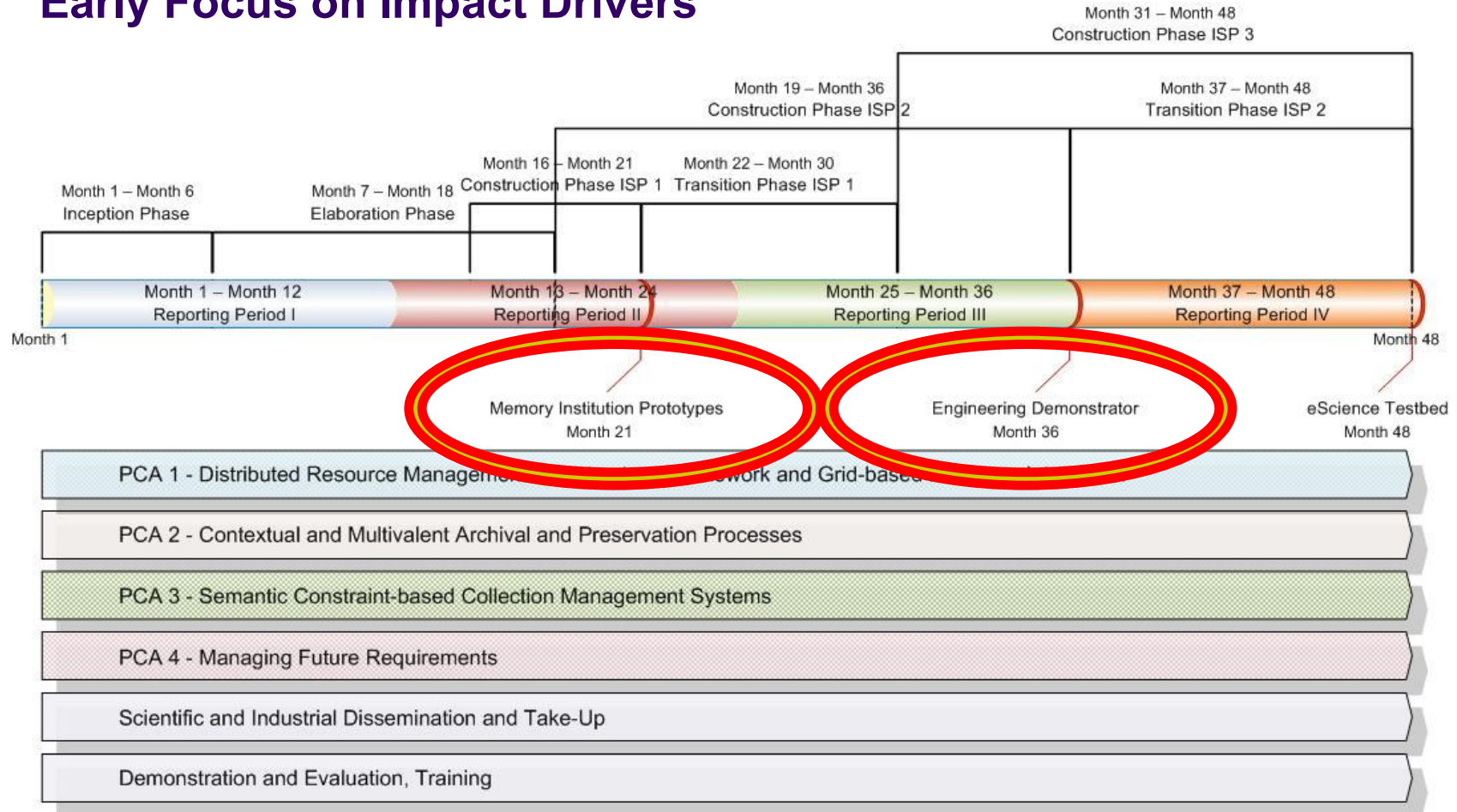


07.11.2008

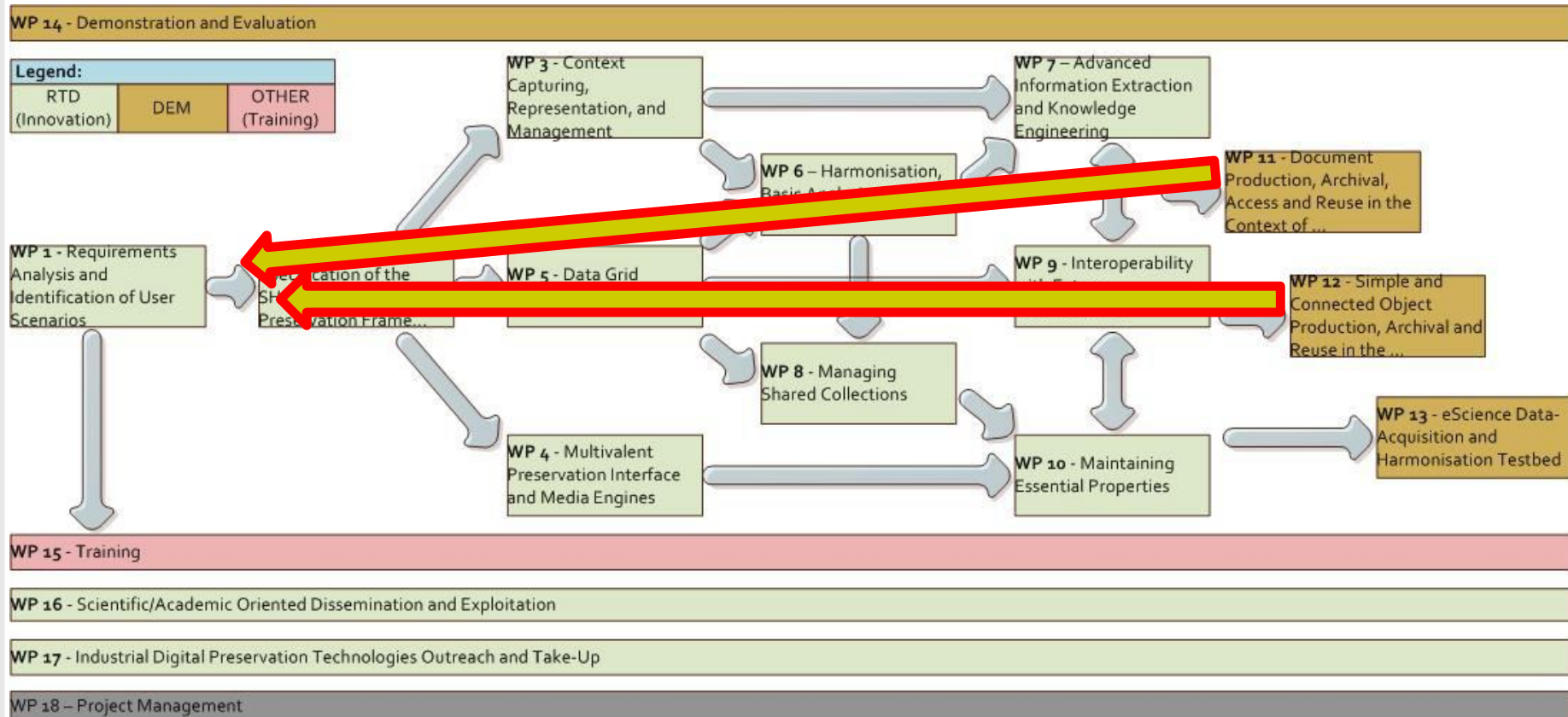
**Fine, that is the SHAMAN Project structure ...
but how will the project work**



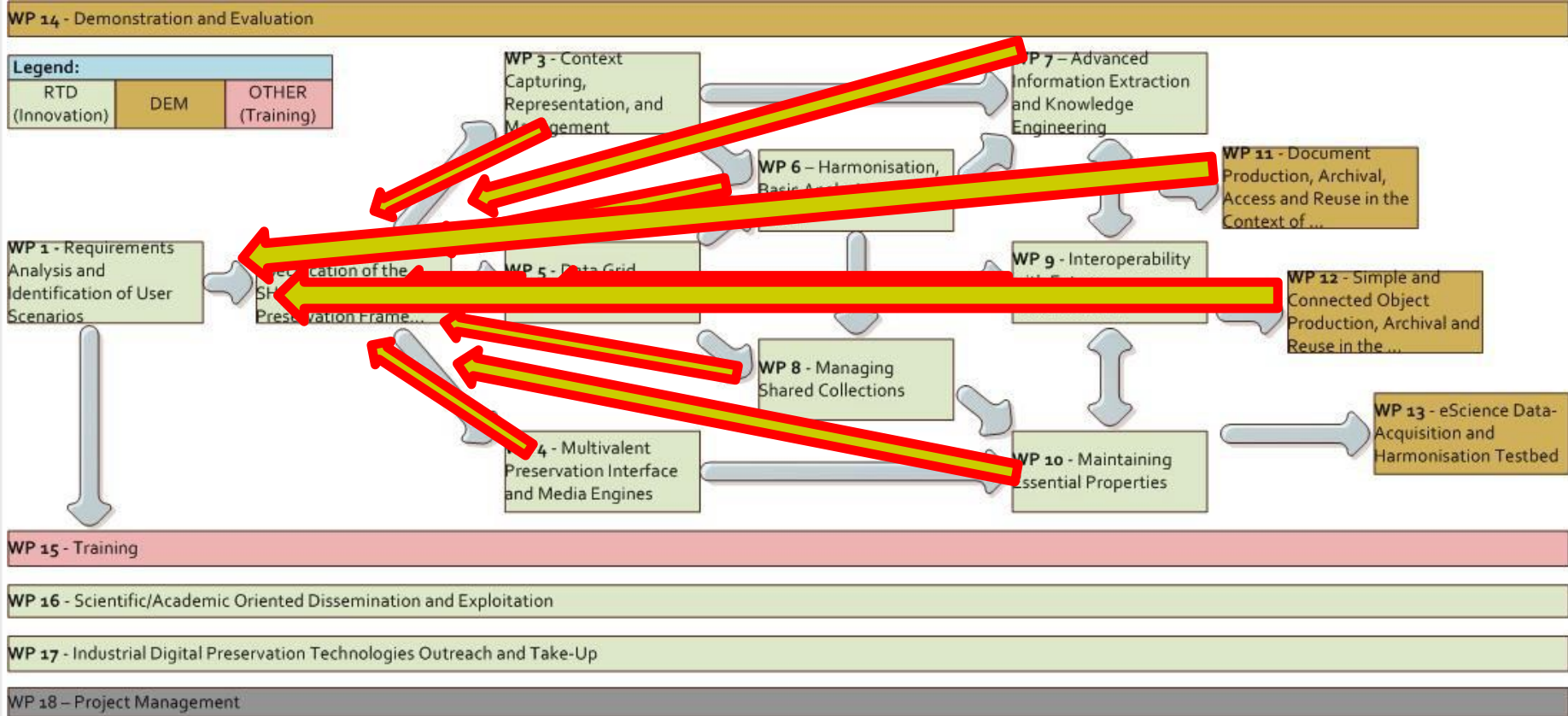
Strategic R&D Impact Steering: Early Focus on Impact Drivers



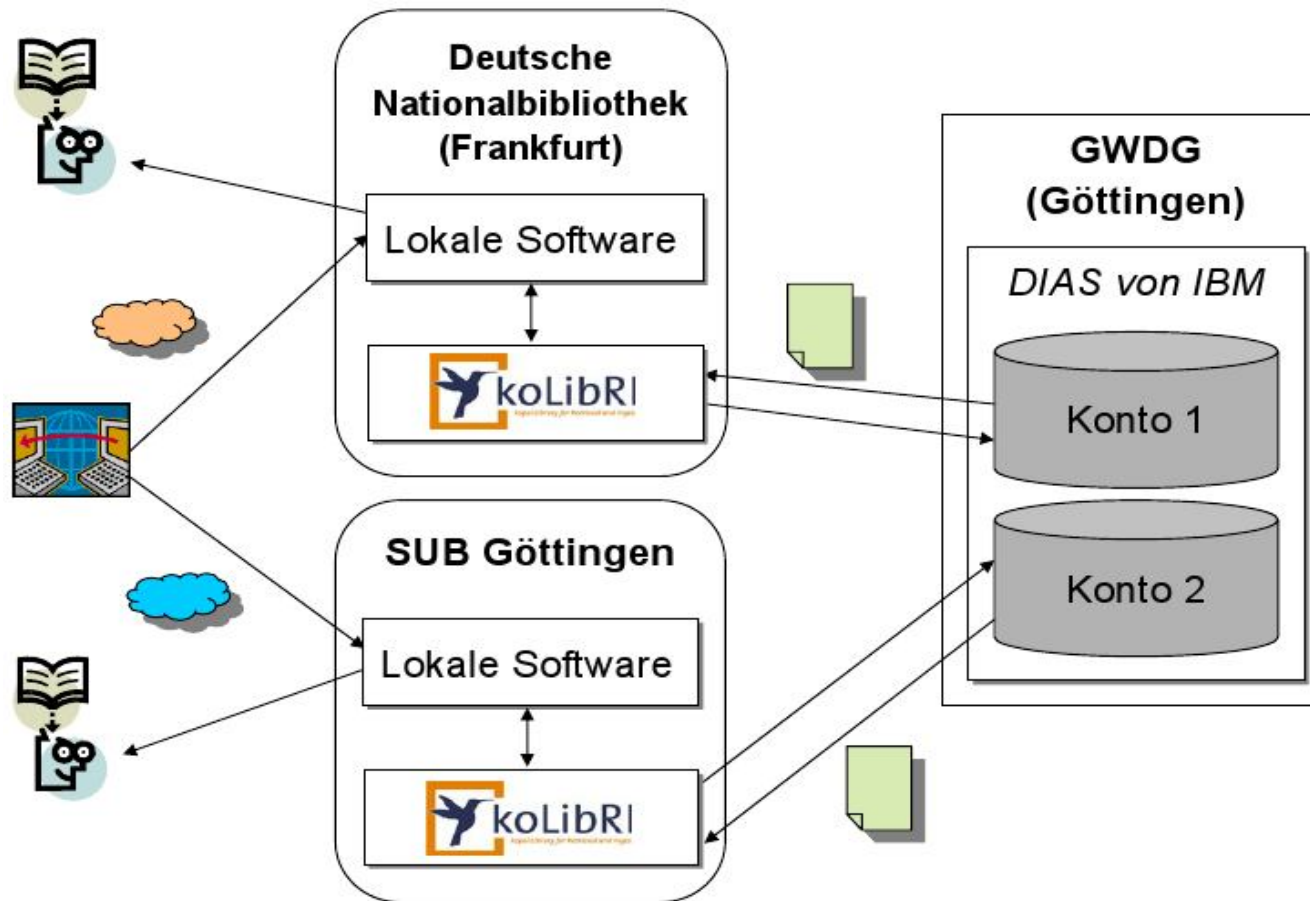
Tactical R&D WP Steering (I): Early Operational Consequences for Requirements Analysis in WP1



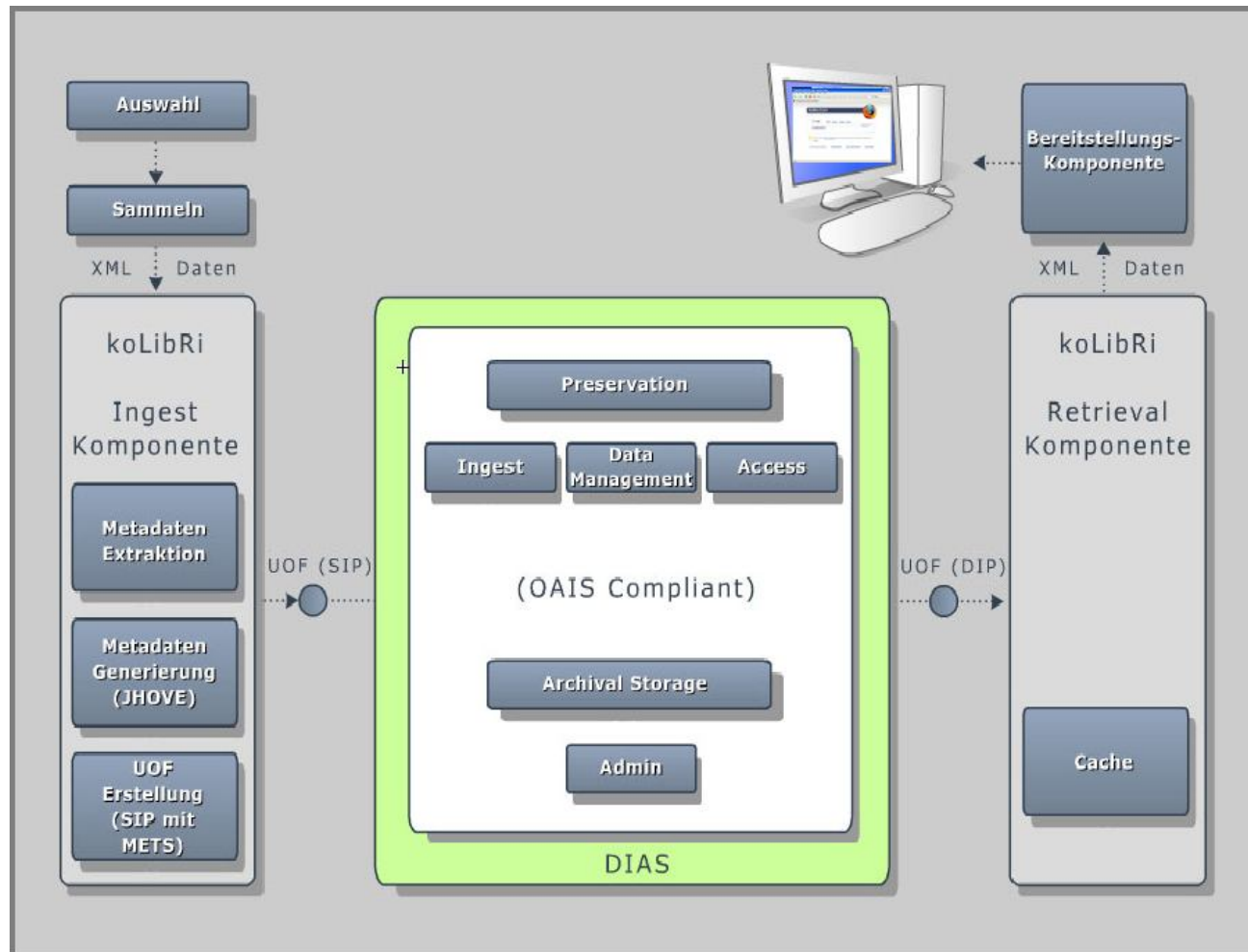
Tactical R&D WP Steering (II): Early Operational Consequences for the Reference Architecture in WP2



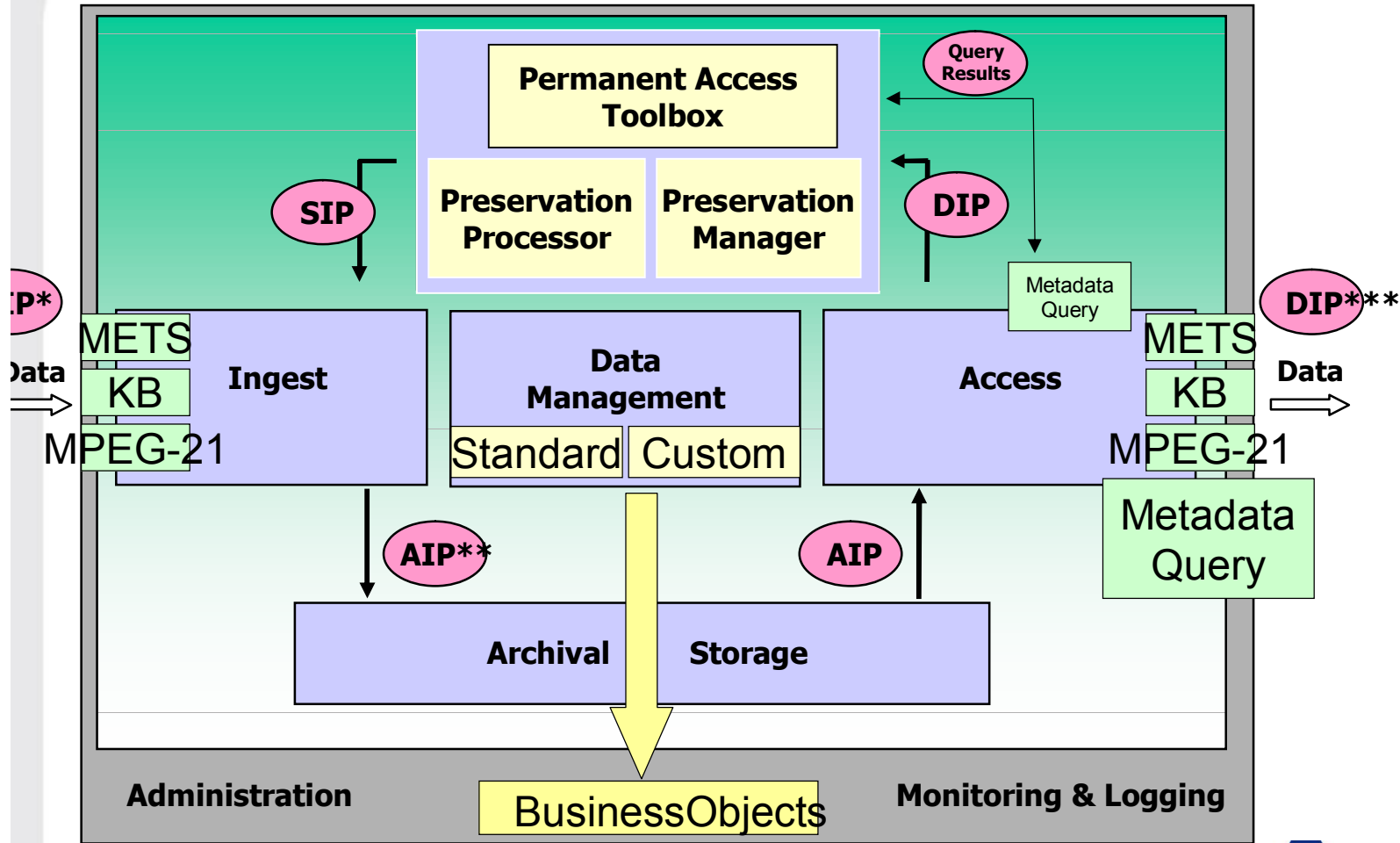
SHAMAN's legacy in ISP-1 (I): KOPAL (DNB, SUB)



SHAMAN's legacy in ISP-1 (II): KOLIBRI (DNB, SUB)



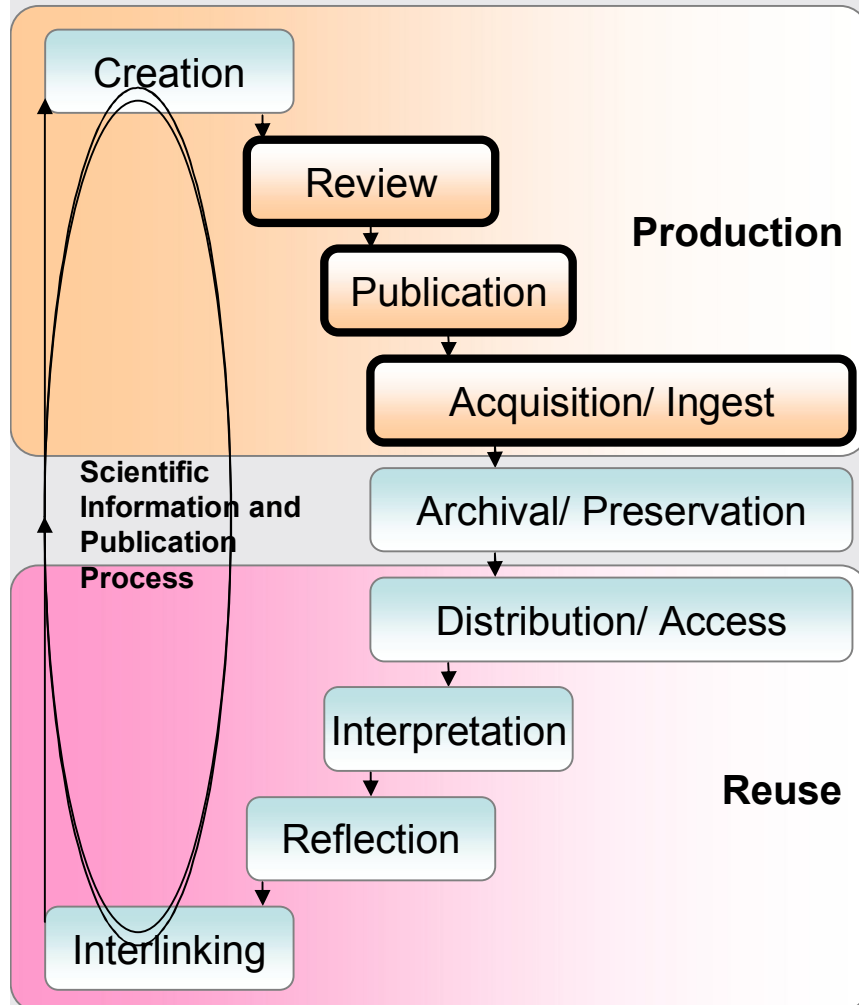
SHAMAN's legacy in ISP-1 (III): DIAS (IBM)



* = SIP: Submission Information Package
 ** = AIP: Archival Information Package

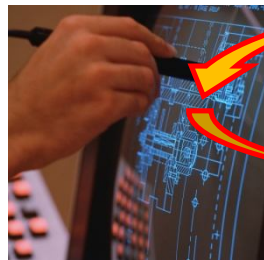
*** = DIP: Dissemination Information Package

SHAMAN's requirements analysis in ISP-1 (IV): Scientific Publication P&R Context

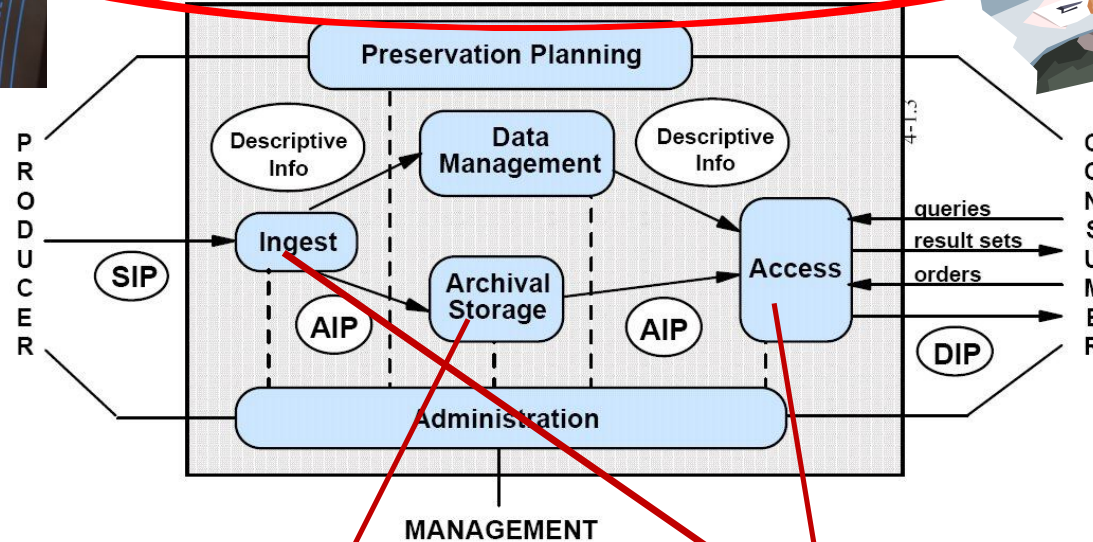


- Scientific congress publication process can make available rich set of information to the *reuse context*
- Scientific community web publishing and DL application CO CONGRESS ONLINE ® can be extended to capture context data beyond the immediate requirements of scientific event organization

Usage Scenario in ISP-2 (I): D&E Scenario and the OAIS Reference Model



Design & Engineering P&R Context

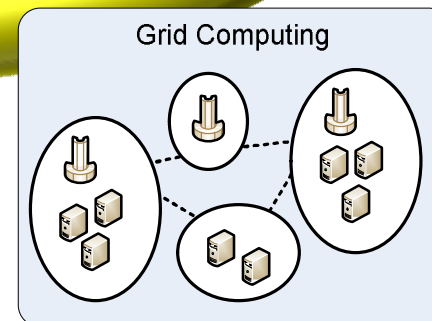
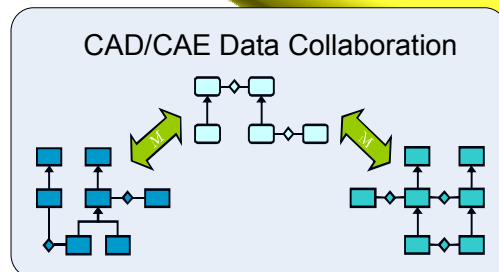
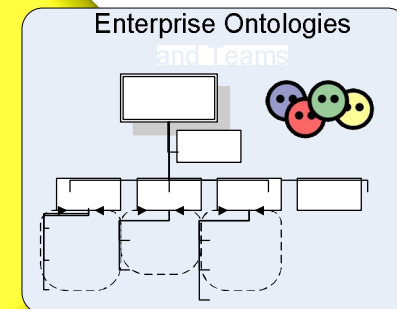
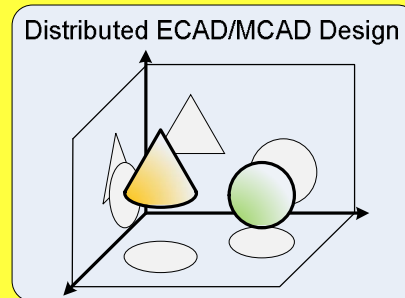
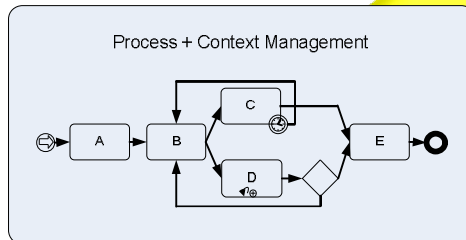
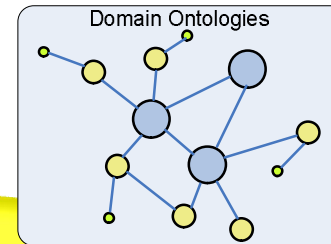
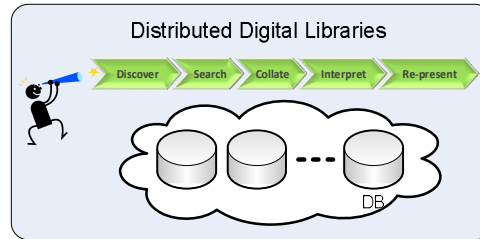


Multivalent approaches preserving CAD models

- migration
- emulation

Providing ingestion and access in distributed heterogeneous industrial engineering szenarios with collaboration support

ISP-2 (II): R&D Dimensions of the Engineering Scenario



Basic Research Challenges in SHAMAN

- **Theory of Preservation:** that may be used to store and access potentially any type of data, based on the integration of digital library, persistent archive, and data management technologies.
- **Infrastructure** for long-term preservation and reuse of data over a decades-long time span.
- **Grid-based production system** that will support the virtualization of data and services across scientific, engineering, document, and media domains.
Identifying Content and Capturing of Context Demonstrate Distributed Ingestion

First Investigations for Embedding Legacy Environments and Application-Domain oriented Use Cases into a Grid-Based Preservation Infrastructure

Towards SHAMAN's Framework Infrastructure and its Reference Architecture

ISP 1 Scenario – Use Cases

- Information Integration
 - Mediated search within distributed repositories
 - Transparent (read) access to legacy systems
 - How to find information on all integrated systems ?
- Distributed Ingestion
 - Local ingest processes are registered for updating federated metadata catalogue/index
 - How to cover local and global data ingestion, such that information can still be found ?
- Managing Distributed Collections
 - Implementation of replication mechanisms
 - User requirements mediated to underlying storage infrastructure
 - How to enable data management over legacy systems?

Challenges of Memory Institutions for SHAMAN within ISP1

Embedding legacy environments into a grid-based preservation infrastructure

Types of legacy systems:

- Data Grids
- Institutional Repositories
- Archival Systems
- Access Systems
- Digital Libraries

Requirements:

- Integrity
- Authenticity
- Search&Browse
- Interpretability
- Virtualization

Types of technology

- Data Grid: iRODS/SRB
- DSpace
- Fedora
- KOPAL
- DAFFODIL

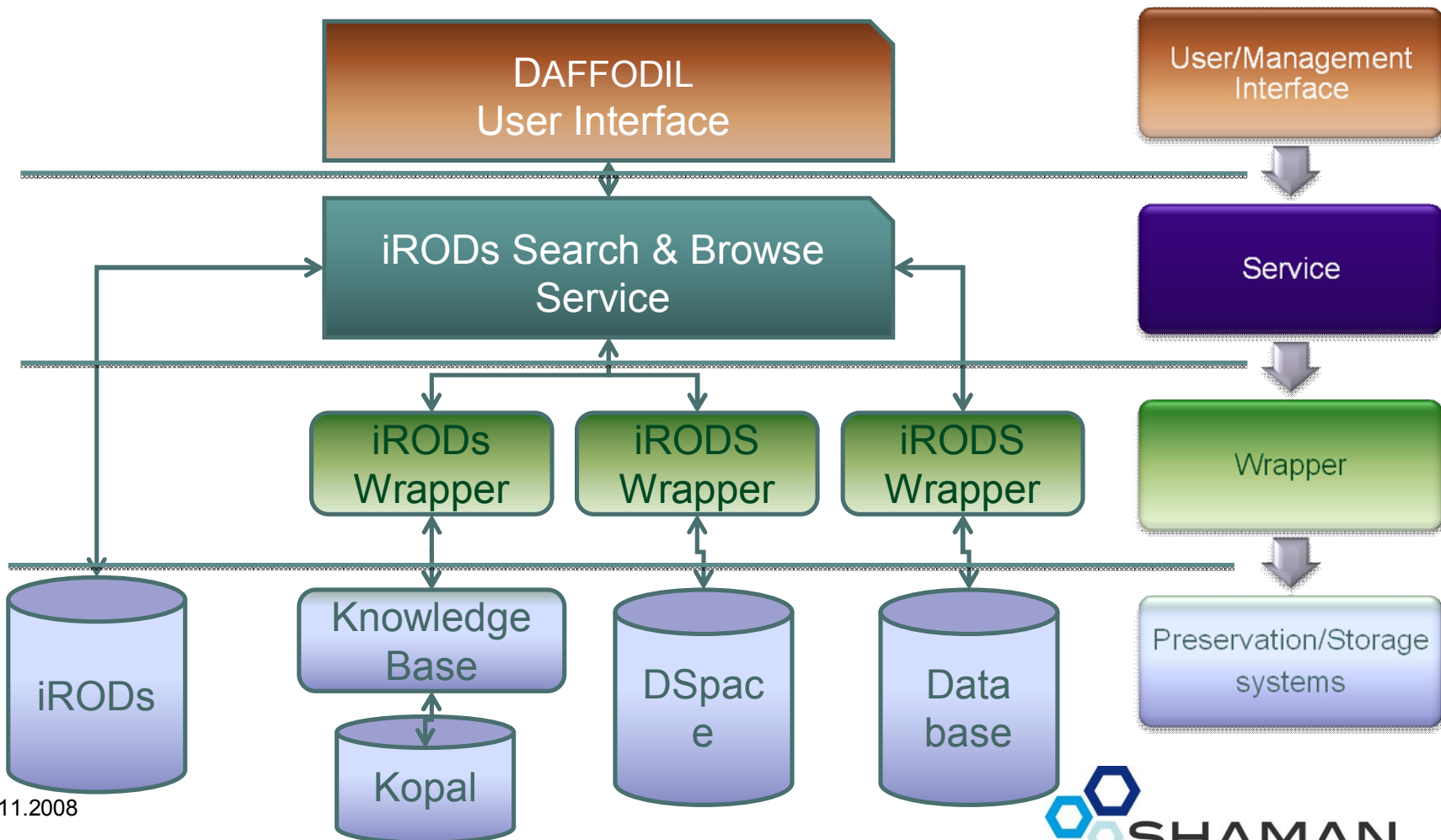
Usecase 1: Information Integration

Search & Browse in distributed heterogeneous resources

- Problem
 - Different access points to all resources
 - Possibly different user interfaces and
 - different query forms
 - Heterogeneous metadata standards
- Solution
 - One access point
 - Central user interface & query form
 - Transparent access to all legacy systems

Information Integration of Legacy Systems (I)

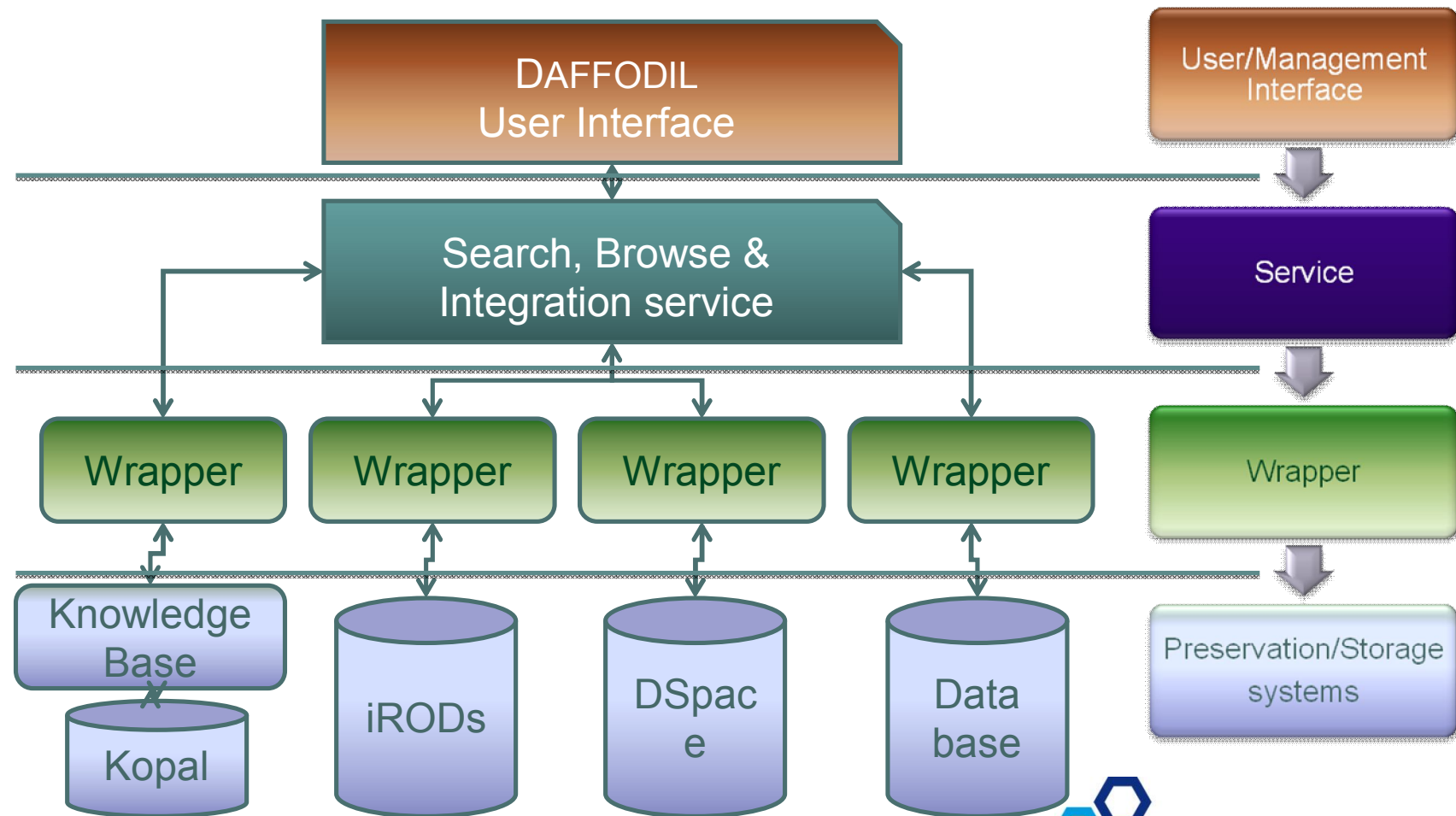
Search and browse in heterogeneous environments



07.11.2008

Information Integration of Legacy Systems (II)

Search and Browse in heterogeneous environments



07.11.2008

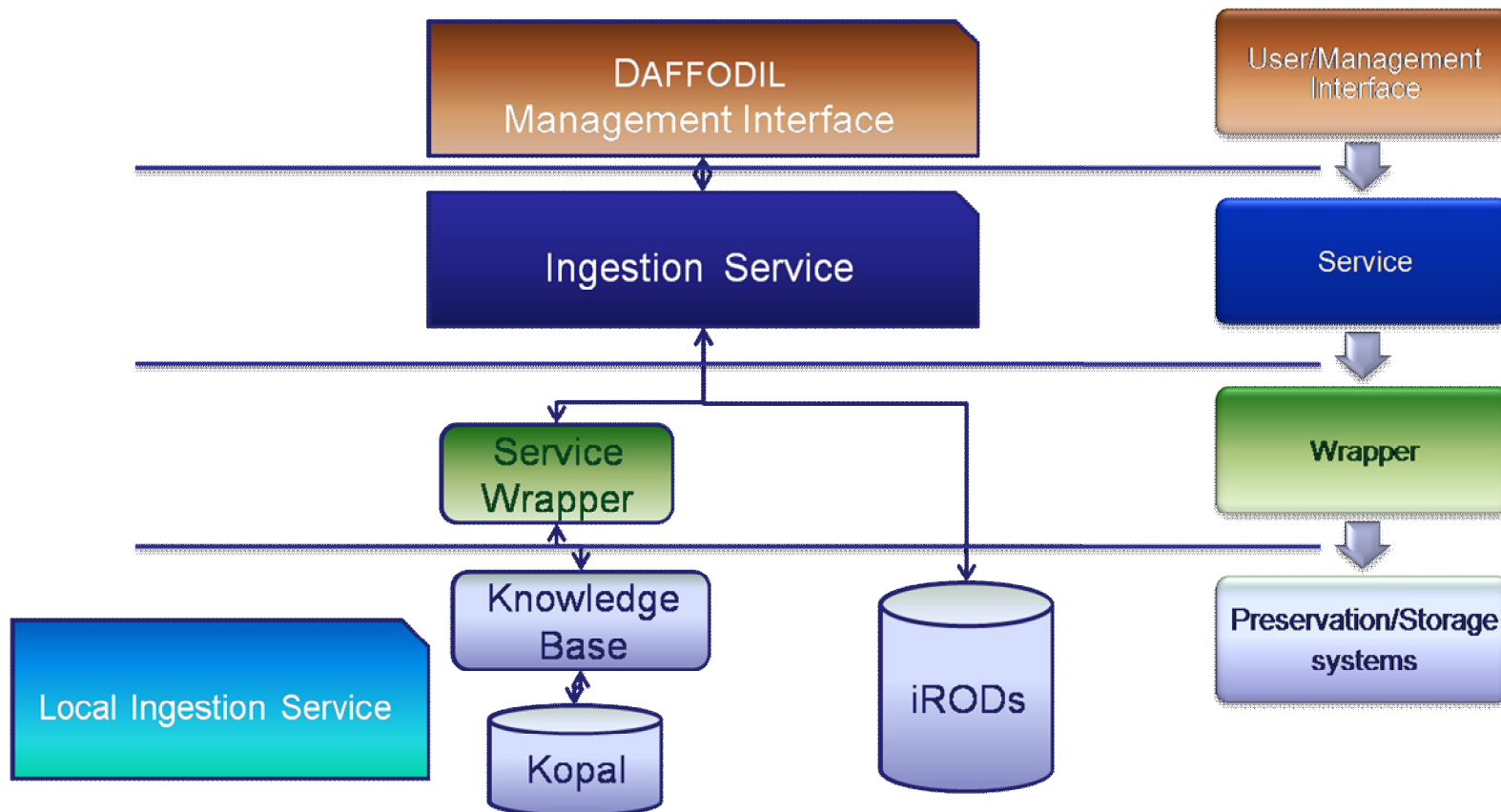
Usecase 2: Distributed Ingestion

Ingestion in distributed heterogeneous resources

- How to cover local and global data ingestion, such that information can still be found ?
- Local ingestion:
 - Notification push/pull
 - Double ingestion (full object)
 - Only metadata ingestion

Distributed Ingestion

Parallel storage, index only, full object store



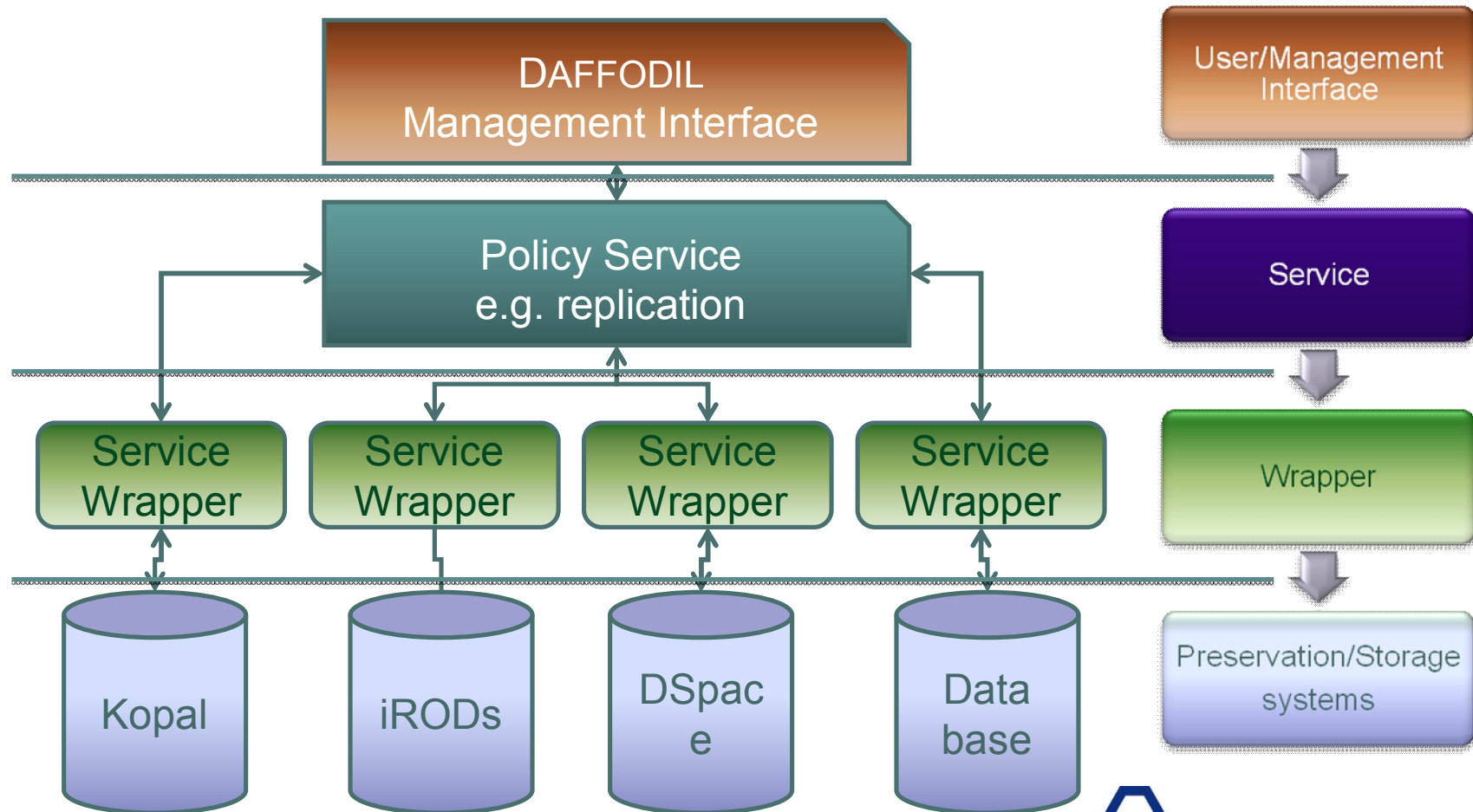
07.11.2008

Resulting Technical Challenges of Managing Distributed Collections

- Management of distributed collections in grid environment with all legacy systems, based on policies.
- Replication of information in world wide distributed data and storage grids to prevent distruction, e.g., in case of disasters
- UC1 & UC2: Wrapper & services need only be aware of local environment
- Here: New mediator level, aware of all repositories and their current status!

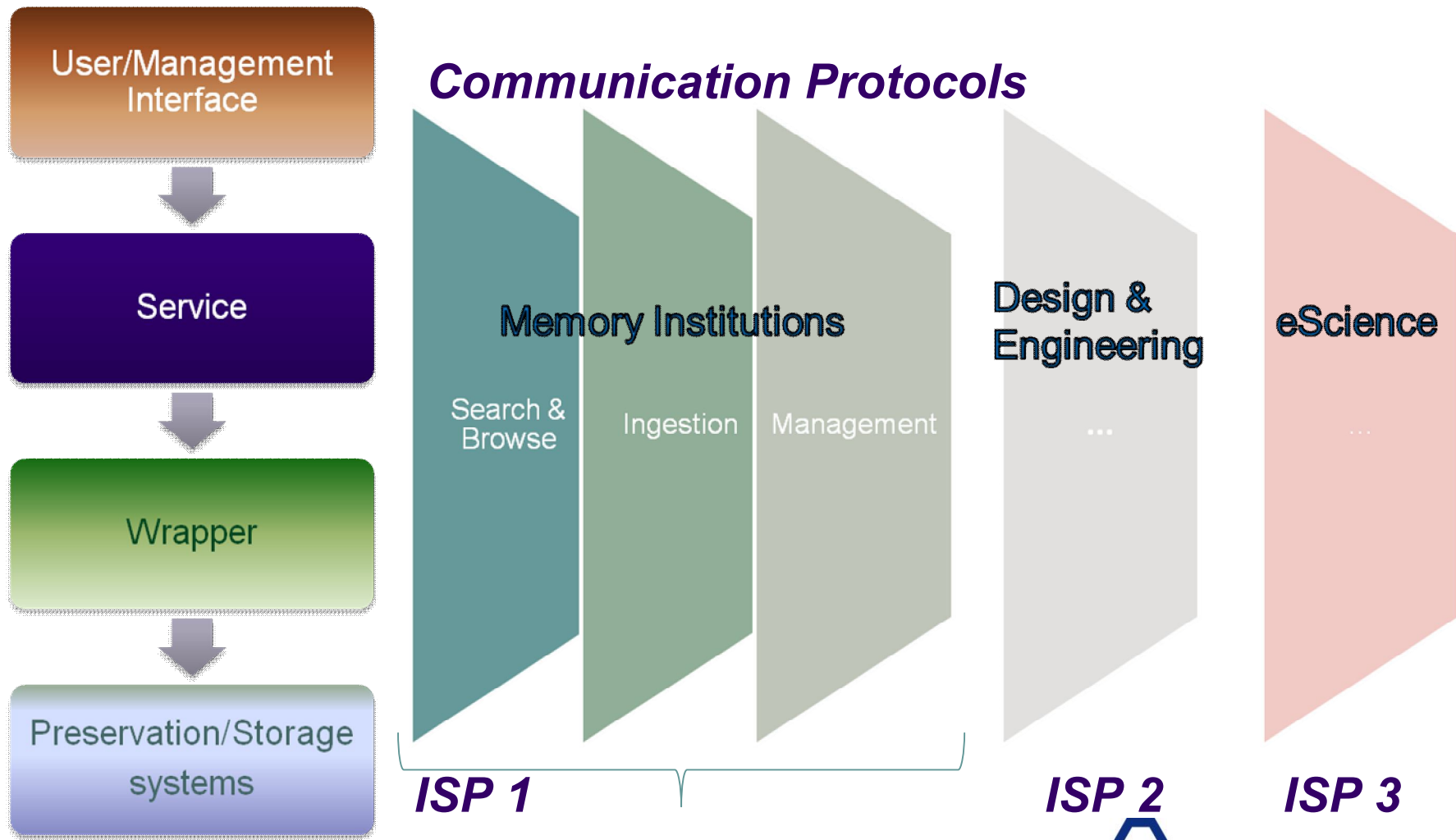
Managing Distributed Collections

Replicate stored objects according to user requirements



07.11.2008

Constructing SHAMAN's Service-Oriented DP Reference Architecture



07.11.2008

Fine.

Thank you very much for your attention

07.11.2008

SHAMAN consortium



Deutsche Nationalbibliothek



SHAMAN collaborators

VLAAMS PARLEMENT



Welcome to the future. Welcome to SHAMAN.

<http://www.shaman-ip.eu/>