



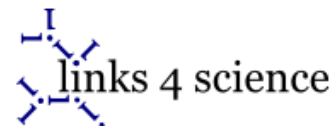
SHAMAN

Sustaining Heritage Access through Multivalent Archiving

Matthias Hemmje, University of Hagen, Germany

Call 3 Information Day, Digital Preservation, Results of Call 1

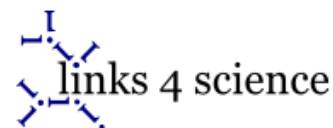
Luxembourg 17th of December 2007



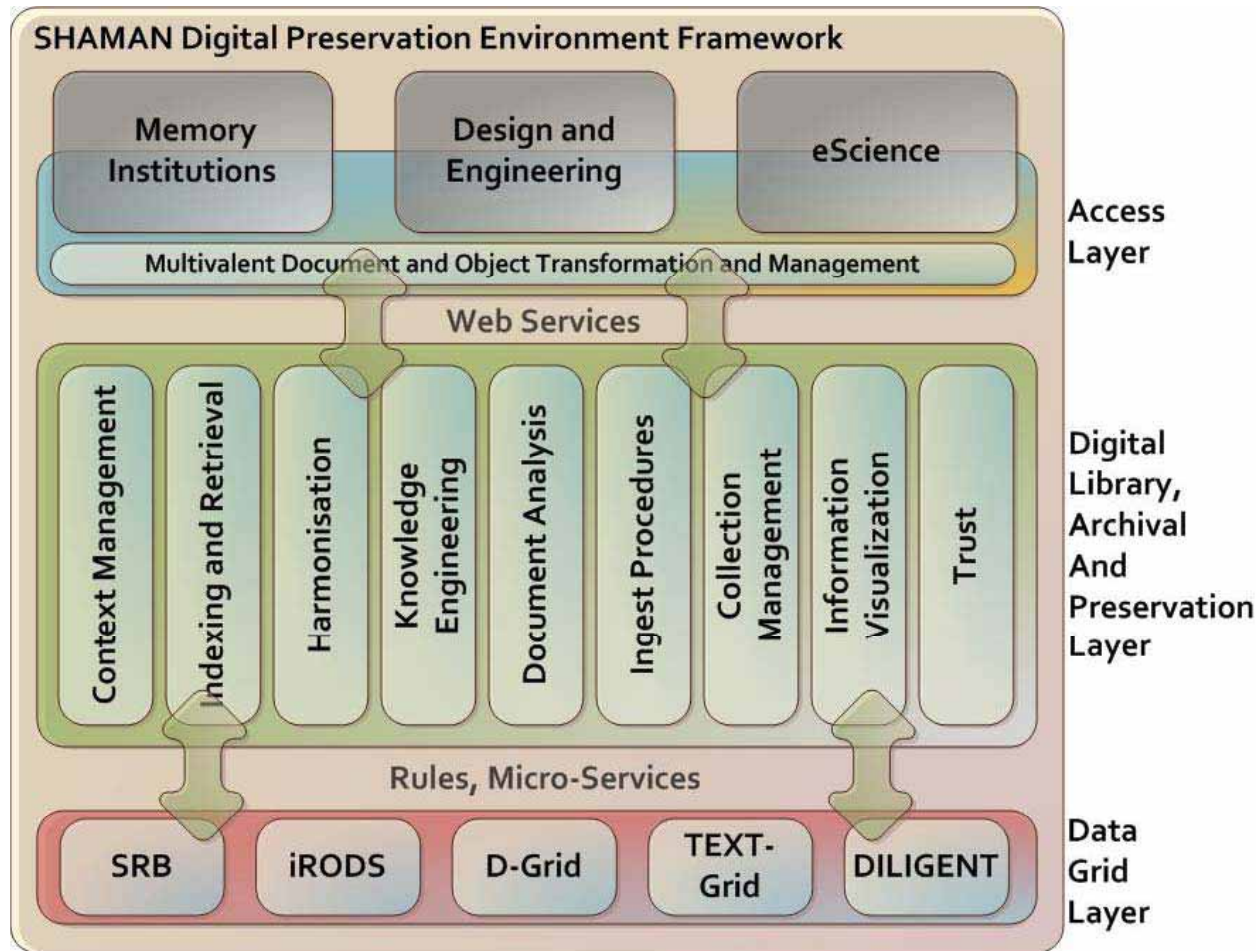


SHAMAN PROJECT PROFILE

- ✓ **Large-Scale Integrating Project**
- ✓ **Main output: Long Term Next Generation DP Framework & Application Solution Environments**
 - 3 Prototype Application solutions:**
 - **Scientific Publishing & Memory Institutions**
 - **Industrial Design & Engineering**
 - **E-Science**
- ✓ **48 months duration**
- ✓ **1.300 PM effort by a team of 60 persons (22W/38M)**
- ✓ **18 Partners from 9 countries**
- ✓ **3 MI + 3 MNC + 3 SME + 9 UNIV (50 - 50 % industry/Academia)**



R&D (I): Conceptual Architecture & Grid Utilization





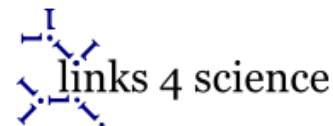
R&D (II): Supporting Storage and Curatorship – Distribution vs. Centralization

Organizational aspects

- the SHAMAN framework can support centralized as well as distributed collection **storage**, indexing, and analysis as well as hybrid forms of these
- the SHAMAN framework can support centralized as well as distributed **curatorship** as well as hybrid forms of both
- All the above mentioned combinations require appropriate global as well as local policies and mutual agreements on the organization level

Possible implications

- the SHAMAN framework can equally well drive application solutions which are tailored towards user community requirements of
 - storage and access distribution as well as
 - individual organizational & curation policies as well as
 - relationships amongst collaboratively curating institutions,
- this means the SHAMAN framework can help bridge gaps between individuals, teams, institutions/organizations, and working cultures within DP application domains





R&D (III): Enabling Migration, Emulation & Hybrid Approaches

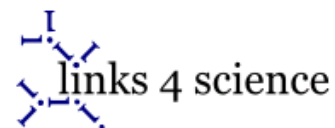
SHAMAN Supports Migration through defining preservation policies which describe:

- the objects of the sub-collection to be migrated
- rules by which the object formats are migrated
- the relevant set of formats related to the meta-data and process contexts
- rules by which the meta-data and process formats are migrated

SHAMAN Supports Emulation policies through defining preservation policies which describe:

- the display technologies of the objects of the sub-collection to be migrated
- the access&re-use technologies for the relevant meta-data contexts
- the access technologies for the relevant process contexts
- rules by which these technologies are emulated

In addition, SHAMAN Supports Hybrid Approaches, too.





R&D (IV): Utilizing UVC and Going beyond with Multivalent

Lorie's UVC approach relies on

- utilizing a simple virtual machine with storage and processing capacities
- and therefore can be used to represent software on a very general level

Multivalent Technologies go beyond the idea of the UVC approach by

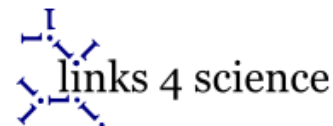
- utilizing abstract representation mechanisms for documents and media-objects including their structural relationships
- providing application-oriented as well as platform- and vendor-independent tools and libraries to visualize, access, & reuse these documents and media objects
- and therefore productively support the application domain of content engineering and preservation

If UVC-based technologies should become available in production quality, **SHAMAN can take advantage of both**

- Multivalent to support Migration and
- UVC to support Emulation

This means, in the ideal case, **future Multivalent Technologies** could be

- migrated to the emerging UVC technologies or
- emulated on the basis of UVC technologies



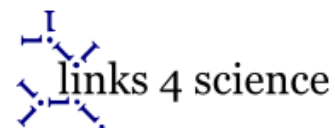
R&D (V): WPs and PCAs supporting Component R&D

WP 1	Requirements Analysis and Identification of User Scenarios	Perla Innocenti (HATII, Univ. of Glasgow), Heike Neuroth (SUB, Göttingen)	PCA1
WP 2	Design and Specification of the SHAMAN Digital Preservation Framework	Gobinda Chowdhury (U. Strathc., Glasgow), Robert Sanderson (Univ. Liverpool)	
WP 5	Data Grid Implementation	Wolfram Schiffmann (Univ. of Hagen), Reagan Moore (SDSC)	
WP 3	Context Capturing, Representation, and Management	Matthias Hemmje (Univ. of Hagen), Marcus Delp (GLOBIT)	PCA2
WP 4	Multivalent Preservation Interface and Media Engines	Paul Watry (Univ. of Liverpool), Thomas Phelps (Industrious Media)	
WP 6	Harmonisation, Basic Analysis and Ingest	Jean-Pierre Chanod (Xerox RCE), Robert Sanderson (Univ. Liverpool)	
WP 7	Advanced Information Extraction and Knowledge Engineering	Jean-Pierre Chanod (Xerox RCE), Robert Sanderson (Univ. Liverpool)	PCA3
WP 8	Managing Shared Collections	Norbert Lossau (SUB, Göttingen) Gunter Schlageter (Univ. of Hagen)	
WP 9	Interoperability with Future Environments	Paul Watry (Univ. of Liverpool), Jose Borbinha (INESC/ID Lisbon)	PCA4
WP 10	Maintaining Essential Properties	Jana Dittmann (Univ. of Magdeburg) Kellie Snow (HATII, Univ. of Glasgow)	

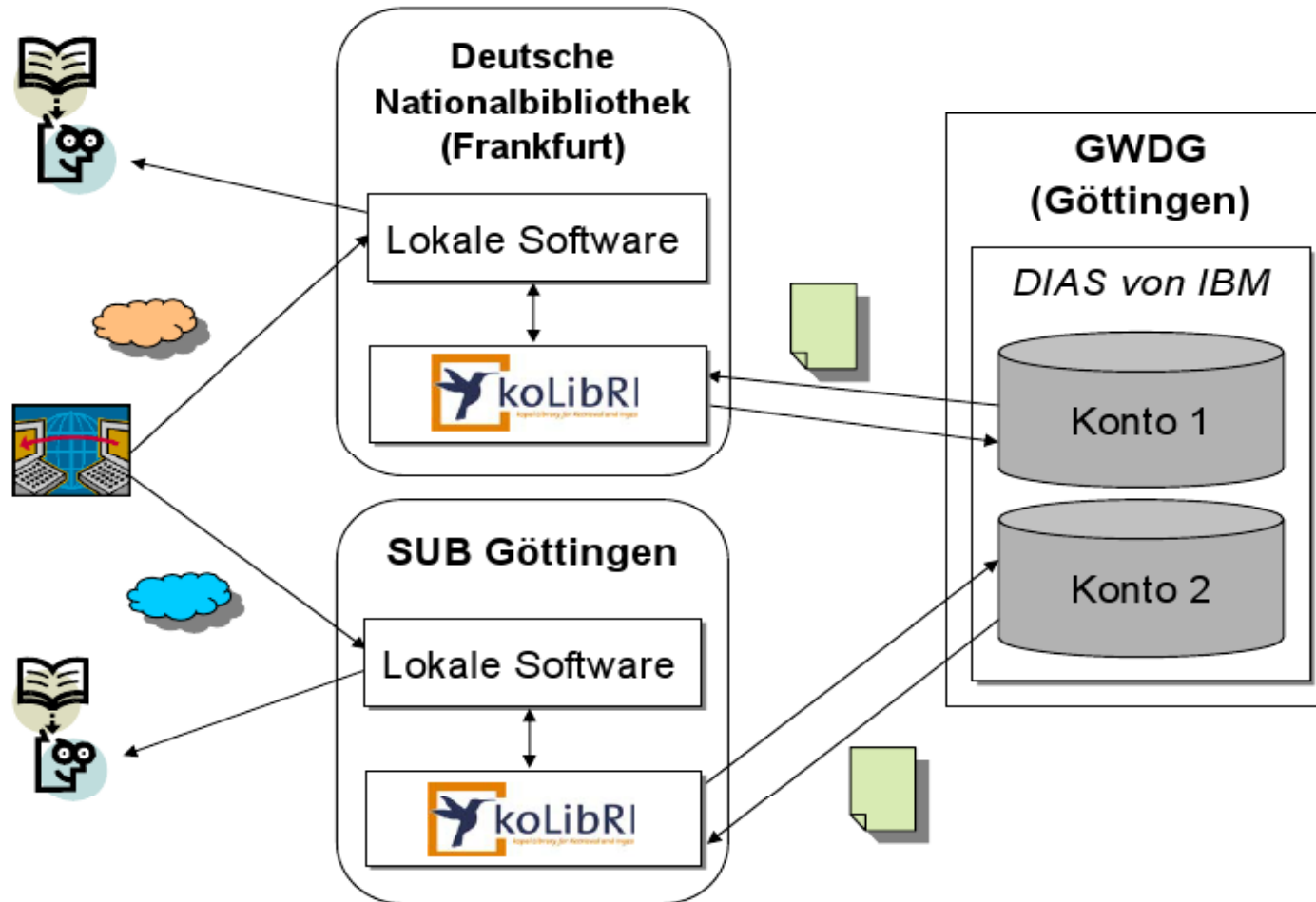


ISPs – Supporting Integration and Cohesion

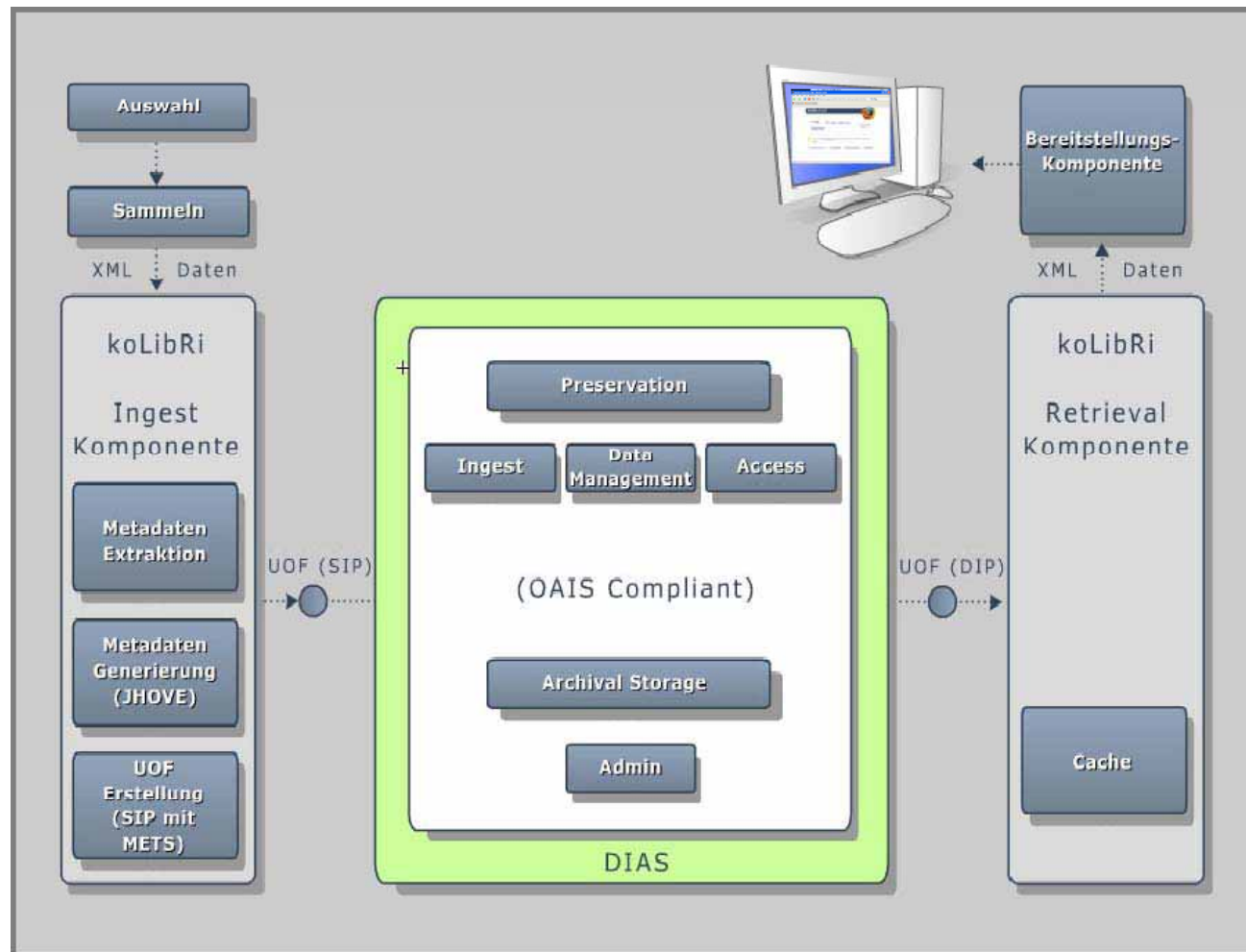
WP 11	Document Production, Archival, Access and Reuse in the Context of Memory Institutions for Scientific and Governmental Collections	Reinhard Altenhöner (DNB), Peter Hunin (Flemish Parliament)	ISP1
WP 12	Simple and Connected Object Production, Archival and Reuse in the Industrial Design and Engineering Domain	Kees Tuinenbreijer (PHILIPS) Wolfgang Wilkes (FUH)	ISP2
WP 13	eScience Data-Acquisition and Harmonisation Testbed	Jose Borbinha (INESC/ID Lisbon) Boris Chidlovski (Xerox RCE)	ISP3



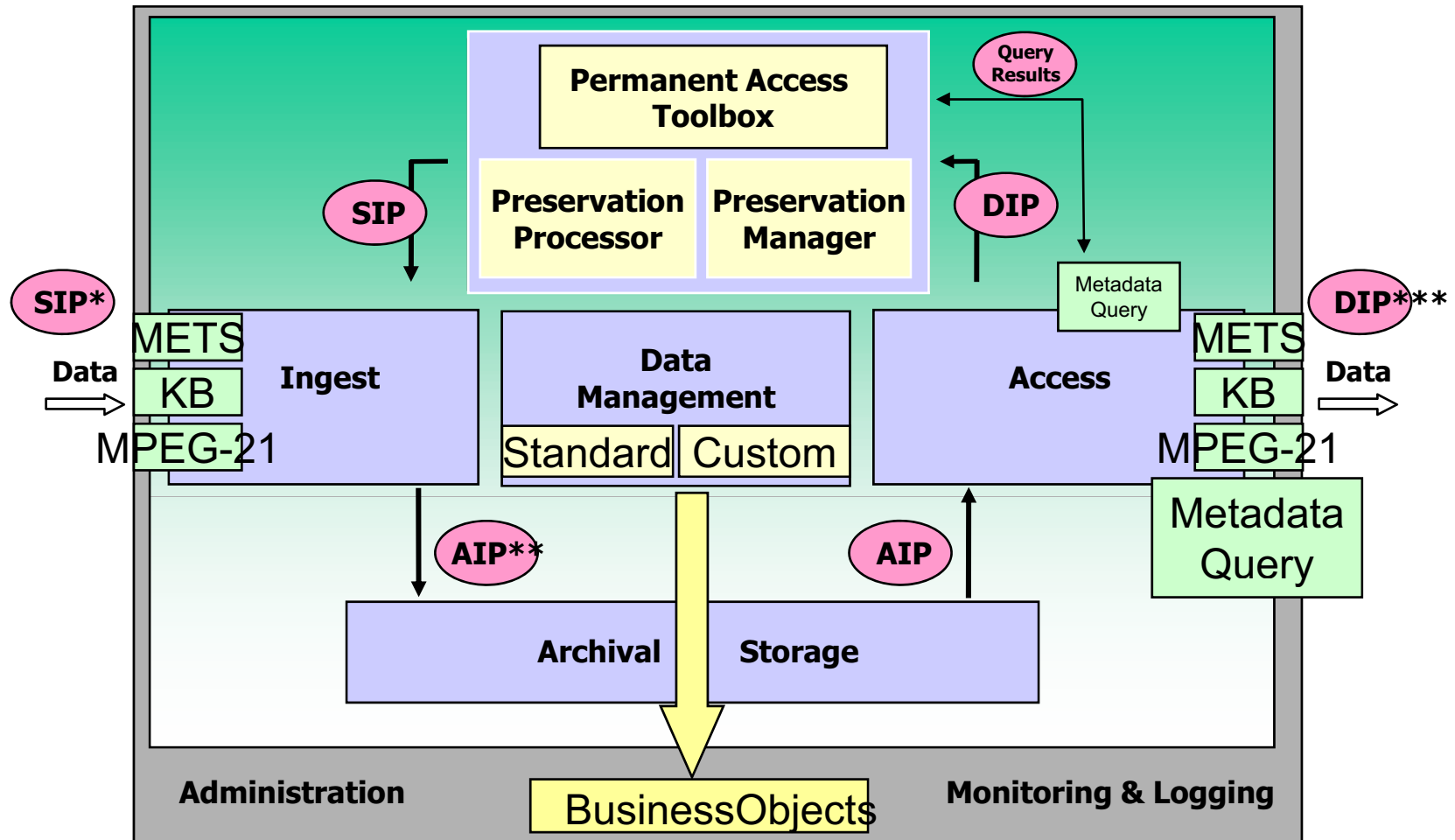
ISP-1 (I): KOPAL (DNB, SUB)



ISP-1 (II): KOLIBRI (DNB, SUB)

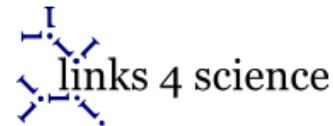


ISP-1 (III): DIAS (IBM)

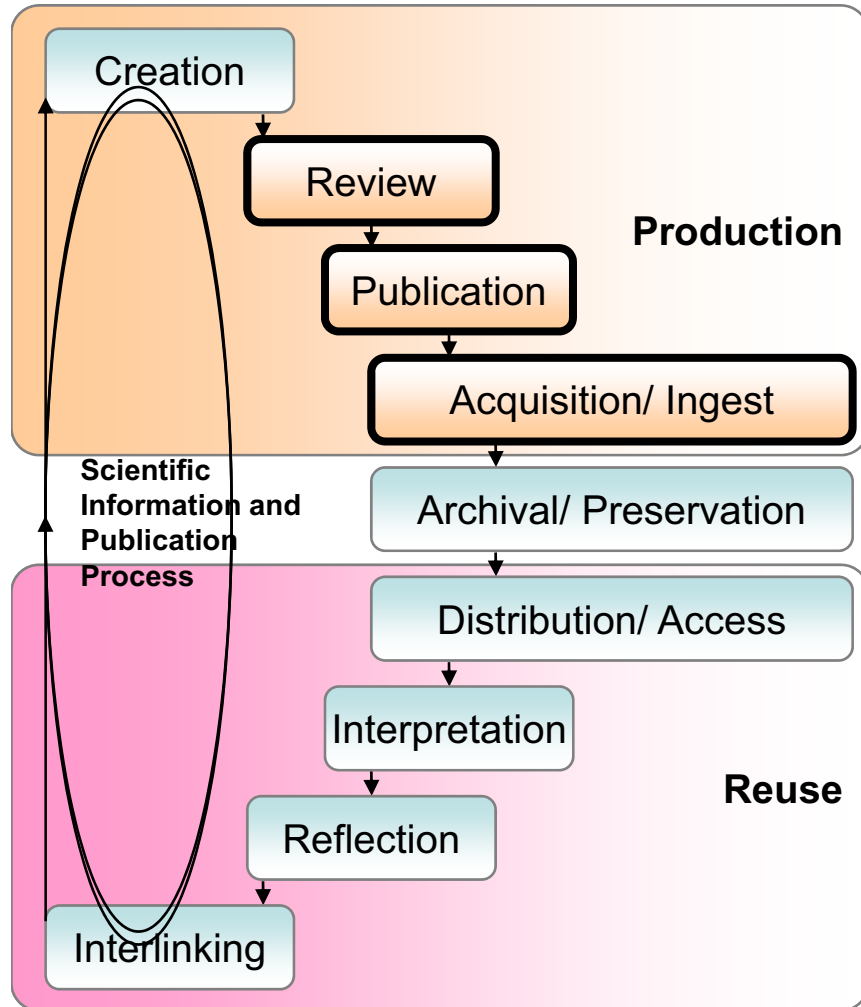


* = SIP: Submission Information Package
 ** = AIP: Archival Information Package

*** = DIP: Dissemination Information Package

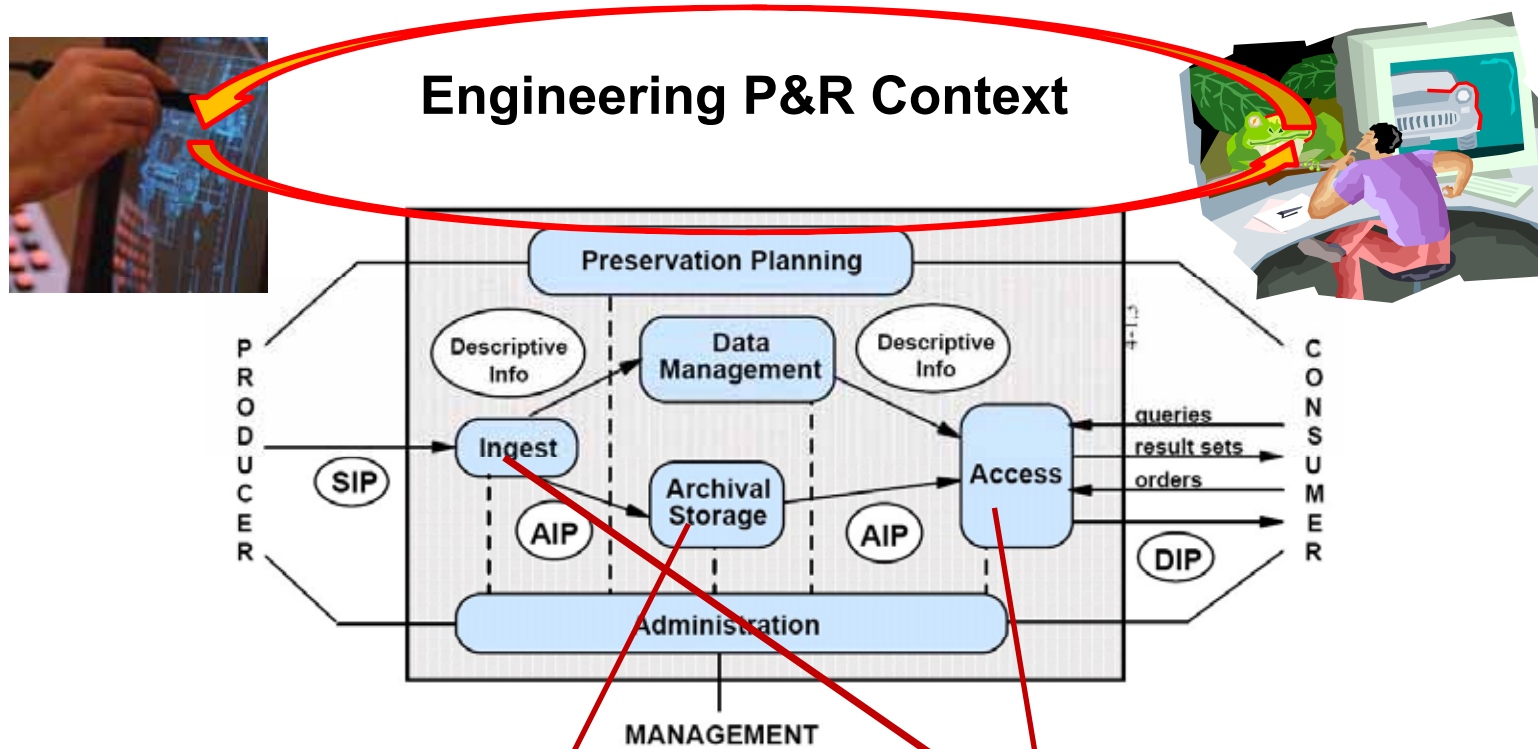


ISP-1 (IV): Scientific Publication P&R Context



- 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

ISP-2 (I): Engineering Scenario and the OAIS Reference Model

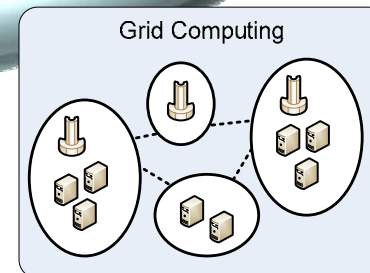
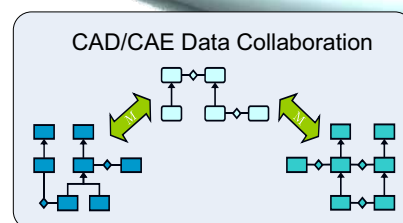
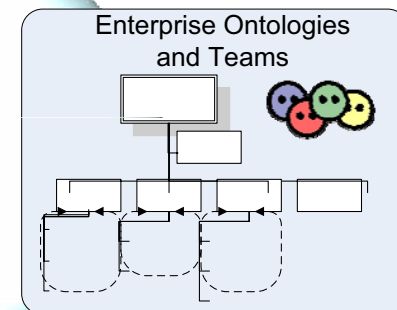
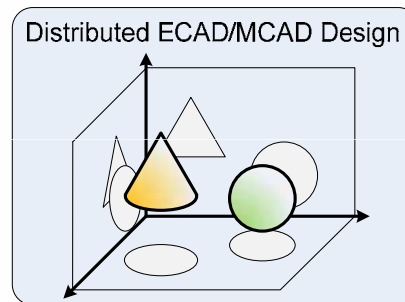
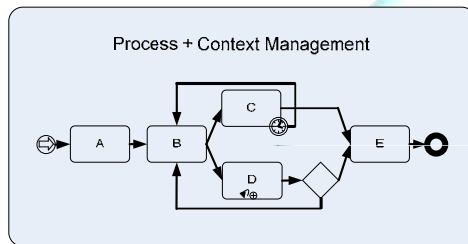
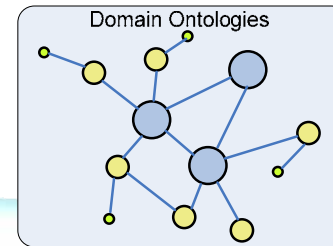
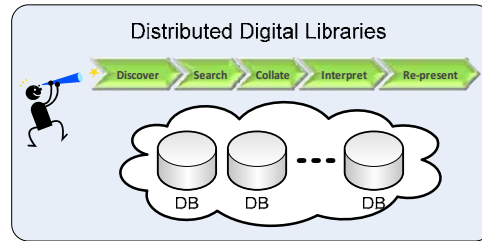


Multivalent approaches preserving CAD models

- migration
- emulation

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

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





IMPACT (I): User Communities & Discussion

SEGMENTED DISSEMINATION & EXPLOITATION ACTIVITIES

Instruments:

Invited Talks, Publication of Papers & Presentations, Themed Conference Tracks

WP16: Scientific/Academic Oriented Dissemination & Exploitation

Scientific Community Target Segments :

- Digital Preservation (iPres, IDCC)
- Digital Libraries (ECDL, JCDL),
- Information Retrieval (SIGIR, ECIR)
- GRID (GGF, Int. Symposium on Grid Computing)

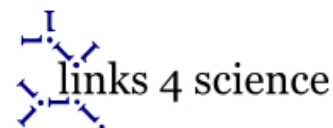
WP 17 Industrial Digital Preservation Technologies Outreach & Take-up

Public User-Community Target Segments :

- **Memory Institution** User Communities (DPE, NESTOR II, DP Coalitions, DP Task Forces, Alliance for the Permanent Access to the Records of Science)
- **E-Science** User Communities (EGEE, D-GRID)

Industrial User-Community Target Segments :

- **Design & Engineering** User Communities (DAT, DAC Conference)





IMPACT (II): User Communities & Engagement

EXPLOITATION DRIVEN DEMONSTRATION & DISSEMINATION

Instruments: Innovation & Exploitation oriented Workshops, Demonstrations, Round Tables, SIGs, Working Groups

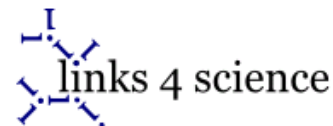
WP 17 Industrial Digital Preservation Technologies Outreach & Take-up

Public User-Community Target Segments :

- **ISP-1 Demonstration Workshops and Follow-Up Round Tables**
 - in Collaboration with NESTOR,
 - for IT, Working Groups, e.g., for ST&M Publishers Association
 - For IT User-Groups, e.g., for the KOPAL User Group in Collaboration with IBM
- **ISP-3 Demonstrations**
 - at EGEE conference in collaboration with D-GRID
 - at Int. Symposium on Grid Computing

Industrial User-Community Target Segments :

- **ISP-2 Demonstration Workshops and Follow-Up Round Tables**
 - at CeBIT, e.g. in Collaboration with EMF
 - at STEP, e.g., General Assembly in Collaboration with ProSTEP
 - at DAT, e.g. within its Exhibition





IMPACT (III): Added Values & Exploitation

EXPLOITATION DRIVEN DISSEMINATION

WP16: Scientific/Academic Oriented Dissemination & Exploitation

Scientific Community Target Segments :

- Added value w.r.t. ongoing research (SHAMAN results tackle and advance parts of DPE Roadmap Challenges and NARA Requirements)
- Take-up by ongoing EU and national R&D projects like PLANETS, CASPAR, NESTOR, DILIGENT, D-GRID possible
- Re-use tangible and intangible outputs as a basis for future DP R&D (projects/proposals)

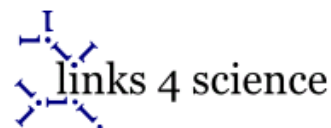
WP 17 Industrial Digital Preservation Technologies Outreach & Take-up

Public User-Community Target Segments :

- Public Memory Institutions are a future business field for Xerox Global Services to sell Digital Archiving and Preservation Products & Services
- Industrious Media is interested in exploiting the Multivalent Technology towards public customers
- Members of the SHAMAN Consortium are interested in exploiting DP functions in the PCO and Scientific Publishing Sectors

Industrial User-Community Target Segments :

- Industrial Manufacturers are target customers for Xerox w.r.t. Product Life Cycle Management Services
- Long Term Asset Management for the Finance, E-Government, and Business Information Management Sectors will also be of interest for XEROX
- Members of the SHAMAN Consortium are interested in exploiting DP functionality in there Data and Product Life Cycle Management Support for Manufacturers
- Industrious Media is interested in exploiting the Multivalent Technology towards industrial customers





Fine

Thank you very much for your attention.

