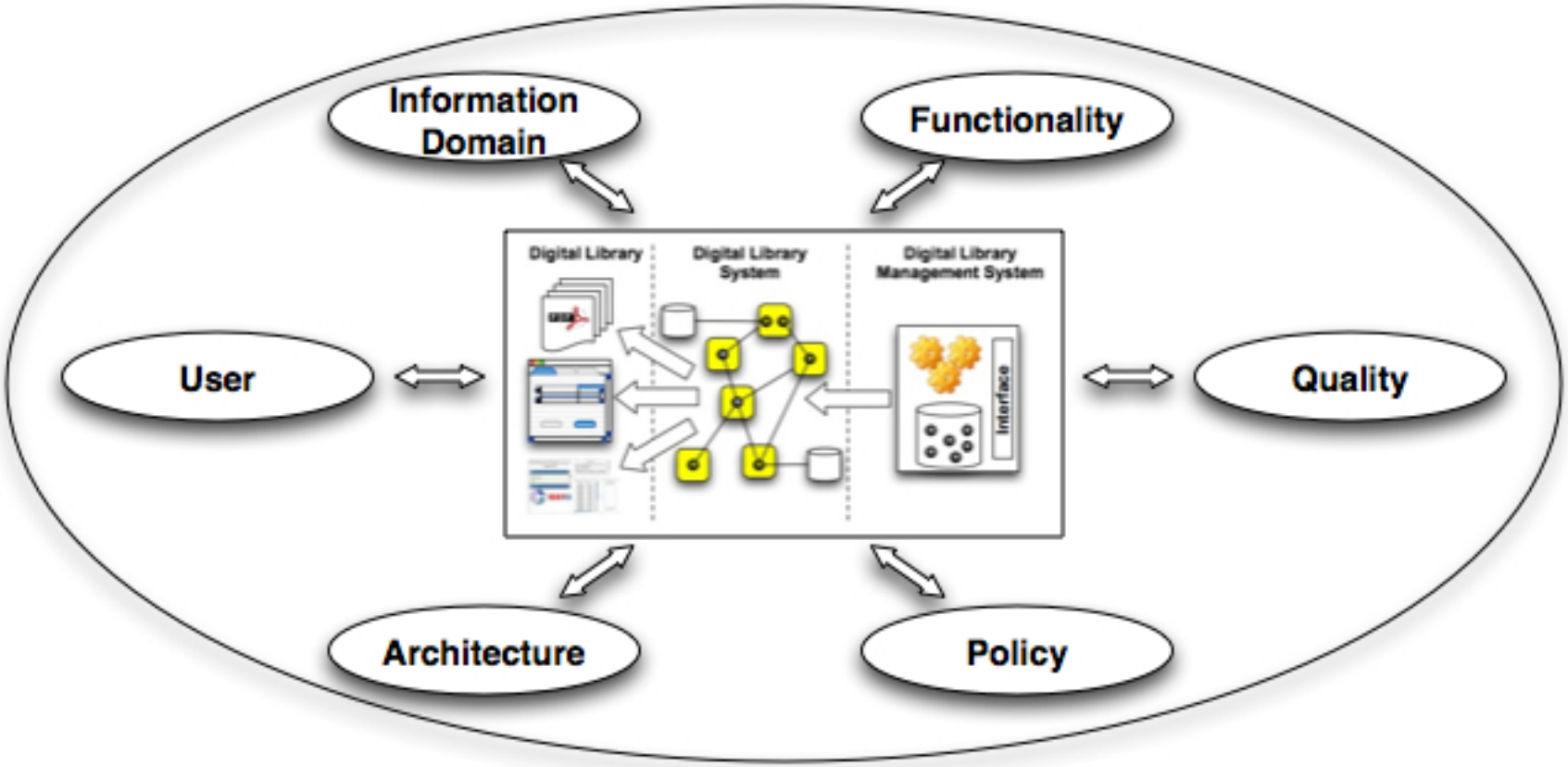
Towards the DelosDLMS, an integrated prototype for future Digital Library Management Systems

Hans-Jörg Schek
Univ. of Konstanz

Heiko Schuldt
Univ. of Basel

The Reference Model and DelosDLMS



DelosDLMS: Partial Implementation of the DL Reference Model

Have a demonstrator for future Digital Library Management Systems that not only shows new **combined text and audio-visual** search functionality and **personalized browsing** by new adaptable information **visualization and relevance feedback** tools at the interface but also proves that generic systems can be built that not only support finding of **relevant** information but also enable to **annotate and process** found information, to integrate **sensor data** stream processing, and – from a systems engineering point of view - allows simple **configuration and adaptation** while being **reliable and scalable**

DelosDLMS Approach

Use a SoA/P2P/Grid infrastructure to coordinate search building blocks and application processes to retrieve and work with dynamic and mobile information.

Start with OSIRIS/ISIS from ETH, extend it, benefit from Diligent/Grid

Add components, replace weak ones by better ones

Use existing knowledge and prototypes and try to integrate it as much as possible

Information Domain

- ETHWorld: 625.026 Images, with features extracted
  - Color Moments (Global + 5 Regions)
  - Texture (Global + 5 Regions)
  - Face (Number of + Area)
  - Term Features (Categories + crawled HTML files) from ETH-Web sites

- ISIS: 53.837 Images, features as ETHWorld,
  - Subcollections:
    Web Cams, TV Streams, ImageNet, Sports Events
    ADAC car images (1.593)
    Nasa space shuttle images (9.921)
    Yorck Paintings (10.422)
    ZEFA Photo Stock (4987) und ImageBank (6.112)
    ARTE Emblemes DB (7.185)
    ImpECt,ESA (48)

Information Domain – contd.

- MED: 50.143 medical images from ImageCLEFMed, features as ETHWorld, no faces, subcollections
  - c@simage: Universitaetsspital Genf (HUG)
  - PathoPic: Universitaet Basel
  - MIRS: Mallinckrodt Institute of Radiology,Washington Univeristy
  - PEIR: University of Alabama at Birmingham, USA

- ISIS Audio:1185 MP3 files from private CD collection, full length, features using the MARSYAS Library: Beat, MPitch, Rafi's Mix, SFX, SVFFT
  Term Features from music info (Artist, Album, Titel)

- ISIS Video:5 movies from private collection, automatically divided into1200 Scenes. Features:
  - Image Features as ETHWorld, no faces
  - Text Features from meta date of the IMDb

Functionality: Demo

Red Car

Diploma Party

Flowers

Fox

ImpeECt

Information Modelling

Functionality

- Strengths
  - Image retrieval, „Google for Images“
  - Relevance feedback: different kinds available
  - Text retrieval (boolean, vector space)
  - Fastmap visualization
  - Video, audio feature extraction and retrieval
  - Predicates, categories, face detection
  - Image segmentation and segment matching
- Main weaknesses
  - Image centered
  - Monolithic indexing component

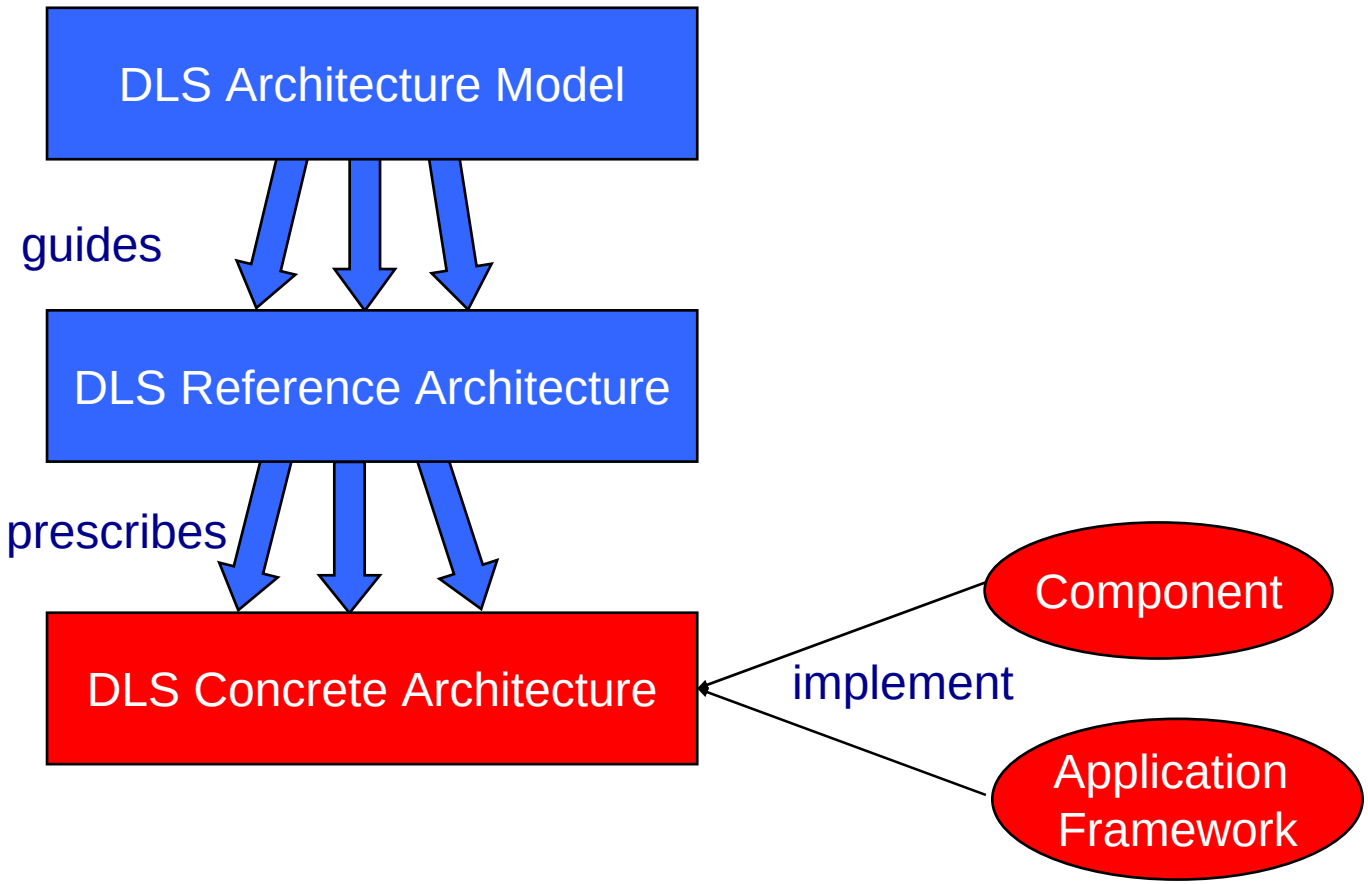
Functional Extensions for DelosDLMS

- Select from many DELOS members
  - Sophisticated term extraction, multi-linguality, query expansion,…
  - Higher-level image/video feature extraction, object recognition
  - Speech retrieval
  - Compound object matching, multi-object, multi-feature
  - More visualizations, navigational access, personalization
  - Multimodal interfaces, mobility
  - Annotations
  - Client Services for collection management, information analysis, mining

Architectural Extension for DelosDLMS

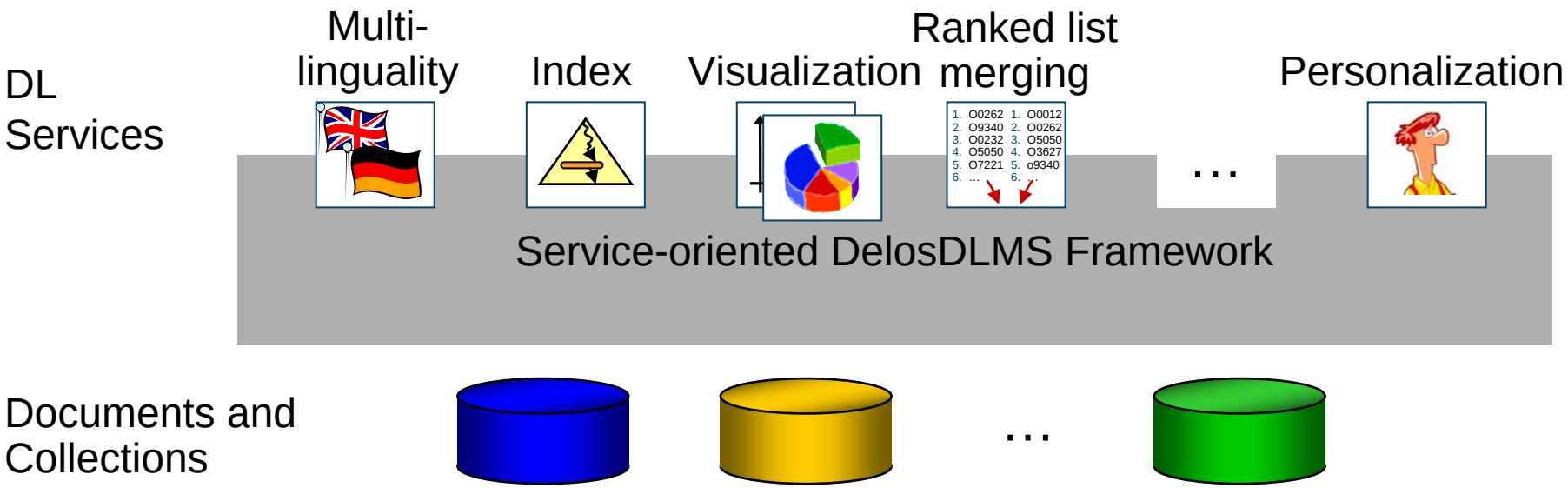
- Select components from many DELOS members
  - Specialized indexing components
  - Ranked list merging
  - Generator and optimizer of search processes
  - New client processes
  - Personalization services
  - Storage services
  - Standardized data transformations
  - Preservation services

DelosDLMS


DLS Architecture Model
guides
DLS Reference Architecture
prescribes
DLS Concrete Architecture
implement
Component
Application Framework

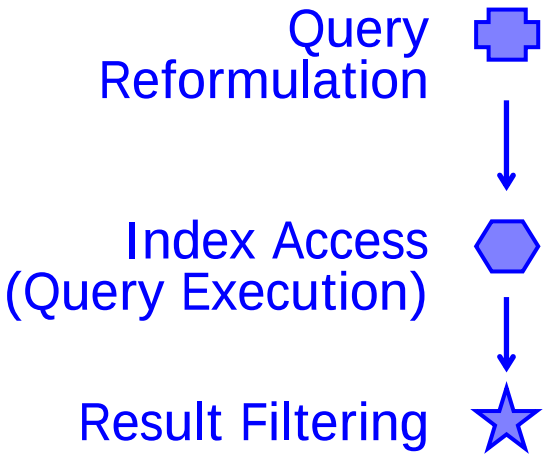
Architecture

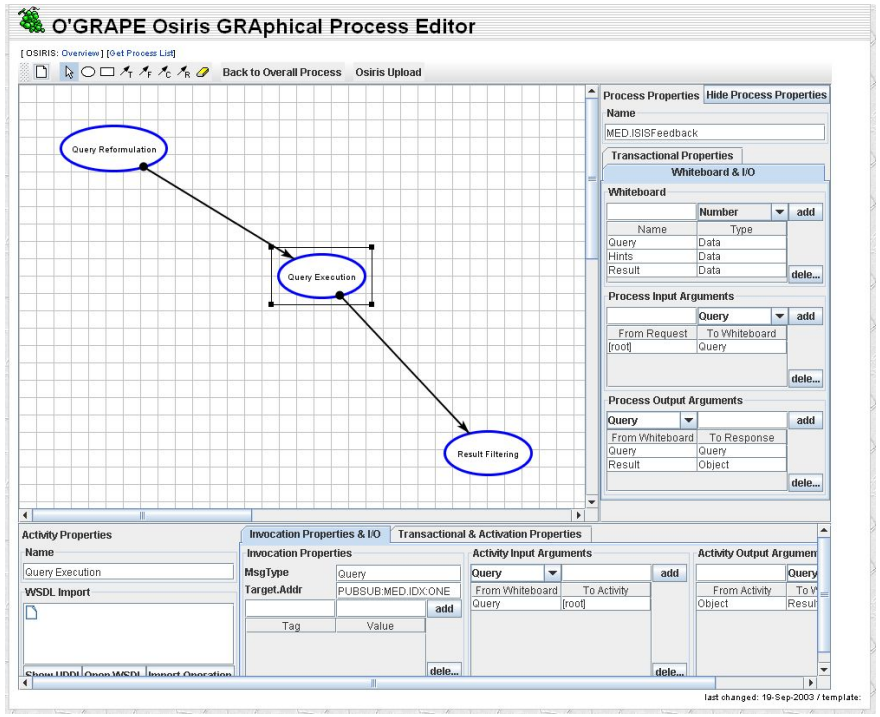
- DelosDLMS follows a service-oriented Architecture
  - DL Functionality is available by dedicated (web) services
  - DL Applications are built by combining existing DL services (possibly from different providers), independent of content
  - Basis: the OSIRIS System of ETH/UMIT/UNIBAS



Defining DelosDLMS Applications

- Development of DL Applications (= Processes) in DelosDLMS
  - combination of existing services („Programming in the Large“)

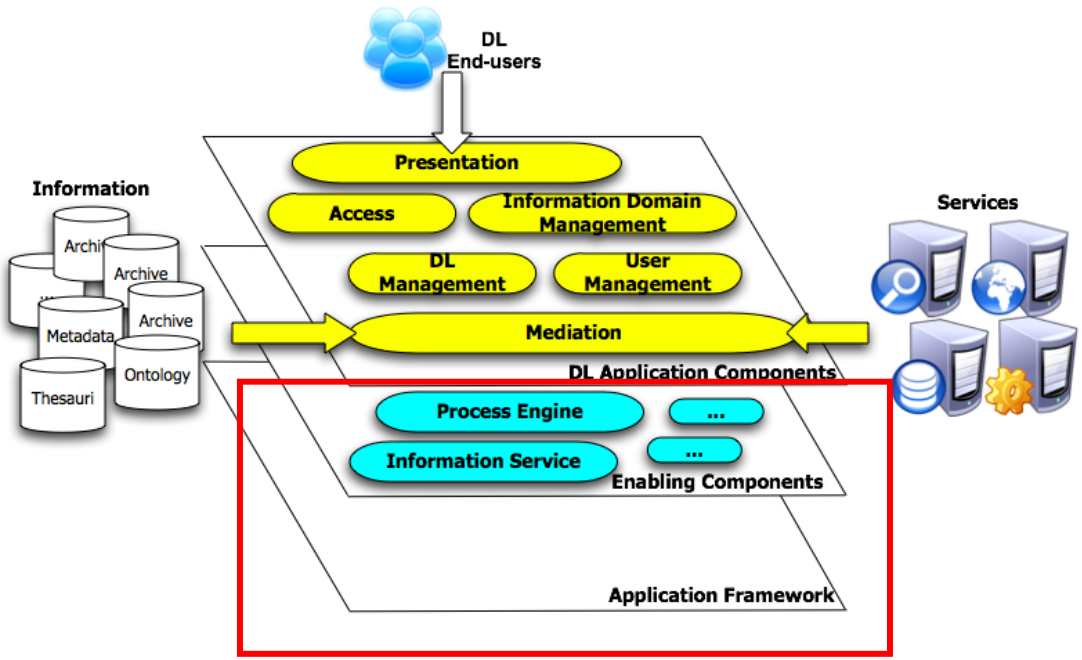




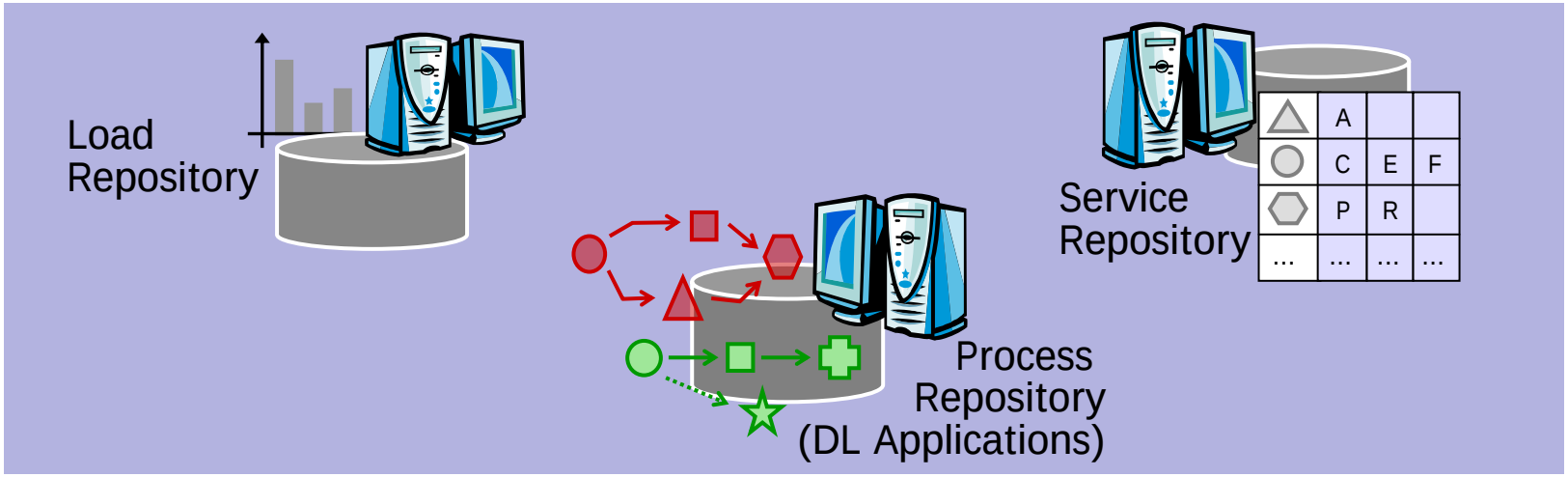
What is DelosDLMS?
DLx Functionality of DelosDLMS

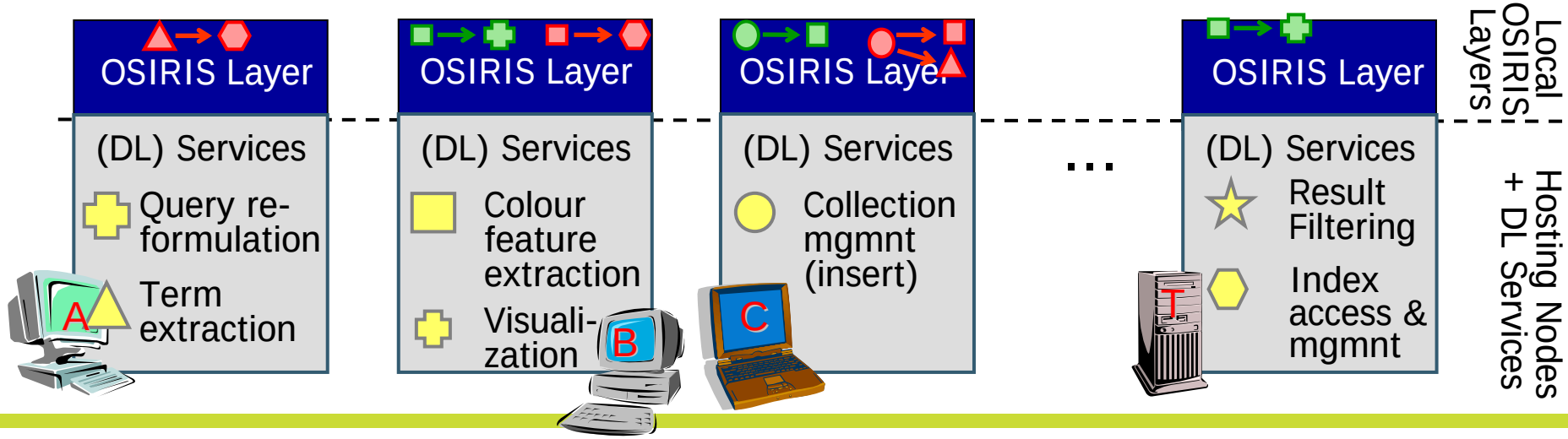
x in {Designer, System Administrator, Application Developer}

- Information Service
- Process Management
- Deployment of Services
- Monitoring Services (load balancing)
- Subscription/ Notification
- …

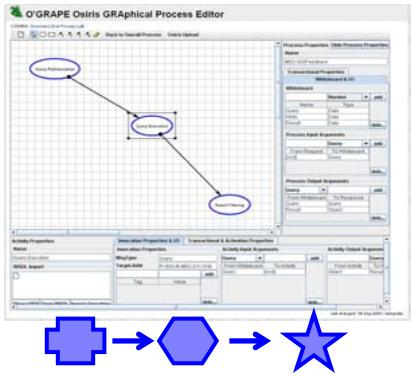


The DelosDLMS/OSIRIS Architecture





Executing DelosDLMS Applications



Process Programs

Execute

OSIRIS Services

OSIRIS Services

OSIRIS Services

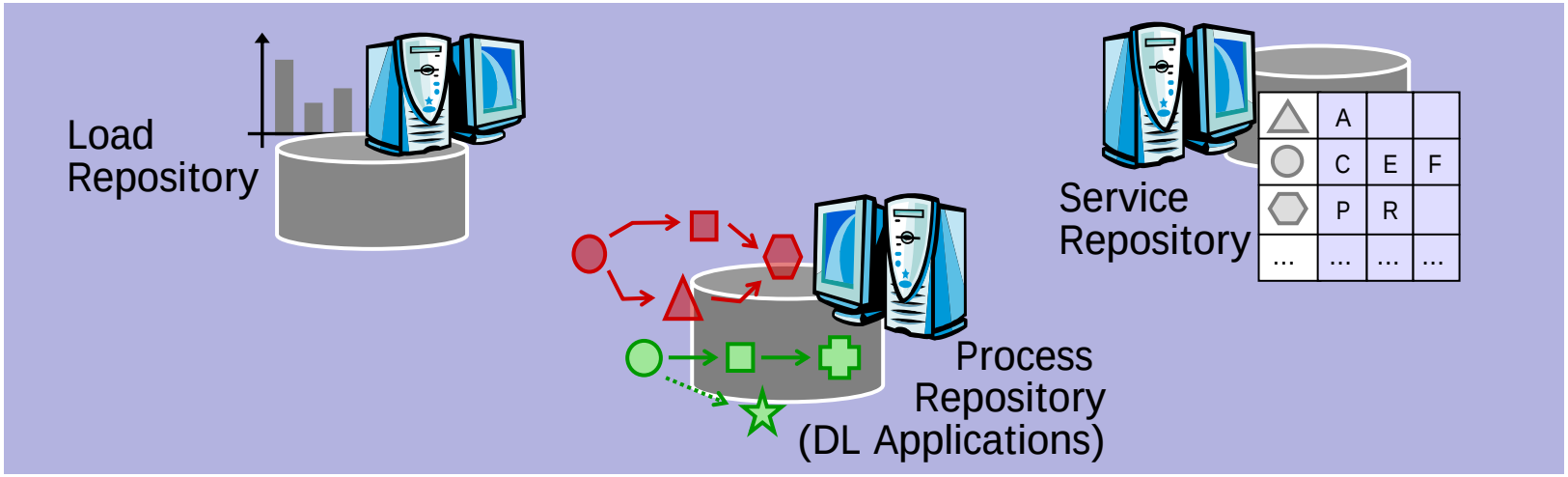
OSIRIS Services

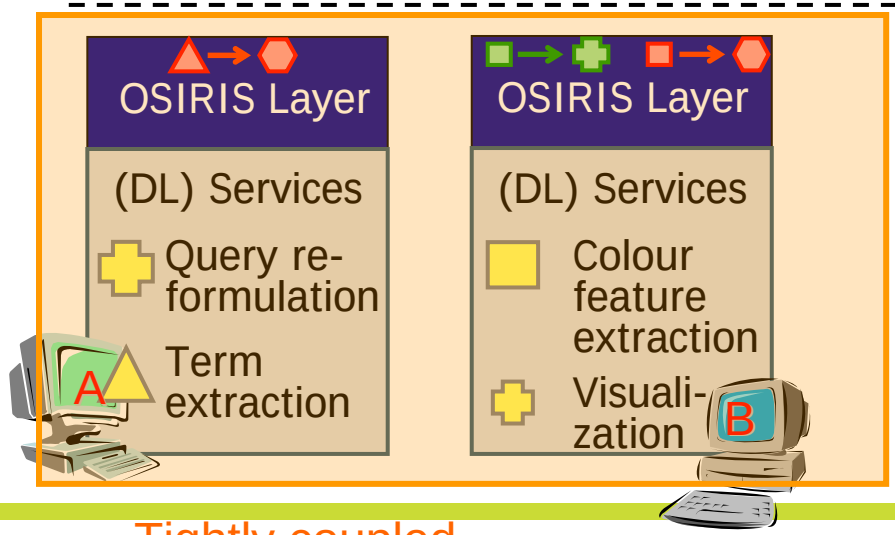
Result Filtering

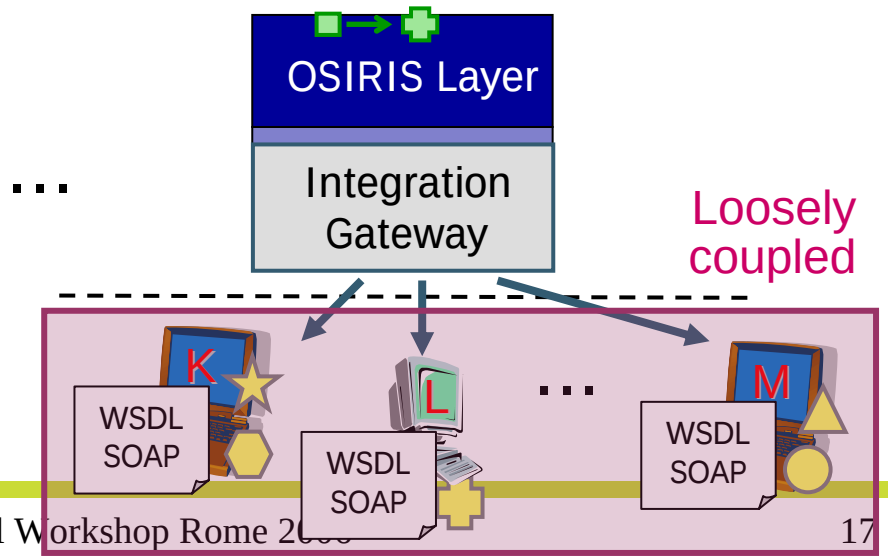
Query Reformulation

Index Access

Integration of Services







Quality of Service

- Tightly coupled services:
  - First-class citizens in the OSIRIS network
  - Integrated into the local OSIRIS layer
  - Quality-of-Service:
    - Provide added value for services (e.g., **compensation**, **restart**, etc)
    - **Load balancing**
    - **Execution guarantees** for applications consisting of such services
- Loosely coupled services
  - Easy to integrate (only web service interface is needed, i.e. description via WSDL, invocation via SOAP)
  - Only “best effort”, no
    - sophisticated failure handling
    - Load balancing, etc.

Conclusion

- Work on the Reference Model is accompanied by systems work
- Practical evaluation of the Reference Model
- Integration of work done in the DELOS network