

Digital Repositories and the Grid

David E. Martin

Program Director, Internet Standards and Technology, IBM

Data Area Co-Director, Open Grid Forum

martinde@us.ibm.com

Open Grid Forum



- In 2001, Global Grid Forum Formed
 - Merger of US, European and Asian Grid Groups
 - Modeled on IETF - No Membership, No Voting, Rough Consensus
- Focus on Grid Middleware
- In 2006, Global Grid Forum Merged with Enterprise Grid Alliance to Form OGF



- Increased Focus on Enterprise and Datacenter Issues
- In 2008, OGF Europe Formed
 - Digital Repositories is a Strong Focus

Areas of OGF Work



Standards

Develop and Document Standards with Community Input and Consensus

-

Interact with other SDOs

Community

Build Communities around Particular Use Cases

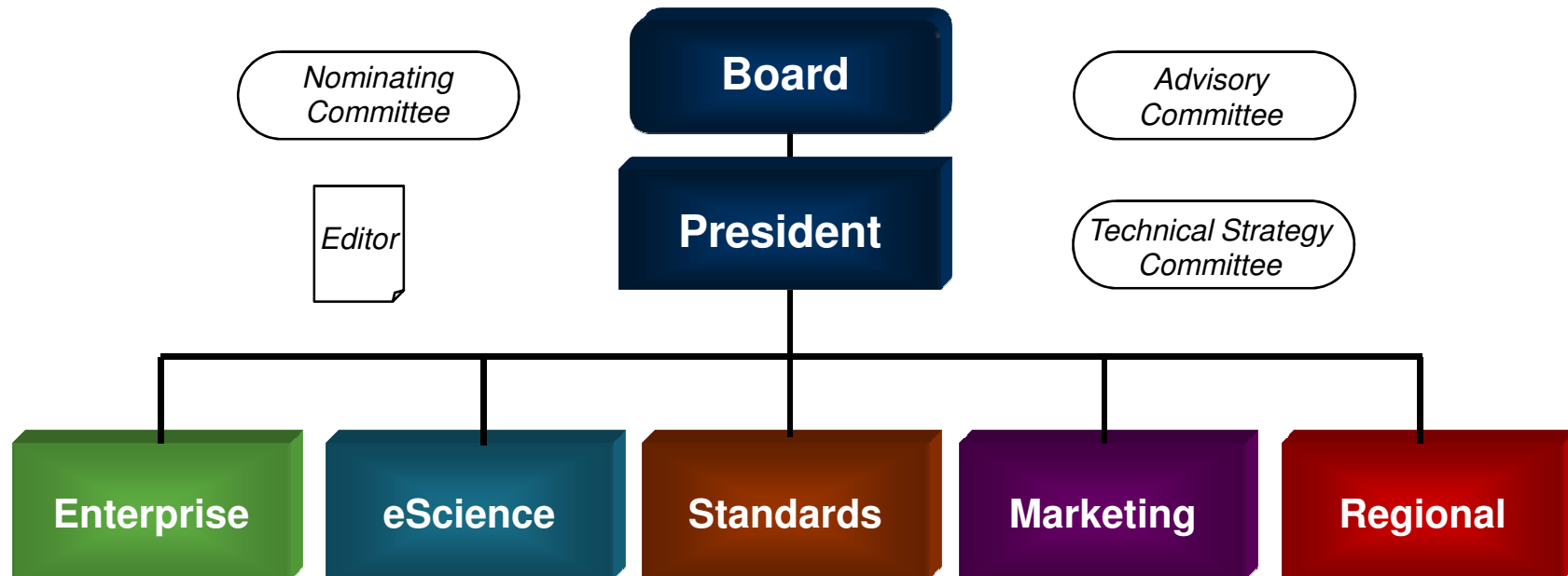
-

Leverage Existing User Communities

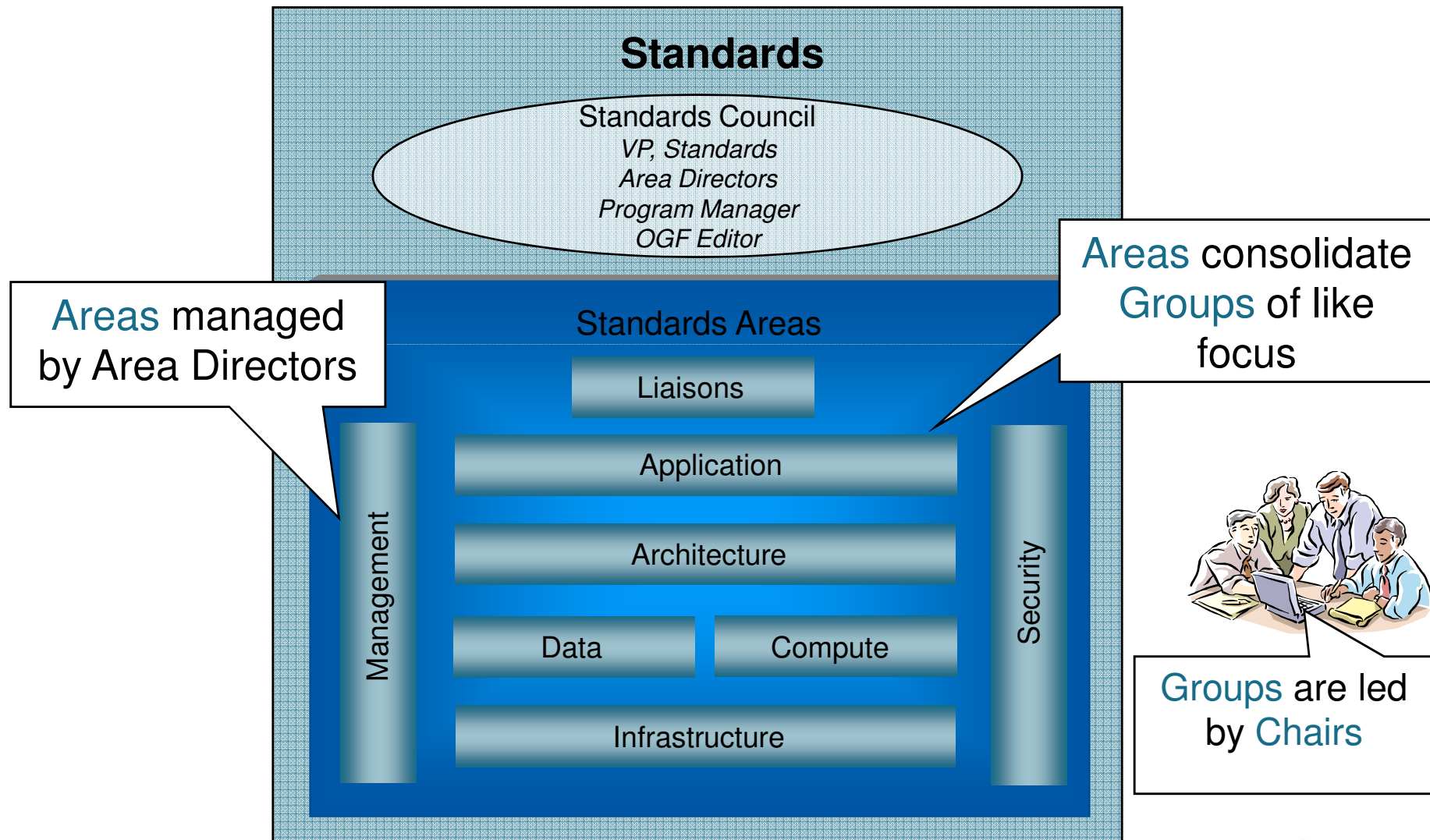
Workshops & Forums

Bring Communities Together to Share, Innovate, Plan and Outreach

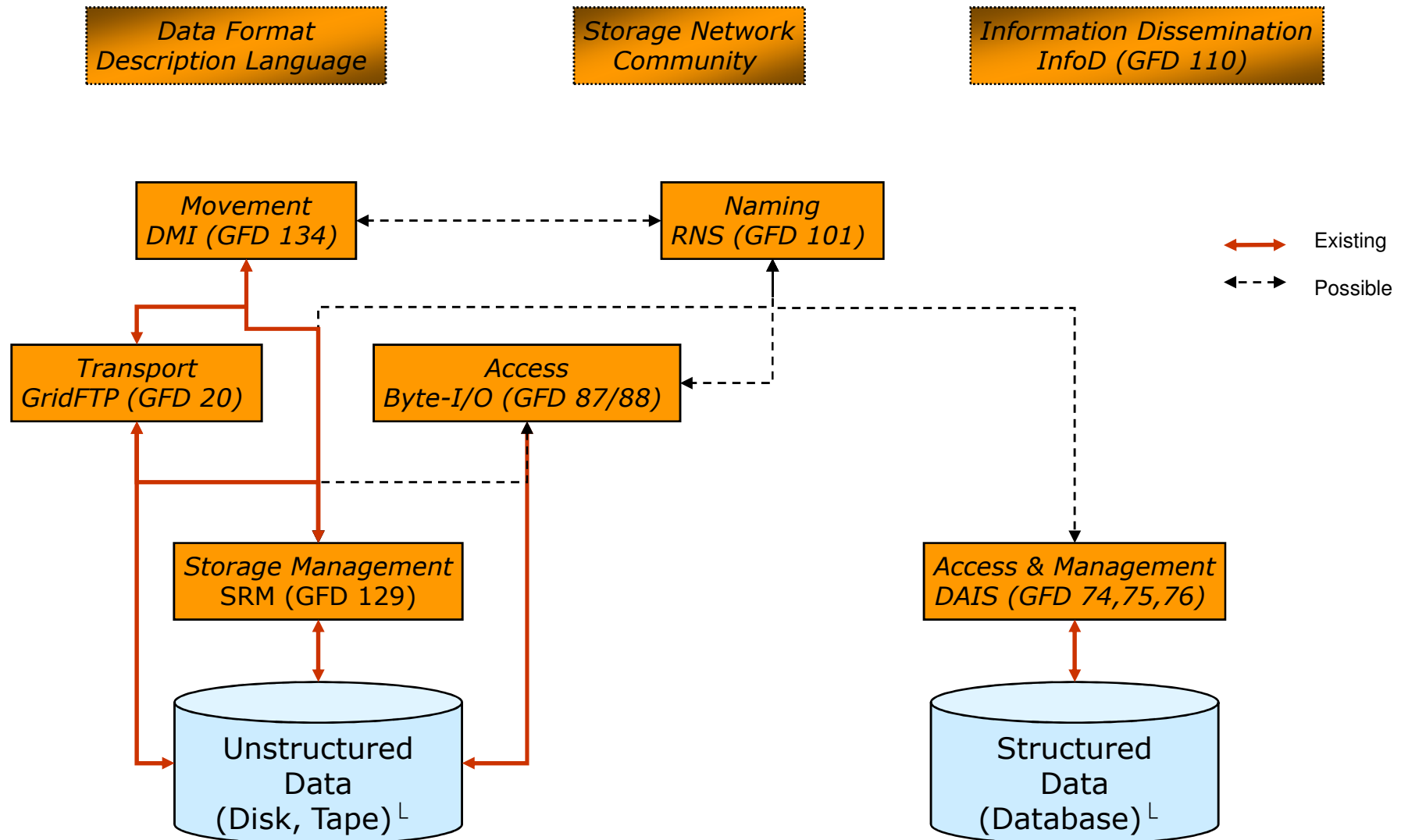
OGF Organization



Standards Function Organization



Data Area Relationships



Some OGF Work of Interest to DR Community



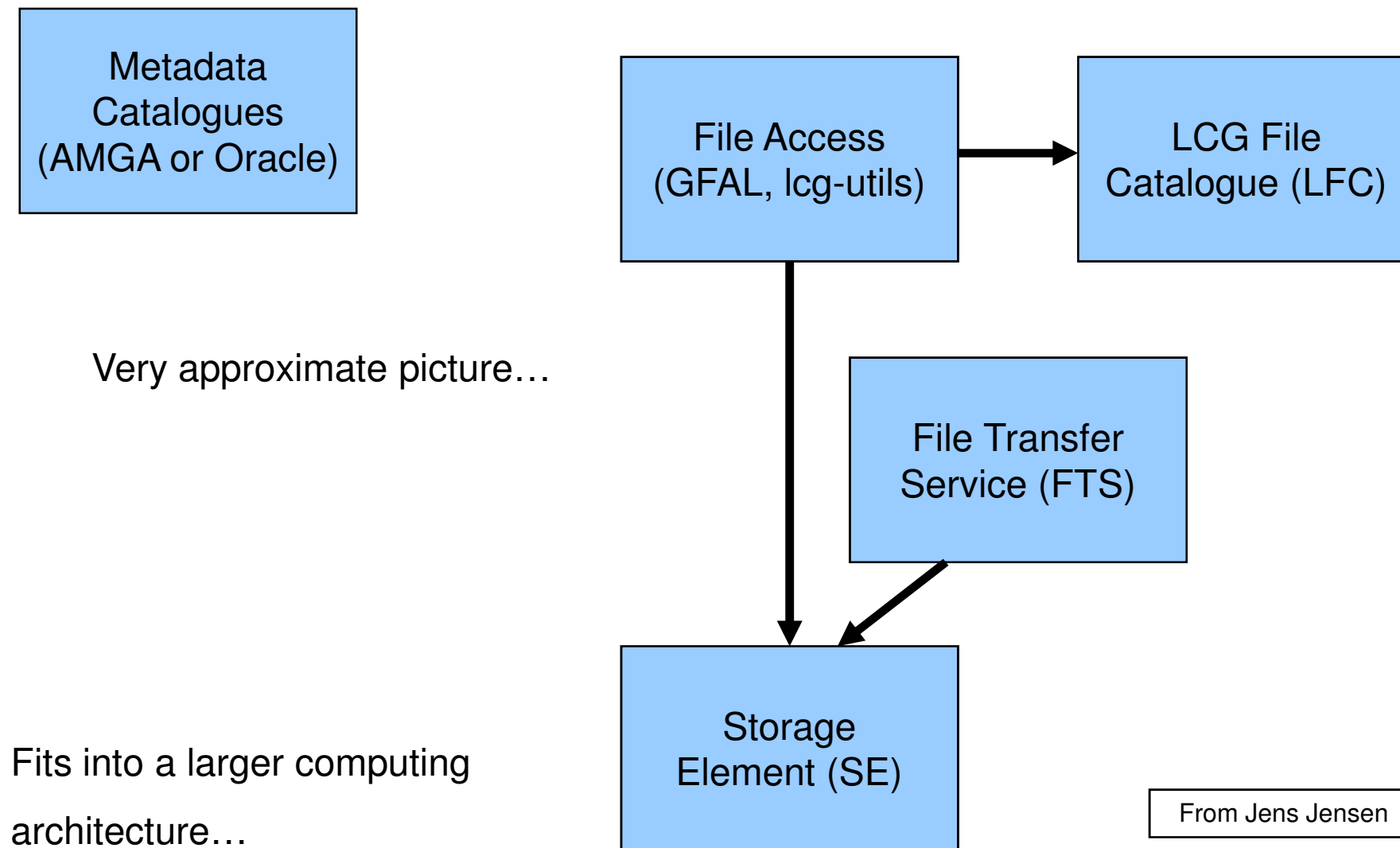
- SRM – Storage Resource Manager
 - Provides Location-Independent Storage Access
- OGSA Data Architecture
 - Provides Grid Data Architecture and Integrated into Overall Service-Oriented OGSA Architecture
- GIN – Grid Interoperability Now
 - Working to Document Interoperability
 - Information Services and Data Subteams are of Most Interest to DR Community

Storage Resource Manager



- OGF Proposed Standard from the Grid Storage Management GSM-WG (GFD-129)
- Storage Resource Manager
 - Storage Element Control Protocol
 - (By abuse of notation, an SE with an SRM interface is often called “an SRM”)
- Designed to manage very diverse storage systems and very large data volumes
- Used in gLite and Elsewhere
 - Six Different (Soon 7!) Implementations

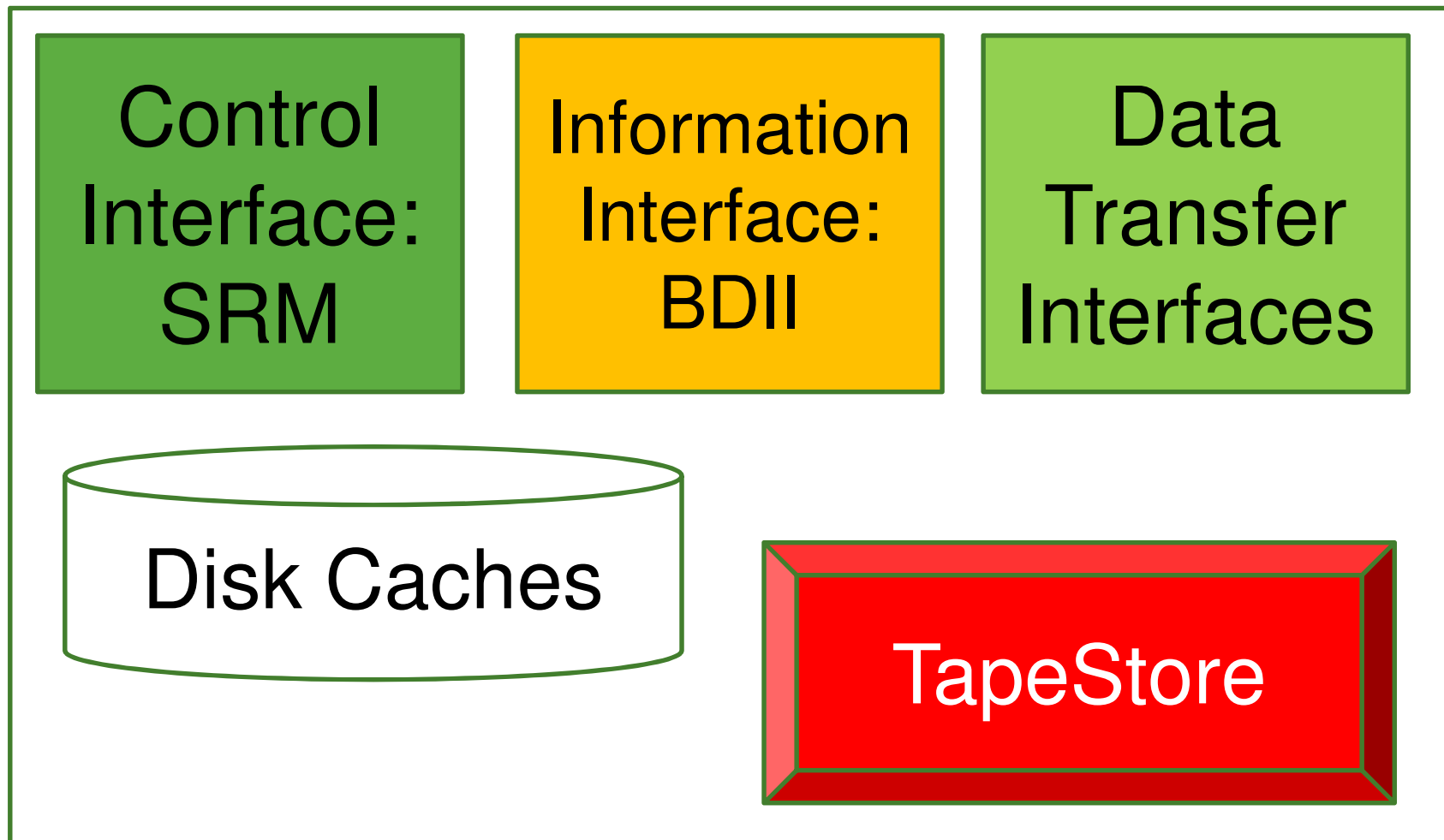
gLite Architecture



Storage Element (SE) Anatomy

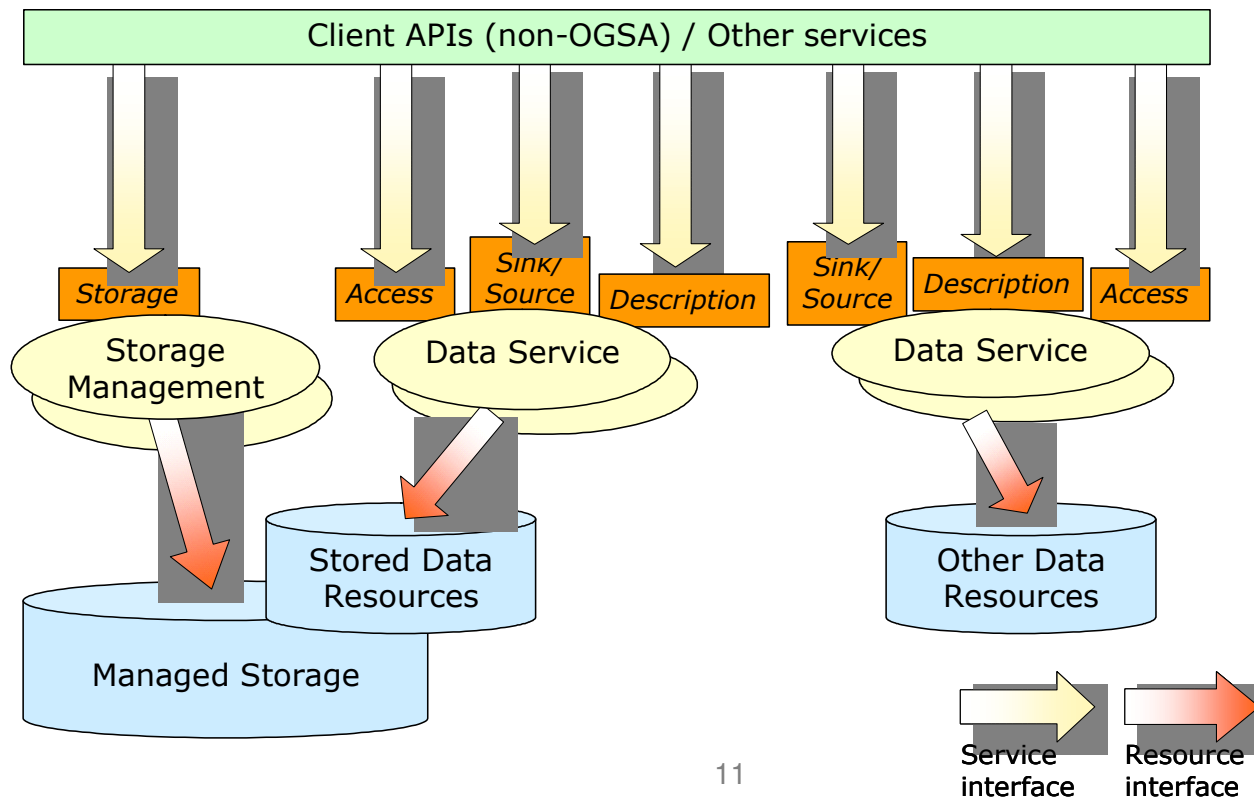


Grid Interface to Disk or Mass Storage Systems



OGSA Data WG

- OGSA Data Architecture Publish as GFD121
- Service Oriented Framework for Data



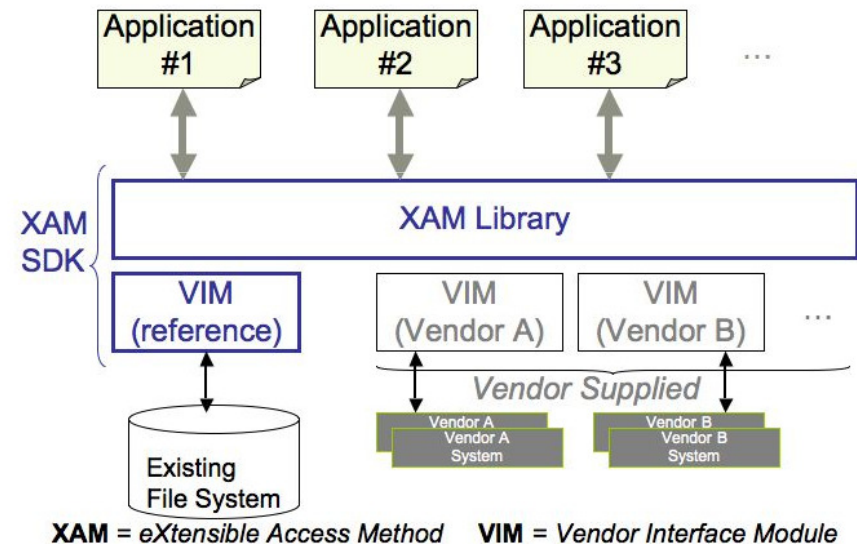
- Grid Interoperability Now (GIN)
- Effort to Span all of OGF and Document How Production Grids Can Interoperate
- 5 Subteams
 - GIN-auth - Authorization and Identity Management
 - GIN-data - Data Management and Movement
 - GIN-jobs - Job Description and Submission
 - GIN-info - Information Services and Schema
 - GIN-ops - Operations experience of pilot test applications

Other Standards – SNIA's XAM



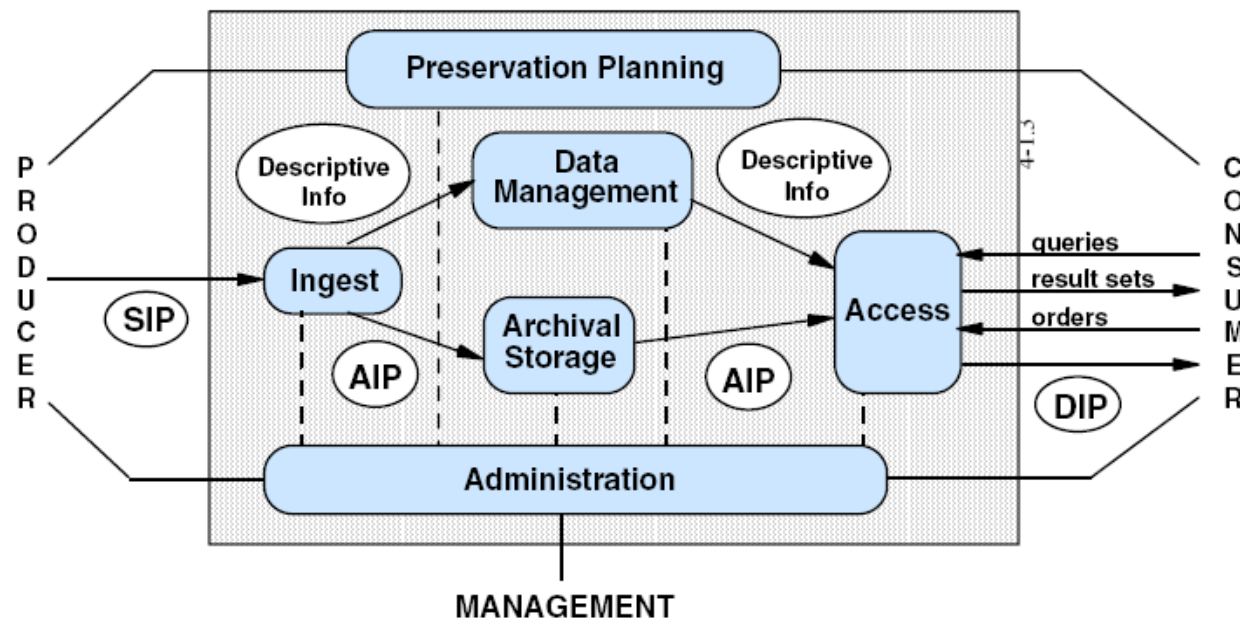
XAM – eXtensible Access Method

- Standard Way of Storing Metadata with Data on a Storage System
- XAM Specifies Handling of Metadata But Doesn't Speak to the Names and Values
- OGF Has a Strong Liaison Relationship with SNIA



Other Standards – ISO's OAIS

- Open Archival Information System (OAIS)
- Reference Model for Specifying Archive System
 - Roles: Producers, Consumers, and Management
 - Functions: Ingest, Archival Storage, Data Management, Administration, Access, Preservation Planning and Common Services



Enterprise Digital Repositories

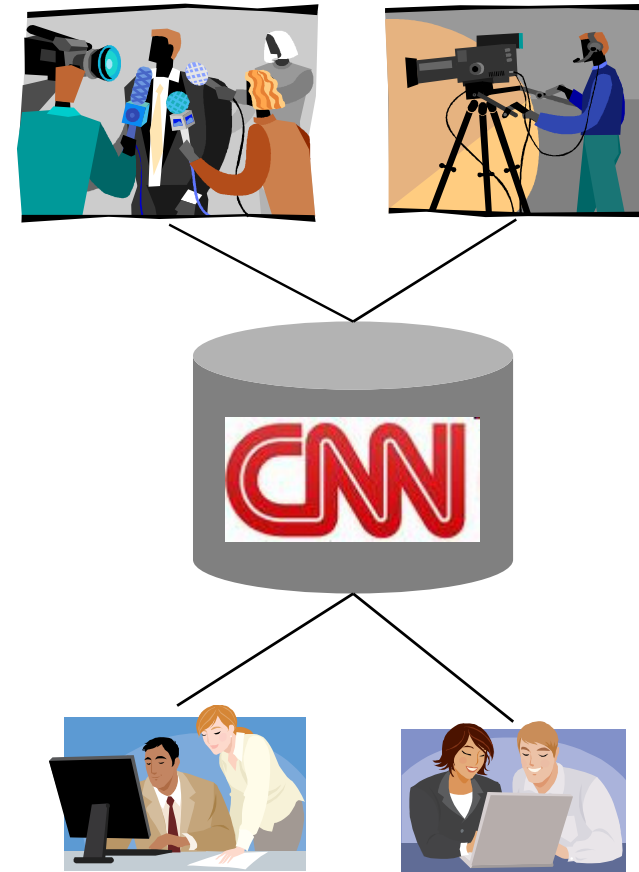


- Often Called Something Else
 - Data Warehouse
 - Enterprise Content Management
 - Document Management System
- Based on Commercial Databases with Extensive Customization
- Highly Centralized Storage with Distributed Retrieval and Processing

Example Enterprise DR

CNN's Media Information Retrieval Application (MIRA)

- Built on top of IBM DB2 Universal Database and DB2 Content Manager VideoCharger
- Manages Over 200,000 Hours of Archived Video
- Handle the Ingesting, Cataloging, Searching, Selection and Fulfillment of the Production Team's Archival Video Needs



Evolution of Digital Repositories

Rise of Cloud Computing is Driving DR to a Grid-Like Model

- Heterogeneous Infrastructure
- Multiple Domains of Control
- Integration with Existing Security, Authentication, Monitoring and Management Systems
- Standards for Interoperability

Conclusion



- Grid and Cloud Computing is Driving Highly Distributed Architectures
- Need Standards to Leverage Existing Work and Accelerate Adoption
 - Issues Around Metadata Seem Key
- OGF Has an Infrastructure and Community That Can Contribute
- The Digital Repository Workshop at OGF24 is the Next Step
 - 2 March 2009, Catania, Italy