

GGF9 OGSA-WG Data Session  
Oct. 06 2003, 6-7:30pm Sheraton 4  
Attendees: 71  
Minutes: Lisa Childers, Andreas Savva

Hiro

- GGF IP Policy
- Cross-WG Session purpose
  - o Goal is to listen

Jay Unger, OGSA Data Services

- taxonomy
  - o program execution (scheduling, micromgmt, etc.)
  - o data services
  - o Grid core services (meta operating services)
- What is a data service?
  - o Service-oriented treatment of data
    - Exploit existing architecture
      - OGSI (dynamic)
      - OGSA
    - Virtualize data sources
      - Location transparent
      - Failure transparent
      - Guaranteed QoX (quality of X where X can stand for a number of things)
        - o Service
        - o Data (consistency, completeness)access
        - o Protection
        - o Retention
  - o A service that “encapsulates” and “abstracts” data
    - Provides location transparency
    - Provides independence from particular resource mgr (file system, dbms, other)
    - Identifies application data
    - ...
  - o base data services
    - datadescription (ogsi service data elements representing key params of the data virtualization encapsulated by the data svc)
    - dataAccess (provides operations to access and/or modify the contents of the data virtualization...)
    - dataFactory
    - dataMgmt
  - o Other dataService interfaces
    - Data replication
    - Data caching
    - Schema transformation

- Data service implement one or more data interfaces and perhaps other ogsa interfaces; extended version of base data interfaces; base data interfaces; ogsi agreement....
  - Supports transformation and federation and is composable!
  - Supports transparent replication and caching
- Savas: Seems complex. Also question on the need for data virtualization.
- Jay: A service is a light abstraction. (A service is not a 'thing'.)
- Fred: With the experience of working for a major vendor of storage systems, data virtualization is essential.

#### Norman Paton, OGSA-DAI

- requirements
  - to provide service interfaces to data resources
    - relational, xml repository, files
  - to support higher-level data integration services
    - comprehensive metadata
    - flexible data delivery
- key concepts
  - service-oriented treatment of structured data access
    - specification collection of interfaces
      - top-level porttypes of ogsa data services proposal
        - relational
        - xml
    - data service port types (on top of agreement and ogsi interfaces)
      - Data description (schema of relational db)
      - Data access (evaluate an sql query)
      - Data factory (virtualize a query result as a data svc)
      - Data mgmt (not db administration; role mapping)
    - Related specifications on the table
      - Grid data distribution service
      - Generic data movement service
    - Relationship to grid data service spec
      - Delivering data to/from data access services is important
      - Knowing where to stop ...
- Relationship with invited groups
  - OREP-wg (overlap; service provider)
    - May want to replicate data from/to dbs
    - Share need to manage and convey datasets
  - PA-RG (service provider)
    - Dbs are an archiving technology
    - Metadata catalogues managed by dbs
  - GFS-WG
- Dave Barry: How does data virtualization map on to this?

- Norm: If DAIS define interfaces in the appropriate way, they won't be super-dependent on what's happening underneath
- Metadata exposes structural characteristics of data; don't need an ontology for that. Metadata does not give a semantical view of underlying data.

OREP, Ann Chervenak

- building on data notion
- topics in the group
  - replicated data items for an equivalence class
  - define replicate set service
  - replica semantic policies
  - higher level index services
  - specification share
- data distribution specification currently living in DAIS group

Reagan More, Preservation data services

- goal
  - build a collection that you will maintain while all underlying implementation will change
    - all system components change over time; must be possible to keep providing access to data.
- requirements
  - need variety of interfaces
  - manage consistency between context (state information resulting from service) and content (digital entities)
  - support transformative migrations between data types
  - manage authenticity
  - support persistent archive
- key concepts
  - automation of all archival processes
    - logical name spaces (close in concept to OGSI GSH)
    - build a persistent service
    - support collection-owned data
    - manage logical name space as a collection hierarchy
    - provide bulk operations
  - archival processes to generate archival context
  - consistency between context and content
    - consistent mgmt of state info generated by services
    - consistency on bulk operations
      - access = manipulation + transport
- GGF standards interactions
  - Data format description language
  - Data transport
  - Grid file system
  - Grid protocol architecture

- Not clear yet whether the data abstractions defined by OGSA are what's needed to manage replica collections; more refinement required.
- Other stds
  - Semantic web ontology web language
  - Digital library federation metadata encoding and transmission std
  - NSFdigital library initiative open archive initiative
  - Nasa/nara open archival information system

#### Osamu Tatebe (AIST), Grid File System

- initial stage, group is almost approved
- grid file system federates and shares virtualized data from file systems in a grid
  - virtual hierarchical namespace with access permission and metadata
  - reliable posix-like i/o interfaces for the grid file system
- requirements
  - address need for mgmt of millions or billions of file-based data dispersed in a grid
  - address need for sharing file mgmt
  - provide posix-like i/o interfaces
- key concepts
  - service-oriented treatment of file system( file data mgmt and access)
    - virtualized global logical hierarchical namespace
    - file system directory tree
      - GSHs for each file system directory
      - Shared file system (sub-)tree and access ctrl
  - interaction with file access services and replica services
- proposed services
  - virtual file system directory services (VFDS)
    - file system metadata mgmt services
      - virtual file system directory
      - owner, file type, acl, access times, size, ...
      - MIME data type
    - Lookup services with acls
  - Grid file system services
    - Extends vfds and file access services
- Relation with other svcs and groups
  - May be included in data svcs and replica svcs (dais, orep)
- Jay: Do you envision building this on top of DAIS or next to DAIS? Will it be integrated with DAIS or as an alternative to DAIS?
  - Depends on Data Services definition(?) It should be possible to map posix i/f to data virtualization currently proposed
- (Worry that if new and old i/f is provided people will use old even if the old i/f is not as good.)
- Why is the posix i/f important? Is it doable/good mapping with ogas data services.

- Complexity of building filesystem on top of SOA; take pragmatical approach of using what is available now (direct file access) and later when SOA proves itself can move to it.
- Not against posix i/f but against posix i/f without DAIS underneath. SOA offers an opportunity to manage not just access to data but also the entities accessing data. (Do we need filesystem concept even?)
- OGSA challenge: Manage interactions between services and maintaining consistency.