# eGee

# GLUE Schema Configuration for SRM 2.2

*Stephen Burke, RAL*

*GSSD, October 9th 2007*

**www.eu-egee.org**

Information Society
and Media

**Enabling Grids for E-sciencE**

- **What is the schema for?**

- **Some history**

- **Details of the schema objects and attributes**

- **Summary**

- **References**

- **The schema is not a model of a system**
- **It publishes attributes which are needed to fulfil specific use cases**
  - Finding services/resources and choosing between them
    - Client tools
    - Direct queries by users
  - Overall Grid monitoring (e.g. gstat)
    - Including summary information for managers
    - (Accounting – but only for a high-level view)
- **Total information volume needs to be limited**
  - Use the information system for a fast query covering all resources in the Grid
  - Use direct queries for detailed information
- **The schema must cope with all implementations & configurations in a reasonable way**
  - May need some simplifying assumptions
  - Formally, nearly all attributes are optional
- **The schema must be implementable in LDAP!**

**Enabling Grids for E-sciencE**

- **The GLUE schema needs to be agreed between many parties, hence agreement can be hard to reach**
  - The list is growing as GLUE is adopted more widely
- **Deploying a schema update takes a long time (1-2 years)**
  - Introduce Capability attributes – arbitrary text tags which give limited scope to add extra information
  - Also GlueServiceData – key/value pairs for services
- **So far we have required upgrades to be backwards-compatible**
  - Sites upgrade gradually
  - Client tools are adapted gradually
  - Accumulation of deprecated items, limits on structural change
- **GLUE 2.0 is now being developed in the OGF framework**
  - Not backwards-compatible
  - Will probably have to deploy in parallel
  - Get involved now if you want to influence it!

Enabling Grids for E-sciencE

- **GLUE schema first defined in 2003**

- **Original SE schema (1.1) was for "classic SE" – simple disk server + gridftp**
  - Plus other access protocols, e.g. rfio, file

- **Version 1.2 defined in 2005**
  - Fairly small changes for SE:
    - New ControlProtocol, GlueService objects
    - Decouple SA ID from path
    - Introduce SAType (permanent, durable, volatile)
    - Introduce Capabilities for protocols
  - Classic SE was still the standard

- **In practice many of the 1.2 changes are still unused**

**Enabling Grids for E-sciencE**

- **Small window of opportunity to upgrade the schema at the end of 2006**
  - SRM 2.2 was the main reason for the upgrade
  - Must be backward compatible
  - Changes should be as small as possible
  - Should be motivated by real use cases
    - User queries, client tools, monitoring/accounting – but not SRM management
- **Many discussions in the SRM mailing list, + two documents with proposals**
  - Documents were iterated several times
- **Face-to-face meeting agreed changes**
  - Some things in the proposals were left out as being too complicated and not clearly motivated, notably the Storage Component (modelling the storage hardware)

**Enabling Grids for E-sciencE**

- **Storage Element (GlueSE)**
  - Summary information for the whole SE
- **Storage Library (GlueSL)**
  - Storage hardware
  - Never used, obsolete since 1.2
- **Storage Space (Area) (GlueSA)**
  - Area of storage assigned to one or more VOs
  - Schema allows multiple SAs per VO or multiple VOs per SA, but client tools generally don't
- **Access protocol (GlueSEAccessProtocol)**
  - gridftp, rfio, dcap, file, …
- **ControlProtocol (GlueSEControlProtocol)**
  - Should be SRM endpoint, or dummy for classic SE
  - Not used in practice so far
- **Service (GlueService)**
  - Endpoints for all services, including SRM

**Enabling Grids for E-sciencE**

- UniqueID: opaque ID (should not be assumed to be the hostname)

- Name: human-friendly name (no semantics)

- InformationServiceURL: URL for the SE GRIS (not used in practice)

- Architecture: disk, tape, multidisk, other (not used in practice)

- ImplementationName: DPM, dCache, CASTOR, … (useful!)

- ImplementationVersion: Overall version number (useful!)

- Status: Queueing, Production, Closed, Draining (useful!)

- TotalOnlineSize, TotalNearlineSize, UsedOnlineSize, UsedNearlineSize: Summary information for monitoring/management (useful, but need not be very accurate)

- NB Deprecated items omitted, but in general should still be published where backward compatibility is required

- Attributes in 1.1, 1.2, 1.3

**Enabling Grids for E-sciencE**

- LocalID: Locally unique ID (no semantics)
- Type: rfio, dcap, … (used by M/W)
- Version: Protocol version (not used?)
- Endpoint: Protocol endpoint (not needed with SRM)
- Capability: Arbitrary text tags (not used so far)
- MaxStreams: Maximum number of parallel streams allowed (not used yet but may be useful?)

- Q: Where is the list of protocol types supported by SRM?

**Enabling Grids for E-sciencE**

- LocalID: Locally unique ID (no semantics)
- Type: SRM, classic (formerly used srm_v1)
- Version: Protocol version (not always published correctly)
- Endpoint: URL of the SRM (not always published correctly)
- Capability: Arbitrary text tags to identify non-standard features (not used yet but may be useful?)

- Can have multiple CPs per SE, e.g. for SRM1 and 2

- Note that clients currently take the SRM endpoint from GlueService. For anything which needs other information from the SE this is not ideal, as there is no good way to link GlueService to GlueSE – assuming a match between the hostname and the SEUniqueID is unreliable.

**Enabling Grids for E-sciencE**

- LocalID: Locally unique ID (no semantics)
- Name: Human-friendly name (no semantics)
- Path: SURL prefix to use when writing files (used by M/W)
- AccessControlBaseRule: Authorisation for writing (VO names or FQANs)
- RetentionPolicy: custodial, output, replica
- AccessLatency: online, nearline, offline
- ExpirationMode: neverExpire, warnWhenExpired, releaseWhenExpired
- Capability: Arbitrary text tags to identify non-standard features (not used yet)
- TotalOnlineSize, UsedOnlineSize, FreeOnlineSize, ReservedOnlineSize, TotalNearlineSize, UsedNearlineSize, FreeNearlineSize, ReservedNearlineSize: Various information about the size of the space (to be discussed …)

- GLUE 1.3 agreement was that GlueSA == SRM space, hence RetentionPolicy et al

- LocalID: Locally unique ID (no semantics)

- Name: Human-friendly name (no semantics)

- Path: SURL prefix to use when writing files (overrides SAPath if present)

- Tag: Space token description (unique per VO)

- AccessControlBaseRule: Authorisation for writing (same as or subset of the SA ACBR)


- Separate object due to LDAP restrictions, but logically part of parent SA

- Paths could be omitted (or /) but there appear to be many requests to be able to specify different default paths for writing

- ChangeSpaceForFiles means that you cannot in general do a reverse lookup from an existing SURL to an SA

- Allows different VOs/groups to share an SA with different space tokens/paths (common use case for sites with many "small" VOs)

- UniqueID: opaque ID (can be tricky to ensure uniqueness)
- Name: Human-friendly name (no semantics)
- Type: SRM (formerly used srm_v1)
- Version: Protocol version (1.1.0 or 2.2.0)
- Endpoint: URL of the SRM
- Status: OK, Warning, Critical, Unknown, Other (currently not useful)
- StatusInfo: Text to explain the Status (currently not useful)
- WSDL: URL of a WSDL document (currently not defined)
- Semantics: URL of a manual or specification (currently not defined)
- StartTime: The service start time (currently not useful)
- Owner: VO name (not relevant for SRM)
- AccessControlBaseRule: Service authorisation (should be set?)

- Also legacy LCG attributes (never part of GLUE)
- Current information provider is static so dynamic attributes aren't useful
- A dynamic information provider is in certification

**eGee**

Enabling Grids for E-sciencE

- **https://twiki.cern.ch/twiki/bin/view/LCG/GSSDGLUEExample**

- **GlueSE: not yet updated for 1.3**
  - Name and Type are wrong, Port should be removed
  - New attributes should be added
- **GlueSL: obsolete, not used by anything**
- **GlueSA:**
  - "Old" SA (LocalID == VO name, no VOInfo) for backward-compatibility (need to check behaviour)
  - "New" SAs generally OK (free space = 0?)
  - Some deprecated attributes (size, lifetime) included for backward-compatibility
- **GlueVOInfo: examples OK**
  - Better to move the VOInfo objects after the corresponding SA
  - Need a VOInfo with Tag=default to mark the default space
- **GlueService: OK, only ACBR is new in 1.3**
  - Remove deprecated attributes?

**Enabling Grids for E-sciencE**

- **GLUE 1.3 was intended to provide the minimum we need to survive for the next 1-2 years**
- **Driven by use cases (see documents)**
  - We decided to exclude management use cases – assumed to be done directly and not via the schema
  - Can anyone give an important use case that is not supported by the current schema?
- **We should publish all meaningful attributes unless there is a strong reason why not**
  - We need to interoperate with other Grids, especially OSG
  - Size attributes will need a dedicated discussion about what is feasible
  - But we should not publish meaningless attributes!
- **Do we have volunteers to write the info providers?**
- **What is planned on the client side – lcg-utils etc?**

- **GLUE 1.2: http://glueschema.forge.cnaf.infn.it/Spec/V12**

- **GLUE 1.3: http://glueschema.forge.cnaf.infn.it/Spec/V13**

- **GLUE 2: http://forge.ogf.org/sf/projects/glue-wg**

- **Input documents to SRM 1.3 discussion:**
- **http://forge.ogf.org/sf/docman/do/downloadDocument/projects.glue-wg/docman.root.background.specifications/doc14619**

- **http://listserv.fnal.gov/scripts/wa.exe?A3=ind0611c&L=srm-devel&P=214145&E=2&B=--Boundary_%28ID_EryUZRmYbVia2QzE4Pzreg%29&N=GLUE-SE-1.3-input-1.03.pdf&T=application%2Fdownload**

- **GSSD schema example:**
  **https://twiki.cern.ch/twiki/bin/view/LCG/GSSDGLUEExample**