GLUE Schema Configuration for SRM 2.2

Stephen Burke, RAL
(with contributions from Maarten Litmaath)

SRM Workshop, November 13th 2007

Information Society and Media

Overview

- What is the schema for?
- Some history
- Details of the schema objects and attributes
- Where are we going?
- Summary
- References

- Detailed schema examples

Schema philosophy

- **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 (lcg-utils/GFAL/FTS)
    - Direct queries by users (lcg-info, ldapsearch)
  - Overall Grid monitoring (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!**

Schema evolution

- **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
  - In discussion now, probably ~1 year from production

History

- **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**

GLUE 1.3

- **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
- **Much recent discussion about how to use this in practice**
  - Now have a proposal for a way forward

Schema structure

- **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, but client tools generally haven't supported this up to now
- **Access protocol (GlueSEAccessProtocol)**
  - gridftp, rfio, dcap, file, …
- **ControlProtocol (GlueSEControlProtocol)**
  - Should be the SRM endpoint(s), or dummy for classic SE
  - Not used in practice so far
- **Service (GlueService)**
  - Endpoints for all services, including SRM

GlueSE

- 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

AccessProtocol

- 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?)

ControlProtocol

- 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.

- 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 …)

VOInfo

- 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
- Allows different VOs/groups to share an SA with different space tokens/paths (common use case for sites with many “small” VOs)
  - **But not supported by SRM implementations?**

GlueService

- UniqueID: opaque ID
- 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

GLUE 1.3 for SRM v2.2

- **GLUE 1.3 SE defined to allow publication of SRM v2.2 properties**
  - Retention Policy
    - 1 per SA
  - Access Latency
    - 1 per SA
  - Expiration Mode
    - Always “NeverExpire” in the WLCG MoU
  - Space token user descriptions
    - 0 or more per SA
  - New size attributes for Online and Nearline
    - SE has {Total,Used}{Online,Nearline}Size (4)
    - SA has {Total,Used,Reserved,Free}{Online,Nearline}Size (8)
    - Nearline sizes do not make sense for Replica-Online (T0D1)
    - Some of these numbers may be hard to provide in practice
      - *At least on some SE implementations*
      - *Sensibility ←→ accuracy ←→ cost ←→ usefulness*
      - *Publish the “easy” ones now, check what remains desirable later*

What is an SA?

- **For now: a storage “basin” with given Retention Policy and Access Latency, out of which SRM v2.2 spaces can be carved**
  - Limiting case: SA == space
    - That extreme encountered technical objections from some implementations
  - Plan: implement simple schema, see if all important use cases are met
    - Enhancements later, when experience has been gained
  - An SE would have at most 3 SAs per VO (T1D0, T1D1, T0D1)
    - Each SA can have multiple space token user descriptions pointing to it

- **Still need to publish GLUE 1.2 SA for backward compatibility with “broken” clients currently in production**
  - GFAL looks for “GlueSALocalID == VO”
    - And “GlueAccessControl(Base)Rule == VO”
  - Bug turned out handy to protect old clients from new info!

Example and testing

- https://twiki.cern.ch/twiki/bin/view/LCG/GSSDGLUEExample
- **Example for atlas, dteam, lhcb and ops**
- **VOs, paths, space token tags, hostname and site name to be adapted**
- **Spaces for dteam may be shared with ops**
- **Dteam/ops tokens**
  - T1 site should at least implement srm2_d0t1 and srm2_d1t0
  - T2 site should implement srm2_d1t0

- **Conformance checked by S2 test suite**
  - http://lxdev25.cern.ch/s2test/bdii/
- **Latest results**
  - http://lxdev25.cern.ch/s2test/bdii/s2_logs/
- **BDII check scripts available**
  - https://twiki.cern.ch/twiki/pub/LCG/GSSD/bdii-publication-check.tar.gz

Where do we go next?

- **Write information providers which publish the new information**
- **Update clients to use it, especially lcg-utils/GFAL**
- **May need to iterate with experience**

- **There is now an OGF working group producing a GLUE 2.0 specification**
  - Not backward-compatible
  - Rationalise structure and attributes
  - In discussion now, hope for agreement by end of year
  - But probably not in production until 2009!
  - Will anyway have to run in parallel with 1.3 for some time

Summary

- **GLUE 1.3 was intended to provide the minimum we need to survive for the next 1-2 years**
- **Driven by use cases**
  - We decided to exclude management use cases – assumed to be done directly and not via the 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 more discussion about what is feasible
  - But we should not publish meaningless attributes!
- **We now have a template for what to publish – need to try it out and see how it goes**

References

- 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

Glue SE record

```
dn: GlueSEUniqueID=ccsrmtestv2.in2p3.fr,mds-vo-
  name=IN2P3-CC,o=grid
[...]
GlueSEUniqueID: ccsrmtestv2.in2p3.fr
GlueSEName: ccsrmtestv2.in2p3.fr
GlueSEPort: 8443
GlueSEArchitecture: tape
GlueSEStatus: Production
GlueSEImplementationName: dCache
GlueSEImplementationVersion: 1.8.0-26
GlueSESizeTotal: 9730000
GlueSESizeFree: 6355000
GlueSETotalOnlineSize: 220000
GlueSEUsedOnlineSize: 198000
GlueSETotalNearlineSize: 9510000
GlueSEUsedNearlineSize: 3177000
[...]
```

← No ":srm_vN"

Glue 1.2 SA

```
dn: GlueSALocalID=lhcb,GlueSEUniqueID=
 ccsrmtestv2.in2p3.fr,mds-vo-name=IN2P3-CC,o=grid
[…]
GlueSAPath: /pnfs/in2p3.fr/data/lhcb ← No more SARoot
GlueSALocalID: lhcb ← VO name
GlueSAPolicyFileLifeTime: permanent
GlueSAStateAvailableSpace: 300000000000
GlueSAStateUsedSpace: 160000000000
GlueSAAccessControlBaseRule: lhcb ← keep old syntax
GlueSAAccessControlBaseRule: VO:lhcb ← add new syntax
[…]
```

Custodial-Nearline SA

```
dn: GlueSALocalID=lhcb:custodial:nearline,
 GlueSEUniqueID=ccsrmtestv2.in2p3.fr,mds-vo-name=IN2P3-CC,o=grid
[…]
GlueSAPath: /pnfs/in2p3.fr/data/lhcb
GlueSALocalID: lhcb:custodial:nearline
GlueSAPolicyFileLifeTime: permanent
GlueSAStateAvailableSpace: 300000000000   ← in kB
GlueSAStateUsedSpace: 160000000000   ← in kB
GlueSAAccessControlBaseRule: lhcb ← old syntax to be dropped
GlueSAAccessControlBaseRule: VO:lhcb ← new syntax
GlueSARetentionPolicy: custodial
GlueSAAccessLatency: nearline
GlueSAExpirationMode: neverExpire
GlueSATotalOnlineSize: 70000 ← disk cache for tape back-end
GlueSAUsedOnlineSize: 50000
GlueSAFreeOnlineSize: 20000
GlueSAReservedOnlineSize: 0
GlueSATotalNearlineSize: 300000   ← tape back-end
GlueSAUsedNearlineSize: 160000
GlueSAFreeNearlineSize: 140000
GlueSAReservedNearlineSize: 0
[…]
```

Custodial-Nearline space 1

```
dn: GlueVOInfoLocalID=lhcb:LHCb_RAW,GlueSALocalID=
 lhcb:custodial:nearline,GlueSEUniqueID=
 ccsrmtestv2.in2p3.fr,mds-vo-name=IN2P3-CC,o=grid
[…]
GlueVOInfoLocalID: lhcb:LHCb RAW
GlueVOInfoName: lhcb:LHCb_RAW
GlueVOInfoPath: /pnfs/in2p3.fr/data/lhcb
GlueVOInfoTag: LHCb_RAW
GlueVOInfoAccessControlBaseRule: VO:lhcb
GlueChunkKey: GlueSALocalID=lhcb:custodial:nearline
GlueChunkKey: GlueSEUniqueID=ccsrmtestv2.in2p3.fr
[…]
```

Custodial-Nearline space 2

```
dn: GlueVOInfoLocalID=lhcb:LHCb_RDST,GlueSALocalID=
 lhcb:custodial:nearline,GlueSEUniqueID=
 ccsrmtestv2.in2p3.fr,mds-vo-name=IN2P3-CC,o=grid
[…]
GlueVOInfoLocalID: lhcb:LHCb RDST
GlueVOInfoName: lhcb:LHCb_RDST
GlueVOInfoPath: /pnfs/in2p3.fr/data/lhcb
GlueVOInfoTag: LHCb_RDST
GlueVOInfoAccessControlBaseRule: VO:lhcb
GlueChunkKey: GlueSALocalID=lhcb:custodial:nearline
GlueChunkKey: GlueSEUniqueID=ccsrmtestv2.in2p3.fr
[…]
```

Replica-Online SA

```
dn: GlueSALocalID=atlas:replica:online,GlueSEUniqueID=
 ccsrmtestv2.in2p3.fr,mds-vo-name=IN2P3-CC,o=grid
[…]
GlueSAPath: /pnfs/in2p3.fr/data/atlas/disk   ← normally VO root !
GlueSALocalID: atlas:replica:online
GlueSAPolicyFileLifeTime: permanent
GlueSAStateAvailableSpace: 10000000000
GlueSAStateUsedSpace: 190000000000
GlueSAAccessControlBaseRule: atlas
GlueSAAccessControlBaseRule: VO:atlas
GlueSARetentionPolicy: replica
GlueSAAccessLatency: online
GlueSAExpirationMode: neverExpire
GlueSATotalOnlineSize: 200000   ← disk space managed by VO
GlueSAUsedOnlineSize: 190000
GlueSAFreeOnlineSize: 10000
GlueSAReservedOnlineSize: 0
GlueSATotalNearlineSize: 0   ← 0 nearline sizes need not be published
GlueSAUsedNearlineSize: 0
GlueSAFreeNearlineSize: 0
GlueSAReservedNearlineSize: 0
[…]
```

Replica-Online space

```
dn: GlueVOInfoLocalID=atlas:DISK,GlueSALocalID=
 atlas:replica:online,GlueSEUniqueID=
 ccsrmtestv2.in2p3.fr,mds-vo-name=IN2P3-CC,o=grid
[…]
GlueVOInfoLocalID: atlas:DISK
GlueVOInfoName: atlas:DISK
GlueVOInfoPath: /pnfs/in2p3.fr/data/atlas/disk  (*)
GlueVOInfoTag: DISK
GlueVOInfoAccessControlBaseRule: VO:atlas
GlueChunkKey: GlueSALocalID=atlas:replica:online
GlueChunkKey: GlueSEUniqueID=ccsrmtestv2.in2p3.fr
[…]
```

(*) Normally VO root !

Custodial-Online SA

```
dn:
   GlueSALocalID=dteam:custodial:online,GlueSEUniqueID=
 ccsrmtestv2.in2p3.fr,mds-vo-name=IN2P3-CC,o=grid
[...]
GlueSAPath: /pnfs/in2p3.fr/data/dteam
GlueSALocalID: dteam:custodial:online
GlueSAPolicyFileLifeTime: permanent
GlueSAStateAvailableSpace: 8000000000
GlueSAStateUsedSpace: 2000000000
GlueSAAccessControlBaseRule: dteam
GlueSAAccessControlBaseRule: ops
GlueSAAccessControlBaseRule: VO:dteam
GlueSAAccessControlBaseRule: VO:ops   ← can share between
   VOs
GlueSARetentionPolicy: custodial
GlueSAAccessLatency: online
GlueSAExpirationMode: neverExpire
[...]
```

Custodial-Online space 1

```
dn: GlueVOInfoLocalID=dteam:srm2_d1t1,GlueSALocalID=
 dteam:custodial:online,GlueSEUniqueID=
 ccsrmtestv2.in2p3.fr,mds-vo-name=IN2P3-CC,o=grid
[…]
GlueVOInfoLocalID: dteam:srm2 d1t1
GlueVOInfoName: dteam:srm2_d1t1
GlueVOInfoPath: /pnfs/in2p3.fr/data/dteam
GlueVOInfoTag: srm2_d1t1
GlueVOInfoAccessControlBaseRule:
   VOMS:/dteam/Role=lcgadmin   ← not for whole VO
GlueChunkKey: GlueSALocalID=dteam:custodial:online
GlueChunkKey: GlueSEUniqueID=ccsrmtestv2.in2p3.fr
[…]
```

Custodial-Online space 2

```
dn: GlueVOInfoLocalID=ops:srm2_d1t1,GlueSALocalID=
 dteam:custodial:online,GlueSEUniqueID=
 ccsrmtestv2.in2p3.fr,mds-vo-name=IN2P3-CC,o=grid
[…]
GlueVOInfoLocalID: ops:srm2 d1t1
GlueVOInfoName: ops:srm2_d1t1
GlueVOInfoPath: /pnfs/in2p3.fr/data/dteam
GlueVOInfoTag: srm2_d1t1   ← private namespace per VO
GlueVOInfoAccessControlBaseRule: VO:ops
GlueChunkKey: GlueSALocalID=dteam:custodial:online
GlueChunkKey: GlueSEUniqueID=ccsrmtestv2.in2p3.fr
[…]
```

Control Protocol records

```
dn: GlueSEControlProtocolLocalID=srm_v2.2,
 GlueSEUniqueID=ccsrmtestv2.in2p3.fr,
 mds-vo-name=IN2P3-CC,o=grid
[…]
GlueSEControlProtocolType: SRM
GlueSEControlProtocolLocalID: srm_v2.2
GlueSEControlProtocolVersion: 2.2.0
GlueSEControlProtocolEndpoint:
   httpg://ccsrmtestv2.in2p3.fr:8443/srm/managerv2
[…]
```

- And similar for SRM v1.1 if supported
- Allow for proper way to detect SRM service at site

SRM Service records

```
dn: GlueServiceUniqueID=httpg://ccsrmtestv2.in2p3.fr:8443/srm/managerv2,
 mds-vo-name=IN2P3-CC,o=grid
[…]
GlueServiceUniqueID: httpg://ccsrmtestv2.in2p3.fr:8443/srm/managerv2
GlueServiceName: IN2P3-CC-SRM-2.2.0
GlueServiceType: SRM   ← not "srm"
GlueServiceVersion: 2.2.0
GlueServiceEndpoint: httpg://ccsrmtestv2.in2p3.fr:8443/srm/managerv2
GlueServiceURI: httpg://ccsrmtestv2.in2p3.fr:8443/srm/managerv2   (*)
GlueServiceAccessPointURL:
   httpg://ccsrmtestv2.in2p3.fr:8443/srm/managerv2 (*)
[…]
GlueServiceAccessControlRule: atlas   ← keep old syntax
GlueServiceAccessControlRule: dteam
[…]
GlueServiceAccessControlBaseRule: VO:atlas   ← add new attributes with
GlueServiceAccessControlBaseRule: VO:dteam      new syntax
[…]
```

(*) Keep for compatibility with old clients

- **And similar for SRM v1.1 if supported**