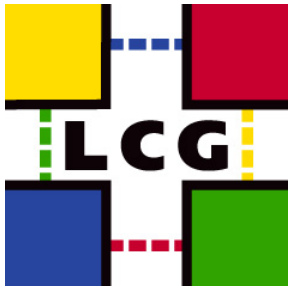
Baseline Services Group



Service Challenge Meeting
Taipei
26th April 2005

Ian Bird
IT/GD, CERN

Overview

- Introduction & Status
  - Goals etc.
  - Membership
  - Meetings
  - Status of discussions
- Baseline services
- SRM
- File Transfer Service
- Catalogues and …
- Future work
- Outlook

Goals

- Experiments and regional centres agree on baseline services
  - Support the computing models for the initial period of LHC
  - Thus must be in operation by September 2006.
- The services concerned are those that
  - supplement the basic services

- Not a middleware group – focus on what the experiments need & how to provide it
  - What is provided by the project, what by experiments?
- Where relevant an agreed fall-back solution should be specified –
  - But fall backs must be available for the SC3 service in 2005.

- scalability/performance metrics.
  - Feasible within next 12 months → for post SC4 (May 2006), & fall-back solutions where not feasible
- When the report is available the project must negotiate, where necessary, work programmes with the software providers.
- Expose experiment plans and ideas

Group Membership

- ALICE: Latchezar Betev
- ATLAS: Miguel Branco, Alessandro de Salvo
- CMS: Peter Elmer, Stefano Lacaprara
- LHCb: Philippe Charpentier, Andrei Tsaragorodtsev
- ARDA: Julia Andreeva
- Apps Area: Dirk Düllmann
- gLite: Erwin Laure
- Sites: Flavia Donno (It), Anders Waananen (Nordic), Steve Traylen (UK), Razvan Popescu, Ruth Pordes (US)

- Chair: Ian Bird
- Secretary: Markus Schulz

Communications

- Mailing list:
  - project-lcg-baseline-services@cern.ch
- Web site:
  - http://cern.ch/lcg/peb/BS
    - Including terminology – it was clear we all meant different things by "PFN", "SURL" etc.
- Agendas: (under PEB):
  - http://agenda.cern.ch/displayLevel.php?fid=3l132
- Presentations, minutes and reports are public and attached to the agenda pages

Overall Status

- Initial meeting was 23rd Feb
- Have been held ~weekly (6 meetings)
  - Introduction – discussion of what baseline services are
  - Presentation of experiment plans/models on Storage management, file transfer, catalogues
  - SRM functionality and Reliable File Transfer
    - Set up sub-groups on these topics
  - Catalogue discussion – overview by experiment
  - Catalogues continued … in depth discussion of issues
  - [Preparation of this report], plan for next month

- A lot of the discussion has been in getting a broad (common/shared) understanding of what the experiments are doing/planning and need
  - Not as simple as agreeing a service and writing down the interfaces!

Baseline services

- We have reached the following initial understanding on what should be regarded as baseline services

- Storage management services
  - Based on SRM as the interface
- gridftp
- Reliable file transfer service
- X File placement service – perhaps later
- Grid catalogue services
- Workload management
  - CE and batch systems seen as essential baseline services,
  - ? WMS not necessarily by all
- Grid monitoring tools and services
  - Focussed on job monitoring – basic level in common, WLM dependent part

See discussion in following slides

- VO management services
  - Clear need for VOMS – limited set of roles, subgroups
- Applications software installation service
- From discussions add:
  - Posix-like I/O service → local files, and include links to catalogues
  - VO agent framework

SRM

- The need for SRM seems to be generally accepted by all
- Jean-Philippe Baud presented the current status of SRM "standard" versions
- Sub group formed (1 person per experiment + J-P) to look at defining a common sub set of functionality
  - ALICE: Latchezar Betev
  - ATLAS: Miguel Branco
  - CMS: Peter Elmer
  - LHCb: Philippe Charpentier

⇨ Expect to define an "LCG-required" SRM functionality set that must be implemented for all LCG sites
  - May in addition have a set of optional functions

- Input to Storage Management workshop

Status of SRM definition

CMS input/comments not included yet

- SRM v1.1 insufficient – mainly lack of pinning
- SRM v3 not required – and timescale too late
- Require Volatile, Permanent space; Durable not practical
- Global space reservation: reserve, release, update (mandatory LHCb, useful ATLAS,ALICE). Compactspace NN
- Permissions on directories mandatory
  - Prefer based on roles and not DN (SRM integrated with VOMS desirable but timescale?)
- Directory functions (except mv) should be implemented asap
- Pin/unpin high priority
- srmGetProtocols useful but not mandatory
- Abort, suspend, resume request : all low priority
- Relative paths in SURL important for ATLAS, LHCb, not for ALICE
- Duplication between srmcopy and a fts – need 1 reliable mechanism
- Group of developers/users started regular meetings to monitor progress

Reliable File Transfer

- James Casey presented the thinking behind and status of the reliable file transfer service (in gLite)
- Interface proposed is that of the gLite FTS
  - Agree that this seems a reasonable starting point
- James has discussed with each of the experiment reps on details and how this might be used
- Discussed in Storage Management Workshop in April
- Members of sub-group
  - ALICE: Latchezar Betev
  - ATLAS: Miguel Branco
  - CMS: Lassi Tuura
  - LHCb: Andrei Tsaregorodtsev
  - LCG: James Casey

fts: *generic file transfer service*
FTS: *gLite implementation*

File transfer – experiment views

Propose gLite FTS as proto-interface for a file transfer service:
(see note drafted by the sub-group)

- CMS:
  - Currently PhedEx used to transfer to CMS sites (inc Tier2), satisfies CMS needs for production and data challenge
  - Highest priority is to have lowest layer (gridftp, SRM), and other local infrastructure available and production quality. Remaining errors handled by PhedEx
  - Work on reliable fts should not detract from this, but integrating as service under PhedEx is not a considerable effort
- ATLAS:
  - DQ implements a fts similar to this (gLite) and works across 3 grid flavours
  - Accept current gLite FTS interface (with current FIFO request queue). Willing to test prior to July.
  - Interface – DQ feed requests into FTS queue.
  - If these tests OK, would want to integrate experiment catalog interactions into the FTS

FTS summary – cont.

- LHCb:
  - Have service with similar architecture, but with request stores at every site
  - Would integrate with FTS by writing agents for VO specific actions (eg catalog), need VO agents at all sites
  - Central request store OK for now, having them at Tier 1s would allow scaling
  - Like to use in Sept for data created in challenge, would like resources in May(?) for integration and creation of agents
- ALICE:
  - See fts layer as service that underlies data placement. Have used aiod for this in DC04.
  - Expect gLite FTS to be tested with other data management service in SC3 – ALICE will participate.
  - Expect implementation to allow for experiment-specific choices of higher level components like file catalogues

File transfer service - summary

- Require base storage and transfer infrastructure (gridftp, SRM) to become available at high priority and demonstrate sufficient quality of service
- All see value in more reliable transfer layer in longer term (relevance between 2 srms?)
  - But this could be srmCopy
- As described the gLite FTS seems to satisfy current requirements and integrating would require modest effort
- Experiments differ on urgency of fts due to differences in their current systems
- Interaction with fts (e.g catalog access) – either in the experiment layer or integrating into FTS workflow
- Regardless of transfer system deployed – need for experiment-specific components to run at both Tier1 and Tier2
- Without a general service, inter-VO scheduling, bandwidth allocation, prioritisation, rapid address of security issues etc. would be difficult

fts – open issues

- Interoperability with other fts' → interfaces
- srmCopy vs file transfer service
- Backup plan and timescale for component acceptance?
  - Timescale for decision for SC3 – end April
  - All experiments currently have an implementation
- How to send a file to multiple destinations?
- What agents are provided by default, as production agents, or as stubs for expts to extend?
- VO specific agents at Tier 1 and Tier 2
  - This is not specific to fts

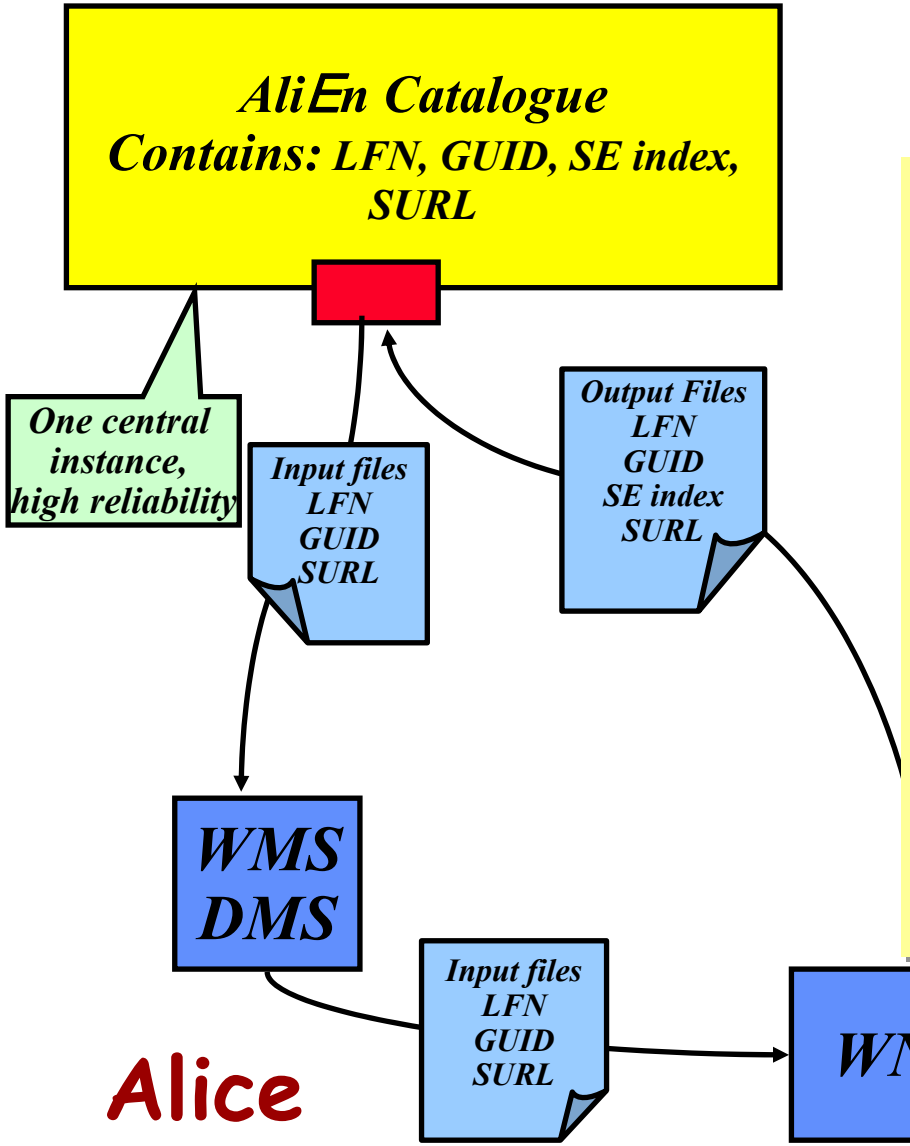
Catalogues

- Subject of discussions over 3 meetings and iteration by email between
- LHCb and ALICE: relatively stable models
- CMS and ATLAS: models still in flux

- Generally:
  - All experiments have *different* views of catalogue models
  - Experiment dependent information is in experiment catalogues
  - All have some form of collection (datasets, …)
    - CMS – define fileblocks as ~TB unit of data management, datasets point to files contained in fileblocks
  - All have role-based security
  - May be used for more than just data files

Catalogues …

- Tried to draw the understanding of the catalogue models (see following slides)
  - Very many issues and discussions arose during this iteration
  - Experiments updated drawings using common terminology to illustrate workflows
  - Drafted a set of questions to be answered by all experiments to build a common understanding of the models
    - Mappings, what, where, when
    - Workflows and needed interfaces
    - Query and update scenarios
    - Etc …
  - → ongoing

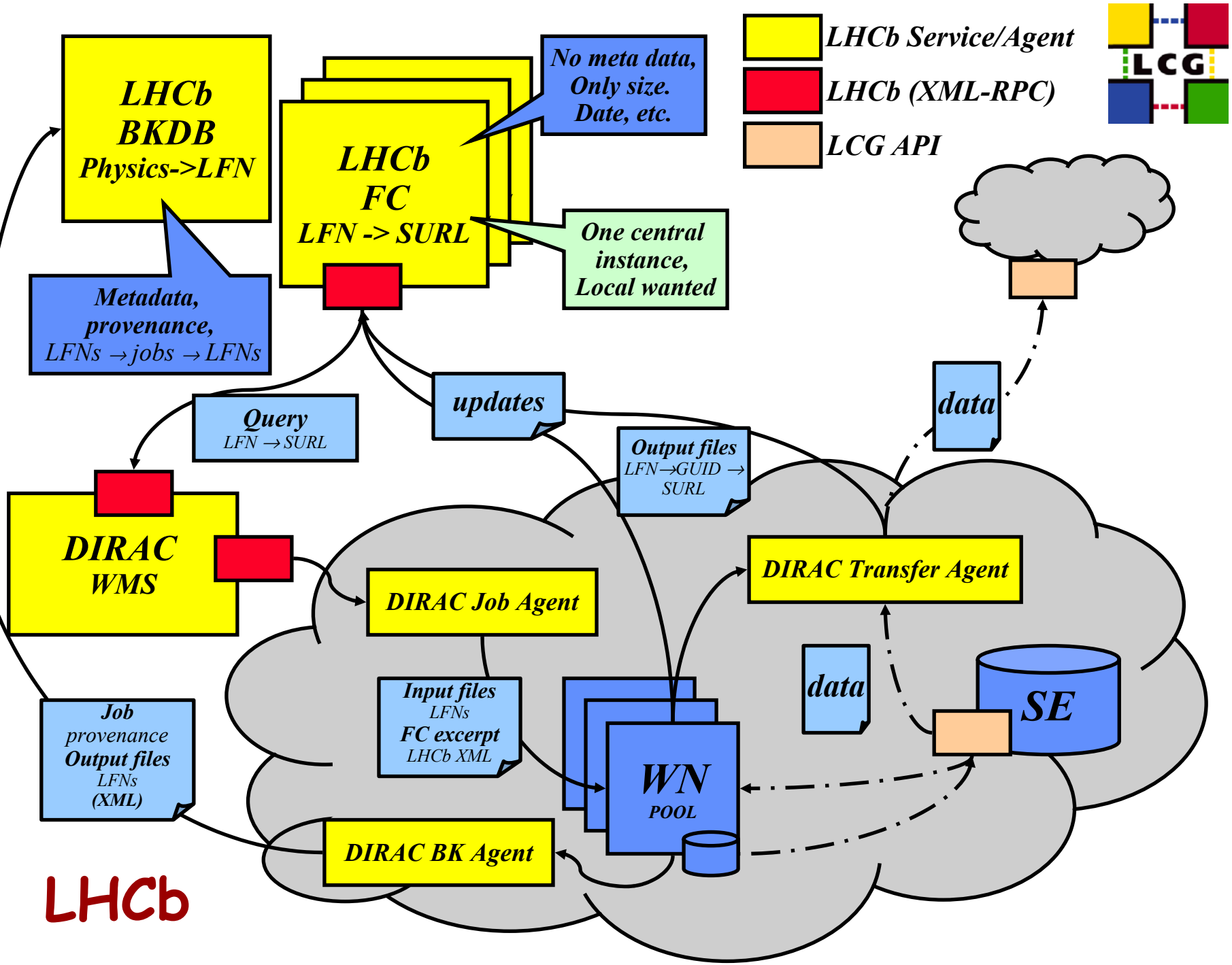


Comments:

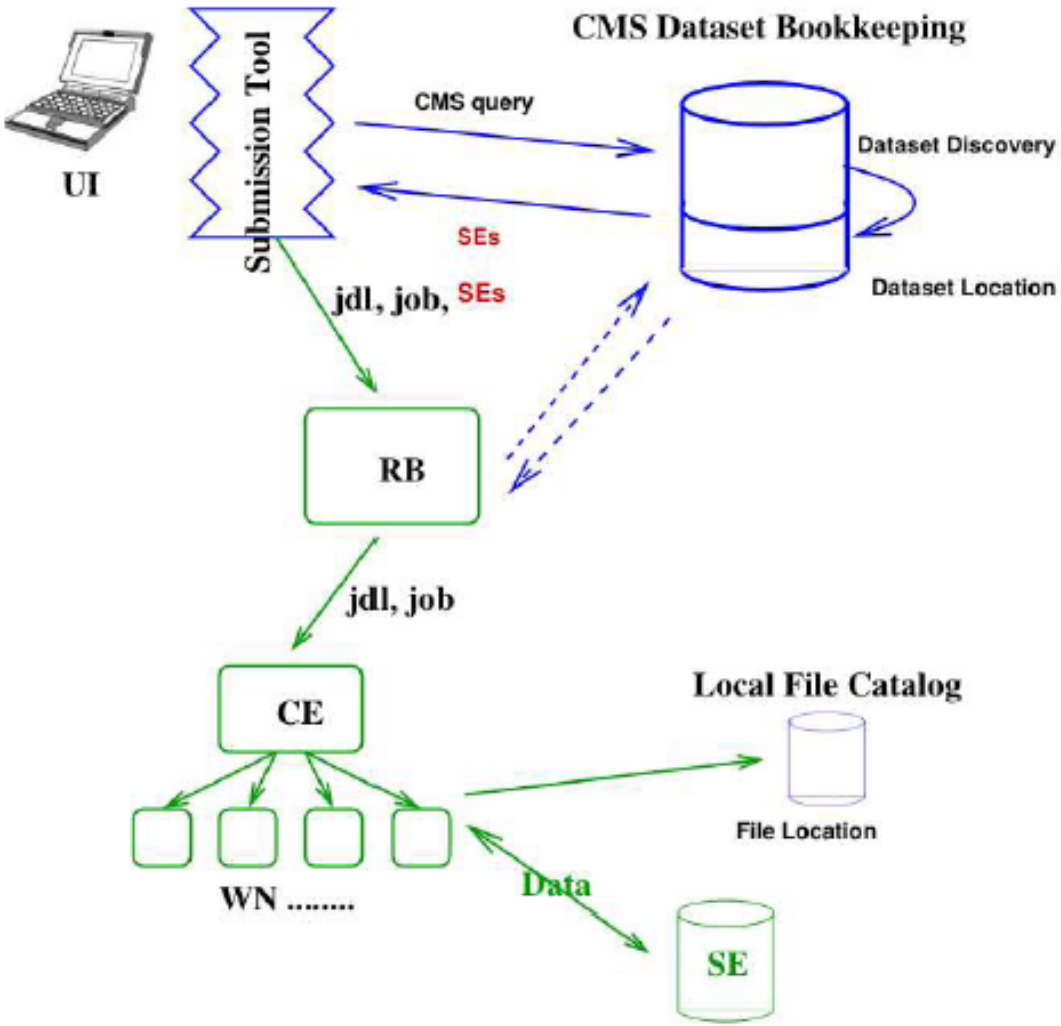
- **Schema shows only FC relations**
- *The DMS implementation is hidden*
- *Ownership of files is set in the FC, underlying storage access management assured by a 'single channel entry'*
- *No difference between 'production' and 'user' jobs*
- *All jobs will have at least one input file in addition to the executable*
- *Synchronous catalog update required*

Job flow diagram shown in:

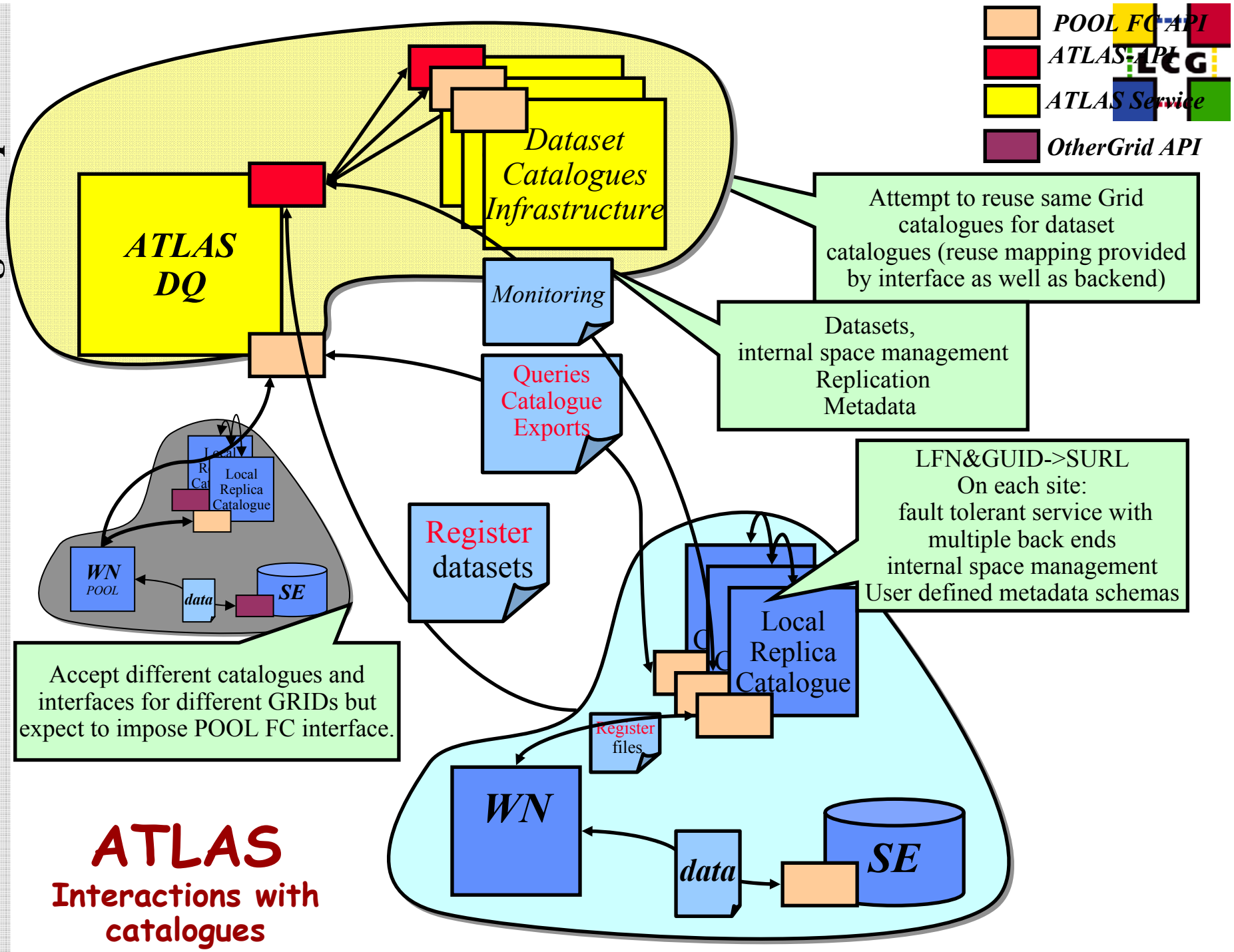
http://agenda.cern.ch/askArchive.php?base=agenda&categ=a051791&id=a051791s1t0/transparencies


LHCb Service/Agent
LHCb (XML-RPC)
LCG API
LCG
LHCb
BKDB
Physics->LFN
LHCb
FC
LFN -> SURL
No meta data,
Only size.
Date, etc.
One central
instance,
Local wanted
Metadata,
provenance,
LFNs → jobs → LFNs
Query
LFN → SURL
updates
Output files
LFN→GUID →
SURL
data
DIRAC
WMS
DIRAC Job Agent
DIRAC Transfer Agent
Input files
LFNs
FC excerpt
LHCb XML
data
SE
WN
POOL
Job
provenance
Output files
LFNs
(XML)
DIRAC BK Agent
LHCb

CMS Baseline



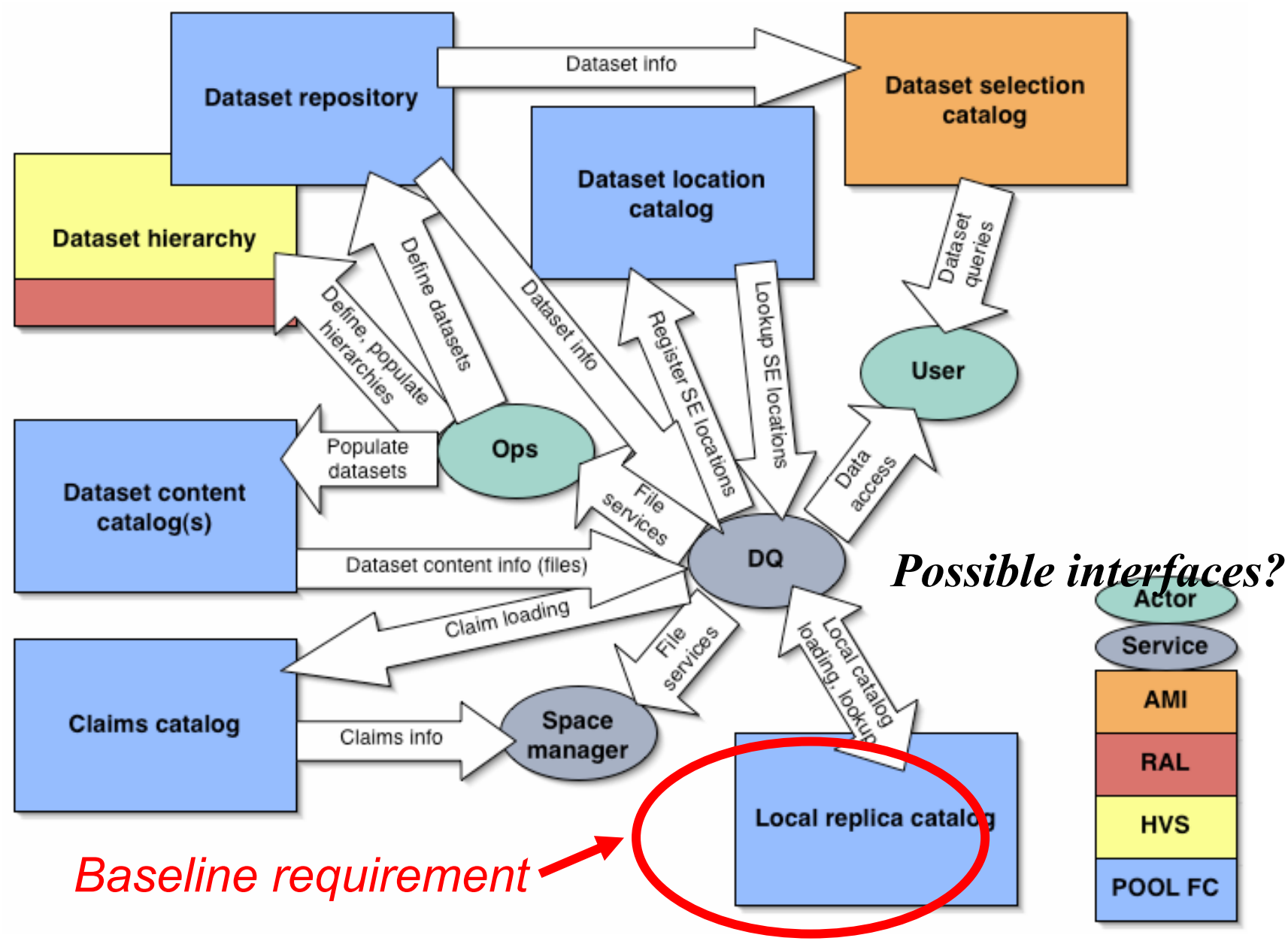
- User on UI:
  - Dataset bookkeeping system
- Either User on UI or RB:
  - Data location service
- Job on worker node:
  - Data access/storage
  - Local file catalog (either "trivial" or standard)
- Output management – no connections from WN outwards, apart from via output sandbox or asynchronous management by Phedex



ATLAS

Interactions with catalogues

Dataset Catalogues Infrastructure (prototype)



Possible interfaces?

Baseline requirement

Summary of catalogue needs

- ALICE:
  - Central (Alien) file catalogue.
  - No requirement for replication
- LHCb:
  - Central file catalogue; experiment bookkeeping
  - Will test Fireman and LFC as file catalogue – selection on functionality/performance
  - No need for replication or local catalogues until single central model fails
- ATLAS:
  - Central dataset catalogue – will use grid-specific solution
  - Local site catalogues (this is their ONLY basic requiremnt) – will test solutions and select on performance/functionality (different on different grids)
- CMS:
  - Central dataset catalogue (expect to be experiment provided)
  - Local site catalogues – or – mapping LFN→SURL; will test various solutions
- No need for distributed catalogues;
- Interest in replication of catalogues (3D project)

Some points on catalogues

- All want access control
  - At directory level in the catalogue
  - Directories in the catalogue for all users
  - Small set of roles (admin, production, etc)
- Access control on storage
  - clear statements that the storage systems must respect a single set of ACLs in identical ways no matter how the access is done (grid, local, Kerberos, …)
    - Users must always be mapped to the same storage user no matter how they address the service
- Interfaces
  - Needed catalogue interfaces:
    - POOL
    - WMS (e.g. Data Location Interface /Storage Index – if want to talk to the RB)
    - gLite-I/O or other Posix-like I/O service

VO specific agents

- VO-specific services/agents
  - Appeared in the discussions of fts, catalogs, etc.
  - This was subject of several long discussions – all experiments need the ability to run "long-lived agents" on a site
    - E.g. LHCb Dirac agents, ALICE: synchronous catalogue update agent
  - At Tier 1 and at Tier 2
  - → how do they get machines for this, who runs it, can we make a generic service framework
  - GD will test with LHCb a CE without a batch queue as a potential solution

Summary

- Will be hard to fully conclude on all areas in 1 month
  - Focus on most essential pieces
  - Produce report covering all areas – but some may have less detail
- Seems to be some interest in continuing this forum in the longer term
  - In-depth technical discussions
  - …