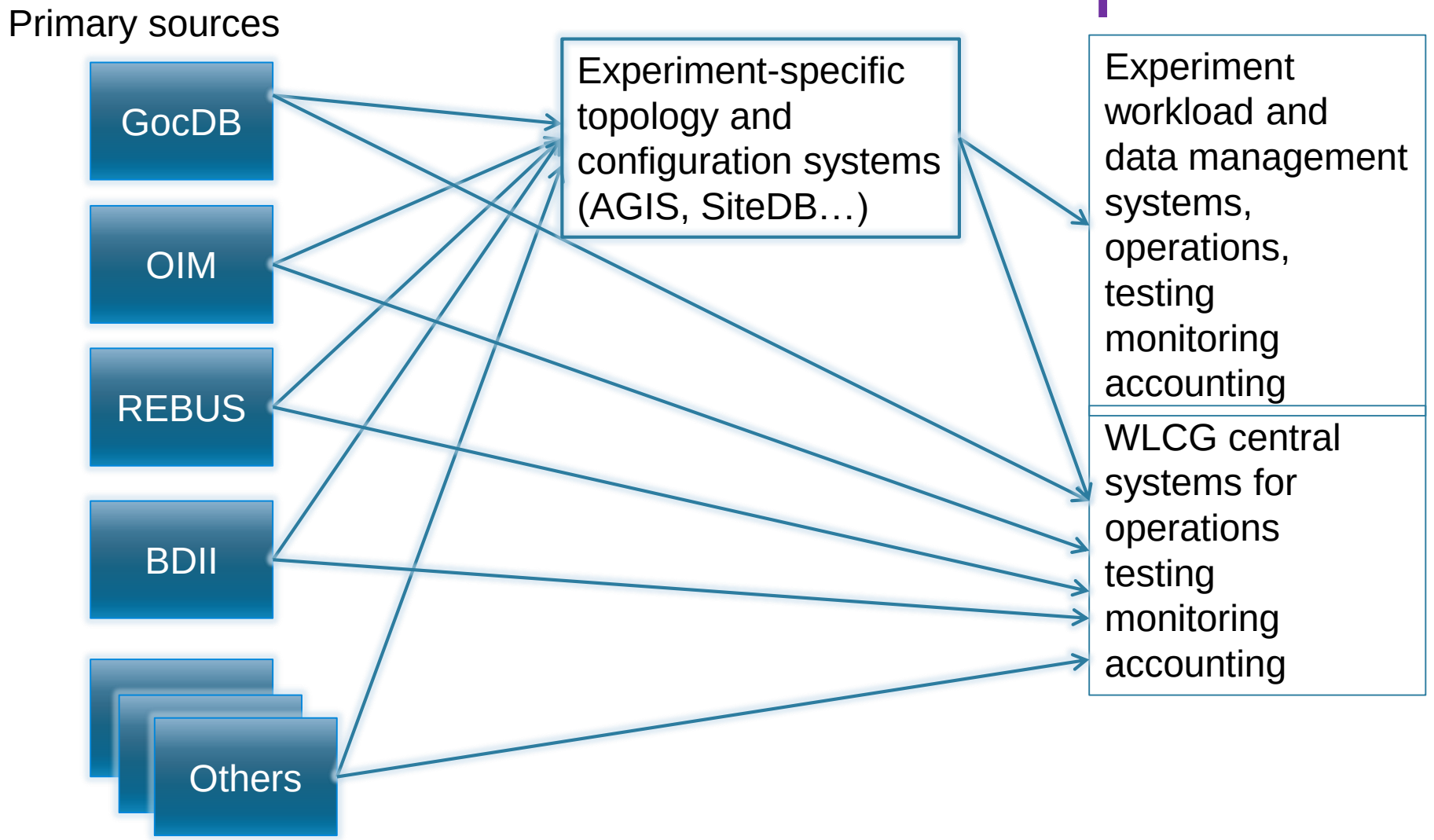
Information System Evolution

WLCG Workshop, Manchester,
21.06.2017

Julia Andreeva CERN IT

WLCG IS Landscape



GocDB and OIM

- Used for static information, mainly for service discovery
- Push more static information into GocDB and OIM. For example queue names

BDII

- Experiments rely on BDII mostly for the description of the computing resources. Attempts to use BDII for complete storage configuration were not successful.
- Contains both static and dynamic information. Which resources at the sites used by a particular VO. LHCb and ALICE use BDII for dynamic information like numbers of pending/running jobs
- OSG stopped using BDII
- The IS Evolution Task Force last year took a decision to make BDII an optional service. Static and semi-static information should move to GocDB/OIM

REBUS

- Management tool
- Contains federation topology and pledges
- For historical reasons REBUS was responsible for generation of the T1 accounting reports. This functionality was moved to the accounting portal.
- Since we foresee that a new component CRIC (Computing Resource Information Catalogue) will provide complete topology and configuration description of the WLCG resources, federation topology should move to CRIC
- Since CRIC will provide administrative UI with fine grained authentication/authorization, pledge definition might move to CRIC in the future as well

Drawbacks and limitations

- Multiple information sources. Sometime contradictory information. No single place where data can be integrated, validated and fixed. No single source providing information for all consumers in a consistent way.
- No complete and correct description of storage resources (with protocols and storage shares). Every experiment has to perform this task on its' own. Therefore, we do not have a global picture on the WLCG scope.
- No clear and common way for integration of non-traditional/non-grid resources (commercial clouds, HPC, etc…). As a rule they are also mainly described in the experiment-specific systems and again missing in the global WLCG resource description.

Clearly, there is a potential to address all these problems in a common way

Lessons learned and moving forward

- During last years we accumulated a lot of operational experience
- Fully distributed IS system like BDII with service information providers looked liked a good technical solution. However, we did not manage to make it a reliable source for all our services, in particular for complex services like storage. Should we think about something more simple but reliable?
- Dependency on the service information providers and therefore often on a particular version of a particular service implementation brings a lot of delays and requires large scale deployment campaigns whenever we need to change anything in our information data sets. If we want to be agile, may be should break or at least relax this dependency? At least for static and semi/static information?
- When we address the common (considering 4 experiments and WLCG operations) issue can we avoid situation when we have to make changes in 4-6 different systems?
- If primary information sources are all distributed and therefore the responsibility is distributed as well, can we effectively debug/fix eventual problems? Apparently not. How to solve this issue? Central intermediate (regarding information flow) service owned by operations can be a place where we perform data integration, validation and correction.