



Science & Technology Facilities Council

e-Science

# GOCDB new model for EGEE-III and beyond

Gilles Mathieu – STFC

GAG meeting, Abingdon

4 December 2008



# Current limitations

- GOCDB3 schema is fine as long as:
  - we don't modify it too much
  - we don't distribute GOCDB, even partially
  - regions don't have specific needs
- But this will happen... and we may face:
  - an increased complexity of the relational model
  - scalability problems
  - Some regions wanting more, and “leaving the ship” to implement their own solution
  - interoperability problems



# Current limitations

- These limitations are mainly due to:
  - The high complexity of the information
  - The unclear definition of future needs
  - The low flexibility of relational DB models
- We must then propose something that:
  - can cope with any level of complexity
  - is flexible enough to be easily modified
  - is customisable to answer regional needs



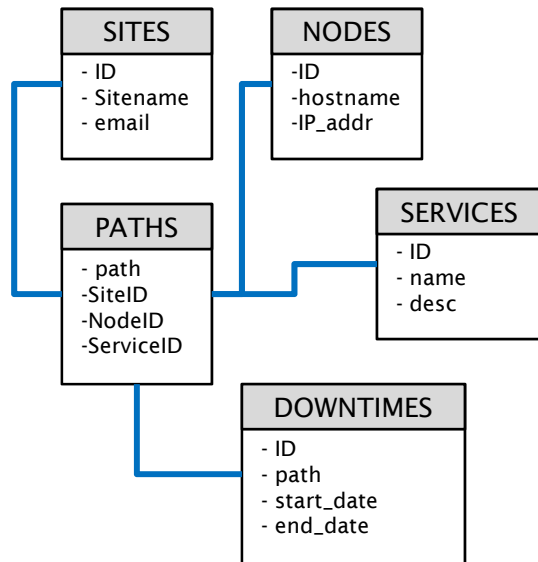
# proposed evolutions

- Keep a central service, not necessarily a central DB
  - There is a need for a central access point, but:
  - the fact that regional DB are distributed or not must not be an issue
- Build a sustainable architecture that allows regionalisation but doesn't force it
  - Not all regions are at the same level
- Propose an implementation where nothing exists, work with existing solutions otherwise
  - Some regions have their own solution and don't want to be forced to use another one



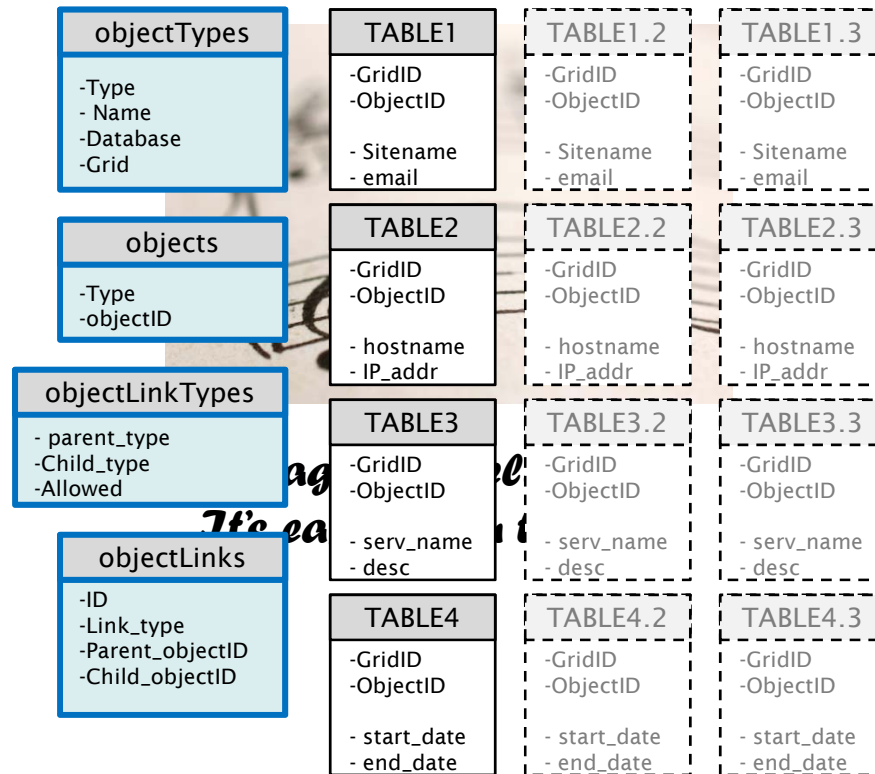
# Proposed DB schema

### Current relational model



- Physical Data Tables
- Hard coded relationships and constraints

### Proposed object model



Core tables  
(relationships)

Data tables

Grid 1

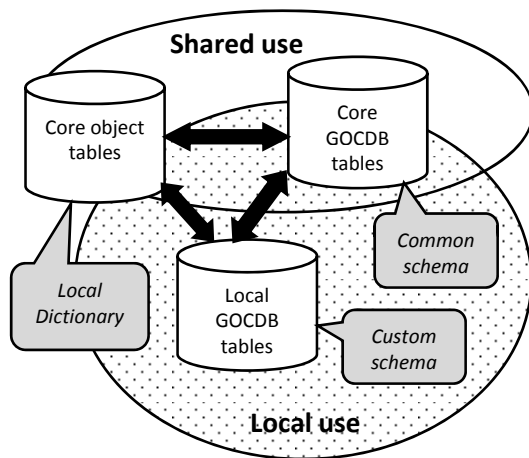
Grid 2

Grid 3

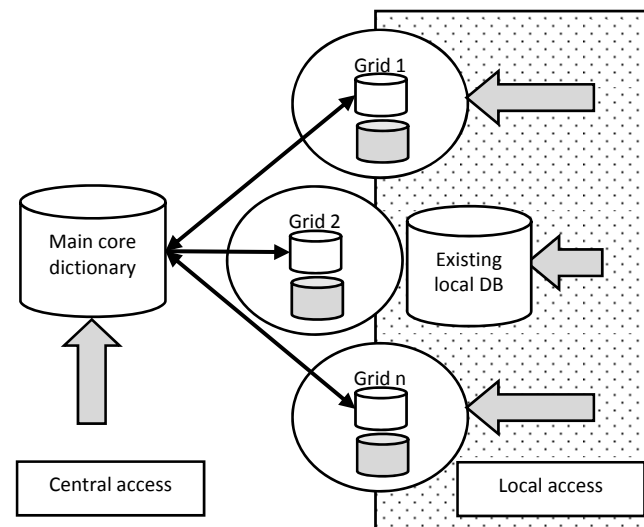


# Proposed architecture

- Logically separated Grids. *We don't care where they are physically.*
- Central/local access
- No data duplication



A logical grid entity  
("regional GOCDB")



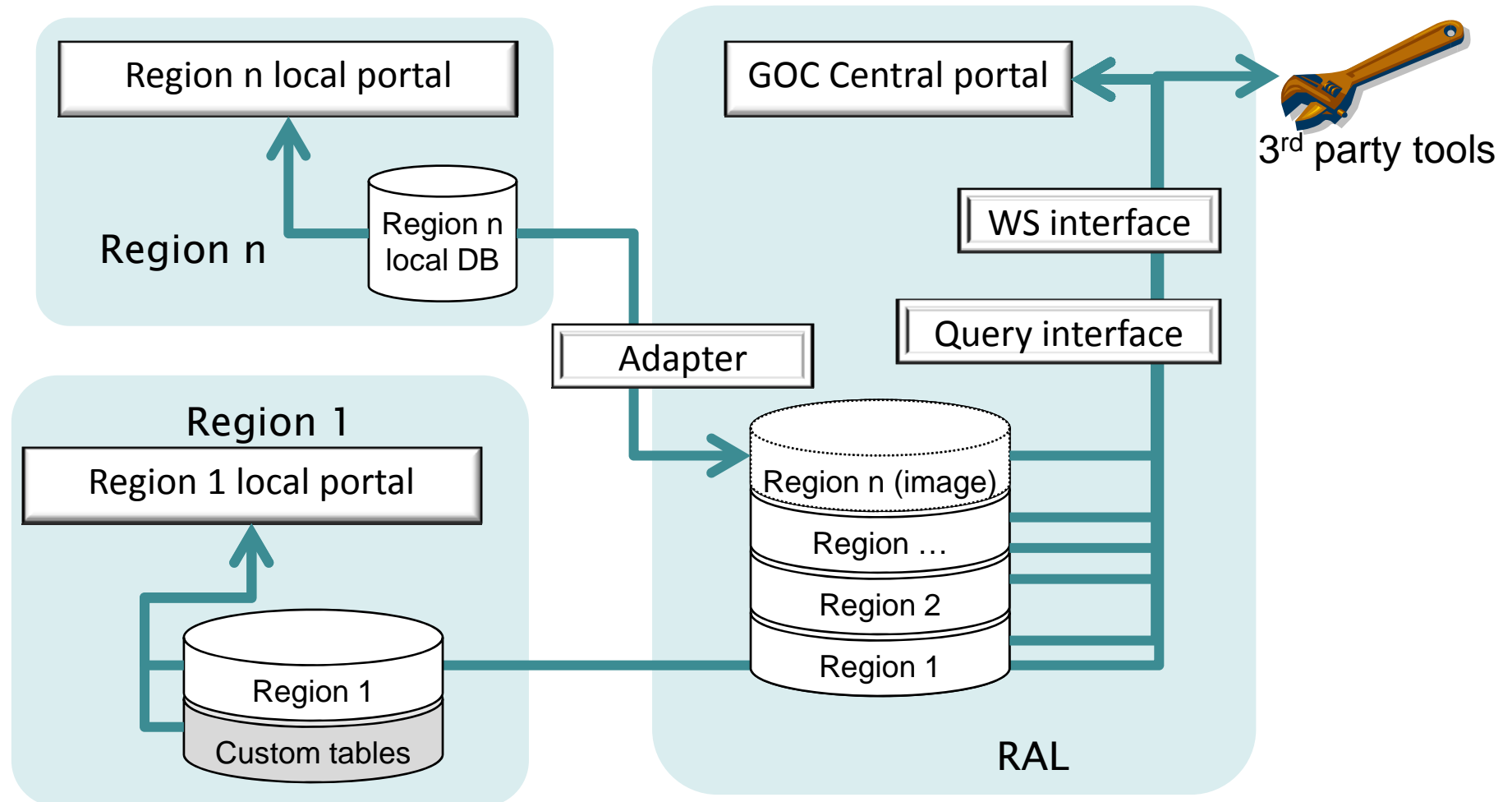
Local entity used directly

Use of an adapter if a different local solution exists

The whole system



# Regionalisation scenario





# Considerations

- Administration of local systems
  - responsibility (and maintenance) comes with hosting
  - GOCDB will:
    - Define the structure of the information to publish
    - Propose an implemented solution
    - Provide tools, APIs, adapters, documentation and support
- Compatibility with existing tools
  - No change required once we have moved to WS
- Hardware/software
  - Primary implementation of the model on Oracle
  - Other solutions possible, but that needs investigations





# Risks and limitations

- Multi-server architecture
  - What if one of the regional DBs is down?
  - What if some regions change the model?
- Developers not getting the idea
  - The model cannot be taken without good understanding



# Workplan and timelines

- ***By January 09***
  - Start prototyping new model
  - 2 regional use-cases definitions (NGS and Grid-Ireland)
- ***By May 09***
  - sustainable prototype implementation of the new model
  - regional use-cases working in parallel with a central DB.
  - More use-cases study
  - External adapters prototyped
- ***By October 09***
  - New model operated and in production
  - More distributed instances, depending on regions choice and/or readiness



# For more details...

- GOCDDB regionalisation
  - <http://www.grid-support.ac.uk/files/gocdb/03-GOCDDB-Regionalisation.doc>
- New architecture and model description
  - <http://www.grid-support.ac.uk/files/gocdb/04-TheModel.doc>
- “A pseudo object database model and its applications on a highly complex distributed architecture”
  - IARA/IEEE Conference on Advances in Databases (DB 2009)  
March 1-6, 2009 - Gosier, Guadeloupe/France