GOCDB decrypted plans for EGEE-III and beyond

Gilles Mathieu – STFC

Follow-up of SA1 meetings, Abingdon

December 2008

Main ideas

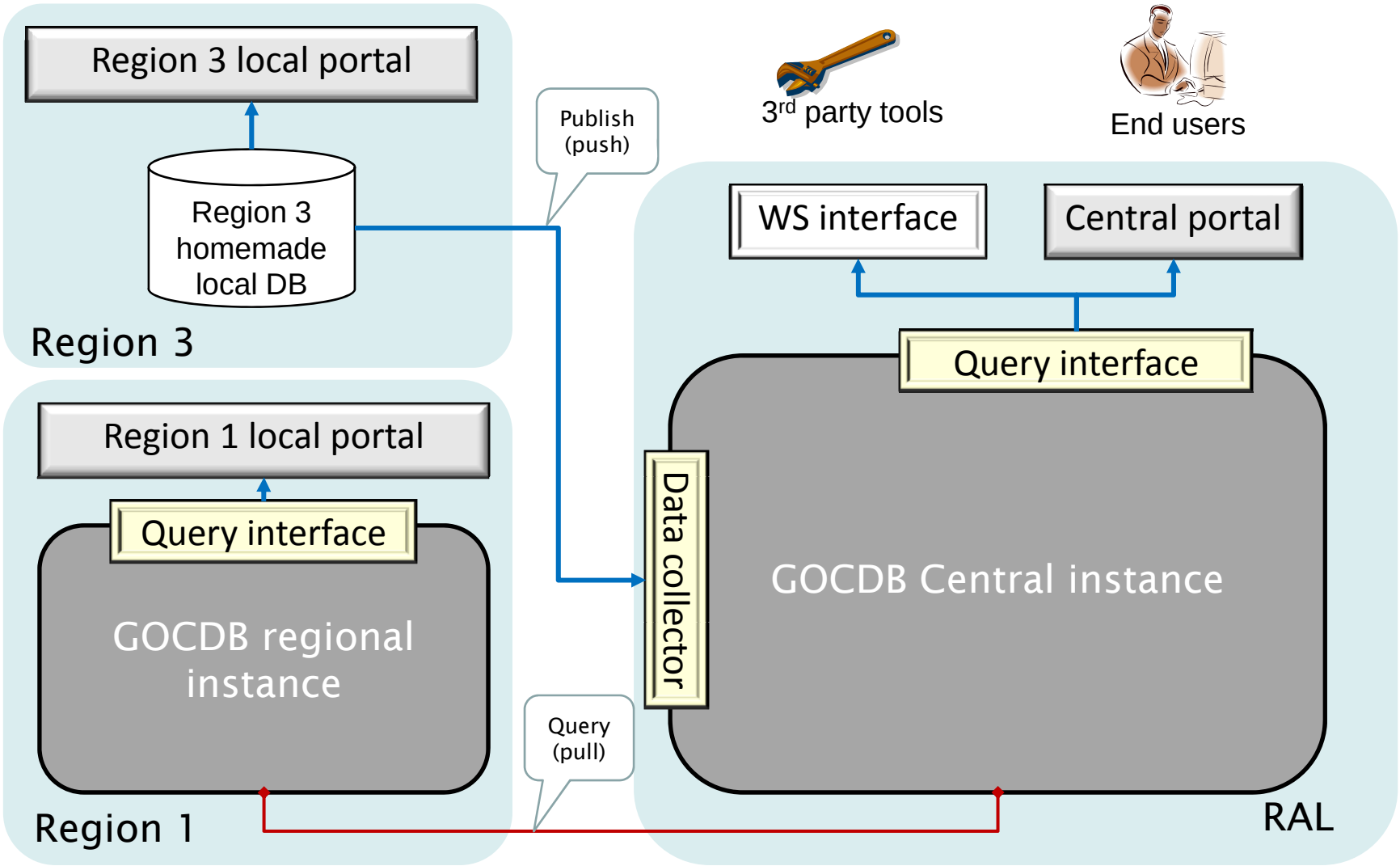
- Keep a central service, not necessarily a central DB
  - There is a need for a central access point, but:
  - the fact that regional DB are distributed or not must not be an issue
- Build a sustainable architecture that allows regionalisation but doesn't force it
  - Not all regions are at the same 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

Use cases

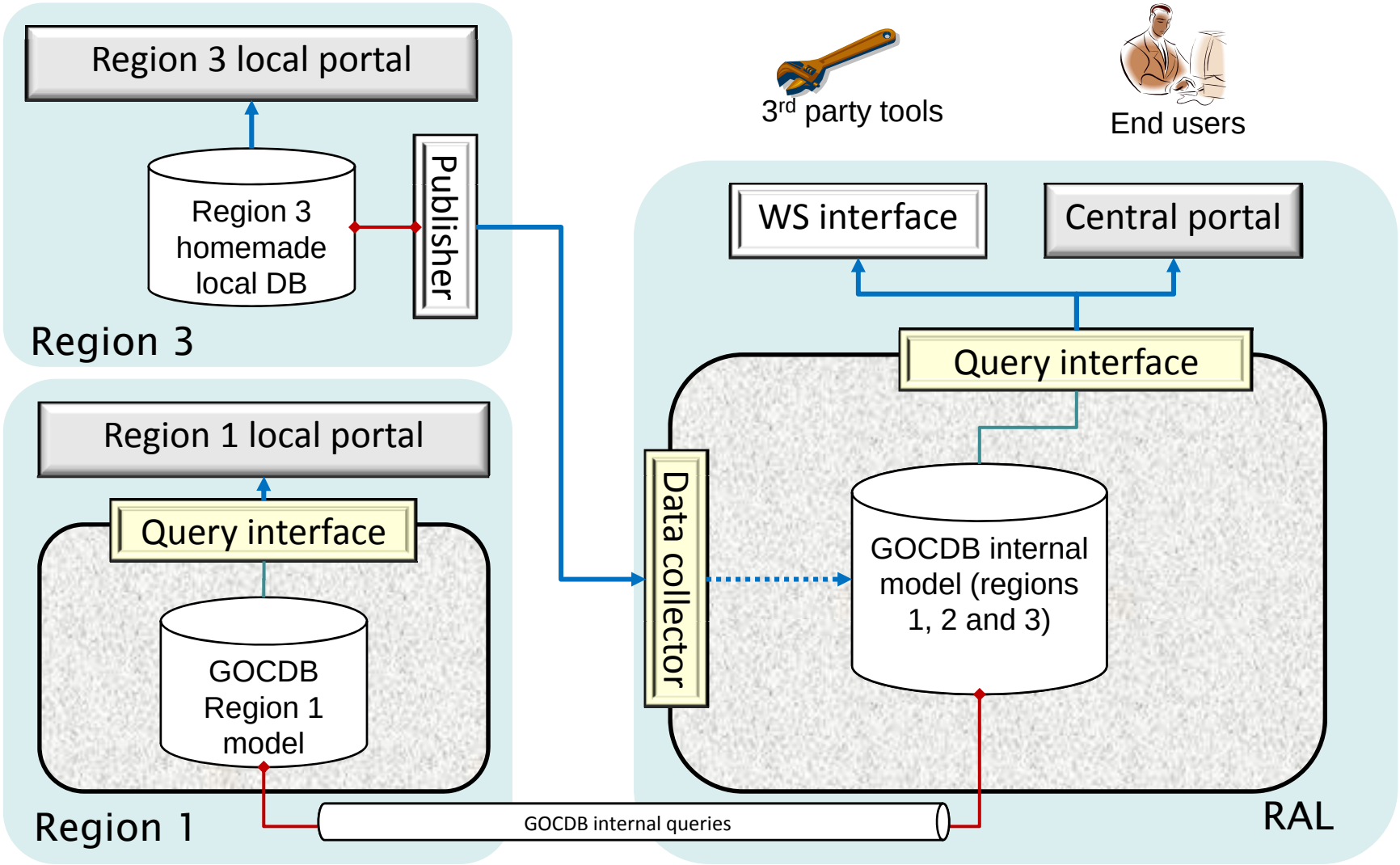
- Region 1
  - Use a distributed GOCDB instance
  - Customise it to their needs with minimal effort
- Region 2
  - Keep on using central GOCDB
- Region 3
  - Use their own model and implementation
  - Publish required data to a central system

Science & Technology Facilities Council
e-Science

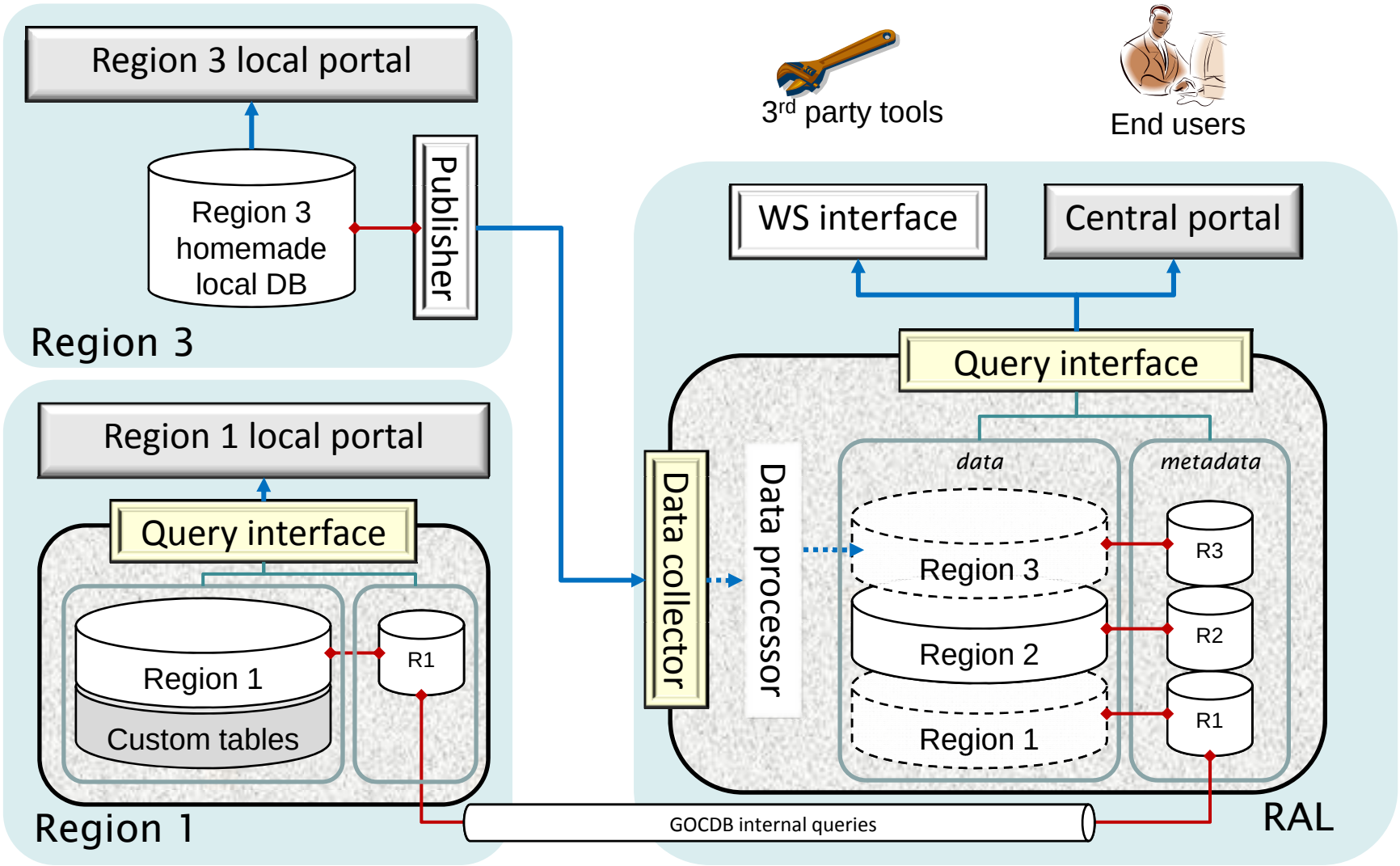
High level architecture



One level down



Details and components

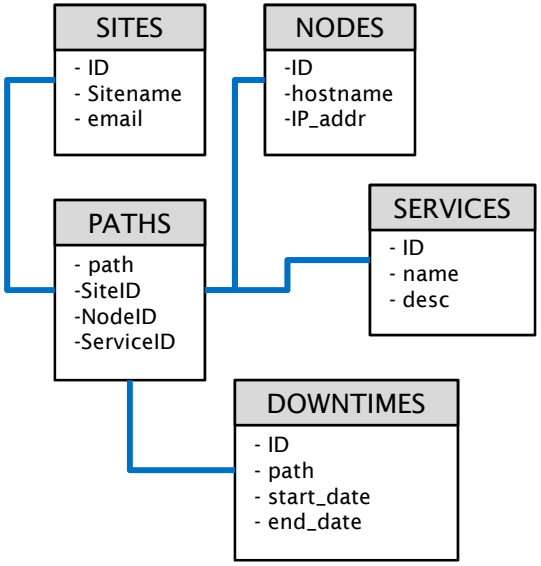


Comparing use cases

- Region 1
  - DB schema and tools to build local portal already provided +
  - Customisable with minimum effort +
  - Oracle needed -
- Region 2
  - No need to care about hosting +
  - No need to care about any development +
  - Not customisable -
  - Central schema not ideal for local use -
- Region 3
  - All latitude in the choice of a solution +
  - Possibly reuse existing infrastructure in region +
  - Need to develop publishing adapter -
  - Development and maintenance effort -

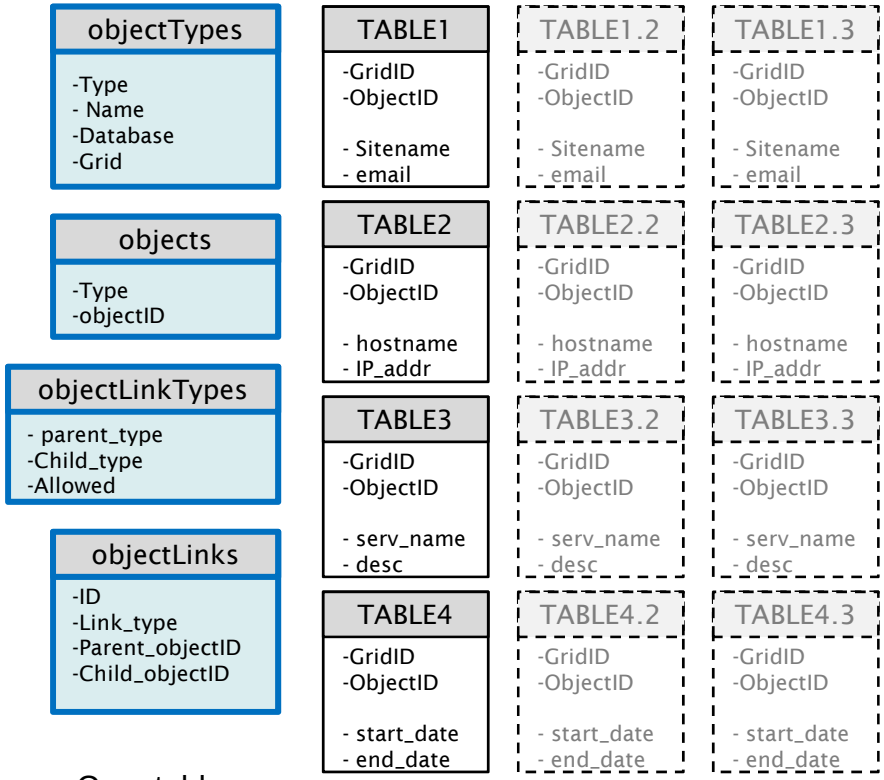
internal DB schema

Current relational model



- Physical Data Tables
- Hard coded relationships and constraints

Proposed object model



Questions & Answers

- Why changing current schema while it works fine?
  - 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
  - So, we need to change…

Questions & Answers

- Why not choosing a standard Object DB model package?
  - Because we don't want an object DB. We want a design that:
    - Benefits from not having hard-coded constraints (object-like)
    - Allows quick access and search through data (relational DB)

- Has this model been used and tested before?
  - Yes. See last slide for more details
  - Prototyping phase should allow for a better validation

Questions & Answers

- Isn't it too complicated and over-engineered?
  - Our solution is more difficult to explain than it really is
  - Trying to distribute current model **would** be complicated.
  - FYI, current relational model is visible here: http://goc.grid.sinica.edu.tw/gocwiki/GOCDB3_development?action=AttachFile&do=get&target=Schema.png

- Wouldn't a simple "central cache" solution be enough?
  - Externally, our solution will give the same service
  - Internally, it is much easier to maintain and operate
  - A simple cache does not allow for a "region1" scenario
  - Immediate development efforts will pay in the long term

Questions & Answers

- What is the benefit of having a “region 1” use case?
  - From a design point of view, region 1 allows to view distributed DBs as a single one.
  - From a technical point of view, it comes with the model and does not need extra work, so why not providing it?
  - Code and methods used to access it can be shared
  - It means less work both for GOCDB and for the region

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…

- GOCDB regionalisation
  - http://www.grid-support.ac.uk/files/gocdb/03-GOCDB-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