SynCat: Catalogue Synchronisation
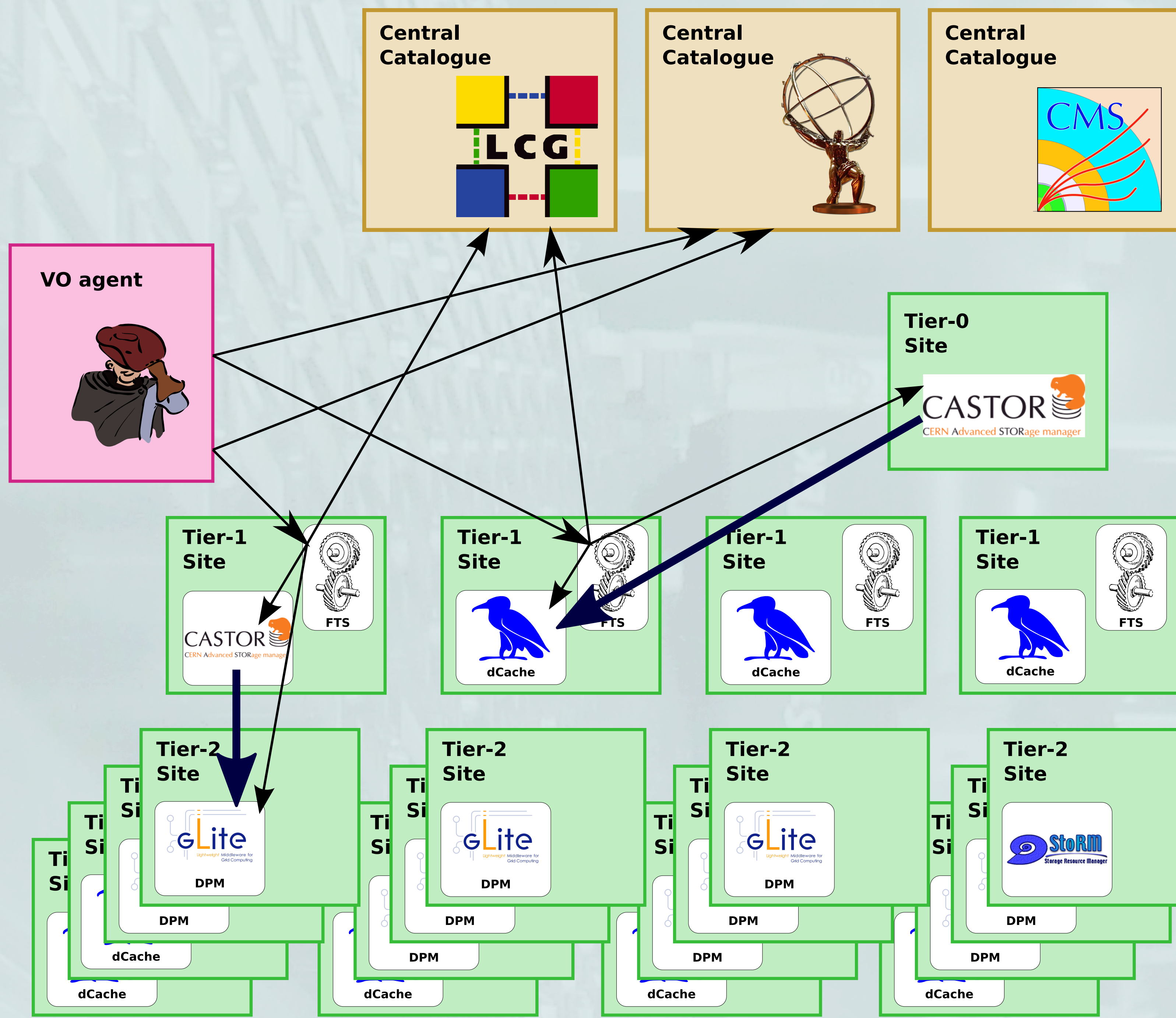
Flavia Donno (CERN), Jens Jensen (STFC), Shaun De Witt (STFC)
Paul Millar (DESY), Giuseppe Lo Presti (CERN)

WLCG Storage

No direct coupling between Storage Element namespace and the Catalogue(s)

The problem

There is (deliberately) no coupling between a Storage Element's namespace and the external catalogues. This is to avoid loops in the architecture.

All catalogue updates must go through some component the Storage Element is unaware of.

There may be many external catalogues, likely one per VO. Each catalogue will have a different external agent that maintains the catalogue.

Local site activity not propagated to catalogues; e.g., a site-admin deleting a file is not propagated, locally created files that are uploaded to the Storage Element must also be registered with VO catalogue.

File status information not propagated: which files are online right now? Relevant for pilot jobs.

Software bugs may leave some data not registered in catalogues. This may be triggered by partially successful transfers.

The only way to re-synchronise a VO catalogue against a site's SE namespace is to stat each file ("srm ls"). Doing this over the entire namespace is expensive.
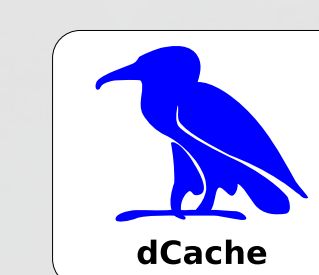
The solution: SynCat

Provide a common language to allow an external agent to follow changes to a Storage Element's namespace.

The SynCat language supports providing information as either:
- a sequence of "full-dump", or
- full-dumps followed by "incremental updates",

The site administrators may choose how often a full dump is undertaken. Dumping larger namespaces will take more resources so can be undertaken less often.

Implementation status

 PNFS available and deployed at multiple sites, Chimera is work-in-progress

 Agreed if sufficient interest from VOs
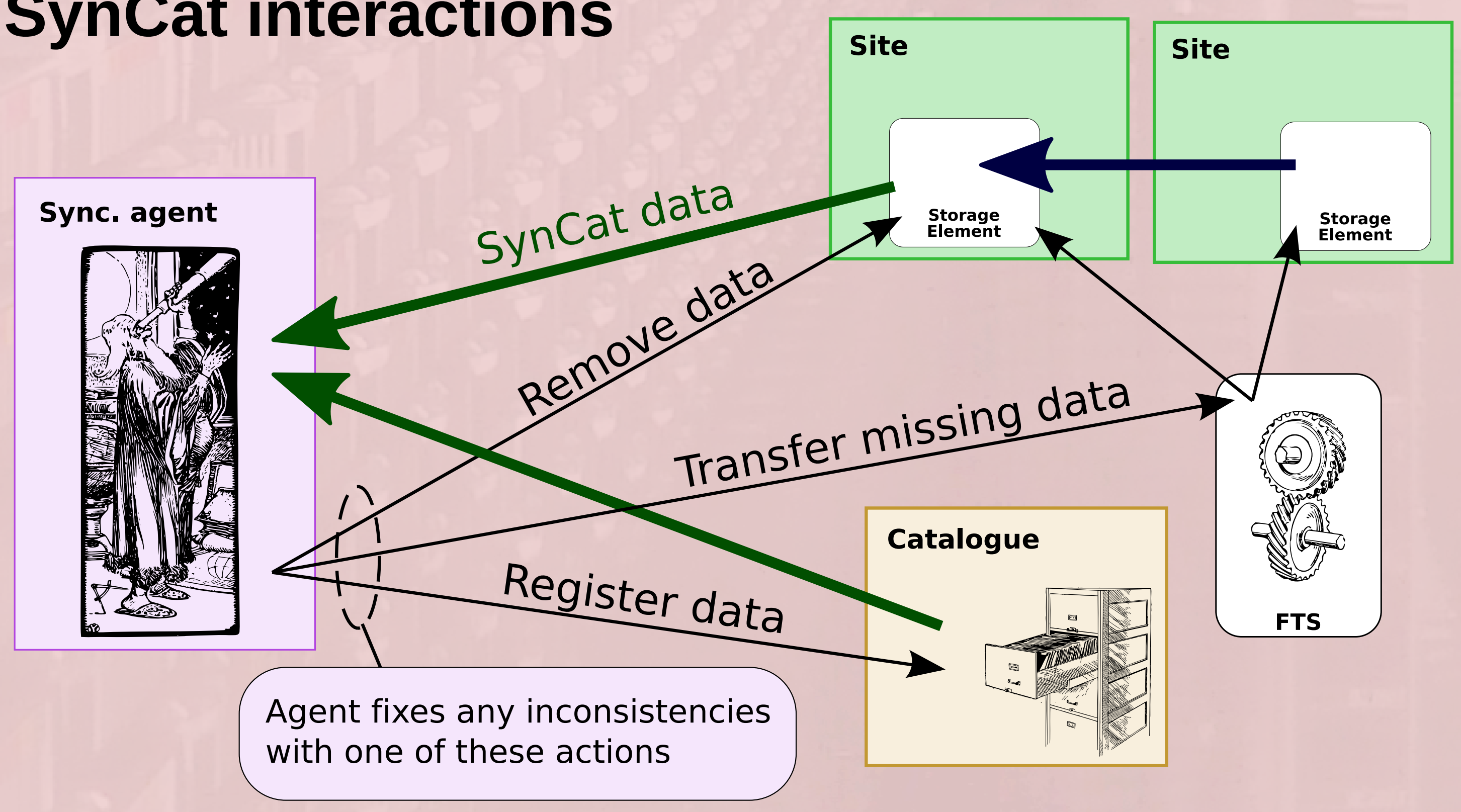
 Planning to implement (open Savannah ticket)

 Planning to implement (open Savannah ticket)

 Agreed to implement

SynCat interactions

SynCat data

Remove data

Transfer missing data

Register data

Agent fixes any inconsistencies with one of these actions

Further details

Website:
http://www.desy.de/~paul/SynCat/

Mailing list:
project-egee-syncat@cern.ch