DATA MANAGEMENT SERVICES OF NORDUGRID

Aleksandr Konstantinov
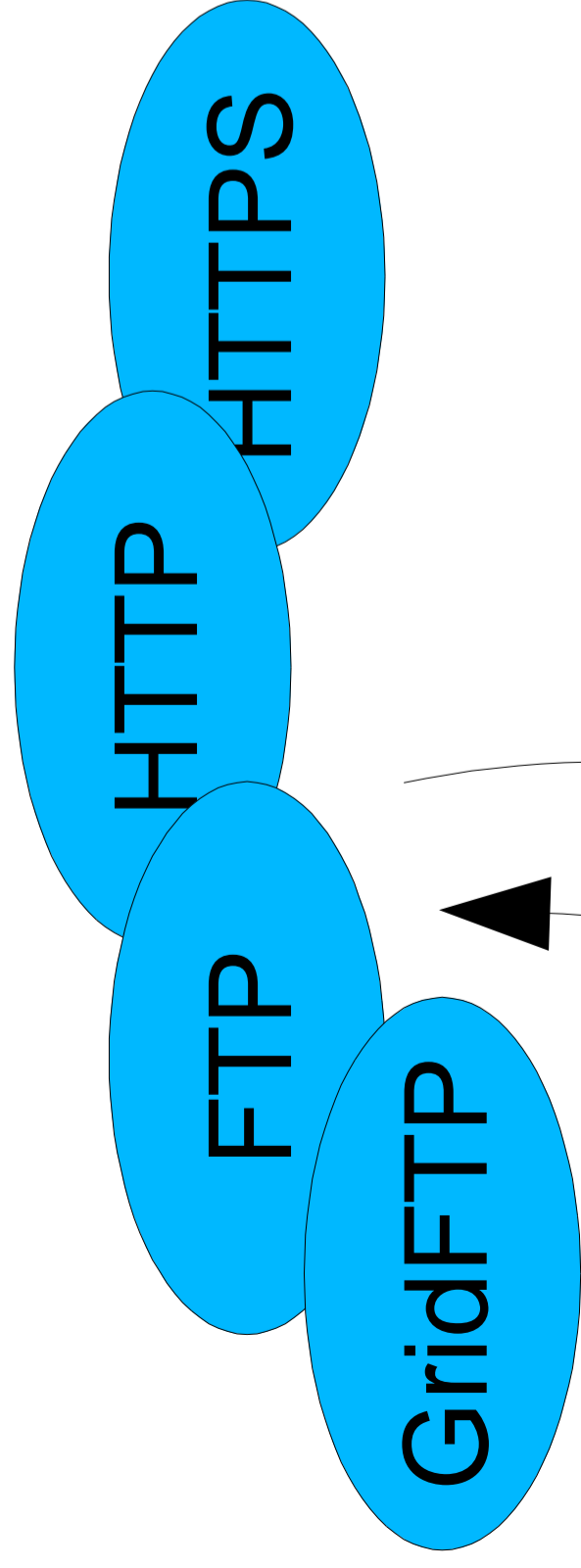
IMSAR of Vilnius University   and   University of Oslo

During existence of NorduGrid grid–like infrastructure technical team have gained some experience.

- For distributed data management system Data Indexing Services are even more important than type of Data Storage Services.

- Experience shows that services fail at much more often rate than any sane expectation can suggest. That leads to these problems:

  - Unsynchronised information in Indexing Services is inevitable.

  - Multi–GB data transfers often fail just due to failure at Indexing Service.

  - Many other problems.

- Proprietary protocols limit number of useful tools and often make it difficult to integrate them. Developers have to rely on single external implementation which can be dropped at any moment in favor of more promising (from developers' point of view) one.

- Using gained experience Data Management services and tools of NorduGrid were evolving (not as rapidly as we would like).

1st *step* – Client driven Data Management

This setup is still mostly used across sites belonging to NorduGrid infrastructure.

- Most Storage Elements (Data Services) are GridFTP compatible servers.
  - based on Globus Toolkit™ libraries
  - more flexible authorisation added
  - grid–identity based data access
- Indexing Services are just simple frontends to databases ("*Passive*" Indexing Service).
  - Whole work of managing data and meta–data is done by client applications.
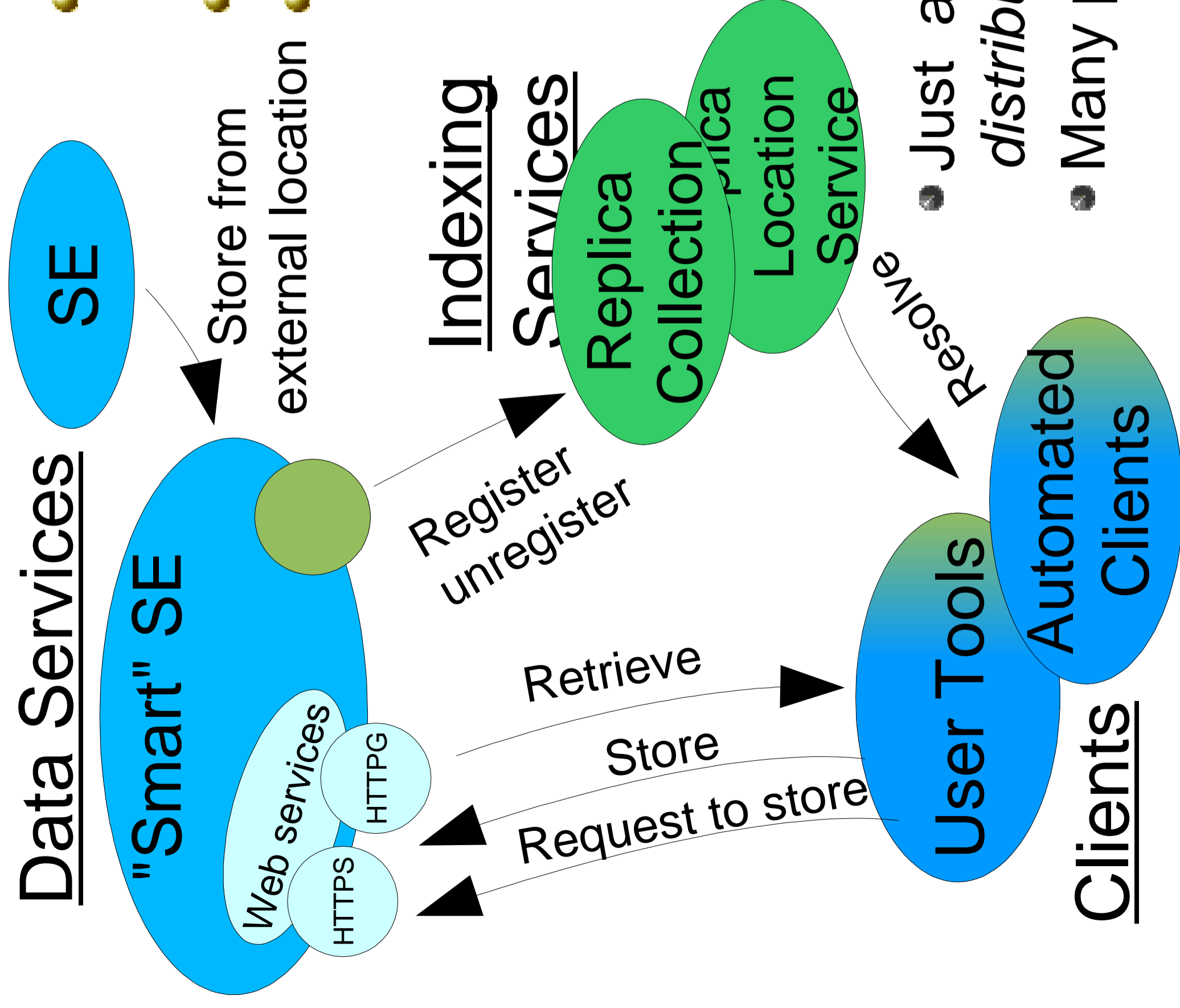 - Requires a lot of *static* configuration by system managers.

Issues

- It is not a trivial task to maintain information in Data and Indexing Services synchronized.
- System can't recover from failure itself.
- In order to make system robust no user tools can be used for important operations due to undefined availability of required services.
- More kinds of services means higher probability of failure.

Data Services

HTTPS

HTTP

FTP

GridFTP

Indexing Services

Replica Collection

Replica Location Service

Clients

User Tools

Automated Clients

Retrieve

Store

Register

unregister

Resolve

2nd step – Service driven Data Management

In order to solve some of problems new Storage element was designed:

- HTTPS/G interface allows *firewall friendly* network setup and *secure* yet *standard* way to transfer data
- Web Services control channel is extensible.
- Direct interaction with Indexing Services creates *self-sufficient* network of services.

- From client's point of view Indexing Service is used only as entry point into the system.
- Operations are reliable.
- Can acquire new data without continuous intervention/control from clients side.

- Just a first step on a way to a *reliable fully automated distributed* data management system.
- Many problems are still not solved.

This setup is being partially used during ATLAS Data Challenge 2 at sites belonging to NorduGrid infrastructure.

Data Services

SE

Store from external location

"Smart" SE

Web services

HTTPG

HTTPS

Indexing Services

Replica Collection

Replica Location Service

Register unregister

Resolve

Clients

User Tools

Automated Clients

Retrieve

Store

Request to store

3rd step – Automated Data Management

Possible roadmap of future development.

- Current implementation of "Smart" SE seems stable and mostly satisfactory. Still many things to be done:
  - Interoperability with other solutions through development of compatible interfaces like Storage Resource Manager and Reliable File Transfer.
  - Use gained experience to extend functionality.
  - Adopt better trust delegation model.
  - Make data access control consistent through full system.
  - Main task becomes "Smart" Indexing Service
  - Distributed – redundancy
  - Active – consistency, replication, monitoring of the system.
- Simplify task for user tools.

Data Services

"Smart" SE

Exchge with external location

"Smart" Indexing service as part of "Smart" SE

Indexing Services

Exchande information and tasks

Distributed "Smart' Indexing Services

Resolve

Data management requests

OPTIONAL

Retrieve

Store

User Tools

Automated Clients

Clients