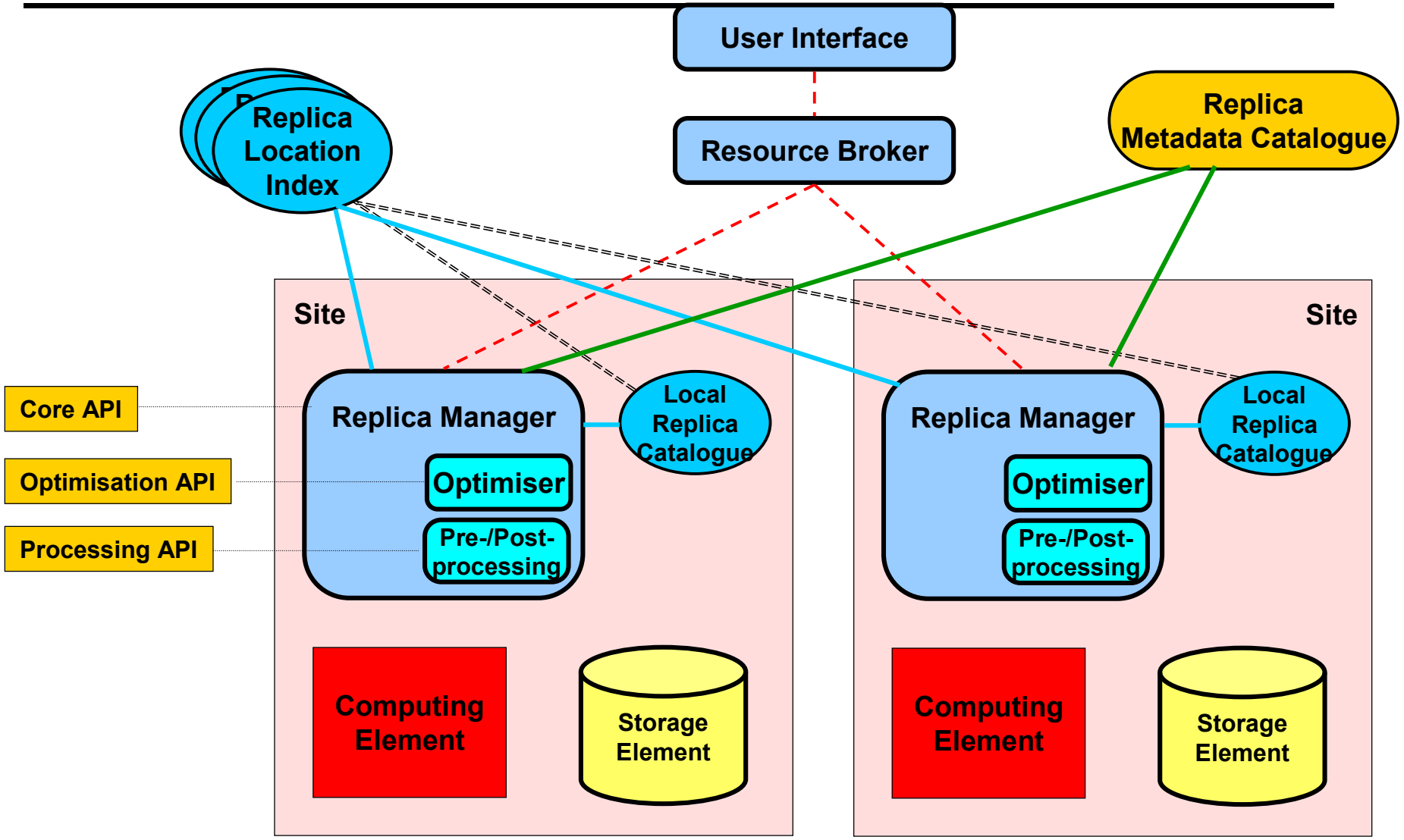
WP2
Architecture overview & dependencies

Leanne Guy

WP2 – CERN

Leanne.Guy@cern.ch http://cern.ch/leanne

Replica Manager Architecture



WP2: External dependencies

Replica Manager:

- ➡ InfoService for StorageElement attributes :**WP5**
  - i.e disk space

Replica optimiser module (Optor):

Seletion of best replica requires information from other WPs' components

- ⬅➡ Resource Broker (RB) **:WP1**
- ➡ network monitoring **:WP7**
- ➡ SE (file transfer time within SE) **:WP5**

WP2 - WP1 Interaction

WP2-WP1 Interface: Provided by WP2

Replica Optimiser <-> Resource Broker

WP2 provides 4 Replica Optimiser functions to WP1:

- getAccessCost() : Calculate cost / LFN
  - called by the RB at scheduling time
  - requires input from WP7: Estmation of transfer time
- initFilePrefetch()
  - initiate transfer of PFNs to local store of CE
- cancelFilePrefetch()
  - in case of job crashes
- getBestFile()

Implementation details at: http://edms.cern.ch/document/337977/

Example usage WP1 - WP2

- Time t0: RB calls getAccessCost() to select the “best” CE with respect to physical file locations. (Therefore call the function for each candidate CE.)
- t1: The best CE has been found, schedule job.
- t1 <= t2 <= t3: Prior to job start on CE, RB can call initFilePrefetch()
- t3: Job starts
- t4: Job access file LFN1 using getBestFile(LFN1,...)
- t5: Job access files LFN2 and LFN3 using getBestFile(LFN[2,3],...)

WP2 - WP7 Interaction

WP2-WP7 Interface: Provided by WP7

getNetworkCosts ()

- Provides an estimate of the network cost between 2 IP addresses.
  - Needed by the QO service to determine best replicas
- Called internally by the RO module
- Requires network cost functions with respect to throughput, round trip time and packet (as already defined by WP1)

Agreed in the WP1-WP3-WP7 meeting in Paris that WP7 will provide the C++ API !

- still much to discuss in terms of details etc

WP2 - WP5 Interaction

WP2-WP5 interface: Provided by WP5

getIOBandwidth()

- Internally required by the QO service, this function should provide a measurement of the bandwidth available between a SE and the outside world. Needed by getAccessCost()

Open questions and issues:

- How closely coupled should the RM and the SE be ?
- Consistency between the SE and the RM
  - entries need to be syncronised, designation of responsibility
- Management of private file on Grid SEs
  - job output not registered with RM, local clean up policy
- Replica manager owns all master and secondary copies
  - Delete file via RM only, original permissions preserved in RC

What is missing ? (1)

Overall integration

- No overall use cases
- No overall view of integration of components & dependencies

Definitions

- Definition of a CE and worker nodes
- How are WN connected within the CEs
- Logical specifiction of CE and hardware.
- Definition of an SE,
  - Diskserver, MSS, tape, who will take responsibility for SE definition
- Filenames: LFN, PFN, TFN :
  - Syntax to be agreed within the EDG : all WPs

What is missing ? (2)

Site definition

- Is there an CE and SE per site ?
- Handling of local cache ?

Implementation time lines of WP5 for reservation API

- Written commitment desired

Extension of JDL

- Hints for access patterns needed for replica optimisation

Uniform package names, interfacesGUIs to be defined

- Currently, all WPs have "their own standards". "ATF-old" has started with some naming conventions.

Integration with PPDG, other Grid projects