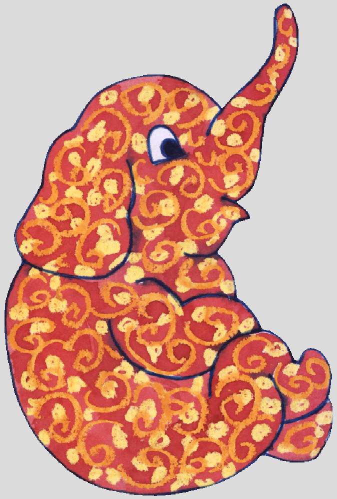
Storage Element Overview

DataGrid Project Conference

Heidelberg, 26 Sep-01 Oct 2003



Jens G Jensen

RAL, EDG WP5

Releases

- Release for Testbed 2.0
  - Web service in secure or insecure mode (or both!)
  - Access control being integrated
  - Disk, CASTOR, HPSS, Atlas Datastore
  - GridFTP, NFS, RFIO
- Release for Testbed 2.1 – 2.0 plus:
  - SRM version 1.1 interface
  - “Proper” queuing system
  - Compiled with gcc 3.2.2

Collaborations

- DataGrid Storage Element
  - Integrate with WP2 Data Replication Services (Reptor)
  - Jobs running on worker nodes in a ComputingElement cluster may read or write files to an SE
- SRM – Storage Resource Manager
  - Collaboration between Lawrence Berkeley, FermiLab, Jefferson Lab, CERN, Rutherford Appleton Lab

Deployment

- Working Storage Elements:

  CERN – Castor and disk

  UAB Barcelona – Castor

  RAL – Atlas DataStore and Disk

  ESA/ESRIN – disk

  CC-IN2P3 – HPSS
- Testing:

  ESA/ESRIN plan tape MSS (AMS)

  NIKHEF disk
- Others:

  SARA building from source on SGI

  INSA/WP10 – to build support for DICOM servers

Installation

- RPMs and source available
- Source compiles with gcc 2.95.x and 3.2.2
- Configures using LCFG-ng
- Tools available to build and install SEs without LCFG:

```
./configure --prefix=/opt/edg
make
make install
./edg-se-configure-all –mss-type=disk
```

Limitations

- Disk Cache
  - Not much protection
  - On disk-only SEs, files are copied into disk cache
  - No proper disk cache management yet (“pinning”)
- Users can only be members in one VO
  - Once VOMS is supported this limitation will go away
  - VOs not properly supported for insecure (anonymous) access
- ACLs fixed – can only be modified by SE admin

Development Priorities

1. EDG TrustManager adopted for web services authentication – **done**
2. Proper queuing system – **done**
3. Delete, exists (not part of SRM) – **done**
4. SRM v.1.1 interface – **being integrated**
5. Access control – use GACL – **being integrated**
   - **No setACL, getACL yet**
   - May be replaced with a different ACL system?
6. Improved disk cache management (including pinning) – **work started**
   - Will also improve file metadata
7. VOMS support – **not yet done**

For TB 2.1

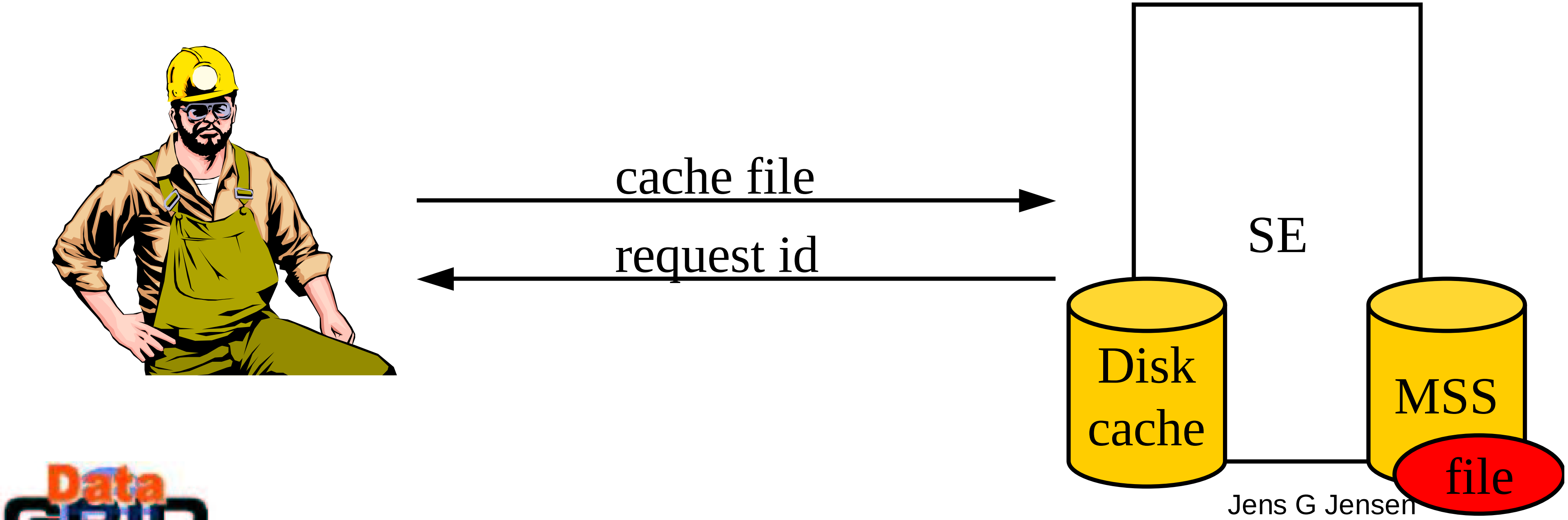
- More secure mode SEs to be deployed
- Access control integrated
- General improvements – error reporting in particular needs improvement !
- Bugfixes…
- Requests asynchronous (SRM for now)
  - Current interface will be synchronous for now
  - Can make current interface asynchronous as well

SRM in TB2.1

- SRM version 1.1
  - Storage Element enables common access to disk, CASTOR, HPSS, ADS, …
  - Slightly different interface but same principle
  - Defined as a web service
  - `srmGet` is the same as `cache`, `srmPut` the same as `create` – each command returns TURLs
  - Commands allow operations on several files
  - Allows users to use any SRM (version 1) client

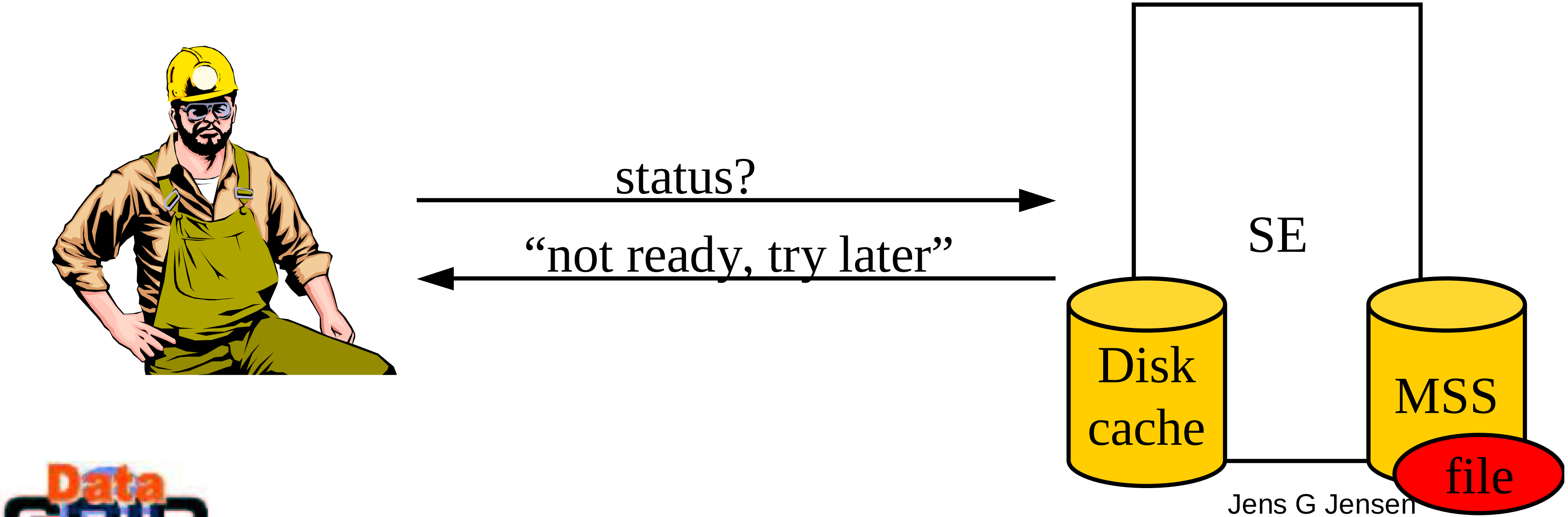
More about SRM

- Users know Site File Names (SFN) or Physical File Names (PFN)
- `lxshare0408.cern.ch/bongo/mumble`



More about SRM

- Client queries the status of a request
- Better that client polls than server callbacks
- Server (ideally) able to give time estimate



status?

“not ready, try later”

SE

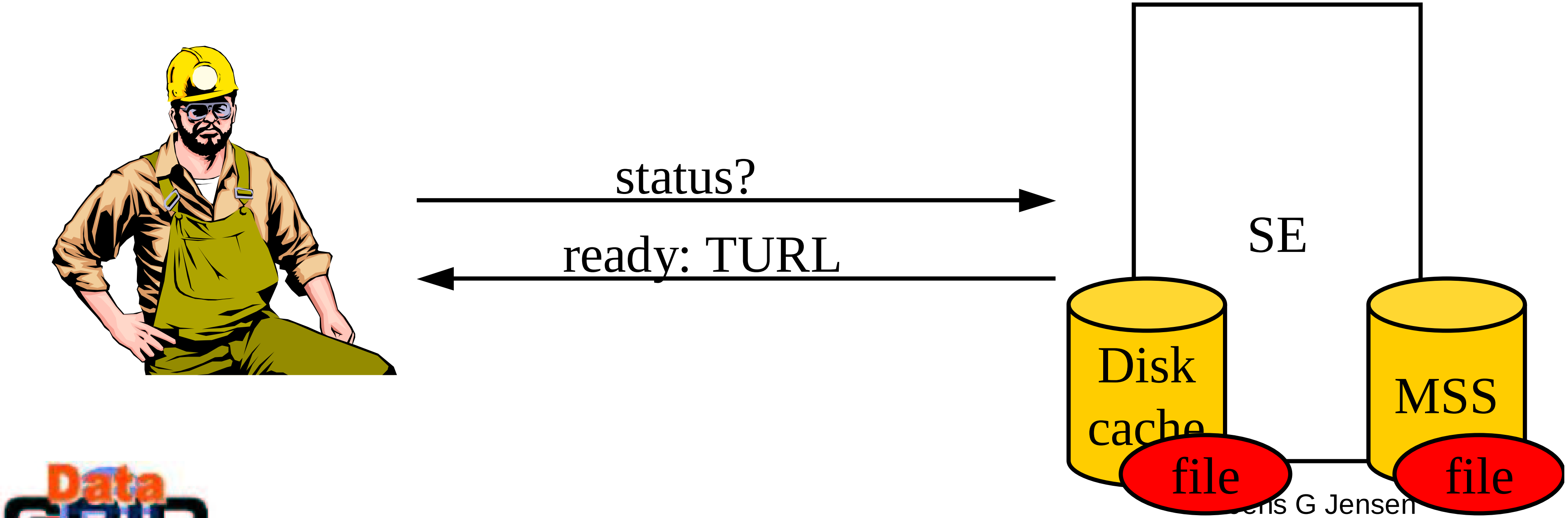
Disk cache

MSS

file

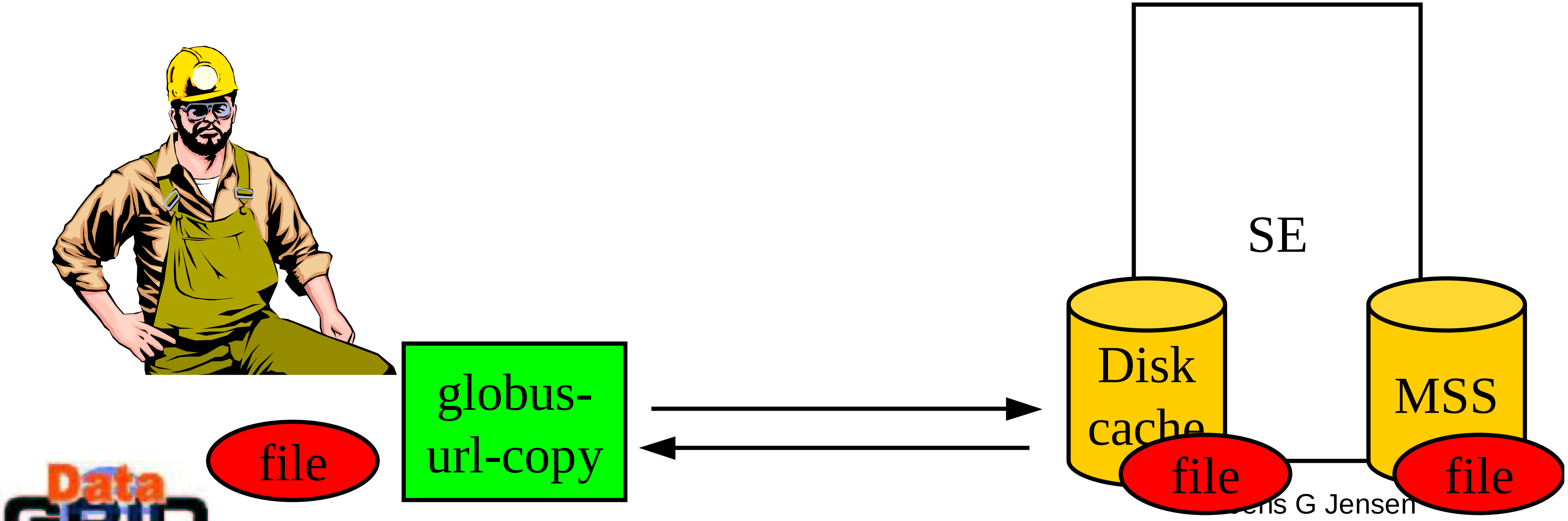
More about SRM

- When request is ready, client gets a Transfer URL (TURL)
- `gsiftp://lxshare0408.cern.ch/flatfiles/01/data/16bd30e2a899b7321baf00146acbe953`



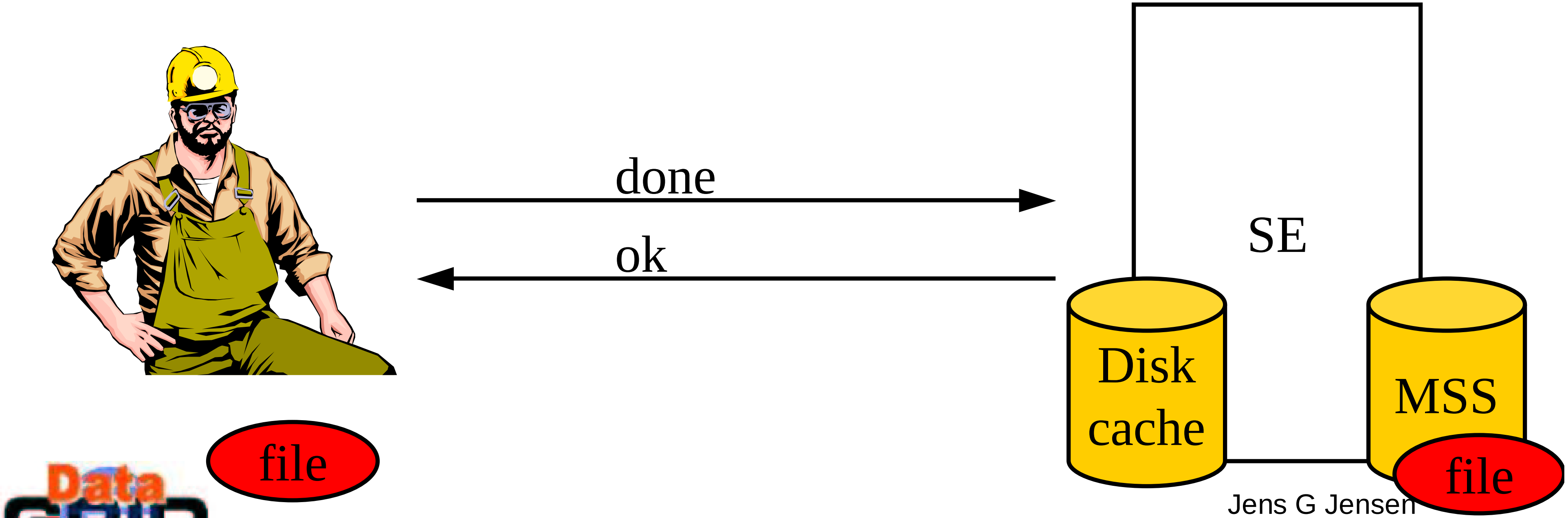
More about SRM

- Client accesses the file in the SE’s disk cache using (usually) non-SE tools



More about SRM

- Finally, client informs SE that data transfer is done
- This is required for cache management etc



SRM command line

- Work started on generic SRM command line tools
  - Initially for testing – one SRM command per command line tool
  - Later “usable” tools doing Get-getStatus-Transfer-Done cycle acting on several files

Data transfer interface

- GridFTP
  - The standard data transfer protocol in SRM collaboration
- Some SEs will be NFS mounted
  - Caching and pinning still required *before* the file is accessed via NFS
- Easy to add new data transfer protocols
  - E.g. http, ftp, https,…

Information Interface

- Can publish into MDS
- Can publish (via GIN) R-GMA
  - Using GLUE schema for StorageElement
  - `http://hepunx.rl.ac.uk/edg/wp3/documentation/doc/schemas/Glue-SE.html`
- Also a file metadata function as part of the control interface

Future Directions – post EDG

- Guaranteed reservations
  - SRM2 recommendations: volatile, durable, permanent *files* and *space*
- Full SRM version 2.1
- Scalability
  - Scalability will be achieved by making a single SE distributed
  - Not hard to do
- Resurrect ele* commands