UNIVERSITY
OF OSLO

ARC Storage Solution, NGIn School

Jon K. Nilsen, Dept. Of Physics, Univ. Of Oslo

UNIVERSITY
OF OSLO

Outline

- KnowARC storage solution
  - Current ARC Data Management – pros and cons
  - Requirements for KnowARC solution
  - Storage solution by KnowARC
- NGIn School
  - Program
  - Experiences
  - Summary

Grid Data Management Services

- What services do we need?
  - Storage Manager
  - Storage Elements
  - Indexing services
  - File transfer services
- Internal and external communication through web services

Current ARC – pros and cons

- Good
  - Easy to install and configure
  - Multi-OS, multi back-end, non-intrusive
  - Stable, high performance services
  - Good error handling
- Bad
  - Non Service Oriented Architecture
  - No native indexing service
  - No consistent data management solution
  - Limited data management user interface
  - No native file transfer service

Required capabilities
(from KnowARC design document)

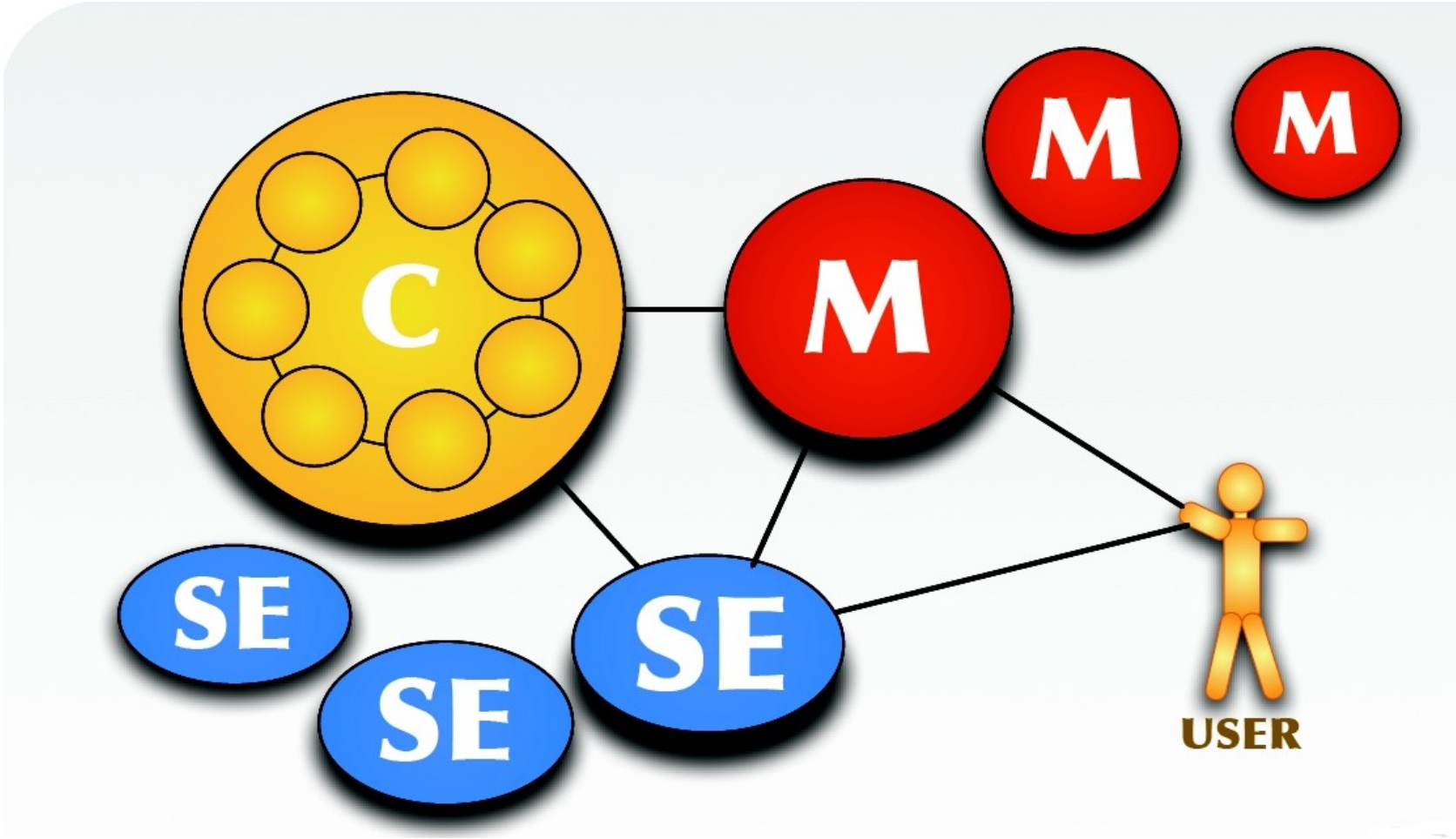
- Scalable, reliable indexing service capable of operating on collections
- Indexing service should not appear as a single point of failure
- Metadata should be extendable to have application-specific information
- Must have standard interfaces and API
- Must be aware of VO-specific storage solutions, and use them for storage
  - E.g., WLCG has SRM as standard solution, so ARC needs to support SRM

Required Capabilities (cont'd)

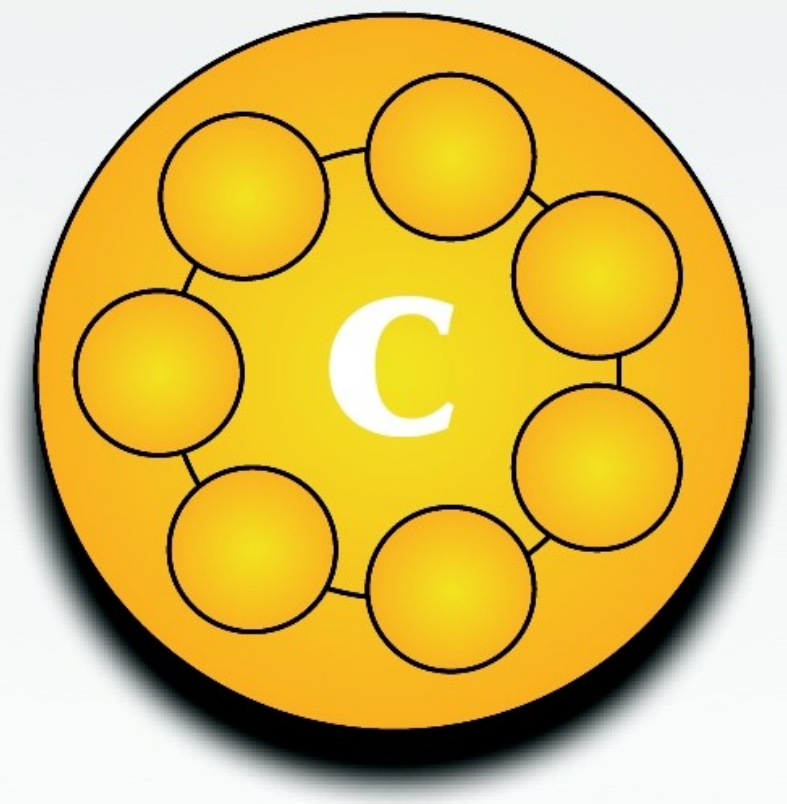
- Must handle bulk data manipulation
- Must be aware of physical file replicas and use this information to optimize file transfer
- Must support secondary storage facilities and flexible data staging
- User should see single access point to data (ref. Google)
- Should provide simple (POSIX-like) access to data for applications
- Easy to install
- Easy to use

UNIVERSITY OF OSLO

KnowARC storage design

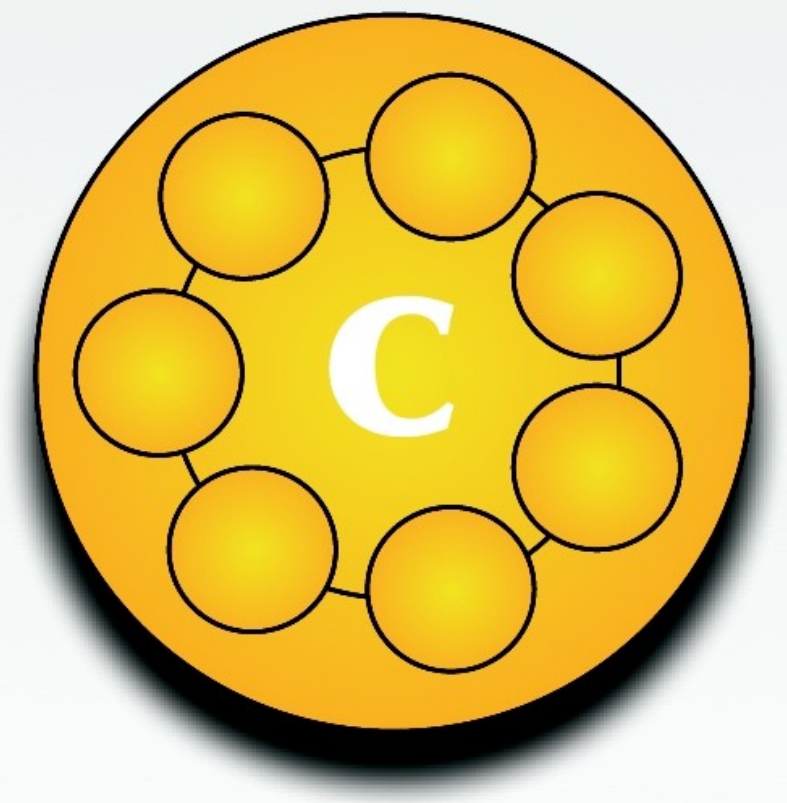


Catalog



- Maintains all metadata
  - GUID
  - Logical -> physical filename
  - Is it replicated
  - Is it marked for deletion
  - Access control list
- Metadata stored in a Distributed Hash Table

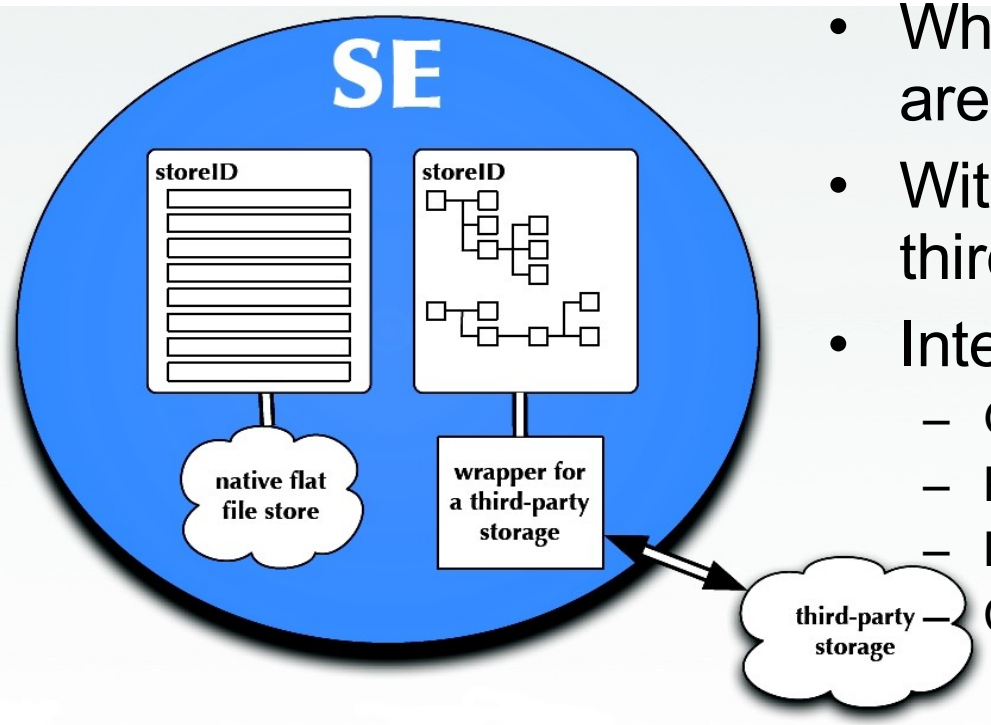
Catalog



- Interface
  - New file
  - New collection
  - Get metadata
  - Remove
  - Change metadata
  - Traverse logical name

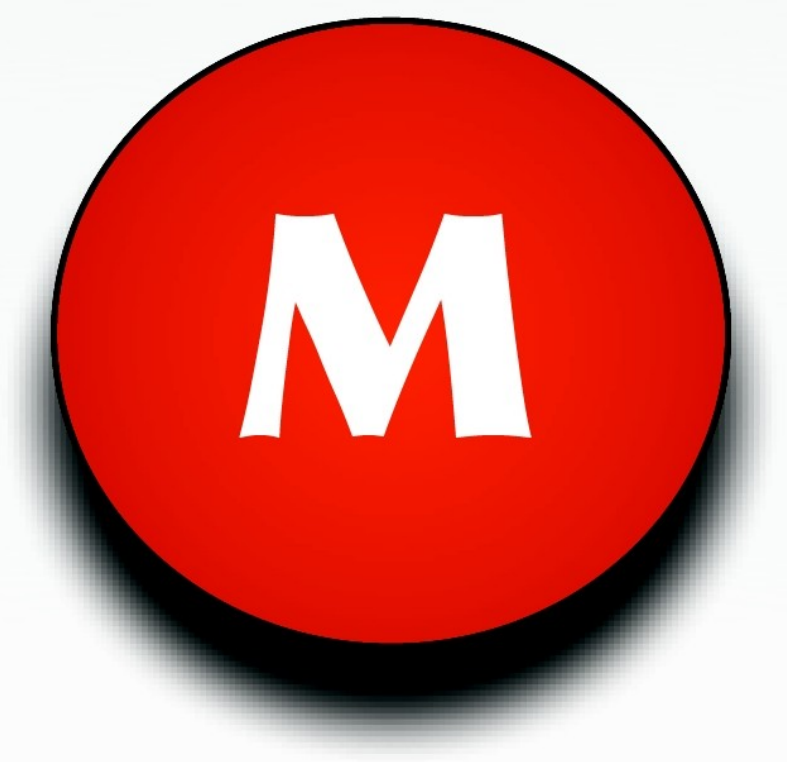
UNIVERSITY OF OSLO

Storage Elements



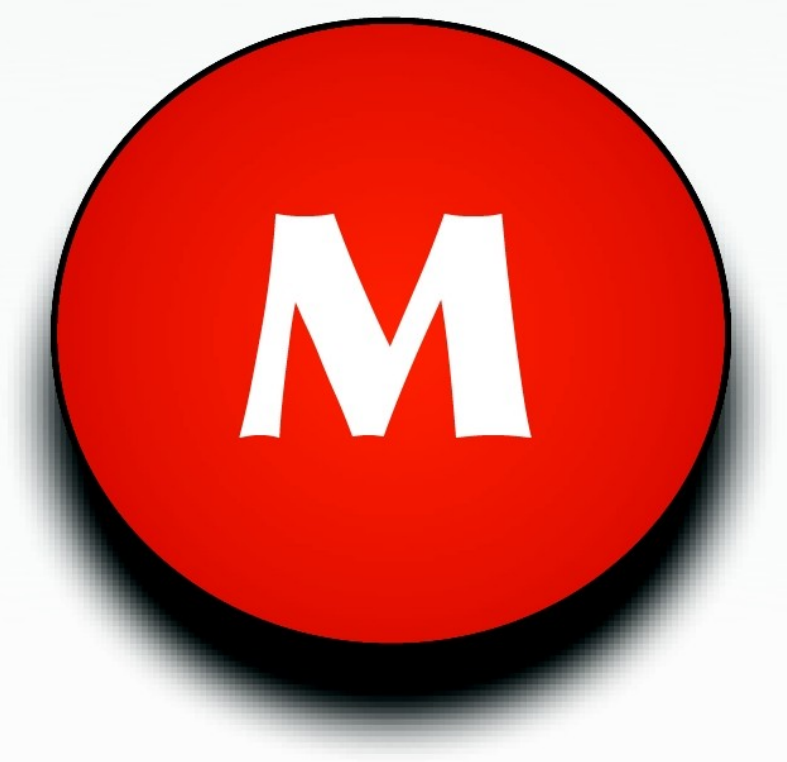
- Where the physical files are stored
- With possibility to mount third-party storage
- Interface
  - Get file
  - Put file
  - Delete file
  - Copy file to a URL

Storage Manager



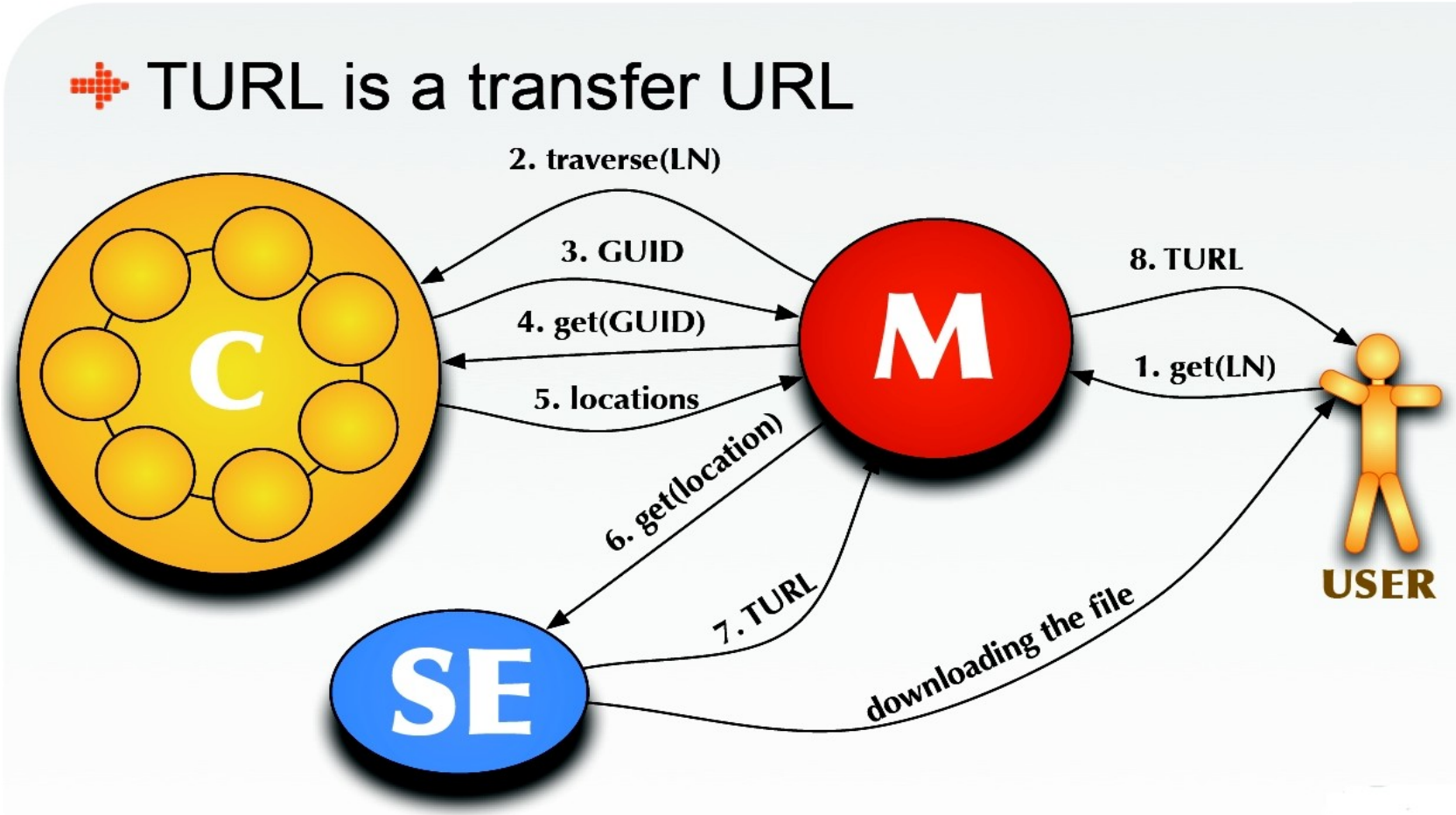
- Provides high-level interface to user
- Communicates with catalog to get metadata and create new files and collections
- Communicates with Storage Elements to initiate transfers

Storage Manager



- Interface
  - Put file
  - Get file
  - Delete file
  - Make new collection
  - List collection
  - Move file or collection
  - Modify metadata
  - Stat (get metadata)

Download


TURL is a transfer URL
2. traverse(LN)
3. GUID
4. get(GUID)
5. locations
6. get(location)
7. TURL
8. TURL
1. get(LN)
downloading the file
C
M
SE
USER

LFN, PFN, GUID, Source URL, Transfer URL

- Logical file name: `some/filename`
- Physical file name: `a.nice.site:4242/tmp/some/filename`
- Global Unique Identifier: `63047220-2dd5-11d9-9669-0800200c9a66`
- Source URL: `srm://a.nice.site:4242/tmp/some/filename`
- Transfer URL: `gridftp://a.nice.site:4242/tmp/some/filename` or `gridftp://yet.another.site:4242/MyDocuments/tuid`

UNIVERSITY
OF OSLO

ARC Storage Discussions

- Builds on an already working solution
  - Grid Underground Storage System
  - Deployed in Hungary – with one catalog, one storage element and one manager (but no problem to expand it)
- The distributed hash catalog is the key element
  - One promising algorithm (Etna) is being considered
  - No suitable implementations of it yet
- What is the intended use of this system?
  - One ring (of catalogs) to rule them all?
  - Or yet another dCache solution?

UNIVERSITY OF OSLO

NGIn School

UNIVERSITY
OF OSLO

NGIn School

- NGin: Innovative tools and services for NorduGrid
- Project funded by the NorduNet3 program
- Goal is twofold
  - Extend existing Grid middleware
  - Train new Grid experts, securing future technology development
- Training program includes Grid PhD positions and a Nordic Grid School, the NGIn School
- First Grid School at NorduGrid07

UNIVERSITY OF OSLO

NGIn School program

- Day 1 (General Grid Introduction)
  - Intro to Grid
  - Intro to ARC
  - First steps with ARC (tutorials)
- Day 2 (Specialized tutorials)
  - HEP distributed analysis
  - ARC in bioinformatics, ARC in medical imaging
  - ARC-gLite interoperability
  - Grid Job Manager
  - Dynamic Runtime Environments
- Day 3 (ARC Development)
  - Sys-admin and developer training

Experiences

- Day 1
  - ~20 participants
  - Interesting introductions to Grid and ARC middleware
  - Very general topics
- Day 2
  - ~8 active participants
  - Quite specific topics -> more fluctuating participants
  - $dN_{participants}/dt < 0$
- Day 3
  - Merged into NorduGrid parallel session
  - Training consisted of Ferenc writing an ARC service on the beamer

UNIVERSITY
OF OSLO

NGIn School Summary

- A very nice school for newcomers
- From newbie to specialized user in two days
- Maybe not so useful for NGIn PhDs
- Could have been more emphasis on developer training
- Who was the audience of the school?
  - Recruiting new Grid users?
  - Training NGIn PhDs?
  - Both?