Towards the next generation ARC middleware

*Early results of the **KnowARC project***

Balázs Kónya

KnowARC Project Leader,Lund University

EGEE07, Budapest, 4 October 2007

KnowARC factsheet

KnowARC

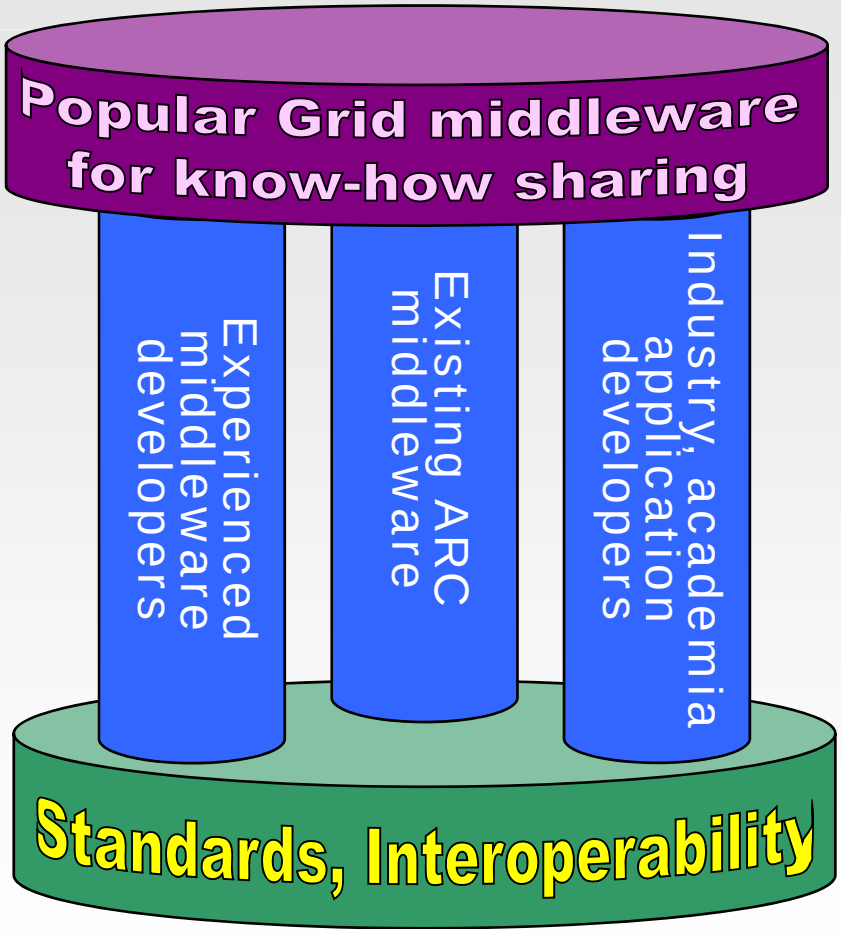
Title of Contract: **Grid-enabled Know-how Sharing Technology Based on ARC Services and Open Standards**

Acronym: KnowARC
Contract Nber: 032691
Instrument: STP - Specific Targeted Research Projects (aka **STREP**)
Thematic priority/domain: Information Society Technologies (IST)
Call title: IST Call 5
Call identifier: FP6-2005-IST-5
Activity: IST-2005-2.5.4 - Advanced Grid Technologies, Systems and Services
Program: **FP6**
Duration: **39 month (June 1, 2006 – August 31, 2009)**
Start date: **June 1, 2006**
Community Contribution: EUR 2 899 494:-
Resources: **17.5 FTEs in total, 12.5 financed**
Coordinator: University of Oslo
Technical Coordinator: **Lund University**
Contact: **Prof. F. Ould-Saada, www.knowarc.eu**

First project period ended on 31 August, 2007

KnowARC Objectives

- develop a **next generation** grid solution based on the **ARC software**
  - moving ARC to **SOA**
  - achieving industrial quality
  - **standards** conformance, **interoperability**, reliable solution
  - not to undermine **existing functionality** of current ARC
  - achieving significant awareness and use of ARC, exploitation in **new areas**



Advanced Resource Connector (ARC)

KnowARC

- **General purpose** Open Source European Grid middleware
  - One of the major production grid middlewares
  - Developed & maintained by theNorduGrid Collaboration
  - Deployment support, extensive documentation, available on most of the popular Linux distributions
- **Lightweight architecture** for a dynamic heterogeneous system following **Scandinavian design** principles
  - start with something simple that works for users and add functionality gradually
  - non-intrusive on the server side
  - Flexible & powerful on the client side
- **User- & performance-driven development**
  - Production quality software since May 2002
  - First middleware ever to contribute to HEP data challenge
- **Middleware of choice** by many national grid infrastructures due to its technical merits
  - SweGrid, SWISS Grid(s), Finnish M-Grid, NDGF, etc…
  - Majority of ARC users are NOT from the HEP community

Illustrations:"Scandinavian Design beyond the Myth"

www.scandesign.org

ARC generations: ”*ARC Classic*”

KnowARC

- Know as **ARC0**, **production ARC** or current ARC
- First production release May 2002
  - No major technology change since then
- **Custom interfaces**, Pre-WS technology:
  - Gridftp-based custom protocol for job management
  - LDAP-based infosys with nordugrid schema
- **Heavy pre-ws globus-(library) dependency**
- Very **powerfull custom made grid solution**
  - E.g.: Grid Manager on the frontend, ARC gridftp server
- ”Classic ARC” will stay **supported** by the NorduGrid Collaboration through various projects
  - KnowARC invested large efforts into the new production ARC Release in its first project period (see Tuesday Talk by Anders)
  - next generation ARC will offer client-side backward compatibility via the new ARCLIB
- ARC Classic technology **white paper**:
  - "Advanced Resource Connector middleware for lightweight computational Grids". Future Generation Computer Systems 23 (2007) 219-240
- Software available: download.nordugrid.org

The dark side of *ARC Classic* (1/2)

KnowARC

- ARC exposed to critical (Swedish) **users:**
  - Why can not I use the grid from Windows?
  - What do I gain running *ngsub* instead of *qsub*?
  - Clumsy input/output data management
  - Why do I have to write my own tool to manage my 1000 jobs on the grid?
  - Too complicated: I have better things to do than master certificates, XRSL, GACL syntax, new set of commands
  - ARC does not work with other grids
  - The grid is only good for big (Physics) projects
  - The grid is slow and unreliable

The dark side of *ARC Classic* (2/2)

KnowARC

- ARC exposed to critical **sysadmins**:
  - Reluctance to open up my firewall for the gridports
  - Clumsy and poorly documented configuration
  - Security concerns due to grid mapping, root services
  - Managing Software Environments is a pain
  - ARC does not work together with other Grids
- ARC exposed to **advanced users** and **developers**:
  - Non-standard based interfaces
  - Monolithic software
  - Middleware-dependent client-side development library
  - Poor support for parallel applications

ARC generations: motivation for a generation change

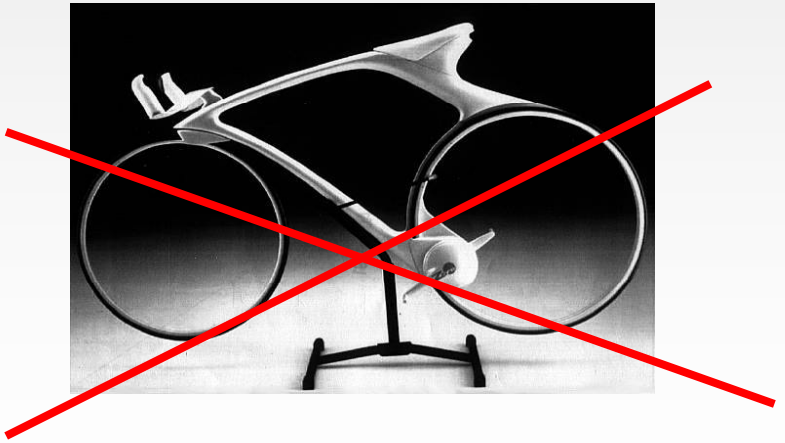
KnowARC

- **Re-engineer interfaces** but **keep functionality**
  - Offer well-undertsood interfaces for 3rd party systems
  - Enable easier interoperability with other standards-compliant middlewares
  - Comply with community embraced emerging standards
- **Improve modularity**, make it easier to develop new services for ARC: follow SOA
- **Improve portability**: reduce globus dependency

ARC generations: *”ARC1”*

KnowARC

- Also known as ”new ARC” or ”**next generation ARC**”
- Under heavy development: first **technology-preview** early prototype has been recently released
- ARC1 tries to **follow mainstream** but **avoid extremes**
- ARC1’s high-level design complies with OGSA (GFD.80)
- Standard-compliant **WS-based interfaces** will support OGF BES, JSDL, Glue2
  - With necessary Nordugrid extensions
  - replace ldap infosys, gridftp-based job management
- Will offer a new **framework for service development**
- Will come **with a new security framework**
- Will come with new **ARCLIB**
- Will come with **new high-level functionality**: advanced brokering, logging, software deployment, data management



ARC1: What is available? (1/3)

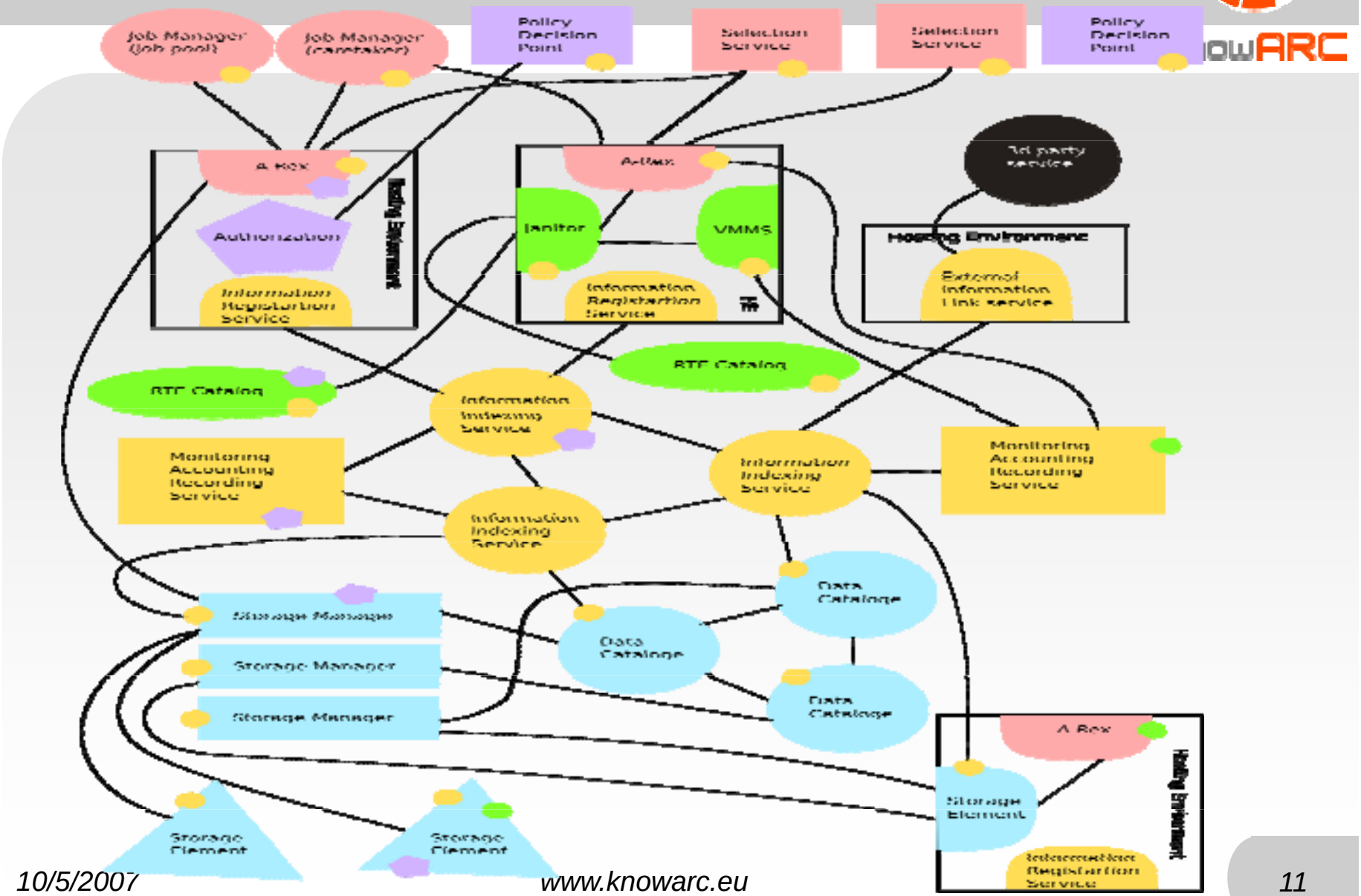
KnowARC

- ARC1 high-level **design process is completed**
  1. **Design document**
     - State of the art survey
     - Requirements
     - Shortcommings of ARC Classic
     - **OGSA compliant** high level service decomposition
  2. **Interoperability** minimal service **survey**
  3. KnowARC **standards comformance roadmap**
     - ❖ Survey of the standardization landscape
  4. **Definition of interfaces** (WSDLs) for the core components
  5. KnowARC **security review**

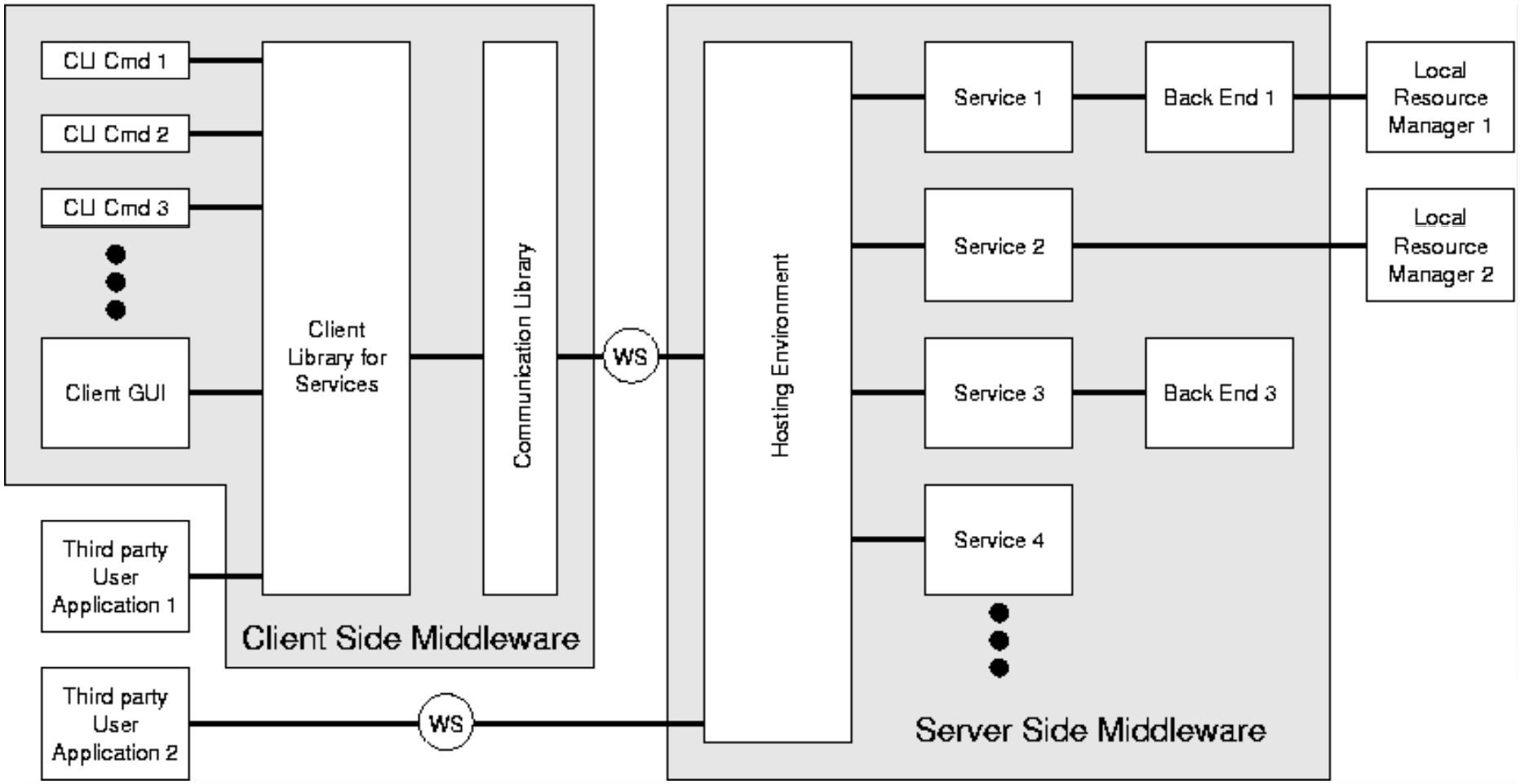
References:

1. http://www.knowarc.eu/documents/Knowarc_D1.1-1_07.pdf
2. http://www.knowarc.eu/documents/Knowarc_D3.1-1_07.pdf
3. http://www.knowarc.eu/documents/Knowarc_D3.3-1_06.pdf
4. http://www.knowarc.eu/documents/Knowarc_D1.2-1_07.pdf
5. http://www.knowarc.eu/documents/Knowarc_D1.6-1_07.pdf

ARC1 Design: service decomposion



ARC1 design: internals



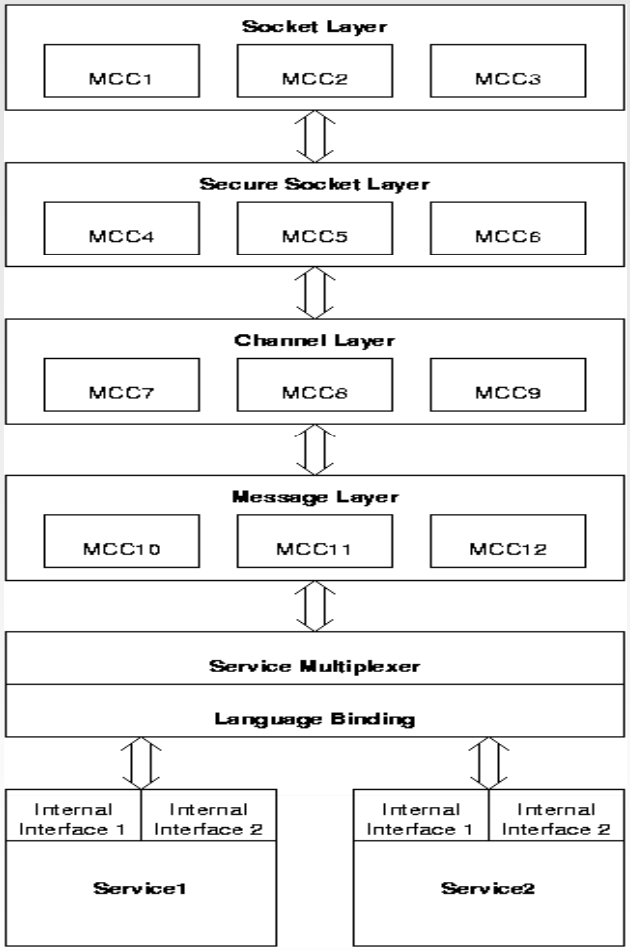
ARC1: What is available? (2/3)

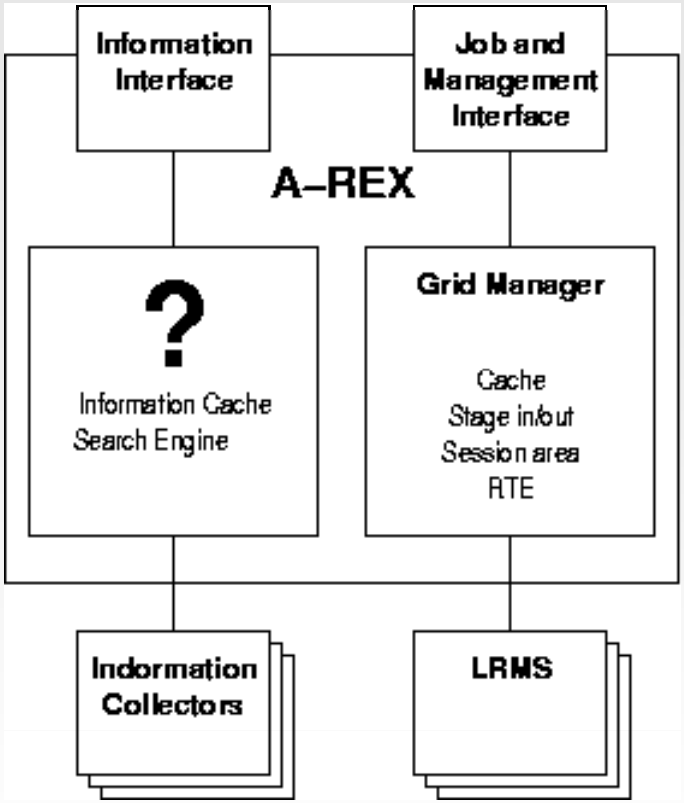
KnowARC

Technology preview code for

- **Hosting Environment Daemon** (HED)
  - ARC container
  - Networking & basic security layer
  - Service development framework
- **A-REX** service
  - ARC Resource Coupled Execution Service
  - BES/JSDL compliant job execution service
  - Extensions were necessary to both BES and JSDL
- **Proof-of-concept clients**
  - Minimal A-REX client to demonstrate WS technology
  - Minimal **ARC →Glite gateway client**
  - These clients are not based on ARC1LIB yet

ARC1: HED and A-REX internals





ARC1: What is available (3/3)?

KnowARC

Reusable developments based on *ARC Classic*:

- Dynamic Runtime Environment extension demonstrator
  - Janitor modul: Dynamic application installation modul
  - RDF-based Catalog: extension to RTE description
- Enhanced batch system support
  - Hardened batch system interface scripts
  - Additional support for LSF, Loadleveler
  - Incorparated into ARC 0.6.1 release
  - More info: ARC Back-end Interface To The Additional Batch Systems, http://www.knowarc.eu/documents/Knowarc_D1.4-1_07.pdf

ARC1: What will be available? (1/2)

KnowARC

- Storage system
  - Detailed component design is completed, coding starts soon
  - Design document is available from svn.nordugrid.org
- Information system
  - Design draft for Information Index Service is available, to be discussed on Wednesday technical session
  - Local Information system subcomponent of the HED is being discussed
- HED & Services Security framework
  - draft is ready
  - ARCPDP implementation has been started
- ARC1LIB
  - Early proposal for generalization of current ARCLIB is available
  - Pluggable structure to enable support for 3rd party middlewares

ARC1: What will be available? (2/2)

KnowARC

- HED Backends
  - Backend refers to HED components shared by services
  - Example: local information engine
- Application software integration:
  - Early-stage ongoing work in automotive industry and bioinformatics
  - flowGuide: workflow management solution by s+c
  - Taverna
  - Medical image analysis on the Grid with MedGift
    - Currently based on Classic ARC and grid job manager by Antti Hyvärinen

Approach to standardization

KnowARC

- **Interoperability efforts**
  - ARC <-> Glite
- Standard **conformance roadmap** revealed
- **Engaged** with those **standardization efforts** where major middleware and infrastructure providers are present
  - GLUE, OGSA-BES, JSDL
- **Avoid** emerging "**de facto standards**" defined **by implementations**
- **Focus on Interfaces, Interfaces, Interfaces**

Contribution to *ARC Classic*

KnowARC

- ARC 0.6.1 release with
  - New external software packages
  - Improved batch system support (Torque, LSF, SGE, LSF, Loadleveler, Condor,fork)
  - Thorough testing
  - Numerous bug fixes
- Support & maintenance
- Resource consumption & Performance profile (work being done by Kosice)

Operation

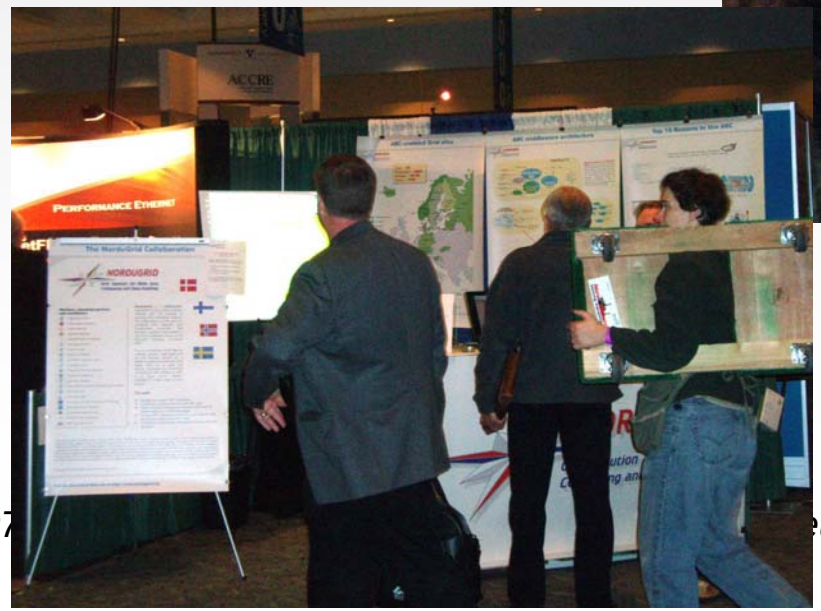
KnowARC is **not** an infrastructure project, nevertheless it operates some NorduGrid community services

- Build system for the ARC software stack
  - Operated by Copenhagen
  - Automated Grid-based build system for "middleware nightlies" including a monitoring component“
    - http://www.knowarc.eu/documents/Knowarc_D5.1-1_07.pdf
- Nordugrid code repository (svn), download area operated by Copenhagen
- NorduGrid Bugzilla operated by Copenhagen manned by collaboration members
- Helpdesk manned by Kosice, Oslo,Lund and other members
- KnowARC Pilot Grid System
  - Small testbed for pre-releases of ARC Classic and ARC1
  - Plans to deploy an Instant CA service

Outreach, events

KnowARC

- Desing Week in Hungary, September 2006
- NorduGrid Technical meetings
- NorduGrid Conferences
- Exhibition booth at
  - SuperComputing events
  - CHEP07
- KnowARC/Nordugrid sessions at EGEE events
- OPEN to everyone!



People behind KnowARC

KnowARC

- KnowARC consortium
  - NorduGrid Collaboration partners
    - Oslo, Copenhagen, Lund, Uppsala
  - Middleware expert
    - NIIF
  - Application specalists
    - Geneva, Lübeck, Kosice,s+c
- Technical group includes original core ARC developers

Further information

- Technical documents (in ”EU deliverable” format):
  - www.knowarc.eu/documents
- Code:
  - svn.nordugrid.org -> arc1 directory
  - technology-preview packages as public deliverables from www.knowarc.eu/documents
- Technical threads on **nordugrid-discuss** mailing list