



<http://www.grid-support.ac.uk>



<http://www.ngs.ac.uk>

OGSA-DAI





Policy for re-use



- This presentation can be re-used for academic purposes.
- However if you do so then please let training-support@nesc.ac.uk know. We need to gather statistics of re-use: no. of events, number of people trained. Thank you!!



Acknowledgments



- Matt Ford, NGS Induction Workshop (Dec. 2004, NeSC)
- Neil Chue Hong , OGSA-DAI Tutorial GGF13
- OGSA-DAI website, www.ogsadai.org



Data services on NGS



Simple data files

- Middleware supporting
 - **Replica files**
 - **Logical filenames**
 - **Catalogue**: maps logical name to physical storage device/file
 - **Virtual filesystems**, POSIX-like I/O
- **Storage Resource Broker**

Structured data

- RDBMS, XML databases
- Require extendable middleware tools to support
 - Move computation near to data
 - easy access, controlled by AA
 - integration and federation
- **OGSA -DAI**



What is OGSA-DAI?



- The Open Grid Services Architecture Data Access and Integration project is concerned with constructing middleware to assist with access and integration of data from separate data sources via the grid.
- The project was conceived by the UK Database Task Force and is working closely with the Global Grid Forum DAIS-WG and the Globus team.



OGSA-DAI Design Principles – I



- Efficient client-server communication
 - Minimise where possible
 - One request specifies multiple operations
- No unnecessary data movement
 - Move computation to the data
 - Utilise third-party delivery
 - Apply transforms (e.g., compression)
- Build on existing standards
 - Fill-in gaps where necessary

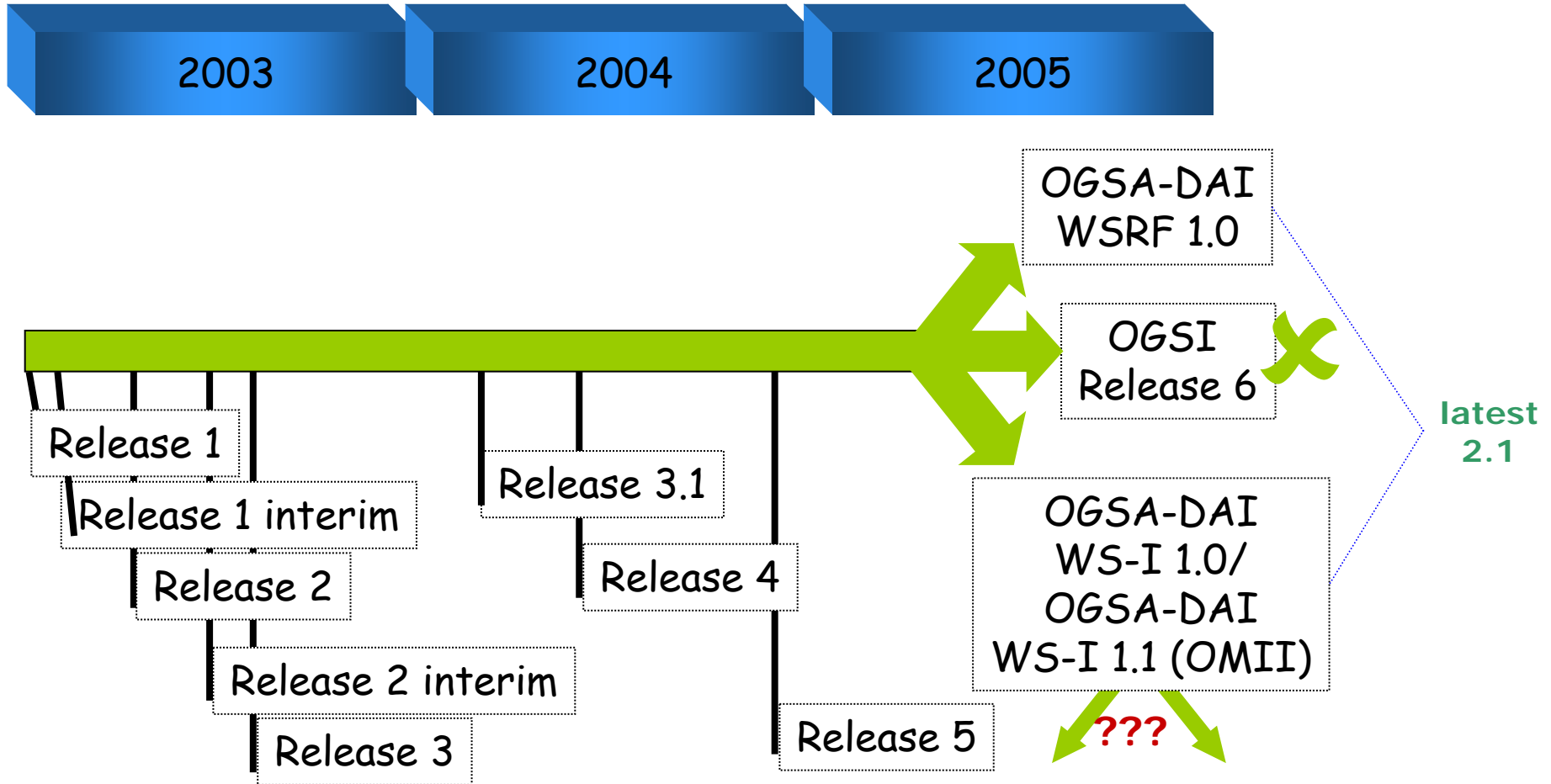


OGSA-DAI Design Principles – II



- Do not hide underlying data model
 - Users must know where to target queries
 - Data virtualisation is hard
- Extensible architecture
 - Modular and customisable
 - e.g., to accommodate stronger security
- Extensible activity framework
 - Cannot anticipate all desired functionality
 - Activity = unit of functionality
 - Allow users to plug-in their own

OGSA-DAI Timeline





OGSA-DAI Motivation



- OGSA-DAI is motivated by the need to:
 - Provide an extensible framework for easily integrating data resources on to Grids.
 - Provide for data discovery from previously unknown locations.
 - Allow different types of data models from distributed data resources to be easily integrated to Grid applications.
 - Allow data to be accessed through uniform interfaces.
 - Facilitate the integration of data from various sources to obtain the required information.
 - ...



OGSA-DAI Provides



- Access to and updating of data resources
- Exposure of Data Resources to the Grid
- Additional data manipulation functionality at the service level
- Uniform access to disparate, heterogeneous data resources
 - Does not hide underlying data model
- Data resources exposed through services
 - Clients interact with these services



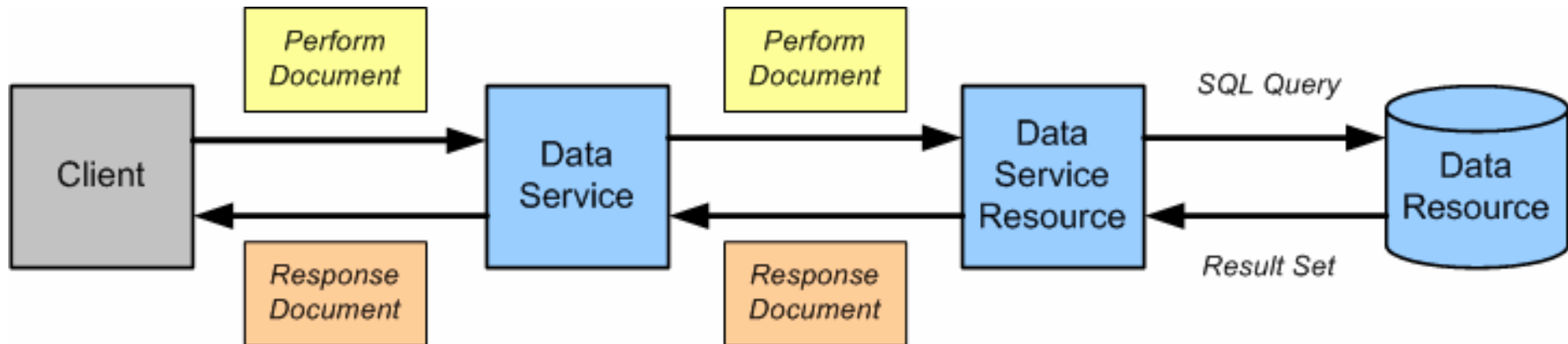
Interacting with Data Resources



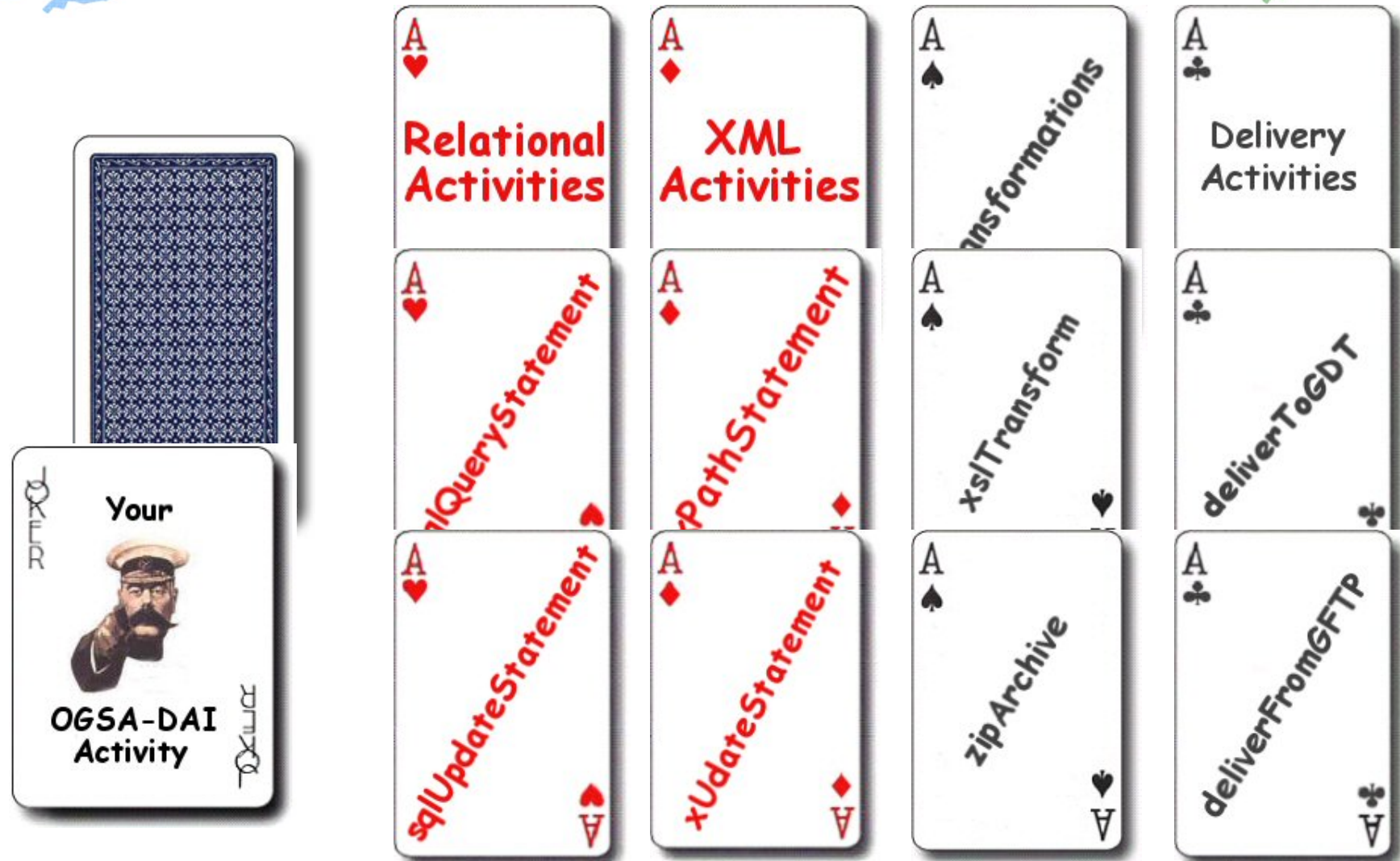
- **Activity:** The data resource manipulation, data transformation and delivery actions that the client wants the service to perform.
 - Think of sending the job to the data not the data to the job.
- **Perform documents:** Used by clients to specify to the services the activities they want executed.
 - Usually don't see/construct explicitly
- **Response documents:** Used by the services to inform clients as to the status of execution of their Perform documents and, often, to also return data to a client.



Interacting with Data Service resources



GOSC OGSA-DAI Deck of Activities





OGSA-DAI and the NGS



- the OGSA-DAI deployment on the NGS is being actively developed
- users should expect that procedures may change – it does not reflect the commitment NGS has to providing a service.
- Initially the Manchester JISC data cluster has been charged with deploying the OGSA-DAI service



Why OGSA-DAI?



- Can embed additional functionality at the service end
 - Transformations, compressions, third party delivery
 - The extensible activity framework
- Avoiding unnecessary data movement
- Common interface to heterogeneous data resources
 - Relational, XML databases, and files
- Language independence at the client end
 - Do not need to use Java
- Platform independence
 - Do not have to worry about connection technology, drivers, etc



OGSA-DAI Practical



- <http://agenda.cern.ch/fullAgenda.php?ida=a061881>