



omii europe
open middleware infrastructure institute

Grids, Grid Data Services and OGSA-DAI

Mike Mineter
NeSC-TOE
mjm@nesc.ac.uk



Acknowledgement

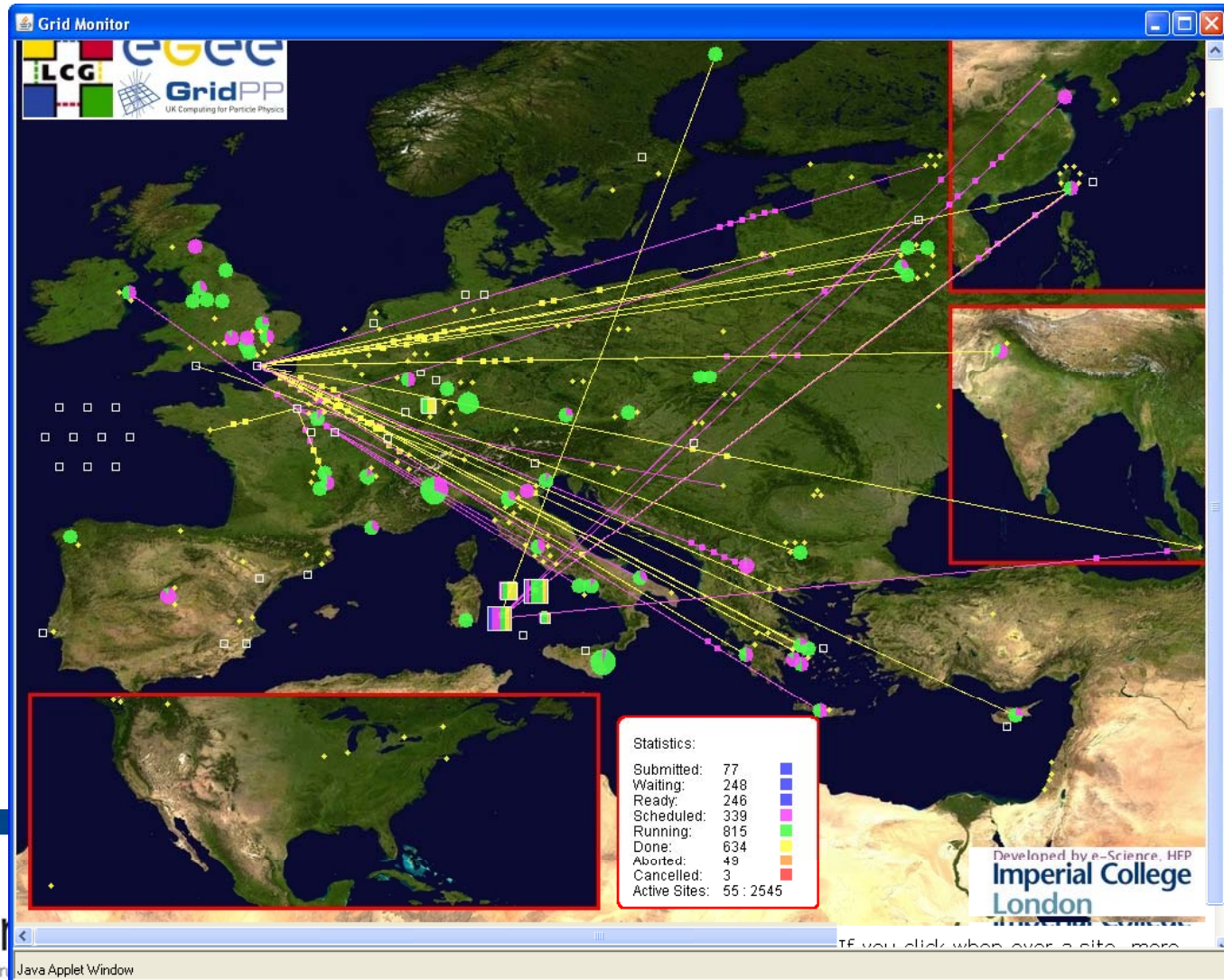
- **Many slides from OGSA-DAI team.**
- **(Some slides from me.)**

Contents

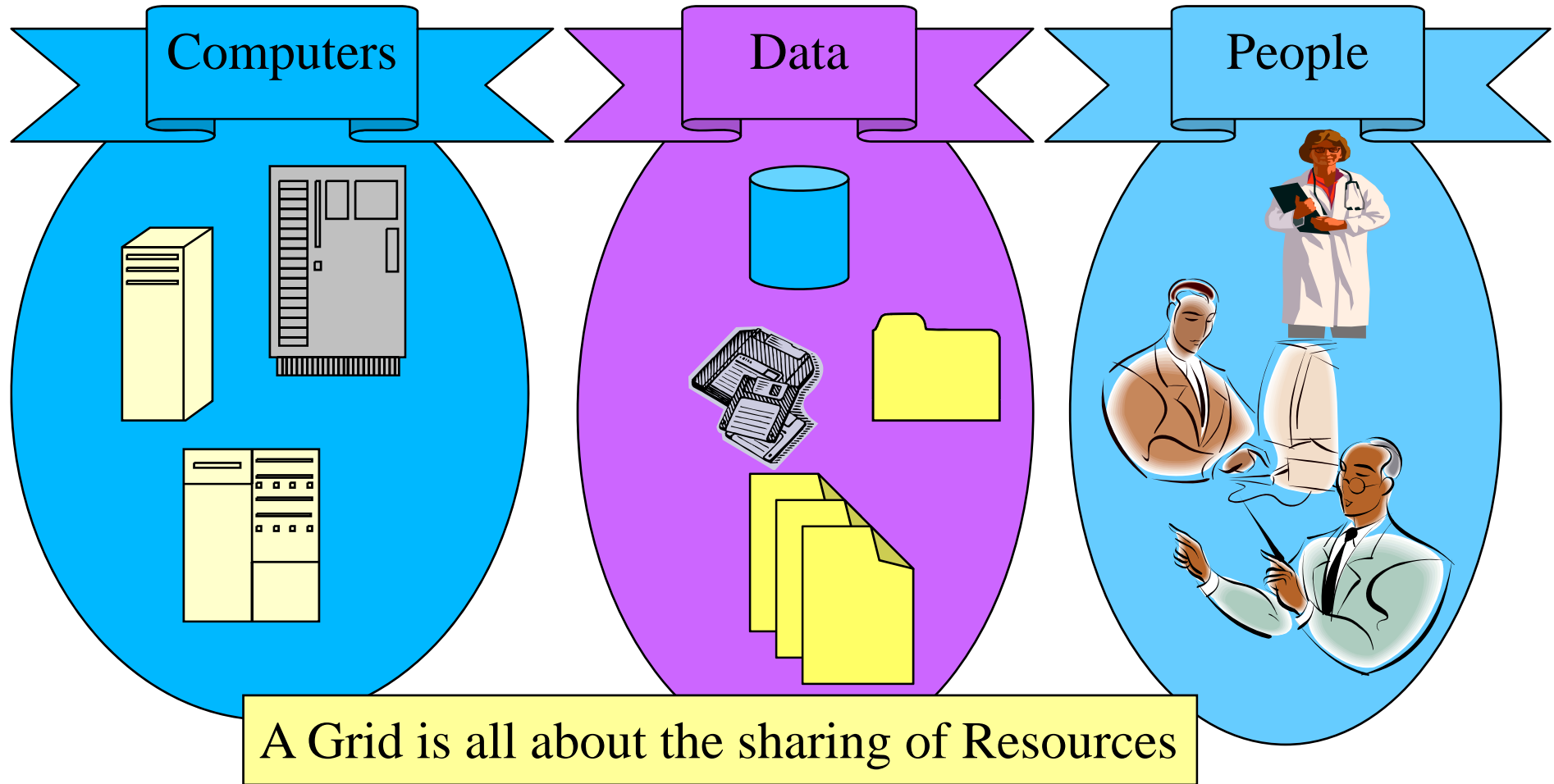
- **What is a Grid?**
- **What is a Grid Data Service?**
- **Why the “OGSA-DAI” acronym?!**
- **Why does OGSA-DAI matter?!**
- **When should we use OGSA-DAI?**

What is a Grid? - 1

- This is



What is a Grid?

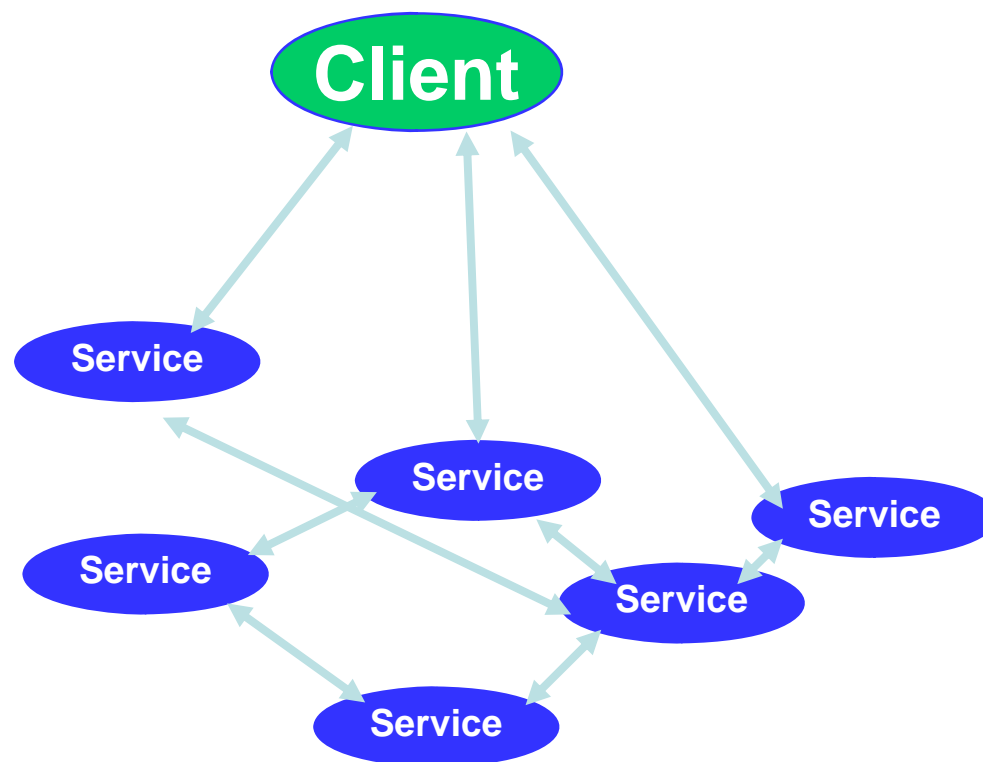


A Grid is..

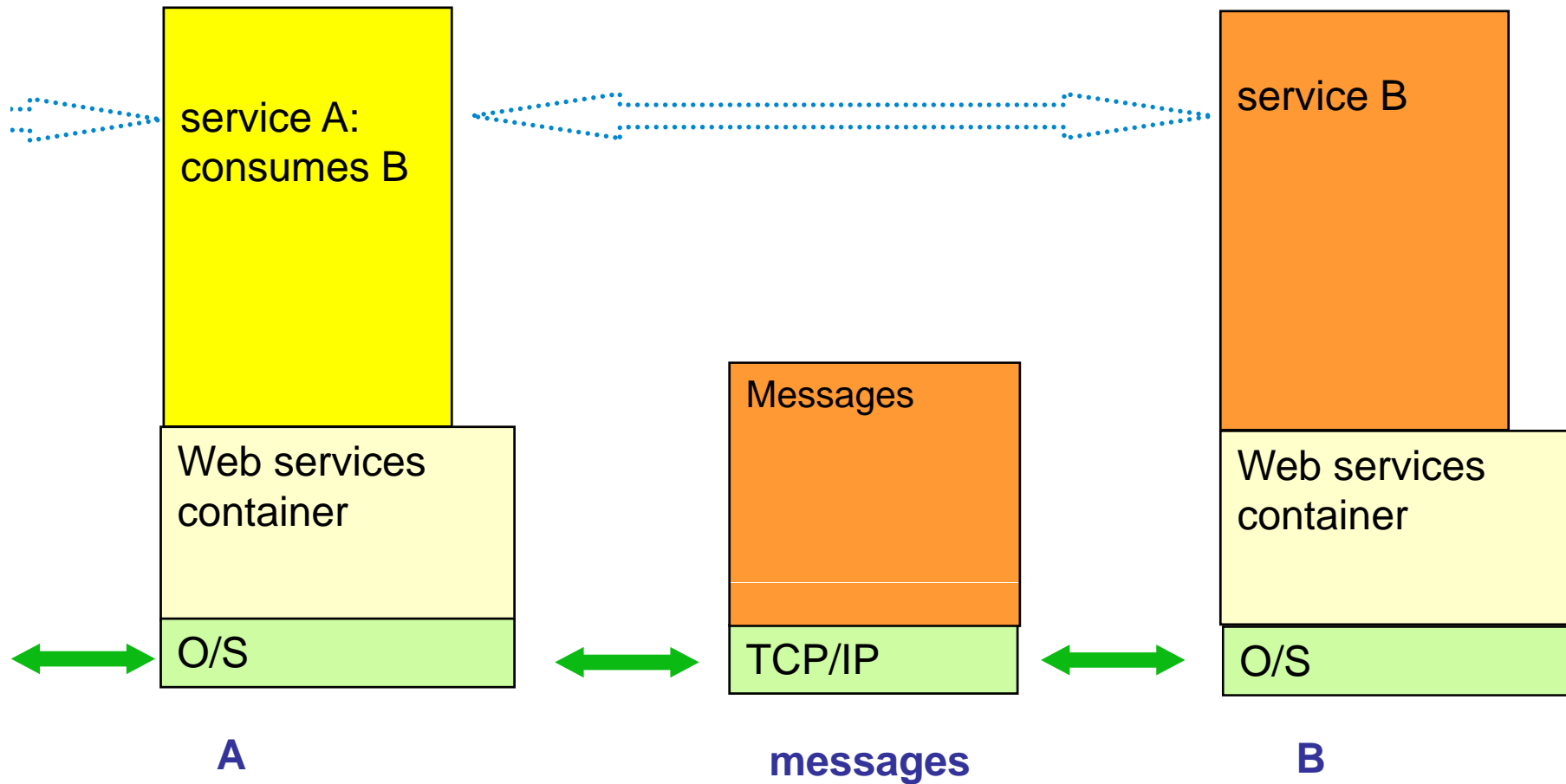
- ... all about the sharing of Resources
 - Within and between virtual organisations (= collaborations)
 - **Resources accessed by abstractions**
 - **User wants a job to run, wants to access data,...**
 - » **Rarely cares where this happens**
- ... a set of resources (and enabling services) that share mechanisms for
 - Authentication: communicate identity of user/provider
 - **X.509 certificate commonly used in “production grids”**
 - Authorisation: what can this user be allowed to do
 - **Member of which VO, which group,...**
 - Underpinned by agreement across VOs and resource providers
- ... infrastructure that builds on the Internet to permit orchestration of services across administrative domains

Web services – software components that are...

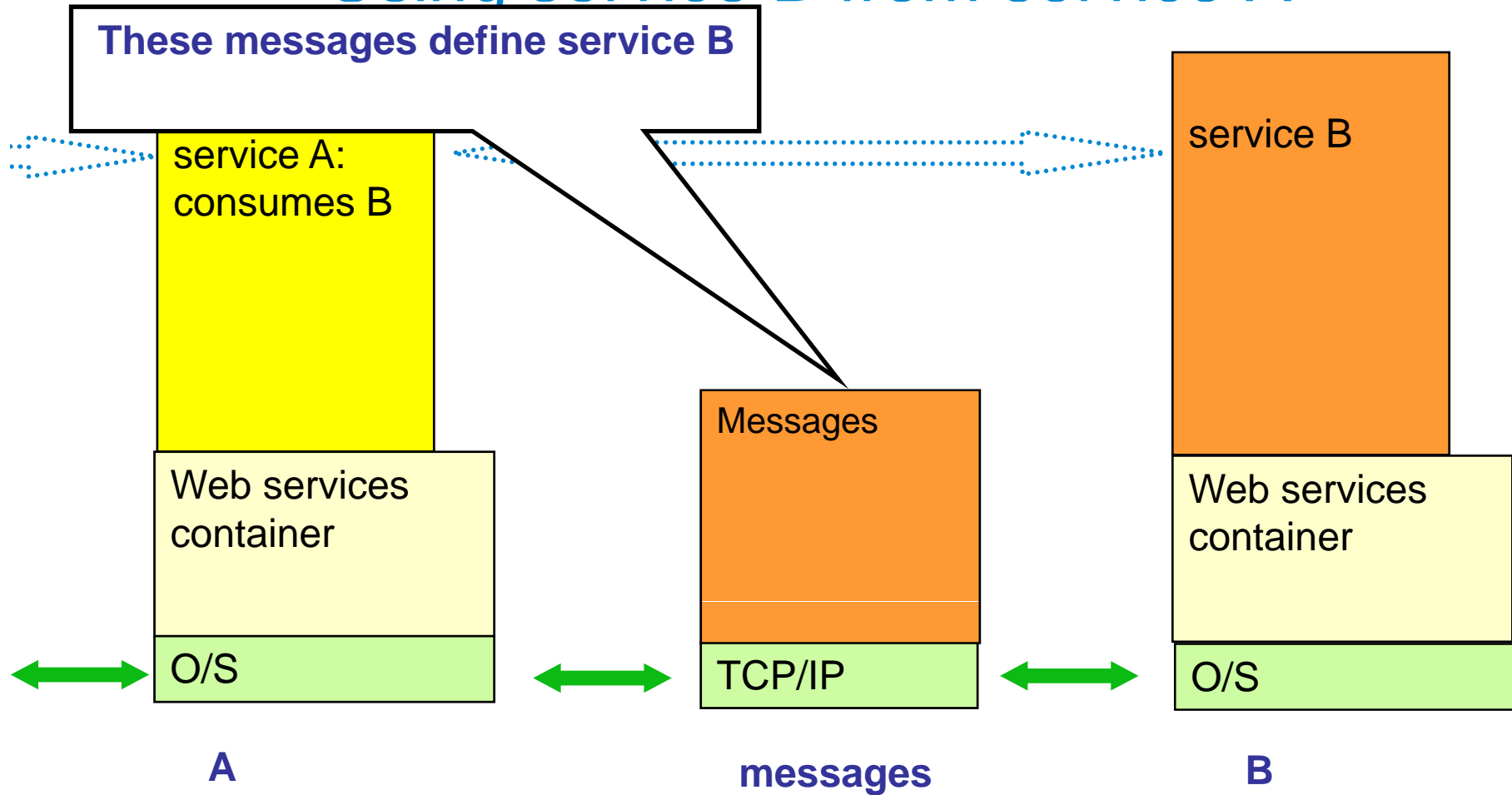
- Accessible across a network
- Loosely coupled, defined by the messages they receive / send
- Service description that can be used to create client software
- Based on standards (for which tools do / could exist)
- Developed in anticipation of new uses



Using service B from service A

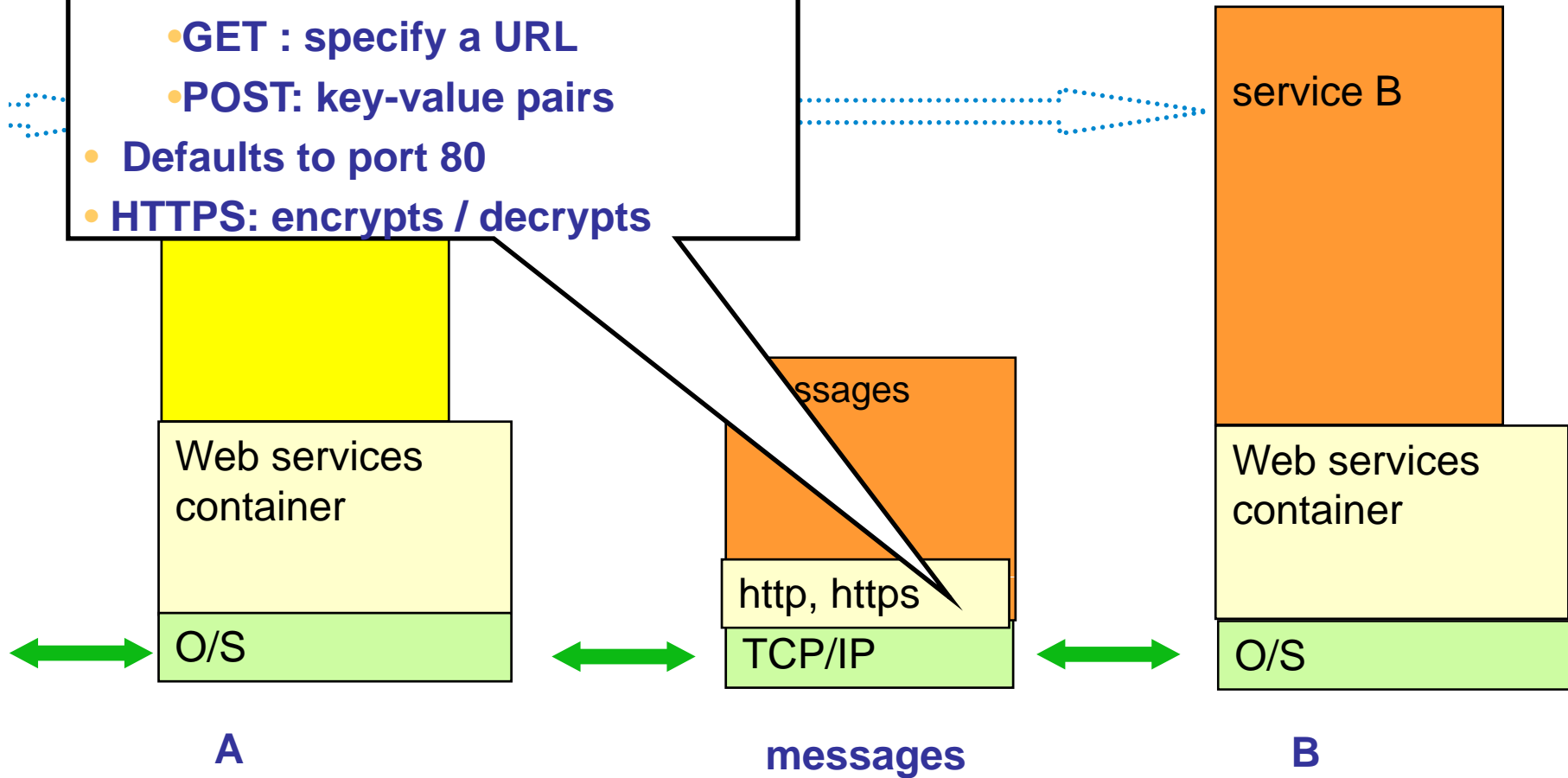


Using service B from service A

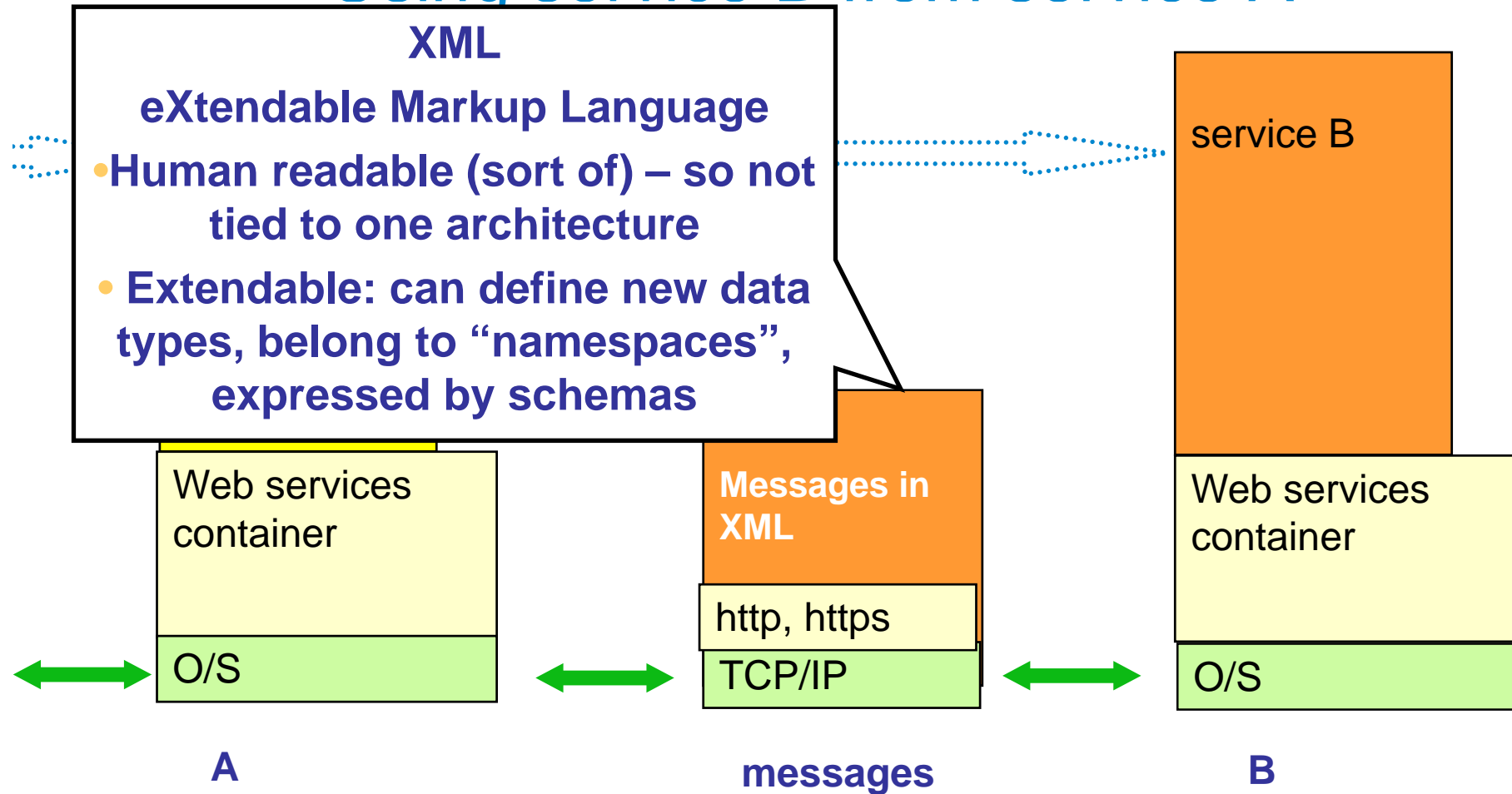


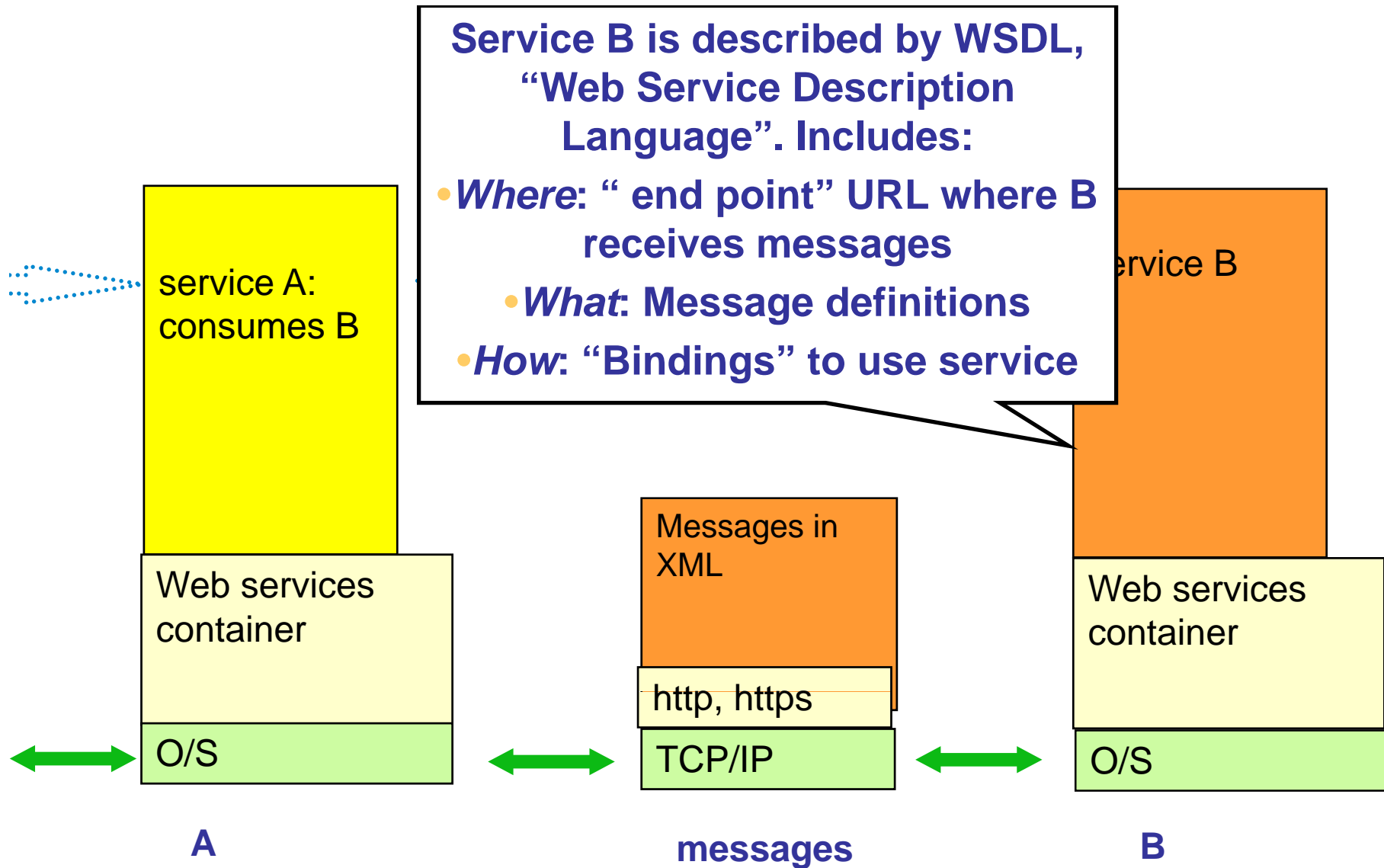
- Commonly used for WS - original purpose: carry HTML
- HTTP request methods
 - GET : specify a URL
 - POST: key-value pairs
- Defaults to port 80
- HTTPS: encrypts / decrypts

B from service A



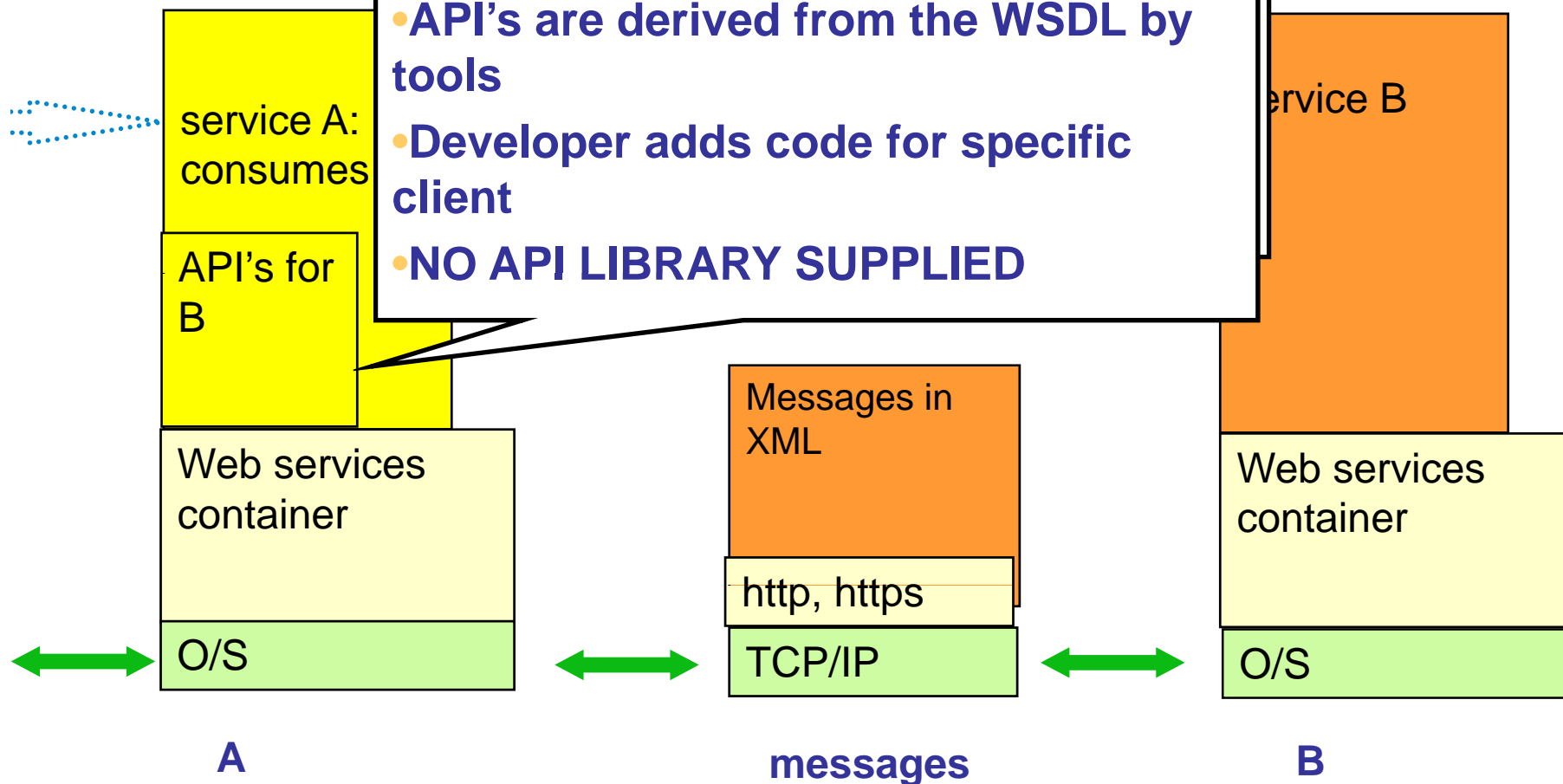
Using service B from service A



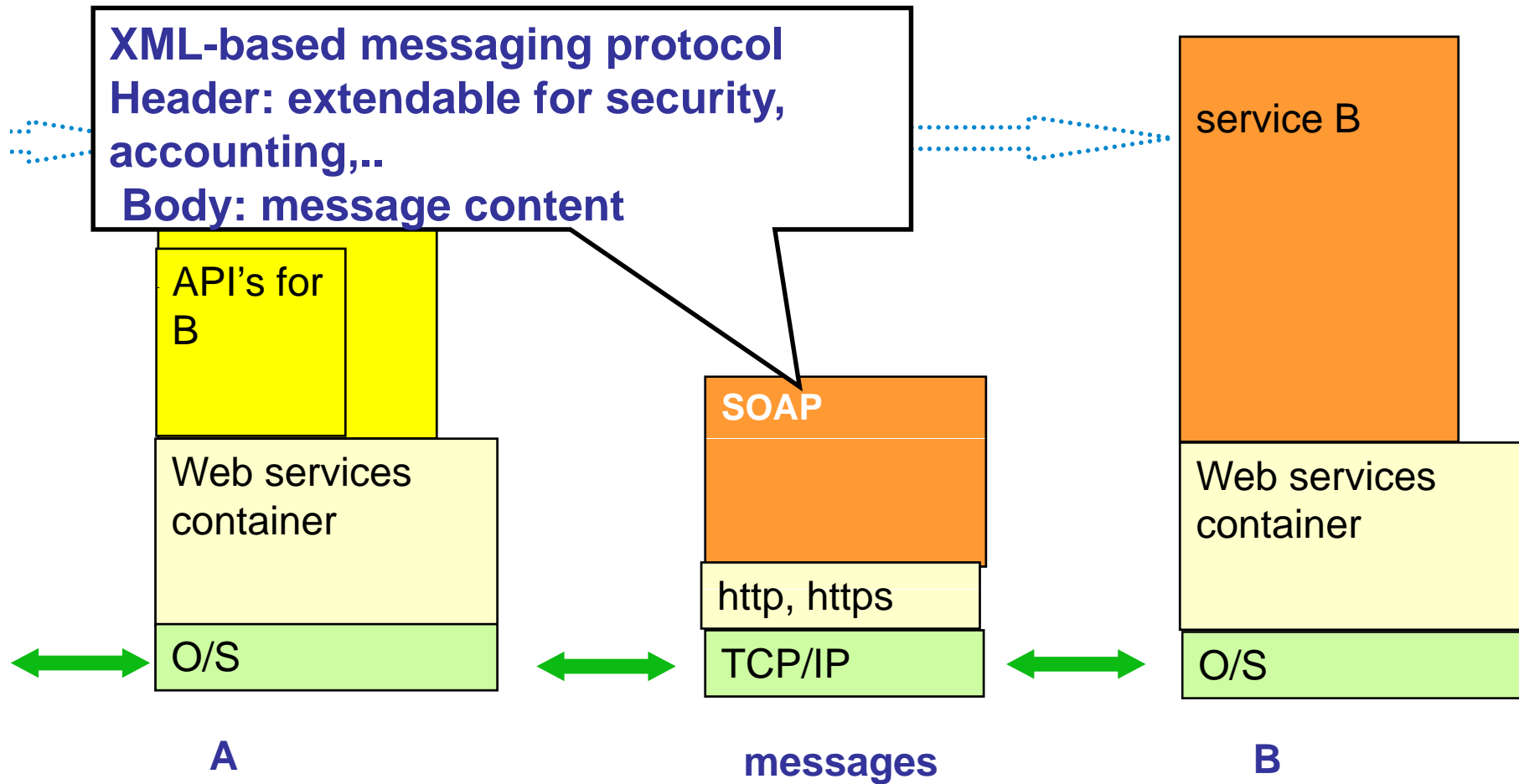


Service B is described by WSDL, “Web Service Description Language”. Includes:

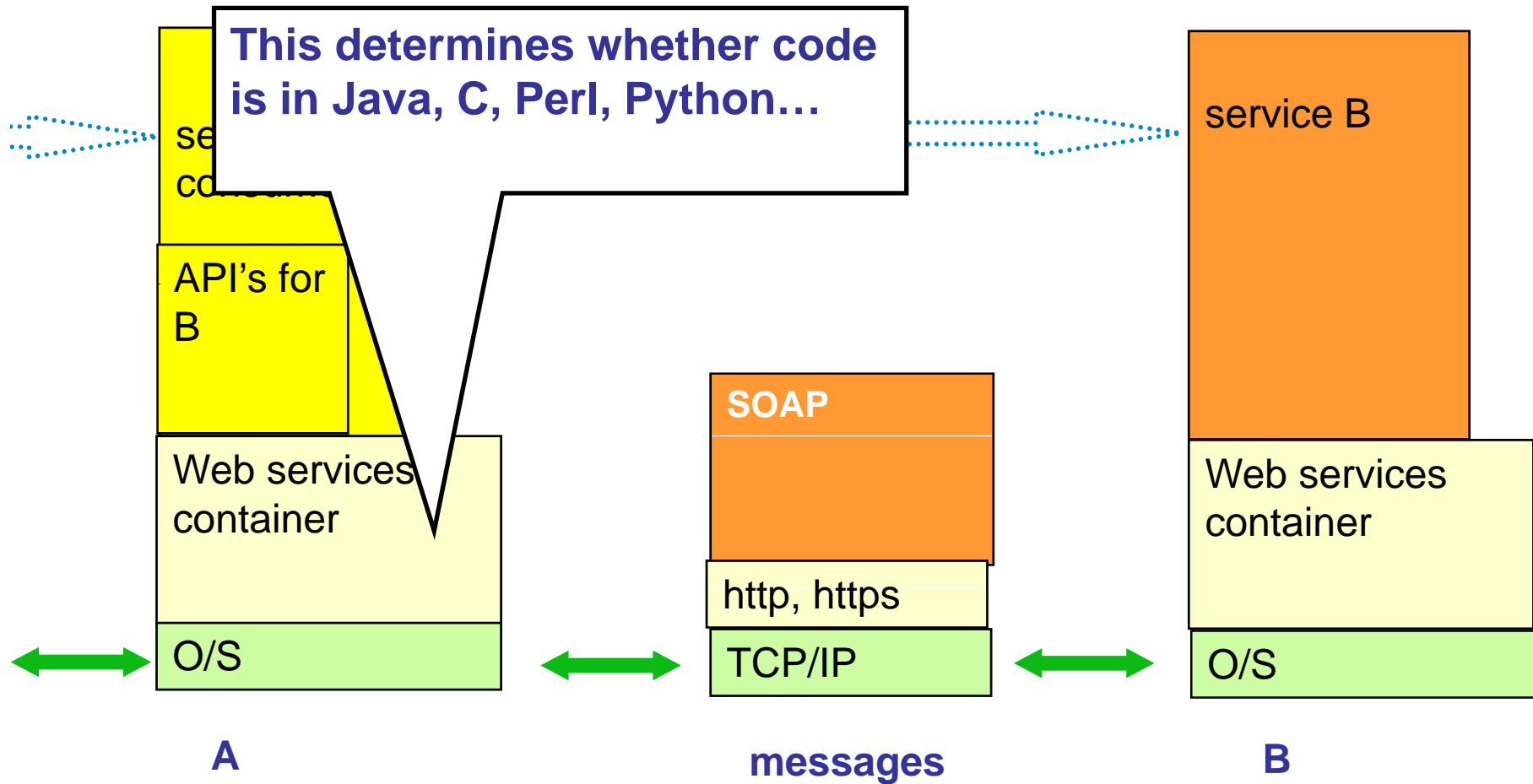
- API’s are derived from the WSDL by tools
- Developer adds code for specific client
- **NO API LIBRARY SUPPLIED**



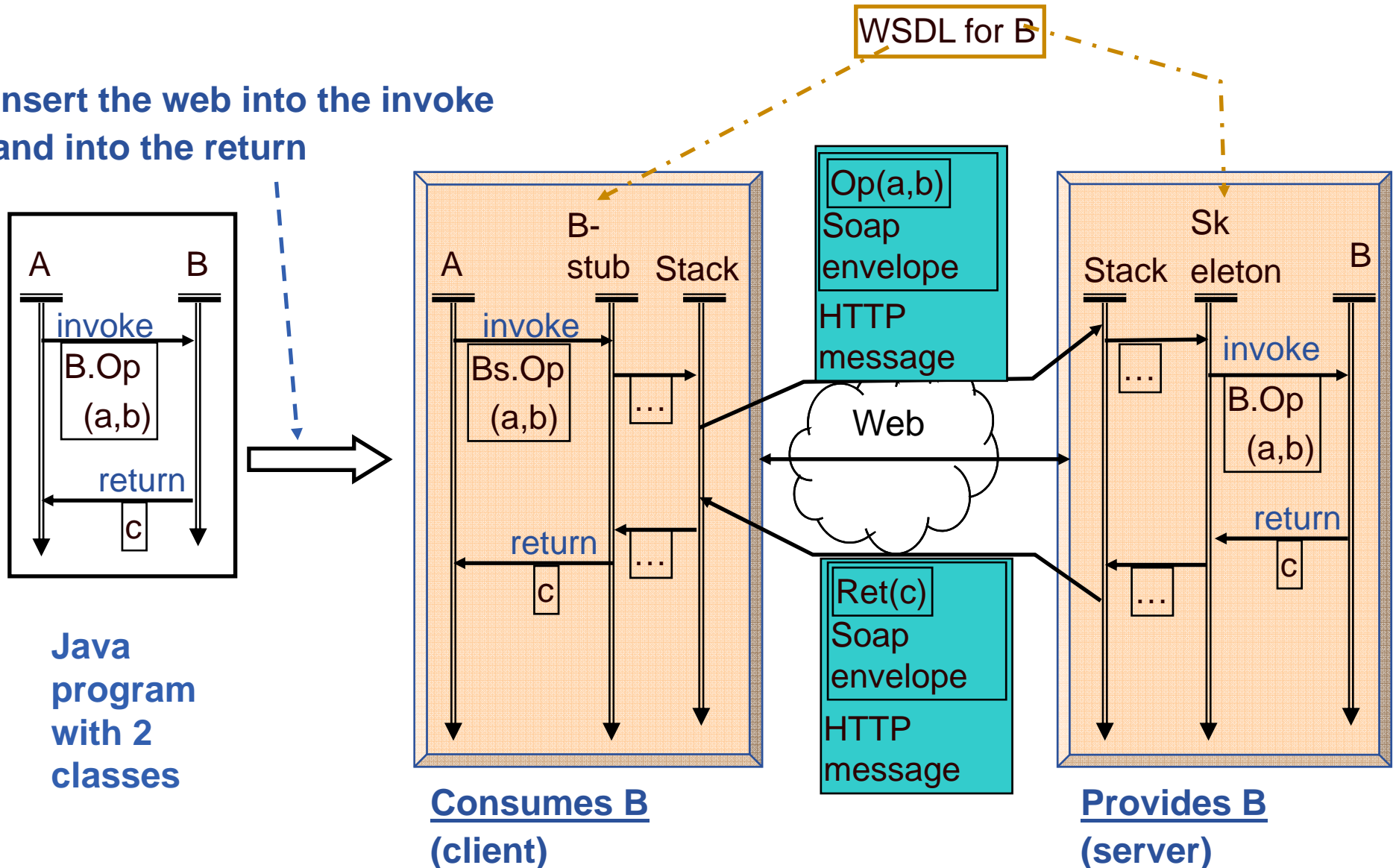
SOAP



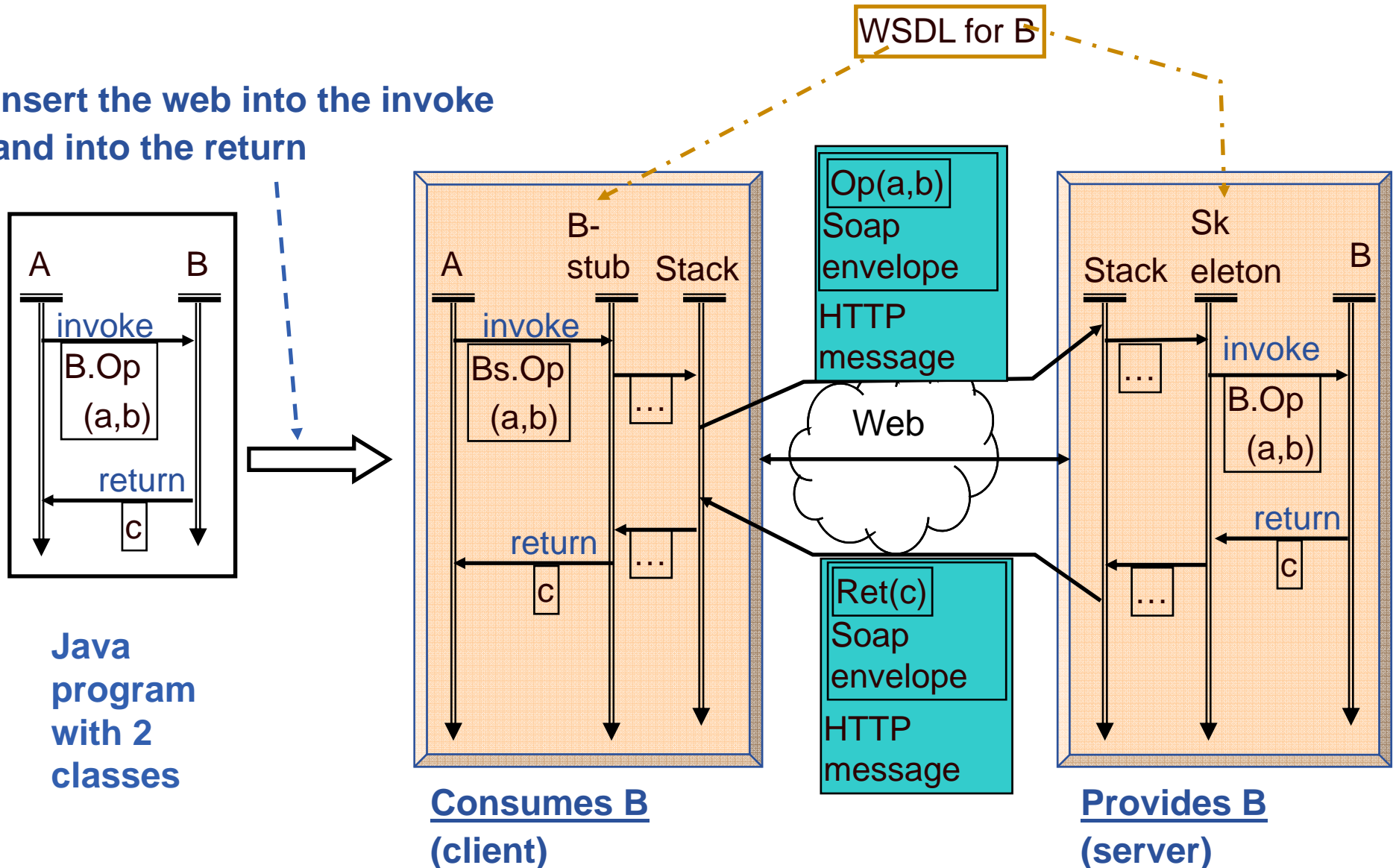
Code languages



Insert the web into the invoke and into the return

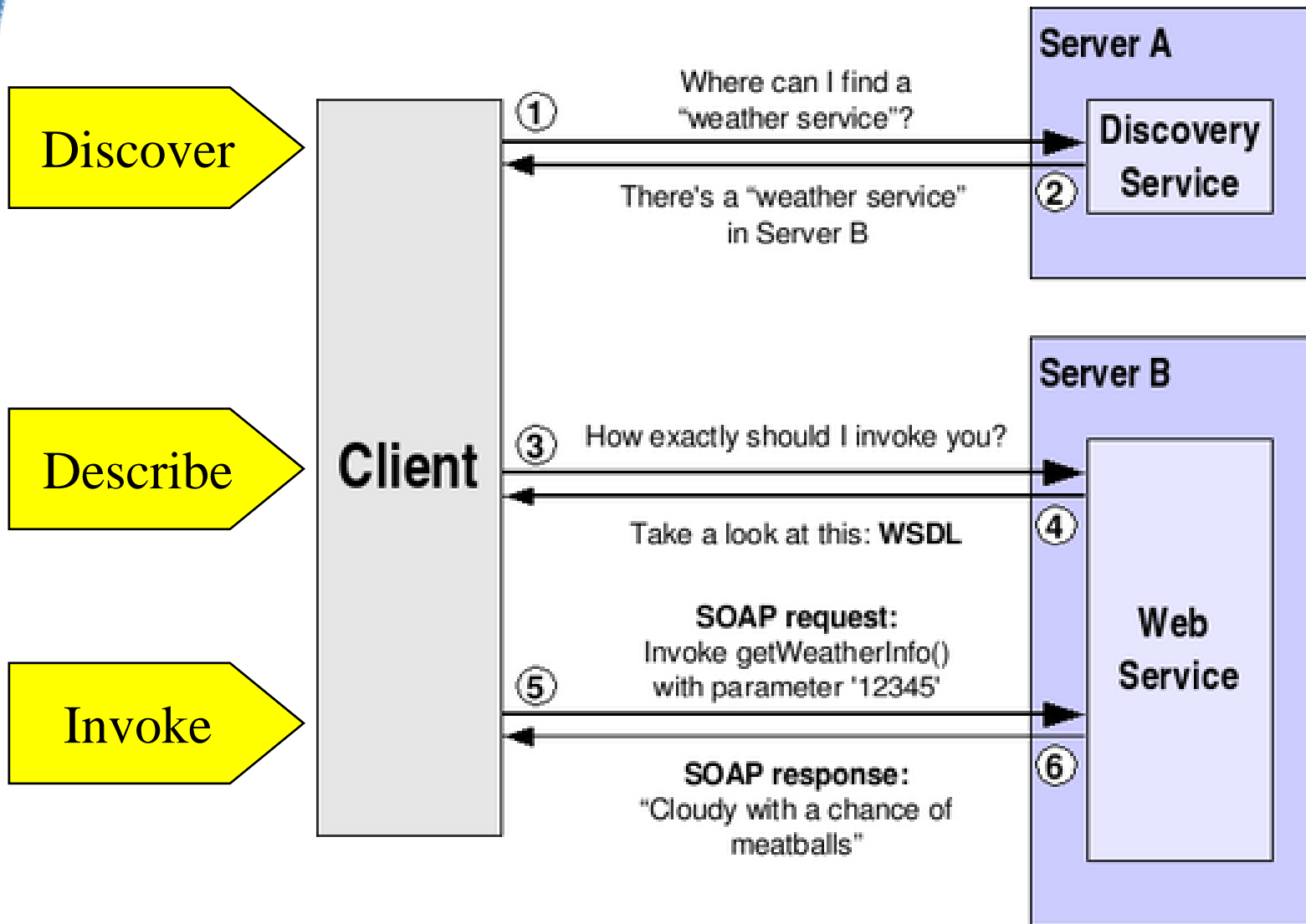


Insert the web into the invoke and into the return





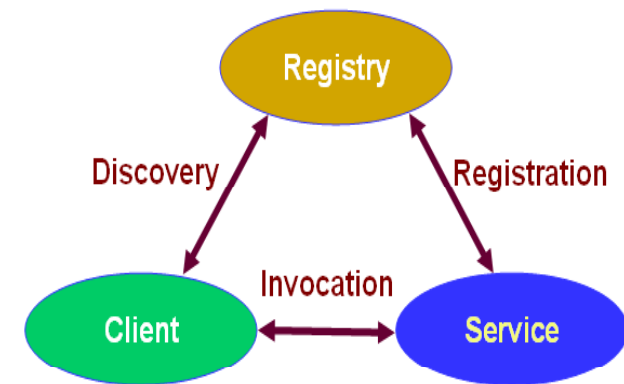
Real Web Service Invocation



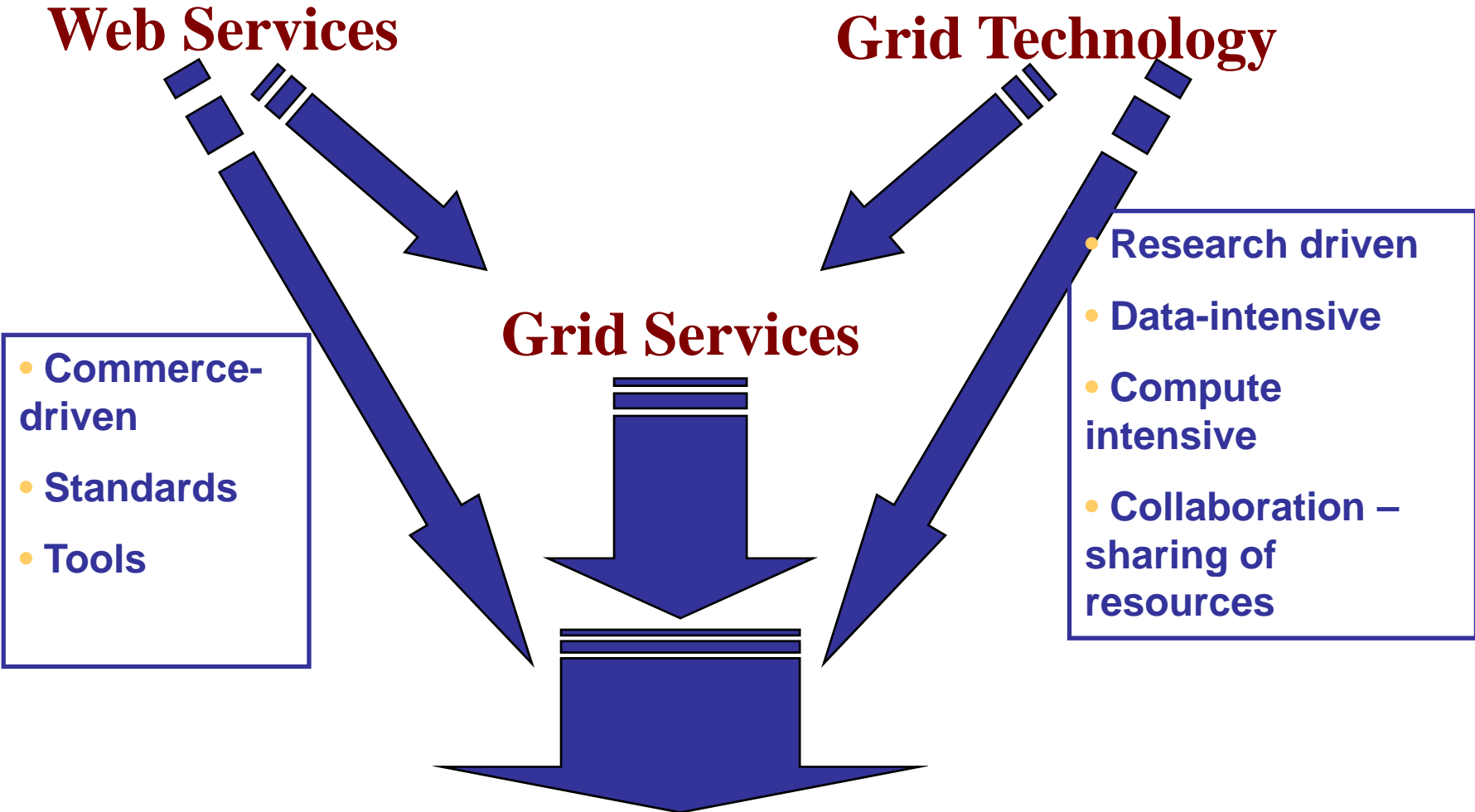
- **WS-I (Interoperability) delivers practical guidance, best practices and resources for developing interoperable Web services solutions.**
- **<http://www.ws-i.org/>**

Open standards:

- **SOAP: protocol for message passing**
- **Web Service Description Language: to describe services**
- **UDDI: Universal Description, Discovery and Integration**
- **WS-Security: incorporates security**



October 2001 View



2007 View

Web Services

- **Basis for defining standards for different services**
- **For services on grids:**
 - Need to manage state – interact with resources
 - Need to be notified of change of state

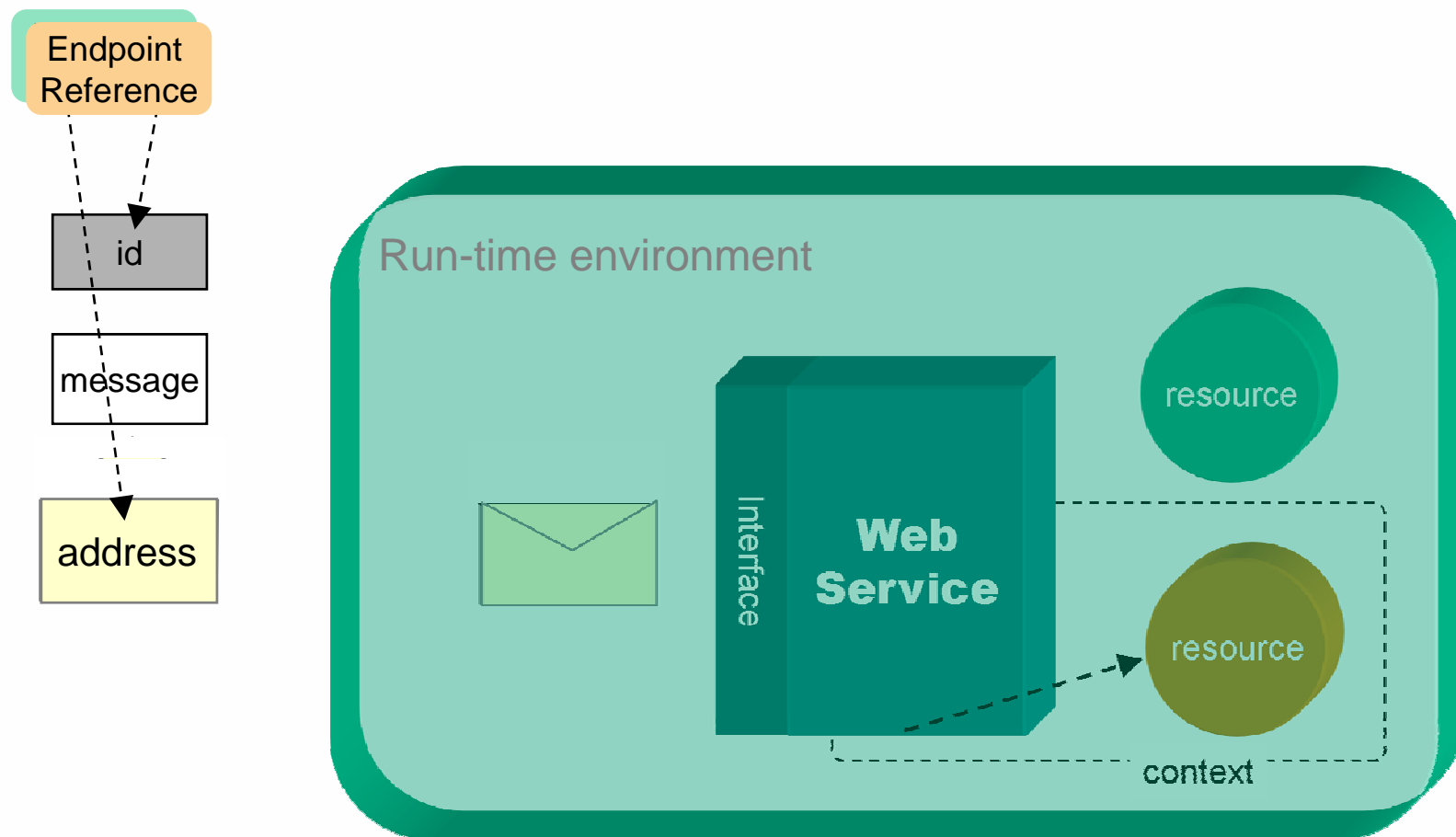


WSRF – Web Services Resource Framework

WS-Notification

The WS-Resource framework model

Using a Web service to access a WS-Resource



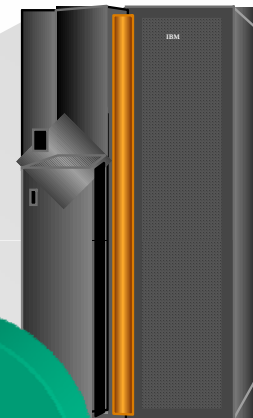
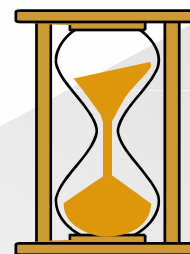
The WS-Resource framework model

■ WS-Resource Properties

- Resource state and metadata
“Projected” as an XML document
- Query and Set operations

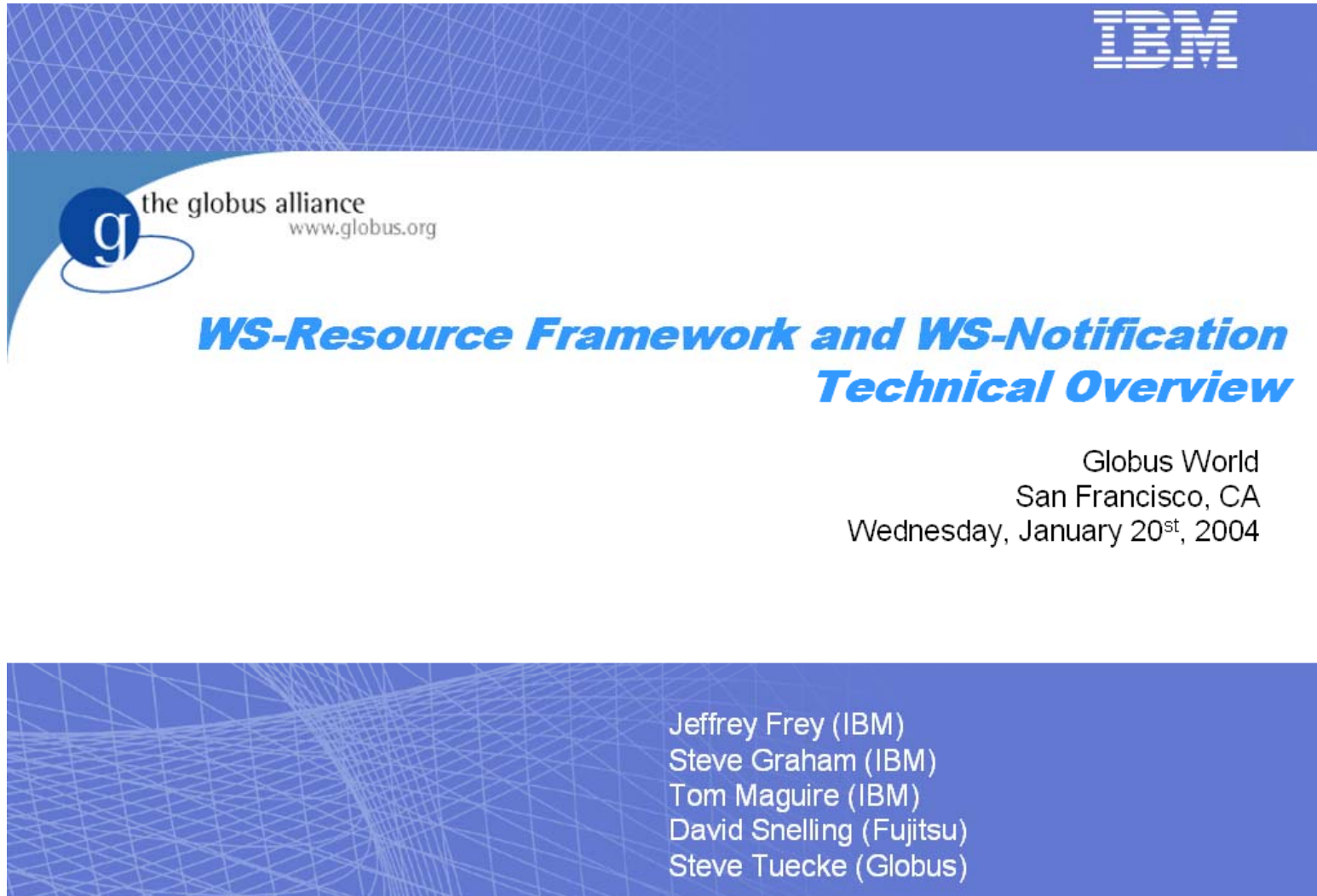
■ WS-Resource LifeTime

- Explicit destruction or
“Soft state” time-to-live
- Provides for cleanup
of resource instances



```
<ProcessorProperties>  
  <ProcID>5A34C1DE03</ProcID>  
  <ProcArchitecture>Power6.2</ProcArchitecture>  
  <ProcSpeedMIPS>400</ProcSpeed>  
  <ProcCacheMB>256<ProcCache>  
  <ProcRunning>1</ProcRunning>  
  
</ProcessorProperties>
```

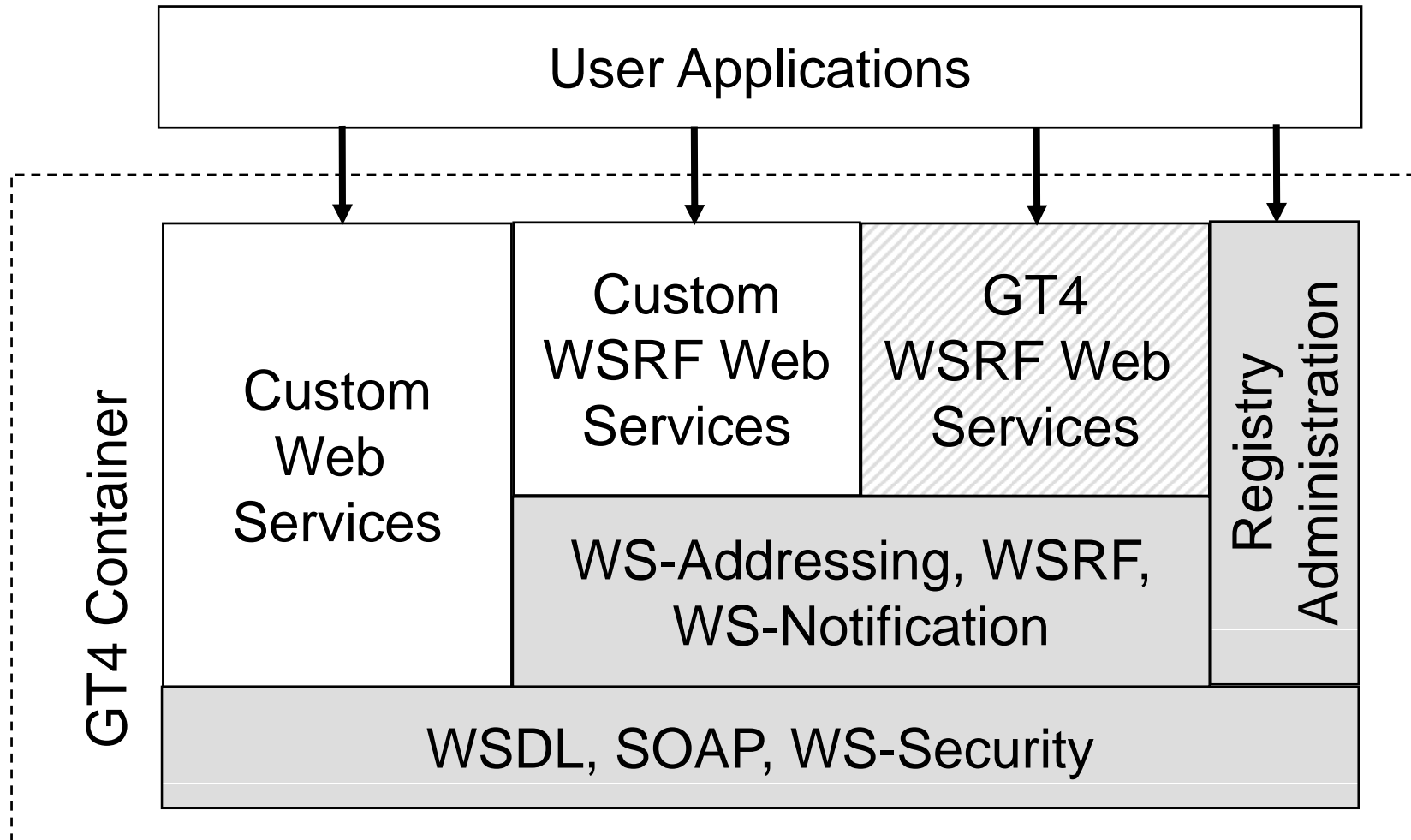
<http://www.nesc.ac.uk/action/esi/contribution.cfm?Title=385>

The slide content is presented on a blue background with a white grid pattern. At the top right is the IBM logo. On the left is the Globus Alliance logo, which consists of a blue circle with a white 'g' inside, followed by the text 'the globus alliance' and 'www.globus.org'. The main title is 'WS-Resource Framework and WS-Notification Technical Overview' in a bold, blue, italicized font. Below the title, the location and date are listed: 'Globus World', 'San Francisco, CA', and 'Wednesday, January 20st, 2004'. At the bottom right, a list of names and affiliations is provided: Jeffrey Frey (IBM), Steve Graham (IBM), Tom Maguire (IBM), David Snelling (Fujitsu), and Steve Tuecke (Globus).

- **Among middleware that uses WSRF/ WS-N:**
 - **Globus Toolkit 4**
 - **UNICORE**



Globus Toolkit 4 Web Services Core

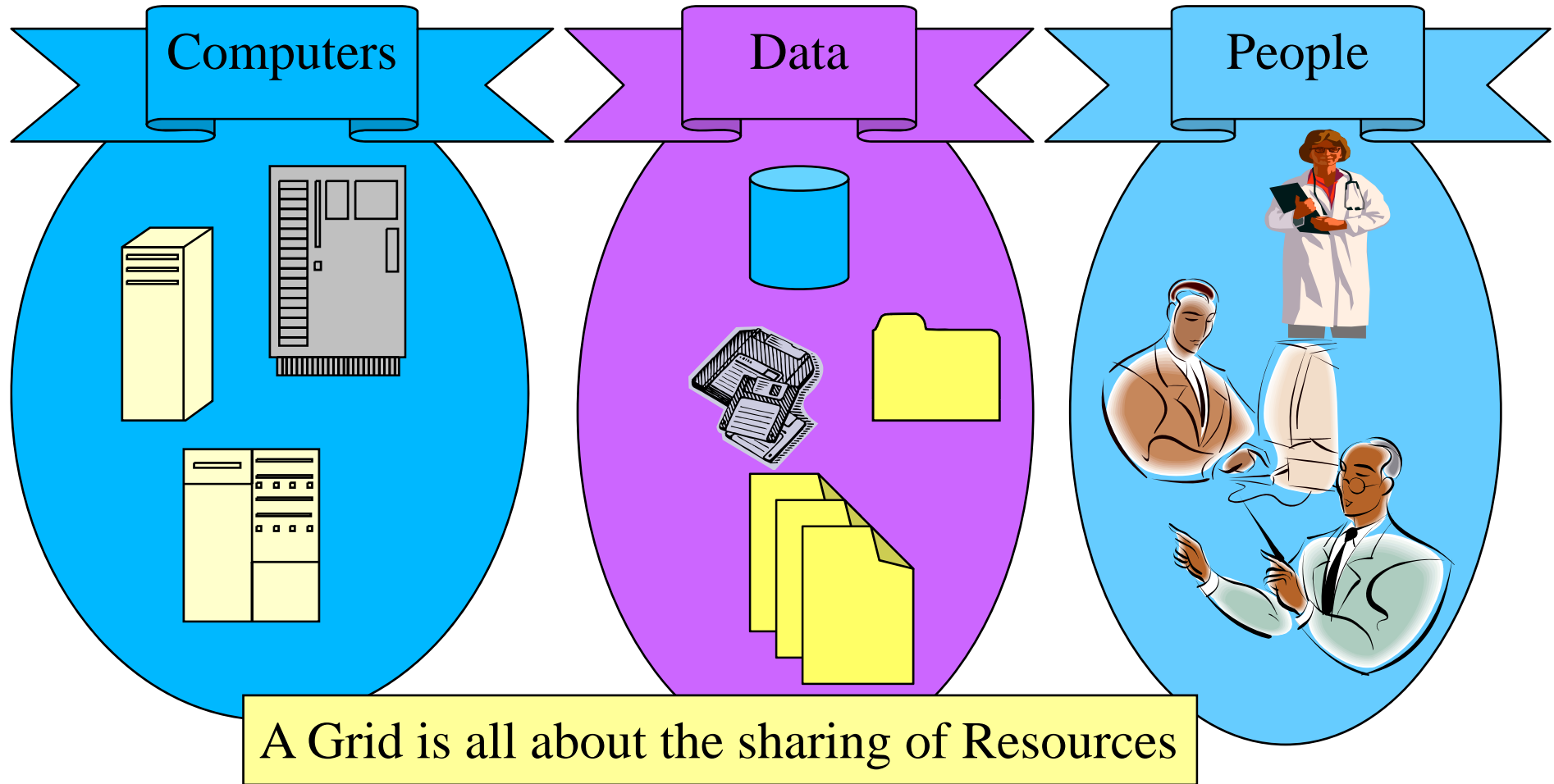


Thanks to J. Schopf, ANL

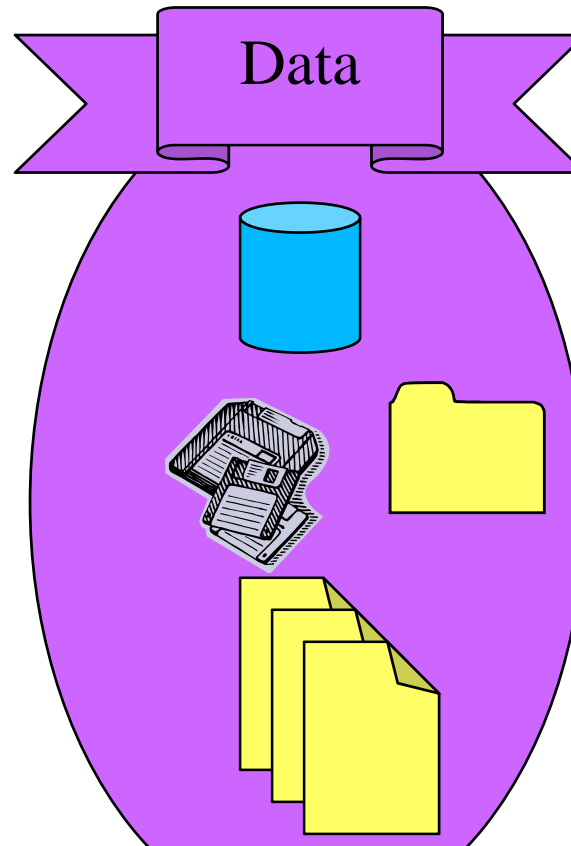
Story so far....

- **Grids orchestrate resources**
- **They do this with many services that (increasingly often) are exposed as web-services**
- **“Grid services” have to manage state, notification: standards for this are WS-Resource Framework and WS-Notification - e.g. Globus Toolkit 4**
- **Where do “grid data services” fit in?**

What is a Grid?



Focus on Data



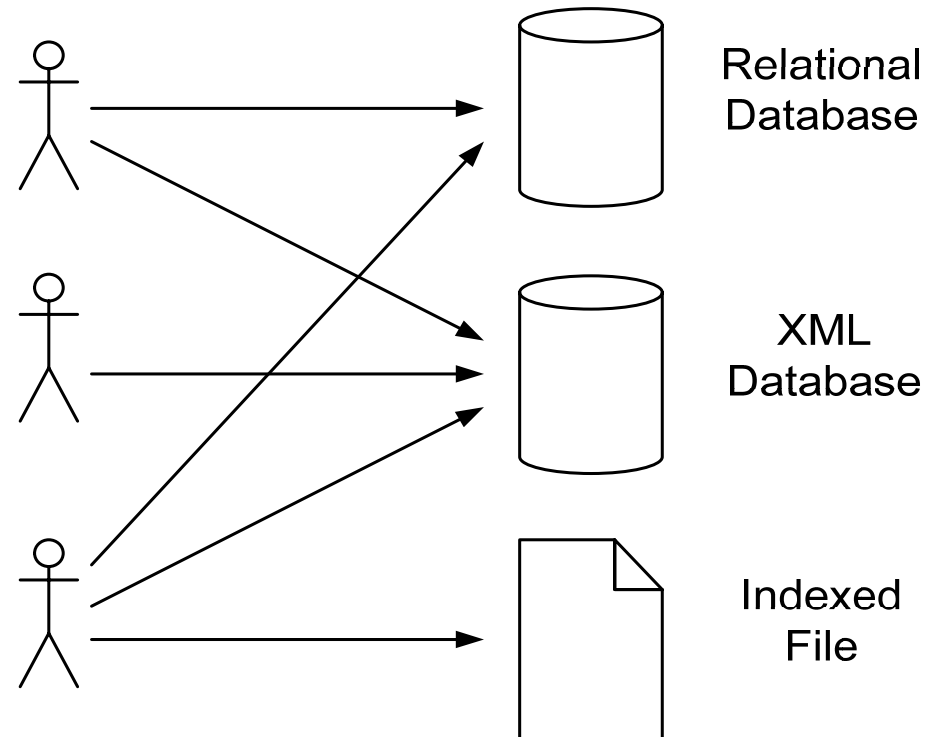
OGSA-DAI enables the sharing of Data Resources

Types of data services

- **Many user communities manage data in grid vaults (aka storage elements)**
 - **Experimental data**
 -
 - **Replicated for resilience**
 - **And to be close to where computation will happen**
- Many new user communities have more diverse data resources
- ***To facilitate new research need data to be accessible from Grid infrastructures***
- Resources:
 - May pre-date Grids
 - Providers may have current ways to distribute data to users
 - May not be able to replicate data
 - ***Need AuthN and AuthZ***

Motivation

- Grid is about sharing resources
- OGSA-DAI is about sharing structured data resources



Life before OGSA-DAI....

- A few examples follow of alternative approaches to sharing data.

Sharing data via web site download

- ZIP up data and put it on a web site
- Pros
 - Easy distribution for providers
 - Easy access for consumers
- Cons
 - Consumers have to download all the data
 - Consumers have to load data into local databases to use it
 - Static snapshot
 - Security

Sharing data via direct access

- **Providers tell consumers**
 - Database URL – mycomputer.epcc.ed.ac.uk:3306
 - Username – userID
 - Password – password
- **Pros**
 - Consumers have direct access
- **Cons**
 - Firewall issues
 - User and password management is hard
 - No consistent security model
 - Hard to use in grid/web service workflows

Sharing data via direct access

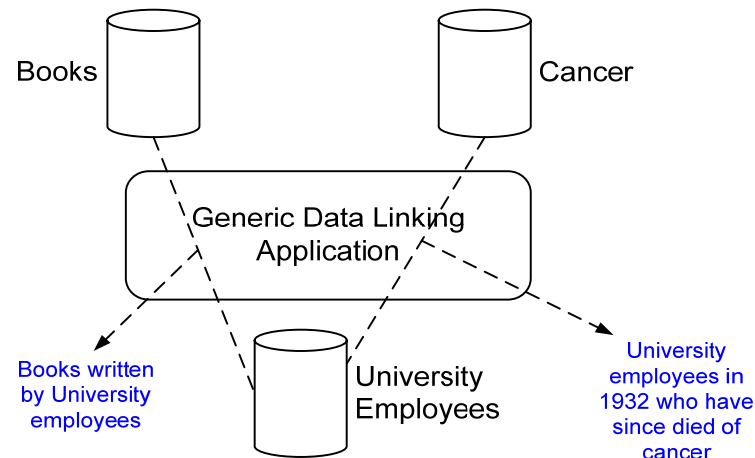
- **Cons (continued)**
 - No server-side layer in which to standardize database heterogeneities
 - Myriad drivers
 - Different APIs across different data types
 - Relational and JDBC
 - XML and XMLDB
 - Indexed files and Lucene

Domain-specific web services

- Manipulate data using domain-specific operations, e.g.
 - `Book findByISBN(ISBN)`
 - `List<Book> findByAuthor(Author)`
 - `List<Book> findByKeyword(Word)`
- Pros
 - Fits with grid/web service approach
 - Abstraction hides back-end database details
 - Web services are programming language neutral
 - Operations likely to map well to authorization policies

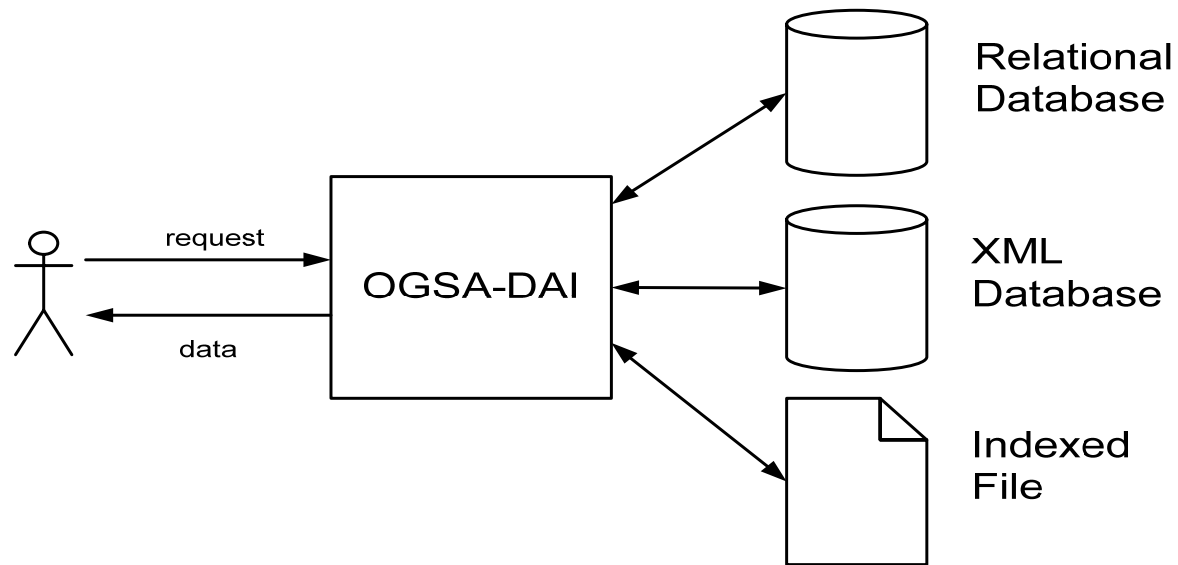
Domain-specific web services

- Cons
 - Slower than direct access
 - Web service layer
 - SOAP transport overhead – especially for large result sets
 - Domain-specific API prevents use of generic data exploration, mining and manipulation tools



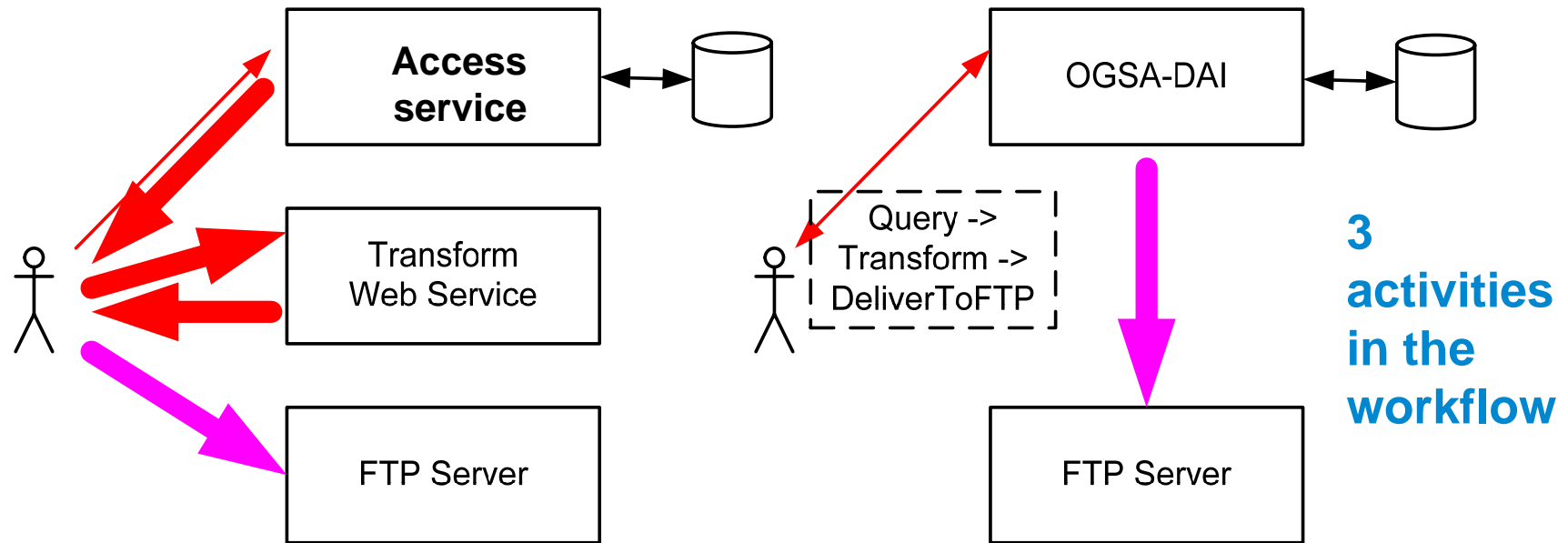
OGSA-DAI generic web services

- Manipulate data using OGSA-DAI's generic web services

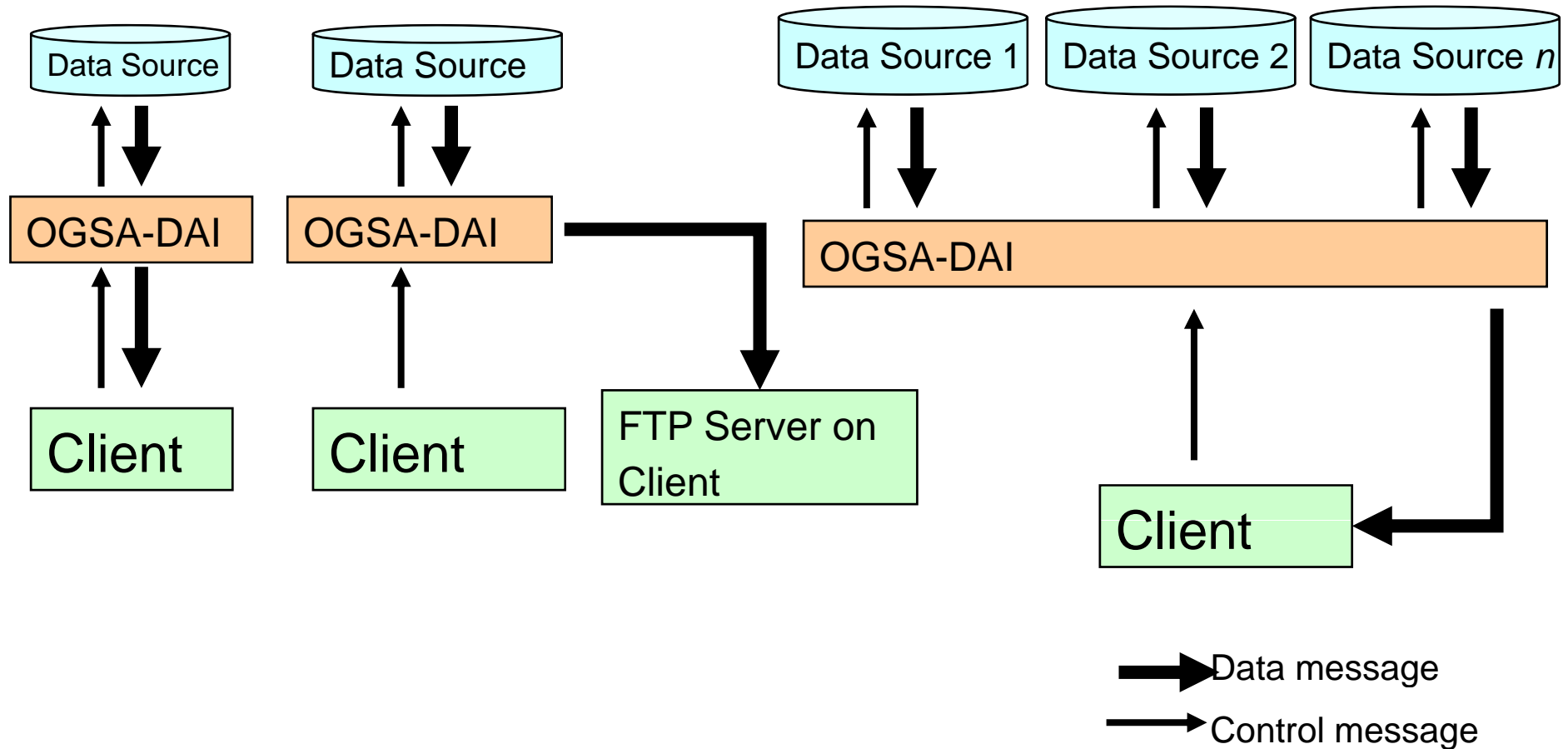


Importance of workflows

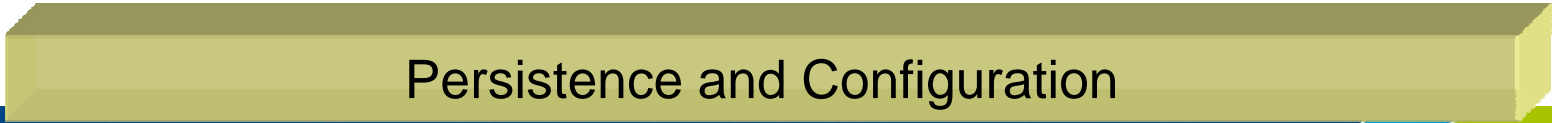
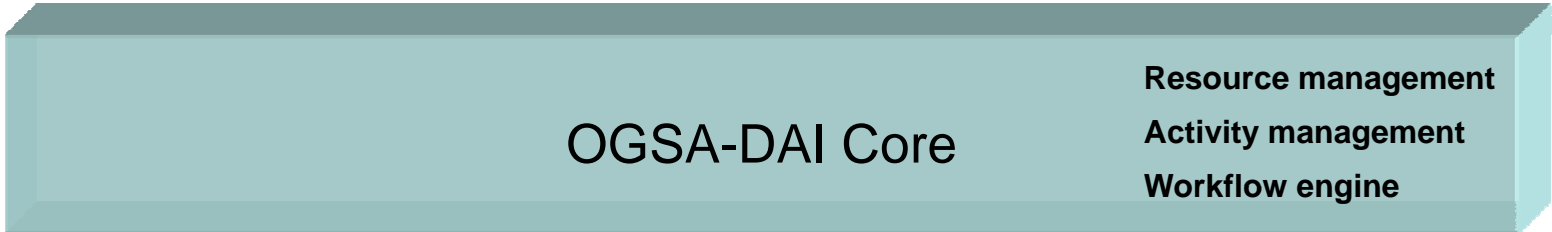
OGSA-DAI server
is close to data



Usage Scenarios



OGSA-DAI 3.0



Typical roles

- **Researcher**
 - Wants to use data from context of known application, easy portal, workflow..
- **Data publisher**
 - Deploys OGSA-DAI server
 - Determines AuthN and AuthZ policies for their data
 - Establishes activities (= workflow components)
- **Informatician / Application developer**
 - Deploys client software
 - Uses Java to build workflow
 - Exposes client for...

OGSA-DAI 3.0

- **OGSA-DAI has evolved constantly since February 2002**
- **OGSA-DAI 2.2 released April 2006**
- **As the number of users grew so did the requirements**
 - **More effective data streaming**
 - **Standardisation of activity inputs and outputs**
 - **Targeting multiple data resources in a single workflow**
 - **Supporting application-specific presentation layers**
- **OGSA-DAI 2.2 was not suitable for addressing these**
- **OGSA-DAI 3.0**
 - **A complete re-design and re-implementation of OGSA-DAI**
 - **A stable framework for the future**
 - **Released September 2007**

Where might OGSA-DAI not be suitable?

- **OGSA-DAI is not**
 - A complete solution to every data-related problem
 - A replacement for or competitor to JDBC
 - Just about accessing relational databases
- **It is not suitable if**
 - You have a single data resource that isn't going to change
 - You have no data transformation requirements
 - You want rapid access to data in a single data resource

What is OGSA-DAI?

- **An extensible framework**
- **accessed via** web services
- **that executes data-centric** workflows
- **involving** heterogeneous **data resources**
- **for the purposes of** data access, integration, transformation and delivery
- **within a grid**
- **and is intended as a toolkit for building higher-level application-specific data services**

Thank you!

<http://www.ogsadai.org.uk>

<http://omii-europe.org>

