



Alexandra Nenadic, **Stian Soiland-Reyes**, Carole Goble
University of Manchester
All Hands Meeting, Oxford 2009-07-12

**BRINGING CABIG SERVICES
TOGETHER USING TAVERNA**

What is caBIG?

- **Cancer Biomedical Informatics Grid**
- Biomedical platform by National Cancer Institute (NCI) in US
- Linking cancer institutes and research organisations across the country
- Improving the US cancer research community's access to bioinformatics data and tools.

caBIG services

- ~120 service providers across US





caBIG services

- 75 data services
- 53 analytical services
- Data types semantically described using a controlled vocabulary – *Silver level compatibility*
- Services registered in an *Index Service*, with service metadata (*caDSR*)
- Security infrastructure (*GAARDS*), delegation service, single sign-on
- **Thus – forming a grid - caGrid**

Taverna workbench

- Open source *workflow* authoring and execution tool

The screenshot displays the Taverna Workbench 2.1-rc-1 interface. The top window title is "Taverna Workbench 2.1-rc-1". The interface is divided into several panels:

- Service panel:** Located on the left, it features a "Filter:" input field and a "Clear" button. Below is an "Import new services" button and a list of "Available services" including "Service templates", "Local services", and various external services like "Biomart", "Biomoby", "Soaplab", and "WSDL".
- Workflow explorer:** Located below the service panel, it shows a tree view for a workflow named "FetchDragonimagesfromBioMoby". It includes sections for "Workflow input ports", "Workflow output ports", "annotations", "images", and "Services". The "Services" section is expanded to show "Decode_base64_to_byte" (with sub-items "base64" and "bytes") and "getDragonSimpleAnnotatedImages" (with sub-items "Object(input)", "SimpleAnnotatedJPEGImage(Collection - 'image' As Simples)", and "SimpleAnnotatedJPEGImage(Collection - 'image')").
- Workflow diagram:** Located on the right, it shows a flowchart of the workflow. It starts with "namespace" and "id" boxes pointing to an "Object" box. The "Object" box points to "getDragonSimpleAnnotatedImages", which then points to "getJpegFromAnnotatedImage". This box points to "Parse_Moby_Data_JPEGImage", which then points to "Decode_base64_to_byte" and "Parse_Moby_Data_SimpleAnnotatedJPEGImage". The "Decode_base64_to_byte" box points to "images" and "annotations" boxes, which are grouped under "Workflow output ports".



Taverna use

- Used by eScientists in fields such as:
 - Bioinformatics
 - Chemistry and cheminformatics
 - Astronomy
 - Text mining
 - Social sciences
 - Music
 - .. and more



Connecting services

- Taverna is oblivious to the science, workflows are built by dragging and dropping how the data should flow between different *services*
 - Web Services
 - Local scripts
 - Java APIs
 - R
 - .. other customized invocation mechanisms




caGrid technology

- WS-Resource Framework (WSRF) grid services
- Based on Globus Toolkit
- Secured services: Extending *Grid Security Infrastructure* (GSI) to form GAARDS (*Grid Authentication and Authorization with Reliable Distributed Services*)



caGrid extensions to Taverna

- Browse and semantic search registered caBIG services
 - caBIG single sign-on support through *GAARDS*
 - Support for WSRF stateful web services
- 

Finding caBIG services from Taverna

Taverna Workbench 2.1.1

File Edit Workflows Advanced Help

Design Results

Service panel

Filter:

Import new services

- ApiConsumer service...
- Biomart service...
- Biomoby service...
- caGrid service...**
- caGrid service from WSDL...
- Soaplab service...
- WSDL service...

WSDL @ http://soap.bind.ca/wsdl/bind.wsdl

WSDL @ http://soap.genome.jp/KEGG.wsdl

WSDL @ http://www.ebi.ac.uk/xenbl/XEMBL.wsdl

caGrid services search criteria

Select grid NCI Production CaGrid 1.3

Service query criteria:	Service query value:
Operation Class	Accession
None	AbstractProbe
None	AbstractProbeAnnotation
None	AbstractStudyDisease
None	Accession
None	AccessionCharacteristics
None	AccessRights
None	AcquisitionProcedure
None	ActionSuccessor

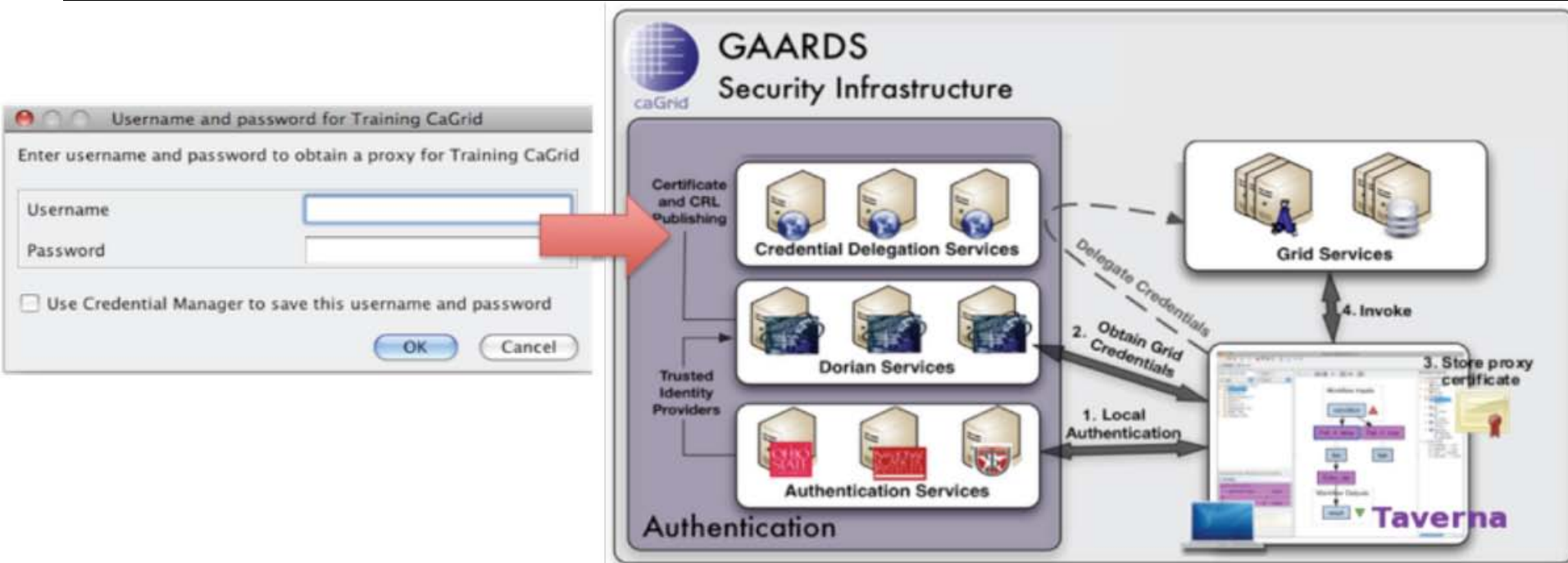
OK Update caDSR n

caBIG single sign-on

Sign in to chosen authentication service


Temporary *certificate* is issued

Credentials delegated to calls to grid services





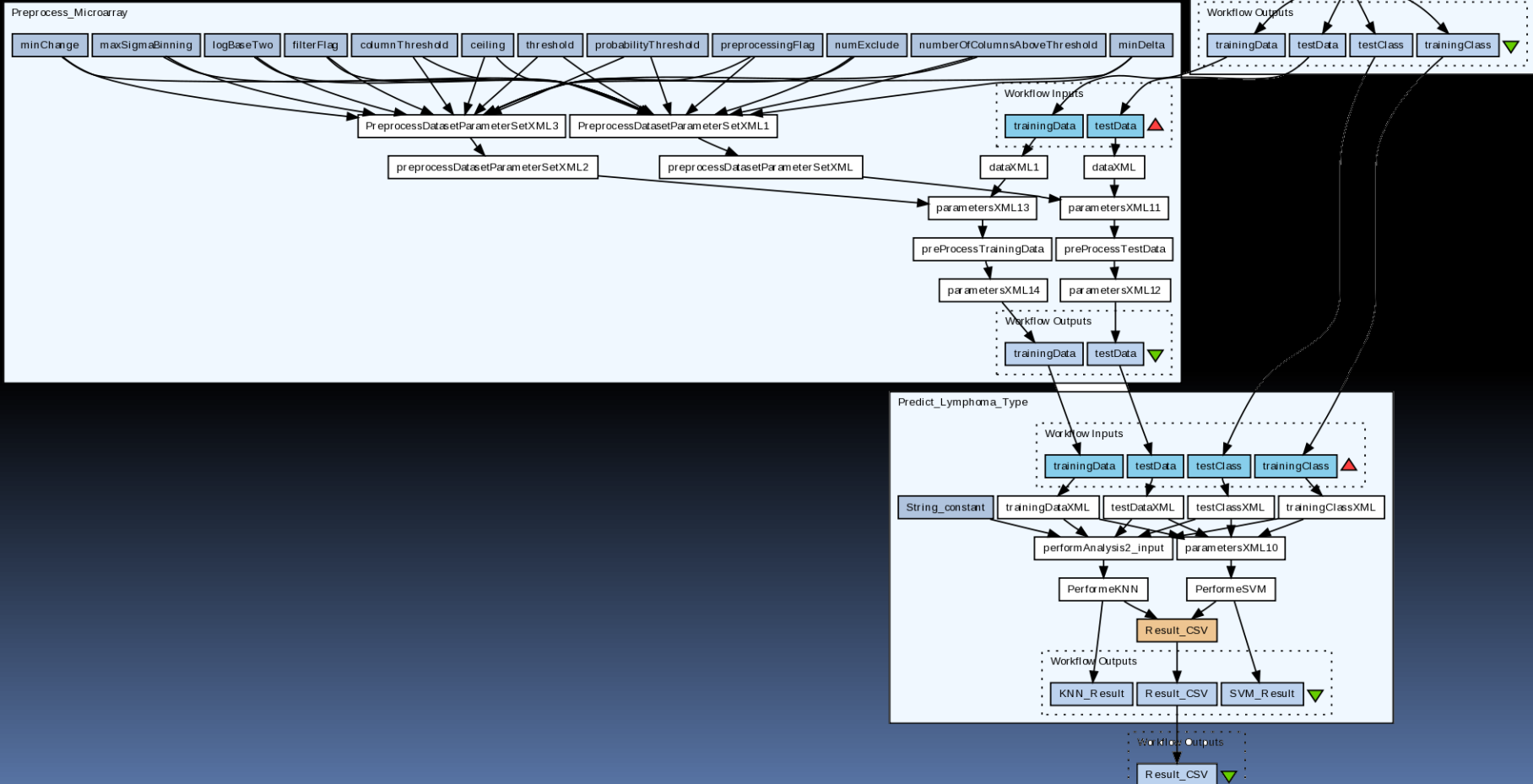
WSRF services

- Most caBIG services can be invoked as normal *WSDL* services from Taverna
 - Taverna extension add supports for stateful *WS-Resource Framework (WSRF)* services
 - Passing *endpoint references* as part of the workflow
- 

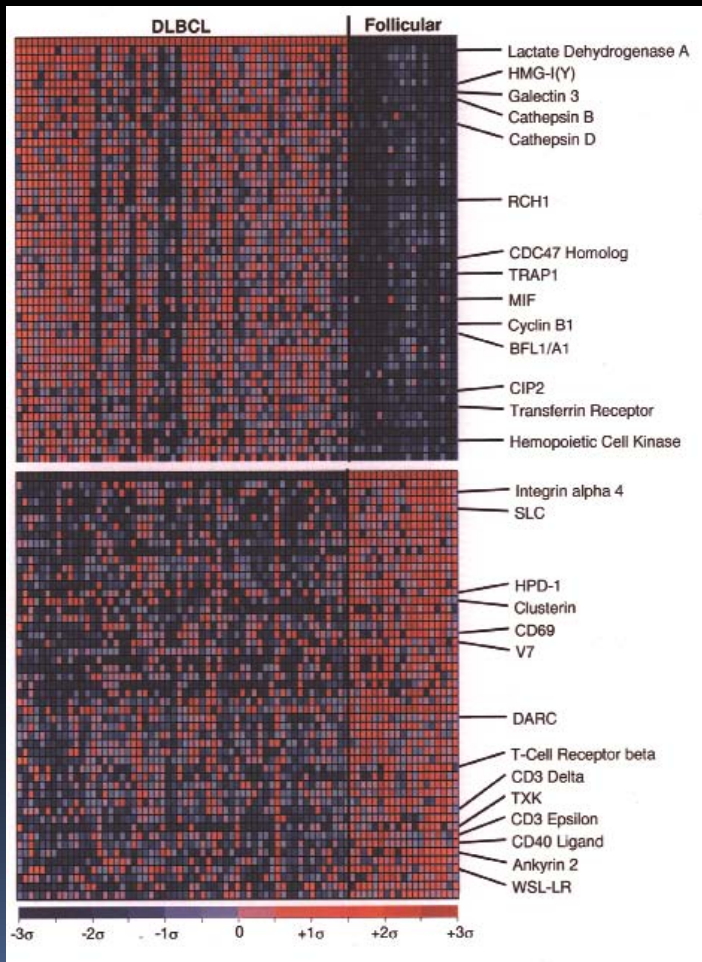
Integrating Taverna and caBIG

- Taverna workflow execution as a caGrid WSRF service
- *caTransfer* – supporting large data
- *CQL* building – customized data queries from within Taverna
 - *University of Chicago / Argonne National Laboratory*

Use case: Lymphoma prediction workflow



Use case: Lymphoma prediction workflow



SampleName	TrueClass	SVMPredClass	KNNPredClass
testSample_0	DLBCL	DLBCL	DLBCL
testSample_1	FL	FL	FL
testSample_2	DLBCL	DLBCL	DLBCL
testSample_3	FL	FL	FL
testSample_4	DLBCL	FL	DLBCL
testSample_5	DLBCL	DLBCL	DLBCL
testSample_6	FL	FL	FL
testSample_7	DLBCL	DLBCL	DLBCL
testSample_8	DLBCL	DLBCL	FL
testSample_9	DLBCL	DLBCL	DLBCL
testSample_10	DLBCL	DLBCL	DLBCL
testSample_11	FL	FL	FL
testSample_12	DLBCL	DLBCL	DLBCL
testSample_13	DLBCL	DLBCL	DLBCL
testSample_14	DLBCL	DLBCL	DLBCL
testSample_15	DLBCL	DLBCL	DLBCL
testSample_16	DLBCL	DLBCL	DLBCL
testSample_17	DLBCL	DLBCL	FL
testSample_18	DLBCL	DLBCL	DLBCL
testSample_19	FL	FL	FL
testSample_20	DLBCL	DLBCL	DLBCL
testSample_21	DLBCL	DLBCL	DLBCL
testSample_22	DLBCL	DLBCL	DLBCL
testSample_23	DLBCL	DLBCL	DLBCL
testSample_24	DLBCL	DLBCL	DLBCL
testSample_25	DLBCL	DLBCL	DLBCL

Taverna demonstration

- Monday 2009-12-07 17:40
- Thursday 2009-12-10 12:50
- **eScience North West Centre**
stand downstairs, booth 6



Questions..?