

Customizing Galaxy for Large Scale Services

Nikolay Vazov

University Center for Information Technologies University of Oslo



UiO **: University of Oslo**



What is a "large scale service"?

A service (a portal) which is:

- operating on national and international level (users)
- open to the entire world (not behind a firewall or on a dedicated network, intranet)
- sharing both the front-end (galaxy framework) and the back-end (e.g. cluster) resources
- sharing data storage resources

What is a "large scale service" setup?

Very much like *usegalaxy.org*, but ... requiring also at least:

1. Authentication & Authorization

- Who is allowed to log in
- What is the authentication procedure
- What tools the different users are allowed to run

2. Resource allocation management (esp. in a cluster setup)

- Projects and Project Managers vs Users
- Allocation of resources to projects
- Accounting for these resources
- Report generation per project basis



Authentication challenges

a) Built-in authentication procedure

OK, but impossible to filter the users and implement SSO

b) Local LDAP

OK, but hard to add external users

c) Trusted Identity Providers (Federations)

OK, but technically quite challenging

Identity Providers solution



Trusted Identity Providers

- a) National scale in Norway (FEIDE)
- b) Scandinavian scope (KALMAR)
- c) World scope (EDUGAIN)
- d) Facebook/Google, etc
- e) Integrated gateway : Dataporten



Trusted Identity Providers *implementation*

SAML 2.0 built into Galaxy

- 1 additional package only required (Galaxy independent)
- SP configuration full control and flexibility (Galaxy independent)
- many changes in the Galaxy code

Difficult to keep pace with Galaxy development

Example of the Lifeportal?

2

	HotSpot (00d 23h 53m 🗙 🚦 Lifeport	al × Elifeportal × 🕂					
5	A https://galaxy-test.uio.no/root		▼ C Google	Q	r 🏦	☆ 🗎	=
K	🛅 Most Visited 👻 🥮 Getting Started 🗟 Lat	est Headlines 🔻					
1	UiO I University of Oslo	Analyze Data Workflow Shared Data - V	isualization▼ Help▼ User▼			Using 0	bytes
	Lifeportal						
	Academic login	About the Lifeportal					
1	Authenticate with your Feide 📫 account.	Lifeportal gives you easy access to the High Performance Computin cluster Abel at the University of Oslo. The Galaxy based Lifeportal ha	g as a				
	Academic Login Login for other users	continuously growing list of services, and among them the most wide tools from the former Bioportal.	ely used				
5	Email address:	More about Lifeportal					
		Getting access					
	Password:	Getting access to the Lifeportal					
	Login Forgot password? Reset here						
8	Partners						
			Colorer				
		ioinformatics Platform	Jalaxy				
	>						<
	20						
(Swiss-German Galaxy Wo	rkshop, October 2016				
	The						

Example of the Lifeportal?

	HotSpot (00d 23h 56m 🗴 🚦 Lifeportal	× : Lifepo	ortal	× +									
F	🔶 🖴 https://galaxy-test. uio.no /root					▼ C	Google ▼ Google	Q	ŧ	⋒	☆	ê	≡
)=	👸 Most Visited 👻 🖲 Getting Started 🛛 Latest Headlines 👻												
0	UiO : University of Oslo	Analyze Data	Workflow	Shared Data -	Visualization -	Help U	lser v				Usin	ig 0 byt	es

Lifeportal

Academic login	About the Lifeportal
Authenticate with your eduGain 🛊 account.	Lifeportal gives you easy access to the High Performance Computing cluster Abel at the University of Oslo. The Galaxy based Lifeportal bas a
AAI@EduHr - Croatian Research and 🔺	continuously growing list of services, and among them the most widely used tools from the former Bioportal.
AAI@EduHr - Croatian Research and	More about Lifeportal
ACOnet	Getting access
AESIR	Getting access to the Lifeportal
ARNES	
Academy of Sciences Library Aristotle University of Thessaloniki	
Art Academy of Latvia - LMA IDP	
Athena - Research and Innovation	Norwegian Bioinformatics Platform
•	
×	

Trusted Identity Providers implementation

SAML 2.0 implemented outside the Galaxy code

- SP configuration out of the box
- Galaxy *absolutely* independent!
- A complex bundle of software involving compilation & dependencies :
 - * PHP (SimpleSamplPhP)
 - Memcached server + 2 C-libraries
 - PHP support for Memcached server
 - Apache module for memcached

Difficult to keep pace with the development of the software bundle

Example of Geoportal

remote_user = True

Select your identity provider for the GEOPORTAL

English | Bokmål | Nynorsk | Sámegiella | Dansk | Deutsch | Svenska | Suomeksi | Español | Français | Italiano | Nederlands | Luxembourgish | Czech | Slovenščina | Lietuvių kalba | Hrvatski | Magyar | Język polski | Português | Português brasileiro | Türkçe | 日本語 | 简体中文 | 繁體中文 | русский язык | eesti keel | עָבְרִית | Bahasa Indonesia | Srpski | Latviešu

Select

Select your identity provider for the GEOPORTAL

Please select the identity provider where you want to authenticate:

 Feide - Norwegian educational institutions (test-IdP)
 Image: Constraint of the second sec

Apply for access to Geoportal

Copyright © 2007-2010 Feide RnD





Trusted IdP (last developments)

OAuth 2.0 implemented outside/inside the Galaxy code – under development

Explored solutions depend on the existing libraries:

- **Apache driven** (the configuration in the Apache config files)
 - tiny but extremely cryptical solution, quite unflexible
- SimpleSamlPhp library
 - very difficult to resolve all PHP dependencies

• Python package

- outside Galaxy code requires a separate server/port to communicate with the oath2 gateway
- Inside Galaxy code needs rewriting and intrusion into Galaxy code
- mod_wsgi (in Apache) still under investigation

Identity Providers solution DATAPORTEN in geoaccessno portal

Choose a user account to login with



In order to login to **GeoAccessNo-test**, you need to select a login provider.



or login with another account



Slide by Trond Thorbjørnsen (uio)

Resource allocation challenges

The principles of resource allocation:

- All jobs on a shared cluster are run "on behalf of" an account (project)
- All projects receive an allocation of CPU hours. When the CPU hours in a project are exhausted, no jobs belonging to this project will be allowed to run on the cluster.
- All users must be members of at least one project in order to run jobs
- Every project has a project manager who manages the CPU allocations of the project and adds/removes users in the project

Resource allocation challenges

In order to implement this model we need:

- A resource allocation manager GOLD**
 - · Defines projects and users and assigns users to projects
 - · Allocates resources to projects
 - Reserves resources to jobs
 - · Charges the projects for used resources (per job)
 - Refunds the projects for reserved but unused resources

Interface from Galaxy to GOLD

GOLD – terminal & web

[root@galaxy-prod01 ~]# /opt/gold/bin/glsuser | grep added abegiz3@vahoo.com True anna.vader@unis.no Anna Vader anna.vader@unis.no True laurindodasilva@fas.harvard.edu True mgrabowski@email.gwu.edu True ovidiu.paun@univie.ac.at True ricivito@hotmail.it True

non-feide user added by magnus.popp@nhm.uio.no Manually added non-feide project manager - nikolaiv non-feide user added by torbjorn.ekrem@ntnu.no non-feide user added by t.f.hansen@ibv.uio.no non-feide user added by a.k.brysting@ibv.uio.no non-feide user added by geirksa@ifi.uio.no

✓ Sold allocation

*

File Edit View History Bookmarks Tools Help

: Lifeportal

💥 🔣 http://docs.ada...ting.com/gold/ 💥 🚺 List Users

A https://gold.galaxy-prod01.hpc.uio.no/index.cgi



Logout

List Users								
Search:	By:							
added	Description	0						

Name	Active CommonName		PhoneNumber	EmailAddress	DefaultProject	Organization	Description	
abegiz3@yahoo.com	True						non-feide user added by magnus.popp@nhm.uio.no	
mgrabowski@email.gwu.edu	True		5 2				non-feide user added by t.f.hansen@ibv.uio.no	
ovidiu.paun@univie.ac.at	True						non-feide user added by a.k.brysting@ibv.uio.no	
anna.vader@unis.no	True	Anna Vader	3 	anna.vader@unis.no			Manually added non-feide project manager - nikolaiv	
ricivito@hotmail.it	True						non-feide user added by geirksa@ifi.uio.no	
laurindodasilva@fas.harvard.edu	True		9 				non-feide user added by torbjorn.ekrem@ntnu.no	

List Users - Mozilla Firefox

₩ 4

Name	Active	CommonName	PhoneNumber	EmailAddress	DefaultProject	Organization	Description
@yahoo.com	True						non-feide user added by magnus.popp@nhm.uio.no
vski@email.gwu.edu	True						non-feide user added by t.f.hansen@ibv.uio.no
aun@univie.ac.at	True						non-feide user added by a.k.brysting@ibv.uio.no
der@unis.no	True	Anna Vader		anna.vader@unis.no			Manually added non-feide project manager - nikolaiv
Photmail.it	True						non-feide user added by geirksa@ifi.uio.no
dasilva@fas.harvard.edu	True		54				non-feide user added by torbiorn ekrem@ntnu.no

Project application form

۲				Lifeportal - Mozil	la Firefox			×
<u>F</u> ile	<u>E</u> dit <u>V</u> iew Hi <u>s</u> tory <u>I</u>	<u>B</u> ookmarks <u>T</u> ools <u>H</u> elp						
1 L	feportal	🗙 🕺 http://docs.ada	ting.com/gold/ 💥 🗍 🛄 L	st Users	※ (4)			
4	https://lifeportal.ui	o.no/project_application			. 20	습 🗸 🕻	Gold allocation	🏩 🦊 🖀
Ui	• University of Oslo	Analyze Data	Workflow Shared Data -	Visualization 🗸 Admin	ProjectAdmin R	Reports Apply for a project H	elp → User →	Using 37.4 KB

Lifeportal

200

Lifeportal project application form

Please fill in the required information

Personal information about the project leader

Project responsible, e.g. John Doe. (permitted chars: capital/small letters and blanks)*:
Job title/position (permitted chars: capital/small letters, digits and blanks):
E-mail address:

Phone no (please, use digits only, e.g. 0047XXX for Norway):

Cell-phone number (please, use digits only, e.g. 0047XXX for Norway)*

Institution (Faculty, Department) (permitted chars: capital/small letters and blanks)*. Country (permitted chars: capital/small letters and blanks)*.

Project information

Project name (permitted chars: capital/small letters and digits)*:	
CPU hours (specify how many CPU hours you need for the project) (permitted chars: digits)*:	
Preferred applications* (Click on the box to display the apps) :	BEAST Blast Gaussian Migrate
Project description (permitted chars: all)*:	
Start date*	YYYY-MM-DD
End date :	2014-09-30
I declare that the project does not contain sensitive data*:	
I have read the Lifeportal requirements and accepted them*	
Note: Fields marked * must be filled out.	
Send application form Cancel	



1		34	in.	١.	
۰.	κ.				
М	2				
		×.			

Project application form (new version)

) i) 🔒 https://nettskjema.uio.no/use	r/form/preview.html?id=66135	C Search	☆
	UiO I Universitetet i Oslo		
	Nettskjema Spørreskiemaer, påmeldinger og bestillinger	Hielo	English Nynorsk
	Forside Mine skjemaer Lifeportal project applicatio) form	· •
	Lifeportal project application fo	Apent for svar? Skjemaet er åpent Steng Avansert	Sist endret 18. januar 2016 12:56 av Nikolay Aleksandrov Vazov
	Vis Bygg skjema Innstillinger Rettighete	r Innhent svar Se resultater	
	Lifeportal project application f	orm	
	Personal information about the project leader		Side 1
	Project responsible, e.g. John Doe: *		
	Job title/position:		
	E-mail address: *		
	Phone:		
	Cell-phone number: *		



Project selection for job execution

					Lite	portai - Mozii	a Firero	x					*
<u>F</u> ile <u>E</u> dit <u>V</u> iew Hi <u>s</u> to	ry <u>B</u> ookmarks	<u>T</u> ools <u>H</u> elp											
🚦 Lifeportal	×)	🛿 http://docs.adating.com/gold/ 💥	List Use	ers	8	👔 🔋 Lifeportal		×	(4)				
https://galaxy	-test. uio.no /root										☆ ►	3 Gold allocation	🕺 👆 🖆
UiO E University of Os	slo	A	nalyze Data	Workflow	Shared Data -	Visualization -	Admin	ProjectAdmin	Reports	Help -	User -		Using 23.3 MB

Lifeportal

Tools	Abel handshake (version 1.0.0)				History	2 0
search tools Upload file		1stTESTafter 23.2 MB	<i>a</i> =			
Tests		<			352: Abel handshake	• / ×
<u>Abel handshake</u>	Projects/Accounts				2E1: Abal bandabaka	
Bioportal Phylogeny Tools	You have logged as user : vazovn@abv.bg. Please	e select the project (account) where you will be runn	ing your current job.		351. Aber handshake	30 / X
R	[Ip9 🗢	350: Abel handshake	• 0 %			
					349: Abel handshake	• 0 ×
Get Data	Number of tasks	Number of tasks per pode	Walltime (ich duration)	Memory per CBU		
Text Manipulation	Enter the number of tasks for the current ich	Enter the number of tasks per node (may 16)	Select how long the job shall be running	Select how much memory you need for the job	348: Abel handshake	@ (/ X
Filter and Sort	(Enter 1, if not sure.)	(Leave 1, if not sure.)	delectriowoolg the job shart be running .	(the allowed format is digit+6b/Hb,	<u>347: 1.fasta</u>	• 0 %
Join, Subtract and Group				e.g. 16GB or 3000Mb):	246: shal data tyt	@ / W
Convert Formats	1	1	DD: 00 HH: 2 MM: 00 SC: 00	4Gb	340. aber-uata.txt	
Extract Features					345: Abel handshake	• 0 %
Fetch Sequences	Execute				344: Abel handshake	• 0 %
Operate on Genomic Intervals						- 0.00
Graph/Display Data	About the tool This tool tests the connection to	the Lifeportal computational backend Abel.			343: Dumper.py	× / ×
FASTA manipulation	Abel home page: http://www.uio.no/hpc				<u>342: 1.fasta</u>	• 0 %
NGS: QC and manipulation					341: Abel handshake	@ / %
NGS: Mapping						
NGS: SAM Tools					<u>340: 1.fasta</u>	@ 0 XX
Workflows					339: Abel handshake	• 0 %
 <u>All workflows</u> 					<u>338: 1.fasta</u>	• 0 %
					<u>337: 1.fasta</u>	• 0 %
					<u>336: lifeportal-</u> leaf-wide1-60x30.jpg	• 0 %
					<u>335: 1.fasta</u>	• 0 %
					334: Abel handshake	• 0 %
					333: Dumper.py	• 0 %

Project selection for job execution (in job_conf.xml)

	Specify job resource parameters
	Processors (threads)
	Number of processing cores, 'ppn' value (1-16). Leave blank to use default value.
	Memory
	Memory size in gigabytes, 'pmem' value (1–900). Leave blank to use default value.
	Time
	Maximum job time in hours, 'walltime' value (1-672) or 28 days. Leave blank to use default value.
	Project
	Default Geoportal 200hrs project
-	Project to assign resource allocation to. Leave blank to use default value.
ſ	- 4 Evocuto
	✓ Execute

Project management

<u>F</u> ile <u>E</u> dit <u>V</u> iew Higtory <u>B</u> ookmarks <u>T</u> ools <u>H</u> elp	
😫 Lifeportal 🗙 😰 http://docs.adating.com/gold/ 💥 📋 List Users 🛛 💥 🗇	
🛊 🖻 https://lifeportal.uio.no/project_admin	💼 🦊 🏙
UiO # University of Oslo Analyze Data Workflow Shared Data - Visualization - Admin ProjectAdmin Reports Apply for a project. Help - User -	Using 37.4 KB

Lifeportal

5

oject Administration	Lifeportal Project Admin	Lifeportal Project Administration								
sers <u>Manage users</u>	Lifeportal projects list sorted by pro	Elfeportal projects list sorted by project name Full Lifeportal local projects list Attention! The Amount of CPU hours below will only be displayed from Start date. The value will be 0.00 before Start and after End								
ojects Manage projects Show pending projects Show rejected projects										
	Project owner	Project name	Users	Active	Amount	Description	Start	End		
	David.Rees@bio.uib.no	Ip20	MEMBERS David.Rees@bio.uib.no	True	155.50	Barnacles	2014-02-01	2014-09-30		
	 Trond.Oskars@student.uib.no 	lp6	MEMBERS Trond.Oskars@student.uib.no	True	0.00	PhylogenyCephalaspidea	2013-10-21	2013-11-05		
	Trond.Oskars@um.uib.no	lp34	MEMBERS Trond.Oskars@um.uib.no	True	941.40	AdditionalAnalysesCephas	2014-05-22	2014-09-30		
	○ a.k.brysting@ibv.uio.no	lp13	MEMBERS a.k.brysting@ibv.uio.no ovidiu.paun@univie.ac.at I.n.dsen@bv.uio.no m.k.brandrud@ibv.uio.no mariekbr@student.ibv.uio.no mildries@student.ibv.uio.no mildries@student.ibv.uio.no	True	8958.28	Cochlearia	2013-12-03	2014-09-30		
	a.k.krabberod@ibv.uio.no	lp2	MEMBERS a.k.krabberod@ibv.uio.no jon.brate@ibv.uio.no	True	42281.51	PhyRad	2013-10-17	2014-09-30		
	 a.k.kristoffersen@odont.uio.no 	lp27	MEMBERS a.k.kristoffersen@odont.uio.no	True	211.77	IKM16samplicon2014	2014-04-10	2014-09-30		
	🔿 anna.vader@unis.no	lp17	anna.vader@unis.no	True	1000.00	unis - Manually Added	2014-01-30	2014-09-30		
	 arthurah@student.ibv.uio.no 	Ip38	MEMBERS arthurah@student.ibv.uio.no	True	20000.00	Xcell	2014-06-23	2014-09-30		
	g.o.s.thomassen@usit.uio.no	lp15	MEMBERS g.o.s.thomassen@usit.uio.no	True	50.00	rektest	2014-01-21	2014-09-30		
	O g.o.s.thomassen@usit.uio.no	lp16	MEMBERS g.o.s.thomassen@usit.uio.no	True	50.00	rkftest21	2014-01-21	2014-09-30		
	g.o.s.thomassen@usit.uio.no	lp19	MEMBERS g.o.s.thomassen@usit.uio.no	True	10.00	test11	2014-02-10	2014-09-30		
	a.o.s.thomassen@usit.uio.no	lp5	MEMBERS	False	10.00	fttest01_				

< ~



Project manager interface

Galaxy integrated vs. GOLD vs. "home made"

