Services for Life Sciences

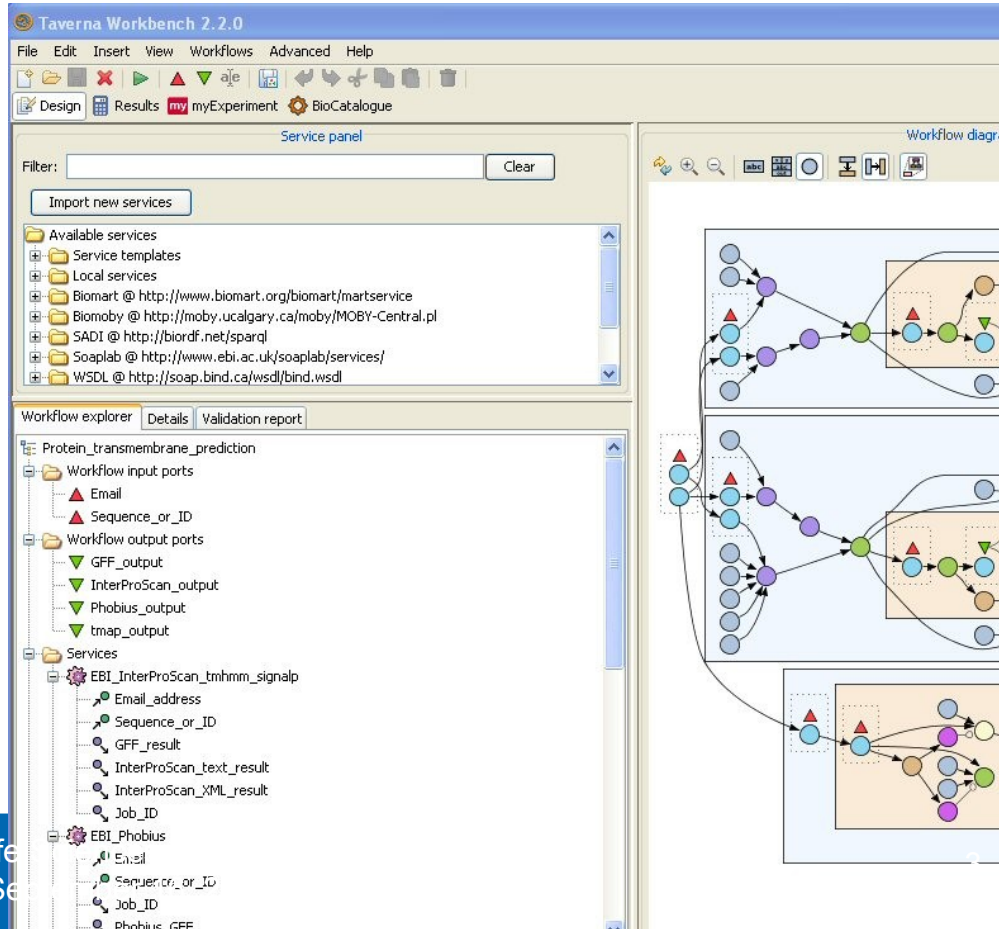
Johan Montagnat (CNRS)
Yannick Legré (HealthGrid)
Peter Rice (EBI)

LS HUG tasks

- Workfows (TSA3.2.4)
  - EBI
- CoreBio services (TSA 3.4)
  - EBI
- Life Sciences VRC (TSA 3.4)
  - CNRS / HealthGrid

Taverna Workbench

- Workbench for scientific workflows
  - Original developer: Tom Oinn, EBI
  - Workflow client for UK e-Science myGrid project
- Open source (Lesser GPL)
  - Linux, Windows, Mac
  - Maintained by OMII-UK
- Web services, grid services, REST, local tools, etc
  - Service APIs plugins
  - 3000 public services
- Support for BioCatalogue registry



Taverna current status

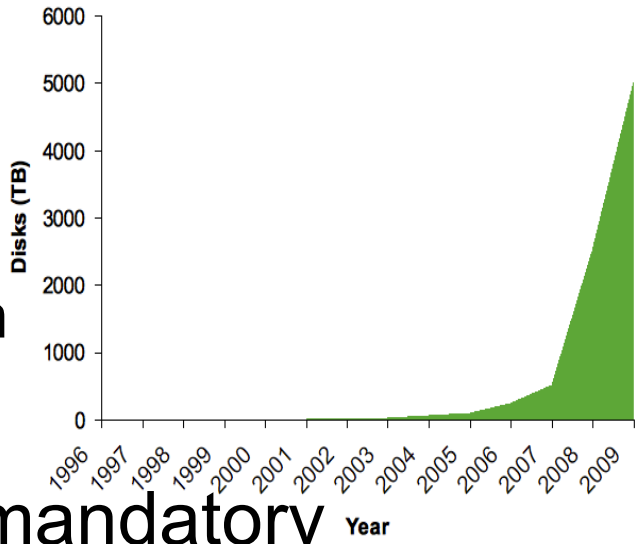
- Taverna 2.0 released 2008 (currently v2.2)
  - Taverna 1.7 remains in use
  - Scalability and extensibility improved
- Taverna 3.0 expected 2011
  - New plugin architecture
  - OSGi platform
  - Alpha release November 2011
- EBI will adapt Taverna to the evolution of EBI services interface
  - e.g. as defined by EMBRACE NoE (REST services)

CoreBio DBs - ELIXIR

- Life Sciences ESFRI project
  - Coordinated by EBI, **www.elixir-europe.org**
  - 32 partners, 13 countries
- To build a sustainable European infrastructure for biological information supporting life science research and its translation to medicine, the environment, the bioindustries, society.
- ELIXIR is formally seeking input from organisations that are interested in hosting one of its nodes.

CoreBio DBs

- Management of biological DBs hosted by ELIXIR
  - Face biological DBs exponential growth
- Example: the European Nucleotide Archive (all DNA sequence data)
  - Currently around 100TB
  - Doubling every 12 months
  - New sequences are 10GB-1 TB each



- Replication on computing sites is mandatory
  - Data transfer is too costly

Life Sciences VRC

- Part of LS VRC operation model implementation
  - Governance model
  - VO management tools
  - VO monitoring
- Governance model
  - See this morning's presentation
  - Wiki site: http://wiki.healthgrid.org/LSVRC:Biomed
  - Document: https://dav.healthgrid.org/lsvrc/LSVRC_proposition_09-08-2010-final.pdf

VO management tools

- Users database design
  - Relational schema defined
  - Relations to VOMS and application DB entities
- User database implementation
  - MySQL server + PHP front-end
  - VOMS and application DB interface to be studied

VO monitoring tools

- Nagios server planed
- Hudson integration server dashboard used

All | Biomed VO Storage Elements | Biomed VO general tests

S	W	Job ↓	Last Success	Last Failure	Last Duration
		All Storage Elements	6 mo 0 days (#6)	6 mo 0 days (#5)	4 hr 26 min
		Failure	N/A	6 mo 1 day (#1)	0.35 sec
		Init jobs	3 days 19 hr (#67)	5 mo 18 days (#21)	2 min 33 sec
		LFC response time	1 hr 10 min (#45)	N/A	0.94 sec
		Proxy infos	1 hr 10 min (#17)	N/A	0.85 sec
		Proxy init	1 hr 10 min (#16)	N/A	4.7 sec
		SE agh15.atlas.unimelb.edu.au	3 mo 11 days (#29)	N/A	4.6 sec
		SE agh3.atlas.unimelb.edu.au	1 mo 2 days (#28)	1 mo 1 day (#29)	2 min 42 sec
		SE aliserv1.ct.infn.it	2 mo 13 days (#29)	2 mo 15 days (#23)	3.7 sec
		SE axon-g05.ieeta.pt	1 hr 10 min (#16)	N/A	1 min 57 sec

Infrastructure tests

- VO-level monitoring
  - Using proxies from users within the VO
  - Two different DNs to identify different problems
- Services monitoring
  - VOMS (proxy generation), LFC
  - Sites information
  - SEs
  - More to be done

Technical team

- Technical team on duty
  - Weekly shifts among participants
- Monitoring activity
  - GGUS tickets submission and follow-up
    - From 35 to 17 tickets / month
  - SE decommissioning
    - Get list of files, associated users, notify and follow-up

Start date	End date	Team on shift
13/09/2010	24/09/2010	UPV
27/09/2010	1/10/2010	ISC-PIF
4/10/2010	8/10/2010	LPC
11/10/2010	22/10/2010	INFN-BA
25/10/2010	5/11/2010	I3S
8/11/2010	12/11/2010	IPHC
15/11/2010	26/11/2010	Creatis
29/11/2010	10/12/2010	UPV
13/11/2010	17/12/2010	ISC-PIF
20/12/2010	24/12/2010	LPC
27/12/2010	31/12/2010	No shift this week
3/01/2011	14/01/2011	INFN-BA
17/01/2011	28/01/2011	I3S
31/01/2011	4/02/2011	IPHC
7/02/2011	18/02/2011	Creatis

Conclusions

- Services for biological data
  - Taverna workflow system and BioCatalogue
  - EBI / ELIXIR databases replication
- Support for the LS VRC implementation
  - VRC users management
  - VO-level monitoring
- Tooling need to be improved for VRC
  - See requirement talk
  - Better information, OLAs, data migration procedures... all non-VRC specific