



# The ER-flow project and Astronomical Workflows

\*\*\*

## Current status and future perspectives

*On behalf of the Astro Community in ER-flow and of  
the AstroCG in OGF*

C. Vuerli, G. Taffoni, G. Castelli, F. Pasian, U. Becciani, A. Costa, E. Sciacca  
INAF – Istituto Nazionale di Astrofisica



# Outline



- The ER-flow project
  - Project overview
- The Astro Community in ER-flow
  - Applications porting (WP5)
  - Data sharing / interoperability (WP4)
  - Training & dissemination (WP2)
  - Established collaborations
- Future perspectives
  - Key Concepts
  - The STARnet federation
  - Empowering the synergy between IVOA and projects, alliances, federations related to DCIs and Clouds
  - Astronomical projects currently in place and the new opportunities in Horizon 2020



# The ER-flow project objectives



- **Community building**
  - **1:** To build a European community of workflow developers and users
- **Workflow interoperability & dissemination**
  - **2:** To migrate workflow based scientific applications of the supported research communities to the European Grid Infrastructure through the SHIWA Simulation Platform
  - **3:** To disseminate the workflow interoperability solution of the SHIWA project among the selected research communities and identify further research communities that need the simulation platform to run their experiments.
- **Interoperability of the scientific data in the workflow domain**
  - **4:** To define requirements of the supported research communities on interoperability of the scientific data in the workflow domain identifying existing and missing protocols and standards needed to support this interoperability.
  - **5:** To write a study on the interoperability of the scientific data in the workflow domain.



# The ER-flow project



- Web site: <http://www.erflow.eu>
- Start: September 1<sup>st</sup>, 2012
- End: August 31<sup>st</sup>, 2014
- Scientific Communities
  - Earth Science
  - Computational Chemistry
  - Heliophysics
  - Astrophysics



# The ER-flow project

- Helps to build a collaborative European Research community
  - Training and dissemination play a key role
- Fosters the reuse of applications within and beyond scientific collaborations
- Technologies:
  - *SHIWA repository*: a database of workflows and related metadata
  - *SHIWA web portal* integrated in the workflows repository and with various workflow engines
  - *SHIWA desktop*: access via the user's computer
  - Various *workflow engines* already integrated with the SSP



# Why Astro in ER-flow

- Modern astro experiments and projects require tight collaborations and imply complex computing models
  - Huge amount of data
    - to process → DCIs
    - to archive and share → Data repositories, data preservation
  - Worldwide applications sharing, reuse and customization
    - Application repositories
    - Smart tools to access them
  - Data interoperability in an application sharing scenario



# Astro Community in ER-flow



## 1<sup>st</sup> Year

- INAF – Istituto Nazionale di Astrofisica (IT)
  - Leads the astro participation to ER-flow
  - Three INAF Institutes involved
    - Osservatorio Astronomico di Trieste
    - Osservatorio Astrofisico di Catania
    - Osservatorio Astronomico di Teramo
- University of Portsmouth (UK)
- Slovak Academy of Sciences (SK)



# Astro Community in ER-flow



2<sup>nd</sup> Year: Participation extended to:

- IAA Granada (ES)

Contacts established with:

- CDS / Observatories of Strasbourg (FR)
- Observatory of Meudon Paris (FR)
- IFCA Santander (ES)





# Participation in ER-flow WPs



- Astro participation meant to provide man power in:
  - WP2 (training and dissemination)
  - WP4 (data sharing and interoperability)
  - WP5 (application porting)
- Collaboration with WP3 to fix some technical problems impacting the successful porting of astro applications on SSP
  - Less effort in WP5
  - Priority to workflows generated through different WF systems



# WP5: First Year

- Six astro applications ported on the SSP during the first year
- Workows coded in WS-PGRADE/gUSE (Web Services Parallel Grid Runtime and Developer Environment/Grid User Support Environment) architecture on SSP; some of them are meta-workows (i.e. composition of smaller workows, that could eventually be the building blocks for the creation of new meta-workows).
- Resources belonging to 4 Grid Virtual Organizations:
  - astro.vo.eu-egee.org (European astronomical catch-all VO)
  - inaf (Italian national VO for astronomy)
  - planck (VO dedicated to the ESA Planck satellite mission)
  - VOCE (Virtual Organisation for Central Europe)
- Use of gLite as Grid middlewere.



# WP5: Second Year



- During Y2 the A&A community collaborated in the consolidation of the infrastructure:
  - A common SSO designed, i.e. same credentials to access different SGWs
  - Distributed filesystem under evaluation: data shared between SGWs
  - Enhancement of the SGWs functionalities



# WP5: Second Year



- New usage scenario motivated by:
  - The great amount of legacy software produced by the community
  - The tools and services developed within the VObs to access and share data
  - The large amount of astro workflows which use Tarevrna and Astro-Taverna able to manipulate data using the VObs services and tools
  - Workflows as the result of a cooperation between astronomers and computer scientists.
    - Some of them are building blocks for more complex operations



# WP5: Second Year



- Goals:
  - Astro Taverna WFs stored in the SHIWA repository
  - Astro Taverna WFs tested in the SSP
  - Production of more complex workflows (meta-workflows) starting from simple (building blocks) WFs
- Steps:
  - WFs executed in their native WF management system
  - WFs uploaded in the SHIWA repository (executables, metadata, docs, samples and test data)
  - External links refer these workflows
  - New users can import WFs in their local execution environment.
    - The STARnet federation offers a set of community portals used to test and run WFs



# WP4: Data interoperability



- Tasks
  - T4.1: Virtual Data Object specification
  - T4.2: Data sets transfer
  - T4.3: Data generation and error recovery
  - **T4.4**: Data semantics and WF specification
- Completed Deliverables & Milestones
  - D4.1 (Virtual Data Objects specification) at M8
  - MS4.1 (Data objects transfer service) at M12
  - D4.2 (Study of VDOs generation and error recovery) at M18
- Coming soon
  - D4.3 (Study of domain semantic data and workflow description, M22) (work in progress)
  - MS4.2: Interoperability recommendations (M24)



# WP2: Training / dissemination



- 1<sup>st</sup> Year: Dissemination events organized in the context of the applications gathering campaign
  - Presentation of the SHIWA technology and of the SSP to a wide audience
  - Restricted meetings with potential application providers
  - Porting activity planning

Date	Description
25-29 Nov 2012	University of Portsmouth (UK) – LaSMoG application
11-13 Dec 2012	SAS Bratislava – COMCAPT and MESTREAM applications
7-9 Jan 2013	INAF Teramo – FRANEC/BaSTI application
11-15 Feb 2013	INAF Catania – Workshop for workflow developers and end users
17-23 Feb 2013	University of Portsmouth (UK) – New applications for the 2 <sup>nd</sup> year
18-19 Mar 2013	CDS Strasbourg – VObs – ER-flow collaboration; applications for the 2 <sup>nd</sup> year
20-21 Mar 2013	OBSPM Paris – applications for the 2 <sup>nd</sup> year
11-17 May 2013	IVOA interoperability Heidelberg – Presentation of the project to IVOA members
16-17 Jul 2013	IAA Granada – Workflows and applications for the 2 <sup>nd</sup> year



# WP2: Training / dissemination



## 2<sup>nd</sup> Year

- VO&Friends training event (Catania, 17 - 19 December 2013)
  - Project presentation
  - Hands-on session (deploy and test of WF)
- STARnet meeting
  - Project presentation
- IVOA interoperability meeting Spring 2014
  - Project presentation





# WP2: Training / dissemination



## Publications:

- A&A overview on application porting (JGC)
- Special issue on PASP
- ADASS 2014
- ER-flow Communities paper  
– *In preparation*



# Established collaborations



- European level partners:
  - STARnet (<http://www.oact.inaf.it/STARnet/>)
  - IVOA : International Virtual Observatory Alliance (<http://www.ivoa.net/>)
- National level partners (in Italy):
  - VObs.it : Italian Virtual Observatory (<http://vobs.astro.it/>)
  - IA2 : National Astronomical Archive Center (<http://ia2.oats.inaf.it/>)



# The future

## Key Concepts

- 1) The strategic importance of workflows in astronomy could increase in the forthcoming years and become a fundamental tool for the scientific production, but this requires powerful distributed infrastructures able to:
  - 1.1) Combine workflows coming from different workflow systems
  - 1.2) Access an adequate amount of resources (hardware, software, data)
  - 1.3) Provide smart tools to end users to access and use such kind of resources (e.g. Science Gateways)
- 2) The data sharing / data interoperability in distributed workflow environments built on top of infrastructures as depicted above could become crucial in the next years



# The STARnet federation



- Conceived at INAF Catania (U. Becciani)
- Science Gateway European Federation
- Liferay, gUSE/WS-PGRADE
- Local clusters + DCIs + Clouds
- ER-flow + SCIBUS Spin Off
- MoUs subscribed by federating partners



# STAR<sup>net</sup> Gateway Federation Definition



## Technology.

- Liferay/WS-PGRADE, gUse, SHIWA Repository
- Local and distributed DCIs

## Maintenance.

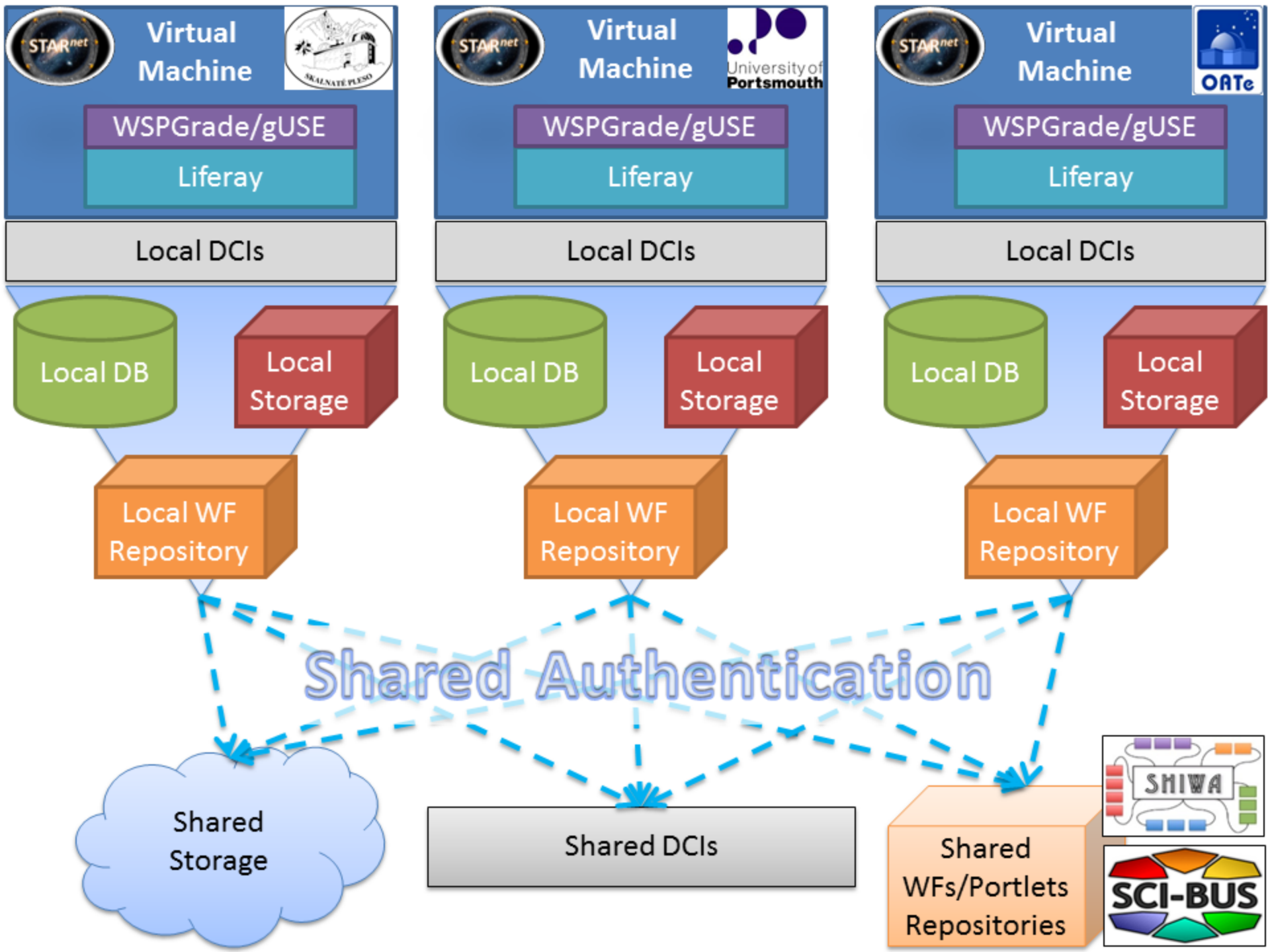
- FrontEnd/BackEnd Virtual Machines
- Master Virtual Machines with Local Customization included (configuration file Enab/Disab. Portlets and services)
- Master maintenance/update → INAF (OACt)

## Shared Services

- Single Sign On (SSO)
- Workflows Sharing (SHIWA repository)
- Cloud Data

## Local account

- Data will be preserved: each federated SG mounts external DB exported by the physical machine (e.g. /mnt/STARnet)





# IVOA, DCIs/Clouds, H2020



- On the IVOA side:
  - The GWS Working Group
- On the DCIs / Clouds side
  - EGI
  - The Astro CG (Community Group) in OGF
  - PRACE
- ...and a number of big Astro Projects
  - Via Lactea
  - SKA and the Radio Community
  - CTA and the HE Astrophysics Community
  - Euclid and the Space Community



# IVOA, DCIs/Clouds, H2020



- New opportunities will be open by Horizon 2020
  - Many H2020 project proposals could be conceived in the framework of the big projects previously mentioned
  - But also alliances and federations like IVOA and STARnet could play the role of promoters / stakeholders of project proposals in H2020
  - Some key questions about H2020 should get an answer in the coming months





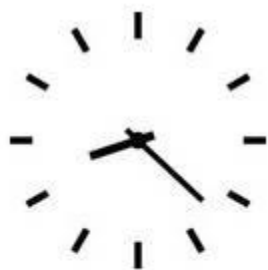
# Some considerations

- A centralized coordination in the landscape that is going to emerge seems to be really hard
- Nevertheless some kind of lightweight coordination for astronomy is desirable, aiming at:
  - Carrying out technological studies concerning the exposed key concepts 1 [1,1, 1.2, 1.3] an 2
  - Monitoring the emerging technologies
  - Producing feasibility studies, technological reports, technological recommendations, papers, etc...to be distributed worldwide
  - Promoting project proposals in response to the H2020 calls trying to involve the big astro projects as project partners



# Some considerations

- Lightweight coordination for astronomy jointly led by data players and DCIs / Clouds players
- On the data side
  - IVOA
- On the DCIs / Clouds side
  - The AstroCG @ OGF ?
  - The Astro Community @ EGI ?
- And what about HPC and PRACE ?
- This liaison role is already recognized to GWS@IVOA on one side and to AstroCG@OGF on the other side but the liaison activity is currently very limited
  - Shall we attempt a relaunch if this activity in the new (and challenging) emerging scenario ?



Q & A time



© Presentation-Process.com