

Curation and Characterization of Web Services

Jose Enrique Ruiz
jer@iaa.es

October 23rd 2012
2012 IVOA Fall Interop Meeting - Sao Paolo



Wf4Ever

Advanced Workflow Preservation Technologies for Enhanced Science



1. Intelligent Software Components (ISOCO, Spain)
2. University of Manchester (UNIMAN, UK)
3. Universidad Politécnica de Madrid (UPM, Spain)
4. Poznan Supercomputing and Networking Center (PSNC, Poland)
5. University of Oxford and OeRC (OXF, UK)
6. Instituto Astrofísica Andalucía (IAA, Spain)
7. Leiden University Medical Center (LUMC, Netherlands)

Web-services based
Workflows



The next generation of archives

Much wider FoV and spectral coverage

- Huge sized datasets (~ tens TB)
- Big Data science highly dependent on I/O data rates
- Subproducts as **virtual data** generated on-the-fly

We are moving into a world where

- **computing and storage are cheap**
- **data movement is death**

The next generation of archives

Much wider FoV and spectral coverage

- Huge sized datasets (~ tens TB)
- Big Data science highly dependent on I/O data rates
- Subproducts as **virtual data** generated on-the-fly

The **move computing to data** paradigm

Archives should evolve from data providers into **services providers**, where web services may help to solve bandwidth issues.

The next generation of archives

Much wider FoV and spectral coverage

- Huge sized datasets (~ tens TB)
- Big Data science highly dependent on I/O data rates
- Subproducts as **virtual data** generated on-the-fly

Data Discovery

Data Access

Data Management

The next generation of archives

Much wider FoV and spectral coverage

- Huge sized datasets (~ tens TB)
- Big Data science highly dependent on I/O data rates
- Subproducts as **virtual data** generated on-the-fly

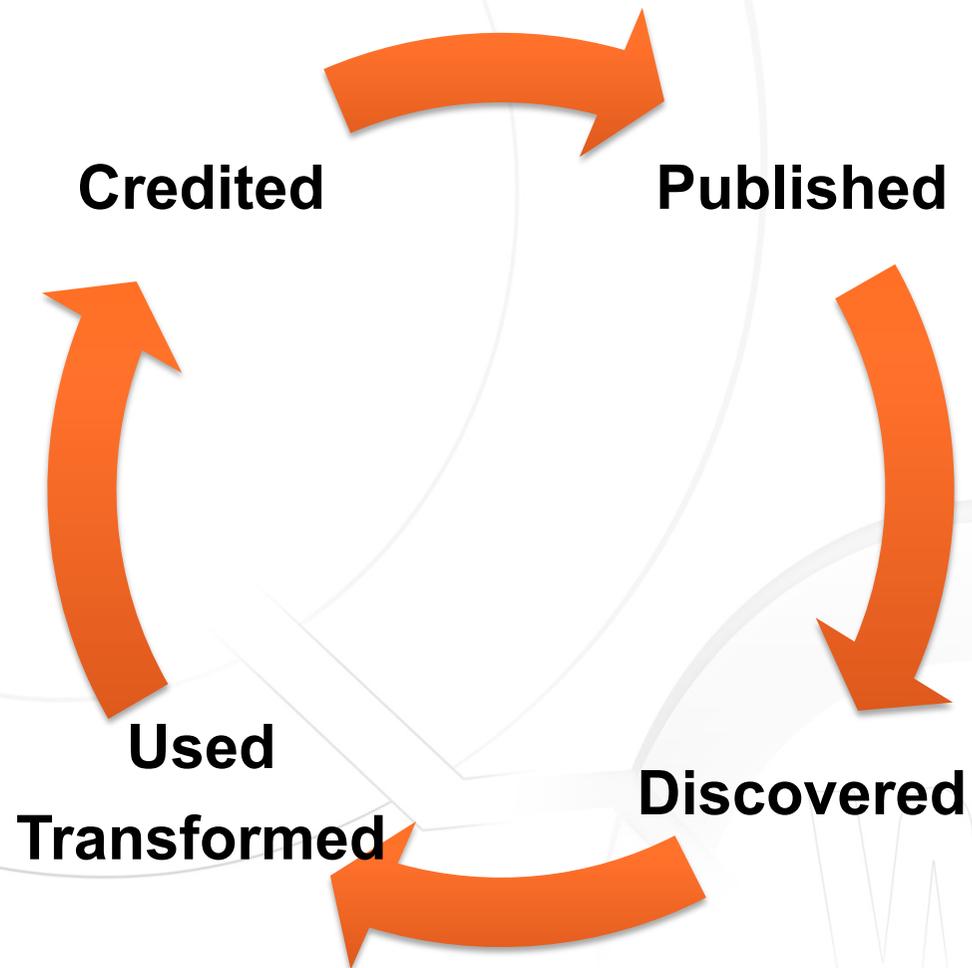
Web Services Discovery

Web Services Access

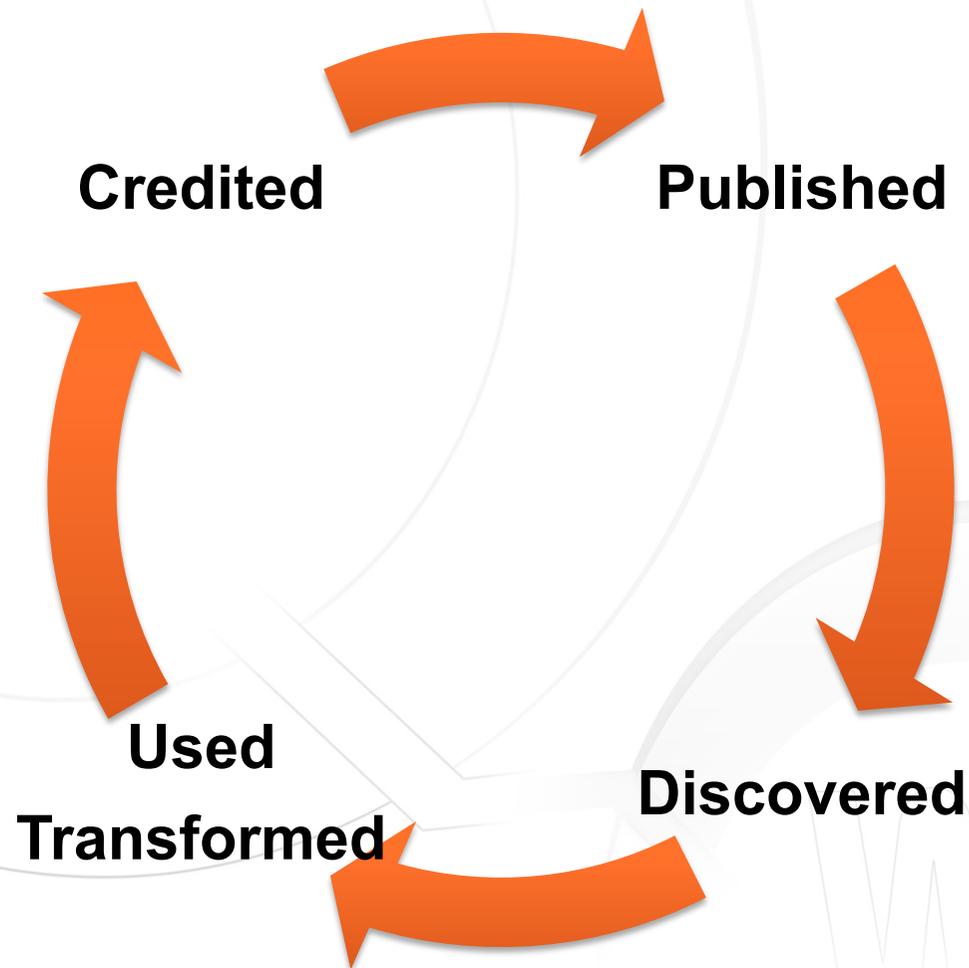
Web Services Management

The Web-Services
Deluge

The Data Lifecycle



The **Service** Lifecycle



Published

- The VO Registry
- Easier to publish services than datasets in the VO ?
- WS are not exclusive property of big data archives

- Publication is not Preservation
- Backup strategies
- Replication/Mirrors
- Versioning

- Software Publishing Platforms

github

runmycode

The screenshot displays the 'Publishing Registry Portal' interface. At the top, there are navigation links: 'New Publisher', 'How to Publish', 'Search the VO Registry', and 'Search the Catalogue Registry'. The main heading is 'Register a New Publisher'. Below this, there is a brief instruction: 'Use this form to describe your organisation, which will serve as the publisher (or Naming Authority) for your resources. When this form is submitted, it will create two resource descriptions in your workspace: an "Organisation" and an "Authority".' A legend indicates that fields with a red asterisk are required, and those with a green asterisk are recommended. The form includes fields for 'Title', 'Identifier' (with a value of 'twc//'), 'Short Name', 'Status' (with radio buttons for 'Active', 'Inactive', and 'Deleted'), and 'Password'. A 'Resource Curation' section contains fields for 'Publisher', 'Publisher Ivo-id', 'Creator', 'Name', 'Name Ivo-id', and 'Logo'. On the left side, there is a sidebar menu for 'The Euro-VO projects: VOTECH EuroVO-DCA EuroVO-AIDA'. The menu items include 'EURO-VO Registry', 'Search Resources', 'Insert Resources', 'Resource', 'Organisation', 'Authority', 'Data Collection', 'Service', 'Registry', 'Table Service', 'Data Service', 'Catalog Service', 'Cone Search (CS)', 'Open Sky Node (OSN)', 'Simple Image Access (SIAP)', 'Proto Spectral Access (PSAP)', 'Simple Spectral Access (SSAP)', 'Simple Line Access (SLAP)', and 'Theoretical Spectral Access (TSAP)'. The 'Service' item is currently selected.

Discovered

- **Search Criteria**

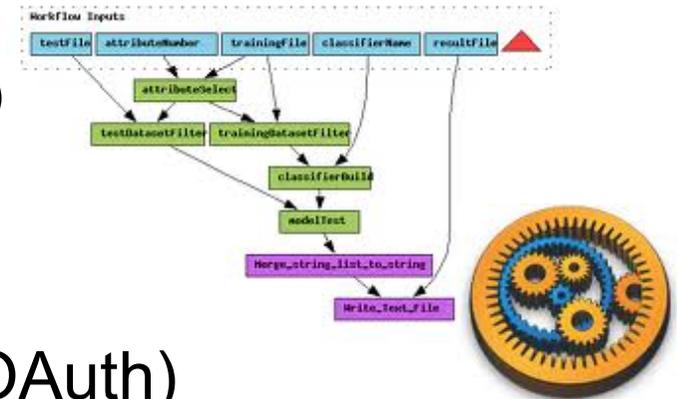
- Relevant Keywords (Semantics)
- Authoring - Institution, Archive
- Waveband, Science
- Function-based
 - VO Services mainly focused on Data Discovery and Access (DAL)
 - Wrapped Legacy Apps and Data Processing (SIAv2, Theory IG)
 - KDD IG
- Input/Output Data (TAP, UTypes, VOSI #tables)
- Access Policy (Authentication – SSO, OAuth)
- A-Synchrony (SOAP, REST) and Stage Data (VOSpace)
- Allocation of CPU/Storage, Estimated Computing Time



- **Proposition of alternatives and similars**

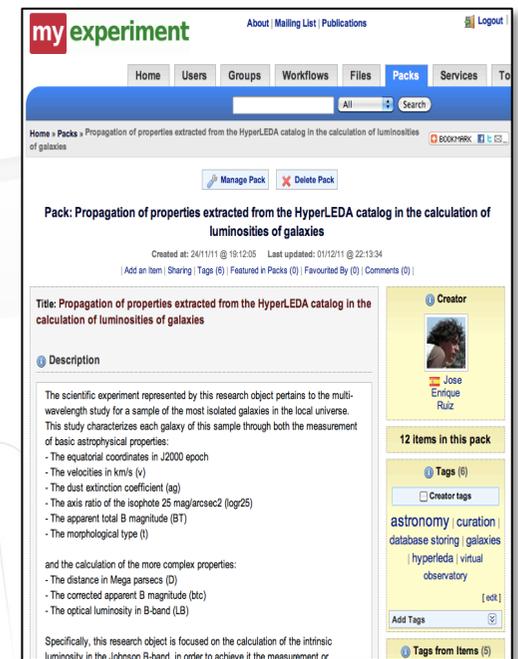
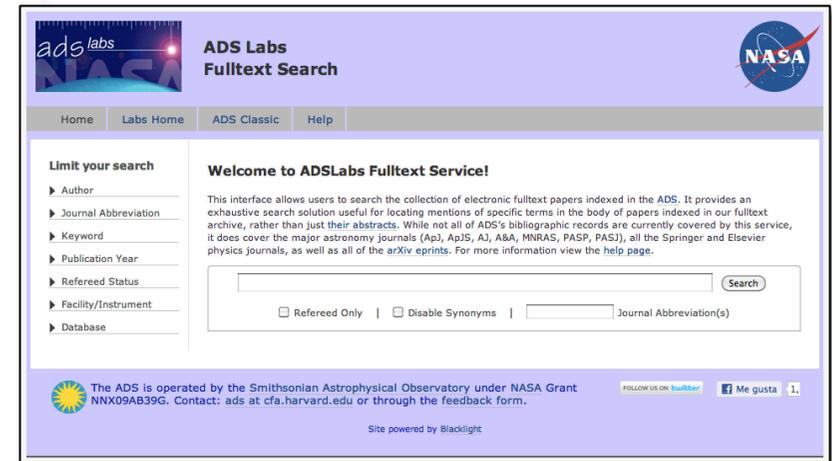
Used and Transformed

- **How to use them ?** (WADL, WSDL – VOSI #capabilities)
 - Input Data -> Parameters needed and formats
 - Self-described WS (PDL, S3, SimDAL, SimDB)
 - Output Data -> Response format - TAP
 - Example Data, Self-Consistency Checking
- Access Policy (**Authentication** – SSO, OAuth)
- **WS orchestration in Workflows** (Data-flow vs. Control-flow)
- How the **community** uses WS ?
- Propositions based on patterns of statistical use or popularity
- **Provenance** of the methods is Wf-evolution by re-use
- Consumed by Humans and Machines - **Interoperable** (WS-I)



Credited

- **Linked to related Artefacts**
 - Data Facilities and Archives
 - Authors, ACSL Software, Wfs
- **Quality Assessment**
 - Technical and scientific
 - Penalize abandoned and award the maintained
- **Automate Monitoring (VOSI #availability)**
 - Decay
 - Performance, WS Analytics
 - Change of interfaces, permissions, etc.
- **Community Curation**
 - Blogging
 - Recommendation
 - Folksonomy



In a cloud of web services and data..

Web Services should benefit of the same privileges acquired by Data until now.

Start thinking on how to provide

- **Detailed curation**
- **Thorough characterization**

