

# Provenance Support for Content Management Systems: A Drupal Example



**Aída Gándara and Paulo Pinheiro da Silva**

agandara1@miners.utep.edu, paulo@utep.edu  
The University of Texas at El Paso

## Motivation

The usefulness of provenance is greatly affected by its accessibility. Isolated environments limit the sharing of information. Web portals differ in how they function and often require manual steps for sharing information. Our solution is to build a structure into content management servers so tools can share and access needed information, data and provenance-related data; moving information from isolated environments without scientists having to access one specific portal or understand the nuances of different portal's interface.

## CI-Server on Drupal

The CI-Server framework on Drupal extends Drupal functionality to support the sharing and reuse of scientific documents. CI-Server extends the functionality within common content management systems by leveraging the semantic information in published documents to help build more knowledge. CI-Server has three primary goals:

1. to enable information sharing by providing tools that scientists can use within their scientific research to process data, publish and share artifacts
2. to build community by providing tools that support building and viewing discussions between scientists about artifacts used or created through scientific processes
3. to leverage the knowledge collected within the artifacts and scientific collaborations to support scientific discoveries.

## CI-Client API

Client tools use the CI-Client API to integrate with the Drupal server to publish and reuse information at the server. By providing this functionality within the tools known by the scientists, CI-Server is better prepared to serve the collaborative needs of scientific teams. Furthermore, the interface to CI-Server systems is generic, so scientists do not need to know the culture of each Web portal in order to share and reuse information. Tools that currently leverage the CI-Client API:

- the PML Data Annotator
- WDO-It! Semantic Abstract Workflow System
- CI Desktop

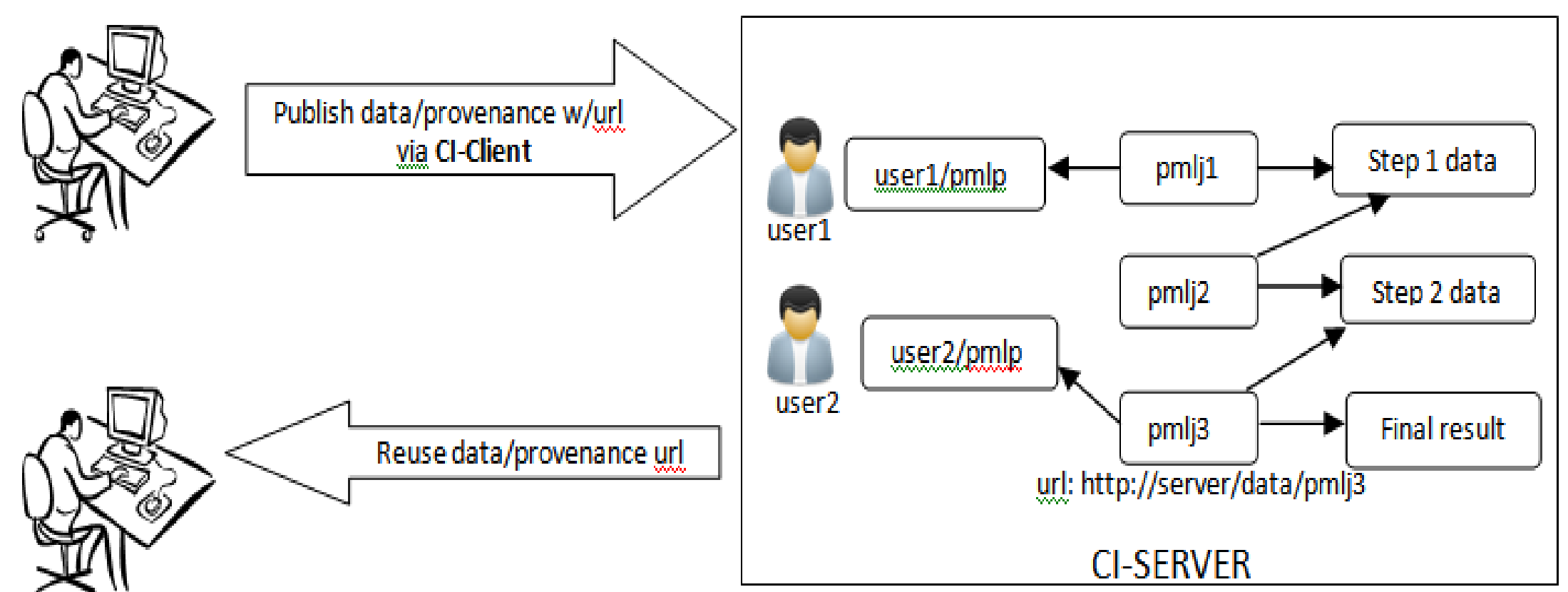
As a result, other tools are able to immediately leverage the published data via a URL.

## Provenance Support

The figure illustrates provenance support enabled by the CI-Server framework. Via the CI-Client API, the top scientist has published pmlj1, pmlj2, pmlj3, together called a PML nodeset, and corresponding data; moving data from the scientist's workstation to a Drupal-based CI-Server. The pmlj documents are semantic documents, written in OWL, built with knowledge about how a scientific process occurred. pmlj documents rely on links to available resources, e.g. the corresponding data published on the CI-Server.

The CI-Server uses Drupal modules to support file management and url aliasing, enabling users to upload content and then access it via url links. Because pmlj3 is identified with a URL, the scientist reusing the information can use the url link to access the entire provenance nodeset once it is published on the CI-Server.

Furthermore, the CI-Server is able to dynamically build content from content it manages. For example, the CI-Server dynamically builds pmlp nodes, source related provenance, for users of the system. Building provenance dynamically with internal CI-Server knowledge aids in the collection of provenance.



## Future Work

- Social Provenance – enhancing the collection of Drupal comments to support the management of scientific discussions and relating them to scientific data.
- Building Collective Knowledge – leveraging the semantic knowledge published in semantically annotated documents as well as other information managed by the CI-Server to support scientific discoveries.
- Building CI-Servers – deploying the CI-Server framework to other environments to open up the CI-Server knowledge base accessible to CI-Client based tools.
- Capturing Scientific Processes – enhance additional scientific workflow tools (e.g. Kepler Workflow System) to support publishing additional process related content on a CI-Server.

## References

1. Gruber, T. 2008. Collective knowledge systems: Where the Social Web meets the Semantic Web. *Web Semant.* 6, 1 (Feb. 2008), 4-13.
2. McGuiness, D. et al. PML2: A Modular Explanation Interlingua. In Proc of the AAAI 2007 Workshop on Explanation-aware Computing, Vancouver, British Columbia, Canada, July 2007.



This work was supported in part by DHS grant 2008-ST-062-000007 and by NSF grant HRD-0734825.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the supports.

