Instruction for the CENSAM Data Provenance System

Last Edited: 2012-11-05

Juebo Wu
Department of Geography, National University of Singapore

Chih-Yuan Chen,
Department of Geography, National University of Singapore
Contents

Table of Contents

Contents ......................................................................................................................... 2

1 Overview ...................................................................................................................... 5

2 Web Service Management .......................................................................................... 7
   2.1 Web Service category ........................................................................................... 7
   2.2 Web Services Management ................................................................................. 9
   2.3 Test the Web Service .......................................................................................... 11

3 Role Changing ............................................................................................................ 13

4 Workflow Creation ..................................................................................................... 14
   4.1 Select Operations .............................................................................................. 15
   4.2 Add Date Source ............................................................................................... 16
   4.3 Preview the workflow ....................................................................................... 17
   4.4 Execute the workflow ....................................................................................... 18
   4.5 Export the workflow ....................................................................................... 20
   4.6 Add Annotations .............................................................................................. 21

5 Provenance management ............................................................................................ 22
   5.1 Search Provenance ........................................................................................... 22
   5.3 Searching provenance by Web Service Operation ............................................. 25
   5.4 Modify the Existing Workflow ........................................................................... 25
Figures

Figure 1 The framework of CCI Data Provenance System ................................................................. 6
Figure 2 Web Service Management .................................................................................................. 7
Figure 3 Web Service category .......................................................................................................... 7
Figure 4 Create Category window .................................................................................................... 7
Figure 5 Modify a category ............................................................................................................... 8
Figure 6 Delete a category ................................................................................................................ 8
Figure 7 Web Service Management interface .................................................................................. 9
Figure 8 Web services descriptions .................................................................................................. 9
Figure 9 Delete the web service ........................................................................................................ 9
Figure 10 publish a web service ...................................................................................................... 10
Figure 11 Test the Web Service ....................................................................................................... 11
Figure 12 Input the parameters for testing the web service ............................................................... 11
Figure 13 Test result of the web service ........................................................................................... 12
Figure 14 Change the Role ............................................................................................................... 13
Figure 15 Information of the user and research project .................................................................. 13
Figure 16 The graphic Interface for the multiple pipeline workflow ............................................... 14
Figure 17 Choose Web Service category .......................................................................................... 15
Figure 18 Choose Web Service ........................................................................................................ 15
Figure 19 Choose operation ............................................................................................................. 15
Figure 20 Choose the Data Resource .............................................................................................. 16
Figure 21 Data source windows ....................................................................................................... 16
Figure 22 Automatically set the Data Source .................................................................................... 17
Figure 23 A sample of the OPM graph ............................................................................................. 17
Figure 24 Execute the workflow ...................................................................................................... 18
Figure 25 Result window of the executed operations ....................................................................... 18
Figure 26 Skip the operations .......................................................................................................... 19
Figure 27 OPM graph for the workflow .......................................................... 19
Figure 28 Export OPM dot file format .......................................................... 20
Figure 29 Export OPM OWL file format ......................................................... 20
Figure 30 Annotations with the key-value coding convention ......................... 21
Figure 31 Add new Annotation ....................................................................... 21
Figure 32 Provenance management ............................................................... 22
Figure 33 Provenance searching interface ...................................................... 22
Figure 34 Advanced searching interface ....................................................... 23
Figure 35 Check Provenance detail ............................................................... 23
Figure 36 Workflow information ................................................................... 24
Figure 37 The annotation result window ....................................................... 24
Figure 38 Searching provenance by Web Service Operation ......................... 25
Figure 39 Workflow modification .................................................................. 26
Figure 40 Rearrange operations .................................................................... 26
Figure 41 Add new Operation into existing workflow .................................... 27
Figure 42 Execute the modified workflow ..................................................... 27
1 Overview

The CENSAM Data Provenance System is aiming to record and share the provenance information related to the activities of data analysis and processing in CENSAM Cyberinfrastructure system (CCI). The CCI has the advantages for allowing users from different domains of specialty and different purposes of research to register and share their data in a centralized environment, and to use a unified CENSAM ontology to keep the data interoperability. To further utilize those advantages for not only sharing the data but also sharing the analytical process and knowledge, we developed the CENSAM Data Provenance System in order to achieve this goal. Four service modules have been developed in our prototype system including Web Service management, role changing, workflow creation and Provenance management as we described below:

1. **Web service management** is developed for the Web Service publishing, testing and maintenance, where CCI user can either connect to external Web Services follow our Web Service protocol standard, or use the internal Web Services provided by the CCI system. The main functions in this module contain category catalog, service publishing, operation testing, and keyword searching for the management purposes.

2. **Role changing** is designed for workflow creation purpose to record the information about user and the research project related to the analytical data analysis or processing. The role is also used in the Open Provenance Model (OPM) as an element of provenance.

3. **Workflow Creation** is provided by connecting CCI, CDR, the first two service modules of the Data Provenance System, and a graphic user interface for the multiple pipeline workflows. All the data in the registered CDR and all the processing functions in the registered Web Services can be used for creating new workflow. The provenance information will be generated automatically while the user is running the workflow process. Detail information including the structure of the workflow, the data sources, parameters of the used functions, and the user role are also recorded and restored into our provenance repository. The user can add the annotation or tag to the workflow manually the key-value coding convention.
4. **Provenance management** is composed of four individual functions including provenance searching, browsing, OPM results export, and workflow modification, to utilize the provenances information. These function modules help the user share and reuse the data provenance.

The framework of CCI Data Provenance System is shown on figure below. We will introduce how to use the Data Provenance System started from the next section.

![Figure 1 The framework of CCI Data Provenance System](image)
2 Web Service Management

For access to the Web Service Management, user can go to “Data Provenance” top menu and click the “Manage Web Service” sub menu.

Figure 2 Web Service Management

2.1 Web Service category

The Main interface shows a catalog of Web Service category including the title, description, and creator of the Web Service:

Figure 3 Web Service category

User can create a new category by clicking “Create Category” and edit the information, or modify and delete the category information as well by clicking respective function buttons.

Figure 4 Create Category window
Click “Modify” to change the information of the category.

Figure 5 Modify a category

Click “Delete” to delete the category.

Figure 6 Delete a category
2.2 Web Services Management

Click one of titles will go into Web Service Management interface.

![Figure 7 Web Service Management interface](image)

For the selected category, all registered web services will be shown on this page with descriptions.

![Figure 8 Web services descriptions](image)

User can delete the web service by clicking “Delete” button. Because the web service publisher defines the service detail, no modification for service is allowed here.

![Figure 9 Delete the web service](image)
For creating a new service, click the “Create Service” button. There are two steps to finish the process of service publishing:

Step 1 is to select a service category for the new service.

Step 2 is to publish the detail information of this new service. Our Web Service Catalog is using the Web Services Description Language (WSDL) as the standard protocol to access the web services. User can publish the new service either by typing the WSDL schema URL address (Method 1) or uploading the wsdl.xml file (Method 2). Only one method should be selected for publishing a new service.

![Publish Web Service](image)

**Figure 10 publish a web service**

After publish a new web service, the detail information of the web service will show on the page including WSDL schema URL address, the name of the operations and its description. All of the information are provided by the service provider.
2.3 Test the Web Service

On the web service page, you can test the individual operation of the service by clicking the “test” button and input the parameters for the specified web service and confirm.
If the web service is available and invoked correctly, the result will be prompt later. The results window will show a text results in XML format. And if the output data is a geographical data and registered in the CCI, an openlayer window and user interface will show on the map result window.

Figure 13 Test result of the web service
3 Role Changing

Go to Web Service unit by click the menu “Change Role”.

Figure 14 Change the Role

Put the information about the user for workflow creation. This function is just a prototype to record the user information because the CCI system doesn't have the access control function.

Figure 15 Information of the user and research project
4 Workflow Creation

After defining the user information, the system will jump to the Workflow Creation page. **Workflow Creation** is composed of single pipeline and multiple pipelines. By default, the page is single pipeline definition after choosing new workflow. The user can decide which operations should be combined into this workflow. The user can compile a multiple pipeline workflow by our prototype user interface. The only difference between the single pipeline and multiple pipeline workflows is the graphic interface. We will only explain the service module using the single pipeline workflow. For the detail of multiple pipeline workflow please see our video demo and sample flow chart on CENSAM website:

Video: http://www.censam.org/presentation/vediodemoofmultiplepipelineworkflow
Sample flow chart: http://www.censam.org/presentation/multiplepipelineflowchart

![Figure 16 The graphic Interface fro the multiple pipeline workflow]
4.1 Select Operations

To select operations, the user should carry out three steps: (1) Choose Web Service category; (2) Choose Web Service; (3) Choose operations.

(1) Choose Web Service category

![Figure 17 Choose Web Service category](image)

(2) Choose Web Service

![Figure 18 Choose Web Service](image)

(3) Choose operation

![Figure 19 Choose operation](image)

Click “Add” to add the operation to this process after choosing the operation. And click “Del” to delete one operation.
4.2 Add Date Source

If the operation needs the data source that has been published in CCI system, the user can click “Choose” button to complete this step. The main data source choose interface:

![Figure 20 Choose the Data Resource](image)

Click “add” to add the data source and save the chosen data source, and the data source information of candidate operation will be updated automatically.

![Figure 21 Data source windows](image)
4.3 Preview the workflow

Click “View workflow” to generate the OPM graph of workflow from the chosen operations.
4.4 Execute the workflow

After operations are chosen, click “Execute” to carry out the chosen process. The process will be performed step by step.

![Execute Processing](image)

Figure 24 Execute the workflow

After each step has been executed, user can click “Check” button to see the result is correct.

![Result window of the executed operations](image)

Figure 25 Result window of the executed operations

If the operation has been executed and the result is reusable, then user can click “Skip” to skip the executed operation in order to avoid the duplication.
Figure 26 Skip the operations

When the process is completely performed, the provenance information will be created. A OPM graph will show after the workflow is executed:

Figure 27 OPM graph for the workflow
4.5 Export the workflow

Click “OPM DOT” to generate OPM dot file according to OPM specification.

Click “OPM OWL” to generate OPM owl file according to OPM specification.
4.6 Add Annotations

Through “Add annotations to this operation” panel bar, user can add some comments and annotations to this process. The comment format is following the key-value coding convention.

![Annotations with the key-value coding convention](image)

Figure 30 Annotations with the key-value coding convention

Annotations editing interface:

![Add new Annotation](image)

Figure 31 Add new Annotation
5 Provenance management

They CCI Provenance management system provides provenance searching, browsing, exporting, and workflow modification. For using the system, go to the “Data Provenance” menu and click the sub menu “Manage Provenance”.

5.1 Search Provenance

After entering the search keywords, click “Search” to retrieve the provenances that have been done before.
Click “Advanced” link to advanced searching interface, where provides more approaches for retrieving provenance.

![Advanced searching interface](image1)

**Figure 34 Advanced searching interface**

### 5.2 Check Provenance detail

For checking the Provenance detail, use can click one of items specified by time and go into the detail of this provenance record. The Provenance detail will show as below:

![Check Provenance detail](image2)

**Figure 35 Check Provenance detail**
User can click “Annotations to this operation” panel bar to view the annotation information.
5.3 Searching provenance by Web Service Operation

Our prototype System provides another way to search for the provenance and workflow by Web Service Operation. Use can simply click the link on individual Web Service operation on the provenance detail page and system will automatically retrieve the related provenance data, which also has the same Web Service operation.

![Diagram](Figure 38 Searching provenance by Web Service Operation)

5.4 Modify the Existing Workflow

Modifying or reproducing the existing workflow provides the reusability and the knowledge sharing of the CCI system. The user can carry out a new workflow by using the provenance information others have done before. And different users can also duplicate the same process, change the parameters, add or remove operations etc. to built up their own workflow to fit different purposes of research. To run the modification of the existing workflow, user will need to click “Modify Workflow” button in provenance detail page to enter reproducing page.
The user can remove, add or reorder the operations in our interface designed for this purpose. All parameters of the operations will be set to the same value as the previous workflow by default.
In this interface user can also add new operations by clicking “Add Operation” and selecting the web service.

**Figure 41 Add new Operation into existing workflow**

After modifying the workflow, click “Start” to the candidate operations page, here the user can choose the new data source from different data sources or just use the same setting as the previous workflow.

**Figure 42 Execute the modified workflow**

Once the new workflow is completed, just click the “Execute” button, and the process will be the same as section 4.4.