



# DataUp Web Application User's Guide



Copyright © 2013 by The Regents of the University of California. All rights reserved.

This manual is licensed under the Creative Commons Attribution-Share Alike 3.0 United States License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/3.0/us/> or send a letter to Creative Commons, 171 Second Street, Suite 300, San Francisco, California, 94105, USA.

## Version History

Version	Date	Notes
1.0	2/15/13	Initial Draft
1.1	2/25/13	First release
1.2	5/3/13	Minor edits

## About the DataUp Project

In today's technologically advanced world, the data generated by researchers is increasingly born digital and subject to intensive transformation and analyses before publication. The various file formats, software, and hardware required to succeed in the modern research landscape can become daunting, especially since education about digital data management has not kept pace with these technological advancements. There is a significant gap between the data management skills needed by modern researchers and their current abilities; the gap is more noticeable given the current increase in funder requirements for data management plans and requests by journal publishers to make supporting data publicly available alongside traditional research articles.

The DataUp project was born out of this need for seamless integration of data management into the researchers' current workflows. We recognized that the large majority of Earth, environmental, and ecological scientists use spreadsheets in the course of their data collection and organization; rather than requiring they learn a new program, we decided to meet them where they already work: Microsoft® Excel and related spreadsheet software.

The DataUp project's goals were to facilitate data management, sharing, and archiving for scientists. The resulting tools from the DataUp project are part of the Investigator Toolkit for DataONE, an NSF DataNet project building cyber-infrastructure that links together existing archives of ecological and environmental scientific data. The OuterCurve Foundation holds the code's copyright and DataUp is part of their Research Accelerators Gallery.

The DataUp project is run through the University of California Curation Center (UC3) at the California Digital Library, a division of the University of California Office of the President. Project funders were the Gordon and Betty Moore Foundation and Microsoft Research Connections.

## Terminology Used

This manual relies on the following terminology:

- A *workbook* is an XSLX or XSL file that is the native format for Microsoft Excel.
- A *data set* is the data in a workbook.
- Each workbook consists of one or more *worksheets*. Individual worksheets are accessed through the tabs at the bottom of the Excel screen.
- An *entity* is a set of related data within a data set. To be compliant with archival Best Practices, a data set must contain only column labels in the first row and individual data points in each subsequent row. A workbook that contains a data set with multiple entities has each entity in a separate worksheet.

## Contents

1	Workflow Outline .....	4
2	Accessing the Web Application .....	4
2.1	Third-Party Authentication .....	4
2.2	Signing Into the DataUp Web Application .....	5
2.3	Selecting a File to Upload.....	7
3	Check for Best Practices .....	8
4	Creating Metadata.....	10
4.1	Entering General Metadata .....	10
4.2	Entering Column Metadata.....	13
5	Creating Citations.....	14
6	Posting to the Repository .....	17
6.1	Check your data .....	17

## 1 Workflow Outline

The DataUp web application takes you through the process of documenting, managing, and archiving your data from start to finish. Subsequent sections in this guide provide detailed information on each step of the workflow identified below.



## 2 Accessing the Web Application

### 2.1 Third-Party Authentication

In place of user names and passwords, the DataUp web application uses third-party authentication to manage logins. This has a number of benefits:

- The application does not store passwords, eliminating a potential security issue.

- DataUp logins are greatly simplified. If you are already logged into the authentication provider, you can easily log into DataUp.
- If you already have a provider account, you don't have to remember an additional password.

You do not need to create the provider account in advance. Registration and sign-in can be done as part of the DataUp register/sign-in process. Currently, you can sign in with the following providers (both free services):

- **Facebook** ([www.facebook.com](http://www.facebook.com)).
- **Windows Live ID**. This service was recently renamed "Microsoft Account". ([account.live.com](http://account.live.com))

**Note:** Windows Live ID allows users to associate an existing email address with their Windows Live ID account.

**Note:** If you're using the DataUp web application on a shared computer, be sure to logout from both DataUp and the provider when finished. To logout from Facebook, go to [www.facebook.com](http://www.facebook.com). To logout from Windows Live ID, go to [account.live.com](http://account.live.com).

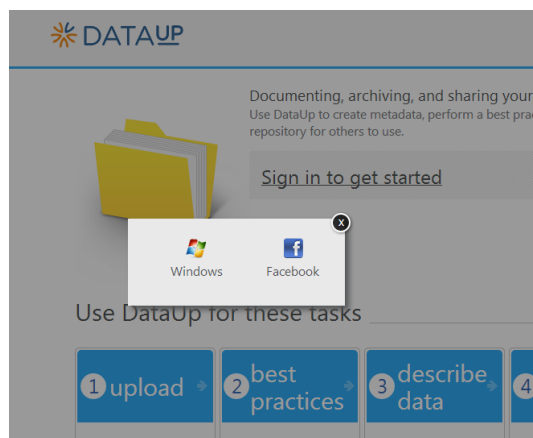
## 2.2 Signing Into the DataUp Web Application

The web application provides a single, simple access procedure for both new and existing users.

1. Point your web browser to [dataup.org](http://dataup.org). The DataUp web application sign in page appears.

**Note:** Do not confuse the DataUp web application at [dataup.org](http://dataup.org) with the DataUp project web site at [dataup.cdlib.org](http://dataup.cdlib.org).

2. Click on the "Sign in to get started" link. A list of third-party authentication providers appears (Figure 1).

**Figure 1: Third Party Authentication Provider Choices**

3. Choose an authentication provider. This can be a provider where you already have an account, or you can create a provider account the first time you log in.
4. If you are not currently logged into the provider, their login screen appears. (If you are already logged into the provider, ignore this step.) If you have an account with the provider, sign in. If you don't have an account with the provider, look for a link that allows you to create one.

Some providers (such as Facebook) will ask your permission to share your personal information with the DataUp web application. Only the specified information is shared, and is only used to pre-fill the DataUp registration form.

5. If this is your first time on DataUp using this provider, a registration page appears (Figure 2). Fill in the form and click the **Done** button. If you've already registered for DataUp using this provider, ignore this step.

**Figure 2: Registration Page**

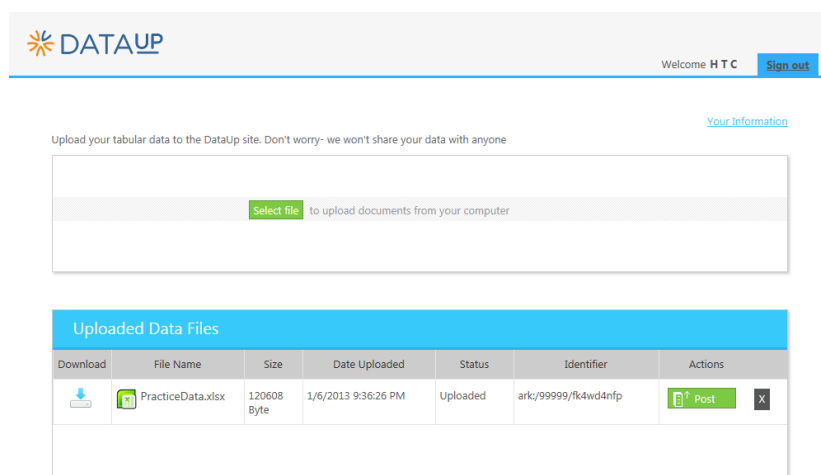
Register with DataUp  
You only need to do this once. We will not send you unwanted emails or sell your information.

**Your Information**

First Name	<input type="text"/>	Middle Name	<input type="text"/>
Last Name	<input type="text"/>	Organization	<input type="text"/>
Email	<input type="text"/>		

After you complete these steps, the main DataUp screen appears. See Figure 3 below. Note: The first time you log into DataUp, there will be no files in the "Uploaded Data Files" section.

Figure 3: Main DataUp Page



## 2.3 Selecting a File to Upload

On the main DataUp page, click **Select file** to choose a file from your computer to upload. You can also use the drag-and-drop feature if your browser supports it.

Choose whether to upload the file in CSV or XLSX format. The application recommends that all files be treated as CSV documents – this is the most archive-friendly format.

**Note:** If the file you have selected has already been uploaded to your repository, a confirmation message appears to confirm you want to overwrite the existing file.

Once the file has been successfully uploaded, the DataUp web application automatically moves to the next step in the workflow and checks the data for best practices (see Section 3). The uploaded file will now appear in your “Uploaded Data Files” list the next time you log into the application.

If you have already uploaded a data file to the DataUp Web Application, it will appear in the “Uploaded Data Files” section of the main page. To restart the DataUp workflow, you need to select “Post” next to the file name under the “Actions” column.

At any point in the application workflow, you can use the **Save** button to save a draft of the information you’ve entered, and complete the workflow at a later time. If you are on the metadata step, you will not be allowed to save unless all of the required fields have been completed (see Section 4). When you return to the web application to continue the posting process, select the **Post** button at the main DataUp page next to the file name.

## 3 Check for Best Practices

The DataUp web application is designed to support archival “Best Practices” for spreadsheets. This helps you store data that can be retrieved and reused. Best Practices errors can cause data to be misinterpreted, misunderstood, or difficult to discover. This section describes the Best Practices errors detected by the web application. The DataUp web application explains why these potential errors should be avoided and suggests possible fixes.



Table 1: Best Practices Checked for by DataUp Application

Potential problem	Why this may be a problem	Suggested remedy
<b>Embedded charts, tables, pictures</b>	These embedded items will not be visible when data are exported as a .csv file. Also, these elements are visible only if the file is opened with Microsoft Excel.	Move embedded charts, tables, or pictures to other tabs in your file or a completely separate file
<b>Embedded comments</b>	Comments will not be visible when data are exported as a .csv file. Also, these elements are visible only if the file is opened with Microsoft Excel.	Create a new column titled "Comments" and add your text there
<b>Commas</b>	Commas are often used to separate multiple piece of information/data (e.g. City, State). Cells should have only one piece of information.	Split pieces of information into multiple columns (e.g. City column and State column)
<b>Special characters</b>	Special characters (\$, %, ~, {, $\Sigma$ , $\partial$ , $\zeta$ , etc.) may cause problems for other programs or may be modified upon export	Use alphanumeric characters only. If needed, describe the symbol in a new column
<b>Color coded text or cell shading</b>	Formatting will not be visible when data are exported as a .csv file. If formatting is used as a coding scheme, all codes will be lost upon export.	Use descriptions or alphanumeric coding schemes in a new column
<b>Columns have mixed data types</b>	Some programs cannot handle mixed data types (e.g. numbers and text in the same column).	Ensure you are using only numbers or only text in a column; split data into multiple columns if necessary
<b>Non-contiguous data</b>	Empty columns or rows tend to be used to separate multiple data tables on the same tab.	Move multiple tables onto separate tabs.
<b>Merged cells</b>	Merged cells will not be maintained when data are exported as a .csv file. Information may be lost when cells are unmerged upon export.	Unmerge cells and annotate appropriately so information is not lost.
<b>Blank cells</b>	Blank cells within a contiguous data table are potentially problematic for reading files in other programs.	Designate a coding scheme for missing data or other explanations for blank cells.
<b>Header row absent or more than one header row</b>	Ideally the first row of a data table contains parameter names for the columns. If there is no header row, your data table may be difficult to use and document. If there are multiple header rows, some software programs may have problems.	Create a header row with unique parameter names that describe the columns' contents.
<b>Multiple sheets (tabs)</b>	Each tab will be exported as a separate .csv file.	

## 4 Creating Metadata

Metadata is key to archival Best Practices. Metadata can be described as data about data. Classical examples include library catalogs and inventories. Examples in modern technology include the technical and descriptive information digital cameras embed in graphic files and the information (owner, access permissions, creation and modification date/time values) that computer file systems record for all online files.

Metadata used by the DataUp project is based on the Ecological Metadata Language (EML). For further information, see the EML web site: [knb.ecoinformatics.org/software/eml](http://knb.ecoinformatics.org/software/eml).

The DataUp Add-in in assists you in creating metadata that describes your data. Two kinds of metadata are supported:

- *General metadata*, which describes the whole collection of data. This can be entered using the “Data Description” Tab in the web application.
- *Column metadata*, which provides individual descriptions for each column. This can be entered using the “Column Description” Tab in the web application.

### 4.1 Entering General Metadata

Below is a list of the metadata elements available for describing data via the DataUp web application. Only those with an asterisk (\*) are required fields.

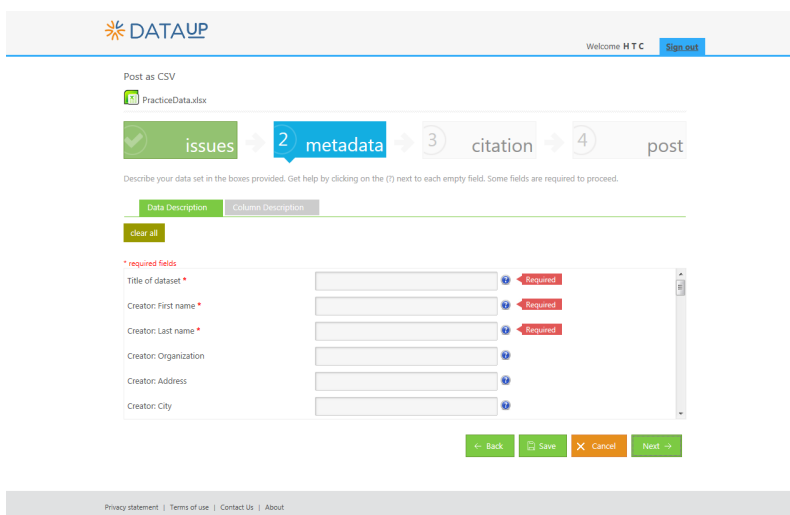
Name of metadata element	Description
Title of dataset*	
Creator: First name*	The creator is the primary data collector/creator.
Creator: Last name*	
Creator: Organization	
Creator: Address	
Creator: City	
Creator: State/province	
Creator: Postal code	
Creator: Country	
Creator: Phone	
Creator: Email*	
Today's date*	The date of metadata creation
Abstract*	Abstract describing the project for which the data were collected and any other miscellaneous pertinent information needed to understand the dataset.
Keyword(s)*	Keywords used to describe the dataset and enable discoverability. Ideally, these would be chosen from a controlled vocabulary (i.e., keyword thesaurus).

<b>Name of metadata element</b>	<b>Description</b>
Keyword thesaurus used	Keyword thesaurus (i.e., controlled vocabulary) used to describe the dataset.
Intellectual rights	
URL for data	If the data were obtained in part from a website, you can add the URL here.
Geographic coverage: Description	Brief description of the dataset's geographic coverage.
Geographic coverage: West bounding coordinate	Value between -180 and +180
Geographic coverage: East bounding coordinate	Value between -180 and +180
Geographic coverage: North bounding coordinate	Value between -90 and +90
Geographic coverage: South bounding coordinate	Value between -90 and +90
Temporal coverage: Beginning date	First date of data collection/creation (YYYY-MM-DD)
Temporal coverage: Ending date	Last date of data collection/creation (YYYY-MM-DD)
Data Contact Person: First name	The data contact person is the best person to contact with questions or concerns about this dataset
Data Contact Person: Last name	
Data Contact Person: Organization	
Data Contact Person: Address	
Data Contact Person: City	
Data Contact Person: State/province	
Data Contact Person: Postal code	
Data Contact Person: Country	
Data Contact Person: Phone	
Data Contact Person: Email	
Data Publisher: repository name	Name of repository where data will be archived
Project title	Title of project for which this dataset was collected
Project personnel: Last name	Contact person for project affiliated with this dataset
Project personnel role	Role of contact person for project affiliated with this dataset
Project description	Description of project affiliated with this dataset
Funding	Funders for project affiliated with this dataset
Data table name	This should match the name of the corresponding worksheet in the Excel workbook.
Data table description	Describe the dataset in the worksheet.
Identifier	Unique identifying URL. See "Creating Citations" on page 14.
Formatted citation	Standard DataCite <sup>1</sup> citation created by the add-in.

<sup>1</sup> [www.datacite.org](http://www.datacite.org)

Use the Data Description tab to create general metadata.

Figure 4: Create Metadata Step



Note the following features of **Data Description** tab:

- Mandatory fields are indicated by a red asterisk after the title.
- Fields with data that don't pass validation tests have red outlines and error icons. These include mandatory fields that are empty.
- Any validation error disables the **Save** button.
- The "Identifier" and "Citation" fields should be left blank, they are automatically populated based on information retrieved from the repository. See "Creating Citations" on page 14.

If you would like to view the metadata information in Excel format, download the saved or posted file to your computer. The general metadata is embedded in this worksheet (Figure 5). Metadata can only be edited using the web application tool.

Figure 5: General Metadata in "Metadata" Worksheet

	A	B
1	Name	Value
2	Title of dataset	Galapagos Rift water collections June 2008
3	Creator: First name	Jane
4	Creator: Last name	Smith
5	Creator: Organization	San Diego University
6	Creator: Address	415 20th St.
7	Creator: City	San Diego
8	Creator: State/province	CA
9	Creator: Postal code	
10	Creator: Country	USA

## Entering Column Metadata

The **Column Descriptions** tab is used to enter column specific metadata.

Use the **+Add Item** button to add column level metadata to your dataset, such as units of measure, variables, or a description of each column of data.

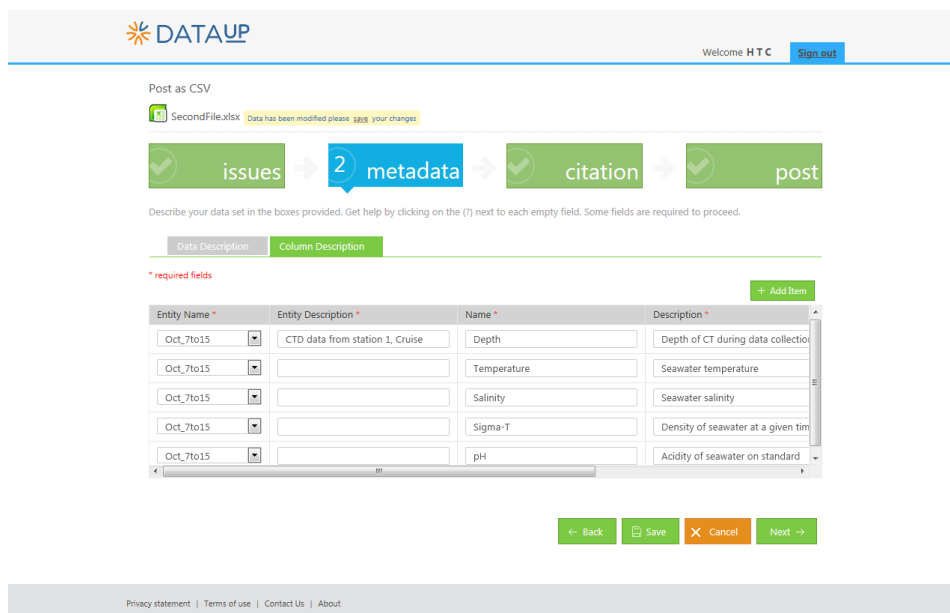
**Figure 6: Column Descriptions Tab**

The screenshot shows the DataUp web application interface. At the top, there is a header with the DataUp logo and a user greeting "Welcome HTC" with a "Sign out" button. Below the header, there is a progress bar with four steps: "issues", "2 metadata", "3 citation", and "4 post". The "2 metadata" step is highlighted. Below the progress bar, there is a section for "Describe your data set in the boxes provided. Get help by clicking on the (?) next to each empty field. Some fields are required to proceed." There are two tabs: "Data Description" and "Column Description". The "Column Description" tab is active. Below the tabs, there is a table with columns: "Entity Name", "Entity Description", "Name", and "Description". There is an "+ Add Item" button to the right of the table. Below the table, there are navigation buttons: "Back", "Save", "Cancel", and "Next".

Fill in the following for each column (Figure 7):

- **Entity Name:** A label for the entities. This defaults to, and should be the same as, the Excel worksheet from which the column labels were copied.
- **Entity Description:** A description of the entity. The Entity Description only needs to be entered once per Entity; the add-in automatically copies this text to the other column metadata.
- **Name:** The name of the column. This is initially copied from, and should be the same as, one of the column labels you copied in the previous step.
- **Description:** Description of the column data.
- **Type:** One of three values, "Numeric", "DateTime" or "Text". A dropdown list is available for quick entry.
- **Units** (only entered for numeric data): unit of measure for the data. Most common units of measure are available in a drop-down list. More complex or uncommon units are entered by hand.

Figure 7: Complete Column Descriptions



If you would like to view the column metadata information in Excel format, download the saved or posted file to your computer. The general metadata is embedded in this worksheet (Figure 8). Metadata can only be edited using the web application tool.

Figure 8: Column Descriptions Embedded in "Metadata" Worksheet

EntityName	EntityDescription	Name	Description	Type	Units
Station1	CTD data from Station 1, Cruise 1435	Depth	Depth of CTD during data collection	Numeric	meter
Station1	CTD data from Station 1, Cruise 1435	Temperature	Seawater temperature	Numeric	celsius
Station1	CTD data from Station 1, Cruise 1435	Salinity	Seawater salinity	Numeric	practical salinity units
Station1	CTD data from Station 1, Cruise 1435	Sigma-T	Density of seawater at a given temperature	Numeric	kilogram per meter
Station1	CTD data from Station 1, Cruise 1435	Chlorophyll-a	Concentration of chlorophyll in seawater	Numeric	microgram per liter
Station1	CTD data from Station 1, Cruise 1435	DO	Concentration of oxygen in seawater	Numeric	milligram per liter
Station1	CTD data from Station 1, Cruise 1435	pH	Acidity of seawater on standard 0-14 pH scale	Numeric	number
Station1	CTD data from Station 1, Cruise 1435	log10 of PAR	Log 10 of the photosynthetically active radiation of photos per meter squared per second	Numeric	number
Station1	CTD data from Station 1, Cruise 1435	PAR	Photosynthetically active radiation of photos per meter squared per second	Numeric	number

## 5 Creating Citations

Once you've finished describing your data, two final metadata must be provided before the Excel worksheet can be archived.

- An **Archival Resource Key** or **ARK**. This is a unique and persistent URL that supports long-term access to your data. The web application helps you request your repository to generate an ARK.

- A **Citation** for your data, based on the other metadata, including the ARK.

For more information on ARKs, refer to <https://confluence.ucop.edu/display/Curation/ARK>.

The Create Citation page appears (Figure 9). Follow the steps below to create a citation (some delays might occur while a remote repository is contacted):

**Figure 9: Create Citation Page**

Post as CSV

SecondFile.xlsx Data has been modified please [save your changes](#)

issues → metadata → 3 citation → 4 post

Steps:

1. Retrieve an identifier from the repository for your dataset (Read this important information about your unique identifier [here](#))
2. Check the rest of the information below for accuracy
3. Make any necessary changes or edits to the information
4. Click 'Generate Citation' to create your data citation

Please Click on [Retrieve an identifier](#) if you need [get a new identifier](#) for this document.

Identifier: ark:/99999/1k4h42dbf [Retrieve an Identifier](#)

Publication Year: 2009

Title: Galapagos Rift water collections for 2009

Version:

Publisher: Smith, Jane

Citation:

[Generate Citation](#)

[Back](#) [Save](#) [Cancel](#) [Next](#)

Privacy statement | Terms of use | Contact Us | About

1. Click on the **Retrieve an identifier** button. The **Get Identifier** dialog appears (Figure 10).

**Figure 10: Get Identifier Dialog**

GET IDENTIFIER

Please provide repository details to get identifier

Repository Name\* ONEShare

I Accept [User Agreement](#)

[OK](#) [Cancel](#)

2. Read and accept the User Agreement, then choose a repository from the drop-down list. Click **OK**.

- The **Create Citation** dialog reappears, with the **Unique Identifier** field filled in (Figure ). Click the **Generate citation** button.

**Figure 11: Create Citation Dialog with Identifier Filled In**

CREATE CITATION

Steps:

1. Retrieve an identifier from the repository for your dataset (Read this important information about your unique identifier [here](#))
2. Check the rest of the information below for accuracy
3. Make any necessary changes or edits to the information
4. Click 'Generate Citation' to create your data citation

Unique identifier: ark:/90135/q1028pfv

Publication year: 2012

Dataset title: Galapagos Rift water collections June 2008

Dataset version:

Publisher (Data owner): Smith, Jane

Citation:

- The **Citation** field is filled in (Figure 12). Click the **Save** button.

**Figure 12: Create Citation Page with Citation Filled In**

DATAUP

Welcome H T C

Post as CSV

SecondFile.xlsx Data has been modified please [save your changes](#)

issues → 2 metadata → 3 citation → 4 post

Steps:

1. Retrieve an identifier from the repository for your dataset (Read this important information about your unique identifier [here](#))
2. Check the rest of the information below for accuracy
3. Make any necessary changes or edits to the information
4. Click 'Generate Citation' to create your data citation

Please Click on [Retrieve an identifier](#) if you need get a new identifier for this document.

Identifier: ark:/99999/fk4h42dbf

Publication Year: 2009

Title: embedded metadata step

Version:

Publisher: c, h

Citation: c, h(2009); embedded metadata step . c, h. ark:/99999/fk4h42dbf

Privacy statement | Terms of use | Contact Us | About

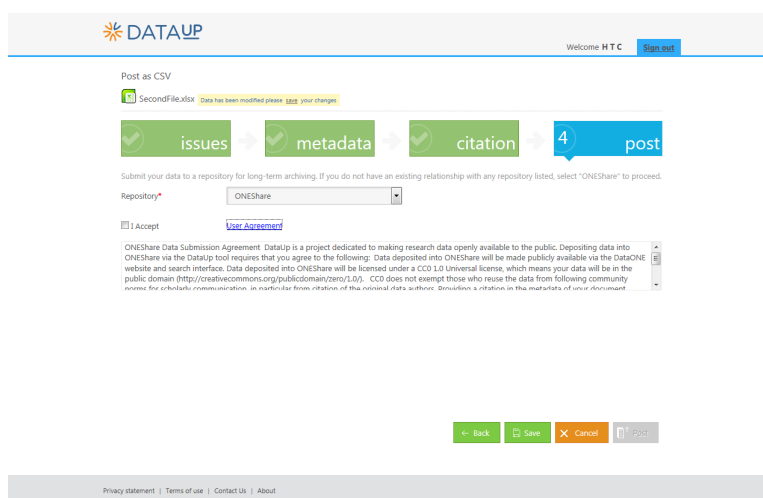


## 6 Posting to the Repository


This section describes how use to use the web application to submit your data to a repository for long-term archiving.

1. At the Post step of the web application, choose a repository to store your data.
2. After you have viewed the User Agreement for the repository selected, click **I Accept**.
3. Click on the **Post** button. Notice that the post button is disabled until you have accepted the User Agreement.

**Figure 13: Post Page with User Agreement Displayed**



### 6.1 Check your data

Once you have posted your data to the repository it's a good idea to go check to make sure everything posted as you would expect. At the main DataUp web application page, use the Download icon (  ) to open the file that you just added to the repository. Notice that the web application has fixed the issues identified earlier and the Metadata tab has been created.