

The background of the slide is a vibrant cosmic image featuring a nebula with swirling clouds of pink, purple, and blue gas, interspersed with numerous bright stars of varying sizes and colors, including white, yellow, and red.

International Data Curation Conference  
Workshop on Data Curation in the Cloud  
February 27<sup>th</sup>, 2014

# PUTTING YOUR USER'S DATA IN THE CLOUD WITH DATAUP

Kristin M. Tolle, Ph.D.

Director, Environmental Science Infrastructure Development  
Microsoft Research Connections

# TRAINING AGENDA

- What is DataUp? (in brief)
- What's new? (in brief)
- How do I become an admin and set up my own repository?
- How do I get the DataUp source code to deploy my own instance?
- What is the admin/user experience based on selections I make in the UI once I set up my repository

Please ask questions as we go!!



# DATAUP OBJECTIVES (CIRCA 2011)

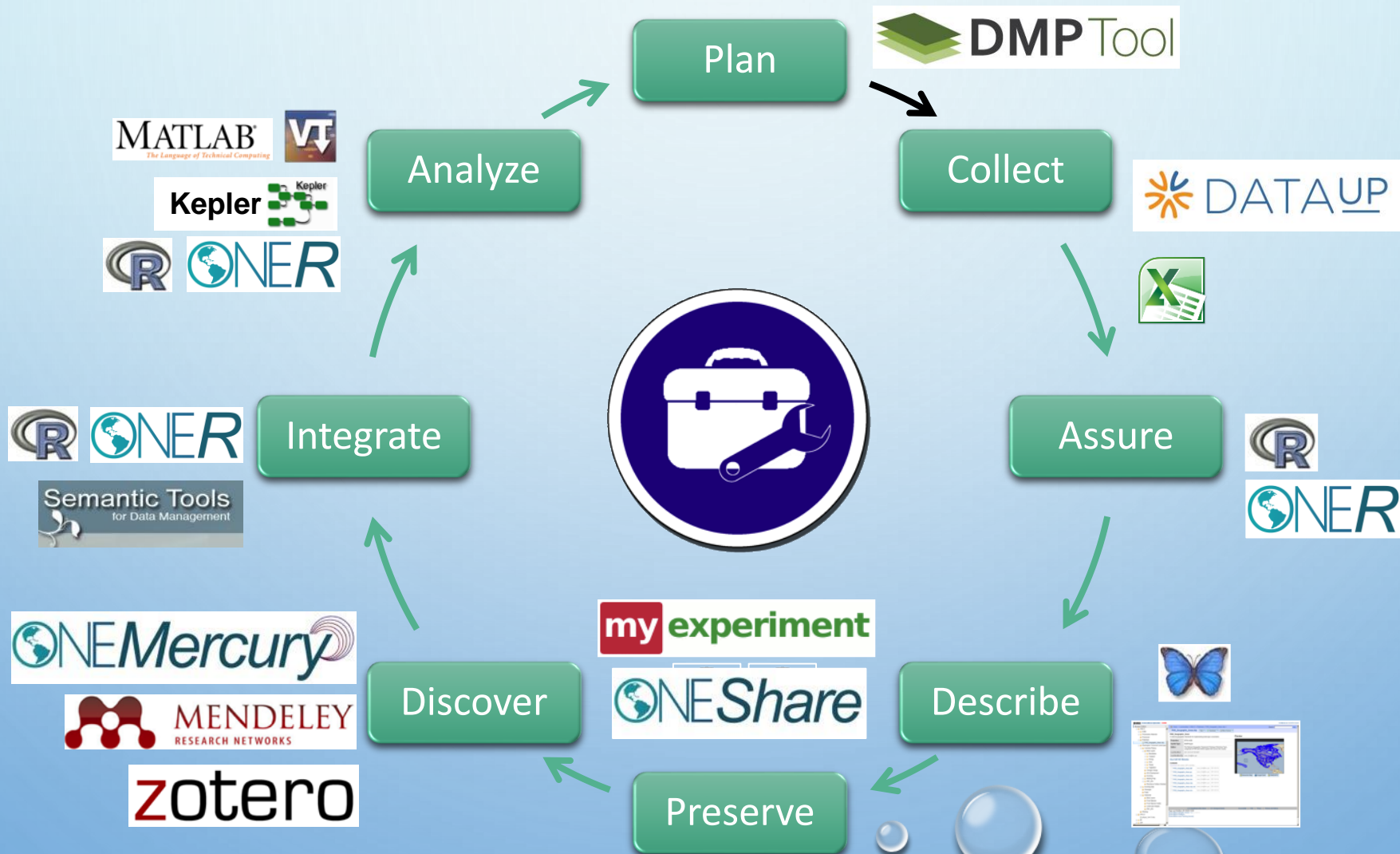
- Create a tool designed to improve data reusability practices within the tool used most\* by (environmental) scientists—Microsoft Excel
- Users stepped through a workflow of steps to create the most reusable data and place that in a repository. Designed to:
  - Encourage uploaders to follow “best practices” for data reusability
  - Encourage users to attach appropriate metadata (EML) to aid in reuse
  - Give users a citation so that they can appropriately cite their data
  - Post to a repository occurs only when (required) conditions are met

\* Based on several surveys conducted by the California Digital Library, supported by an AGU survey taken by MSR

# DATAUP V1

- Collaborative project with University of California, California Digital Library, The Gordon and Betty Moore Foundation, and Microsoft Research.
- Released October 2012 at DataONE Global Meeting—as this was a DataONE-centric project.
- Provided both an Excel Add-in and an Azure web service.
- Source code released on OuterCurve and BitBucket as Apache 2.0 open source license.
- Planned to support all DataONE repositories (via collaborator API). Ended up only supporting ONEShare (Merritt).
- Private funding concluded—application submitted for NSF funding (awarded August 2013—10 month gap)

# INVESTIGATOR TOOLKIT SUPPORT



# WHAT IS SEQUIM??

- Microsoft Research wanted to repurpose DataUp for internal needs (e.g., downstream data processing).
- We also interacted with many folks at AGU and other events and saw gaps in the DataUp experience for other repositories and organizations.
  - Need UI for administrators to manage their repositories
  - Need flexibility in metadata (EML is too limited—expand to Dublin Core and make dynamic)—support any type of scientist.
  - Need data validation to ensure that data will work in downstream processes and the Environmental Science tools (e.g., MatLab scripts, FetchClimate, Node Atlas, Distribution Modeller, etc.)
  - Need for Data Validation on a more atomic level.
  - Need to set up more repositories—including personal ones like ONEDrive where interim processing steps can be stored with auto-generated metadata



# A FAMILY OF TOOLS FOR SCIENCE

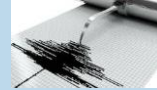
## Natural Disasters



### Telonics



### Richter



## Water-Power



### QinetikQ



### NOAA



## Species Migration



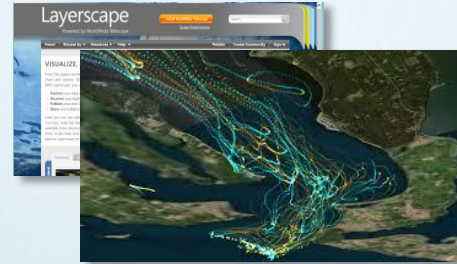
### Mataki



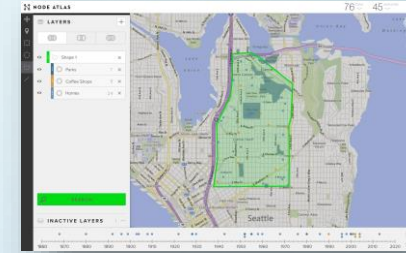
### CLEO



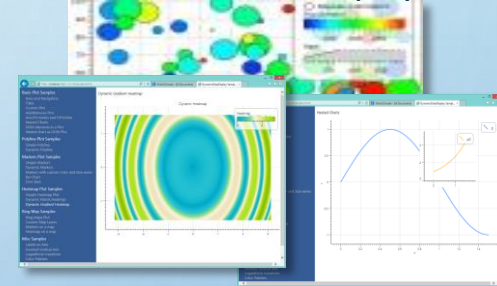
## Layerscape



## Node Atlas



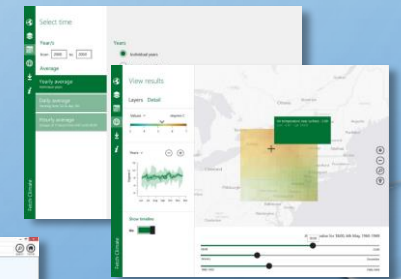
## Interactive Data Display



## DataUp



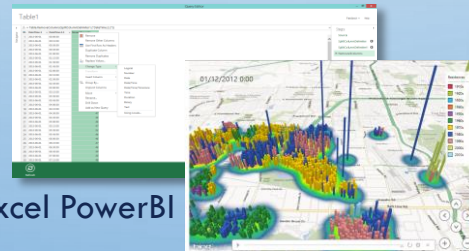
## FetchClimate



## Distribution Modeller



## Excel PowerBI



# SEQUIM => DATAUP V2

- Restarted collaboration with CDL Fall 2013.
- Released February 24<sup>th</sup>, 2014 in conjunction with IDCC 2014.
- NSF funded programmer started last week to do future updates in BitBucket directly.
- In discussions with DataONE to develop mechanisms for creating a “vibrant” open source contribution community.
- Replaced V1 on <http://dataup.org> site.
- Add-in deprecated (no planned future release).



# BRASS TACKS

- DataUp was/is developed:
  - HTML5 frontend
  - C# backend developed and optimized to run on .NET and Windows Azure
- <http://dataup.org> is the California Digital Libraries (CDL) Azure instance of DataUp—not Microsoft Research
- <https://bitbucket.org/dataup/dataup2> BitBucket repository is also managed by CDL
- But... you can stand up your own version ...more on this later



# WHAT CAN I DO WITH THIS VERSION?

<http://www.dataup.org/>



# REQUEST ADMIN ACCESS



DataUp is a data documenting, archiving and sharing tool that allows you to:

- Upload tabular data to your own repository in the cloud
- Perform a best practices check on your data
- Add metadata and descriptions to your data
- Share your data with others

Sign in with...



Microsoft Account



Facebook Account



Contact is the place to request access



## Contact Us

For information on the DataUp project partners, visit [About the Project](#).

We welcome your questions, comments, and/or feedback!

By email: [uc3@ucop.edu](mailto:uc3@ucop.edu)

Via Facebook: [www.facebook.com/DataUpCDL](https://www.facebook.com/DataUpCDL)

Or use this feedback form:

Name:

Email:

Your questions,  
comments and  
feedback:

Information submitted via this form is subject to [California Digital Library Privacy Policy](#).

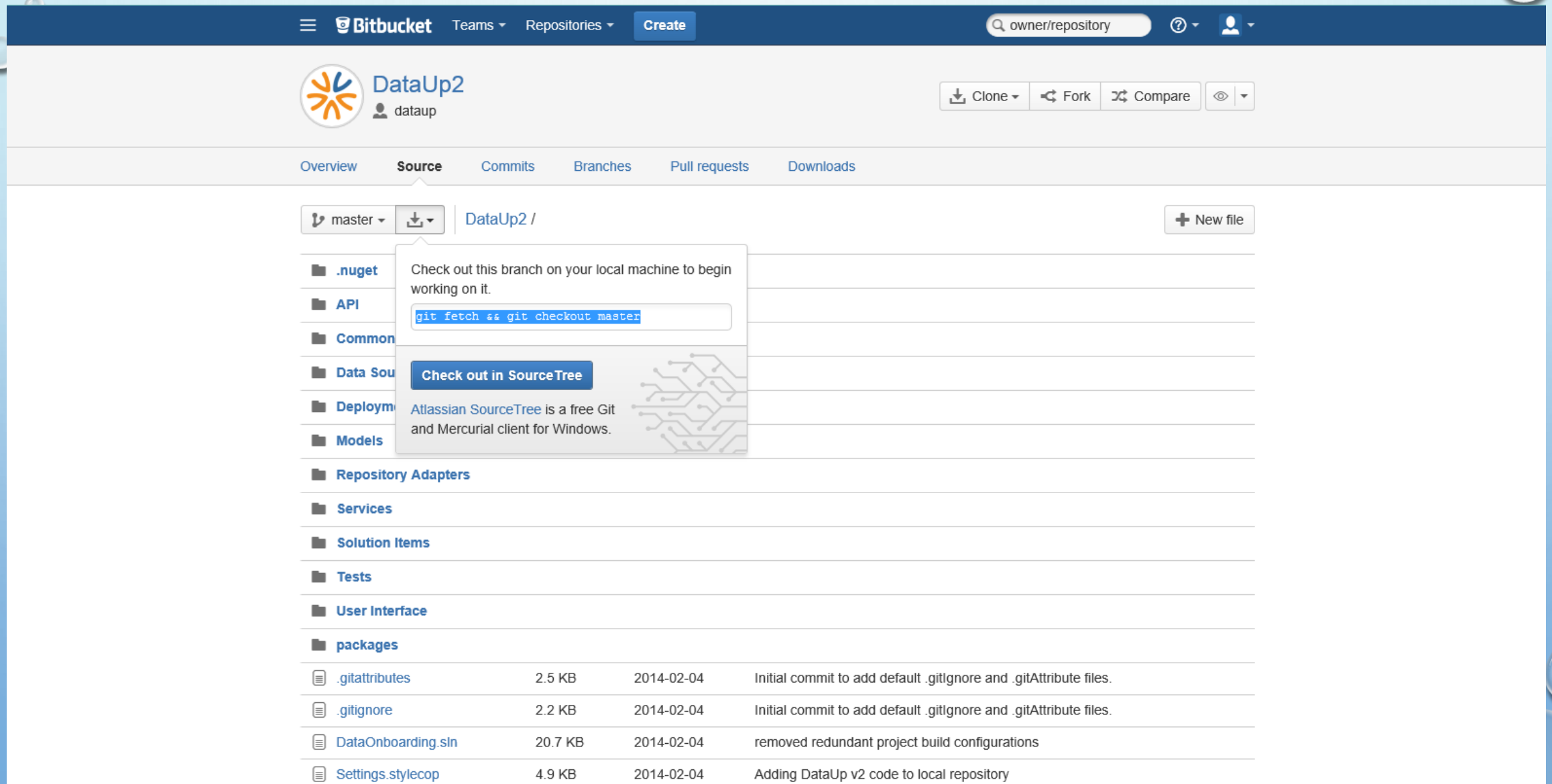
[Back to DataUp Home](#)

Presently this points to the CDL website set up for DataUp—which is appropriate since you are asking to use their instance of DataUp

# SOME TERMINOLOGY

- **Repository type:** new repository you want to integrate into DataUp. DataUp v2 Feb 2014 release supports Merritt ONEShare and OneDrive out of the box.
- **Repository Adapter:** Software component that needs to be implemented for every repository type to facilitate bidirectional data transfer. This is how you make your repository available to your users.

# HOW DO I STAND UP MY OWN INSTANCE?



The screenshot shows the Bitbucket web interface for the 'DataUp2' repository. The top navigation bar includes the Bitbucket logo, 'Teams', 'Repositories', and a 'Create' button. A search bar contains 'owner/repository'. The repository name 'DataUp2' and user 'dataup' are displayed. Action buttons for 'Clone', 'Fork', 'Compare', and a view toggle are present. The 'Source' tab is selected, showing a file tree on the left and a commit history table at the bottom. A modal window is open over the 'master' branch, providing instructions on how to checkout the branch using Git or SourceTree.

Bitbucket Teams Repositories Create owner/repository ?

DataUp2 dataup Clone Fork Compare

Overview Source Commits Branches Pull requests Downloads

master DataUp2 / + New file

.nuget API Common Data Sou Deploy Models Repository Adapters Services Solution Items Tests User Interface packages

Check out this branch on your local machine to begin working on it.

```
git fetch && git checkout master
```

Check out in SourceTree

Atlassian SourceTree is a free Git and Mercurial client for Windows.

.gitattributes	2.5 KB	2014-02-04	Initial commit to add default .gitignore and .gitAttribute files.
.gitignore	2.2 KB	2014-02-04	Initial commit to add default .gitignore and .gitAttribute files.
DataOnboarding.sln	20.7 KB	2014-02-04	removed redundant project build configurations
Settings.stylecop	4.9 KB	2014-02-04	Adding DataUp v2 code to local repository



# INITIAL STEPS

- Sign up for a BitBucket account.
- Browse to the DataUp v2 repository on BitBucket located at <https://bitbucket.org/dataup/dataup2>
- Fork the DataUp v2 repository under your account.
- Clone the forked repository from the DataUp v2 repository onto your local disk.
- Now you can start developing or deploying (as is) your own instance.

# AUTHENTICATION CONSIDERATIONS

- What security scheme does the repository type conform to?
- DataUp v2 has two reference implementations:
  - Basic authentication (user name + password combo). ONEShare repositories are secured by this scheme.
  - Windows live authentication (token based auth). OneDrive repositories are secured by this scheme.
- If the repository leverages one of the above schemes, you can reuse the code to facilitate the new repository type.
- If not extend the existing AuthN scheme to implement the new mechanism.

# DATA TRANSFER PROTOCOLS

- What data transfer protocol do I want the repository type conform to?
- All repositories in DataUp v2 follow HTTP/HTTPS protocol to transmit files and data.

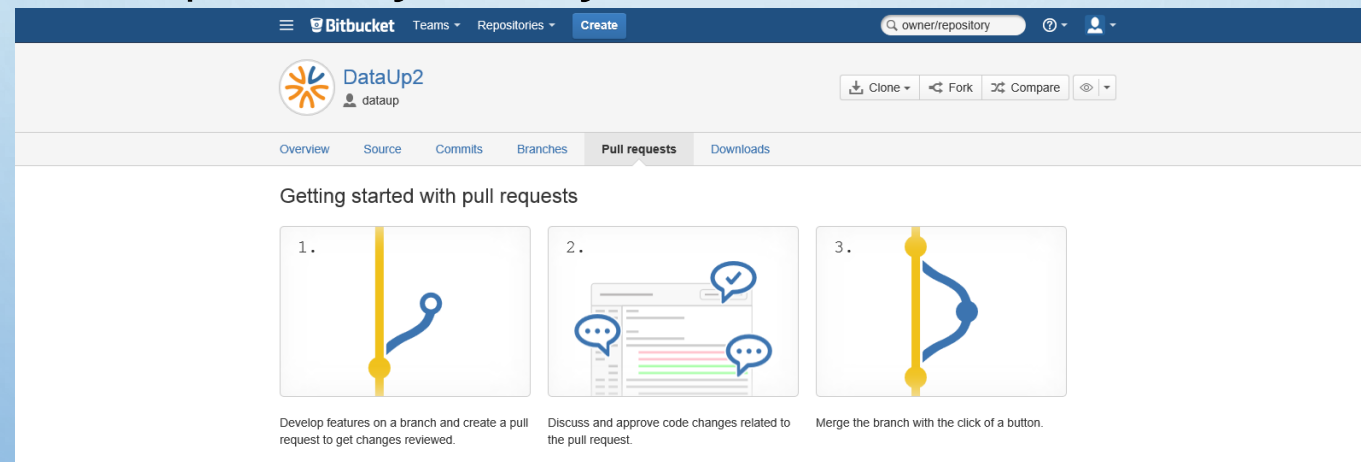


# NEXT STEPS

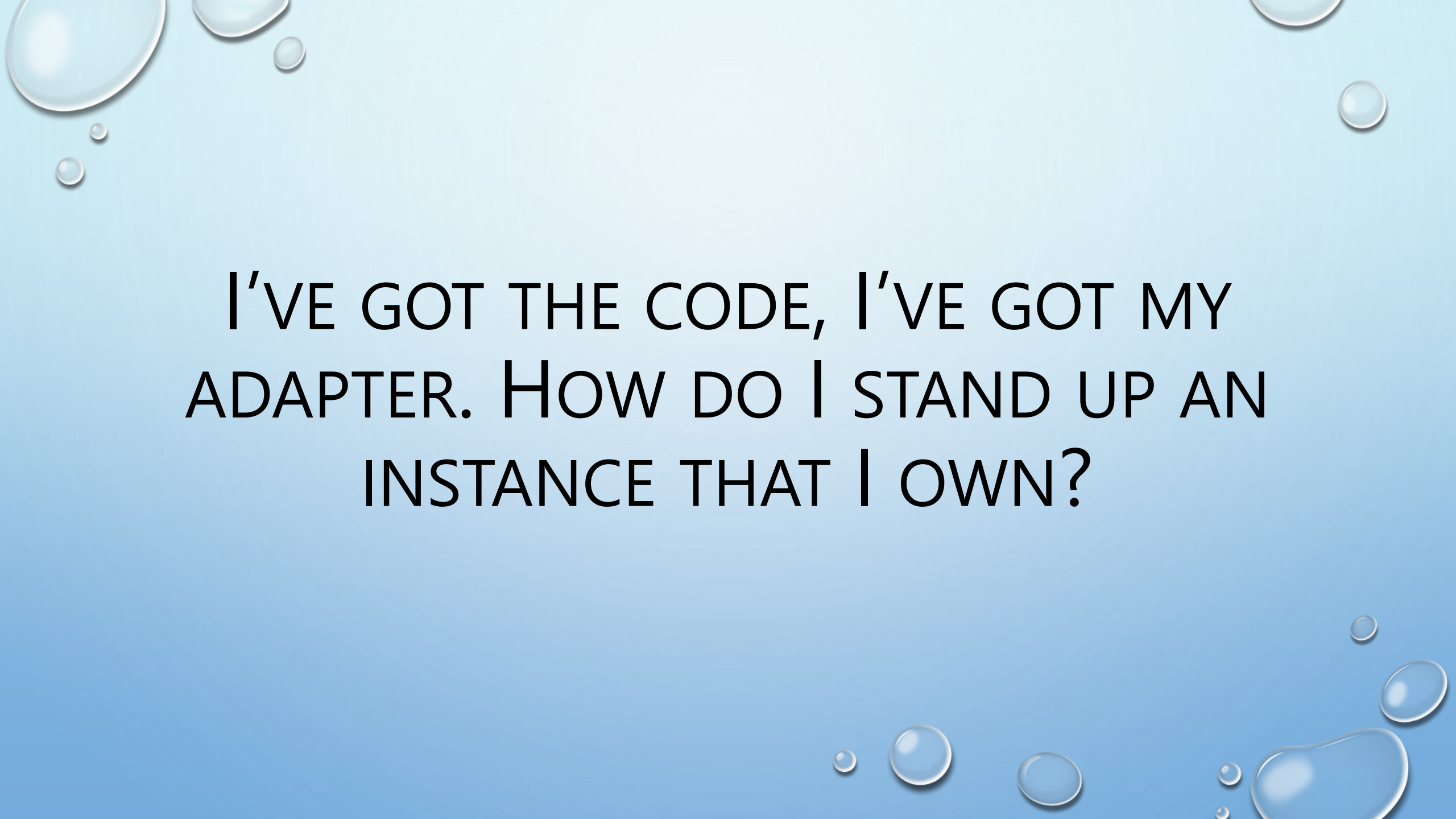
- Add a new repository adapter by extending the repository adapter façade. (Requires implementing [IRepositoryAdapter](#) interface and extend [RepositoryAdapterFactory](#) class)
- Create (or reuse/customize if you support HTTP/HTTPS) the administrative user interface for admins to configure a new repository instance in DataUp (example, UI to provide uri supporting HTTP verbs GET, POST, PUT and DELETE).
- Re-use existing authentication scheme or implement new one to secure data transfer.

# SHARING WITH OTHERS

- If you create a repository adapter that is specific to your repository you may wish to share that with others.
- After testing the changes on your local box, submit a “Pull” request to the DataUp v2 repository that you forked from.



- Wait for the repo owner (CDL) to accept or reject your pull request.

The background is a light blue gradient. In the top-left corner, there are several water droplets of varying sizes, some with highlights. In the bottom-right corner, there are more water droplets, including a large one and several smaller ones. The text is centered in the middle of the slide.

I'VE GOT THE CODE, I'VE GOT MY  
ADAPTER. HOW DO I STAND UP AN  
INSTANCE THAT I OWN?



# SHORTEST PATH

- DataUp source on BitBucket is optimized for .NET/Azure the easiest thing is to set up an Azure instance of your own and push the bits out there.
- First Step—get an Azure account:
  - (if academic) Apply for an Azure4Research one-year grant  
<http://research.microsoft.com/en-us/projects/azure/>
  - (if not) Request a free 1-month subscription from  
<http://www.windowsazure.com/en-us/pricing/free-trial/>

# LONGER PATH

- Porting the code to another platform (e.g., Linux)
- ...Then standing it up on Azure ☺
- We'll teach you how to set up a Linux VM at an Azure training near you <http://research.microsoft.com/en-us/projects/azure/training.aspx>

# WHAT'S NEXT?

1. You must be an admin on the Azure instance
2. (recommended) HTTPS Certificate to secure web application and the API
3. Deployment script templates are checked into BitBucket  
<https://bitbucket.org/dataup/dataup2/src/e26c62b12c6f935190fc5db47076f905b13d3f81/Deployment/?at=master>
4. Publish to Azure

# A CLOSER LOOK AT STEP #3

- Create a new ACS namespace under Azure Active Directory Services (or use existing)
- Configure Identity providers. DataUp2 (presently) supports Microsoft Account (Windows Live Id) and Facebook:
  - Configuring Facebook requires registering the deployment with Facebook developer network.
  - No additional configuration required for Windows Live
- Configure DataUp2 API as a "Relying party" application
- Create new SQL Azure Database
- Create new Azure blob storage account (use an existing)
- Update the configuration in DataUp2 deployment scripts with the keys provided in the Azure subscription (requires some technical knowledge and working with XML)
- Package the deployment scripts



# CREATE A NEW ACS NAMESPACE UNDER AZURE ACTIVE DIRECTORY SERVICES

The screenshot displays the Azure portal interface. The top navigation bar shows 'Windows Azure' and the user's profile. The left sidebar lists various services: ALL ITEMS, WEB SITES, VIRTUAL MACHINES, MOBILE SERVICES, CLOUD SERVICES, SQL DATABASES, STORAGE, and HDINSIGHT. The main content area is titled 'active directory' and shows a table of active directories. Below this, a 'NEW' modal is open, allowing the user to create a new ACS namespace. The modal includes a list of services on the left, a selection of 'ACCESS CONTROL' in the center, and a form on the right to specify the namespace name, region, and domain.

active directory

NAME	STATUS	SUBSCRIPTION	LOCATION
dataup	✓ Active	Pay-As-You-Go	West US
CDLTest	✓ Active	Pay-As-You-Go	West US

NEW

- MEDIA SERVICE
- SERVICE BUS
- VISUAL STUDIO ONLINE PREVIEW
- BIZTALK SERVICE
- ACTIVE DIRECTORY

DIRECTORY

ACCESS CONTROL

MULTI-FACTOR AUTH PROVIDER

QUICK CREATE

NAMESPACE

new-acs-namespace ✓

REGION

West US ✓

CREATE ✓

# CONFIGURE IDENTITY PROVIDERS

The screenshot shows a web browser window with the URL `https://dataup.accesscontrol.windows.net/v2/mgmt/web/IdentityProvider/Select`. The browser's address bar includes a search icon, a lock icon, and two tabs: 'Active Directory - Windows Az...' and 'Add Identity Provider'. The browser's menu bar shows 'File', 'Edit', 'View', 'Favorites', 'Tools', and 'Help'. The page header displays the 'Windows Azure Platform' logo and the text 'Access Control Service' with a breadcrumb 'Service Namespace: dataup > Identity providers >'. A left-hand navigation pane lists several categories: 'Home', 'Trust relationships' (with sub-items 'Identity providers', 'Relying party applications', and 'Rule groups'), 'Service settings' (with sub-items 'Certificates and keys' and 'Service identities'), 'Administration' (with sub-items 'Portal administrators' and 'Management service'), and 'Development' (with sub-item 'Application integration'). The main content area is titled 'Add Identity Provider' and contains the instruction 'Select an identity provider to add to this ACS service namespace.' Below this, there are two sections. The first, 'Add a custom identity provider (prerequisites required)', contains two radio button options: 'WS-Federation identity provider (e.g. Microsoft AD FS 2.0)' (which is selected) and 'Facebook application'. Both options have a 'Learn more' link. The second section, 'Add a preconfigured identity provider', contains three radio button options: 'Windows Live ID (already added)', 'Google', and 'Yahoo!'. At the bottom of the main content area are 'Next' and 'Cancel' buttons.

Windows Azure Platform

Access Control Service Service Namespace: dataup > Identity providers >

**Home**

**Trust relationships**

- Identity providers
- Relying party applications
- Rule groups

**Service settings**

- Certificates and keys
- Service identities

**Administration**

- Portal administrators
- Management service

**Development**

- Application integration

## Add Identity Provider

Select an identity provider to add to this ACS service namespace.

### Add a custom identity provider (prerequisites required)

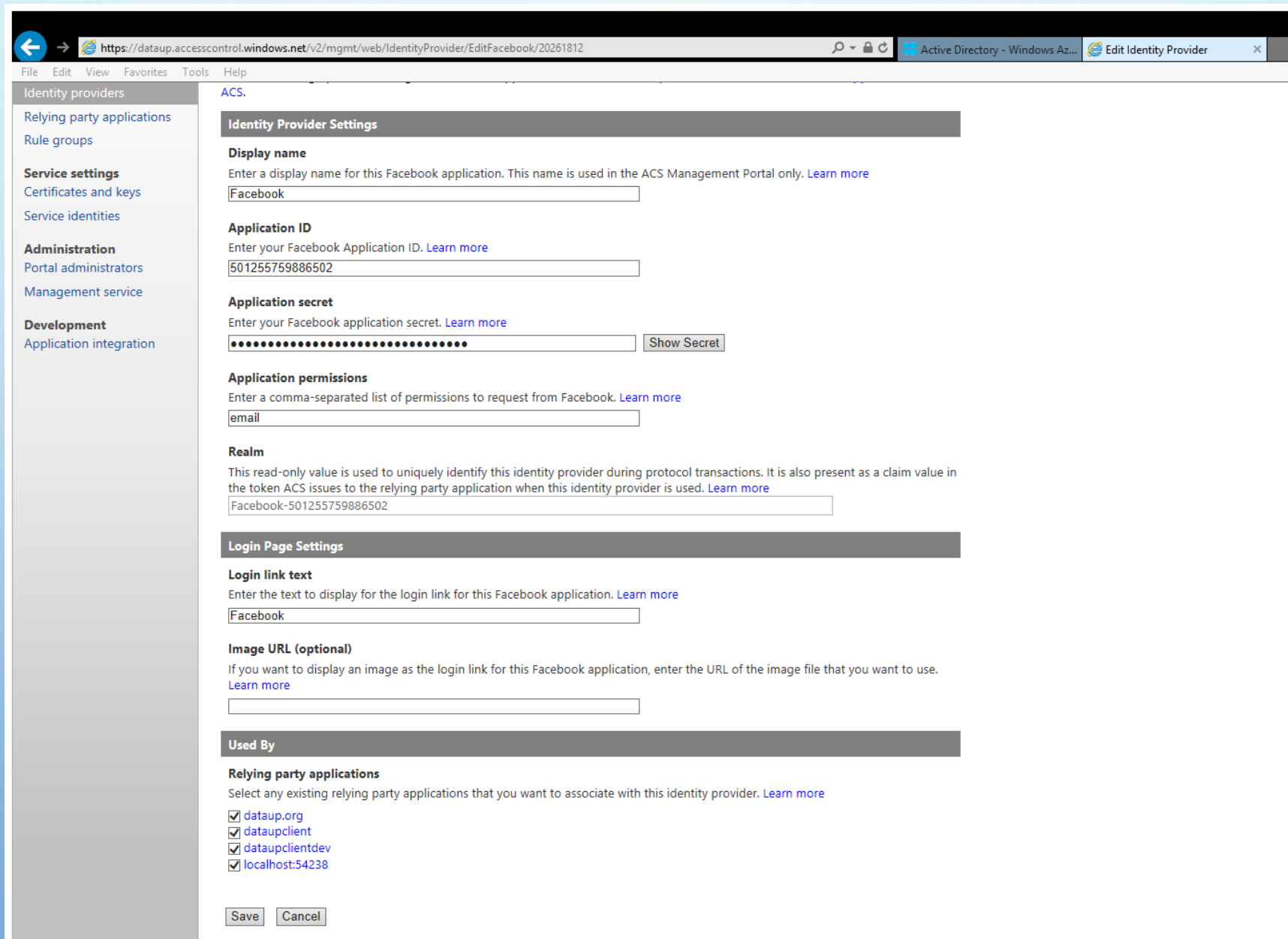
- ☒ WS-Federation identity provider (e.g. Microsoft AD FS 2.0) [Learn more](#)
- ☐ Facebook application [Learn more](#)

### Add a preconfigured identity provider

- ☐ Windows Live ID (already added)
- ☐ Google
- ☐ Yahoo!

[Next](#) [Cancel](#)

# RELYING ENTITY PARTNER SETUP



The screenshot shows the 'Edit Identity Provider' page in the ACS Management Portal. The browser address bar shows the URL: `https://dataup.accesscontrol.windows.net/v2/mgmt/web/IdentityProvider/EditFacebook/20261812`. The page has a left-hand navigation menu with categories: Identity providers, Service settings, Administration, and Development. The main content area is titled 'ACS.' and contains two sections: 'Identity Provider Settings' and 'Login Page Settings'. The 'Identity Provider Settings' section includes fields for 'Display name' (Facebook), 'Application ID' (501255759886502), 'Application secret' (masked with dots), 'Application permissions' (email), and 'Realm' (Facebook-501255759886502). The 'Login Page Settings' section includes 'Login link text' (Facebook) and an 'Image URL (optional)' field. At the bottom, the 'Used By' section lists existing relying party applications with checkboxes: dataup.org, dataupclient, dataupclientdev, and localhost:54238. 'Save' and 'Cancel' buttons are at the bottom.

Identity providers

- Relying party applications
- Rule groups

**Service settings**

- Certificates and keys
- Service identities

**Administration**

- Portal administrators
- Management service

**Development**

- Application integration

ACS.

### Identity Provider Settings

**Display name**  
Enter a display name for this Facebook application. This name is used in the ACS Management Portal only. [Learn more](#)  
Facebook

**Application ID**  
Enter your Facebook Application ID. [Learn more](#)  
501255759886502

**Application secret**  
Enter your Facebook application secret. [Learn more](#)  
..... [Show Secret](#)

**Application permissions**  
Enter a comma-separated list of permissions to request from Facebook. [Learn more](#)  
email

**Realm**  
This read-only value is used to uniquely identify this identity provider during protocol transactions. It is also present as a claim value in the token ACS issues to the relying party application when this identity provider is used. [Learn more](#)  
Facebook-501255759886502

### Login Page Settings

**Login link text**  
Enter the text to display for the login link for this Facebook application. [Learn more](#)  
Facebook

**Image URL (optional)**  
If you want to display an image as the login link for this Facebook application, enter the URL of the image file that you want to use. [Learn more](#)

### Used By

**Relying party applications**  
Select any existing relying party applications that you want to associate with this identity provider. [Learn more](#)

- ☒ dataup.org
- ☒ dataupclient
- ☒ dataupclientdev
- ☒ localhost:54238

[Save](#) [Cancel](#)

# CREATE A SQL AZURE DATABASE

The screenshot displays the Windows Azure portal interface. At the top, the header shows 'Windows Azure' and a user profile for 'rajashekhar.k@outlook.com'. The left sidebar contains navigation links for various services: ALL ITEMS, WEB SITES (0), VIRTUAL MACHINES (0), MOBILE SERVICES (0), CLOUD SERVICES (2), SQL DATABASES (2), STORAGE (2), and HDINSIGHT. The main content area is titled 'sql databases' and shows a table of existing databases.



NAME	STATUS	LOCATION	SUBSCRIPTION	SERVER	EDITION	MAX SIZE
adminservice	✓ Online	West US	Pay-As-You-Go	v12h47zv64	Web	1 GB
dataupclient	✓ Online	West US	Pay-As-You-Go	v12h47zv64	Web	5 GB


Below the table, the 'NEW' section is visible, featuring a grid of service categories. The 'SQL DATABASE' category is selected, leading to a 'QUICK CREATE' wizard. The wizard includes input fields for 'DATABASE NAME' (set to 'dataupdb') and 'SERVER' (set to 'v12h47zv64 (Location=West US)'). A 'CREATE SQL DATABASE' button with a checkmark is at the bottom right.





# SETUP AZURE BLOB STORAGE


Windows Azure


Subscriptions  rajashekhar.k@outlook.com 


 ALL ITEMS


 WEB SITES  
0


 VIRTUAL MACHINES  
0

 MOBILE SERVICES  
0

 CLOUD SERVICES  
2

 SQL DATABASES  
2


 STORAGE  
2


 HDINSIGHT  
0


storage


NAME	STATUS	LOCATION	SUBSCRIPTION	
datatupstorage	→ ✓ Online	West US	Pay-As-You-Go	
dataupdev	✓ Online	West US	Pay-As-You-Go	


NEW


 COMPUTE


 DATA SERVICES


 APP SERVICES


 NETWORK SERVICES


 STORE  
PREVIEW


 SQL DATABASE

 STORAGE

 HDINSIGHT

 CACHE  
PREVIEW

 RECOVERY SERVICES

 QUICK CREATE

URL

dataupstore ✓  
\*.core.windows.net

LOCATION/AFFINITY GROUP

West US ▼

REPLICATION

Geo-Redundant ▼

CREATE STORAGE ACCOUNT ✓

# WHAT ARE SQL AND THE BLOB STORAGE FOR?

- As you know, DataUp doesn't hold files. We want people to put them in the repository and we do not want to pay for storing them twice.
- SQL holds the file properties (e.g., owner, location, status, etc.) and is used to populate the UI that users see.
- Blob storage (used 'as needed') is the temporary storage location for the files users upload.

# NEW ADMIN UI EXPERIENCE

# WHAT CAN I DO AS AN ADMIN?

- Files—same as users
- Add/edit repositories
- Create data validation rules—repository independent!!

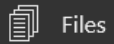
The screenshot displays the DataUp web application interface. The browser address bar shows the URL: <https://web.dataup.org/Home/index?jwt=eyJ0eXAiOiJKV1QiLCJhbGciOiJIUzI1I>. The page header is orange with the DataUp logo on the left and user information 'Welcome, Sequim Test | Settings | Sign Out' on the right. A dark sidebar on the left contains three menu items: 'Files' (selected), 'Repositories', and 'Quality Checks'. The main content area is titled 'My Files' and includes a 'View All' dropdown and a 'Date Uploaded' dropdown. Action buttons for 'Post', 'Download', and 'Delete' are visible. The file list is organized in a grid of cards, each representing a file with its name, repository, and a status indicator (orange checkmark). The files listed are:

File Name	Repository	Status
+ Add a New File OR Drag & Drop Files Here		
DataFile4 - with i... .xlsx	Test Repository	✓
s1.csv	deptest_sd_imp	✓
excel_good_exam... .xlsx	deptest_sd_imp	✓
copes-weekly_data.xlsx	deptest_merritt_imp02	✓
copes-PinkDays.xlsx	Test Repository	✓
copes-scope-of-d... .xlsx	deptest_sd_imp	✓
copes_log4.xlsx	Test Repository	✓
Copes-RedDays.xlsx	OneShare	✓
test file.xlsx	OneShare	✓
excel_good_exam... .xlsx	OneShare	✓
s1.csv	deptest_sd_imp	✓

The footer contains links for 'Privacy Statement', 'Terms of Use', 'Contact', and 'About Dataup', along with logos for 'Microsoft Research' and 'Powered by Windows Azure'.



# MANAGING REPOSITORIES



Files



Repositories



Quality Checks

## Repositories

[+ Add Repository](#)

Repository Name	File Type	Visibility	Date Created	Created By	Actions
Test Repository			2/25/2014 11:27:18 AM	Susan Borda	
deptest_merritt_imp02			2/13/2014 2:50:08 PM	Sequim Test	
OneShare			2/11/2014 4:15:39 PM	Susan Borda	
deptest_sd_imp			2/11/2014 1:32:32 PM	Sequim Test	

# THINGS TO NOTE

- Admins of an instance of DataUp can see all of the “visible” repositories in that instance—even if they are not visible to users.
- Three major functions on the page that an admin can do:
  - Add repository
  - Edit an existing repository
  - Remove a repository

# VISIBILITY OF A REPOSITORY



Files



Repositories



Quality Checks

## Repositories

[+ Add Repository](#)

Repository Name	File Type	Visibility	Date Created	Created By	Actions
Test Repository			2/25/2014 11:27:18 AM	Susan Borda	
deptest_merritt_imp02			2/13/2014 2:50:08 PM	Sequim Test	
OneShare			2/11/2014 4:15:39 PM	Susan Borda	
deptest_sd_imp			2/11/2014 1:32:32 PM	Sequim Test	

# WHAT AN END USER SEES

The screenshot displays the DATAUP web application interface. At the top, an orange header bar contains the DATAUP logo on the left and the text "Welcome, Kristin Tolle | Sign Out" on the right. Below the header, the main content area is titled "My Files". To the right of this title are two dropdown menus: "Uploaded Files" and "Date Uploaded". Further right are three action buttons: "Post" (with a plus icon), "Download" (with a download icon), and "Delete" (with a delete icon). A tooltip box is visible over the "Post" button, containing the text "Test Repository" and "OneShare".

On the left side of the "My Files" section, there is a dashed box containing the text "Add a New File" with a plus icon, followed by "OR" and "Drag & Drop Files Here" with a cursor icon. An orange callout box with a white border and a shadow points to this area, containing the text "Users can upload single, multiple or zip files".

Below the "Add a New File" section, there is a form for uploading a file. The form has four fields: "File Name" (containing "Tibet5"), "File Extension" (containing "CSV"), "File Size" (containing "515.60 KB"), and "Uploaded Date" (containing "2/25/2014 15:21:30"). To the right of these fields is a "Status" field containing the text "Uploaded".

At the bottom of the "My Files" section, there is a row of five file thumbnails: "Tibet1.csv", "Masterlist.csv", "Configuration.csv", "Chicago Narcotics.xlsx", and "Earthquake Data f... .xlsx".

The footer of the page contains a row of links: "Privacy Statement", "Terms of Use", "Contact", and "About Dataup". On the right side of the footer, there are logos for "Microsoft Research" and "Powered by Windows Azure".



# EDITING A REPOSITORY



Files



Repositories




Quality Checks

## Repositories

[+ Add Repository](#)

Repository Name	File Type	Visibility	Date Created	Created By	Actions
Test Repository			2/25/2014 11:27:18 AM	Susan Borda	
deptest_merritt_imp02			2/13/2014 2:50:08 PM	Sequim Test	
OneShare			2/11/2014 4:15:39 PM	Susan Borda	
deptest_sd_imp			2/11/2014 1:32:32 PM	Sequim Test	

# NAME, TYPE, USER LOGIN

 DATAUP

Welcome, Sequim Test | [Settings](#) | [Sign Out](#)

Files

Repositories

Quality Checks

←

Edit Repository

Repository Name \*

KrisTest\_sd\_imp

Repository Type

Merritt

SkyDrive

☒ Impersonate

User Agreement

File Type


xslx;csv;shp;nc;cdf


Visibility

Visible to Admin

▼

Create Metadata

 Add Item


 Get Data from XML

Privacy Statement | [Terms of Use](#) | [Contact](#) | [About Dataup](#)

Microsoft Research

Powered by Windows Azure

# FILE TYPES, VISIBILITY

Welcome, Sequim Test | [Settings](#) | [Sign Out](#)

Files

Repositories

Quality Checks

Repository Name \*

Repository Type

☒ Impersonate

User Agreement

File Type

Visibility

deptest\_sd\_imp

SkyDrive

xls;xlsx;txt;csv;cdf;nc;shp

Visible to All

Create Metadata

+

 Add Item

↑

 Get Data from XML

Mapping	Field *	Description *	IS-Required	Type *	Range
First Name	First Name	enter your First Name	<input checked="" type="checkbox"/>	Text	
Last Name	Last Name	enter your Last Name	<input checked="" type="checkbox"/>	Text	

[Privacy Statement](#) | [Terms of Use](#) | [Contact](#) | [About Dataup](#)

Microsoft Research

Powered by Windows Azure

# METADATA

DATAUP

Files

Repositories

Quality Checks

Welcome, Sequim Test | Settings | Sign Out

Repository Type

☒ Impersonate

User Agreement

File Type

Visibility

Create Metadata

+ Add Item

Get Data from XML

Click on +Add Item or Get By XML button to provide Meta Data Fields

Mapping	Field *	Description *	IS-Required	Type *	Range

Cancel

Submit


Privacy Statement | Terms of Use | Contact | About Dataup

Microsoft Research

Powered by Windows Azure



# SETTING TYPES

Welcome, Sequim Test | [Settings](#) | [Sign Out](#)

Files

Repositories

Quality Checks

File Type

Visibility

File Type: xls;xlsx;txt;csv;cdf;nc;shp

Visibility: Visible to All

Create Metadata

+

 Add Item

↑

 Get Data from XML

Mapping	Field *	Description *	IS-Required	Type *	Range
First Name	First Name	enter First Name	<input checked="" type="checkbox"/>	Text	
Last Name	Last Name	enter Last Name	<input checked="" type="checkbox"/>	Text	
DateTime	Date and Time	enter the Date/Time	<input checked="" type="checkbox"/>	DateTime	
Latitude	Latitude	if applicable enter the Latitude	<input type="checkbox"/>	Text	0to90
Longitude	Longitude	applicable enter the Longitude	<input type="checkbox"/>	Numeric	0to180

×

 Cancel

✓

 Submit

Text

Numeric

Email

Range

Phone

DateTime

# WHAT AN END USER SEES

← excel\_good\_example2.xlsx

OneShare [Change]

Metadata >

! Best Practices Check

✓ Quality Check

Citation

☐ I have read the Terms and Conditions

× Cancel

↑ Post

File Metadata

Field Metadata

Some repositories may require additional metadata to be persisted along with the data. Following section(s) cover required and optional fields that need to be filled in before submitting the file to the selected repository.

First Name \*



Last Name \*



Email \*



Date \*



Title \*



Keywords \*



Abstract \*



# METADATA IS REPOSITORY SPECIFIC

← excel\_good\_example2.xlsx

deptest\_sd\_imp [Change]

Metadata >

! Best Practices Check

✓ Quality Check

Citation

× Cancel

↑ Post

File Metadata

Field Metadata

Some repositories may require additional metadata to be persisted along with the data. Following section(s) cover required and optional fields that need to be filled in before submitting the file to the selected repository.

First Name \*



Last Name \*



Date and Time \*



Latitude



Values between 1,90

Longitude



Values between 1,180

Save

# BASIC TYPE VALIDATION

DATAUP

Welcome, Kristin Tolle | Sign Out

excel\_good\_example2.xlsx

deptest\_sd\_imp [ Change ]

Metadata >

Best Practices Check

Quality Check

Citation

Cancel

Post

File Metadata

Field Metadata

Some repositories may require additional metadata to be persisted along with the data. Following section(s) cover required and optional fields that need to be filled in before submitting the file to the selected repository.

First Name \*



Last Name \*



Date and Time \*

Please specify a valid date in the YYYY-MM-DD format.



Latitude



Please enter a value between 1 and 90.  
if applicable enter the Latitude



Values between 1,90



# BEST PRACTICES (XLSX FILES ONLY)

We have checked your data file for potential issues that might make your data less understandable to others. Some of the issues are related to test compatibility, and others are simply best practices for data management and organization.

Some of the potential issues listed here must be manually altered to comply with best practices. **Download** modified files to fix errors. You can Ignore these errors and continue posting.

deptest\_sd\_imp [ Change ]

Metadata  
[ Saved ]

Best Practices Check >

Quality Check

Citation

Cancel

Post

MetaData (4)

20120726 (2)

20120726INITIAL (2)

ECS\_20120726\_Calcurve (1)

Remove Selected

Remove Selected

	Potential Problem	Why this may be a problem	Suggested Remedy	Location(s)	
<input checked="" type="checkbox"/>	Merged cells	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.	B24:C24,	^
<input type="checkbox"/>	Commas	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)	C8,B24	^

# ONE PAGE PER SHEET

We have checked your data file for potential issues that might make your data less understandable to others. Some of the issues are related to test compatibility, and others are simply best practices for data management and organization.

Some of the potential issues listed here must be manually altered to comply with best practices. **Download** modified files to fix errors. You can Ignore these errors and continue posting.

deptest\_sd\_imp [ Change ]

Metadata  
[ Saved ]

Best Practices Check >

Quality Check

Citation

Cancel

Post

MetaData (2)

20120726 (2)


20120726INITIAL (2)

ECS\_20120726\_Calcurve (1)

Remove Selected

Potential Problem	Why this may be a problem	Suggested Remedy	Location(s)
<input type="checkbox"/> Commas	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)	C8,B24
<input type="checkbox"/> Special characters	Special characters may cause problems for other programs or may be modified upon export	Use alpha-numeric characters only. If needed, describe the symbol in a new column	B2,C4,C8,C13,C20,B24

# CREATING DATA VALIDATION RULES

Welcome, Sequim Test | [Settings](#) | [Sign Out](#)

Files

Repositories

Quality Checks

←

Add New Rule

Rule Name \*

Node Atlas

Description

Files that are processed in Node Atlas need a minimum of DateTime Latitude and Longitude in any order

Maximum 300 characters.

Header Names

☐ Consider headers order

	Header Name *	Type	Range Start	Range End	Actions
⬆	Latitude	Numeric	0	90	✕
⬆	Longitude	Numeric	0	180	✕
⬆		None			✕


Add Header Name

[Privacy Statement](#) | [Terms of Use](#) | [Contact](#) | [About Dataup](#)

Microsoft Research

Powered by Windows Azure

# DATA VALIDATION RULES

Welcome, Sequim Test | [Settings](#) | [Sign Out](#)










Files

Repositories



Quality Checks

Manage Quality Check Rules

+ Add New Rule

Rule Name	Description	Visibility	Date Created	Created By	Actions
Node Atlas	Files that are processed in Node Atlas need a minimum of DateTime Latitude and Longitude in any order		2/26/2014 2:08:47 PM	Sequim Test	 
QARule0212	QARule0212		2/11/2014 9:50:23 PM	Sequim Test	 
masspec			2/6/2014 11:06:13 AM	Sequim Test	 

[Privacy Statement](#) | [Terms of Use](#) | [Contact](#) | [About Dataup](#)





# QUALITY CHECK USER EXPERIENCE (ALL FILE TYPES)

← excel\_good\_example2.xlsx

deptest\_sd\_imp [Change]

Metadata

Best Practices Check

✓ Quality Check >

Citation

✕ Cancel

⬆ Post

You can validate your data against pre-configured rules. Select the rule, data on which the rule should be applied and click validate to see the results. The report will show you all the detected anomalies.

Rule Name

Node Atlas

Sheet Name

Select Sheets

✓ Validate

- ✓ **MetaData**
- ✓ 20120726
- ✓ 20120726INITIAL
- ✓ ECS\_20120726\_calcurve

# CITATION

← excel\_good\_example2.xlsx

OneShare [\[ Change \]](#)

Metadata

Best Practices Check

Quality Check

Citation >

☐ I have read the [Terms and Conditions](#)

Cancel

Post

Publication Year

2014

Title

excel\_good\_example2

Version

v1

×

Publisher

Tolle, Kristin

Save

Save

# TERMS AND CONDITIONS

← excel\_good\_example2.xlsx

OneShare [\[ Change \]](#)

Metadata

Best Practices Check

Quality Check

Citation



☒ I have read the Terms and Conditions

Cancel

Post

Publication Year

2014

Title

excel\_good\_example2

Version

v1

Publisher

Tolle, Kristin

Save

# POSTED

## My Files

View All ▾ Date Uploaded ▾

Post Download Delete

+ Add a New File

OR

Drag & Drop Files Here

Ireland Water Sa... .csv

Expires in 65hrs.

excel\_good\_exam... .xlsx

deptest\_sd\_imp

Ferns.csv

Expires in 65hrs.

Tibet6.csv

Expires in 42hrs.

Tibet5.csv

Expires in 42hrs.

File Name

excel\_good\_example2

Repository Name

deptest\_sd\_imp

File Extension

xlsx

File Size

45.92 KB

Uploaded Date

2/26/2014 13:36:3

Status

Inqueue

Tibet4.csv

Expires in 42hrs.

Tibet3.csv

Expires in 42hrs.

Tibet2.csv

Expires in 42hrs.

Tibet1.csv

Expires in 42hrs.

Masterlist.csv

Expires in 42hrs.

Configuration.csv

Expires in 42hrs.

Chicago Narcotics.xlsx

Earthquake Data f... .xlsx



# WHAT HAPPENS NEXT?

- Posted or not, the actual file is deleted from the Azure instance to reduce storage costs.
- Timing is shown so users can see where they are in the “timeline” to delete
- 24 hours to delete, the expiration time is bolder and in red (unless posted)
- If posted the file remains visible in the file store and users can still download that version (from the repository) including the citation and metadata

QUESTIONS??