

# **CEDA Data Quick Start User Guide**

# **Dataset Collection:**

Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (1853-current) http://catalogue.ceda.ac.uk/uuid/220a65615218d5c9cc9e4785a3234bd0

02/04/2024

Version 1.5

# Introduction

This document aims to give additional information to aid use of the data within the full MIDAS land and marine surface station data to accompany information already available within the CEDA data catalogue pages within this collection. This document gives further information including:

- How Station Data are connected in MIDAS (Station source id)
- Extracting data into columns in Excel
- State indicators
- Quality control (QC) guide
- Met element name \_j

For information about accessing data, geographic and temporal coverage, please refer to the MIDAS data collection catalogue and the related individual dataset pages using the link above.

# How Station Data are connected in MIDAS (Station source id)

The Met Office collates data from a number of observation networks both in the UK and internationally for various purposes, including storing these data in their long term database – MIDAS. As a station may issue meteorological data within a variety of networks, within which it may have a different identifier (e.g. WMO station ID), the Met Office allocate a specific station ID for all data for that station – known as the src\_id. The Met Office also allocates an Id type for identifying a recording instrument and a message type giving an indication of the range observations that are recorded.

**Source id** (src\_id) is an integer running from 1 upwards unique to all stations. All observations in the database are stored with their source identifier.

**Id\_type** identifies a recording instrument, e.g. a raingauge (RAIN), an anemometer (WIND), a range of weather recording instruments at a UK Met Office Station (DCNN), a range of recording instruments at an airport (ICAO), a range of recording instruments at a WMO station (WMO), etc.

**Met Domain or message types** give an indication of the range of observations that can potentially be recorded at the station. For example, WADRAIN is the message type or Met Domain which will be used to send out daily rainfall data as observed at the station of interest from a station operated by a Water Authority (hence the WA prefix). The distinctive message types indicate that the stations are reporting within different networks and each may include different parameter set.

This document shows where all the elements from the SYNOP, NCM, HCM, SREW, METAR and DLY3208 (daily climate) message types are stored in MIDAS. One notable exception is the marine met domains (SHIP, RIG etc.) - all of these met domains have their data stored in the marine\_ob (and possibly marine\_ice\_ob) table only.

# Finding relevant stations - MIDAS Station Search Tool

To help find a relevant station or to see detailed station details within MIDAS, CEDA have produced the MIDAS Station Search tool:

# https://archive.ceda.ac.uk/tools/midas\_stations

This allows a user to search for a station by name, postcode, county or station 'src\_id'. Additionally, the service links to an interactive map and Google Earth files to give additional search options.. The table below shows the strengths and weaknesses of using each method.

<b>•</b> • • • • •							
CEDA Archive	Search Catalogue	Get Data	Help	Tools	Deposit	My Account	News
Home / Search for Met Office MIDAS statio	ns						
Search for Met	Office	Δ	Sst	atio	าร		
Manc			550	atioi	15		
View stations on interactive map Also availab	le as a Google Earth down	load.					
Soarch for station par							
Finds all station names contain the given string	The string is matched any	where within the	station nam	e and is not c	ase sensitive.		
Search for:	Search Reset						
Open within year range: to: Or	show current stations only	y 🗉					
Search for stations by Finds all stations with postcodes starting with t	postcode he given string. Enter up to	o 5 significant ch	aracters. For	example. "O>	". "OX11" or "O	X11 5"). Most stati	ons do not have the
full postcode recorded, so adding more charact	ers is unlikely to return an	y matches. See P	ostcode area	map or List o	f postal areas fo	or more information	on postcodes.
Search for: Search Reset							
Open within year range: to: Or	show current stations only	y 🗆					
Search for stations by	LIK county						
Select a county to display all stations with that	county. Please note that th	e county names	are the name	s used by the	Met Office and	may not reflect the	current county
divisions within the UK. Map of traditional cour	ties						
ABERDEENSHIRE	Search Reset						
Open within year range: to: Or	show current stations only	y 🗆					
Display station details							
To display details for one or more stations, ente	r the station source ID (sro	c_id) values in the	box below. U	Jse spaces or	commas to sep	arate multiple value	<del>1</del> 5.
Source IDs:	Search Reset						

Search option	Strengths	Weaknesses
Name	Likely to match against a name for	Some stations may be given a different
	the area	name within MIDAS than would be
		expected.
Postcode	All the stations in the postcode area	This will miss nearby stations in
	will appear in the options list.	neighbouring postcode areas
County	All the stations in the county will	The Met Office has used historic county
	appear in the options list.	borders, therefore places may appear in
		different counties. The link below
		shows the county boundaries used.
		https://archive.ceda.ac.uk/tools/traditio
		nal_counties_map/l. Limited to UK
		coverage only.

Interactive Map	All the stations in the location can be seen on a map. Stations can be searched by both message types and geographical areas. Non-UK stations also displayed.	The stations will not appear in a list, they will appear on a map.
Google Earth	All the stations in a location can be seen on a map. The stations can be searched by message type or geographical area seen on the map in different colours. Non-UK stations also displayed.	Files have to be downloaded according to your search requirements and required Google Earth. The stations will not appear in a list, they will appear on a map.

# Please note that not every station reports the same types of data as the instrument in operation will vary from site to site. The MIDAS system will bring together all the data from a particular station regardless of which network it is reporting on.

Once the relevant station has been found, click on it to display more information. The record below shows the information about Birmingham University (src\_id: 56950).

A list of the different message types and descriptions can be found in the link below.

CEDA Archiv	'e	arch Catalogue	Get Data	Help	Tools	Deposit	My Account	News
			New searc	h				
Name: src_id Geographic area: Latitude (decimal degrees): Longitude (decimal degrees): Grid ref: Grid ref: Grid ref: Postcode: Elevation: Drainage stream: Hydrological area ID: Station start date	BIRMINGHAM UNIVERS 56950 WEST MIDLANDS 52.4806 (WCS 84 value: 52 - 1.90493 (WCS 84 value: 52 - 1.90493 (WCS 84 value: -1 SP 064870 (Easting: 40645 OS B1 139 meters Not specified 280 2011-06-06 2015-01-07	ITY (Locate on: Bing 4809) .90637) 6 Northing: 287007)	Googlemap St	reetmap Gooj	<mark>le Eart</mark> h (Requi	es Google Earth si	yftware)) View all ste postcode	tions in B1
Measuremen	its made		Massage	tout data		Mar	unano and data	
CLBD 4407	AWSHRLY		2011-06-0	)6		201	1-06-06	
CLBD 4407	AWSDLY		2011-06-0	06		201	1-06-06	
CLBN 440701	AWSHRLY		2011-06-0	)6		201	1-06-06	
CLBR 095142	AWSHRLY		2011-06-0	06		201	1-06-06	
CLBR 095142	AWSDLY		2011-06-0	06		201	1-06-06	
CLBW 99209	AWSDLY		2011-06-0	)6		201	1-06-06	
CLBW 99209	AWSHRLY		2011-06-0	06		201	1-06-06	
Remarks								

### http://artefacts.ceda.ac.uk/badc\_datadocs/ukmo-midas/met\_domain.html

The message types are important as they represent each type of observation and measurements that are taken at the different stations. For example, Birmingham University (shown above) has message

type AWSHRLY, which means it is an Automatic Weather Station recording hourly rainfall, temperature and weather observations.

# **Extracting data for stations**

The MIDAS data can be downloaded directly as yearly files, but as these files contain all station data for the given type (e.g. these files are too large for most users therefore a guide has been produced. To extract the data CEDA has developed a tool within the CEDA Web Processing Service (WPS), by using this link below, the MIDAS data can be extracted for one or more weather stations at a time. **This process is strongly advised as there are data from a large number of stations within the MIDAS collection.** 

# https://ceda-wps-ui.ceda.ac.uk/

Guidelines on how to use the CEDA WPS to extract MIDAS data is available at:

### https://help.ceda.ac.uk/article/4977-ceda-wps

NOTE: to use the MIDAS tools in the CEDA WPS service you must first have access to the full MIDAS dataset collection. Once you have this you can use the CEDA WPS service as follows:

#### A step by step guide

1. Login in



This will take you to a page asking you to log in with your CEDA user account credentials and agree to sharing information with the CEDA WPS service.

2. Select 'Processes' from the top menu bar

Processes Choose one of the pinned processes or a Web Processing Service.

Available Web Processing Services	
<ul> <li>Section 2 Compliance Checker ■</li> <li>A WPS for checking file compliance with standards such as CF</li> </ul>	¢°7 ★ 11
☆ Data Subsetter ▲ A WPS for subsetting data sets such as the CRU Time Series and HadUK-Grid	¢°7 ★ 11
✿ MIDAS Extract ▲ A WPS which supports filtering and extraction of MIDAS data.	¢°7 ★ 11
* NAME Model A swallow is a WPS to run the Met Office NAME model.	¢°7 🖈 11

Powered by Birdhouse | Get the code on GitHub | Version v0.11.0

3. Click on MIDAS Extract option

You will be presented with a range of options below the 'MIDAS Extract' option. To extract data use either the 'Extract UK Station Data' or 'Extract UK Station Data with Data Input' option (this one uses a text field for start/end date input as opposed to a sliding bar in the first option that some users may prefer).

The 'Get Weather Stations' option provides a route to find potential stations to use only based on various search criteria. The output of this tool may be used to feed into the data extraction tools if desired.

TD	Daily temperature data
WD	Daily weather data
RD	Daily rainfall data
RH	Hourly rainfall data
RS	Sub-hourly rainfall data (up to 31st March 2005)
ST	Soil temperature data
WН	Hourly weather data
WM	Mean-wind data
RO	(solar) Radiation Observation data

4. Fill in the following fields with the appropriate parameters. The 'Observations tables' are:

5. It is optional to fill in bounding box, counties or station source ids, but one option must be filled in as highlighted in the diagram below.

טו	
	¥
e name of the database tab	le used in the MIDAS database to identify a particular selection of weather observations.
te Range	
-01-01	2024-02-
e date range to search for s	station data.
ounding Box	
	North 61.00
West -12.00	East 3.00
	South 49.00
Island	Norma
•	Finland
3	Sverige
	United Kingdom
	Беларусь
	Deutschland
	Украї
	France România
10 °	France România
10 *	France România
200m out Zoom in F	România © CoercitientMas contributes. Reset zoom
10 ° Zoom out Zoom in F ease select a valid bounding	Prance România © © Constituente contributer. Reset zoom 3 box within the following geographical boundaries: northern extent: 61, southern extent: 49, eastern extent: -12,
Zoom out Zoom in F ease select a valid bounding estern extent: 3.	România e e contenente outebook Reset zoom g box within the following geographical boundaries: northern extent: 61, southern extent: 49, eastern extent: -12,
Zoom out Zoom in F ease select a valid bounding estern extent: 3. e spatial bounding box with	Prance România Peset zoom g box within the following geographical boundaries: northern extent: 61, southern extent: 49, eastern extent: -12, nin which to search for station data. This input will be ignored if counties are provided.
10 °       Zoom out     Zoom in       F       ease select a valid bounding estern extent: 3.       ue spatial bounding box with bounties	Prance România P © Constituent a contributer. Reset zoom g box within the following geographical boundaries: northern extent: 61, southern extent: 49, eastern extent: -12, nin which to search for station data. This input will be ignored if counties are provided.
10 *       Zoom out     Zoom in       F       ease select a valid bounding estern extent: 3.       ne spatial bounding box with punties       ABERDEENSHIRE	France România © © ConstituentAles contributes. Reset zoom g box within the following geographical boundaries: northern extent: 61, southern extent: 49, eastern extent: -12, nin which to search for station data. This input will be ignored if counties are provided.
10*       Zoom out     Zoom in       F       ease select a valid bounding estern extent: 3.       ie spatial bounding box with <b>sunties</b> ABERDEENSHIRE ALDERNEY	Prance România © © ConstituentAle outburn. Reset zoom g box within the following geographical boundaries: northern extent: 61, southern extent: 49, eastern extent: -12, nin which to search for station data. This input will be ignored if counties are provided.
10 °       Zoom out     Zoom in       Passe select a valid bounding estern extent: 3.       ie spatial bounding box with punties       ABERDEENSHIRE ALDERNEY ANGUS ANTONA	Prance România Peset zoom g box within the following geographical boundaries: northern extent: 61, southern extent: 49, eastern extent: -12, n'n which to search for station data. This input will be ignored if counties are provided.
10 *       Zoom out     Zoom in       F       ease select a valid bounding estern extent: 3.       te spatial bounding box with <b>punties</b> ABERDEENSHIRE ALDERNEY ANGUS ANTRIM	France România © © Constituent la contributer. Reset zoom g box within the following geographical boundaries: northern extent: 61, southern extent: 49, eastern extent: -12, nin which to search for station data. This input will be ignored if counties are provided.
10 °       Zoom out     Zoom in       F       ease select a valid bounding setern extent: 3.       e spatial bounding box with bounties       ABERDEENSHIRE ALDERNEY ANGUS ANTRIM       ADDERLY       ANTRIM       ADDOLL       ADDOLL       ADDOLL	France România e e Constructive ouverburn. Reset zoom g box within the following geographical boundaries: northern extent: 61, southern extent: 49, eastern extent: -12, nin which to search for station data. This input will be ignored if counties are provided.
10 *       Zoom out     Zoom in       F       aase select a valid bounding estern extent: 3.       e spatial bounding box with punties       ABERDEENSHIRE ALDERNEY ANGUS ANTRIM       ADOXL_COLLEGE within which ist of counties within which ation Source IDs	France       Romania         Reset zoom       Reset zoom         g box within the following geographical boundaries: northern extent: 61, southern extent: 49, eastern extent: -12, nin which to search for station data. This input will be ignored if counties are provided.         nin which to search for station data. This input will be ignored if counties are provided.         to search for weather stations.
200 out Zoom in F ease select a valid bounding estern extent: 3. e spatial bounding box with punties ABERDEENSHIRE ALDERNEY ANGUS ANTRIM AROUS ANTRIM ADOUT COUNTIES within which ation Source IDS	France       Romania         Reset zoom       g box within the following geographical boundaries: northern extent: 61, southern extent: 49, eastern extent: -12, nin which to search for station data. This input will be ignored if counties are provided.         In which to search for station data. This input will be ignored if counties are provided.         It is search for weather stations.
10 °       Zoom out     Zoom in       F       base select a valid bounding stern extent: 3.       e spatial bounding box with punties       ABERDEENSHIRE ALDERNEY ANGUS ANTRIM       ADOVEL (ALLIGUE AND DE ist of counties within which ation Source IDs       ist of weather stations source	France       Romania         Peset zoom       e o Constituent/Lee courbume.         Reset zoom       g box within the following geographical boundaries: northern extent: 61, southern extent: 49, eastern extent: -12, nin which to search for station data. This input will be ignored if counties are provided.         In which to search for station data. This input will be ignored if counties are provided.         In which to search for weather stations.         In to search for weather stations.         In the ignored if an input job ID is provided.
10 *       Zoom out     Zoom in       F       aase select a valid bounding estern extent: 3.       e spatial bounding box with punties       ABERDEENSHIRE ALDERNEY ANGUS ANTRIM       ADOXL_COLLEGE within which ation Source IDs       ist of weather stations source sut Job Id	France       Romania         Reset zoom       Reset zoom         g box within the following geographical boundaries: northern extent: 61, southern extent: 49, eastern extent: -12, nin which to search for station data. This input will be ignored if counties are provided.         nin which to search for station data. This input will be ignored if counties are provided.         to search for weather stations.         ce IDs. This input will be ignored if an input job ID is provided.
10 *       Zoom out     Zoom in       F       ease select a valid bounding estern extent: 3.       e spatial bounding box with punties       ABERDEENSHIRE ALDERNEY ANGUS       ANGUS       ANGUS       ANGUS       ANGUS       ANGUS       Stor of counties within which ation Source IDs       ist of weather stations source put Job Id	France       Romania         Reset zoom       Reset zoom         g box within the following geographical boundaries: northern extent: 61, southern extent: 49, eastern extent: -12, nin which to search for station data. This input will be ignored if counties are provided.         nin which to search for station data. This input will be ignored if counties are provided.         to search for weather stations.         ce IDs. This input will be ignored if an input job ID is provided.

6. Select your desired delimiter and press submit. The job will then be queued to run asynchronous and the following screen will show you current progress on the Job Monitor page

Delimiter		
comma		~
The delimiter to be used in the output	files.	
	Submit	

You can choose to remain on the page for the job to complete to close the screen and await an email notification when the job has been completed.

7. Once completed, on the Job Monitor page you can select the 'Details' button next to your completed job to see the status and download associated files

Data extracted	Webpage to the relevant column headings
UK Daily Temperature	
	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo
	<u>-midas/TD_Table.html</u>
UK Soil Temperature	
	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo
	<u>-midas/ST_Table.html</u>
UK Daily Rainfall data	
	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo
	-midas/RD_lable.html
UK Daily Weather Observation data	
	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo
	- <u>midas/wD_lable.ntml</u>
OK Houriy Rainfall data	http://artafacta.coda.co.uk/hada.datadaca/ukma
	midas/PH Table html
LIK Hourly Weather Observation data	
	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo
	-midas/WH_Table html
LIK Mean Wind data	
	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo
	-midas/WM Table.html
UK Soil Minimum Temperatures (1959-1970 only)	
	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo
	-midas/TMSL_Table.html
UK Sub- hourly Rainfall data only to April 2005	
	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo
	-midas/RS_Table.html
Global Radiation Observations	
	http://artefacts.ceda.ac.uk/badc_datadocs/ukmo
	<u>-midas/RO_Table.html</u>

These links below define the column headings depending on which data has been extracted.

# Handling extracted data

# Extracting data into columns in Excel

The link below demonstrates how to convert a text file to Excel

- 1. Highlight all the data
- 2. Click on the data tab and into Text to columns

C	) 🖬 🤊	- (° - )	) =	last and a second		-	ĥ			_		ten
C	Home	Inse	rt Page	Layout Fo	ormulas	Data	Reviev	N Y	View			
采 Fr Act	om From vess Web	From F Text	From Other Sources *	Existing Connections	Refresh All *	Diagonal Conn Prope Se Edit L	ections erties inks	A Z↓ Z↓	AZA Sort	Filter	K Clear Reapply Advanced	Text to F Columns Du
		Get Exte	rnal Data		C	onnection	IS		So	rt & Fil	ter	
	A1		- (	$f_{x}$								
							А					
1												
2	2014-03-01	09:00, 0	DCNN, 302	7, 12, 1, AW	5DLY, 30	476, 101	1, 2.8, -	1.2, -3	.2, , 6, 6	6, 6, , 2	2014-03-01 09:	03, 0, , , ,
3	2014-03-01	21:00, 0	DCNN, 302	7, 12, 1, AW	5DLY, 30	476, 101	1, 7.4, -	1.1, , ,	6, 6, , ,	2014-	03-01 21:04, 0	
4	014-03-02 (	09:00, D	CNN, 3027	, 12, 1, AWS	DLY, 304	76, 1011	, 5.9, -1.	.3, -5.8	8, , 6, 6,	6, , 20	014-03-02 09:0	6, 0, , , ,
5	2014-03-02	21:00, 0	OCNN, 302	7, 12, 1, AW	5DLY, 30	476, 101	1, 9.4, 5	.9, , ,	6, 6, , ,	2014-0	03-02 21:03, 0,	
6	2014-03-03	00.00 F	CNN 302	7 12 1 ΔΜ		476 101	1 6 9 3	0.3	6 6 6	201	4-03-03-00-03	0

3. Choose the 'Delineated' option

e Text Wizard has determined that your data is Fixed Width.	
this is correct, choose Next, or choose the data type that best describes your data	a.
Driginal data type	
Choose the file type that best describes your data:	
<u>Delimited</u> - Characters such as commas or tabs separate each field.	
Fixed width - Fields are aligned in columns with spaces between each field	
Preview of selected data:	
Preview of selected data:	A
Preview of selected data: 1 2 2014-03-01 09:00, DCNN, 3027, 12, 1, AWSDLY, 30476, 101 3 2014-03-01 21:00, DCNN, 3027, 12, 1, AWSDLY, 30476, 101	1, 2.8,
Preview of selected data: 1 2 2014-03-01 09:00, DCNN, 3027, 12, 1, AWSDLY, 30476, 101 3 2014-03-01 21:00, DCNN, 3027, 12, 1, AWSDLY, 30476, 101 4 014-03-02 09:00, DCNN, 3027, 12, 1, AWSDLY, 30476, 1013	1, 2.8, 1, 7.4, , 5.9, -
Preview of selected data: 1 2 2014-03-01 09:00, DCNN, 3027, 12, 1, AWSDLY, 30476, 101 3 2014-03-01 21:00, DCNN, 3027, 12, 1, AWSDLY, 30476, 101 4 014-03-02 09:00, DCNN, 3027, 12, 1, AWSDLY, 30476, 101 5 2014-03-02 21:00, DCNN, 3027, 12, 1, AWSDLY, 30476, 101	1, 2.8, 1, 7.4, 1, 5.9, - 1, 9.4,
Preview of selected data: 1 2 2014-03-01 09:00, DCNN, 3027, 12, 1, AWSDLY, 30476, 101 3 2014-03-01 21:00, DCNN, 3027, 12, 1, AWSDLY, 30476, 101 4 014-03-02 09:00, DCNN, 3027, 12, 1, AWSDLY, 30476, 101 5 2014-03-02 21:00, DCNN, 3027, 12, 1, AWSDLY, 30476, 101 4 III	1, 2.8, 1, 7.4, ., 5.9, - 1, 9.4,

4. Choose a delimiter to be 'comma'

Convert Text to Columns	Wizard	- Step 2	of 3	-				8	63
This screen lets you set the the preview below.	e delimite	rs your o	lata co	ntains.	You can see	e how you	r text is a	affected	d in
Delimiters									
Tab									
Semicolon	] T <u>r</u> eat o	onsecutiv	ve delin	iters a	s one				
Comma									
Space Ie	xt <u>q</u> ualifi	ier:			•				
Other:									
Data preview									
									<b>^</b>
2014-03-01 09:00	DCNN	3027	12	1	AWSDLY	30476	1011	2.8	
2014-03-01 21:00	DCNN	3027	12	1	AWSDLY	30476	1011	5 9	
2014-03-02 21:00	DCNN	3027	12	1	AWSDLY	30476	1011	9.4	-
۰ III								+	
	ſ	Cano	el	<	Back	Next >		Einis	h
				<u> </u>					

# 5. Select finish

is select lets you select	each col	umn and	set the	Data P	Format.				
Column data format									
General		10			maria unh ma	to our house			_
General converts numeric values to numbers, date values to dates, and all remaining values to text.						.0			
Date: DMY				Advanc	ced				
Do not import column	n (skip)								
estination: \$4\$1									
Data preview									
Data preview									
Data preview	Genera	Genera	Gener	Gene	rGeneral	General	Genera	Gener	
General	Genera	Genera	Genez	Gene	rGeneral	General	Genera	Gener	<u> </u>
Ceneral	Genera DCNN	Genera 3027	Gener	Gene:	rGeneral AWSDLY	General 30476	Genera 1011	Gener 2.8	Ĺ
Part         Part           Peneral         2014-03-01         09:00           2014-03-01         21:00         014-03-02         09:00	Senera DCNN DCNN DCNN	Genera 3027 3027 3027	Gener 12 12 12	Gene 1 1	rGeneral AWSDLY AWSDLY AWSDLY	General 30476 30476 30476	Genera 1011 1011 1011	Gener 2.8 7.4 5.9	Â
Ceneral 2014-03-01 09:00 2014-03-01 21:00 014-03-02 09:00 2014-03-02 21:00	Genera DCNN DCNN DCNN DCNN	Genera 3027 3027 3027 3027 3027	Gener 12 12 12 12	Gene: 1 1 1 1	AWSDLY AWSDLY AWSDLY AWSDLY AWSDLY	General 30476 30476 30476 30476 30476	Genera 1011 1011 1011 1011	Gener 2.8 7.4 5.9 9.4	Ĥ
Ceneral           2014-03-01         09:00           2014-03-01         21:00           014-03-02         09:00           2014-03-02         21:00	Genera DCNN DCNN DCNN DCNN DCNN	Genera 3027 3027 3027 3027 3027	Gener 12 12 12 12	Sene 1 1 1 1	rGeneral AWSDLY AWSDLY AWSDLY AWSDLY	General 30476 30476 30476 30476 30476	Cenera 1011 1011 1011 1011	Cener 2.8 7.4 5.9 9.4	^ -
Conta greview Ceneral 2014-03-01 09:00 2014-03-01 21:00 014-03-02 09:00 2014-03-02 21:00 ∢ III	Genera DCNN DCNN DCNN DCNN	Genera 3027 3027 3027 3027	Sener 12 12 12 12	Sene: 1 1 1 1	rGeneral AWSDLY AWSDLY AWSDLY AWSDLY	General 30476 30476 30476 30476	Senera 1011 1011 1011 1011	Gener 2.8 7.4 5.9 9.4	

# **Cleaning up the data**

At first glance there may be more than 1 entry for a given dataline for a given station's report type and timestamp. This is due to a couple of factors which need to be accounted for:

- 1. When data are initially received by the Met Office's system they are noted as 'version 1' of the data
- 2. Subsequently, if the entry is to be updated, e.g. following quality control, the *original* line is retained with a version number of zero and the new entry becomes the new version 1 line.
- 3. Within the Met Office's system subsequent changes to the data line will then replace the version 1 entry only, meaning that intermediate changes are not stored by the MIDAS system.
- 4. However, as CEDA's copy is obtained by periodic snapshots extracting the previous 12 months' worth of data and only entirely duplicate lines are removed, it is possible that the intermediate version 1 record states are captured where changes are more than a month apart.

Consequently, further steps should be undertaken to try and prepare the data ahead of any analysis and to understand the limitations of the data presented. It may also be possible that true duplicate lines may have appeared in the archive.

### Step 1 - remove 'duplicate' lines:

To remove the duplicates select all the data and select Data tab, in the Data Tools group, click Remove Duplicates (Highlighted in yellow).

0	) H "	• (4 •	)•	1	temperature march 2014.xlsx - M crosoft Excel							
6	Home	Ins	ert Pag	ge Layout	Formulas Data Re							
*	2 🔁			Connections		K Clear		3	🧯 😰	•	1	
- Fri Au	on From Less Web	From From From Other Existing Refresh All - Se Edit Links		Z Sort Fille	Advanced	Text It Remove Column's Duplicates V	Dala	Consolidate What-If Analysis *	Group U	ngroup		
		Get Ext	ernal Data		Connections	Sort &	D			(		
	A3	A3 🔻 🔄 🎉 01/03/2			/03/2014 21:00:00					$\checkmark$		
	A		A B C		D	E		F G		н		
1	Date		id type	id	observation hour count	version number	met domain na	me (message type)	Station id	state indicator for	the record	Maxi
2	01/03/201	4 09:00	DONN	3027		12 1	1 AWSDLY		30476		1011	L
3	01/03/201	4 21:00	DCNN	3027		12 1	AWSDLY		30176		1011	L
4	014-03-02	00-00	DONN	3027	_	10 1	AWCDIM		20475		1011	

- 1. Do one or more of the following:
  - Under Columns, select one or more columns.
  - To quickly select all columns, click Select All.
  - To quickly clear all columns, click Unselect All.

To delete duplicate values, select on	e or more columns that contain duplicates.	
≸≣ Select All 3≣ Unselect /	My data has heade	s
Columns		*
01/03/2014 09:00		
V DONN		Ξ
3027		
12		
V 1		
AWSDLY		
☑ 30 <del>1</del> 76		
1011		-
E		-

2. Click OK. A message is displayed indicating how many duplicate values were removed and how many unique values remain, or if no duplicate values were removed. Click OK.

# Step 2 - Version Number and Quality control

The data arrives at the Met Office and it undergoes quality control to check that the data are correct and consistent with the surrounding data points.

As noted above, 'version 1' data should contain the latests state of the data line as held within the MIDAS system, thus removing version 0 lines will ensure that the original values, i.e. before any changes have been applied, are removed. However, this may still result in more than one 'version 1' entry. Here the record time stamp should act as a sufficient delineator to find the latest entry.

The Met Office MIDAS system undertakes systematic quality control checks on the data automatically as well as occasional manual alterations. For each data column in the data there should be an associated QC entry - denoted by the \_q and \_j values. Use of these flags can help to remove unwanted datalines that would otherwise contaminate data analysis. However, the values to use may not be immediately obvious:

### Quality control (QC) guide

Quality control is when the data are checked to make sure it is correct and consistent with the surrounding data points.

These data are displayed with a QC code composed of up to 5 digits e.g. 1006, 15006 or 6

Each of the 5 digits represents an entry in the 'MESQL' quality control flagging entries, filled in from the right. The letters 'MESQL' stand for the following types of QC information:

M- marker E- estimate S- status Q-query L-level

Not all types of QC entries are required and, as stated above, the value should be read from the RIGHT. For example:

Exam	ple 1			
М	Е	S	Q	L
	1	0	0	6

Exan	nple 2	2		
Μ	E	S	Q	L
				6

The numbers represent different information for each of the QC types and can be found in detail on teh following page:

https://data.ceda.ac.uk/badc/ukmo-midas/metadata/doc/QC J flags.html

For example, a QC of 1006 has the following values from the QC codes:

For example									
Μ	E	S	Q	L					
	1	0	0	6					
E (1) - Estimate/correction derived automatically from a program with no manual intervention									
S (0)-Observed and not suspect									
<b>Q (0)</b> -Original value is/was not queried, or no information available									
L (6)-Final (or only) areal or buddy job run and queries processed									

### Met element name \_j

This attribute is a single character code which either describes the method of measurement, or further qualifies the meteorological values. The meaning of any value depends on the element being qualified.

#### For example

This data below is taken from the UK Daily rainfall data

005349, RAIN, 1920-01-01 00:00, 1, DLY3208, 900, 1, 310, 1001, 14, , 0, , , D

This therefore means we need to look under the Precipitation codes. The **D** therefore implies the rainfall has been converted from inches.

Full listing is available in the link below <a href="https://data.ceda.ac.uk/badc/ukmo-midas/metadata/doc/QC\_J\_flags.html">https://data.ceda.ac.uk/badc/ukmo-midas/metadata/doc/QC\_J\_flags.html</a>:

# **State Indicators**

A state indicator is an attribute (rec\_st\_ind) of each table that is used to describe the current stage in the life of a particular record, from creation to deletion.

Full listing is available in the link below: <a href="https://data.ceda.ac.uk/badc/ukmo-midas/metadata/doc/state\_indicators.html">https://data.ceda.ac.uk/badc/ukmo-midas/metadata/doc/state\_indicators.html</a>