



TECHNICAL TIPS AND TRICKS

A TIP FOR ANALYSING AND UNDERSTANDING DATA COLLECTIONS

(...or any other great mess of technical data, sitting on your system.)

In this game, we deal with increasingly huge collections of files, usually in mountains of folders spread far and wide.

Large projects, or a large number of projects lead invariably to large numbers of files and folders on our systems. Because we often work with others on these projects, there will be copies and/or different versions of files in different parts of the data collection.

*How do **you** approach the sorting and 'distillation' of large data collections such as these?*

THE PROBLEM

How often have you –

- Been provided with a **significant 'data warehouse' for a project**, consisting sometimes of hundreds or thousands of files, that you need to digest in order to provide an analysis, evaluation, an audit / IGR / JORC compliance report, a recommendation?
- Been handed a similar data collection that you had to understand, clean up, sort and validate for the purpose of **adding it to a database**?
- Wanted to compile a list of all the files in specific directories to **construct a data package**, for documentation or as a basis for work lists?
- Needed to pull together all the **files of a particular type**, for example core photos or lab results documents that were spread far and wide in a multitude of sub-folders?
- Lost a file or directory that you just **know** was somewhere on your huge file system?
- Needed to **clean up your data system** to remove duplicates, temporary, old or intermediate work files? (Particularly the really big ones!)

What we need then, is a usable list of all the files or directories under a specific point on our file systems that we can search, sort, filter or print.

Or, more powerfully, use as a basis for creating a script that will copy, move, rename, delete or perform some other operation on our list of selected files.

APPROACHES

There are many different approaches to sieving through the data files and directories, and I dare say you each have your own.

COLLECT - MANAGE - ANALYSE

But hard experience has taught me that if I can **create a list of all of the files, with their directory path, size, created and modified dates, and the file type**, and use this in (say) Excel, to keep track of what I have looked at, cleaned up, identified, ranked etc., it makes life a lot easier.

Unfortunately, in Windows Explorer, there is no easy way to create such a list. For the masochistic amongst us, it is possible to use a 'DOS' command window, and execute a command such as –

“DIR /W /S > some_text_file.txt”

- which will create a messy list of the files and directories for us to clean up.

Which is why we have created a neat little utility for our 'GeoUte' collection that does just that. It's called, uList.



Give it a starting point, and an output filename, and it will generate a list of all the files and directories under that point in CSV format.

Open the file in Excel, and you can start searching, sorting, annotating and identifying until your heart's content.

As you sort through the piles of files, you can track what has been done, where the important files are and so forth.

	A	B	C	D	E	F	G	H	I
1	Level	File Name	Root	File Type	Full Path	Resolved	Size	Created	Modified
2	1	add_rep.ins	V:\Porgera\Ddhs	.ins	V:\Porgera\Ddhs	\\OB1\Clie	61,484.00	13/10/2005 9:49	4/02/1997 8:31
3	1	ard_blasthole.xls	V:\Porgera\Ddhs	.xls	V:\Porgera\Ddhs	\\OB1\Clie	720,384.00	13/10/2005 9:49	18/02/2002 9:50
4	1	ard_blasthole.zip	V:\Porgera\Ddhs	.zip	V:\Porgera\Ddhs	\\OB1\Clie	196,241.00	13/10/2005 9:49	18/02/2002 9:52
5	1	ard_by_lithology.xls	V:\Porgera\Ddhs	.xls	V:\Porgera\Ddhs	\\OB1\Clie	2,319,872.00	13/10/2005 9:49	18/02/2002 8:34
6	1	ard_by_lithology.zip	V:\Porgera\Ddhs	.zip	V:\Porgera\Ddhs	\\OB1\Clie	598,208.00	13/10/2005 9:49	18/02/2002 8:35
7	1	coarse_rejects_notes.doc	V:\Porgera\Ddhs	.doc	V:\Porgera\Ddhs	\\OB1\Clie	69,632.00	1/06/2006 3:41	1/06/2006 3:50
8	1	CompletionList_Wantok_Marvin.doc	V:\Porgera\Ddhs	.doc	V:\Porgera\Ddhs	\\OB1\Clie	67,584.00	26/05/2006 3:58	26/05/2006 4:51
9	1	ddhs2.spc	V:\Porgera\Ddhs	.spc	V:\Porgera\Ddhs	\\OB1\Clie	741.00	13/10/2005 9:49	4/12/2001 7:58
10	1	ddhs_get_all.xls	V:\Porgera\Ddhs	.xls	V:\Porgera\Ddhs	\\OB1\Clie	1,758,720.00	13/10/2005 9:49	4/11/2001 11:55
11	1	ddhs_get_test.xls	V:\Porgera\Ddhs	.xls	V:\Porgera\Ddhs	\\OB1\Clie	959,488.00	13/10/2005 9:49	12/05/2001 6:14
12	1	ddhs_prd_060307.zip	V:\Porgera\Ddhs	.zip	V:\Porgera\Ddhs	\\OB1\Clie	84,223,981.00	15/03/2011 5:34	7/03/2006 11:52
13	1	ddhs_prd_060307a.zip	V:\Porgera\Ddhs	.zip	V:\Porgera\Ddhs	\\OB1\Clie	101,404,816.00	15/03/2011 5:42	7/03/2006 11:39
14	1	ddhs_sql_060307.zip	V:\Porgera\Ddhs	.zip	V:\Porgera\Ddhs	\\OB1\Clie	84,477,542.00	15/03/2011 5:34	7/03/2006 11:53
15	1	ddhs_sql_060307a.zip	V:\Porgera\Ddhs	.zip	V:\Porgera\Ddhs	\\OB1\Clie	98,632,560.00	15/03/2011 5:42	7/03/2006 11:38
16	1	install.ins	V:\Porgera\Ddhs	.ins	V:\Porgera\Ddhs	\\OB1\Clie	59,799.00	13/10/2005 9:49	4/02/1997 1:43
17	1	install2.ins	V:\Porgera\Ddhs	.ins	V:\Porgera\Ddhs	\\OB1\Clie	65,530.00	13/10/2005 9:49	11/02/1997 12:15
18	1	Porgera.vsd	V:\Porgera\Ddhs	.vsd	V:\Porgera\Ddhs	\\OB1\Clie	2,815,488.00	22/02/2006 1:21	22/02/2006 1:22
19	1	Thumbs.db	V:\Porgera\Ddhs	.db	V:\Porgera\Ddhs	\\OB1\Clie	23,040.00	15/03/2011 5:33	15/03/2011 5:35
20	1	weighted_means.mdb	V:\Porgera\Ddhs	.mdb	V:\Porgera\Ddhs	\\OB1\Clie	249,856.00	4/05/2006 8:53	4/05/2006 9:24
21	1	weighted_means.zip	V:\Porgera\Ddhs	.zip	V:\Porgera\Ddhs	\\OB1\Clie	16,148.00	4/05/2006 9:25	4/05/2006 9:25
22	2	OtherComponents	V:\Porgera\Ddhs\Ddhs_Dev		V:\Porgera\Ddhs	\\OB1\Clie	1,469.00	13/10/2005 9:49	22/04/2005 6:53
23	3	fixsvy.f	V:\Porgera\Ddhs	.f	V:\Porgera\Ddhs	\\OB1\Clie	2,246.00	13/10/2005 9:52	15/12/1997 10:19
24	3	lodsvy.f	V:\Porgera\Ddhs	.f	V:\Porgera\Ddhs	\\OB1\Clie	10,617.00	13/10/2005 9:52	15/12/1997 10:19

One of the major benefits of this approach is that you can use Excel to create lists of commands to Copy, Move, Rename or Delete (for example) large numbers of files at once. (Make sure you have a backup first...!)

COLLECT - MANAGE - ANALYSE

IMAGINE ...

- If someone required a copy of all the core photos, and each drillhole had a separate directory in which the photos were stored?
 - Sort your list by file type (.jpg), and select only those that were in a directory under the Drillhole folders, that contained '...core_photos...' as an example.
 - Use the Excel 'CONCATENATE' function to create a set of COPY commands, then save this in a '.bat' file and run it!
- If you could pull together all the CSV files from the folders in a data warehouse to build a database.
- If you could find all those duplicate 100GB satellite images, spread across your system and remove them.
- If you could instantly create an index of all the files in a resource estimation project for the JORC report.

With the **GDD uList** utility, you can! And what's more, it's free!! Download it from our website [here](#).

Try it for yourself! If you work with data, it's one of those utilities you will end up using every day, discovering more and more things you can do with it.

BUT WAIT, THERE'S MORE!

GeoUte 'uList' has a big brother, '**uIndex**'. The general functionality is similar, but **uIndex** creates the file and directory lists in a database, which means you can use all our data analysis grid features to give your lists the 'Sara Lee' treatment; roll it and fold it, sort, search, slice and dice!

It also means you can keep a complete index of your entire systems in there, updated each night if you want to.

Apart from the fact it makes a great data administration tool, you can locate files you've lost in a tiny fraction of the time it takes using Windows Explorer.

Ask us for more details when you come off hold.....

Now, imagine you had used this tool to pull together the **400-odd CSV and Excel files** from a thousand folders for a data analysis project, and you needed to know **what was in each of those files**, which ones had the same data format and so on, without having to open each one?

You'll have to wait till next month when we will introduce you to the **GDD Data Previewer – uLook**.

See you then,



Dudley the GeoData Dude.

New Tips and Tricks come out every month. Keep an eye on our website for the January Tip!

COLLECT - MANAGE - ANALYSE