Paper 131-2013

Creating ZIP Files with ODS

Jack Hamilton, Division of Research, Kaiser Permanente, Oakland, California

ABSTRACT

ZIP files are a convenient way to bundle related files together, and can save storage space at the same time. The ZIP format is used internally by SAS® for SAS® Enterprise Guide® projects, but until SAS® 9.2 there was no native way to create a ZIP file with your own SAS® program. Starting in SAS® 9.2, you can create your own ZIP files using ODS PACKAGE statements. This presentation describes how to create simple ZIP archives, and discusses how to create an archive file with an internal directory structure.

INTRODUCTION

The ZIP file format was devised in the late 1980's by Phi Katz for his PKZIP utility. It has been updated over the years, and now incorporates several different compression algorithms. The ZIP algorithms are not always the most effective or efficient - there are several competitors that might work better in any given situation, that might compress better or faster or both - but it's a good overall performer. As a result, the ZIP format now used internally by many products (Java JAR files, SAS Enterprise Guide project files, etc.), in addition to being a standard format for distributing groups of files that a consumer might extract and use directly.

Despite its wide use in the computer industry, reading and writing ZIP files has not historically been supported by SAS, even though DSD uses ZIP format files internally. Starting with SAS 9.2, it is possible to create ZIP files using SAS code. There is still no supported method for unzipping files (ask Mr. Google about "SAS saszipam " for an unsupported, undocumented method that often works). Rumor has it that an upcoming version of SAS will support reading and writing ZIP files through a new FILENAME engine.

ODS PACKAGES

Support for writing ZIP files in SAS 9.2comes as part of ODS, the source of so many great new features in SAS over the last few years. Its stated purpose is to create file packages as part of the SAS Publishing Framework, which is part of SAS's extensive product offerings for information distribution. But it's available in base SAS.

CREATING A PACKAGE AND ADDING FILES

To create a package, use the ODS PACKAGE statement with the OPEN option. Typically, you will give the package a name that indicates its purpose, as with any other name in SAS. Here's an example of creating a package named ProdOutput:

```
ods package(ProdOutput) open nopf;
```

The NOPF option tells SAS not to create an additional ZIP entry describing the contents of the ZIP file; you will not want that file, and in some versions of SAS omitting the NOPF option will cause an error if the Publishing Framework is not licensed.

Next, you will add files to the package using the ADD FILE option:

```
ods package(ProdOutput) add file='c:\temp\renal.xps';
```

You can specify either a complete file name and path, or a path relative to the current active directory.

At this point, SAS is collecting information, but hasn't started to create the ZIP file yet.

Finally, you tell SAS that it's time to write the ZIP file using the PUBLISH ARCHIVE option. The name of the ZIP file is also specified in this statement:

```
ods package(ProdOutput) publish archive
        properties(archive_name='ProdOutput.zip'
                   archive_path='c:\temp');
```

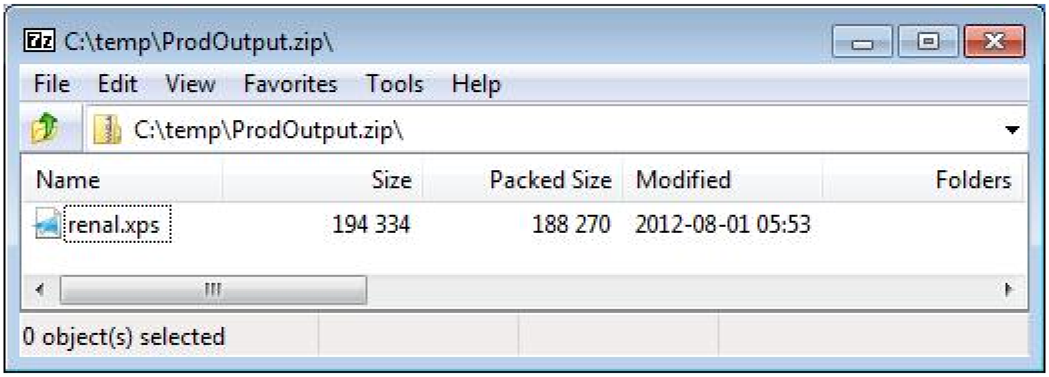
You will see a message in the log confirming that the ZIP file has been written:

```
   NOTE: Writing zip package c:\temp\ProdOutput.zip.
```

Finally, tell SAS you're through adding files, using the CLOSE option:

```
   ods package(ProdOutput) close;
```

If you look at the ZIP file (in this case, using the free utility 7-Zip), you can see the contents of the ZIP file (Figure 1):



Name	Size	Packed Size	Modified	Folders
renal.xps	194 334	188 270	2012-08-01 05:53	

0 object(s) selected

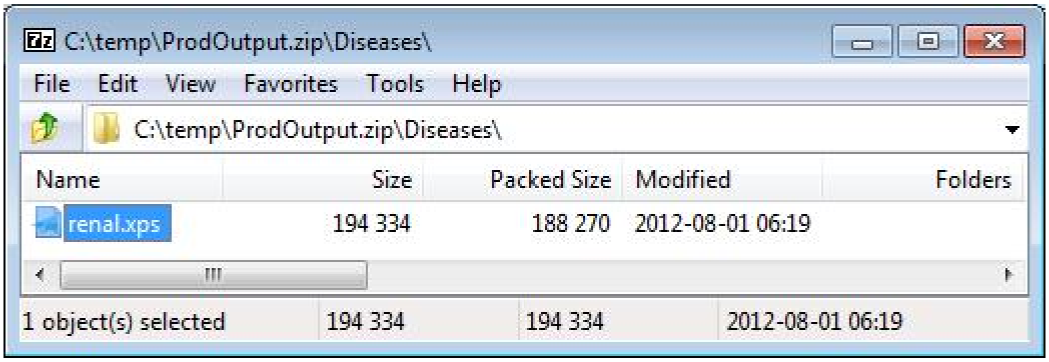
Figure 1. Simple ZIP file shown in 7-Zip

CREATING AN INTERNAL DIRECTORY STRUCTURE

An additional option, PATH, in the ADD FILE allows you to specify an internal directory structure within your ZIP file:

```
   ods package(ProdOutput) add file='c:\temp\renal.xps'
                               path='Diseases';
```

If you use this statement in place of the one above, the "renal.xps" file will be placed inside a directory in the ZIP file (Figure 2):


C:\temp\ProdOutput.zip\Diseases\
File
Edit
View
Favorites
Tools
Help
C:\temp\ProdOutput.zip\Diseases\

Name	Size	Packed Size	Modified	Folders
renal.xps	194 334	188 270	2012-08-01 06:19	

1 object(s) selected 194 334 194 334 2012-08-01 06:19

Figure 2. ZIP file with directories shown in 7-Zip.

AUTOMATING ZIPPING

The examples above assume that you know the names of all the files you will want to add to the ZIP file.

If the file names vary, but you have the names in a SAS data set, you can write the ADD statements with CALL EXECUTE.

```
filename dirlist pipe 'dir /b y:\wuss2012\*.sas';
data _null_;
    infile dirlist pad lrecl=80;
    input @1 filename $80.;
    call execute
        (catx
            (' ',
            'ods package(vary)',
             'add file=',
             quote('y:\wuss2012\' || trim(filename)),
             ';'
            )
        );
run;
```

You can create the data set containing file names while you are creating the files themselves, or in whatever other way you choose. For a platform-independent method of creating a list of files and directories in SAS, see

http://www.sascommunity.org/wiki/Obtaining_A_List_of_Files_In_A_Directory_Using_SAS_Functions

CAUTIONS AND AMPLIFICATIONS

It does not seem to be possible to create an empty directory inside a ZIP file.

It is not possible to encrypt a file while ZIPping it.

You cannot use wild cards to specify files in an ADD statement. This might be a bug, or might be a feature, depending on how you look at it.

You can't add files to an existing ZIP file. Using an existing ZIP file will delete it and create a new version without the previous entries.

The files you want to add have to exist when the PUBLISH statement executes.

You can use a fileref instead of a physical file name.

There's no documented or supported way to read a ZIP file in SAS, even though the software obviously has the capability internally.

The compression method SAS uses is compatible with every ZIP reader I have tried. It is not, however, always the most effective compression method.

FURTHER READING

A version of this paper is available online at

http://www.sascommunity.org/wiki/Creating_a_ZIP_File_Using_Native_SAS_Features

REFERENCES

SAS Institute. Using ODS to Publish Packages”. March 2013. Available at <http://support.sas.com/rnd/base/ods/odspackages/index.html>.

ACKNOWLEDGMENTS

I would like to thank Jamila Gul at the Kaiser Permanente Division of Research and Tom Billings at Union Bank for their help and encouragement.

CONTACT INFORMATION

Your comments and questions are encouraged. Contact the author at:

Jack Hamilton
jfh@acm.org