Silbury Hill Conservation Project – Preparation of the Digital Archive: Final Report

Title:	Silbury Hill Conservation project- Preparation of the Digital archive: Final Report
Author(s):	Ryder, J.
Derivation:	silbury_hill_archive_interim_report_3_mar_2013_v3.doc
Origination Date:	02/09/2013
Reviser(s):	Catherine Hardman
Date of last revision:	25/09/2013
Version:	1.3
Status:	Final
Summary of changes:	Amendments as suggested by C. Hardman
Circulation:	Duncan Brown et al
Required Action:	Review and comment
File name/Location:	silbury_hill_archive_final_report_4_sept_2013_v1.doc/located in \PR661-workingSilbury\Jenny Ryder\Archive reports ADS contract
Approval:	

Preparation of the Digital Archive: Final Report

Jenny Ryder (JR): ADS

Version 1.3: September 2013

Summary of the Project

Work undertaken:

- A comprehensive audit of the digital data;
- Recommending files for selection or deletion;
- Re-organising the directory structure;
- Renaming files according to consistent convention;
- Creating metadata using ADS templates and the Guides to Good Practice;
- Compiling supporting documentation

Results:

- Final dataset less than half the size of the original;
- Final dataset contains less than half the different file formats as the original;
- Dataset is sorted into logical structure based on EH and ADS division of data-Images/Graphics and other data types, then both datasets divided into site and post-excavation data.
- Archive is fully documented with Project Metadata; file-level metadata; file type-specific metadata and supporting documentation.

Recommendations for further work:

- Geophysics data is still to be archived with the ADS;
- ADS to accession, curate and disseminate data;
- Silbury Digital Archive could be used as a case study to inform future data management plans;
- Silbury Digital Archive has great potential for further research particularly through related English Heritage project ‘Later Silbury’ when that is made available and through revisiting the photogrammetry projects.

Background

This document is the final report of the English Heritage / Archaeology Data Service (ADS) project to secure the digital archive created by the Silbury Hill Conservation project undertaken by English Heritage 2000-2008. This report will summarise the final dataset and review the processes used in the preparation of the digital data for deposition with a view to assisting English Heritage in the development of a new digital data management strategy.

The main aims of this report are:

- To provide an overview of the dataset;
- To summarize the work done in preparing the digital archive for deposition;
- To highlight issues encountered and suggest solutions where possible.

Contents

1. The need for the Silbury digital archive preparation project.

The Silbury Hill Conservation Project originated when a 14 metre deep crater unexpectedly opened on the summit. The hill was monitored, surveyed and examined over several years, with the work culminating in the excavation works in 2007/8. During the 2001 excavations and the 2007/8 works, the various deposits revealed as the tunnels were re-excavated and the summit examined were extensively sampled and recorded. The results were assessed and analysed giving rise to a vast and mixed dataset which grew organically resulting in a digital archive without structure, documentation or consistency.

After a general audit carried out in 2009, which provided an overview of the state of the Silbury digital archive, it was proposed that a temporary member of staff should be appointed by English Heritage for 3 months with the following aims, objectives and tasks:

"The aim is to produce a comprehensive, well-ordered digital archive with full metadata entries, properly indexed and with supporting documentation. A temporary project assistant will be employed to carry out the following tasks:

1. A comprehensive audit of all the digital material, detailing types and sizes of files, file-naming conventions and the directory structure
2. Identifying different versions of the same file and making recommendations for selection and deletion
3. Re-organising the directory structure as appropriate
4. Adding metadata to individual files as appropriate
5. Compiling an index to the digital archive
6. Compiling supporting documentation such as data entry conventions and standards
7. Communication with repositories over procedures for archive transfer…

"

Proposal for Pr661 Silbury Hill Compilation and Transfer of the Digital Archive, Aug 2011, Duncan H. Brown

Upon commencement of the work, however, it became apparent that the sheer quantity of files and the amount of work to be done to audit, compile and complete the archive would take a much greater amount of time than originally supposed. After a meeting with staff from the ADS, it was established that, far from the original estimate of 60 days, the work was likely to still take a further year or more to complete.

2. An overview of the original dataset (2012)

The total number of files in the Silbury Project folder was, when first assessed: 30, 082 files; Size: 217 GB (233,355,277,359 bytes).

As the project continued, this grew to 34, 518 files as new digital data was created or found by English Heritage and added to the project folder.

The contents of the dataset were as follows:

Photogrammetry Topcon projects and miscellaneous other related files	9, 777 files	**File formats:** aut; bak; bm$; bm$w; bmp; bmpw; cmr; cnt; cs$; CSV; dat; dll; doc; dwg; dx$; dxf; exe; ext; GID; he$; hlp; imc; ini; isu; jp$; jpg; jpgw; le$; lng; NEF; or$; ORT; pdf; pi4; prj; psd; rel; STE; ti$; tif; tin; txt; wrl; xls; zip
Images those in the images folder- not necessarily all of the images as some image files were also located in other folders	13, 981 files	**File formats:** adf; ai; asc; aux; AVI; bak; bmp; cdc; cdr; cpg; csv; dat; dbf; dir; doc; dwf; dwg; dwl2; dxe; dxf; emf; eps; err; exe; GSI; htm; idx; img; indd; JGw; JPG; kgm; lft; lnk; log; lyr; mov; movie; mpg; mxd; nit; over; pdf; pgw; png; ppt; prj; psd; rrd; rtf; sbn; sbx; scr; shp; shx; skb; skp; tab; taw; tfw; tga; tif; tmp; txt; ucs; wbk; wmv; wrl; xls; xml; zip
New files added during the archive preparation process	4, 436 files	**File formats:** doc; jpg; mdb; tif; xls
All other files miscellaneous finds/enviro data, documents, reports, survey data etc.	6, 324 files	**File formats:** adf; asc; aux; bak; BAS; bmp; cdr; css; CSV; dat; dir; dmp; doc; docx; dwg; dxf; emf; enl; eps; err; EXE; frm; FRT; gif; GSI; hex; htm; jpeg; jpg; js; kgm; lnk; log; mdb; MPF; mpp; msg; mtt; MYD; MYI; nit; obj; opt; PCD; pdf; ply; png; ppt; ps; psd; ram; rar; rdb; rrd; rtf; SCR; SEC; sim; SYM;

		tfw; tif; tmp; txt; vsd; WAV; wbk; wmv; XCF; xls; xml; xyz
Totals	34, 518 files	132 different file extensions

As well as there being a large number of different file formats, the folder structure itself was fairly disorganised. There were hundreds of different folders and it was often the folder that was used to either describe the content of the files within or the person who created the file; such information was missing from the file names or any file metadata.

3. The archive preparation processes

The original plan was to audit the files, then rename and sort them, then compile the metadata and prepare them for deposition. However the complex folder structure and lack of descriptive file names meant that it was almost impossible to assess whether files were draft versions, later versions or duplicates of other files without sorting them and in some cases renaming them, while completing the auditing process.

Auditing the data

In order to do this, a spreadsheet was created which displayed the following information:

- **Original File Path** (so that if any files were in use, or their location referred to in other documents, they could be easily found if relocated).
- **File Name** (the original file name)
- **Renamed** (the new file name, if renaming had taken place)
- **File ext** (the extension, signifying file type)
- **Active/ADS/NMR/Not Archive/Delete** (this signified what may be an appropriate action to take with each: whether the file was active, and therefore to be left unsorted/examined for the present, to be saved and sent to the ADS or NMR, to be retained but was not considered relevant to the Digital Archive, or whether the file was of no further use or relevance and could therefore be discarded).
- **New Location** (if the file was relocated, the path of its new location within the Project folder).

- **Comments/Reason for New Location** (this was used mainly for the files to be deleted and explained the reason behind the action taken with the file, e.g. why it could be discarded, and in most cases, on whose authority/advice)
- **Needs Renaming?** (this column was added later as it became necessary to try and estimate the extent to which the file names needed to be changed, Y for Yes, N for No, N/A if the files were to be discarded and Done, if the files needed to be renamed in order to be relocated)
- **Size (bytes)** (the file size, as time began to run out and priorities had to be made, not every file size was displayed, in some cases, where the same treatment was made to an entire folder, the total size of the files within that folder was given).

This spreadsheet was the basis for all of the other processes as it helped to keep track of a file's location, name changes and any other descriptive information that could be added, for example, if a survey file was seen to be used in an AutoCAD drawing, this could be noted, making it easier for the two files to be linked in the final metadata.

Each action was documented using this spreadsheet, the list was divided first into four separate sheets: Photogrammetry; Images; New Material and Everything else. This was in order to keep large sections of work separate and helped to measure progress as well as to try and do an initial sort of material into groups.
The photogrammetry was left unchanged as the Topcon Projects would not retain their integrity if the files were changed in any way or if the material in the sub-folders were separated; they function as a bundle of data and so have been left in that state.
For all other files, these were sorted into four separate directories: ADS; EH Archives; DISCARD and NOT_ARCHIVE according to whether the file was worth keeping (if so the image/graphics files being sorted to the EH folder or any other files to the ADS folder) and, if not, it was either discarded, or, if the file needed to be kept but was not suitable for a public repository (for example, if the file contained financial or administrative information) this was sorted to the NOT_ARCHIVE folder to be sorted through and dealt with by the English Heritage Archaeological Archives team at the end of the project.

Making recommendations for selection and deletion:

The first stage in looking at which files to keep in the archive and which to discard was straightforward, an overview of the folders had already been created by the English Heritage Archaeological Archives Curator, Claire Tsang. This identified which folders held files that were largely 'NOT ARCHIVE:' those files which included project management information, admin or health and safety information for example. So all of the files identified as 'NOT ARCHIVE' in this earlier audit were separated to a separate folder as mentioned above.

The second stage was to identify any obvious duplicates, though this was not as simple as it might have been as files could be located in any of the hundreds of folders without documentation or file names to identify them. Many of the directories had been used as personal working files and the file names changed to suit whichever person may be working on them at the time, so, from this point on, assessing a file's suitability for archiving meant opening each individual file and working from there. This is the main reason why the selection process took such a long time: the lack of coherent structure or descriptive naming convention meant that, in order to know what a file contained, why it was created and therefore how useful it may be, I sometimes had no clues from either its placing in the archive or its title. During the course of this work, a blog summarising the Silbury Hill work was written: http://archaeologydataservice.ac.uk/blog/2013/07/the-silbury-hill-archive-the-light-at-the-end-of-the-tunnel/ and was followed by a subsequent blog as part of the Day of Archaeology: http://archaeologydataservice.ac.uk/blog/2013/08/jenny-ryders-day-of-archaeology-at-the-ads-a-silbury-hill-update/#more-895

The following criteria for selecting files to be included in a digital archive have been taken from the ADS/EH Guides to Good Practice, available from the Archaeology Data Service website http://guides.archaeologydataservice.ac.uk/:

- **"Reuse Cases**—This is probably both the most important criterion and occasionally the most difficult to judge. Where the data is in a form that can obviously be used by other researchers, or in other contexts, then the question is simply whether reuse is likely to occur. The other complication is that, for

certain types of data, a reuse case can be imagined as feasible even if it is currently not being enacted. An example of this would be a form of data that might lend itself to a post-processing technique under development, or merely envisaged as possible in the future (or an enhancement to an existing technique).

- **Repeatability**—Is the process that created this data repeatable? If so, an earlier stage may be an appropriate PIP (Preservation Intervention Point); if not, then this intervention point should be selected.
- **Retention policy**—The data should match the retention policy of the target archive.
- **Value**—The cost of intervening to preserve data at this particular point, given that no project has an unlimited budget. "Value" here also means the value of the material to be archived, e.g., it might be worth preserving data produced by a repeatable process if that process were particularly expensive and difficult to reproduce. Value, therefore, has to do with balancing the perceived worth of the data against the cost of archiving."

'Data Selection: Preservation Intervention Points'. Edited by Kieron Niven with contributions by Tony Austin, Jonathan Bateman, Stuart Jeffrey, Jen Mitcham Archaeology Data Service / Digital Antiquity (2011) *Guides to Good Practice*

The data was selected with these principles in mind, duplicates and working files were removed, primary data saved, where useful and final outputs were saved. In addition to the issue of the file names and structure, there was also the issue of the lack of specialist knowledge, a file may contain useful information created by specialists for other specialists, but this may not be apparent to the person responsible for compiling the archive without sufficient documentation and this was the case with many of the Silbury files. In these cases it was impossible to assess a file's usefulness to the archive without seeking further assistance, in order to do this it was necessary first to try and work out when it was created and by who, then find out if the relevant staff member was still working for English Heritage. If the original creator was no longer available, the opinions of their colleagues and of the Project Manager had to be sought before a decision was made.

In general, then, it was mostly the lack of documentation that proved the biggest obstacle in terms of time and needing the repeated input of other members of staff, which further added to the overall cost of the project.

Though the selection process was on-going throughout the archive preparation project, before documentation of the files could begin, the sorting took around 8 months.

Re-organising the directory structure

This stage of the process was actually done in tandem with the selection of files. As files were selected, they were renamed and relocated to a more structured directory. The proposed content of the digital archive had grown organically, for example although originally there was a dedicated 'IMAGES' folder, in the end images were scattered in various forms of primary data, working drafts and copied files in various folders, either moved or created by individuals for working purposes and never discarded or re-organised. As files were renamed (according to a file-naming convention in development by EH at the beginning of 2012) and sorted into a logical folder structure it became clearer that there were more duplicates or working versions than the original sift had shown, as like files were more easily compared once grouped together and given a descriptive name.

The final folder structure was divided first into EH Archive and ADS main directories (this equates essentially to their being an 'Images' folder and a 'Supporting data' folder. Both of these were divided into 'Site' and 'Post-excavation' sub-directories and, as this effectively gave a chronological order to the files this also helped in the selection process as raw data was separated from processed data from the assessment and analysis phases of the project.

As each file was moved, this was documented on the original spreadsheet from the audit stage. The documenting of the files is not just beneficial to those intending to reuse the files, nor just to the repositories who need to know what they are preserving, but it is an important way of charting the process and essential in being able to backtrack should something go wrong, or, as was the case a few times in this project, a decision about the file's worth be changed.

Adding metadata to individual files as appropriate

Once the files had been selected, sorted and renamed, the documentation stage could begin. Again, this process was made much more difficult given that some of the files had been created years ago by people who were now no longer working for English Heritage. Some of the information useful to repositories, such as the ADS, was no longer available and where information was missing, once again, the specialists who created the files, where available, had to be consulted or else gaps had to be left.

Where the information was available, however, the creation of the metadata was extremely straight-forward and, though time-consuming as a large task at the end of the process, if each file had been documented as it was created, by the creator, the time involved would be negligible. The ADS provides templates to fill out and these are generally simple and intuitive, again, though an onerous task when documenting 15,000 files with gaps in knowledge, had the documenting been completed as the project progressed, a matter of a few minutes a day on site for example, all processes of selecting and structuring the project data ready for archive would have taken much less time and therefore been much more cost-effective.

As it was, the documenting of the archive: including the files that have been discarded or set aside, creating new explanatory documents and time spent in meetings, seeking advice and concurrently reformatting and resorting the files where appropriate, the documentation stage of the archive preparation process has taken approximately four and a half months.

Compiling an index to the digital archive

This is the file-level metadata which was started as part of the first process of auditing the data. By creating that first spreadsheet, effectively the only other information needed was to add descriptive information about each file and how the files related to each other. This process also revealed either superfluous files, or in some cases gaps where discarded files may still have use, these files were discarded or retrieved where appropriate and again, their new locations were documented.

Compiling supporting documentation

In addition to the templates specified by the ADS, there was also a need for further explanatory documents to be created, such as compiling a record of the notes regarding the Photogrammetry projects and the issues encountered in the Silbury tunnels; sketch plans of the survey point locations; a table of events-given that the Silbury conservation project spanned nine years-and the relating the archive material to those events; a copy of the file-naming convention used; a summary of the archive contents and this report.

Communication with repositories over procedures for archive transfer

As with most of the other processes, this task was on-going throughout the year with both repositories: EH Archives and the ADS being kept informed of progress and discussing plans for final transfer.

4. An overview of the final dataset (September 2013)

The final dataset, despite the Photogrammetry project files being largely unchanged, is now less than half the size it was originally, a reduction of over 20,000 files.

The final file count is 13, 739 which is split as 13, 557 files to be deposited with English Heritage Archives and 182 files to be deposited with the Archaeology Data Service. In addition there are 30 metadata and archive information files: copies of which will be kept by both repositories.

From the original 132 different file extensions listed, the archives now contain only 49 different file formats, most of which are within the Photogrammetry directory which has had to remain the same. The ADS archive now has 6 different file formats and the EH archive, only 2.

The table below summarises the archive contents in more detail:

ADS Archive contents:	
Archive information folder	• a summary of the archive contents (i.e. the final version of this document); • a report detailing the archiving process;

(5 files):	• a table of the Silbury Hill events as identified by the Digital Archivist; • a copy of the file-naming convention used-in diagram form • a report containing all other information regarding the photogrammetry files and the working process.
Metadata folder (25 files): Project-level metadata: this is the top level metadata-effectively an overview of the project and the work and people involved. Then three sub-directories:	
ADS Metadata (5 files):	• ADS Metadata File Level: essentially an index of the all the files to be deposited with the Archaeology Data Service. Sub-folder:- ADS File Types (4 files containing file type-specific information): • 2 doc files-database data dictionary and spreadsheet metadata in table form; • 2 tiff files-two versions of the database entity relationship diagram.
EH Metadata (3 files):	• EH Metadata File Level: essentially an index of the all the files to be deposited with the English Heritage Archives. Sub-folder:- EH File Types (2 files containing file type-specific information): • 2 excel spreadsheets-information regarding the Raster images and Vector images using the ADS templates.
Topcon Projects (16 files): As these are slightly more complicated, they have more metadata and so these files have been given their own folder, the files are:	
Topcon Projects	• 11 tiff files-scanned images of plans of the survey points used to undertake the photogrammetry, (these were created during the archiving process using the stereo pair photographs) • 3 excel spreadsheets-information pertaining to the different file types using the ADS templates. These are: Camera data; Image data; and survey data. • Excel spreadsheet-file level-effectively an index of the Topcon project files • Excel spreadsheet-project relationships: in addition to the ADS

	suggested templates, I have created a file relationship spreadsheet as so many of the photogrammetry files relate to each other. This basically states what image files were used with what survey files to create which project which was then digitised into AutoCAD.
The rest is the actual Silbury data, for the ADS archive this is: 182 files (963MB): • The 182 files are comprised of 127 .csv files; 31 .doc files; 1 .docx; 1 .mdb; 6 .pdf; 16 .xls they are sorted as follows:	
Post-Excavation Data (52 files):	• 16 Finds and Enviro data files: 14 spreadsheets and 2 word documents: 7 from 2001-2004 6 of which are related to the core samples and 9 from the 2007 work; • 33 Reports and Documents: 27 word documents and 6 pdf files, various evaluation, assessment reports and summaries from the various different episodes of work within the Silbury Hill Conservation Project. • 3 files - Site Interpretation Data: 2 excel spreadsheets: the latest context index and site Harris matrix and 1 word document: the Silbury archaeological Phase Summary.
Site Data (130 files):	• 2 Site Records Database files: the database itself in mdb format and a copy of the original user guide in word. • 128 Survey files: 127 csv files containing the spatial data, and 1 word document with notes regarding the crater survey in 2008.
English Heritage Archive contents:	
Archive information folder (5 files):	• a summary of the archive contents (i.e. the final version of this document); • a report detailing the archiving process; • a table of the Silbury Hill events as identified by the Digital Archivist; • a copy of the file-naming convention used-in diagram form • a report containing all other information regarding the photogrammetry files and the working process.
Metadata folder (25 files): Project-level metadata: this is the top level metadata-effectively an overview of the	

<table>
<tr><td colspan="2">project and the work and people involved.
Then three sub-directories:</td></tr>
<tr><td>ADS Metadata (5 files):</td><td>• ADS Metadata File Level: essentially an index of the all the files to be deposited with the Archaeology Data Service.
Sub-folder:- ADS File Types (4 files containing file type-specific information):
• 2 doc files-database data dictionary and spreadsheet metadata in table form;
• 2 tiff files-two versions of the database entity relationship diagram.</td></tr>
<tr><td>EH Metadata (3 files):</td><td>• EH Metadata File Level: essentially an index of the all the files to be deposited with the English Heritage Archives.
Sub-folder:- EH File Types (2 files containing file type-specific information):
• 2 excel spreadsheets-information regarding the Raster images and Vector images using the ADS templates.</td></tr>
<tr><td>Topcon Projects (16 files):</td><td>As these are slightly more complicated, they have more metadata and so these files have been given their own folder, the files are:
• 11 tiff files-scanned images of plans of the survey points used to undertake the photogrammetry, (these were created during the archiving process using the stereo pair photographs)
• 3 excel spreadsheets-information pertaining to the different file types using the ADS templates. These are: Camera data; Image data; and survey data.
• Excel spreadsheet-file level-effectively an index of the Topcon project files
• Excel spreadsheet-project relationships: in addition to the ADS suggested templates, I have created a file relationship spreadsheet as so many of the photogrammetry files relate to each other. This basically states what image files were used with what survey files to create which project which was then digitised into AutoCAD.</td></tr>
<tr><td colspan="2">The rest is the actual Silbury data, for the EH archive this is:
13,557 files (190GB):
• 6,230 Images;</td></tr>
</table>

• 7,327 Photogrammetry Project files. The 6,230 image files are comprised of 6,095 .tiff files and 135 .dwg files, they are sorted as follows:	
Site Images (5840 files):	• 1,123 Photogrammetry Photos: 4 orthoimages; 10 calibration images; 1,109 Topcon images (i.e. the stereo pair photographs); • 133 Site Drawings: scanned hand-drawn drawings: 6 from 2001; 127 from 2007; • 4,584 Site Photos: 21 photos of the summit shaft; 2,307 site photos from 2000-2006; 383 Site Photos from 2007 (DS Camera); 1,873 site photos from 2007 (EH Camera)
Post-excavation Images (390 files):	• 21 AutoCAD drawings-mostly working drawings, plus some showing locations of boreholes, contour models, and badger setts; • 30 Digitised site drawings (AutoCAD)-mainly the summit excavations • 84 Digitised Topcon Projects (AutoCAD)-mainly the tunnel elevations • 255 finds and enviro images: 40 photographs of antler fragments; 25 photographs of the BBC time capsule during conservation; 180 images of 2001-2003 core samples (154 cropped photos and 26 diagrams); 6 photographs of stone fragments; 4 images (one sketch, 3 photos) of turf sample.

The Photogrammetry Topcon PI3000 Projects

These files are all of the Topcon projects created on site using the stereo pair photographs, survey files, and camera calibration files which have all been saved separately elsewhere in the archive.

Most of these files were created by the Topcon software as part of the processing of the images to create a 3D surface. These are being retained as there has not yet been a finished final product of the photogrammetry work and they could do with being revisited. By keeping the projects, it may be easier to reuse these projects than to start the process again using the RAW data, but the user has both options as all have been retained.

N.B. The Photogrammetry-specific metadata has been duplicated and included with the Topcon Projects so that, should they be stored separately to the rest of the archive, the relevant information should still be readily accessible.

The Topcon file types, in order of frequency, are as follows:

File type	File Count	Description
.imc	1115	Image Coordinates file-This file is internally generated by the Topcon PI3000 software.
.bmp	1052	This file is an image internally generated by the Topcon PI3000 software-manipulated versions of the stereo pair tiffs
.ext	1045	This file is internally generated by the Topcon PI3000 software.
.jpg	971	Generally these are the surface files created internally by the Topcon software (though a few of the original stereo pair images were jpegs as well.
.tif	834	Stereo pair images, these are duplicates of those saved separately in the EH Archive. 11 of these tiffs are part of the metadata
.rel	534	This file is internally generated by the Topcon PI3000 software.
.TIF	245	Stereo pair image
.ste	230	This file is internally generated by the Topcon PI3000 software. It contains data relevant to the stereo pairs.
.STE	220	This file is internally generated by the Topcon PI3000 software. It contains data relevant to the stereo pairs.
.txt	153	Generally the 'bundleresult.txt' files, internally generated by the Topcon PI3000 software, it is a key file for identifying how Topcon processed the information as it contains the relevant survey, camera calibration and image data for each particular sub-project
.cmr	85	The main camera calibration file used in the Topcon projects for the tunnel, Also saved elsewhere in the ADS archive.
.cs$	83	This file is internally generated by the Topcon PI3000 software.
.dx$	83	This file is internally generated by the Topcon PI3000 software.
.he$	83	This file is internally generated by the Topcon PI3000 software.
.le$	83	This file is internally generated by the Topcon PI3000 software.

.ti$	83	This file is internally generated by the Topcon PI3000 software.
.pi4	81	Project file-the main file used to view the Topcon project-in this case the Main tunnel, East elevation, bays 01 to 03
.bm$	78	This file is internally generated by the Topcon PI3000 software.
.bm$w	78	This file is internally generated by the Topcon PI3000 software.
.or$	78	This file is internally generated by the Topcon PI3000 software.
.JPG	27	Copies of stereo pair images (created by Topcon generally)
.IMC	10	Image Coordinates file-This file is internally generated by the Topcon PI3000 software.
.wrl	9	This file is internally generated by the Topcon PI3000 software - a general file for recording TIN and texture data.
.ort	7	This file is internally generated by the Topcon PI3000 software. (ortho-image)
.xls	5	These are the metadata spreadsheets.
.bmpw	4	This file is internally generated by the Topcon PI3000 software.
.BM$	4	This file is internally generated by the Topcon PI3000 software.
.jpgw	3	This file is internally generated by the Topcon PI3000 software.
.pdf	3	Working images of orthoimages and points.
.dxf	3	This file is internally generated by the Topcon PI3000 software.
.CSV	3	Survey file used by the Topcon PI3000 software to locate images using target points, it is also saved elsewhere in the ADS archive.
.CV$	3	This file is internally generated by the Topcon PI3000 software.
.DX$	3	This file is internally generated by the Topcon PI3000 software.
.DXF	3	This file is internally generated by the Topcon PI3000 software.
.HE$	3	This file is internally generated by the Topcon PI3000 software.
.LE$	3	This file is internally generated by the Topcon PI3000 software.
.ORT	3	This file is internally generated by the Topcon PI3000 software.
.PI4	3	Project file-the main file used to view the Topcon project
.TI$	3	This file is internally generated by the Topcon PI3000 software.
.TXT	3	Positional information file internally generated by the Topcon PI3000 software.
.dwg	2	test working drawings
.OR$	2	This file is internally generated by the Topcon PI3000 software.

.tin	1	This file is internally generated by the Topcon PI3000 software. It is a Triangulated Irregular Network-a series of triangles that map a surface.
.BMP	1	This file is internally generated by the Topcon PI3000 software.
.csv	1	Survey file used by the Topcon PI3000 software to locate images using target points, also saved elsewhere in the ADS archive.
.bak	1	This file is internally generated by the Topcon PI3000 software.
Totals	47 file types	7327 files

5. The next steps

As planned and stated above, the digital images and graphics have been sent to the English Heritage Archives at Swindon, with the rest of the data: the spreadsheets, reports and site database, being held by the ADS.

English Heritage have agreed that it would be beneficial for the ADS to have the geophysics data from the Silbury Hill project as this may form a more complete dataset for the ADS and would allow the geophysics data to be curated and disseminated effectively with the rest of the Silbury data. A time-frame has not yet been set for when the geophysics data can be prepared for deposition, so the accession of the ADS portion of the archive is dependent upon whether or not the geophysics data can be deposited with the ADS and if so, when.

Once the above has been determined, the ADS can accession and begin curating the data and disseminate it on their website; the idea being that solid links will be established between the ADS datasets and the EH images and graphics.

It is then hoped that the data may be reused for continued research into Silbury Hill, particularly, perhaps, when the subsequent project: Later Silbury - Archaeological evaluation of the fields south of Silbury Hill, Wiltshire. (Pr 5980) becomes available. The photogrammetry projects in particular have further potential for research into recording methods; visualisation and the tunnel deposits themselves through the 3D

images that could be created. A much more visual and perhaps interactive, more tangible archive could be created from the current stored data once again increasing the potential for further research.

6. The impact of the Silbury Hill digital data archive preparation project: conclusions

It is hoped that the Silbury Digital Archive could be used as a case-study to inform future data management plans. The digital landscape has changed dramatically over the period from which the Silbury Hill project first began. For example, in the first stages of the Silbury project, photographs were taken on traditional 35mm cameras and/or black and white cameras rather than digital and generally there was much less digital data being created in the field outside of survey. Now photographs are digital and records themselves are becoming digital, in the Silbury Hill 2007 works, it was intended that the drawings would be replaced by the 3D projections provided through the photogrammetric recording and for the Later Silbury Project in 2010, the digital recording system Intrasis was used as the primary record rather than the paper context sheets. This shifts the emphasis of born digital data from being largely based in the realms of post-excavation to being created throughout the project from conception.

There is a shift in responsibility as well, as digital archives grow in size, as well as for the reasons above, it becomes unrealistic and costly to view the preparation of the archive as the final step of the project as it becomes too unwieldy and too technical for one archivist to undertake. Instead it is necessary for the creators of the data to be responsible for its future, ensuring that others can reuse their work increases the potential of their research and, in the current economic climate where costs need to be justified, goes far in proving the value of their work. Therefore, a well-ordered and documented archive is essential in ensuring that the information is given relevance beyond the life-cycle of the project itself and those who created it.

Related to this is also the economic impact of depositing data and making it accessible to a wider audience. The Archaeology Data Service has recently completed an ‘Impact Study’ with the following aims:

“The project will analyse and survey perceptions of the value of digital collections held by the Archaeology Data Service and how those perceptions of value can be measured. As part of this work, we will assess and quantify the economic impact of those collections with the ultimate objective of improving their prospects for sustainability.” (http://archaeologydataservice.ac.uk/research/impact)
The top three benefits as seen by users were:

- The ability to find data from a single point of access;
- Data beyond the scope to collect myself;
- Long-term preservation of data;

The results of the study are a reflection of the benefits of digital data deposition in general, the study suggested: “that ADS’ value has increased over time with the growth of collections and achieving critical mass.” *Archaeology Data Service Impact Study’* leaflet. The more archives that are widely available to the public, the greater the potential the individual collection has as part of an increasingly rich digital resource. In terms of economic value, the ‘Impact Study’ results showed that over 30 years, a “£1 investment in data and related infrastructure arising from additional use facilitated by the ADS may provide up to £8.30 return” *Archaeology Data Service Impact Study’* leaflet.