I'm briefly hoping to help us all understand two key issues regarding the storage of data online

Question number one: How should researchers archive their websites so that material we upload can be retrieved in, perhaps, 50 years time?

Here is an example of one of the latest battles in history circles: given a fair wind which of these is likely to survive its format the longest?   This letter is already nearly a century old; and what is contained on this (a USB drive) has had to be reformatted twice – since being created in 2005.

But in the next 15 minutes we are going to consider one aspect of digital documents: websites.

Before considering a website as a unitary tool we should first discover what it actually contains; its many individual parts.

Recently I was sent what purported to be an archive of family photos, which the owner was intending to attach to a website. It was based on commercial software but built on an early Excel spreadsheet. On investigation I discovered the software lost its support in 2011 after a 10-year life. Think what computers we were probably using (if any) in 2001. There were about 20 family photos, one per row, with other text data, such as caption, reference, location, in other columns. The images in the spreadsheet ranged from 5KB to 12KB; too small even as seen on the selection page to be of any use. I understand that the original prints are no longer in existence. What use would it be attached to a website? Indeed, what use is it as a standalone?

So before dealing with the presentation end – the website – we should have a clear idea of where each element has come from and where it is now, outside of the website, in another part of the digital world. If something goes wrong at the presentation end we need to have a means of reconstructing the site and/or its elements as a last resort.

Any Plan A would surely be to ensure we can keep the publication of the site in the public domain through the twists and turns of digital development.

Any Plan B is to ensure that if plan A fails at any point, someone can pick up the pieces and attempt to reconstruct the elements: the text, the images, the tabulations, the links, in whatever form is considered appropriate at the time.

Here is a collection of useful stuff for those of us who have built websites, are building websites, or aspire to build websites which contains data that needs to remain available.

1. *Definitely not a nice cheap software with pretty though restrictive themes where you input data online and you have no local copy. Probably cross that one off the list.*
2. *Not take up the offer of a friend who knows someone who can code and can make you up exactly what you want – except when you're told it's not possible; s/he won't be around when the construction goes wrong in a few years; ditto receiving bespoke software passed on by a third party.*
3. *As Jack Latimer, of Community Sites fame, keeps imploring, if the software does not have a built-in export function its main insurance benefit is missing.*
4. *Even if the software does not handle it automatically ensure the administrator works via an Assets folder. In other words, every item to be imported (image, table, text) is imported first to a suitably sub-divided Assets folder, with appropriate file names and notes. From there it is imported to the website.*
5. *Import images to the Assets folder in the highest resolution you can. Then pass a copy of this via compression software. This ensures the website contains manageable file sizes while the Assets folder contains full resolution versions. More to the point, all the files you use for the website are in one place. That's a huge insurance protection if there is a problem with the website.*

All of the above is good website management which we should all aspire to achieve, though I know we often fall short.

To take it a step further, we can/should store multiple copies of the administrator's version of the website and its Assets folder.

But the unanswered point from the question I asked at the beginning still hasn't been answered. If a community group retires its data-rich website today, yet wishes visitors to be exploring its content in fifty years time, is that realistic? Imagine a comprehensive science book published fifty years ago; its design and layout will, of course, reflect the date of publication; its content, of course, will be missing fifty-years-worth of progress in science. Yet its value is no less diminished. So perhaps, instead, we should be focusing on the content – the text, images and tables – rather than the design. So the question then becomes, where should its content be stored? and not where should the webpages be archived?

Not all of our websites are data-rich, of course. If our local history website is maintained mainly to inform members of the date of the next meeting, the names of the current officers, and brief details of the next project, the site is useful while our

organisation is active, but when everyone loses interest the site can disappear along with its previous members.

Question two: How should researchers manage the risk of storing their data with third parties?

We have already discussed some of these risk measures, such as

1. *Having a full record of where all elements used in the website or archive are, and who legally owns them.*
2. *Ensuring we have maximum resolution versions of images, even if sub-optimal versions appear on the site.*
3. *Having an Assets folder which contains a copy of all the bits the site contains.*
4. *Not already mentioned, it is important to use the safest file formats. The biggest issues can arise from file formats used by bespoke image software and some open access page layout and word processing software. There are almost always standard alternatives (sometimes using Export instead of Save), but if an **image.psd** is used, for example, be aware that even Photoshop may not last forever!*

So this question really addresses how much trust we can have in the archive or website software on which we have or are about to, commit our data.

Unless our site went live within the past 5 or 6 years, we will be working with legacy software with rather less standardisation and fewer features. I acquired mine – completely non-standard but very simple to create and use as an administrator – in 2007. It has a free-and-easy approach, which unfortunately does not encourage good discipline. It was no longer supported from 2013, making it a high risk issue for me. More to the point, it could never have been possible to include an Export function. So last year I began to recreate the site on new software. Starting again, I told myself, is good discipline in its own right, forcing us to be tidier and more systematic in our approach second time around.

Now let's deal with data storing and site hosting separately, because they almost always are.

It is common for data to be input online into a themed site, and the cost is often dependent on the number of pages or amount of data. Most major hosting companies include the theme option for good commercial reasons. But we would need to be online to use it – not convenient in slow dataspeed locations. More importantly, our data is also stored online, and only online, unless we create our own Assets folder. We may have no local copy of our website and there may be no easy

way of exporting the online data if we want or need to move the content elsewhere, for cost reasons, for example. Be aware of the term FREE in the advertising.

Purchasing software (not subscription purchase) for use on our local computer or network enables us to work entirely offline and be in control of the number of local copies there are and where else we may keep updated back-ups.

So big question number one is: what happens if the company which created our software goes out of business? If it is a major company such as Wordpress, it would probably be acquired by another successful business. There is always a critical mass of users which enables succession to be the commercial imperative. Smaller firms may be more vulnerable, but if we bought wisely the content can be EXPORTED onto new software; that's why Jack Latimer bangs on about it so often. And in any case, as long as it continues to function fault-free we may have plenty of time to make the change (mine survived 5 extra years).

Now the second big question: what happens if our hosting company goes out of business?

Hosting companies are essentially service providers. They store a copy of the most recent version of our website for a fee and link it to our URL (our web address) so that the world can access it. Which hosting company we use will depend on cost and level of (reliable) service we receive. There will always be hosting companies because without them the internet doesn't stay connected. What happens then if the host we use goes out of business? As with the first big question, they are likely to get taken over or sold to another hosting company – their servers are valuable assets in their own right. That could mean we pay our hosting subscription to a new provider, or the same provider under a different name.

But whatever happens to our hosting company, we'll be OK. We have our current version locally – with x number of backups in different places – so we won't be the ones who will lose our data!

The short answer to our question about *"the risk of storing our data with third parties such as hosting companies"*, is an unquantifiable one if we rely on keeping the only copy with a third party; but an almost zero risk if we maintain x number of copies, including copies of the application software itself.

More cost
If we imagine we have rooted out all of the essential issues about websites we may be responsible for, there is one more to consider.

If the website belongs to us – that is, we own it – we also own the URL or website address.

www.this-site-is-mine.co.uk.   Or whatever you call it.

But to keep owning it we must continue to pay its subscription. The moment we stop paying, the content stops being hosted and therefore seen by the world. I have a list of URLs which I have collected for future reference. A few of those which I bookmarked a few years ago are no longer available, presumably because their owners closed their sites by no longer paying the fee, or not passing responsibility on to another party.

Who says what is on the internet will go on, and on? We will need to make provision for what we have created so that it CAN continue to be utilised into the future.

Mike Neighbour