

1. Testing of the DAAT V4 database

The report below is by Colin Love at ULCC, who was asked to look at the questions in the DAAT V4 database from a sysadmin point of view. He was also given a copy of the *Guidance Notes for Initial Test*.

1.1 File format

When applying the fields to the way we run the TNA [The National Archives] archive they make sense, but I don't think they will in a lot of cases. There seems to be a big assumption that open source formats will be used, but I would have thought files may be in a variety of formats, including proprietary ones (take the OS map format for example). The first question asks if royalties are payable for the format, and everything else assumes an open format.

It may be more useful to drop the term 'source code' in favour of 'file format' and only use source code where you really do mean the source code of tools used to create a particular file format which is more important.

Of course, within the Access database you could have dynamic fields that appear depending on the choices made previously which would cause less confusion.

I note that the other form 'Software' covers the same sort of things, so do you need both anyway?

1.2 Collection Servers: System Management

Generally good. Some questions could do with some sub-questions which may affect the weighting of the outcome. For example it asks if there is a backup policy and where it is stored, but it doesn't ask if the backups are actually checked regularly. The final question about the state of the hardware should probably be covered under 'Assessment of hardware' part.

1.3 Collection Servers: Asset Inventory

No problems.

1.4 Collection Servers: Accommodation

It would be helpful to indicate where the standard can be found for the condition of the server room.

For the fire protection, change halon to 'inert' as halon is no longer in use for computer suites due to its CFC content.

1.5 Assessment condition of hardware

You ask if the new specifications are compatible with older versions, but I think this needs to be clarified. Looking purely at hardware, mostly new versions of kit are not backward compatible with probably the only exception being tape drives which can normally read older formats. You also go on to ask if the specification is complex, large or ambiguous but again this needs to be qualified as we would not see the NDAD setup as any of these, but to somebody running a single server it might be.

Not sure about the question on whether the specification uses digital rights, etc. Perhaps this means something to an archivist but afraid it doesn't mean anything to me.

Lastly, the question of external formatting is spurious on the screen-shot. Can't tell if it is supposed to be there or not as it does not appear in the guide.

1.6 General comments

I think this is quite a useful tool if aimed at smaller organisations that maintain a small archive of data, and perhaps larger ones who don't have much of an idea about what they should be doing.

However, one area we spoke about before was the weighting issue. I think it would be very helpful to weight some of the questions (perhaps not all or not so they make a significant effect as some of the questions are personal choice of working). This will allow the user to judge where they come on a scale, and if done automatically as they complete the form they can see what simple steps they could do to improve their rating.

For example, not having backups at all would be a major concern but simply doing a backup locally would improve the rating, with an offsite copy improving the rating even further. With this sort of feedback the tool would become very useful.

The overall tool should also include an idea as to how good a rating is.

Without some kind of feedback or graph, the user won't know if the rating they receive is good, bad or indifferent. Obviously you need something to create the rating scale initially, and that will take some effort to obtain what a good average marker would be.

My final concern is the number of questions being asked. For a large data centre such as us we can cover all of the questions fairly easily, but for a small organisation with perhaps limited IT skills I think they would struggle, especially if external consultants configured the system or they bought some bespoke software. Some simple questions to ascertain if they manage the setup themselves or use external consultants / support would perhaps change the weighting and add risk if the third-party support company disappeared.

2. Testing of the DAAT V5 database

The comments below are by Gareth Knight of the AHDS, who was assigned the task by Andrew Wilson. Gareth is more proficient with Access and he also looks after the operational preservation activities at AHDS.

2.1 General suggestions

The Access database is generally well designed. However, it is difficult to build an understanding of the interface layout. It would be useful to provide access to specific underlying sections through the main page. If the Access DB is intended to be the final form of the assessment tool (as opposed to a prototype in preparation for a database-delivered web site), it will require user friendly features, such as navigation buttons.

It should be possible to add a new record (the default is to edit an existing record) and lock a record by default, requiring the user to click on an edit button to change an existing value.

Finally, it would improve usability if the ID values were hidden from the user. Unless there are plans to allow the user to perform an ID search, the ID values are intended for internal use only and have little use to the user.

The rest of this report examines particular menu pages.

2.2 Organisation details

Require "other" as an option for organisation type.

The drop-down menu allows entry of other descriptions. Should this be restricted?

2.3 Institutional management

Q3 asks if there is someone with clear responsibility for management. Does this need to be a single person? Several people in the AHDS are responsible for management of different aspects of the asset.

2.4 Section 2: Target collection area

I was confused by the "mostly one file format" attribute. Is it one format or more? It was also possible to select two contradictory attributes, e.g. mostly one format and mixed format.

Should item-level metadata be mentioned on this page? Perhaps it could ask if there is metadata associated with each asset.

The 'size of target collection' should indicate some form of measurement, e.g. number of bytes. Many people will be tempted to enter 2.2TB or other alphanumeric values.

The 'size of parent collection' field appears to mangle values. The value '2418925581107' became '2.418925581107E+12' when I tabbed to the next question.

2.5 Section 3: Hardware assessment

I'm unsure how some of these questions apply to hardware. What is the specification for the hardware? Do we need to know the layout diagram for reproducing the hardware or simply the hardware that is installed in a machine? Many of the questions do not appear to apply to hardware. What does "new version of specification compatible with the old version" refer to in the context of hardware?

The questions suggest that hardware is being discussed for storage and delivery. If this is the case, I wonder if a one-to-one table relationship is appropriate. The department is likely to have several machines that are dedicated to a specific function. Would it be useful to distinguish between different hardware? A warning could also be raised if the institution is performing preservation action, preservation storage and distributing assets on the same machine.

Would it be useful to include a 'don't know' answer to some questions? The software could flag it as an area for concern.

The checkboxes do not appear to be working correctly for the questions "is the specification complex/large/ambiguous" and "which of the following is not used by the specification". They cannot be de-selected.

Please can you clarify the need for signed envelopes and watermarking when discussing hardware?

2.6 System Management

The question that asks if a management system generates an audit trail is too clear cut. Is it possible that users will answer "to an extent", suggesting they have audit trails for some, but not all tasks?

The backup location checkbox requests the user to select one option only. However, both "locally and offsite" and "online offsite" options are selected at the same time.

Would it be useful to assess the adequacy of the backup policy? Many of the assessment criteria are mentioned on a later page.

2.7 Section 7: Preservation Action

Many users will be bewildered by the fixity (4) question, regarding AIP encapsulation or storage as a separate file. I suggest the question is reworded to distinguish between the two options.

It is unclear if the preservation page is discussing preservation policy as a whole or preservation action performed on the identified collection. The questions on the left suggest the former, while the transformation questions suggest the latter. Most notably, I do not think it is possible to indicate if the transformation process is reversible and repeatable for every asset in the collection. Therefore, the YES/No answer will have only limited value.

2.8 Section 8: Identification

My interpretation of the ID information page is that it is intended to store details regarding a specific asset, e.g. a web page, MS Word file, etc. The majority of questions confirm the view, with the exception of "how many multiple file formats?". If I have interpreted the question correctly, it may be better located on the "target collection area" page as an addition to the mixed file format attribute.

The metadata in "location metadata" and "environment metadata" may cause confusion. Perhaps it could be reworded to "storage location" and "operating environment"?

2.9 Section 11: Software assessment

I'm uncertain of the usefulness of some of these questions.

The question "are new versions of software source code compatible with old versions" is ambiguous. I suspect I may have misinterpreted the question.

It does not matter if the source code is the same. It could have been completely rewritten, it only matters that the new version is able to read data produced by the old version. An additional question may focus on this aspect of the software application, "how compatible is the application with older versions of the format". Possible answers could be "compatible with all versions of the format", "compatible with recent versions of the format", N/A.

How useful is the question "Does software source code contain embedded metadata which cannot be mapped into other formats"? This seems like a file format question.

Similarly, I'm not certain that software source code needs digital rights, encrypted sections, watermarks, etc.

2.10 Section 12: Media Condition

"Can the media be read by the CD/DVD drive or tape drive into which it is inserted?", "does it work in a different drive" and "disk surface" requires a N/A option. The questions do not apply to hard disks. What about USB memory sticks?

"Does the disk surface show signs of surface damage" and "the item has been assessed and is not damaged" appear to ask the same question from opposing viewpoints. Perhaps the second question could be reworded to indicate if the disc is readable or not.

2.11 Section 13: Media System Management

It would be useful to establish some visual distinction between questions related to the media carrier and those related to the management system.

The question order is also a little confusing. Many institutions will use media carriers for backup only. Some distinction may need to be made between media carriers intended for access and media carriers used for backup only.

The first question could apply to access and backup media carriers.

I'm not certain of the expected answer for the last question, "is your system computer hardware in a generally good state". If it is asking if the carrier reader hardware works or not, an alternative question could be, "has your media reader hardware (CD-ROM, tape drive, etc.) been checked recently for hardware faults." and "Did it pass all criteria established in the test?"