Defra experiences of using DROID

Mike Kaye

Kate Pritchard

Tony Wallis

Stephen McGrath

The Wider Electronic Information Picture

- Defra is looking to gain proper control of all its electronic information; much of it unmanaged since desktop IT was introduced in 1994.

- This currently resides in Shared Drives, Accredited Shared Drives, CD's, SharePoint Teamsites, Datasets and Databases,

- Considerable progress here:

  o Information Audit programme started in August 2009; captures details of all information assets owned by each Business area i.e. Project and Programme – to be completed March 2011.

  o Analysis is confirming who owns what, where it is, value, high profile information and orphaned information.

  o Appraisal Teams have been successfully re-skilled to apply macro appraisal techniques to physical information "top down Research based review" – this can be applied also to electronic information, in line with TNA's Appraisal Policy.

Total storage on 21 file servers, accumulated by Year, gBytes as at October 2011

Getting DROID installed
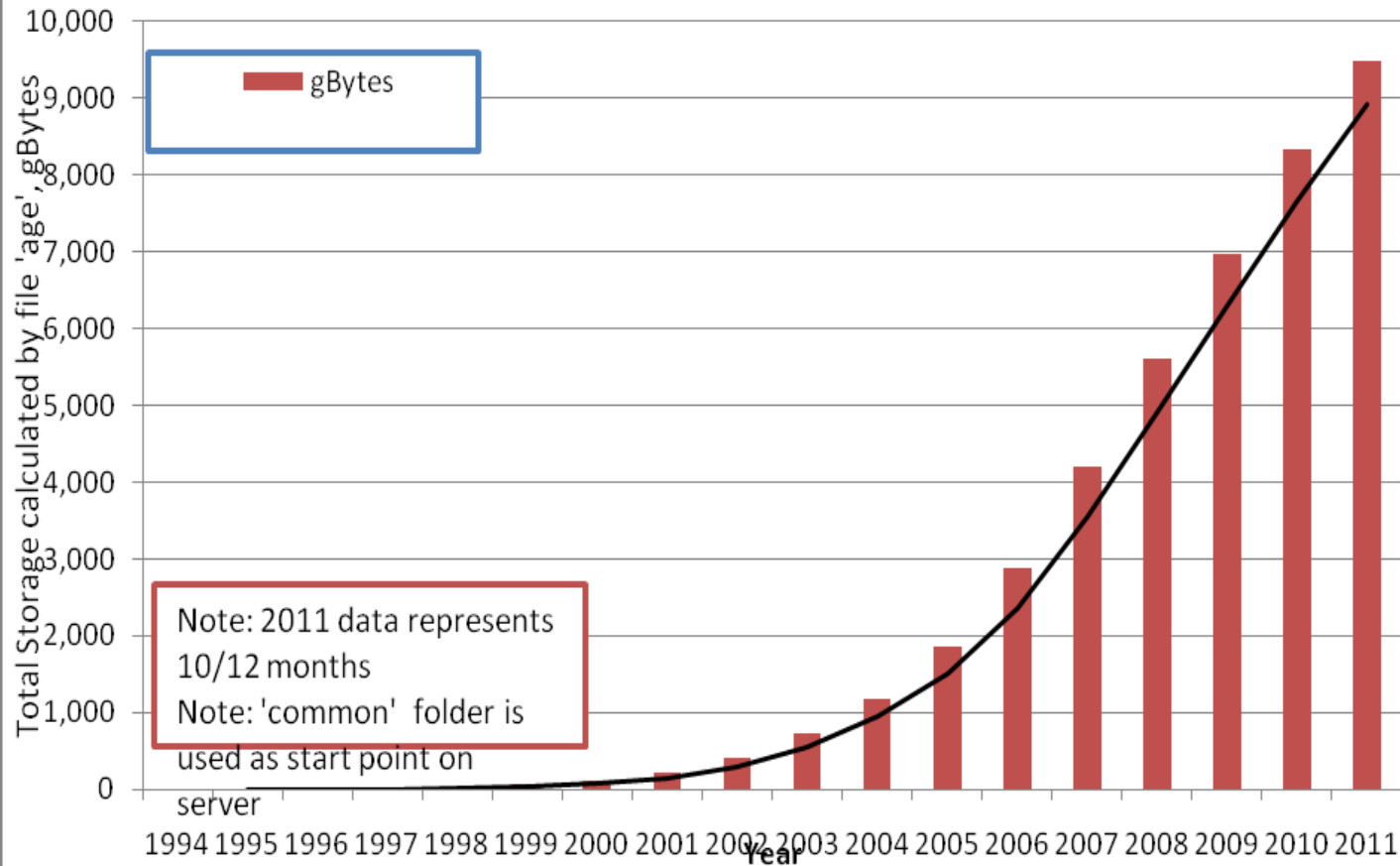
- Getting permissions from our IT provider to install:
  - This took over 6 months
  - Required testing security and impact on server performance
  - Issue with installing Open Source Software so required exception.
  - Incurred payment to install free product!
- Problems with our version of Java meant that it would not load.
- Visited TNA to overcome this and the programmer gave advice.
- Subsequent installations handled by our IT provider.
- DROID profile held on hard drive of machine due to the time taken to run from a server. This was a TNA recommendation.

Running DROID

- We ran DROID within the Information Management Team rather than through our IT provider to keep the cost down. This meant that we gathered first hand experience of the content on our servers.

- Extremely slow possibly due to:
    - The speed of our host laptop
    - The large volume of information (10Tb)
    - The location of our servers (spread across the country).

- Needed to run overnight and disable the hibernation on the laptop due to the time it would take to run DROID.

- IT Provider ran tests to assess the impact of running DROID on the traffic on the servers. They confirmed that this would not be a problem.

- We needed to secure two additional laptops dedicated to running of DROID. As this took up so much processing power and could not be used on our own machines.

- DROID failed in running when areas were very large – over 100Gb and also when running over 4 days. DROID would provide the point where processing got to but could not continue from that break point.

- It is possible to run two or three DROID scans at the same time. Any more might lead to completion time issues.

Challenges

- Installing DROID

- Using DROID 5 the filter did not work for us. Improvement in DROID 6.1. the filter enables us to manipulate the output much more easily without having to export to Access etc.

- Identification of duplicates dramatically slows DROID 6.1 down. We have had to run with this turned off.

- Can be slow to open the area you wish to analyse particularly if the server is not local again this may possibly due to our lap top.

- Slow running across servers. This means that the information on the servers may have changed since analysis began.

- Getting access to our information! Working with our IT provider to gain overall read access across all servers. We have ongoing issues regarding our permissions.

- Need to watch what is happening behind the scenes where our IT provider or business areas may move data post our analysis.

- Important to assess the size and the number of objects of the servers before running DROID. We couldn't look at an area over 100GB because the results were not manageable. CSV reader is Excel and the volume of data was too large.

- Security issues associated with having laptop logged on continuously. We used a secure room so we could run it overnight. Also need to consider the output and the sensitivity of the data and where this is stored.

Potential developments and user needs for future phases of DROID

- Be quicker! Particularly for identifying duplicates where quick wins can be achieved for reducing redundant information.

- A progress bar % complete would be a useful addition rather than a bar.

- DROID has a friendly user interface and is straightforward to use. We are also using Robocopy a lot now because it is so much faster. Looking at Treesize.

- Linking to PRONOM to inform the deletion of digitally obsolete information.

- Carrying out actions on files identified.

  - We have developed a deletion script to enable us to delete redundant information e.g. Duplicates. We had 10% of space on one server taken up with duplicate copies of the same power point presentation!

Thank you

- Questions?

Kate.Pritchard@defra.gsi.gov.uk

Mike.Kaye@defra.gsi.gov.uk