

MIXED: standards for realizing digital preservation

One of the obstacles to preserving data is software obsolescence. Digital data is produced by means of software tools, and these tools leave their mark on the data. It turns out more often than not that data is not accessible by tools of a different vendor, or even by different versions of the same tool. This stumbling block to preservation is usually tackled by either continuously migrating the data or by emulating the software tools. MIXED proposes a third way: preservation by migrating data to standard formats.

Because there are many kinds of digital data – documents, spreadsheets, statistical tables, databases, emails, images, etc. – it is not feasible to establish one standard format. The scope of MIXED is tabular data, namely spreadsheets and databases. The task of MIXED is to select or create a standard format for spreadsheets and databases and to build converters between the vendor formats and the standard formats.

This approach is generic and can be applied to other kinds of data, for example statistical tables. For every kind of data, a standard can be selected or created for interchange and preservation. To make it all work, converters between other formats and these standards have to be built.

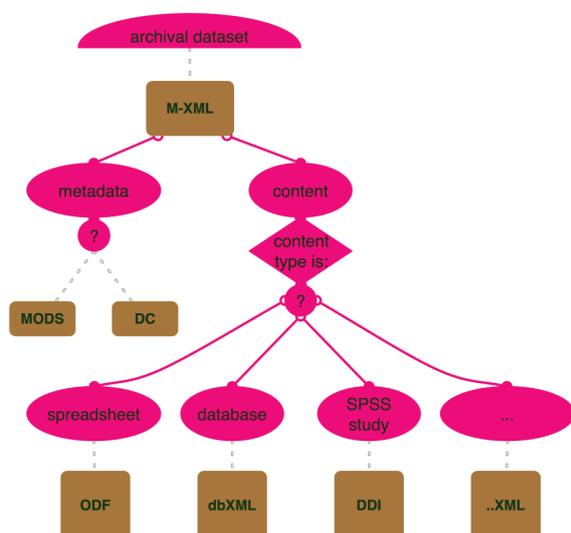
The intention of MIXED is to define an umbrella format for data preservation and interchange. The umbrella format will comprise a wrapper with metadata and provenance data, suitable for archives. It will also select an appropriate standard for each kind of data.

This umbrella is both open – meaning that new kinds of data can be added to it – and closed, meaning that if some kind of data has been added to the umbrella, it

will not be possible to add other standards for the same kind of data. The software framework created by MIXED is open, meaning that it is designed as an open source framework, accepting converters as plug-ins.

As the umbrella grows and the conversion software matures, we will have a new and powerful data preservation strategy: MIXED – Migration to Intermediate XML for Electronic Data.

web site: <http://mixed.dans.knaw.nl/>



Postadres Postbus 93067, 2509 AB Den Haag

Bezoekadres Gebouw Borneo, Anna van Saksenlaan 51, 2593 HW Den Haag

www.dans.knaw.nl, info@dans.knaw.nl, Telefoon 070 349 44 50, Fax 070 349 44 51

DANS is een initiatief van de KNAW en NWO