MICE Data Management Plan (DMP)

The full MICE Data Management Plan covers all aspects of MICE data custody, movement, cataloguing & publishing, reconstruction and simulation, as well as data access both within the collaboration and by people outside of MICE. Only the briefest summary of this plan is described here.

By the standards of High Energy Physics, MICE is a small experiment and data management is organised in such a way as to ease the burden on collaborators as well as meet the open data requirements of our funding agencies.

MICE Data policy

MICE believes in open data. Being UK based we have specifically followed the guidance of the UK funding agencies (which can be found at http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx), however we believe these to be completely in line with those of our other funding agencies.
MICE is still in the process of reviewing our data policy statement but the currently active version (since March 2012) reads:

"The Muon Ionisation Cooling Experiment (MICE) has a policy of open data. This means that all of our data and meta-data is freely available to the general public and, more specifically, other members of the scientific community. However, we do place conditions on your use of our data. These are:

1. Any publication which uses MICE data but that is not produced by the MICE collaboration or has not gone through the MICE internal review procedure must contain the following statement:

"We gratefully acknowledge the MICE collaboration for allowing us access to their data. We also acknowledge that our use of these data to produce this publication does in no way make this publication sanctioned by the MICE collaboration. We further accept responsibility for any errors in our understanding of the MICE data"

2. Support for non-MICE users examining our data will be at a best efforts level.

We would consider it courteous if you informed us of any planned analysis publication using our data at the time paper submission.

The MICE Collaboration."

The nature of MICE data means that we have no privacy issues and can make all our data freely available. MICE also supports open access journals and the MICE policy for the public presentation of MICE data is described in MICE-note 352 which is available at

http://hep04.phys.iit.edu/cooldemo/micenotes/public/pdf/MICE0352/MICE0352.pdf

MICE is also committed to the open access of bibliographical meta-data as described in the UK funding policy to be found at http://www.stfc.ac.uk/rgh/rghDisplay2.aspx?m=s&s=64 .

MICE Data Flow, Data Custody and Data Availability

The primary objective of data management is the safe custody of the data collected by the experiment. If the data are lost the money spent on performing the experiment is wasted. Within MICE this is achieved by carefully controlling the passage of data as it leaves the experiment.

Data collected by the experiment at RAL are aggregated on a disk system that is shared between the online and offline systems of MICE. As soon as a data-collecting run is complete, the data from that run are copied from this disk system to the central tape vault in RAL's e-science data centre. Here they are copied onto two separate tapes, which diminishes the possibility that tape failure would result in data loss. The data are then also copied to a separate disk system, outside of the e-science data centre and made available to the physicists for analysis. At each stage checksums are taken of the files to ensure that there has been no data corruption during the transfers, which ensures that all copies are indeed faithful copies of the original. It is only at this stage i.e. when the data are on two tapes and a separate disk system that a flag is set which will allow these data to be deleted from the original disk system. Data are then asynchronously (typically overnight) copied to a selection of MICE sites using a data moving tool, again checksums are taken to ensure the fidelity of these copies. One of the sites to which all data are copied is Imperial College London and there all data are made free available on a publicly available website. As well as complying with our open data policy, this makes it easy for MICE collaborators to download the data that they want to analyse. This has become the method by which most MICE scientists access the data. All MICE data is available at http://www.hep.ph.ic.ac.uk/micedata/

Data types
Currently MICE only has the raw data written from the experiment. A reconstructed data type is being developed which stores data in the commonly used root format. Monte Carlo simulations of the experiment are also being developed. Once available these data types will be stored and made available in the same way as the raw data is currently.

Software
The MICE software framework, called Maus, is open source under the GNU GPL v3 license and is available at https://launchpad.net/maus

Meta-data
MICE meta-date and conditions data are all stored in the Configurations Data Base (ConfigDB). Access to the ConfigDB is via webservices and any user using the Maus software framework may access these data automatically and transparently when needed. Data stored in the ConfigDB is multiply backed up.

Archival and Future Use
The MICE data is automatically archived through writing it to tape as soon as it leaves the experiment. The Maus software framework is stored on "launchpad", which is a site specifically for the safe handling and distribution of open source software. RAL have a commitment to maintain the ConfigDB for the lifetime of the experiment. Beyond the lifetime of the experiment, there will be no new meta-data or conditions data and so a flat file will be made of the data stored in the ConfigDB, which will be stored in the same repository as the Maus software.

Through our choices and plans we believe that MICE data will be safe and analysable long after the experiment has been completed.