



ARC Guide to Archiving and Managing Complex Data

Prepared by Lara Perez-Felkner



Center for Advancing Research and Communication in Science, Technology, Engineering, and Mathematics

NORC at the University of Chicago

1155 East 60th Street

Chicago, Illinois 60637

arc-info@uchicago.edu

arc.uchicago.edu



Common Database Complex and/or Longitudinal Database Issues for REESE PIs

Data Archiving and Storage

Popular programs for managing complex databases are Oracle, Microsoft SQL Server, and Microsoft Access. Oracle and SQL Server are comparable programs intended for very large complex databases. These programs tend to require users with high technical skills and are used primarily to archive and store files, not for analysis. Universities and survey research institutes such as NORC tend to use these programs to manage their longitudinal studies. Microsoft Access tends to be easier for researchers to use who do not have specialized expertise in database management. Significant price differences exist as well. SQL Server and Oracle are considerably more expensive than Access, which is an off-the-shelf program which comes with the Microsoft Office Suite. Alternately, add-ons to statistical and other analysis software might make archiving data within those programs more feasible.

Interfacing between Database Software and Statistical and Other Analysis Software

These database programs are for storage and retrieval of data, and are not meant to be interfaces for researchers. They have moderate statistical functionality however it is more common and powerful to use separate analysis software, either by exporting between software programs or using a program like StatTransfer to link between database software and analysis software. Some programs have interfaces or extensions that provide more seamless interaction between database software, like Access or Oracle, and commercial statistical packages, like Stata or SAS?

Managing Video or Other Large Data File Types within Databases

Both analytic and database management software tend not to be designed for managing video and similar file types, which can make single database files too large for easy and reliable use. Rather, IT professionals at one's university or research institution can write scripts to link video, picture, and similar files to respondents' data within the primary database. They would write a script that would serve like a hyperlink to a file on the hard drive or server, which would open with the appropriate program (Internet Explorer, Word, etc.) and it would do the task you want



(i.e. show video).¹ Although might involve considerable work on the programming and data loading end, it should be easy to use for analysis once the programming has been completed.

Managing an Inter-University or Other Multi-User Database

Numerous services are available for making data available for multiple users and even public use. For example, the Data Enclave at NORC provides services to make de-identified private data available on a secure server for others to download and use. The Interuniversity Consortium for Political and Social Research (ICPSR) and other data centers provide similar services.

Where to Go from Here

The ARC Center will offer more in-depth information on database management at the 2010 REESE PI Meeting. We encourage you to seek out resources at your home institution. Researchers commonly face some challenges when asking for database management help from IT professionals. Specifically, IT specialists in database management tend to be highly specialized in particular domains of database management and tend to use these programs for administration rather than for analysis. Efforts to frame your questions in specific, technical language can help the process toward a solution that will be useful.

¹ The IT professional would create a hypertext, giving properties to the link to tell it what program to use to open the file.