Making the most of data

Dr Barry Radler is working to help social scientists better document their methodologies and data. Here, he discusses some of his approaches to solving the problems faced by data managers in handling complex longitudinal datasets

In what way has your focus evolved over the years?

My career has centred on understanding how people process information, make decisions and behave in social, political and marketing contexts. While obtaining degrees in mass communication, psychology and consumer research, I explored different perspectives and – quite unintentionally – developed a thorough and broad skillset in experimental, survey and marketing research methods.

I realised some years ago that I have become a 'boundary' person between research and IT. My forte is discovering and exploiting opportunities where IT can benefit social science methods. My interests in social psychology, marketing and persuasion continue, and the methodological tools I have developed have allowed me to move between disciplines. During the past 10 years, my work has highlighted the next big challenge: research data management in the context of large multidisciplinary science.

How does the Midlife in the US (MIDUS) study utilise metadata?

Metadata have been around for a long time; but now researchers are operating in increasingly data-rich electronic environments where storage, curation, reproducibility, harmonisation and discoverability issues are becoming as critical as the data themselves. Metadata help researchers to better distinguish the content and quality of datasets available to them in a digital age.

MIDUS adopted the Data Documentation Initiative (DDI) metadata standard for its electronic codebooks, so its metadata are human-readable but also machine-actionable. MIDUS content traverses many disciplinary boundaries, which necessitates high quality metadata to facilitate analyses among researchers investigating integrative questions that may exceed their specific areas of expertise.

Can you explain how the DDI standard works?

DDI metadata describe each variable in a dataset – such as question text, interviewer instructions, response options and value labels – and also the larger study itself, including information such as funding source, investigators and methodologies.

With the advent of the internet, the fields within codebooks could be standardised and structured via a powerful web-based markup language called XML, which is what DDI does. Simple stylesheets can then convert XML to display the DDI codebooks on the web. So machines can read, search and intelligently use DDI codebook contents. Codebooks that exist as .txt files or Word documents can also be searched, but not intelligently, as they lack DDI's formal structure.

In what ways does your work with MIDUS contribute to the development of DDI?

The characteristics that attract researchers to MIDUS – multidisciplinary data on longitudinal samples – are the same that pose unique challenges to data managers. Documenting the relationships between versions of variables is of paramount interest to a longitudinal study like MIDUS. My previous work with MIDUS and DDI resulted in a metadata portal that contains electronic DDI codebooks for all publicly available MIDUS datasets, as well as links to official data, instruments and documentation.

I currently work on a project that aims to clarify the related nature of longitudinal and cross-cohort versions of MIDUS variables, and develop a harmonised custom data extract, all driven by DDI. My project will use a combination of concordance tables and new DDI functionality to explicitly link longitudinal or cross-project variables of comparable content. It will also support internal harmonisation of data by documenting differences in question wording, administration or coding that compromise the assumption of strict equivalence among related variables. All such information will be 'baked' into the DDI codebooks that reside on the portal so researchers will have this information when they download a custom dataset.

The custom data extract is a major innovation that will enable researchers to combine, merge and manipulate data from different MIDUS samples, projects and longitudinal waves. Other systems provide custom sub-setting of downloaded datasets, but this tool will support accurate merges across different datasets and will provide harmonisation information for related variables, as well as custom DDI codebooks

To what extent has data management benefited MIDUS?

MIDUS data are among the most frequently downloaded datasets from the US National Archive for Computerized Data on Aging at the Inter-university Consortium for Political and Social Research (ICPSR), which houses one of the largest digital data archives in the world. Since 1999, well over 64,000 datasets have been downloaded through ICPSR and more than 24,000 individuals from across the world have accessed MIDUS. What sits behind this high volume usage and high-quality productivity is the committed data management philosophy of MIDUS.

Managing midlife metadata

The **University of Wisconsin** Institute on Aging is leading a longitudinal study to improve understanding of how Americans age in midlife – a study that uses unique data management techniques to document complex research datasets

TODAY, DATA ARE available in greater quantities than ever before. The advancement of technology has provided access to increasingly complex and rich datasets that were unthinkable just 20 years ago. In the context of research, this change provides new challenges regarding the proper management of data. Fundamentally, the question is how to improve the discoverability, storage and quality of the overwhelming amount of data now at our fingertips.

As a result, there is a growing need for interdisciplinary efforts to bridge the gap between the technical demands of data management and the goals of scientific research. Dr Barry Radler, a social scientist by training, is used to operating in this zone. Over the past 20 years, Radler has applied technology to enhance social science research methods. Now working at the University of Wisconsin Institute on Aging, he is managing data within the Midlife in the United States (MIDUS) study, which aims to better understand healthy, long-term ageing using information collected from thousands of Americans.

ADDRESSING MIDLIFE QUESTIONS

Childhood, adolescence and the final years of life have been extensively studied, yet the longest section of life – midlife – remains poorly understood. Recognising this, a multidisciplinary team of scientists set out to understand how health changes across these decades, and how behavioural, psychological and social factors interact to influence health outcomes.

The story began in 1995 when over 7,000 adults in the US aged between 25 and 74 participated in phone interviews and completed questionnaires. The study broke new ground on multiple levels; it included unprecedented assessments of psychological and social factors, linked to demographic parameters. It also had wide scope in assessing health as well as work and family life. Finally, it facilitated additional studies around the main sample. These 'studies within a study' offered in-depth assessments of other significant factors such as daily stress.

Thanks to funding from the US National Institute on Aging awarded in 2002, MIDUS became a longitudinal study when a follow-up of the original 1995 sample was initiated. The content was expanded to include in-depth biological and neuroscience measures, and a sample of 600 African-Americans was added to improve understanding of race and health. In 2011, the study expanded once again by adding a further 3,500 participants and an additional focus on the impact of the economic recession. Soon after, a third round of interviews with the original respondents began. Today, MIDUS includes data from over 13,000 people, comprising thousands of variables across different scientific areas, and multiple samples and cohorts.

MANAGING METADATA

This rich dataset requires proper management. Drawing on his extensive experience in social science research, Radler is responsible for overseeing the survey data collection in MIDUS, ensuring the quality and integration of its many datasets. He is also responsible for the management of metadata, which is appended to all of MIDUS' datasets. Metadata are essential for the storage, discovery, management and analysis of data. In the context of MIDUS, metadata help researchers to quickly identify the content, relevance and quality of data.

But metadata have another use: supporting MIDUS' dedication to providing clear documentation with all publicly available data. Radler explains this commitment to public sharing: "We are federally funded with a mandate to make our data available to the wider scientific community. We do this through an official archive, the Inter-university Consortium for Political and Social Research". As a result, there are now thousands of secondary users of MIDUS' data, making accurate documentation vital.

The complexity of ageing, and the multidisciplinary approach it necessitates, also make metadata important for this study. "Accurately understanding profiles of health requires integrating data from diverse fields, which necessitates high quality metadata to facilitate analyses among researchers," Radler elaborates.

DOCUMENTING RESEARCH

To accommodate all these factors, MIDUS requires a robust documentation system. Indeed, the innovation at the core of the project is a metadata standard – the Data

RESEARCH DATA MANAGEMENT

OBJECTIVE
To utilise the Data Documentation Initiative (DDI) metadata standard in the Midlife in the US (MIDUS) study to improve data quality and facilitate research investigation and analysis.

KEY COLLABORATORS
Dr Carol D Ryff, Director, Institute on Aging, and Hilldale Professor of Psychology, Department of Psychology, University of Wisconsin, USA

PARTNERS
Colectica Software, USA

University of Michigan Inter-university Consortium for Political and Social Research, USA

FUNDING
US National Institute on Aging

John D and Catherine T MacArthur Foundation Research Network on Successful Midlife Development

CONTACT
Dr Barry T Radler

University of Wisconsin-Madison
Institute on Aging
2245 Medical Science Center
1300 University Avenue
Madison, Wisconsin 53706-1532
USA

T +1 608 262 2056
E bradler@wisc.edu

www.midus.wisc.edu

RG **Barry T Radler**

in **Barry T Radler**

BARRY RADLER is a researcher at the University of Wisconsin Institute on Aging. With a background in marketing and psychology, he obtained his PhD in Mass Communication at the University of Wisconsin-Madison. Since 2005, he has managed the MIDUS study, documenting it with a metadata standard called DDI.

Documentation Initiative (DDI). The DDI formally structures all information that is generated by the study and makes it machine-actionable so that different software and network systems can understand and use that information intelligently.

At its core, this metadata standard aims to make research data understandable. "DDI metadata describe each variable in a dataset as well as the larger study itself, including information such as funding source, investigators and methodologies," Radler details. As well as human users, it is essential that computers can interpret the metadata. Therefore, DDI codebooks contain fields that are standardised using the web-based markup language XML. XML allows DDI codebooks to be displayed on the web, such that both humans and machines can read, search and use them.

Importantly, because DDI is open source, anyone can use and improve it. Under Radler's supervision, the MIDUS study has contributed to the evolution of DDI in recent years, allowing the standard to develop from an electronic codebook format for documenting individual studies to a method that can document the entire research data lifecycle.

INTEGRATIVE ANALYSES
Just as the DDI has grown and developed, MIDUS too has evolved to become a resounding success, gaining the support of the wider scientific community. The number of publications using MIDUS data continues to increase year on year; more than 600 publications from the study now appear in over 150 scientific journals.

Shedding new light on a little understood stage of life, MIDUS has also become a case study for good data management. Those same characteristics that attract scientists to the study, ie. the wide-ranging multidisciplinary data on diverse longitudinal samples, also provide challenges for data managers. By developing innovative tools, Radler has capably managed these challenges, and helped the project's scientists to navigate the varied datasets produced by the study.

An essential part of Radler's DDI work is the metadata portal, a 'one-stop shop' for all MIDUS data. "The portal has been innovative in itself, but also highlights areas in DDI that were inadequate for documenting a study like MIDUS," Radler adds. Thus, Radler's current project aims to build on these areas, utilising novel DDI functionality to link longitudinal or cross-project variables of comparable content. Moreover, by documenting any differences in coding, transformation or administration across comparable variables, the portal will support harmonisation.

Another groundbreaking function of the portal will be the ability to extract customised datasets from all available MIDUS data. This

WHAT MAKES MIDUS UNIQUE?

Researcher Barry Radler highlights the characteristics that make this study one of a kind:

- The baseline national sample included over 7,000 people. In addition to a randomly selected sample, MIDUS included one of the largest national samples of twin pairs in the US and as a sample of respondents' siblings

- MIDUS has a wide age range (25-74). This allows researchers to study individuals as they transition through two key phases of life: from young adulthood to midlife, and from midlife into old age

- Quality psychological and social assessments were rigorously measured and modified before use and many cases represent first use outside of small samples. The survey assessments are substantial, including a 45-minute phone interview as well as two questionnaires almost 100 pages long

- It is integrative and multidisciplinary. Data are collected across multiple domains, from the same respondents, and considered alongside laboratory assessments. The end result is extensive data from many different domains of life

- MIDUS is continuously evolving. The third round of survey and cognitive data collection has just been completed. The study has expanded with new samples, including a new national sample of 3,577 people, a sample of 1,100 African Americans and a cross cultural sample of 1,027 adults in Japan

will enable researchers to do something never before possible: obtain customised, cross-project downloads of harmonised MIDUS data that are DDI-compliant. "Ultimately, this tool will produce a tighter relationship between MIDUS data and metadata, facilitate the cross-fertilisation of different types of data, and allow researchers to devote more time to integrative data analyses," Radler concludes.