



## Planning Grant: Developing a National Higher Education Student Unit Record Database

**Dr. Matthew W. Ohland, Purdue University and Central Queensland University**

Matthew W. Ohland is Professor of Engineering Education at Purdue University and a Professorial Research Fellow at Central Queensland University. He has degrees from Swarthmore College, Rensselaer Polytechnic Institute, and the University of Florida. His research on the longitudinal study of engineering students, team assignment, peer evaluation, and active and collaborative teaching methods has been supported by over \$12.8 million from the National Science Foundation and the Sloan Foundation and his team received Best Paper awards from the Journal of Engineering Education in 2008 and 2011 and from the IEEE Transactions on Education in 2011. Dr. Ohland is past Chair of ASEE's Educational Research and Methods division and a member the Board of Governors of the IEEE Education Society. He was the 2002–2006 President of Tau Beta Pi.

**Mr. Russell Andrew Long, Purdue University, West Lafayette**

Russell Long is Director of Project Assessment at the Purdue University School of Engineering Education. He has extensive experience in assessment and student services in higher education and has worked for eight years as the Data Steward of the MIDFIELD project.

# Planning Grant: Developing a National Higher Education Student Unit Record Database – NSF REE Grant 1232740

## Project goals

Retention is the dominant metric in studying student success in engineering education and in higher education in general, yet available national datasets do not facilitate establishing national benchmarks. This project sought to build on the earlier development of a large longitudinal dataset to design a national, longitudinal, student unit-record database that would make it possible to calculate retention and other metrics consistently. This resource would permit benchmarking, peer comparisons, and the design of new metrics backed by community support.

## Major activities

Faculty, institutional researchers, and college and university administrators provided input to the design and protection of a national unit-record data system. Two distinct data use agreements were designed—a “restricted data use agreement for research” and a “restricted data use agreement for institutional analysis”. These agreements are attached as supplementary documents.

The contract to administer these agreements was signed by Purdue, is in draft form at Virginia Tech, and is under review elsewhere. Twenty-two institutions have provided letters of support that promise to release extraordinary amounts of data for study by the engineering education research community. *This sends a resounding message that these institutions understand the potential impact of this database and are eager to learn from it to benchmark and improve their programs.*

## Specific objectives

Design a national student unit-record database while addressing the issues of privacy, confidentiality, cost, logistics, and trust.

*Data should be accessible to a broader community of researchers.* Institutional representatives interviewed recognized the benefits of allowing the research community to have access to a national student unit-record data system. In addition to accelerating the work of current engineering education researchers, permitting access to a broader research community would attract the research interest of demographers, sociologists, statisticians, and others to research questions of interest to engineering education. A few access restrictions were suggested, however. The responsible investigator on any research project should have a PhD or other terminal degree and hold a faculty appointment or research position at an institution of higher education or other not for profit agency that has a demonstrated record of using confidential data.

*Partner institutions must not be affected negatively by published research results.* To protect the partner institutions, names of MIDFIELD institutional partners should not be associated with specific statistics or calculations. Tables and figures displaying results should use labels that

mask the identities of institutions in the data. Institution names may be used only when data is aggregated across more than one institution, and only then so long it is not possible to deduce information about a single institution.

***Partner institutions should have special access to conduct peer comparisons.*** Institutional representatives were clearly interested in the opportunity to use MIDFIELD data to conduct peer comparisons in greater detail than they have access to with currently available data. At the same time, they were unwilling to allow other institutions to have that level of access to their data without some indication of shared risk and trust. Further, findings from such studies should not have the opportunity to have a negative effect on institutions. Such peer comparisons can be directed by investigators who hold a full time appointment at an institution that has contributed data to MIDFIELD. The results from such peer comparisons must be used solely for institutional analysis and only information pertaining to the institution itself may be made public.

***All institutions should have equal access to benefit from the MIDFIELD partnership.*** To ensure that MIDFIELD does not become a resource that further privileges schools that have the resources to participate, but that is out of reach of institutions with less resources devoted to institutional research efforts, it is necessary to identify external funding for institutions to extract historical data and join the MIDFIELD partnership. Yet admission to the partnership is not sufficient to level the playing field. Well-resourced institutions are more likely to have highly skilled researchers who conduct research and publish findings based on MIDFIELD. This benefit cannot be granted to MIDFIELD partners, but a corollary benefit can be assured – that less-resourced institutional partners benefit when other institutions conduct research using MIDFIELD. For this reason, while published research that generates institutional findings must mask institutional identity, institutions must privately be informed of their own identity. Thus researchers at all institutions using MIDFIELD provide an institutional research benefit to the MIDFIELD partners.

## **Major accomplishments**

The key outcome of this work was developing a plan for expanding participation in and access to the MIDFIELD database. The Interuniversity Consortium for Political and Social Research (ICPSR) specializes in handling and sharing large datasets. The PI negotiated a complex restricted-use data dissemination agreement that will be executed among three parties: the MIDFIELD partner institutions provide institutional data, the MIDFIELD team converts the institutional data to the MIDFIELD common format and transmits the common format data to ICPSR, and ICPSR archives the data, administers and enforces data use agreements, and provides access to the data to investigators who execute data use agreements. Two distinct data use agreements have been developed to implement the requirements above: a “restricted data use agreement for research” and a “restricted data use agreement for institutional analysis”.

The Restricted-use data agreement is a major outcome in itself. It is a legal document designed to make it easier for institutions to recognize the value of joining MIDFIELD and to remove legal barriers to the extent possible.

In summary, the process for adding institutional data to MIDFIELD is:

- once an institution executes the MIDFIELD Memorandum of Understanding, Purdue will seek support for the extraction of historical data;
- institutions provide institutional data to the MIDFIELD team;
- project personnel convert partner institution data into the MIDFIELD common format while negotiations begin to execute the restricted-use data dissemination agreement between ICPSR, MIDFIELD, and the institution;
- common format data will be placed in ICPSR; and
- investigators can access common format data under a data use agreement.

## Dissemination

- Huerta, E., M.W. Ohland, R.A. Long (2013). *The Impact of Social Integration on First Time in College Engineering Students Persistence, Longitudinal, Interinstitutional Database Analysis*. American Society of Engineering Education Annual Conference. Atlanta, GA.
- Ohland, M.W., and Long, R.A. (2014). *Planning Grant: Developing a National Higher Education Student Unit Record Database*. American Society of Engineering Education Annual Conference. Indianapolis, IN.
- Ohland, Matthew W., Lindsay, Euan D., Long, Russell A. (2013). *Designing a data partnership to understand the engineering education system.. Australasian Association for Engineering Education Annual Conference" Work Integrated Learning – Applying Theory to Practice in Engineering Education*. Gold Coast, Queensland, Australia.
- Matthew W. Ohland, Russell A. Long, George Alter, Felice Levine (2013). *Restricted-Use Data Dissemination Agreement*. A legal document describing how data will be provided to the MIDFIELD group, adapted to a common format, and archived by the Interinstitutional Consortium for Political and Social Research and giving permission for the transfer of data. This document includes separate appendices for a Research agreement and an Institutional Analysis agreement..

In addition to these publications, the data use agreement and the model of a national student unit-record database designed of this project was shared with a large number of institutions (many of which had been contacted in the design process).

The work of the project through a panel and two seminars as well:

- Ohland, M.W., “How Institutional Policy, Curricular Structure and Program Culture Affect Engineering (and Other) Students”, Learning and Teaching Education Research Centre, Central Queensland University, 3:00 pm, November 8, 2012.
- Atman, Cynthia J., Ohland, Matthew W., and Paretti, Marie C. (2013). Multi-Institution Research Best Practices, Panel T665, moderators Kelly J. Cross and Kathryn F. Trenshaw, sponsored by the Student Division and Educational Research and Methods Division, June 25, 2013, 4:00-5:30 PM, Atlanta, GA.
- Ohland, M.W., “Developing, Studying, and Sharing a Longitudinal Database of Student Educational Records”, Engineering Education Seminar Series, Purdue University, February 14, 2013.

## Impact on engineering education

With input from university data managers, engineering education administrators, specialists in high-volume data archiving, and engineering education researchers, a plan has been developed that will build on the successful development and study of an existing longitudinal dataset containing student records for more than 1,000,000 unique students. This proposal would triple the number of institutional partners contributing whole-population data to the Multiple-Institution Database for Investigating Engineering Longitudinal Development and, at the same time, negotiate the release of MIDFIELD common format data to a data archive where researchers and institutional researchers could access the data under data use agreements that protect the rights of students and institutions. The promise made by 22 institutions to release extraordinary amounts of data for research by the engineering education community is testament to the anticipated impact of this work. ***The primary remaining barrier to impact is context***—institutions and programs must be able to see how the findings from MIDFIELD apply to them. We have already begun to address disciplinary context by publishing studies that focus on demographics, trajectories, and outcomes in specific disciplines. ***The only way to overcome the barrier of institutional context is to add more institutions and more diverse institutions to the database.*** The 22 institutions willing to join will result in a database that has data from almost 10% of US institutions with engineering programs. The expansion of access to the MIDFIELD database represents a tremendous resource for other researchers.

## Impact on other disciplines

While the focus of the MIDFIELD project has always been engineering student pathways, the database is whole population, and includes non-engineering students for comparison. As a result, the database itself permits a much broader range of studies, and a much broader research community will be able to address that broader research agenda through the new access that archiving will provide.