The MURDOCK Integrated Data Repository (MIDR): An integrative platform for biomarker research

Swati Chakraborty1, Colette Blech2, Matthew Gardner3, Brian McCourt1, Meredith Nahm3, Jessica D. Tenenbaum3
1 Duke Clinical Research Institute, 2 Duke Center for Human Genetics, 3 Duke Translational Medicine Institute

Abstract
Background: The first phase of the MURDOCK Study (Measurement to Understand the Reclassification of Disease) established a framework that can help predict disease progression and response to treatment in 4 distinct diseases: cardiovascular disease (CVD), osteoarthritis (OA), obesity and hepatitis C (HCV). Following the release of MIDR 1.1, data from four studies have been successfully loaded into the MIDR. More incorporation of additional datasets in the future.

Methods: We designed and implemented a flexible and scalable architecture that facilitates export of arbitrary data elements from the different studies to be reused in a standardized format. The MIDR is a central data warehouse, with a cooperative approach to data sharing as well as an effort to ensure meaningful and available clinical data as well as to provide access to a comprehensive database.

Results
The MIDR includes data contributed to the MIDR from multiple disparate organizations at Duke, including 2 or more of the original studies into variables across studies as a subset. In order to integrate these disparate data types from multiple data sources, we developed the MURDOCK Integrated Data Repository (MIDR), a centralized data warehouse, with a cooperative approach to data sharing as well as an effort to ensure meaningful and available clinical data as well as to provide access to a comprehensive database.

Discussion
A paragon of team science, the MURDOCK Study has brought together interdisciplinary talent and expertise from multiple disparate organizations at Duke. Central to the success of this collaborative endeavor has been a cooperative approach to data sharing as well as efforts to ensure meaningful and available clinical data as well as to provide access to a comprehensive database.

Figure 1. Overlap among data elements. A. Three studies were integrated into the MIDR: CVD, OA and HCV. B. Among the 3 studies, common data elements are identified as pooled data elements between any number of individual studies.

Figure 2. The MURDOCK Integrated Data Repository. Three pooled data queries will be available through this interface, and comprehensive datasets will be available through customized queries after RR approval.

Figure 3. Data queries. Three pooled data queries will be available through this interface, and comprehensive datasets will be available through customized queries after RR approval.

Table 1. Individual Horizon 1 study information. The first phase of the MURDOCK Study (Measurement to Understand the Reclassification of Disease … datasets as well as data from additional study “horizons,” and will enable user-configured graphical queries.

Table 2. Mapping of metadata. Clinical information was collected independently for the 4 original H1 studies respectively, in 4 different study environments, using diverse tool sets for data collection. Across each of the studies, data elements were mapped across standardized domains. This canonical data model was designed to enable compatibility with multiple public standards, such as those of CDISC BRIDG, CDASH, and SDTM.

Table 3. Summary of data elements. A paragon of team science, the MURDOCK Study has brought together interdisciplinary talent and expertise from multiple disparate organizations at Duke. Central to the success of this collaborative endeavor has been a cooperative approach to data sharing as well as an effort to ensure meaningful and available clinical data as well as to provide access to a comprehensive database.

Table 4. Summary of data elements. A paragon of team science, the MURDOCK Study has brought together interdisciplinary talent and expertise from multiple disparate organizations at Duke. Central to the success of this collaborative endeavor has been a cooperative approach to data sharing as well as an effort to ensure meaningful and available clinical data as well as to provide access to a comprehensive database.

Table 5. Summary of data elements. A paragon of team science, the MURDOCK Study has brought together interdisciplinary talent and expertise from multiple disparate organizations at Duke. Central to the success of this collaborative endeavor has been a cooperative approach to data sharing as well as an effort to ensure meaningful and available clinical data as well as to provide access to a comprehensive database.

Table 6. Summary of data elements. A paragon of team science, the MURDOCK Study has brought together interdisciplinary talent and expertise from multiple disparate organizations at Duke. Central to the success of this collaborative endeavor has been a cooperative approach to data sharing as well as an effort to ensure meaningful and available clinical data as well as to provide access to a comprehensive database.

Table 7. Summary of data elements. A paragon of team science, the MURDOCK Study has brought together interdisciplinary talent and expertise from multiple disparate organizations at Duke. Central to the success of this collaborative endeavor has been a cooperative approach to data sharing as well as an effort to ensure meaningful and available clinical data as well as to provide access to a comprehensive database.

Table 8. Summary of data elements. A paragon of team science, the MURDOCK Study has brought together interdisciplinary talent and expertise from multiple disparate organizations at Duke. Central to the success of this collaborative endeavor has been a cooperative approach to data sharing as well as an effort to ensure meaningful and available clinical data as well as to provide access to a comprehensive database.

Table 9. Summary of data elements. A paragon of team science, the MURDOCK Study has brought together interdisciplinary talent and expertise from multiple disparate organizations at Duke. Central to the success of this collaborative endeavor has been a cooperative approach to data sharing as well as an effort to ensure meaningful and available clinical data as well as to provide access to a comprehensive database.

Table 10. Summary of data elements. A paragon of team science, the MURDOCK Study has brought together interdisciplinary talent and expertise from multiple disparate organizations at Duke. Central to the success of this collaborative endeavor has been a cooperative approach to data sharing as well as an effort to ensure meaningful and available clinical data as well as to provide access to a comprehensive database.

Table 11. Summary of data elements. A paragon of team science, the MURDOCK Study has brought together interdisciplinary talent and expertise from multiple disparate organizations at Duke. Central to the success of this collaborative endeavor has been a cooperative approach to data sharing as well as an effort to ensure meaningful and available clinical data as well as to provide access to a comprehensive database.

Table 12. Summary of data elements. A paragon of team science, the MURDOCK Study has brought together interdisciplinary talent and expertise from multiple disparate organizations at Duke. Central to the success of this collaborative endeavor has been a cooperative approach to data sharing as well as an effort to ensure meaningful and available clinical data as well as to provide access to a comprehensive database.

Table 13. Summary of data elements. A paragon of team science, the MURDOCK Study has brought together interdisciplinary talent and expertise from multiple disparate organizations at Duke. Central to the success of this collaborative endeavor has been a cooperative approach to data sharing as well as an effort to ensure meaningful and available clinical data as well as to provide access to a comprehensive database.