

Enabling a Learning Health System through a Unified Enterprise Data Warehouse: The Experience of the Northwestern University Clinical and Translational Sciences (NUCATS) Institute

JUSTIN B. STARREN, M.D., PH.D.^{1,2}, ANDREW Q. WINTER, B.A.¹, AND DONALD M. LLOYD-JONES, M.D., SC.M.^{1,2}

The Northwestern University Clinical and Translational Sciences Institute (NUCATS) was launched in 2007 to create a central hub supporting clinical and translational science (CTS) across numerous schools at Northwestern University, our three main clinical partners (Northwestern Memorial Healthcare Corporation, Ann and Robert H. Lurie Children's Hospital, and The Rehabilitation Institute of Chicago), community and industry stakeholders, and beyond. NUCATS is designed to support the entire spectrum of CTS, from basic discovery through clinical trials to community-based research, dissemination and implementation, regardless of disease area. Funded by a Clinical and Translational Award (CTSA) from the NIH, and supported by extensive institutional resources, NUCATS has become an innovation leader in several key areas of CTS. Here we highlight the critical contributions an enterprise data warehouse can make to advancing translational science and quality clinical care to enable a learning healthcare system. Central to the success of our EDW has been the governance model and data structure that have enabled rapid advances in CTS, several examples of which we provide.

NUCATS' Center for Data Science and Informatics (CDSI)

Informatics platforms that enable CTS have been a major focus of activity in NUCATS. The Northwestern University Biomedical Informatics Center was created to bring together informatics activities across NUCATS' partners. In 2015, the role of the center was expanded to explicitly include big data/data science, and the name of the center was changed to the Center for Data Science and Informatics (CDSI). CDSI brings together biomedical informatics researchers and clinical informatics leaders from NUCATS' partners into an organization to coordinate biomedical informatics across the NU academic medical enterprise. To meet this mission, CDSI has culled the necessary expertise and resources to enable and facilitate the application of informatics solutions to clinical and translational research. CDSI-coordinated infrastructure is a crucial component of translational research at Northwestern.

About the Northwestern Medicine Enterprise Data Warehouse

A central component of the informatics infrastructure of CDSI is the Northwestern Medicine Enterprise Data Warehouse

(NMEDW). Created in 2007 as part of the original NUCATS formulation, the NMEDW serves as the primary vehicle for data integration and transfer for both research and clinical operations. The NMEDW was created with an initial \$4.6 million, 3-year investment shared among the Feinberg School of Medicine (FSM), the Northwestern Medical Faculty Foundation and Northwestern Memorial Hospital; that investment has grown to \$18 million over 8 years. The latter two members have since merged, creating NMHC. From the beginning, the NMEDW was designed to serve both research and clinical needs from a single, unified warehouse. This dual-use model is one of the major strengths of the NMEDW, and one that enables it to function as a unique bridge, integrating healthcare and research, as well as ensuring support from both the research and clinical partners. The NMEDW currently stores over 67 billion observations on 2.9 million unique patients. Each night it loads 44 million new data elements from 76 separate sources including electronic health records (EHR), pathology data from the hospital and research laboratories, biomarker data from research databases, and research transactional data from our eIRB and other institutional systems. It then transforms source data into integrated versions providing access to biological data along with patient demographics and clinical observations, outcomes, and clinical trials protocols. The NMEDW uses the clinical-grade network, computational, and security infrastructure of NMHC to guarantee data security, and is security-audited every year. Use of the NMEDW continues to grow rapidly, showing 261% growth since 2011 (see *Figure 1*). Bringing together research and clinical data has been critical to the success of phenotyping in the Electronic Medical Records and Genomics (eMERGE) project.¹

Developing a sustainable governance model has been essential to the continued success of the NMEDW. Some of the main governance principles are discussed below:

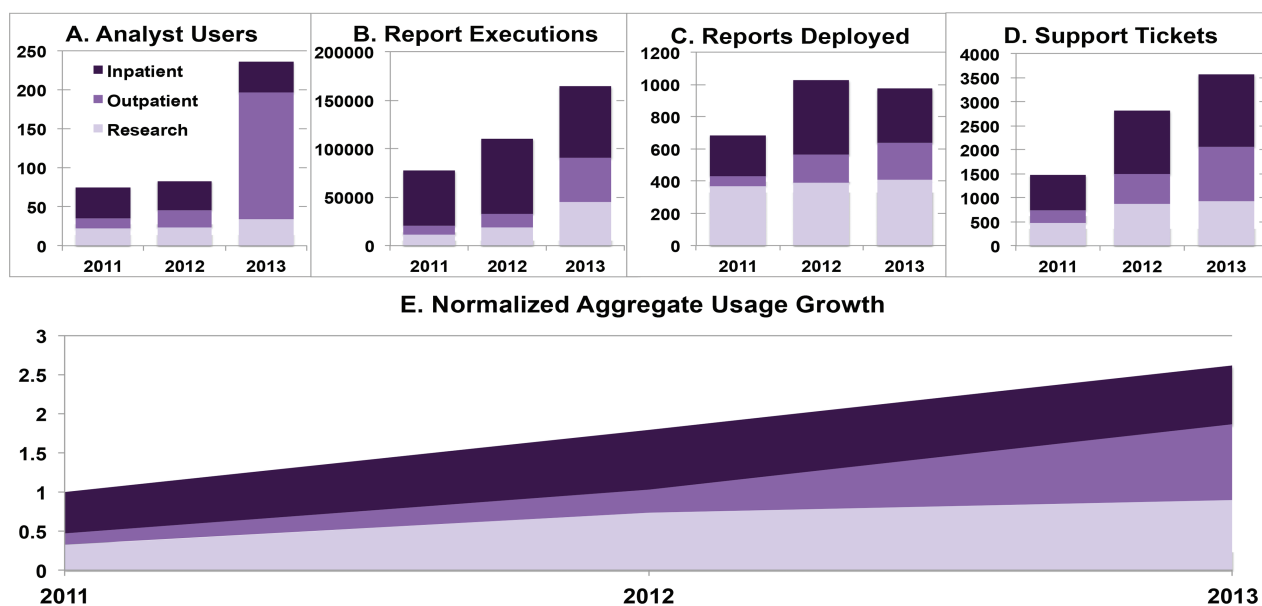
- **Single EDW Instance:** By combining research and care in a single EDW instance, there are economies of scale. In addition, any data structuring done to support care, such as Meaningful Use, becomes immediately available for research. In turn, research results can rapidly be translated into changes in care.
- **No Data Ownership by EDW:** Ownership of data within the warehouse remains with the source institution. Each institution is responsible, through the data steward process, for approving all data releases. This was critical for building

¹Northwestern University Clinical and Translational Sciences Institute, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA; ²Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA.

Correspondence: Donald M. Lloyd-Jones (dlj@northwestern.edu)

DOI: 10.1111/cts.12294

Figure 1. NMEDW Usage Growth. The NMEDW measures usage by a variety of parameters including: (A) the number of analysts developing queries; (B) the number of report executions; (C) the number of discrete reports deployed; (D) and the number of requests to NMEDW staff from users. These are combined into an aggregate usage metric (Panel E), which has shown greater than a 2.5-fold increase from 2011 to 2013.



interinstitutional trust. It also addresses the “minimum necessary” requirement of HIPAA.

- **Location within NUCATS:** The CTSA represents a truly multi-institutional structure. Thus, it was the obvious location for the EDW when each member institution is concerned about control of its own data.
- **Shared funding model:** The NMEDW is jointly funded by all member institutions. As a result, each institution has “skin in the game.”

The NMEDW is governed by a nine member steering committee representing both clinical and research users, and chaired by the NUCATS informatics lead. Annual major project prioritization is conducted using a Delphi process by the steering committee. Each member organization submits a number of proposed projects, NMEDW staff then estimates approximate effort required for each, and the steering committee members create a composite priority ranking. In addition to the steering committee, an NMEDW Advisory Committee, consisting of the CIOs of the member organizations and the NUCATS informatics lead, meets biweekly.

Simultaneously Supporting Research and Quality Improvement

The NMEDW is central to clinical operations, supporting, for example, Meaningful Use, outcomes, quality, compliance and revenue cycle reporting. In fact, the NMEDW was the first warehouse to achieve certification for both Meaningful Use Stage 1 and Stage 2 reporting. The NMEDW also serves the role of an “Honest Broker”² for clinical information to the research community. Access to NMEDW data requires approval from the appropriate oversight body. For quality improvement initiatives, projects must be approved by the respective institutions’ quality improvement committee. For research projects, IRB approval

is required. The NMEDW also supports an i2b2 instance for feasibility counts without IRB approval. To facilitate the rapid review and approval of data requests, the NMEDW has developed an efficient data steward model, supported by electronic workflow. Each member institution has a designated data steward who reviews all data requests. For access to research data sets, the project Principle Investigator serves as data steward. To support the research needs of individual investigators, the NMEDW staff creates “sandbox” data marts. These data marts contain IRB-approved or quality-committee-approved data elements so researchers are empowered with the data they need without the burden or risk of having data that are not required for a particular project. Since 2009, the NMEDW has supported 858 research projects and has proved to be an enabler of numerous clinical quality improvement initiatives by all contributing clinical affiliates.

Vignettes of NMEDW Projects Advancing research and care for HFpEF

One example of the value of an integrated research/care EDW has been the experience of Dr. Sanjiv Shah. Heart failure with preserved ejection fraction (HFpEF) remains one of the most common, and most vexing, problems faced by clinicians today. Dr. Shah joined NU in 2007 and initiated the NHLBI-funded TOPCAT study (spironolactone vs. placebo for patients with HFpEF) as site principal investigator in 2008. From the outset, he worked with NUCATS on his recruitment strategy. We designed a customized daily NMEDW-generated report of patients hospitalized at Northwestern Memorial Hospital who fit prespecified criteria based on free text, laboratory, imaging, and medication data. Despite starting enrollment 2 years after the trial began, Northwestern was the top US enrolling site (among 233 total sites in 6 countries), with 77 participants (2.2% of total).³

The same approach is now being used in the NEAT-HFpEF study, funded by the NHLBI Heart Failure Clinical Research Network. Again, Northwestern is the top enrolling site to date.

As important as it was to support Dr. Shah's research, the NMEDW-enabled research is also changing care at NMHC. Dr. Shah's clinical HFpEF Program was the first of its kind and uses the same NMEDW-based strategy to identify patients who would benefit from follow up in the HFpEF Clinic. Using the NMEDW, he recently published the first study to conduct high-density phenotypic classification ("pheno-mapping") of HFpEF, defining three discrete, clinically relevant phenogroups of HFpEF patients with significant differences in etiology/pathophysiology and risk for adverse outcomes. Compared with patients in Group 1, those in Groups 2 and 3 are at 3.0- and 4.4-fold higher risk, respectively, for hospitalization or death.⁴ This discovery is driving research into changing the therapeutic approach to HFpEF.

Improving care for sepsis

Dr. Emilie Powell wanted to evaluate quality metrics for sepsis care in the emergency department. Sepsis patients are not well identified by diagnostic codes alone, making identification of patients difficult and labor-intensive to process manually. Working with the NMEDW team, Dr. Powell developed an algorithm combining diagnosis codes with clinical parameters to identify 376 severe and/or septic shock patients from 2006 to 2008. This information was used to develop typical patient cases that were the basis for an *in situ* simulation of sepsis treatment in the Northwestern Memorial Hospital ED, and educational simulations for emergency medicine residents-in-training. The NMEDW sepsis analysis was extended into ongoing projects for Master in Public Health students, medical students and research fellows and has resulted in two publications to date.^{5,6}

Enabling real-time monitoring of transplant outcomes

Managing transplant quality is challenging, as CMS transplant scorecards lag current data by up to two years. To help solve this problem, Dr. Bing Ho, transplant nephrologist at NMHC, enlisted the NMEDW to develop a dashboard to analyze kidney transplant data at NMHC. The Real-time Analytics & Process Improvement

Dashboard (RAPID) was released in January, 2013. RAPID pulls information from the NMEDW relevant to patient outcomes and quality metrics, replacing a slow manual process. Dr. Ho and his team used RAPID for a root-cause analysis that led to an improvement in patient and allograft survival. The American Society of Transplant Surgeons now encourages the use of RAPID by all member transplant centers. As of mid-2014, 53 transplant centers had registered and downloaded RAPID.

Conclusions

Enterprise Data Warehouses that simultaneously support both research *and* clinical care as joint primary missions, as opposed to warehouses tailored primarily for research *or* care, are not only viable, but provide significant advantages. These advantages include economies of scale, as well as rapid translation of research discoveries back into clinical practice. However, our successful experience to date suggests that attention to governance models and data stewardship models are critical to ensure stability and efficiency.

CTS

REFERENCES

1. Kho AN, Pacheco JA, Peissig PL, Rasmussen L, Newton KM, Weston N, Crane PK, Pathak J, Chute CG, Bielinski SJ, et al. Electronic medical records for genetic research: results of the eMERGE consortium. *Sci Transl Med*. 2011; 3(79): 79re71.
2. Dhir R, Patel AA, Winters S, Bisceglia M, Swanson D, Aamodt R, Becich MJ. A multidisciplinary approach to honest broker services for tissue banks and clinical data: a pragmatic and practical model. *Cancer*. 2008; 113(7): 1705–1715.
3. Pitt B, Pfeffer MA, Assmann SF, Boineau R, Anand IS, Claggett B, Clausell N, Desai AS, Diaz R, Fleg JL, et al. Spironolactone for heart failure with preserved ejection fraction. *New Engl J Med*. 2014; 370(15): 1383–1392.
4. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghide M, Bonow RO, Huang CC, Deo RC. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation*. 2015; 131(3): 269–279.
5. Powell ES, Sauser K, Cheema N, Pirotte MJ, Quattromani E, Avula U, Khare RK, Courtney DM. Severe sepsis in do-not-resuscitate patients: intervention and mortality rates. *J Emerg Med*. 2013; 44(4): 742–749.
6. Venkatesh AK, Avula U, Bartimus H, Reif J, Schmidt MJ, Powell ES. Time to antibiotics for septic shock: evaluating a proposed performance measure. *Am J Emerg Med*. 2013; 31(4): 680–683.