10-25-2019

An Integrated Oncology Data Warehouse for Clinical Decision Support and Complex Patient Cohort Identification in a Hybrid Cancer Center

Michelle Keller
Miami Cancer Institute, MichelleK1@baptisthealth.net

Peter McGranaghan
Baptist Health South Florida, PMcGranaghan@baptisthealth.net

Bill Morgan
Baptist Health South Florida, BillMo@baptisthealth.net

Julie Gorman
Miami Cancer Institute, JulieLG@baptisthealth.net

Annette Balboa
Miami Cancer Institute, AnnetteBa@baptisthealth.net

See next page for additional authors

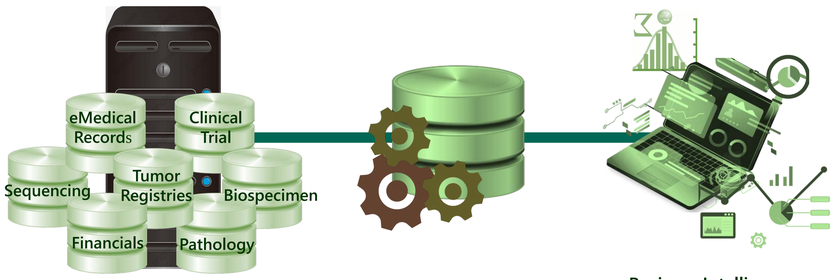
Authors
Michelle Keller, Peter McGranaghan, Bill Morgan, Julie Gorman, Annette Balboa, Muni Rubens, and Paul Lindeman

An Integrated Oncology Data Warehouse for Clinical Decision Support and Complex Patient Cohort Identification in a Hybrid Cancer Center

Michelle Keller MS, Peter McGranaghan PhDc, Bill Morgan BS, Julie Gorman BS, Annette Balboa MPH, Muni Rubens PhD, Paul Lindeman MD
Miami Cancer Institute, Miami, Florida

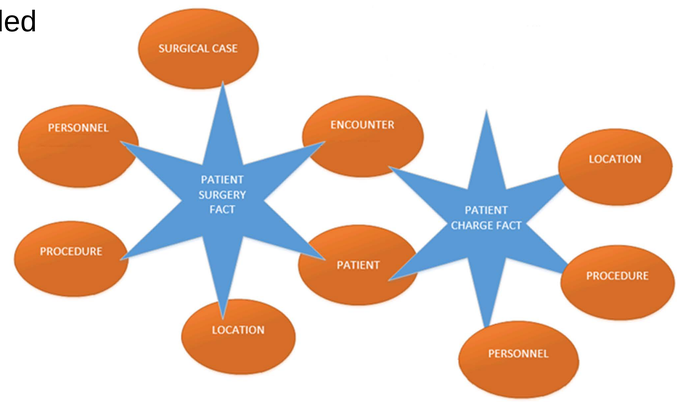
INTRODUCTION

- A data warehouse is a repository that centralizes and integrates data from disparate systems to provide the ability to easily access historical, consistent data
- Integrating disparate source systems into one centralized location, known as the Oncology Data Warehouse (ODW), enables our colleagues at Miami Cancer Institute (MCI) to rapidly identify more robust research cohorts and enables data-driven decision making



METHODS

- We integrated structured and unstructured data from disparate sources into one centralized data model by developing an automatic extraction, transformation and loading (ETL) process from all source databases into a new set of tables
- The ODW is modeled as a star schema, with fact and conformed dimension tables, which expands to a galaxy schema that can snowflake to other data models as needed



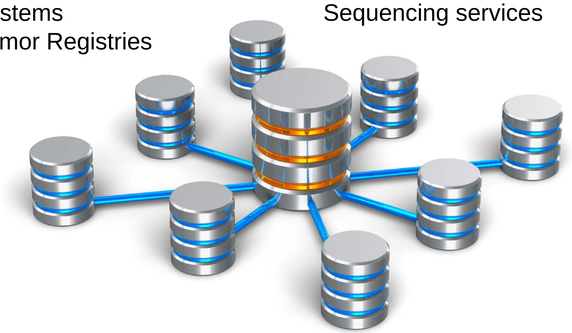
An integrated data warehouse which provides the ability to quickly access historical and current data for clinical decision support and complex patient cohort identification in a hybrid cancer center

Michelle Keller MS, Peter McGranaghan PhDc, Bill Morgan BS, Julie Gorman BS, Annette Balboa MPH, Muni Rubens PhD, Paul Lindeman MD
Miami Cancer Institute, Miami, Florida

- The ETL code performs incremental loads at regularly prescribed intervals into two parallel storage areas, a relational database management system (RDMS) as well as a Big Data file storage system.

RESULTS

- An interdisciplinary team of physicians, engineers, scientists, and subject matter experts at the MCI, designed, developed, and implemented the ODW with information originating from different data sources, including:
  - Electronic Medical Record Systems
  - Financial Systems
  - Clinical Trial Management Systems
  - Tumor Registries
  - Biospecimen Repositories
  - Pathology synoptic reports and archives
  - Next Generation Sequencing services



- Structurally it is a subject-oriented, integrated collection of data leveraging conformed dimensions
- The ODW is capable of connecting most business intelligence (i.e. Tableau) or statistical (i.e. SAS) tools for automated or static report development.

CONCLUSION

- We organized enterprise information about oncology patients which can be utilized for clinical decision support and precision medicine use cases by implementing an innovative combination of technology tools and methods
- The dynamic ODW enables physicians, clinical management teams, and medical analysts to systematically mine and review the molecular, genomic, and associated clinical or administrative information of patients, and identify patterns that may influence treatment decisions and potential outcomes.