JOHNS HOPKINS
MEDICINE

Project 8: UI for Radiation Therapy Cohort Selection

Seminar Presentation

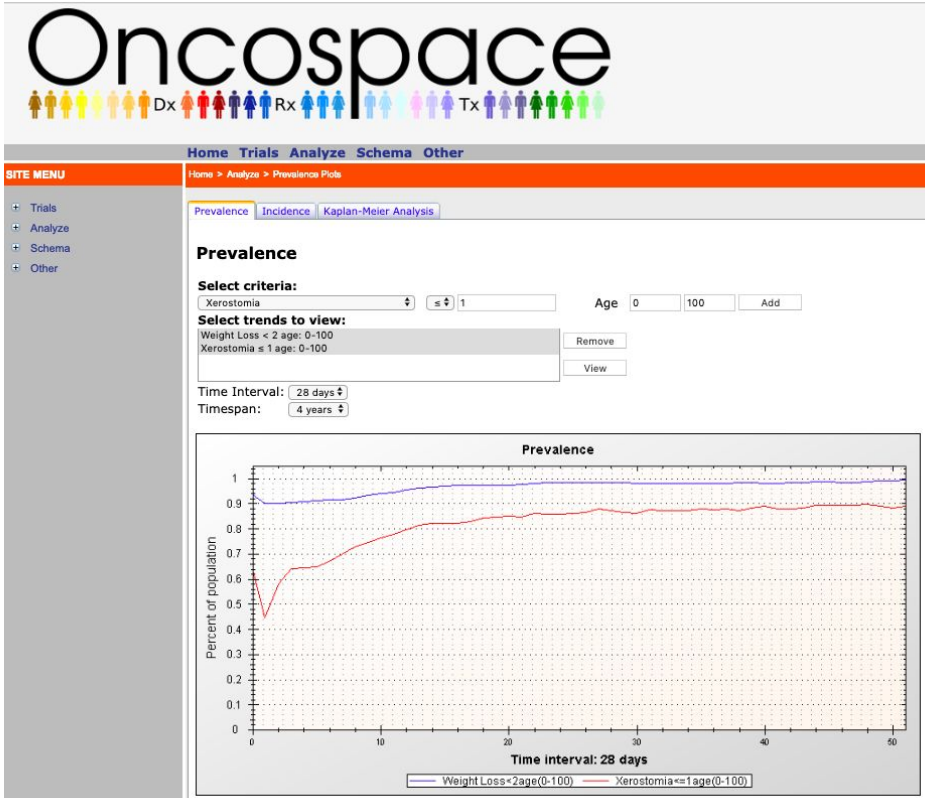
Members: Domonique Carbajal, Keefer Chern
Mentors: Todd McNutt,Pranav Lakshminarayanan

Oncospace™
Dx Rx Tx

Project Objective

Develop a User Interface that will allow user the ability to:

- Select a patient cohort based upon any number of variables (SQL)
- Perform statistical analysis on the extracted data
- Display the data in an easily comprehensible way (C# and JavaScript)
- Load and save parameters in a database query call



Prevalence

Select criteria:

Xerostomia ≤ 1 Age 0 100

Select trends to view:

Weight Loss < 2 age: 0-100
Xerostomia ≤ 1 age: 0-100

Time Interval: 28 days

Timespan: 4 years

Paper Selection

The Big Data Effort in Radiation Oncology: Data Mining or Data Farming?

Mayo, Charles S., et al. “The Big Data Effort in Radiation Oncology: Data Mining or Data Farming?” *Advances in Radiation Oncology*, vol. 1, no. 4, 2016, pp. 260–271., doi:10.1016/j.adro.2016.10.001.

Addresses conceptualization of handling data specifically for radiation oncology and poses warnings for handling errors in the database

Complete Author List:

Charles S. Mayo PhD [a,*], Marc L. Kessler PhD [a],
Avraham Eisbruch MD [a], Grant Weyburne BS [a], Mary Feng MD [b],
James A. Hayman MD [a], Shruti Jolly MD [a], Issam El Naqa PhD [a],
Jean M. Moran PhD [a], Martha M. Matuszak PhD [a],
Carlos J. Anderson PhD [a], Lynn P. Holevinski BS [a],
Daniel L. McShan PhD [a], Sue M. Merkel MSA RT(R)(T) [a],
Sherry L. Machnak MBA RT(T) [a], Theodore S. Lawrence MD PhD [a],
Randall K. Ten Haken PhD [a]

Purpose

Describe data related issues in ROIS, impart vision for solutions and key data elements that need to be addressed for fully utilizing available information

.

Introduce metaphor of “data farming” and necessary distinctions from data mining

Terminology Used and Defined

PQI - Practical Quality Improvement

ROIS - Radiation Oncology Information System

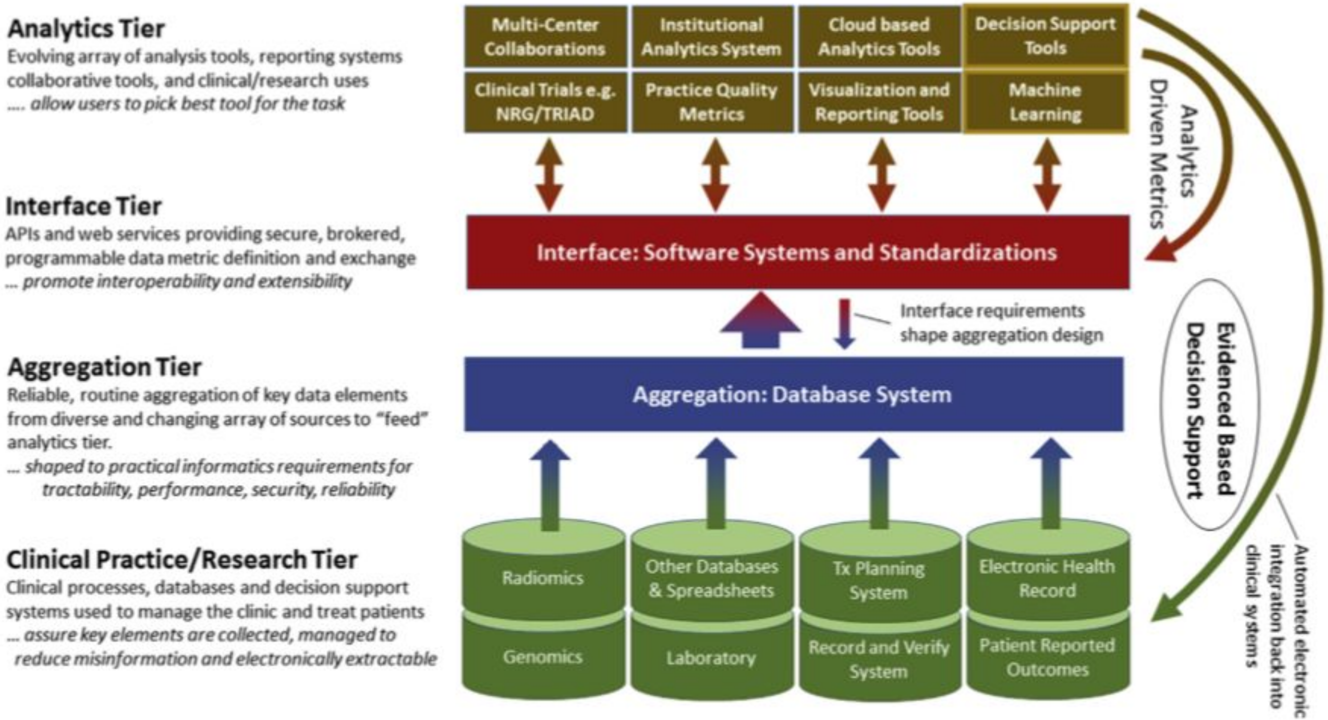
ETL - Extract, Transform, Load

Data Mining - (As defined by paper)

Data aggregation and analysis efforts, “mining” creates expectations data elements needed already exist in electronic system

Assumes data allows for accurate linkage to patients, identification of relationships among data elements, and extraction of reliable values

M-ROAR - University of Michigan instance of a Radiation Oncology Analytics Resource



Analytics Tier

Evolving array of analysis tools, reporting systems collaborative tools, and clinical/research uses

.... allow users to pick best tool for the task

Interface Tier

APIs and web services providing secure, brokered, programmable data metric definition and exchange

... promote interoperability and extensibility

Aggregation Tier

Reliable, routine aggregation of key data elements from diverse and changing array of sources to "feed" analytics tier.

... shaped to practical informatics requirements for tractability, performance, security, reliability

Clinical Practice/Research Tier

Clinical processes, databases and decision support systems used to manage the clinic and treat patients

... assure key elements are collected, managed to reduce misinformation and electronically extractable

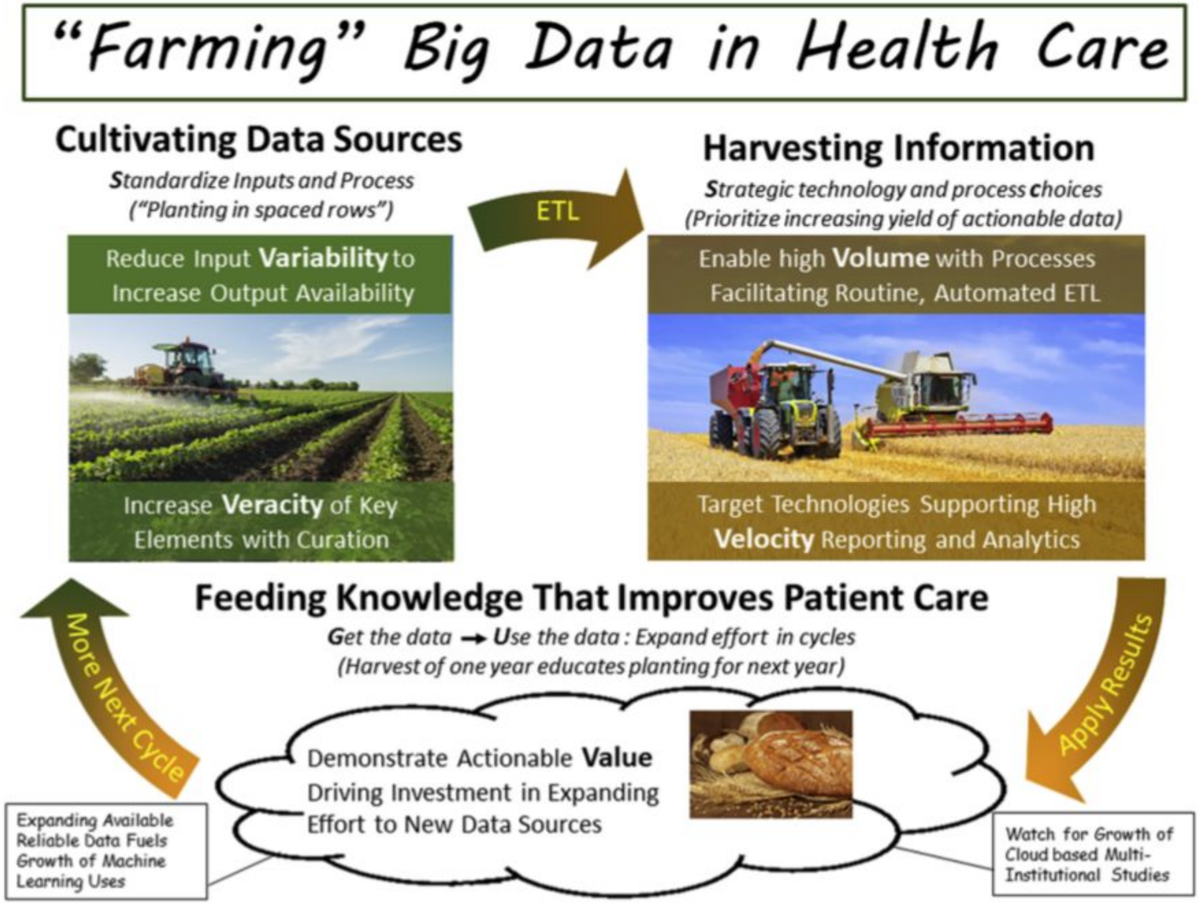
Multi-Center Collaborations
Institutional Analytics System
Cloud based Analytics Tools
Decision Support Tools
Clinical Trials e.g. NRG/TRIAD
Practice Quality Metrics
Visualization and Reporting Tools
Machine Learning
Analytics Driven Metrics
Interface: Software Systems and Standardizations
Interface requirements shape aggregation design
Aggregation: Database System
Evidenced Based Decision Support
Radiomics
Other Databases & Spreadsheets
Tx Planning System
Electronic Health Record
Genomics
Laboratory
Record and Verify System
Patient Reported Outcomes
Automated electronic integration back into clinical systems

Figure 1 The systems required for construction of a knowledge-guided radiation therapy system that supports machine learning, reporting, and participation in trials and other clinical efforts can be conceptualized in 4 tiers. The foundational clinical processes and aggregation tiers enable the benefits of the analytics tier. The integration tier promotes interoperability even when multiple technologies are used.

Mayo, Charles S., et al. "The Big Data Effort in Radiation Oncology: Data Mining or Data Farming?"

Data Farming: Highlighting 5 V's of Big Data in relation to ROIS

- **Variability**
  - Various data types need to be combined from multiple sources (different locations or practitioners) by criteria like time range
- **Veracity**
  - Incorrect and missing values cannot be avoided as PQI efforts focus on tails of distribution
- **Volume**
  - Storage and processing requirements drive decisions, image storage
- **Velocity**
  - Speed of system analytics and visualizations impact integration into clinical workflow
- **Value**
  - Obtaining support depends on cost-benefit to PQI and research efforts



"Farming" Big Data in Health Care

Cultivating Data Sources

Standardize Inputs and Process
("Planting in spaced rows")

Reduce Input **Variability** to Increase Output Availability

Increase **Veracity** of Key Elements with Curation

ETL

Harvesting Information

Strategic technology and process choices
(Prioritize increasing yield of actionable data)

Enable high **Volume** with Processes Facilitating Routine, Automated ETL

Target Technologies Supporting High **Velocity** Reporting and Analytics

Apply Results

Feeding Knowledge That Improves Patient Care

Get the data → Use the data : Expand effort in cycles
(Harvest of one year educates planting for next year)

Demonstrate Actionable **Value** Driving Investment in Expanding Effort to New Data Sources

Expanding Available Reliable Data Fuels Growth of Machine Learning Uses

Watch for Growth of Cloud based Multi-Institutional Studies

More Next Cycle

Figure 2 Farming is a useful metaphor for envisioning the issues in creating outcomes databases in health care.

Mayo, Charles S., et al. "The Big Data Effort in Radiation Oncology: Data Mining or Data Farming?"

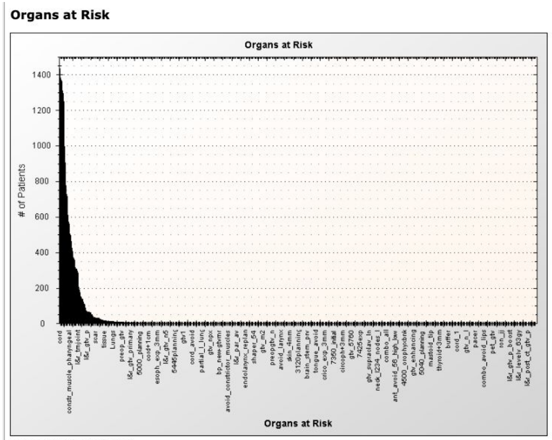
Availability of Key Data Elements

- Free text entry in EHR make for extremely variable data complicating staging and outcomes

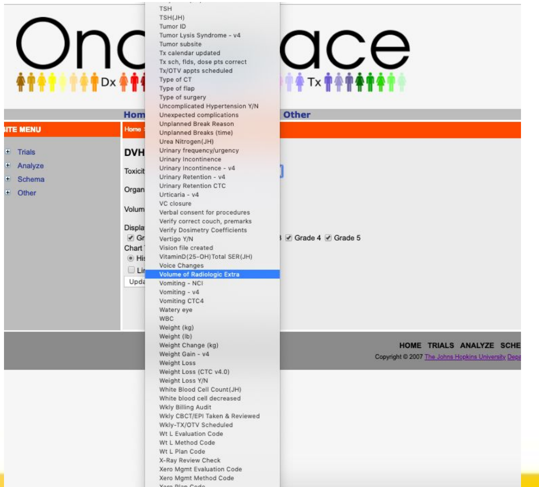
- Recurrence and toxicity information are often entered into the EHR as free text notes because it is the fastest means of proceeding with the demands of a busy clinical day

- NLP methods aren’t fully developed/can work even better in addition to altered practices

- Recommends: Practice changes to use standardized, quantified entry of key data elements; enables gathering the data now; and will enhance the accuracy and reduce costs of NLP methods when they evolve in the future.



Total: 33014 organs at risk



TPS-Treatment Planning System
ATPS- as treated plan sum

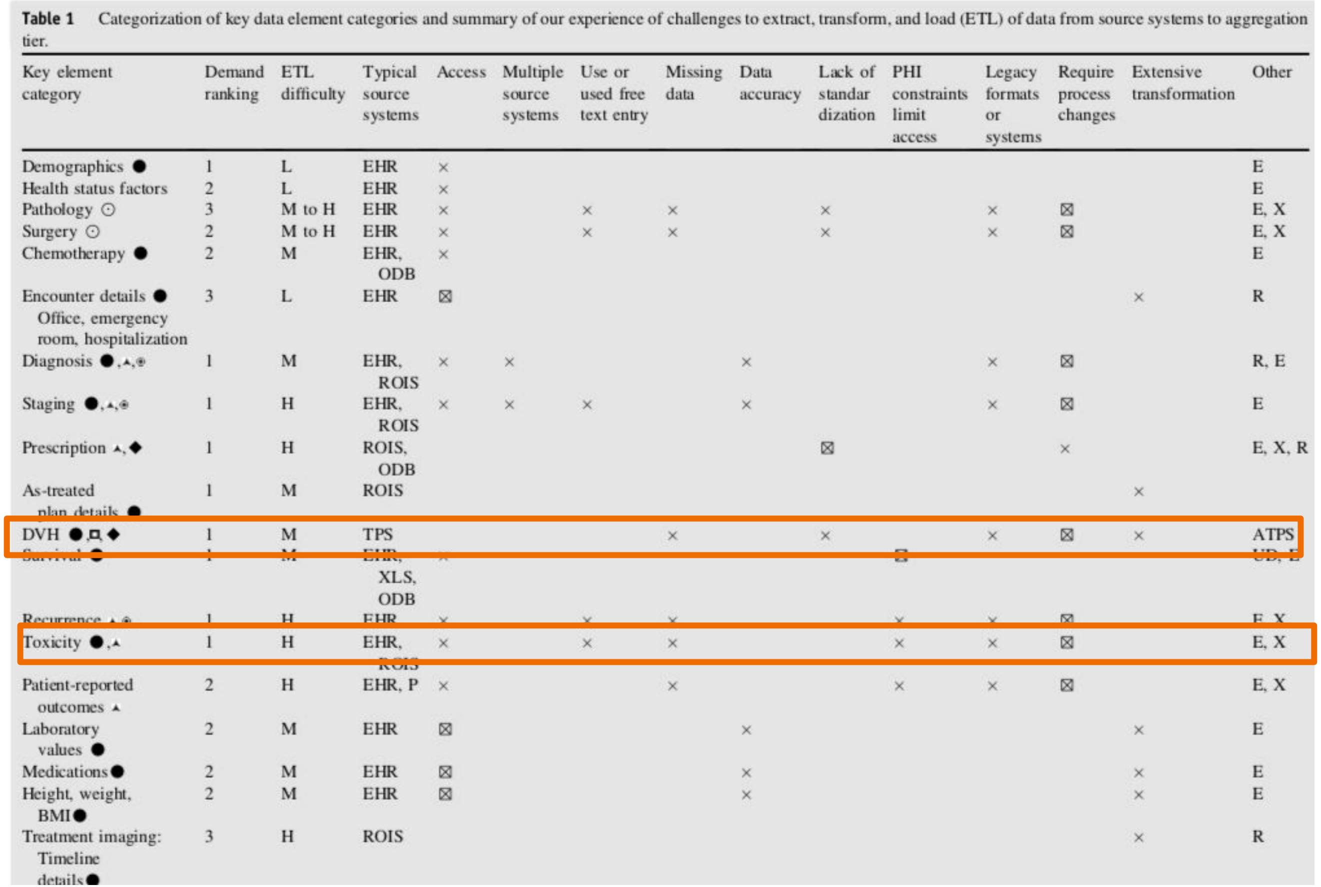


Table 1 Categorization of key data element categories and summary of our experience of challenges to extract, transform, and load (ETL) of data from source systems to aggregation tier.

Key element category	Demand ranking	ETL difficulty	Typical source systems	Access	Multiple source systems	Use or used free text entry	Missing data	Data accuracy	Lack of standardization	PHI constraints limit access	Legacy formats or systems	Require process changes	Extensive transformation	Other
Demographics ●	1	L	EHR	×										E
Health status factors	2	L	EHR	×										E
Pathology ⊙	3	M to H	EHR	×		×	×		×		×	☒		E, X
Surgery ⊙	2	M to H	EHR	×		×	×		×		×	☒		E, X
Chemotherapy ●	2	M	EHR, ODB	×										E
Encounter details ● Office, emergency room, hospitalization	3	L	EHR	☒									×	R
Diagnosis ●,▲,⊛	1	M	EHR, ROIS	×	×			×			×	☒		R, E
Staging ●,▲,⊛	1	H	EHR, ROIS	×	×	×		×			×	☒		E
Prescription ▲,◆	1	H	ROIS, ODB						☒			×		E, X, R
As-treated plan details ●	1	M	ROIS										×	
DVH ●,◘,◆	1	M	TPS				×		×		×	☒	×	ATPS
Survival ●	1	M	EHR, XLS, ODB	×						☒				UD, E
Recurrence ▲,⊛	1	H	EHR	×		×	×			×	×	☒		E, X
Toxicity ●,▲	1	H	EHR, ROIS	×		×	×			×	×	☒		E, X
Patient-reported outcomes ▲	2	H	EHR, P	×			×			×	×	☒		E, X
Laboratory values ●	2	M	EHR	☒				×					×	E
Medications ●	2	M	EHR	☒				×					×	E
Height, weight, BMI ●	2	M	EHR	☒				×					×	E
Treatment imaging: Timeline details ●	3	H	ROIS										×	R

E-susceptible to errors
X- manual effort required to extract

Mayo, Charles S., et al. "The Big Data Effort in Radiation Oncology: Data Mining or Data Farming?"

Building Data Curation in Practice Process

- The concept that inaccurate data values are acceptable because large volumes of data undermines the ability to carry out cohort discovery for rare combinations of factors.

- Errors or omissions for high-grade toxicities make it difficult to implement automated solutions to characterize distributions and correlate to contributing factors.

- “Assuring compliance with nomenclature standards for target and organ-at-risk structures and the existence of “as treated” plan sums dramatically increases the reliability of automated processing of DVH data” (267)

Approaching Technologies for radiation oncology big data

In considering the value of a new technology, it is important to look at:

- performance of query operations
- ability to integrate into existing systems to carry out ETL operations
- ability to integrate into development of clinical applications to use the data in practice
- ability to interact with standard analytics or machine learning systems
- implications for availability of staff required to implement the technology
- cost (hardware, software, training, staff, time)

Mayo, Charles S., et al. “The Big Data Effort in Radiation Oncology: Data Mining or Data Farming?”

Important for our project as we must also consider our implementations impact on the above criteria

In Review

Positive Points

- Used relevant information to radiation oncology without over specifying
- Took a multi-level look at the implementation of a system(funding to error handling)
- Gave specific examples of undermined cohort selection due to data error

Negative Points

- many recommendations oversimplified ease of establishing change to existing data collection in clinical setting
- adherence to the “farming” metaphor became overextended
- could have used more examples from their specific system (M-ROAR) implementation

Future steps for work:
More comprehensive look at attempt to alter data collection methods and influence on speed of queries and clinician response to changes

Citations

Mayo, Charles S., et al. "The Big Data Effort in Radiation Oncology: Data Mining or Data Farming?" *Advances in Radiation Oncology*, vol. 1, no. 4, 2016, pp. 260–271., doi:10.1016/j.adro.2016.10.001.