

**6:30~8:00 PM Panel**

# **Data Quality: Practices, Technologies and Implications**



**Louiqa Raschid**

*Professor,  
Smith School of Business,  
Dept. of Computer Science,  
UMIACS, Univ. of Maryland*



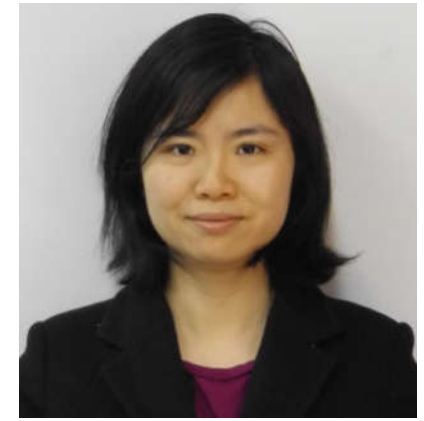
**Kathy Grise**

*Senior Program Director,  
Big Data Initiatives,  
Future Directions,  
IEEE*



**John Novak**

*Senior Director,  
Architecture and Gov.  
Tech, Healthcare Delivery  
Systems iLab, New Jersey  
Innovation Institute, NJIT*



**Yi Chen**

*Henry J. Leir Chair in  
Healthcare, Associate  
Professor, Martin  
Tuchman School of  
Management, NJIT*

# Big Data in Healthcare

- Data with exponential growth are observed in Healthcare
- Big data analytics hold big promises on
  - Early risk warning and disease prediction
  - Clinical decision support
  - Personalized treatment and precision medicine
  - Public health
  - Patient engagement

Improve outcomes



Lower down costs

# Data Quality is Critical

- Cleanness Data
  - Incomplete data
  - Duplicate data
  - Invalid data
  - Inaccurate data
- Usefulness of data
  - Trustworthiness
  - Timeliness
  - Traceability

# How Dirty Data Are Generated?

- Data Entry
- Data Processing  
e.g. Information extraction from unstructured data  
can generate errors
- Data Transport  
Data are generated from multiple sources and are  
presented in multiple formats. Data integration can  
generate errors

# Panel Structure

- What are data quality issues in healthcare?
  - led by Mr. John Novak
- What are the state-of-art technologies to address data quality challenges?
  - led by Dr. Louiqa Raschid
- What are community resources (e.g. standards) to address data quality problems?
  - led by Ms. Kathy Grise
- What are recommended action items?

# Gaps and Action Items

## - Novak

- What are the gaps between needs and practices?
- Consistent data
  - Sharable
  - Consistent
  - Normalized
- What are the recommended actions to take on data quality problems in healthcare?
  - Define minimum required information for patient matching
  - Publish and champion minimum required
  - Encourage the sharing of data

# Gaps and Action Items

## - Raschid

- First generation tools work well.
- Next generation needs to focus on processes and workflows and HumInt.
- Scientists still spend huge amounts of time on cleaning. How can we fix this problem?
- Is Open Knowledge Networks a solution?
- An unexpected case study ...



# Gaps and Action items

## - Grise

1) What are the gaps between needs and practices?

- Lack of standards and governance
- Inconsistent, unpredictable data entry
- Vendors may be concerned about proprietary and monetary value which inhibit ability to share

2) What are the recommended actions to take on data quality problems in healthcare?

- Define and establish consistent standards and governance applicable to overall data formats, data generation, data collection, data analysis
- Establish agreement, and if not agreement, at least conformance of compatible instruments, software applications – requires cooperation among manufactures and suppliers
- Provide consistent training of healthcare professionals for improved data entry and interpretation



# Gaps and Action items

## - Yi

- Techniques to assist user data entry are in high demand, especially in EHR
- Establishing incentives for providing high-quality data

# Questions?



New Jersey's Science & Technology University

*THE EDGE IN KNOWLEDGE*