

# *Symposium on Data Science for Healthcare (DaSH)*

*The Leir Retreat Center*

*Ridgefield, CT*

*Sponsor: Leir Charitable Foundations, Supporter: NJIT*

## **Summary**

The aims of the DaSH symposium were to: i) understand the grand challenges in developing data science technologies for using big data to improve outcomes and reduce costs in healthcare; ii) discuss research directions and priorities; iii) foster collaboration and community building between data science and healthcare.

A range of topics have been discussed by DaSH participants. Below is a summary of topics that are considered highly important to be addressed in the near future for using big data in healthcare.

- **System usability, real-life applications and human-in-the-loop:** While technologies can bring a lot of benefits to healthcare, it is important to make sure that technologies are designed to solve real problems, and the developed systems are user-friendly to end users. Technologies should help users rather than impose burden or create frustration. Also, given the diverse needs of users, it's important to personalize the system to match user needs.
- **Data integration and data interoperability:** Healthcare related data are scattered in diverse distributed heterogeneous data sources. It is important to connect different data sources in order to make analytics and to obtain insights.
- **Data quality and provenance:** Data are generated by different people, under different contexts, bearing different qualities. It is important to assess data quality and trustworthiness. Thus it is helpful to associate annotations on data to "explain" the data, and to understand where the data come from (the provenance).
- **Data availability and privacy.** There are concerns about user privacy protection and that users may not be willing to share data. However, younger generations may be willing to share their data, especially if they know others can benefit from it. Also, different people may have different sensitivity on different subsets of data. Being aware of individual concerns can achieve public utility and individual privacy at the same time.
- **Example of using big data for improving healthcare:** Effective models of electronic health record can provide risk modeling, diagnosis support, and documentation assistance. Social media data can be used to sense public health status and beliefs, study post-market pharmacovigilance, and report infectious disease status. Also, processing data generated by infectious disease propagation simulations can provide early warning and help to understand intervention. Sensor data and mobile technology can provide convenient and real-time monitoring, and make it possible for point-of-care technologies. Effective models developed for analyzing experimental and lab data, such as medical images and brain responses on blasts, have the potential to reduce the amount of clinical trials, which are expensive, invasive and sometimes infeasible.

The DaSH participants conclude that the objective of using big data of high *volume*, high *velocity*, high *variety* and high *veracity* is to create *value* in healthcare by delivering better healthcare outcomes to more people at lower costs. This requires an interdisciplinary approach with collaborations of data scientists, healthcare experts, social scientists, and policy makers from academia, industry, and government.