Stanford MEDICINE

Data Science Hub Overview

Manisha Desai, Michael Halaas, Daniel Rubin, Nigam Shah
August 12, 2016
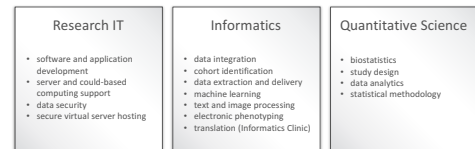

Data Science Hub (DaSH)

- Critical enabler to:
  - Establish our preeminence in the impactful and emerging field of biomedical data science
  - Build a renowned learning health system
  - Realize our vision for Precision Health

Data Science Hub (DaSH)

- Data Science Hub will enable/provide:
  - An environment to **securely access and use data** (security, privacy and compliance issues)
  - Access to **large-scale data** that will create new opportunities for innovative collaborations
  - The ability to **integrate** diverse data sources – (text, images and wearables data to claims and EMR data)
  - **Consultation and collaboration** in using and/or developing effective methods to acquire, represent, store and analyze biological and clinical data at all scales

---

Data Science Hub Vision

- Creation of a world-class collaboration and service unit integrating technology infrastructure, informatics, and quantitative science

Research IT	Informatics	Quantitative Science
• software and application development • server and could-based computing support • data security • secure virtual server hosting	• data integration • cohort identification • data extraction and delivery • machine learning • text and image processing • electronic phenotyping • translation (Informatics Clinic)	• biostatistics • study design • data analytics • statistical methodology

- Bring together EMR, imaging, multi-omics, population health, mHealth, biobanking, registry, etc. data to enable innovative research, and attract and retain top faculty and trainees
- Provide integrated support and consultation across multiple areas of data science and biomedical disciplines with **faculty-level expertise**

Levels of Engagement: Baseline

1. **Baseline Services**
- Includes direct integrated data science **consultation** and **pre-award assistance** with faculty-level expertise, professional staff support and access to technology infrastructure
- Available to all researchers at a fixed number of hours (~8 hours/year/project)
- **Consultation/pre-award support** includes:

Integrated data science consultation	Study design and cohort identification	Data vending	REDCap support
Data de-identification	Secure virtual server hosting for translational projects	Unstructured data utilization	Deployment and translation to the clinic ("Informatics Clinic")
Data linking and systems integration	Data access and analytics tools	Server and cloud-based computing support	Joint support for implementation research
Data security for projects with regulated data	Data analysis support	Guidance on statistical programming	Interpretation of findings
Machine learning	Text processing and image analysis	Digital phenotyping	Translation of statistical methods

---

Levels of Engagement: Collaboration

2. **Collaboration**
- **Post-award involvement** in team-science projects and development of **customized research programs** within departments/institutes/centers
- Available on a case by case basis for research-funded projects or through partnerships
- **Collaborative engagement** includes:

Development of custom extractors (e.g., extracting non-Stanford ED visits)	Development of machine learning based approaches for risk stratification (e.g., identifying eligibility for palliative care)
Development of new digital phenotyping approaches (e.g., identifying patients with Familial Hypercholesterolemia)	Collaboration on manuscripts and grant proposals
Development of custom image feature extractors or new digital image phenotyping methods (e.g., finding imaging markers of disease or those that correlate with clinical/molecular data)	Development and delivery of short courses
	Mentoring of junior faculty and fellows
	Full integration of data scientists into research teams

- Examples of entities with QSU partnerships: Medicine, Neurosurgery and SCI
- Examples of post-award projects that utilize the Data Science Hub infrastructure: CHOIR, MyHeart Counts, High Risk Infant Follow-up, Oncoshare, GenePool, etc.

Advisory Committee: Accountability and Governance

- Advisory Committee comprised of key SOM faculty chairs and stakeholders, health system representatives, main campus representatives, and Stanford Medicine leadership
  - Provide high level guidance, input and evaluation
  - Help define success metrics and hold leadership accountable
  - Meet quarterly

- Annual review with Stanford Medicine Leadership

Achieving and Sustaining Preeminence – Assessment Measures

- Metrics
  - Funding (from NIH and other government sources, Foundations, and/or Industry) attributable to interactions with Data Science Hub
  - Publications resulting from interactions with Data Science Hub
  - Faculty/customer satisfaction
  - Hit rate on relevant grant proposals
  - Success at integrating across data science components
  - Metrics will be published for transparency