Policy Issues in the Clinical Development and Use of Immunotherapy for Cancer Treatment

Session 4: Expanding opportunities for collaboration and information exchange
Moderator: Amy Abernethy, Flatiron Health

Trans-IT interoperability
Eric Perakslis, Takeda Pharmaceuticals
Disease Knowledge Networks: multi-level mechanistic hypotheses

- Syndrome
- Symptom/Traits
- Cellular Systems/Signaling Pathways
- Proteome/Metabolome
- Genome

Modified from http://www.phenomics.ucla.edu/phenomics/cognitive_phenomics.htm
Health Science Today... and Tomorrow

Medicine today is built on hundreds of years of individual observations...

What if we could integrate and model them instead?
What is trans-IT interoperability?

In healthcare, interoperability is the ability of different information technology systems and software applications to communicate, exchange data, and use the information that has been exchanged.\(^1\) Data exchange schema and standards should permit data to be shared across clinicians, lab, hospital, pharmacy, and patient regardless of the application or application vendor.\(^2\)

*Healthcare Information and Management Systems Society (HIMSS) 2013*
What should be interoperable?
There are Three Major Primary Sources of Patient Data

**Community Care Setting**
- Longitudinal data
- Small-to-large e-HR
- Probably most Comprehensive and least available
- Little IT/Informatics capabilities

**Academic/Government Medical Center**
- Disease/injury record
- Large e-HR (EPIC/Cerner)
- May have CTMS
- Likely has data warehouse or other research-centric structures
- More IT/Informatics but also a much more complex environment

**Registry Clinical Trials**
- Protocol-based
- Extremely narrow data slice
- Not longitudinal
- Most-highly regulated
- Cross-trial data may be sketchy
- Stronger IT/Informatics

*(sidebar: this is an access issue we should fix)*
What should be interoperable?
There are many additional sources of Translational Data

Text, Language & Standards
- NCIthesaurus
- CDISC SHARE
- HL7 STANDARDS
- ICD-10
- SNOMED CT

High-Content Data
- dbGaP GENOTYPES and PHENOTYPES
- The Cancer Genome Atlas
- GenomeConnect The Clinical Genomics Portal
- CANCER IMAGING ARCHIVE

Patient Powered
- patientslikeme
- MMRF Multiple Myeloma Research Foundation
- pcor
- Million Veteran Program: A Partnership with Veterans
- 8,000 Enrolled!
Linguistics of Immuno-informatics and cancer Informatics. How to make patient-centric?

Solving the interoperability challenge of a distributed complex patient guidance system: a data integrator based on HL7’s Virtual Medical Record standard.

Enabling semantic interoperability in multi-centric clinical trials on breast cancer.

...ad infinitum
Data Lakes

A repository for large quantities and varieties of data, both structured and unstructured.

Data generalists/programmers can tap the stream data for real-time analytics.

The lake can serve as a staging area for the data warehouse, the location of more carefully “treated” data for reporting and analysis in batch mode.

The data lake accepts input from various sources and can preserve both the original data fidelity and the lineage of data transformations. Data models emerge with usage over time rather than being imposed up front.

Data scientists use the lake for discovery and ideation.

Data lakes take advantage of commodity cluster computing techniques for massively scalable, low-cost storage of data files in any format.
In computing, a data warehouse (DW or DWH), also known as an enterprise data warehouse (EDW), is a system used for reporting and data analysis. DWs are central repositories of integrated data from one or more disparate sources. 
https://en.wikipedia.org/wiki/Data_warehouse
There are many In-flight efforts that should be considered. They are highly diverse and, sometimes, more competitive than cooperative.

**Consortia/Federal:**
- ASCO CancerLinQ
- tranSMART Foundation
- Sage Bionetworks
- eMerge Network
- Project Data Sphere
- CMS

**Commercial:**
- FLATIRON
- athenahealth
- Aetna
- OPTUM
- Moffitt Cancer Center
- M2Gen

**Community Care Setting**
**Academic/Government Medical Center**
**Registry Clinical Trials**
**Payors (physician behavior)**
1. Do we have consistent questions and use cases? (alignment & consistency)
2. Are there certain things we always need to know? (automation)
3. Who will be asking the questions? How often? (roles, security, partnerships)
4. What do we want the data to tell us? (hypothesis-free vs. hypothesis)
5. What should we be asking? (artificial intelligence)
Disease map
(p value <= 10^{-5})
Disease-drug map
(p value <= 10^{-5})
We can be aspirational and smart....

http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp#