Project Data Sphere

Convener, Collaborator, Catalyst

Bill Louv, PhD; President, Project Data Sphere
Project Data Sphere

What?
- Facilitate research and collaborations

How?
- Open source data platform

Why?
- Mission to accelerate cancer research
Project Data Sphere®: Background

● An initiative of Life Sciences Consortium of the CEORT
● Launched April 8, 2014
● Open-access data sharing
● Integrate data from diverse sources
● Powerful SAS analytic tools
PDS Landscape: Data to Research

Data Providers

- Govt.
- Pharma Industry Biotech
- Academia

Data Types

- Images
- Clinical Trials
- Molecular Data
- Registries
- Drug labels, Ontologies

Research Programs

- Images and Algorithms (RECIST)
- External Control Arm (SCLC) Prostate DREAM
- Breast Cancer (Alliance)
- Rare Tumor Registry (Merkel Cell)
- Immune Related Adverse Events
Platform Metrics: Utilization & Proof of Principle

- Industry and NCI’s NCTN Commitment
  - >100,000 patient lives of data from >28 data providers

- Densely used by researchers
  - >2,400 researchers have accessed >19,500 files

- Prostate Cancer DREAM Challenge
  - >500 participants; 58 teams; 21 countries
  - More participants than first 32 DREAM Challenges

- Proof of Principle
  - 26 publications in top-tier peer-reviewed journals
# Dataset Providers

## # of Dataset Providers by Year

Data Growth

*As of October 31, 2019*
164 Datasets by 28 Data providers

- **NCI**: 43
- **Alliance**: 22
- **Sanofi**: 19
- **Amgen**: 13
- **Janssen**: 13
- **AstraZeneca**: 11
- **Eli Lilly**: 7
- **Bayer**: 7
- **Pfizer**: 7
- **RTI**: 4
- **Celgene**: 4
- **Others**: 11

*As of October 31, 2019*
24 Tumor Types from >100,000 Patient Lives

*As of October 31, 2019*
Conclusions

- **Convener**
  - Bring together experts from academia, industry, and regulatory agencies
  - Organize taskforces, workshops, and symposia

- **Collaborator**
  - Conceive novel research questions
  - Execute research initiatives

- **Catalyst**
  - DATA to accelerate cancer research
  - Create enriched datasets
  - Convener + Collaborator $\approx$ Cascading catalytic effect on the outcomes

- **Challenges**
  - Remove barriers to sharing data
  - Enable interoperability of data sharing platforms