Roche Data Commons – pRED implementation

Kuentzer, Jan
Roche Pharma Research and Early Development
pRED Informatics, Data Science
Roche Innovation Center Munich
Personalized Healthcare

Focus areas for informatics

- Smarter, more efficient R&D
- Improved access & personalized patient care

1. Database & Analytics Interface
   - Potential biomarkers selected
   - New R&D hypothesis generated
   - Clinical trial data generated
   - Real world data generated
   - Treatment plan selected
   - Comprehensive diagnosis performed
   - Evidence-based treatment options outlined

2. Personalized Healthcare
   - Patient matched to clinical trial
   - New R&D hypothesis generated
   - Improved access & personalized patient care
Personalized Healthcare
High dimensional clinical data types

GENETICS DATA
- e.g. mutation and expression information

SENSOR DATA
- e.g. smartphone-based clinical biomarker data

REAL-WORLD DATA
- e.g. patient data from electronic health records (EHR)

IMAGE DATA
- e.g. IHC of tumor tissue

Database & Analytics Interface
Sensor Data

*Digital Biomarker Discovery in Parkinson Disease*

> 5000
ACTIVE TESTS

28,000
HRS OF CONTINUOUS MONITORING

1
EXPLORATORY TRIAL

Different Blue's: Data accumulated by different patients

Sensor data collected by patients

June 15
Time
Sep 2016

0 TB
0.8 TB
Sensor Data

Digital Biomarker Discovery in Parkinson Disease

Active Test

Patients are asked to complete six tests on the app at the same time each day, ideally in the morning.

Voice Recording  Balance Test  Gait Test  Dexterity Test  Rest Tremor Test  Posture Tremor Test

Data upload

In both the Active Test and Passive Monitoring test, smartphone data is captured & automatically uploaded...

Passive Monitoring

Patients carry their smartphone around as they go about their daily activities
Frequent measurement are key

Importance of Longitudinal Measurement:
Patient tremor severity varies from day to day (on/off)

Sensor Data

Deuschl et al (1998), Movement Disorders

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Frequency (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physiologic tremor</td>
<td></td>
</tr>
<tr>
<td>Dystonic tremor</td>
<td></td>
</tr>
<tr>
<td>Parkinsonian tremor</td>
<td></td>
</tr>
<tr>
<td>Cerebellar tremor</td>
<td></td>
</tr>
<tr>
<td>Holmes tremor</td>
<td></td>
</tr>
<tr>
<td>Palatal tremor</td>
<td></td>
</tr>
<tr>
<td>Neuropathic tremor syndrome</td>
<td></td>
</tr>
<tr>
<td>Drug-induced &amp; toxic tremors</td>
<td></td>
</tr>
<tr>
<td>Psychogenic tremor</td>
<td></td>
</tr>
</tbody>
</table>

Power on smartphone
(amount of energy at different frequencies)

Frequency (Hz)

No visible tremor

Plainly visible tremor
Image Data

Understanding the Immune Context of Tumor Treatment

- **10,000 SLIDES/YEAR**
- **100,000 OBJECTS/SLIDE**

- Overlay of imaging results on the original high resolution image
- Interactive summary statistics for user-selected areas
- High-throughput tissue imaging that allows quantitative analysis of tumor immune context
Genetics Data

Correlating cell drug resistance with underlying NGS data

= 600+

EXOME SEQUENCED CELL LINES

10% of cancer cell lines sensitive to treatment with drug X

A gene expression signature predicts cell-line sensitivity to drug treatment
Real-World Data

Identification of exclusion criteria in CIT trials

Real-World Data

Creation of matched cohorts
**UDIS – Understanding Disease Informatics System**

*pRED’s 1st generation MI Platform*

---

**Collaborations**
- New York Genome Center
- Broad Institute
- NIH (National Cancer Institute, National Human Genome Research Institute)

**Public**

**Specialized End-user Tool**

**UDIS**
*Understanding Disease Informatics System*

---

**Computer Scientists**

---

**Spotfire**
Roche Data Commons

Driver for new holistic strategy

Data Curation
Standard terminology at data acquisition phase

Analytics to Data
Support data federation
Distributed analysis

Data Integration
Integrate different data types (e.g. Imaging & Omics)

Omics-goes-Clinical
Data Provenance
Data and workflow mgmt

Externalization

Data Volume
Store only once
Minimize data transfer
Roche Data Commons
Five-layer model

Physical Hardware

File System & Workflow

Single Point of Truth

Integration/Data Mart

User Interface

Virtual File System

Reference Data

Workflow Y (v1.0)

Workflow Y (v1.2)

Workflow Y (v2.0)

Automated Workflows

Virtual File System

V1.0

V1.5

V2.0

Genome Information

Terminology

Pathways

Sample

Proteomics

Study

Transcriptomics

Genetics / Genomics

Image Analysis results

External compound & assay data

CRO Metainfo

Omics

Imaging

Oomics

Chemical Biology

Chemical Compound Integration

Aggregate Data

Workflow Webinterface

Spotfire

Cytoscape

R

Genome Browser

Gene-centric UI

Study-centric UI

Compound centric UI

Pharmacology centric UI
Roche Data Commons

Service oriented modular approach
File & Workflow Management

Layer 2

Generic File Management & Workflow System

- Allows data federation...
- Data de-duplication...
- Enables data provenance...
- Supports high data volume...
- Leverage internal compute infrastructure (HPC)
- Additional Cloud option
- Provides Expert User platform together w/ layer 1
File & Workflow Management

Arvados platform

Layer 2 Realized by Arvados Open Source Technology

- Open source infrastructure for managing, processing, and sharing genomic and other biomedical data
- Enables data federation, data & workflow provenance, managing and processing petabytes of data
Single Point of Truth (SPoT)

Layer 3

**SPoT**
*(Andy Hunt, Dave Thomas)*

(i) Every data element stored only once;
(ii) Any possible linkages to this data element by reference only;
(iii) All other locations of the data just refer back to the primary "source of truth" location. Consequently updates to the data element in the primary location propagate through the entire system.

**Single Point of Truth**

- Single, unambiguous data representation...
- Structured storage by data type...
- Enable integration of different data types...
- Support data integration across organizations...
- Continuous extension of new data types...
- Most important concept for 5 layer model...
Terminology Reference SPoT

Reference SPoT Example

pREDi Terminology Service (RTS)

- 18 Reference Terminologies
- Almost 110’000 concept, 300’000 synonyms
- More than 30 Applications connected
- More than 300 Application Terminologies
- Serving pRED, PD, PT, GPS, Partnering and DIA
- Supporting Data Curation, Data Capture, Text Mining, Semantic Search & TouchPoint
Roche Data Commons

Support pharma research area

- Distributed
- Scalable
- Modular
- Integration
- Flexible
- Agile

- Physical Hardware
- File System & Workflow
- User Interface
- Integration/Data Mart
- Single Point of Truth

Single Point of Truth (Integration/Data Mart)
Cancer Immunotherapy
System Architecture and processing workflow

UDIS
Understanding Disease Informatics System

LIMS
Sample Metadata

IRIS
Image Metadata

Sub-sampling & barcode

Metadata transferred for search

Image Analysis

Standard Image Analysis Workflow

High-quality sample metadata

salesforce

CROs

Slides with barcode

Integration/D Data Mart

File System & Workflow

Single Point of Truth

Integration/D Data Mart

User Interface

Physical Hardware
IRIS Architecture

First realization of Roche Data Commons

Physical Hardware

File System & Workflow

Single Point of Truth

Integration/Data Mart

User Interface

Images in Virtual File System

Reference Data

Tissue Imaging

Workflow Webinterface

Kaiseraugst

Penzberg

Indianapolis

Images in Virtual File System

Image Analysis Workflows

Image Analysis Workflows

Clinical Sample

Genetics /Genomics

Chemical Compound Integration

External compound & assay data

Transcriptomics

Image Analysis results

Omics

Cross-study prevalence tool

Study centric view

Clinical Sample

Patient

Study

Biomarker

USER “PLAYGROUND” – YOU CUSTOMIZE

Spotfire

IRIS Viewer

Study centric view

Cross-study prevalence tool
I would like to thank **ALL pRED colleagues who have contributed to the work summarized in these slides!**

**Special Thanks to:**
Thomas Thies, Juergen Hammer, Angelika Fuchs, Marc Litherland, Moritz Gilsdorf, Christian Gossens, Francesca Milletti, Florian Lipsmeier, Fabian Schmich, Fabian Gaire ...
Doing now what patients need next