tranSMART

CTSA Informatics Key Function Committee
Directors Call
April 6, 2012

Brian D. Athey, Ph.D.
Kevin A. Smith, M.S.
University of Michigan
Agenda

• What is tranSMART?
  – Early adaptors to be aware of
• What will be there in July?
• Future Potential?
• tranSMART and the CTSA Community?
What is tranSMART?

- An emerging **open-source non-commercial community** involved in clinical and translational research that includes pharma, non-profit, government, and academic groups who collaboratively build, share, and use the tranSMART platform to break down technical and cross-organizational barriers that prevent critical analyses across related data sets from being conducted.
What is tranSMART?

History

tranSMART wins CIO100 Award

Project Description: This subsidiary of Johnson & Johnson created tranSMART, a data warehouse entirely hosted in Amazon's Elastic Computing Cloud (EC2). It brings together R&D data from a variety of sources, allowing information to cross boundaries between scientific disciplines, development phases and business groups. The tranSMART warehouse takes data from across the entire research organization and puts it in the hands of scientists supporting translational medicine, which analyzes research from the laboratory to help create safer and more effective clinical trials. Started in the summer of 2008 and built over 18 months, the data warehouse adds role-based security to control access to the data. More than 150 users have already been trained on the system, which has been queried 2,000 times in last six months.

http://www.cio.com/cio100/detail/2014
What is tranSMART?

**Content**
- Reference data (Entrez, GEO)
- Curated data (vendors, open)
- Internal data sets
- Experimental platforms
- RDF sources
- Syndicated references

**Platform**
- Data schema/info. models
- Packaged ETL
- Application core
- APIs/SDK – services (i2b2)
- Standards (CDISC, etc.)
- Data exchange
- Security

**App store**
- Kaplan Meier
- ANOVA
- Gene Pattern + apps
- Galaxy + apps
- Ariadne
- Commercial tools
What is tranSMART?

- The tranSMART platform includes:
  - set of data models,
  - shared data sets,
  - data transformation utilities,
  - analytical web applications,
  - and i2b2 enabled

that can accelerate discoveries within complex biological systems by creating a standardized and semantically integrated database of research results linked to reusable and scalable self-service analytics.
Early Adaptors: Innovative Medicines Initiative - Vision

European Union and EFPIA CEOs joined forces to make drug development in the EU more innovative and efficient by forming the biggest PPP in Life Science

(adopted on December 20, 2007 by the European Council)
Early Adaptors: IMI SAFE-T Consortium

- Safer And Faster Evidence-based Translation
- tranSMART is being implemented on private cloud
- Validation is planned
Early Adaptors: IMI OncoTrack Consortium

- Colorectal cancer consortium
- tranSMART is being installed for KM
Early Adaptors: One Mind for Research

PTSD-TBI Vision diagram

The PTSD-TBI Ecosystem will be an iterative, dynamic network that continually optimizes research paradigms, incentive models, and outcomes.
What is tranSMART?

- **Board (Governance/Oversight)**
- **Work Groups**
  - Operations
    - Communications / Events
    - Code Governance & Open Infrastructure
  - Strategic/Program Mgmt.
    - Roadmap
    - Scientific Computing
    - Data Sharing
- **Projects**
  - Galaxy integration
  - VCF data interface/model
  - Postgress port
- **Task Forces**
  - Next Generation Sequencing integration/use
  - TBD? – ask group?
What is tranSMART?

**Current Project Board**

Expected to change ~June 2012 with new members/leadership structure

Volunteers from the community comprised of representatives from different perspectives/industries working to organize governance and provide leadership towards achieving goals

Current Members

**Non-Profit Research**
- Keith Elliston (CHDI)
- Mike Kellen (Sage)

**Academic Medical Centers**
- Brian Athey (U. Michigan)
- Yi-Ke Guo (Imperial College of London)

**Pharmaceutical R&D Informatics**
- John Shon (J&J), Sandor Szalma (J&J)
- Anne Westcott (Astra Zeneca)
- Christophe Gibault (Sanofi)

**Commercial Software/Legal**
- Peter Emerson (Recombinant)
What is tranSMART?

**Workgroups & Chairs**

- **Product Roadmap** Sandor Szalma (Janssen)
- **Communications & Events** Dan Housman (Recombinant), w/Keith Elliston (CHDI)
- **Open source/code governance** Terry Weymouth (U-M)
- **Scientific Computing/Bioinformatics** Yi-Ke Guo (IC-London)
- **Data sharing, data governance** Brian Athey (U-M) w/Mike Kellen (Sage)

More volunteers wanted to participate in each workgroup
Adoption – current trends (pilots)
We need help -- tracking adoption and documenting/sharing use cases

Adoption at relevant translational research organizations is a key factor in the long term success of the tranSMART open source project. We are actively tracking deployment at sites and helping the community to learn from experiences. In 2011 an early version of tranSMART was provided to multiple sites for review in advance of the full open source license. The following list is an initial list of sites that have active tranSMART initiatives for evaluation or implementation.

<table>
<thead>
<tr>
<th>Organization</th>
<th>Start Date</th>
<th>Notes</th>
<th>Stage</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Johnson &amp; Johnson</td>
<td>2008</td>
<td>J&amp;J is the initial development site for tranSMART. Used initially in immunology and oncology therapeutic areas.</td>
<td>Production</td>
<td>Pharma</td>
</tr>
<tr>
<td>Recombinant Data</td>
<td>2008</td>
<td>Recombinant maintains test and development versions of tranSMART. Recombinant was the initial development partner and works with multiple institutions to deploy tranSMART.</td>
<td>N/A</td>
<td>Development</td>
</tr>
<tr>
<td>Sage Bionetworks</td>
<td>2010</td>
<td>Sage deployed tranSMART as a component of the Synapse open discovery platform.</td>
<td>Production</td>
<td>Consortium</td>
</tr>
<tr>
<td>UBIOPRED</td>
<td>2010</td>
<td>A European multi-institution collaboration to share biomarker data relating to Asthma/COPD.</td>
<td>Production</td>
<td>Consortium</td>
</tr>
<tr>
<td>SAFE-T</td>
<td>2011</td>
<td>SAFE-T is an IMI initiative focused on Safety biomarkers. It was deployed a pilot of tranSMART. The project is expected to proceed.</td>
<td>Pilot</td>
<td>Consortium</td>
</tr>
<tr>
<td>University of Michigan Comprehensive Cancer Center</td>
<td>2011</td>
<td>The University of Michigan Cancer Center is using tranSMART to support a collaboration across many research groups for a multi-site oncology SNP study.</td>
<td>Development</td>
<td>Academic Medical Center</td>
</tr>
<tr>
<td>University of Michigan, Bioinformatics Core</td>
<td>2012</td>
<td>The Bioinformatics Core at the University of Michigan Department of Computational Medicine and Bioinformatics is enhancing tranSMART by 1) incorporating National Center for Integrative Biomedical Informatics (NCIBI) tools and data into tranSMART (<a href="http://portal.ncbi.org/gateway">http://portal.ncbi.org/gateway</a>) and 2) assisting in the support of tranSMART as an open source community.</td>
<td>Development</td>
<td>Academic</td>
</tr>
<tr>
<td>One Mind for Research</td>
<td>2012</td>
<td>One Mind for Research is doing a case study and a prototyping effort for data integration and sharing of clinically relevant data sets based around tranSMART.</td>
<td>Pilot</td>
<td>Consortium</td>
</tr>
</tbody>
</table>

Please provide what can be shared regarding your deployment / evaluation status!
Share with Kevin Smith (kasmith@umich.edu)
Initial Timeline

Jan 15
tranSMARTproject.org live

March 5 AMI and Github released

March 20
AMIA 2012

Q4 2011 eTRIKS review

Q1-4 2011 multiple private deployments (~8)

Feb 10 .9 GPL Open source code released

March 6 initial project board meeting

April 24 Bio-IT World 2012 launch 1.0 RC 1

Open Source Timeline
What will be there in July?

Roadmap 2012: Committed Updates

tranSMART 1.0 Release Dates (GPLv3) @ tranSMARTproject.org

• April 24 2012 - Release Candidate 1
• June 15 2012 - Release Candidate 2
• July 15 2012 - tranSMART 1.0 release
• Postgres port in progress Q3 (ideally in 1.0)
What will be there in July?

- i2b2 1.6 internal upgrade
- Inclusion of features from 'adopter sites'
- R query interface and plugin framework
- New statistical analysis pipelines including: Survival analysis, Heatmap, Hierarchal-clustering, K-mean clustering, PCA
- High dimensional biomarker support in Scatter plot and ANOVA
- Integration of biomarker demo data sets
- Documentation (End user, installation ETL, Admin)
- Demonstration version/reference updates
- Amazon AMI image
- Sandbox
Future Potential?

- Next-generation open source analytical and data sharing informatics platform—allowing sharing between pharma and academics
- Accelerating discovery by enabling collaboration among clinicians, researchers and informaticians
- Sustained through academic, government, and industry public-private partnerships
- Coalesces / builds upon NIH-funded initiatives -- NCBCs, CTSA, national data sharing and policy mandates
- Leverages existing and emerging national and international research networks

-------------------------

PPP and Core team still in formation. More details forthcoming...
Future Potential?
Michigan Internal Discussion
tranSMART and the CTSA Community? Questions and Discussion

• We need help in evaluating potential
• Looking for interested CTSA sites for discussion (leading to IKFC Working Group?)
Please consider getting involved ...

- [www.transmartproject.org](http://www.transmartproject.org)
- [LinkedIn tranSMART group](https://www.linkedin.com/groups/Transmart)
  social networking for community members
- [transmart-announce Google group](https://groups.google.com/group/transmart-announce)
  event and release announcements
- [transmart-discuss Google group](https://groups.google.com/group/transmart-discuss)
  post and discuss items