The tranSMART Foundation: tranSMART-Pro Project Introduction
Meeting Outline

• Introduction

• Roadmap
  – 16.1
  – 16.2
  – 17.1

• Business Analysis Project
  – Requirements Analysis
  – Use Cases

• Development proposals
  – Harvard
  – Hyve
  – Deloitte

• Funding Process
  – Commitments and Timeline

• Q&A
## tranSMART Foundation Priorities

<table>
<thead>
<tr>
<th>Innovative platform</th>
<th>Innovate efforts</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Improve the API</td>
<td>• ‘Full genome’ variants</td>
</tr>
<tr>
<td>• Support research projects</td>
<td>• ‘Wearables sensor’ data</td>
</tr>
<tr>
<td>• Integrate technologies</td>
<td>• ‘Commercial grade’ platform</td>
</tr>
<tr>
<td>• Support innovative efforts</td>
<td>- Installation process</td>
</tr>
<tr>
<td></td>
<td>- Stability and reliability</td>
</tr>
<tr>
<td></td>
<td>- Data loading</td>
</tr>
<tr>
<td></td>
<td>- Data availability</td>
</tr>
</tbody>
</table>
tranSMART Development and Release Roadmap

Project Name

17.2 Specification Development Release

17.1 Specification Development Release

16.2 Development Release

16.1 Release

Today

Longitudinal Data Scalable Genomics

Improved ETL, XNAT, SmartR

Improved Installation, code governance

3Q15 4Q15 1Q16 2Q16 3Q16 4Q16 1Q17 2Q17 3Q17 4Q17
16.1 Release Project Summary

• 16.1 Project Status
  – A new wiki area has been set up for the testing instances (under ‘Developing the platform’ on the main page)
  – Development and Release-candidate test instances

• 16.1 Release Summary
  – Scripted Install (supports Ubuntu/Postgres)
  – Digitally signed release artifacts (.tar files)
  – Extensive bug fixes and enhancements to existing functionality

• 16.1 is now open for beta testing
  – see the wiki page for links to test instances, instructions for reporting bugs etc.
  – Beta test period extends until 4 Apr, followed by time to bugfix & build release artifacts

• 16.1 Production Release planned for April 25th
16.2 Development Project Summary

• 16.2 Project Status (Completing Development Phase)
  – Planned functional upgrades (completed developments):
    • SmartR (Luxembourg/ITTM)
    • XNAT Plugin (Erasmus)
    • Improved ICE tool and ETL Loading (Sanofi)
    • GWAS enhancements (Pfizer)
  – Awaiting Legal Approval
    • Omics Cohort Selection (JNJ)
  – Ubuntu/Postgress and RHEL / Oracle installation support

• 16.2 Release planned for Sept. 30th
17.1 Specification Project Process

• 17.1 Project Requirements Gathering
  – Community Interviews
  – Use Case Development

• Technology Evaluation

• 17.1 Development Proposal Solicitation
  – Harvard, Hyve, Deloitte

• Foundation Recommendations
17.1 Business Analyst Project Interviews

- Membership Community Interviews
  - Academic
    - U.Mich. (Gold)
  - Pharma
    - Sanofi (Gold)
    - Pfizer (Gold)
    - Roche (Gold)
    - Takeda (Gold)
    - AbbVie (Silver)
    - Boehringer Ingelheim (membership in process)
  - Other Members
    - IOI (Silver)
    - The Hyve (Silver)
    - Perkin Elmer (Silver)
    - Thompson Reuters (Silver)
    - Rancho Biosciences (Silver)
Requirements Analysis

- Requirements definitions were driven by interviews
  - Members identified key requirements based on their use cases
- Use cases were collected and formalized
  - Each member was asked to provide use cases for key requirements
  - Foundation staff formatted use cases in standardized template
- Resulting gaps led to new requirements that were fed back to the community.
  - Collated and rationalized requirements, with rankings are being made available to the community
## tranSMART Community Priorities (1 of 2)

<table>
<thead>
<tr>
<th>Topic</th>
<th>Description</th>
<th>Priority</th>
</tr>
</thead>
</table>
| Platform Robustness and Performance | ▪ Make the platform more supportable and upgradable  
▪ Make queries more reliable and have better performance on scaling  
▪ Robust and well-documented APIs  
▪ Tune queries for Oracle backend  
▪ Support for federated data model  
▪ Improved ETL performance  
▪ Security and access control | 3.77     |
| Longitudinal Data Support    | ▪ Ability to align clinic or data capture visits across trials  
▪ Support unscheduled data (e.g., EHRs)  
▪ Support for i2b2-style 'encounters' table  
▪ Support for relative / elapsed time queries and comparison operators (e.g., patients with AE within 24 hours of dosing)  
▪ Support for date and event selections in advanced workflows and visualizations  
▪ Backward compatibility with existing study data, such that existing studies do not have to be reloaded | 3.23     |
| Cross Study Support          | ▪ Ability to merge data from multiple trials for analysis without losing data origins  
▪ Allow concepts to be independent of studies so that the same concept can be utilized and compared across studies | 3.10     |
### 17.1 tranSMART Community Priorities (2 of 2)

<table>
<thead>
<tr>
<th>Topic</th>
<th>Description</th>
<th>Priority</th>
</tr>
</thead>
</table>
| Support for High Volume Variant Data     | - Allow loading, querying, exporting, and cohort-building based on genetic variation data, whether derived from sequencing or other platforms (e.g., array-based assays)  
- Support for whole-genome volumes of variants (10s millions per patient) from 100s of thousands of samples  
- Support all variant types and annotations supported by VCF 4.2 spec (including structural variants)  
- Support for ARVADOS, and GA4GH API                                                                                                                                                      | 3.00     |
| Upgrade path / i2b2 integration          | - Backward compatibility from previous versions of tranSMART  
- Ensure date support from previous versions of i2B2  
- i2b2 integration must not mean that data have to be loaded twice                                                                                                                          | 3.00     |
| Continuation of SmartR or other plugin visualization/analytic tool (e.g., Spotfire) | - Harmonize workflows: Need to maintain SmartR workflows and tranSMART workflows  
- Ability to create a cohort and data set, then be able to apply multiple workflows against it                                                                                               | 2.85     |
| Support use of standard and internal proprietary ontologies | - Embedded support for ontologies to help make data curation (ETL) and cross study data queries and normalization easier                                                                                   | 2.85     |
| Better support for flexible ETL          | - tranSMART should recommend preferred ETL tool  
- Improved error handling                                                                                                                                                                         | 2.23     |
## Summary of Use Cases

<table>
<thead>
<tr>
<th>Topic</th>
<th>Description</th>
<th>Submitter</th>
<th>tf</th>
<th>Author</th>
<th>File</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clinical Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Longitudinal Data</td>
<td>BMI rate of change for genomic variants</td>
<td>Jay Bergeron</td>
<td>Ken Kubota</td>
<td></td>
<td><a href="https://docs.google.com/a/transmartfoundation.org/document/d/1j8gGCI7xkgV/">https://docs.google.com/a/transmartfoundation.org/document/d/1j8gGCI7xkgV/</a></td>
</tr>
<tr>
<td>Cairo user study [TraIT]</td>
<td>Combine experimental and clinical data to perform</td>
<td>Evert van den Broek</td>
<td>Ken Kubota</td>
<td></td>
<td><a href="https://drive.google.com/open?id=DB_J1ST33lbiOQk4vI5Z3N1VYVE">https://drive.google.com/open?id=DB_J1ST33lbiOQk4vI5Z3N1VYVE</a></td>
</tr>
<tr>
<td>Cairo user study [TraIT]</td>
<td>Combine experimental and clinical data to perform</td>
<td>Ken Kubota</td>
<td></td>
<td></td>
<td><a href="https://drive.google.com/open?id=DB_J1ST33lbiOQk4vI5Z3N1VYVE">https://drive.google.com/open?id=DB_J1ST33lbiOQk4vI5Z3N1VYVE</a></td>
</tr>
<tr>
<td><strong>Genomic Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Import VCF Files</td>
<td>Importing VCFs into transSMART</td>
<td>Sherry Cao</td>
<td>Keith Nangle</td>
<td></td>
<td><a href="https://drive.google.com/open?id=DBvhep0FHeba8uZQaM4bkmvMk">https://drive.google.com/open?id=DBvhep0FHeba8uZQaM4bkmvMk</a></td>
</tr>
<tr>
<td>Create Variant Set</td>
<td>Create a set of variants of interest to a study</td>
<td>Sherry Cao</td>
<td>Keith Nangle</td>
<td></td>
<td><a href="https://drive.google.com/open?id=DBvhep0FHeba8uZQaM4bkmvMk">https://drive.google.com/open?id=DBvhep0FHeba8uZQaM4bkmvMk</a></td>
</tr>
<tr>
<td>Create patient cohort based on variant data</td>
<td>Use genotype data to define patient cohorts</td>
<td>Sherry Cao</td>
<td>Keith Nangle</td>
<td></td>
<td><a href="https://drive.google.com/open?id=DBvhep0FHeba8uZQaM4bkmvMk">https://drive.google.com/open?id=DBvhep0FHeba8uZQaM4bkmvMk</a></td>
</tr>
<tr>
<td>Summary Statistics for variant data</td>
<td>Show summary statistics for a set of variants and a</td>
<td>Sherry Cao</td>
<td>Keith Nangle</td>
<td></td>
<td><a href="https://drive.google.com/open?id=DBvhep0FHeba8uZQaM4bkmvMk">https://drive.google.com/open?id=DBvhep0FHeba8uZQaM4bkmvMk</a></td>
</tr>
<tr>
<td>Export genotype data</td>
<td>Export genotypes for a patient cohort and variant at</td>
<td>Sherry Cao</td>
<td>Keith Nangle</td>
<td></td>
<td><a href="https://drive.google.com/open?id=DBvhep0FHeba8uZQaM4bkmvMk">https://drive.google.com/open?id=DBvhep0FHeba8uZQaM4bkmvMk</a></td>
</tr>
<tr>
<td>Perform genotype association analysis</td>
<td>Perform genotype association for a variant set and</td>
<td>Sherry Cao</td>
<td>Keith Nangle</td>
<td></td>
<td><a href="https://drive.google.com/open?id=DBvhep0FHeba8uZQaM4bkmvMk">https://drive.google.com/open?id=DBvhep0FHeba8uZQaM4bkmvMk</a></td>
</tr>
<tr>
<td>Percentile rank analysis (PhenWAS)</td>
<td>Perform percentile rank analysis for a variant set, set</td>
<td>Sherry Cao</td>
<td>Keith Nangle</td>
<td></td>
<td><a href="https://drive.google.com/open?id=DBvhep0FHeba8uZQaM4bkmvMk">https://drive.google.com/open?id=DBvhep0FHeba8uZQaM4bkmvMk</a></td>
</tr>
<tr>
<td>Query transSMART for cancer specific genotypes</td>
<td>The genetic variant is present in cancer samples but Tom Quaiser</td>
<td>Keith Nangle</td>
<td></td>
<td></td>
<td><a href="https://drive.google.com/open?id=DBvhep0FHeba8uZ1oMy0zdmSvMVU">https://drive.google.com/open?id=DBvhep0FHeba8uZ1oMy0zdmSvMVU</a></td>
</tr>
<tr>
<td>Query transSMART for gene fusion events</td>
<td>An analyst queries the data in transSMART to retrieve Tom Quaiser</td>
<td>Keith Nangle</td>
<td></td>
<td></td>
<td><a href="https://drive.google.com/open?id=DBvhep0FHeba8uZ1oMy0zdmSvMVU">https://drive.google.com/open?id=DBvhep0FHeba8uZ1oMy0zdmSvMVU</a></td>
</tr>
<tr>
<td>Query transSMART for copy number variation</td>
<td>The CNV may be specific to disease samples over no Tom Quaiser</td>
<td>Keith Nangle</td>
<td></td>
<td></td>
<td><a href="https://drive.google.com/open?id=DBvhep0FHeba8uZ1oMy0zdmSvMVU">https://drive.google.com/open?id=DBvhep0FHeba8uZ1oMy0zdmSvMVU</a></td>
</tr>
<tr>
<td>Apply gene expression signature to predict novel indication for drug</td>
<td>The CNV may be specific to disease samples over no Tom Quaiser</td>
<td>Keith Nangle</td>
<td></td>
<td></td>
<td><a href="https://drive.google.com/open?id=DBvhep0FHeba8uZ1oMy0zdmSvMVU">https://drive.google.com/open?id=DBvhep0FHeba8uZ1oMy0zdmSvMVU</a></td>
</tr>
<tr>
<td>Derive Gene Expression Signature to predict drug efficacy</td>
<td>The CNV may be specific to disease samples over no Tom Quaiser</td>
<td>Keith Nangle</td>
<td></td>
<td></td>
<td><a href="https://drive.google.com/open?id=DBvhep0FHeba8uZ1oMy0zdmSvMVU">https://drive.google.com/open?id=DBvhep0FHeba8uZ1oMy0zdmSvMVU</a></td>
</tr>
<tr>
<td>Microarrays: Samples with Specific aberration</td>
<td>Identify which samples have an ERBB2 amplification</td>
<td>David Merberg (Takeda)</td>
<td>Ken Kubota</td>
<td></td>
<td><a href="https://drive.google.com/open?id=DB_J1ST33lbiN3hMV3RTMVVMc">https://drive.google.com/open?id=DB_J1ST33lbiN3hMV3RTMVVMc</a></td>
</tr>
<tr>
<td>NHTMP TIMA User Story</td>
<td>Calculate survival times in colorectal cancer, test bic Jeroen Goos</td>
<td>Ken Kubota</td>
<td></td>
<td></td>
<td><a href="https://drive.google.com/open?id=DB_J1ST33lbiZ1ICN57VH46b6x">https://drive.google.com/open?id=DB_J1ST33lbiZ1ICN57VH46b6x</a></td>
</tr>
<tr>
<td>Proteomics Use case</td>
<td>bio marker discovery and validation for colorectile c</td>
<td>Thang Pham</td>
<td>Ken Kubota</td>
<td></td>
<td><a href="https://drive.google.com/open?id=DB_J1ST33lbiUk6w1TBPQv3eW">https://drive.google.com/open?id=DB_J1ST33lbiUk6w1TBPQv3eW</a></td>
</tr>
<tr>
<td>Proteomics Use case</td>
<td>Discovery and validation of markers in blood, tissue</td>
<td>Melike de Wit</td>
<td>Ken Kubota</td>
<td></td>
<td><a href="https://drive.google.com/open?id=DB_J1ST33lbiMhq5Xk9YDqF">https://drive.google.com/open?id=DB_J1ST33lbiMhq5Xk9YDqF</a></td>
</tr>
<tr>
<td>Microarrays: differential expression</td>
<td>Find out if the STIKS gene is differentially expressed</td>
<td>Ken Kubota</td>
<td>Ken Kubota</td>
<td></td>
<td><a href="https://drive.google.com/open?id=DB_J1ST33lbiKXYKDKvMVUS%D0%B0%D1%8E%D1%82">https://drive.google.com/open?id=DB_J1ST33lbiKXYKDKvMVUSают</a></td>
</tr>
<tr>
<td><strong>Ontology</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ETL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Security</strong></td>
<td>Partitioned Access to trial data use case</td>
<td>Administrators are able to grant access within and ac Ken Kubota</td>
<td>Ken Kubota</td>
<td></td>
<td><a href="https://drive.google.com/open?id=DB77f8Qe6fFJQzT7aVRbRx9mZ1k">https://drive.google.com/open?id=DB77f8Qe6fFJQzT7aVRbRx9mZ1k</a></td>
</tr>
<tr>
<td><strong>Data</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Summary will be made available on our Wiki
**Individual use cases will be made available via public share.
Use Case Example

• Collected approximately 2 dozen use cases related to clinical & longitudinal data and genomics.

• For example:
  – An analyst queries the data in tranSMART to retrieve cancer specific genotypes from a study or a group of studies, e.g. the genetic variant is present in cancer samples but not in normal samples (from the same patient). In addition, when outcome data from drug treatment is available, analyst should be able to optionally retrieve the genotypes along with associated outcome data linked to the patients.
  – User identifies cohort(s) of interest using a combination of clinical attributes and variant attributes, including:
    - Gene name(s)
    - Variant ID(s) (rsid, probeset ID)
    - Specific mutation, e.g. V600E of BRAF gene
  – Query performance should be within the norm of database size with proper indexes and the hardware it runs on.
Technology Evaluation

• Work Flow Systems (as a backend options)
  – Adam / Spark
  – Broad Institute GATK and Google Cloud
  – Google Genomics platform
  – Curoverse/ARVADOS (Keep / Crunch / Lightning)

• Standards activities
  – Global Alliance 4 Genomics and Health (GA4GH)
  – Workflow systems support for GA4GH APIs

• Results:
  – tranSMART will standardize on GA4GH and associated APIs
DEVELOPMENT PROPOSALS
Solicited various development firms for proposals. Three firms have bid on 17.1 development for both system integration and technology development.

<table>
<thead>
<tr>
<th>Project</th>
<th>Deloitte</th>
<th>Harvard</th>
<th>Hyve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systems Integration and Platform Evolution</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Longitudinal Data</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Scalable Genomics</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>I2b2 integration</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Harvard Proposal

Phase | ID | Summary | Estimate (Days) | Estimate (Hours) |
--- | --- | --- | --- | --- |
0 | t-1 | Discussions around upgrade path for new i2b2 and tranSMART release | 1w | 40 |
0 | t-2 | Determine merge strategy for getting our code back into tranSMART main codebase | 3d | 24 |
0 | t-3 | Ramp up new hires to i2b2 tranSMART ecosystem | 1w | 40 |
--- | --- | --- | --- | --- |
Total | | | 104 |

### Phase 1 - t-4

- Determine Database Schema Integration strategy | 2w | 60 |
- Create local database to develop off of | 3d | 24 |
- Get current release working on developer’s machine | 3w | 120 |
- Install plugin from Hackathon into current branch and fix issues | 8d | 64 |
- Install local i2b2 via Virtual Machine/Docker and connect to tranSMART plugin | 1w | 40 |
- Review plugin architecture | 2w | 80 |
- Get new 1 branch building via Bamboo | 3d | 24 |
- Update Ansible to use new tranSMART Codebase | 8d | 64 |
- Build base database schema for tranSMART, different from clean i2b2 schema | 2w | 80 |
- Move database schema into i2b2base | 2w | 80 |
- Set up deployment of database in ansible | 1w | 40 |
- Documentation around database Schemas | 3d | 24 |
- Regression Testing / Bug Fixing | 4w | 160 |
--- | --- | --- | --- | --- |
Total | | | 1150 |

### Phase 2 - t-17

- Refactor plugin code to break out logic for generating i2b2 messages | 3w | 120 |
- Develop automated tests for new plugin code | 3w | 120 |
- Finish Modifier implementation within interface for generating i2b2 queries | 2w | 80 |
- Finish Backend Modifier implementation | 2w | 80 |
- Finish Temporal Queries implementation within interface for generating i2b2 queries | 2w | 80 |
- Finish Backend Temporal Queries implementation | 2w | 80 |
- Documentation on Modifier/Temporal Revisions | 1w | 40 |
- QA Testing on new functionality | 4w | 160 |
--- | --- | --- | --- | --- |
Total | | | 800 |

### Phase 3

- Investigate current install procedures for industry standards | 2w | 80 |
- Enhance/Stabilize tranSMART Docker | 1w | 40 |
- Create R Docker to run with tranSMART and integrate | 2w | 120 |
- Configure i2b2 Docker to work with tranSMART Docker | 1w | 40 |
- Create i2b2 database installer based on Liquibase | 2w | 80 |
- Package whole project as an AMI/Docker Environment / All in one | 3w | 120 |
- QA Different Installation Scenarios | 3w | 120 |
--- | --- | --- | --- | --- |
Total | | | 720 |

### Phase 4

- Integration fact validation UI and backend changes | 3w | 120 |
- Integrate fact validation schema changes | 2w | 80 |
- Integrate system summary dashboard | 1w | 40 |
- Integrate system summary schema changes | 1w | 40 |
- Integrate Sample Explorer Linkage UI | 2w | 120 |
- Integrate Sample Explorer Linkage Schema changes | 2w | 80 |
--- | --- | --- | --- | --- |
Total | | | 400 |
1. **Clinical Data Improvements:** Changes to the data model and APIs for clinical data to allow, amongst others, Electronic Health Record data and sample data to be stored and queried in the tranSMART UI and (other) REST clients and to improve the cross-study data support.

2. **Scalable Genomics Backend:** Changes to the data model and APIs to allow the storage and querying of large scale genomics data.

3. **Creation of the tranSMART Core:** Taking the above improvements into account, we will then build, package and version the tranSMART Core, as described in detail in the TranSMART 2.0 Architecture Recommendation whitepaper that was released last year.
Deloitte Proposal
## Summary of Development Proposals

<table>
<thead>
<tr>
<th>Service Provider</th>
<th>Hours</th>
<th>Budget</th>
<th>Blended Rate</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Harvard</td>
<td>3,264</td>
<td>$610K</td>
<td>$187/hr</td>
<td>I2b2 Integration</td>
</tr>
</tbody>
</table>
| Hyve             | 6,048  | $1,067K  | $176/hr      | Systems Integration  
Scalable Genomics 
Longitudinal Data |
| Deloitte         | 17,000 | $2,300K  | $135/hr      | Systems integration  
Clinical data integration  
Scalable genomics  
Longitudinal data  
Cross-study analysis  
Performance tuning, regression testing and documentation |

Expense: $300k

* There is significant overlap between proposals. The Governing Board will decide which specific projects are funded.*
Foundation Recommendations

- Work with funding partners to define a budget
- Work with development partners to determine what priorities can fit in the 17.1 budget.
  - Platform Robustness and Performance
  - Longitudinal Data
  - Scalable Genomics
  - Cross Study Support
  - Platform Enhancement and Evolution (Ontology, SmartR)
  - Full i2b2 Integration
- Execute refined service provider proposal(s) and execute development project
Collaborative Development Funding Models

• **Industry Collaborative Alliance**
  - Example: CBDD – Thomson Reuters
    - Group of companies contracting with TR to develop shared, proprietary tools
    - Advantage: Single source developer and commercial terms with for-profit vendor
    - Disadvantage: closed access

• **Academic Collaborative Alliance**
  - Example: NDEX Project – UCSD
    - Group of companies, augmented by grant funding, to develop open-source platform
    - Advantage: Single source developer, but with academic funding model
    - Disadvantage: Academic project

• **Non-profit Collaborative Alliance**
  - Example: Linux Foundation Projects
    - Sets of projects, with individual stakeholders, producing open-source software
    - Advantage: Diverse developer resources (contract and volunteer), commercial funding model, can be augmented by grant funding
    - Disadvantage: requires minimal critical mass of financial contributors
tranSMART Foundation Proposed Funding Model

• Modeled after Linux Foundation
  – Non-profit Collaborative Alliance
  – Advantages:
    • Enables the use of diverse software development resources
      – Contract, Volunteer, grant funded
    • Provides key advantages to stakeholders
      – Control budget, spending, deliverables and timelines
      – Prioritize features and requirements
  – Disadvantages
    • Distributed software development resources (contract and volunteer)
    • Minimal critical mass required to initiate
  – Examples:
    • openDaylight, R-Consortium, CloudFoundry, Automotive Grade Linux, etc.
    • Over 30 successful programs at Linux Foundation
The tranSMART-Pro Alliance

• Mission:
  – As a collaborative, open-source community, to create the leading, commercial grade translational research analytics platform, to provide access to all functionality and data through open APIs, and to ensure the sustainability of the tranSMART platform.

• Purpose:
  – To invest in the development of the commercial grade tranSMART platform
  – To prioritize features based upon the needs of the Alliance Members and the community
  – To provide a sustainability model for tranSMART
    • And a model for other open-source and open-data projects at the Foundation
tranSMART-Pro Alliance Structure

- **tranSMART Foundation**
  (management and oversight, and enabling infrastructure)

- **Governing Board**
  (drives business decisions and prioritization for Alliance)

- **End User Advisory Group**
  (develops use cases & recommends new features and enhancements)

- **Technical Steering Committee (TSC)**
  (drives Alliance technical direction)

**transSMART-Pro Alliance**
tranSMART-Pro Alliance Overview

• Governing members contribute $250K per year
  – There is a 2 year minimum commitment, with at least 3 members to start

• Investments go toward the development of tranSMART
  – To fund developers to contribute code
  – To fund a systems integrator to ensure consistency across contributions
  – To fund necessary infrastructure for production and support of the platform

• All projects managed by Governing Board, that manages all investments, timelines and milestones
  – Each contributor has a seat on the governing board
  – The Foundation chairs the board
  – Up to three seats can be allocated to community contributors
    • A ratio of 1 board seat per 10 contributors
    • Substantial investment is required… no free riders

• Development managed by a Technical Steering Committee, responsible for details of the project
  – Overseen by the Governing board
  – Each governing board member has a seat on the TSC
  – The TSC is chaired by the Foundation
  – Other members by merit, as approved by the Governing Board
Participation Model for tranSMART-Pro Alliance

- Anyone may participate; the tranSMART-Pro Project will encourage companies to become sponsoring participants and support the mission of the tranSMART-Pro Project with resources
  - Corporate participation levels include Governing, Community and Associate Entity levels.
- A Governing Board will oversee business decision making and resource allocation
  - Composed of Governing participant appointees, elected Community representatives and the TSC Chair
  - The Governing Board will address the business needs associated with budgeting, planning for any large meetings (e.g. a Summit), marketing (if any), feature prioritization and contracting.
- A Technical Steering Committee will focus on overseeing the technical community
  - Approves new tranSMART-Pro Project projects, sets development process guidelines for the community, sets release guidelines, oversees technical collaboration with upstream projects, etc.
  - The initial TSC will be bootstrapped with representatives from the Governing Members level participants with the goal of adding projects with Maintainers to the TSC
  - Anyone can contribute, become a project committer or maintainer and thereby serve on the TSC based on the success and quality of their contributions as recognized by peers.
- An End-user Advisor Board will provide end-user insights and use cases
  - Provides end-user input into the project
  - Anyone can participate, based upon merit, and approval by the Governing Board
# tranSMART-Pro Alliance Participation Levels

<table>
<thead>
<tr>
<th>Participation Level *</th>
<th>Annual Fee</th>
<th>Governing Board Seat</th>
<th>TSC Seat</th>
<th>End User Advisory Board</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Governing Member</td>
<td>Flat fee: $250,000</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Initial 2-year participant commitment</td>
</tr>
<tr>
<td>Community Member</td>
<td>$20,000 - $100,000 depending on size**</td>
<td>1 per 10</td>
<td>By Merit</td>
<td>Yes</td>
<td>Capped at 3 board seats</td>
</tr>
<tr>
<td>Associate</td>
<td>Free</td>
<td>No</td>
<td>By Merit</td>
<td>Yes</td>
<td>Non-profits, open source projects</td>
</tr>
</tbody>
</table>

Anyone (including individuals, students and developers in open source project communities) may participate in the technical community work without becoming a member of the tF or tranSMART 2 Alliance.

- The tranSMART-Pro Alliance is setup as a Collaborative Project under The tranSMART Foundation legal entity. All participants must be members of The tranSMART Foundation to participate. Please visit [www.transmartfoundation.org](http://www.transmartfoundation.org) to learn more about tranSMART Foundation membership.

** Community Member Annual Fee Scale**

- > 3,000 employees = $100K
- 1,000 < 3,000 employees = $50K
- 500 < 1,000 employees = $25K
- < 500 employees = $20K
tranSMART-Pro Alliance Member Expectations

Active participation contributes to a vibrant ecosystem

- Participate in Working Groups on the definition of new Service Frameworks
- Potential to contribute and share new reference implementations and compliance tests of Service Frameworks
- Provide input/feedback on proposals for new Service Frameworks
- Propose use cases that help define product interface definitions
- Promote Alliance certified products via Alliance PR, member PR, events, member white papers, etc.
tranSMART-Pro Alliance Governing Board Expectations

Business Leadership: Governance, Operations

• Governing Membership provides 1 tranSMART-Pro Project Governing board seat
• Oversee project milestones and deliverables
• Propose and approve budgets
• Attend monthly Project Governing board meetings/calls
• Work towards the Collaborative Alliance Objectives
• Promote the Collaborative Alliance and its ecosystem
• Encourage expanded use of Alliance products and services
• Help protect the ecosystem and make strategic decisions to further the future direction of Service Frameworks
• Participate in fundraising and project membership development
tranSMART-Pro Technical Steering Committee Expectations

Senior Technical Leadership

- Governing Membership provides 1 TSC seat
- Guarantee an open decision making process on business and technical issues by senior technical decision makers
- Define requirements for contributing code (code style, documentation, test cases, use cases)
- Define programming languages that are acceptable for contributions and which set of language bindings are required
- Responsible for targeting simultaneous release dates from working groups
- Define release quality standards and enforce technical best practices
- Approve technical work groups around various Service Frameworks to guide expansion of existing services and/or creation of new services
- Attend bi-weekly meetings/calls
Founding Members
Licensing of new developments

- Some code is currently GPLv3
  - Some contributions have Apache 2 license
- Development project will work to isolate GPLv3 code
  - Using a modular approach to development
- All new contributions will be under Apache 2 license
  - Inbound code contributed under open license
  - All contributors execute developer certification of origin (DCO)
- All new distributed code will be made available under Apache 2
  - With exception of components subject to GPLv3
  - Or an alternative open license as determined by the Governing Board
How to Join the tranSMART-Pro Development Project

- Research other tranSMART Foundation Collaborative Projects:
- Find out more the Project at:
- Fill in the online form for more info:
- Contact Keith Elliston at the tranSMART Foundation:
  keith.elliston@transmartfoundation.org