DATASIM Stakeholder Analysis Report

Data and Training Analytics Simulated Input Modeler

27 May 2021

This work was supported by the U.S. Advanced Distributed Learning (ADL) Initiative (HQ003419C0061). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the ADL Initiative or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes.

Stakeholder Analysis Report Nov 2020

DATASIM

Data and Training Analytics Simulated Input Modeler

Prepared by Cliff Casey, Yet Analytics, Inc.
Submitted to ADL Director on December 8, 2020

Purpose of this Document

The purpose of this document is to present a report and summary documentation regarding the Stakeholder Analysis for DATASIM. The report begins with an overview describing how DATASIM will help in the implementation, testing and validation of TLA projects and continues to the benefits of using DATASIM for Profile development and stress testing with a real world use-case, STEEL-R. This report will also outline the feedback received from this stakeholder group and the value this engagement could provide for the continued iterative cycles in the development of both projects.

Table of Contents

DATASIM Overview

DATASIM is an open source platform and set of open source specifications which can be used to generate simulations resulting in realistic xAPI data. The platform uses xAPI Profile Patterns, Templates, and Concepts to model behavior for a cohort of simulated actors in the form of xAPI datasets.

These datasets offer insight into the available Patterns and paths available in an xAPI Profile, and can be used to model complex learner behaviors to evaluate the effectiveness of xAPI data design. The output of these simulations can be used as a design tool to iteratively improve upon xAPI Profiles themselves. Because DATASIM is capable of producing a high throughput of records, these datasets may additionally be used to benchmark and stress-test components of the Total Learning Architecture (TLA) and distributed learning projects. DATASIM requires valid xAPI Profiles in order to generate datasets, so as part of the platform it has the capability to validate xAPI Profiles and other inputs against specification.

DATASIM is open source under the Apache 2.0 License and is funded by the Advanced Distributed Learning Initiative at the United States Department of Defense. The source code of the project[1], as well as the source code of an accompanying frontend user interface / client[2] can be found on GitHub.

The Alpha version of DATASIM, released in March 2020, has demonstrated the capability to generate conformant Profile-aligned xAPI datasets at scale. It has also demonstrated the ability to model and guide simulations based on Actor Alignments and other parameters.[3]

TLA Data Flow Validation

DATASIM has the capability to perform simulated generation for multiple layers of the TLA and also to report out on what data was injected into the test infrastructure at which level. DATASIM can also replicate the behavior of a high volume producer at the Noisy, Transactional and Authoritative layers. This can be used to validate that the correct statements made it to the correct layer(s) of the TLA implementation and validate the overall data flow.

[1] DATASIM Code Repository: https://github.com/yetanalytics/datasim
[2] DATASIM-UI Code Repository: https://github.com/yetanalytics/datasim-ui
[3] DATASIM Requirements Report By Shelly Blake-Plock Yet Analytics Inc.

Profile Validation

Using DATASIM as a Profile design tool will assist in uncovering validation issues. As Profiles are being developed, they can be tested and iterated upon. All inputs to DATASIM are validated and errors are presented to the user. Assumptions can be validated about the shape of data which are indicated by a given Profile.

Group and Role Alignments

In DATASIM, there is the capability to designate an individual actor to a group of actors and/or role. Alignment specification to specify any one of the following as a Personae for a given Alignment: Group, Role, Actor. Add to the Personae Specification the ability to create and name multiple Groups and place each Actor within exactly one Group. An Actor must be in a Group, but if the Group does not have an Alignment it does not impact the simulation.

STEEL-R Stakeholder Use Case

The STEEL-R project seeks to transform the way simulation-based and game-based technologies are used to train in the DoD. This project seeks to improve Soldier and Team/Unit development by providing persistent performance tracking that can be used to facilitate more data-driven recommender engines. The goal is to use prescriptive AI to challenge teams to become experts efficiently and effectively using sports and military psychology.

The STEEL-R project is currently in a phase where it could benefit from a prototyping and xAPI Profile editing tool that will allow them to design better xAPI Profiles and alignments through high capacity simulated datasets. The main reasons that make the STEEL-R project an excellent candidate as a test case are:

- Defined xAPI Profile Components ready for immediate use
- STEEL-R is being architected as a TLA implementation, generating data at all three layers
- There is a need to validate architecture without necessarily having all high volume data sources instrumented
- This feedback can assist in the further development of the Master Object Model (MOM) Profile Specifications.

Approach

This approach will use DATASIM to assist in the overall goals of the project, but also collect feedback from those experiences to improve and iterate upon DATASIM itself. STEEL-R is an ideal use case for this because this project has defined xAPI Profile components that are

available for immediate initial testing. The data structure was designed using TLA implementation and has statements at three different layers of the TLA. This allows for validation without the need for manually generating high volume source data. By using DATASIM as a prototyping tool, this will highlight issues caused by xAPI Profile design. This allows visibility into problems during the development and staging phases before they become problems in production.

Extension of MOM Profile

There is a need for a robust vocabulary in regards to statements generated for these DoD training scenarios. This presents an opportunity to both use the Master Object Model (MOM) Profile as the basis for components in the STEEL-R domain Profiles, but also to suggest new ideas for the MOM Profile based on real life use cases. DATASIM would allow this team to develop DoD specific Profile components to collect the level of granularity the stakeholder is seeking.

- The STEEL-R extension of the MOM Profile will be an overarching secondary xAPI Profile with specific ontology to this project and eventually other DoD needs.
- The MOM Profile can be used in DATASIM along with a xAPI Profile specifically created for STEEL-R use that aligns with and references elements within the MOM Profile, but with additional domain-specific layers of granularity.
- This Profile would be a separate xAPI Profile that can conform and be used to influence updates to the MOM Profile. The stakeholders want to use DATASIM to generate examples of the shape the xAPI data could look like and tailor the ontology using those datasets.

Top-Down Design

The STEEL-R team expressed the desire to move toward a top-down approach to design. Top-Down design is defined by moving from a high-level design concept towards the lower levels. Instead of focusing on the more complex technical development challenges, start with what data is currently output by their systems and what needs to be collected. Using DATASIM, the project will be able to identify the high level questions that still need to be answered and design the Profiles and components based on those needs. This would allow stakeholders to see what would potentially be created and reverse engineer those datasets to create better xAPI Profiles and components. The goal for this engagement is to generate these xAPI simulated statements from the GIFT and After Action Review, and then translate these into MOM conformant statements. This will allow STEEL-R stakeholders to see what statement could be coming from the Authoritative LRS and make data driven decisions on the direction of the project.

Metadata Alignments

STEEL-R would like the ability to test performance factors between outcome and response, both emotionally and cognitively. Sensory and behavioral data being collected from multiple

systems need to be tracked through statements. DATASIM will allow the team to test how data will feed into those overarching competency frameworks and how they are managing the affective component. The goal is to find the alignments between the metadata and the MOM Profile, make these data points into Profile components, and then test these components in DATASIM. The behavior component could be a contribution of performance factors between outcome and response. The desired outcome would be to find what method, medium or area of performance that particular assessment pertained to.

Volume Testing

As the STEEL-R project reaches the next phase where they are building out the proof of concept, the project will need better understanding of the system capacity. DATASIM can create usable testing data up to 5,500 xAPI statements per second, allowing it to test all of the components of learning data infrastructure in the STEEL-R implementation.

Impact

This engagement would allow the STEEL-R team to test their current xAPI Profiles and show STEEL-R stakeholders what data could potentially be generated by their systems. This would also demonstrate how statements relate to each other within the STEEL-R data architecture. DATASIM can be used to validate downstream systems before changes are made to the source system.

This will benefit the DATASIM project, STEEL-R project, as well as all DATASIM users. DATASIM gains valuable feedback from a key stakeholder, while STEEL-R gains access to deterministically repeatable datasets to test Profile design and the efficacy of the data architecture. The lessons learned about the use of xAPI Profiles and specifically the techniques involving DATASIM during this engagement serve as a proof of concept of the application of these technologies in the STE environment.

Risks

The risk of not using a Profile design and testing tool would be to allow bad data design such as unviable verb ontology. This would create bottlenecks within the development process and delay the current sprint schedule. DATASIM's ability to verify the quality of an xAPI Profile through datasets composed of any collection of xAPI Profiles could prevent time and energy being spent building out unviable Profiles. DATASIM can show the user what areas of a learning data environment are being used incorrectly and highlight undesirable data design.

Without using DATASIM, STEEL-R would not have a consistent approach to evaluate their existing and newly created xAPI Profiles without constant access to live training data. This project can solve common road-blocks using initial data simulation to design and inform new

xAPI Profiles. DATASIM brings testable iterative design capabilities for both the Profiles and the data structure itself.

This approach also allows the testing of high volume throughput of the infrastructure before the use of live data, allowing for the prediction of scaling considerations in enterprise implementations. Without access to realistic high volume test data, the risk to the project is improper architecture which can cause performance degradation under real load conditions. Extensive testing before using live data will allow STEEL-R to make critical changes to development and infrastructure before it affects the budget and timeline. Using DATASIM at the beginning stages could assist in getting enough data points to establish a data strategy that could eventually be applied at scale across all DoD training operations.

Timeframe

The following is the work that must be done on DATASIM's Use-Case Scenario.

- Gather Additional Feedback and Requirements
- Identify and execute actions for which DATASIM can assist the STEEL-R project by April 2021.
    - Profile Design Development will begin in December and continue throughout the proof of concept phase and beyond.
    - Integration Testing will begin as soon as the subsequent LRS are implemented in early 2021.
- Dependencies for STEEL-R - Profile and Profile Components
    - Gap Analysis for GIFT - STEEL-R must pivot what data they are generating, agree on resolution of what is not being output by GIFT, but are needed in the STEEL-R implementation. The STEEL-R team can then design Profile components around that ontology of that set of data and feed into the ADL sandbox and other places this project exists.

How to Evaluate Success

- The STEEL-R and DATASIM engagement will provide the ability to develop, simulate, and test data aligned to the xAPI Profiles specification prior to publication.
- This will increase STEEL-R's capacity to support the development and iteration of new xAPI Profiles that are aligned or reference the MOM Profile.
- Ensures xAPI Profiles that will be published are both valid and tested.
- DATASIM will be used in each phase of development to expand upon the existing Profile components and volume testing in the future.
- The STEEL-R project will have a proof of concept demo in February 2021.

STEEL-R Stakeholder DATASIM Feature Ideas

During our initial stakeholder needs analysis with the STEEL-R team a number of additional features and capabilities came up in conversation. While they are not necessary to implement to support the STEEL-R effort we will assess during this engagement where these capabilities may be useful to help guide future platform development.

Statement Invalidation Tagging

The stakeholders would like the ability to tag a statement in such a way that it can be invalidated in a simulated instance. Currently, the registration information within a simulation correlates to patterns detailed in the Profile. These are the only distinct unique identifiers for the actor and pattern within each statement. This will pinpoint where in the simulation and pattern the actor is. This does not currently allow the user to precisely remove certain statements from a dataset.

Statement Generation by Role or Group

Every statement is currently being generated by an individual actor. However, a group of actors can be designated and those groups can be aligned within DATASIM. An alignment is a combination of an actor, group of actors or specific role, a weighted probability, and an xAPI Profile component that infers whether they are more or less likely to engage in an action within the simulation. The action can be a whole activity unto itself or a verb within a statement.

The stakeholders expressed a potential need to have statements generated by the groups/teams themselves. In each assessment, groups and individuals currently do not influence each other. The need to distinguish between those statements could assist in the testing of role and group alignments within DATASIM. In the case of the STEEL-R scenario, both the squad leader (individual actor/role) and the fireteam (group of actors) would both be able to generate statements as an individual and as part of a set group of actors. The user would be able to link statements between the team and the individual associated with that team.

Experience Testing or Behavior Prediction Capabilities

Predicting human performance is not in the overall scope of the DATASIM project. The user can take multiple groups and adjust their modeling to get interesting novel results, but the simulation generator controls the outcome of the dataset. The user can experiment with how data will react to certain alignments, personas and profile components. DATASIM does have the ability to predict what data shapes in which Profiles will cause problems in given configurations of the TLA. In the case of the STEEL-R demo scenario, this could show challenges with modeling specific competencies such as muzzle flagging. The Profile alignments can be adjusted to create the desired results.

Conclusion

The STEEL-R team can start using DATASIM right away. This team will create DoD ontology from Task statements, Army Training Network and existing After Action Reviews using examples generated of what the xAPI could look like.The current state of the DATASIM prototype meets or exceeds the expectations of a Readiness Level 5 Software Application in all areas of evaluation, except being tested in a relevant environment and meeting predicted performance.

This real scenario testing will move DATASIM into its next phase of development. DATASIM will assist in the creation of a DoD specification xAPI Profile that is aligned to the MOM Profile. The STEEL-R team will be able to iterate and improve upon the existing xAPI Profiles and inform the development of new Profiles. These datasets will assist in the top down development of STEEL-R by showing gaps in the simulated array of statements, helping to acquire the correct requirements. Some of these requirements include collecting and storing metadata, both sensory and behavioral. DATASIM can assist in the creation of alignments that can verify assertions of this metadata.