Round-Trip: An Automated Pipeline for Experimental Design, Execution, and Analysis

Daniel Bryce[1], **Robert P. Goldman**[1], **Matthew DeHaven**[1], **Jacob Beal**[2], **Tramy Nguyen**[2], **Nicholas Walczak**[2], **Mark Weston**[3], **George Zheng**[3], **Josh Nowak**[4], **Joe Stubbs**[5], **Niall Gaffney**[5], **Matthew Vaughn**[5], **Chris Myers**[6]
[1]SIFT, LLC., [2]Raytheon BBN Tech., Inc., [3] Netrias, LLC., [4]Strateos, Inc., [5]Texas Adv. Computing Center, [6]Univ. of Utah
dbryce@sift.net

1 INTRODUCTION

Lack of automated tools for experimental design, execution, and analysis is a key barrier to quickly evaluating biological designs. Experimentalists often face challenges describing experimental plans in a format laboratories can act upon, while labs have difficulty linking experimental data to metadata and plans, making this an expensive manual undertaking [2]. Metadata describing contents, conditions, and context of experimental samples is key to gaining insights from experimental data, but acquiring, tracking, and maintaining it is tedious, expensive, and error-prone.

We alleviate these difficulties with automation. As part of the Synergistic Discovery and Design (SD2) project, we have addressed these needs by developing a *Round-Trip* (RT) architecture simplifying this process by automating several steps where human intervention was previously required. Figure 1 illustrates the major RT components, proceeding left-to-right on top creating an automation-assisted experiment. RT attaches metadata to raw data as it returns from right to left on the bottom. Key steps (cross-referenced in Figure 1, with supporting components detailed in Section 2) include:

(1) **Authoring an Experiment Request**: Experimentalists, starting with a "notional" experiment in mind, author a semi-structured *Experiment Request* document.
(2) **Annotating the Experiment Request**: RT generates an *Annotated Experiment Request* that hyperlinks experimental constructs to SBOL[6] definitions through a simple Data Dictionary.
(3) **Structuring the Experiment Request**: Linked constructs and experiment sample tables are converted into a *Structured Request Template*, formally defining a set of minimal requirements for the experiment.
(4) **Experimental Design**: RT expands these to an *Experiment Design* with metadata for every measurement of every sample, plus a set of *Lab Parameters* configuring a machine-executable laboratory protocol.
(5) **Experiment Execution**: The laboratory (Strateos) conducts the experiment, generating *Experiment Data*.
(6) **Check Expected vs. Actual Data**: RT generates *Expected Experiment Data* and compares it with the Experiment Data, checking that the laboratory fulfills the design and identifying any discrepancies.
(7) **ETL**: The RT then stores experiment data, conducts ETL processing, and annotates the Experimental Request with the final data products.

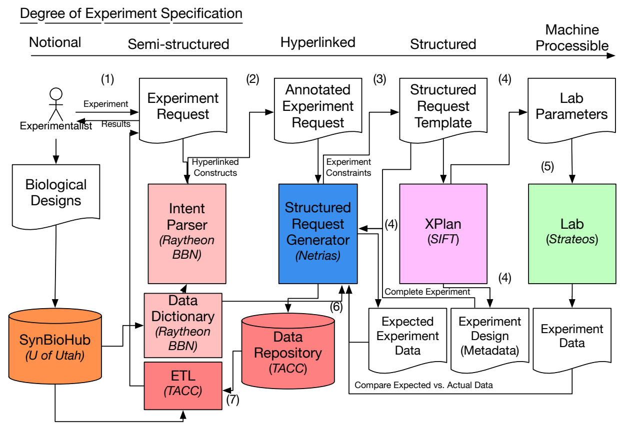

Figure 1: Round-Trip Architecture. (Component developer's affiliation noted in parentheses.)

By adopting "make metadata easy" design principles—providing human affordances, automating tedious metadata design and encoding, and reacting and repairing as deviations arise—this architecture provides many benefits to experimentalists. For example, it connects experimental data and subsequent analyses with deeply-represented experimental constructs by resolving user-friendly construct names. Experiment Requests can be partial, and RT will fill in details in experiment planning. RT also flags mismatches between expected and actual data for follow-up and diagnosis.

2 ROUND-TRIP ELEMENTS

Here we highlight key aspects of software components and data artifacts implementing the major stages of RT. Each component is involved in building up metadata through reducing

This work was supported by Air Force Research Laboratory (AFRL) and DARPA contracts FA875017CO184, HR001117C0092, and HR001117C0094. This document does not contain technology or technical data controlled under either U.S. International Traffic in Arms Regulation or U.S. Export Administration Regulations. Views, opinions, and/or findings expressed are those of the author(s) and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government. Approved for public release, distribution unlimited (DISTAR Case 32707).

Media	Dilution	B-Est. @ 16h	Rep.	Strain
SC Media	50x	0.0, 0.05 uM	10	UWBF_24926, UWBF_24952, UWBF_24959
SC Media	50x	0.0 uM		UWBF_24960, UWBF_25784, UWBF_24962

Table 1: Experiment Request Measurement Table

high-level experiment descriptions to machine representations, and then attaching measurement data to the metadata.

Semi-Structured Experiments: Experiment Requests (ERs) are documents including both prose and tabular descriptions of an experiment. The prose provides context, motivations, and anticipated results, along with lists of strain and reagent descriptions. The ER also includes a measurement table and a parameter table. The measurement table lists experimental factors by column, and constraints on their values by row. Table 1 shows an example measurement table indicating the experiment should contain ten replicates of the strains in the first row, each induced with either 0.0 or 0.05 uM beta-estradiol at 16 hours. It also states that the number of replicates of strains in the second row is still to be determined, and that these are not induced (i.e., 0.0 uM beta-estradiol)

Hyperlinking Experiments: The Intent Parser (IP) [5] processes the ER to identify constructs appearing in the Data Dictionary, then links them to SBOL descriptions in SynBioHub [4]. The Data Dictionary [1] maps each of (potentially many) construct common names to a canonical definition URI in SynBioHub. For example, the term "B-Est" in Table 1 is a common short-hand term that will be linked to the beta-estradiol reagent definition. IP likewise links strain identifiers (e.g., UWBF_24926) and media to their definitions. This provides experimentalists flexibility with common terms and shorthand, while unifying them across experiments (e.g., another ER might use Beta-Est. instead of B-Est.).

Structuring Experiments: Structured Requests (SRs) formally represent the set of samples an experiment will generate. SRs take two forms: templates and expected samples. The SR Generator creates a template capturing constraints from the ER, but this may not map directly to samples (e.g., the second row in Table 1 omits replicate count). After experimental planning (below), the SR Generator expands the SR and an Experiment Design into expected samples, which can be checked against actual samples on experiment completion, also matching lab-specific identifiers (e.g., LIMS inventory IDs) with ER common names via the Data Dictionary.

Machine Processible Experiments: The Experiment Planner (XPlan) [3] uses an SR template to create machine processible experiments that are suitable for laboratory execution. XPlan uses this to constrain its search for an Experiment Design, which in turn describes the expected measurements of each aliquot in the experiment. XPlan dispatches this with a set of Lab Parameters, instructing the lab how to configure and run the experiment. XPlan decides not only how to allocate samples to physical containers, but also which samples to use. For example, XPlan will choose the number of replicates for the strains in the second row of the ER in Table 1 based upon the available containers.

Laboratory Execution: RT submits experiments to the Strateos cloud laboratory for automated execution. Here, RT selects from one of several Strateos experimental protocols, such as growth curves and time series. In these protocols, Strateos measures samples with a plate reader and flow cytometer over several time points, including multiple induction and dilution steps, and returns both raw measurement data and protocol execution traces. In future work, the RT will also interface with laboratories via Aquarium[1].

Metadata Validation and ETL: The SR Generator validates data products by aligning metadata descriptions with expected data. It flags and explains any discrepancies to the experimentalist and lab technicians. If able to successfully match the data, the RT performs a series of ETL steps that summarize results for the experimentalist, organized in terms of the metadata on sample contents, conditions, and context.

3 VALIDATION

Over a four month period, we applied RT to process and execute twenty three ERs, totaling fifty nine 96-well plates of samples and approximately 10 measurements per well. The ERs span three distinct experimental protocols. With RT, we can plan and attach metadata to experimental samples within approximately four hours (not accounting for experiment execution time), whereas before it took approximately three weeks to attach metadata to six 96-well plates worth of data. This has allowed us to reduce laboratory idle time (due to dependent experiments) from several weeks to a few days.

ACKNOWLEDGEMENTS

The authors acknowledge Ben Keller, Peter Lee, and Narendra Maheshri for initial development of ER documents.

REFERENCES

[1] Beal, J., et al. Collaborative terminology: SBOL project dictionary. In *International Workshop on Bio-Design Automation (IWBDA)* (2020).
[2] Jessop-Fabre, M. M., and Sonnenschein, N. Improving reproducibility in synthetic biology. *Frontiers in Bioeng. and Biotech. 7* (2019), 18.
[3] Kuter, U., et al. XPLAN: Experiment planning for synthetic biology. In *ICAPS Workshop on Hierarchical Planning* (2018).
[4] McLaughlin, J. A., et al. SynBioHub: A standards-enabled design repository for synthetic biology. *ACS Syn. Bio. 7*, 2 (2018), 682–688.
[5] Nguyen, T., et al. Intent parser: a tool for codifying experiment design. In *International Workshop on Bio-Design Automation (IWBDA)* (2020).
[6] Roehner, N., et al. Sharing structure and function in biological design with sbol 2.0. *ACS Syn. Bio. 5*, 6 (2016), 498–506.

[1] https://www.aquarium.bio