LabKey Biologics: Tracking lineage of biological entities, sample derivation, and structured data within a new biologics registration software platform

Ryan Luce, Ph.D. [1]; Kevin Krouse [1]; Josh Date [1]; Nick Arnold [1]; Chris Joosse [1]; Jessi Murray [1]; Frank Lee [1]; Randal R. Ketchem, Ph.D. [2]; J. Alex Taylor, Ph.D. [2]; Russell E. Williams [2]

[1] LabKey, 617 Eastlake Ave. E #400, Seattle, WA 98109 USA;

[2] Just Biotherapeutics, 401 Terry Ave N, Seattle, WA 98109 USA

ABSTRACT

In late 2015, LabKey began a collaboration with Just Biotherapeutics to create a tool focused on the data management needs of biologics research by leveraging and extending LabKey's existing open source biomedical data management platform, LabKey Server.

To support the needs of large molecule development, mechanisms were developed for tracking design lineage, sample derivation, and structured data related to proteins and a powerful querying engine was built that allows any sample, regardless of how many sample derivations it may have undergone, to be quickly connected back to related entities. In March 2017, LabKey released *LabKey Biologics* as a standalone application.

CHALLENGE

Traditional methods of data capture lack the structure, relationships, and queryability scientists need to effectively explore their data.

OBJECTIVES

1. Capture structured information about biological entities (sequences, cell lines, expression systems, etc.) and their relationships with one another that ensures uniqueness and allows for querying of experimental data.
2. Easy capture of sample lineage as well as entity design lineage, allowing for the connection of assay data to related biological entities and media recipes / batches.
3. Creation of performant queries that track lineage or other sample information, allowing the joining of assay data based on sample lineage and the viewing of multiple assay types simultaneously.

BIOLOGICAL ENTITY REGISTRATION

Expanding on the foundation of LabKey Server, the first step was to develop a new object type called "DataClass". DataClasses were added as a new object-type within LabKey Biologics, allowing for the representation of virtual entity definitions. This allowed for tracking of lineage, development of specific fields around entities, and definition of relationships between entity types (see below).
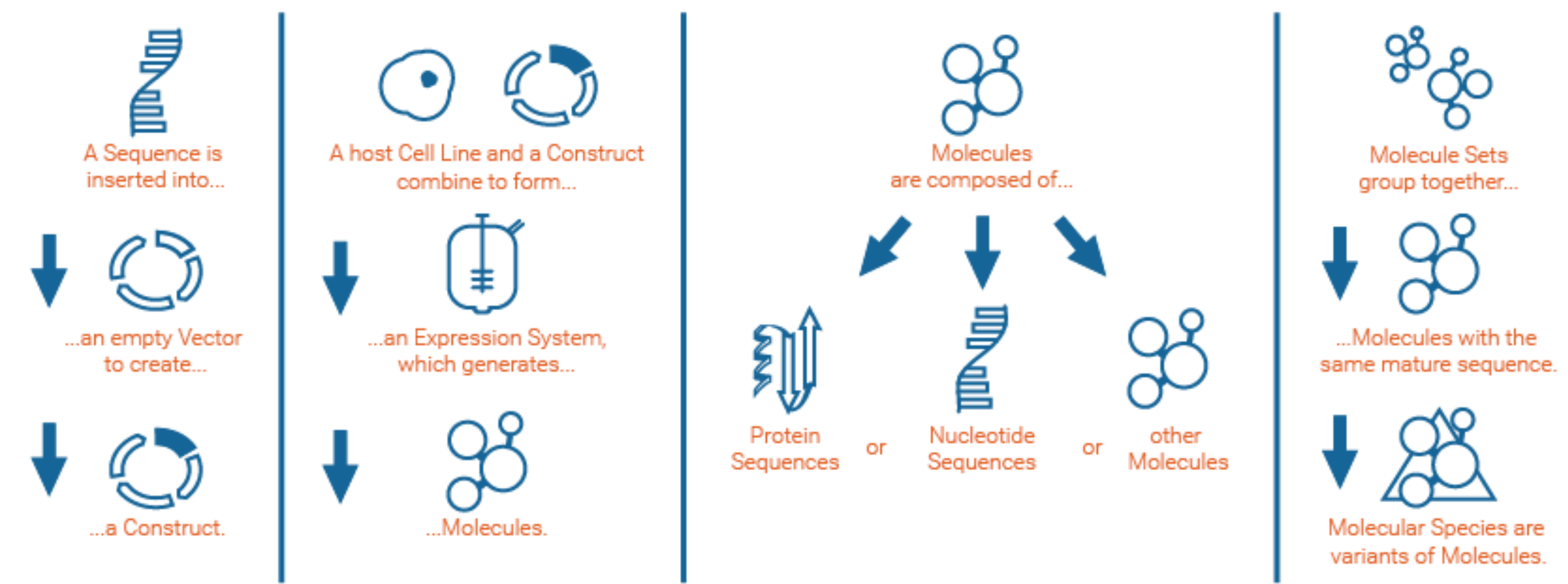
SAMPLE DERIVATION AND LINEAGE

The querying engine underlying LabKey Biologics is dependent upon the lineage of DataClasses as well as physical samples being tracked within the system. To accommodate this, we built functionality within LabKey Biologics to create sample IDs, either singly or in bulk, that are connected to "parents" which may be comprised of a variety of types. For example, a bioreactor run would be one sample type. It would have a parent of DataClass "Expression System" as well as a "Media Batch". The "Media Batch" would then be connected to its "Media Recipe".

Example: Sample ID Generation

ENTITY DATA WITH LINEAGE

Example: Protein Sequence with one "child" sequence and nine "grandchild" sequences

Protein sequence information about lineage, aliases and description, and automatically calculated physical characteristics.

It also has sequences annotations as well as links and relationships to other biological entities.

QUERYING ASSAY DATA WITH LINEAGE
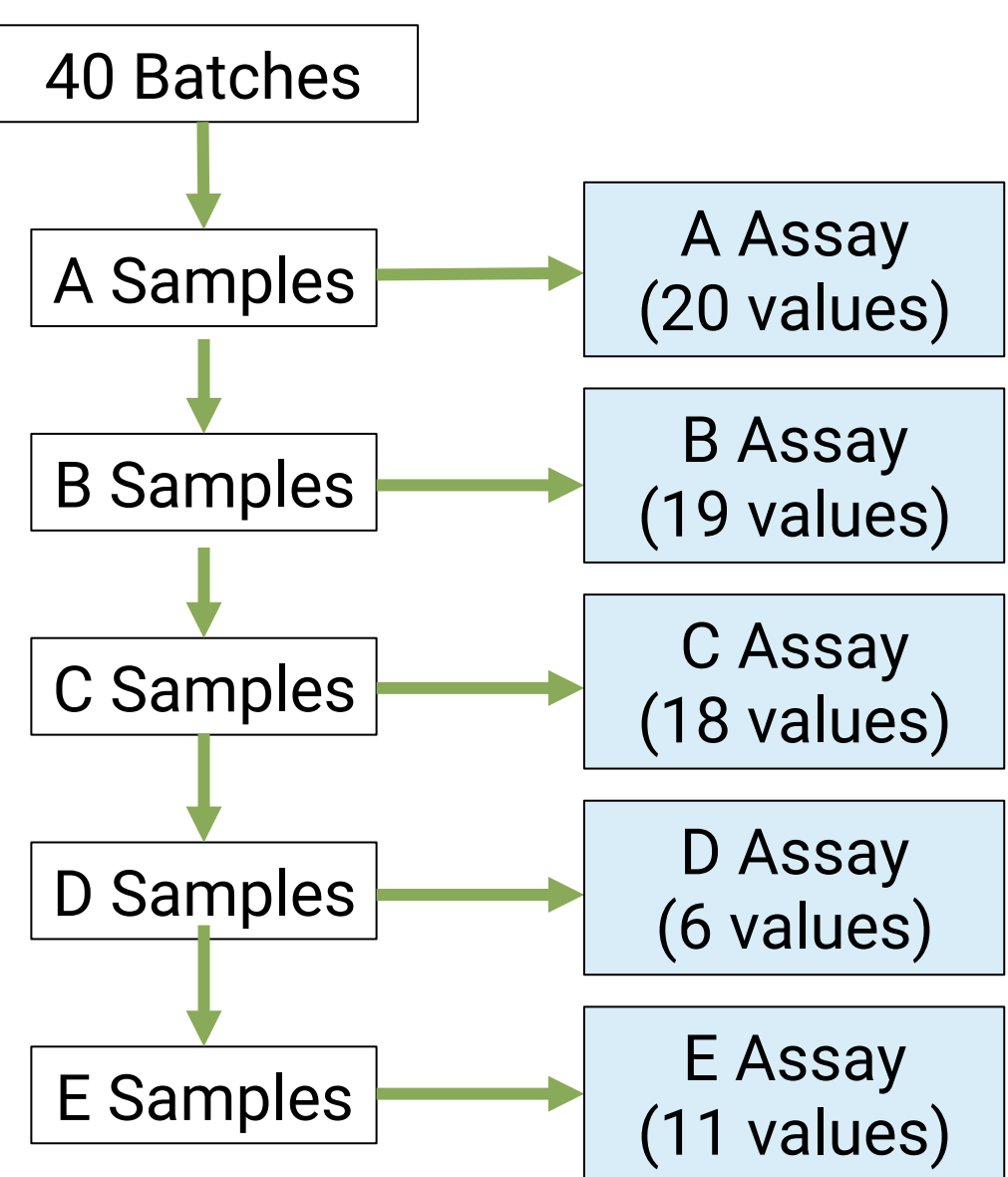
Mock Data Set
- 40 Batches of Expression Runs
- 200 Derivative Samples (40 A Samples, 40 B Samples, Etc.)
- 74 Assay Measures

Query Development
- Outer join assay data measurements based on sampleID.materialinput.batchID
- Columns for Batch ID, Expression System, 74 Measure Columns
- Query Time – 1.829 sec

Batch ID	Expression System	A Meas01	A Meas02	E Meas10	E Meas11
Batch01	ES-101	xxx	xxx	xxx	xxx
Batch02	ES-102	xxx	xxx	xxx	xxx
Batch 39	ES-103	xxx	xxx	xxx	xxx
Batch 40	ES-103	xxx	xxx	xxx	xxx

40 Batch IDs by 74 Assay Measures

LABKEY AND JUST BIO COLLABORATION

Founded in 2003, LabKey is a software and professional services provider that specializes in helping organizations overcome the unique data management, collaboration, and workflow challenges faced in the scientific research environment.

Just Biotherapeutics (founded in 2014) is led by an experienced team in the fields of protein, process and manufacturing sciences. The Just team came together to solve the scientific and technical hurdles that block access to life changing protein therapeutics; from the design of therapeutic molecules to the design of the manufacturing plants used to produce them.

LabKey and Just signed a multi-year agreement in 2015 to develop LabKey Biologics, a product built on the foundation of LabKey Server, to help biotechnology R&D teams optimize the development process through structured data capture. www.labkey.com/biologics