



# Managing Sample-to-Answer Multi-omics Workflows

Informatics in Context –  
USDA Case Study



We often characterize scientific data analysis and management as pure engineering challenges that can be solved through technology.

Science is a collaborative, social process *accelerated* by technology.

# People + process drive Technology

*Regardless of the science, understanding the people, process, and technology is essential to developing comprehensive solutions.*

## People

- Who is doing the work? Who is *requesting* the work? Who needs the results?

## Process

- How do samples move through workflows?
- What types of measurements are needed? What types of data is collected? What analysis is required?

## Technology

- What options are available, what options make sense given the people and process? How do the pieces fit together to form a cohesive solution?

## Case Study

# U.S. Meat animal research center USDA-ARS



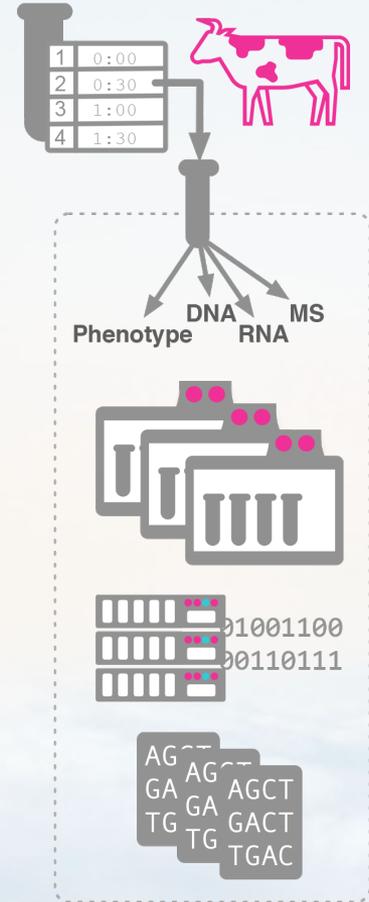
# USMARC multi-omics workflows

- Mammalian and microbial genomics on a wide range of food producing species
- NGS, microarrays, mass spectrometry, et al
- Combined to form complex “multi-omics” workflows
- Constantly evolving new methods
- Informatics system manages and track *everything*



# Example: bioreactor workflow

1. Animal samples are collected in the field
2. A specimen is placed in bioreactor, samples taken at multiple time points
3. Each sample is characterized by phenotype, structure (proteomics via mass spec), and sequenced
4. All time points are collated and summarized
5. Each step generates wet lab artifacts (new samples) and data artifacts
6. Results are available for manual downstream analysis



# People

## Farmers

- Manage the animals and collect samples

## Scientists

- Design and run experiments
- May handle sample and library prep
- Perform analysis using a mix of open source and commercial tools

## Lab Technicians

- Handle most of the wet lab work related to the instruments

## IT

- Strong on-site hardware IT team with HPC expertise
- No on-site software development capabilities

## PEOPLE

FARMERS



SCIENTISTS



LAB TECHS



IT



# Process

## Samples

- Samples sourced on-site and from partners
- Track farm, lab, and analysis data
- Complex parent/child relationships

## Workflows

- Standard methods (phenotyping, assembly, RNA-Seq, etc)
- Variations and new methods



## Instruments

- Multiple vendors – PacBio, Illumina, Biolog, Waters

## Analysis

- Vendor tools
- Pipelines via Galaxy and Lab7 ESP
- Manual analysis using R and Matlab

## PEOPLE

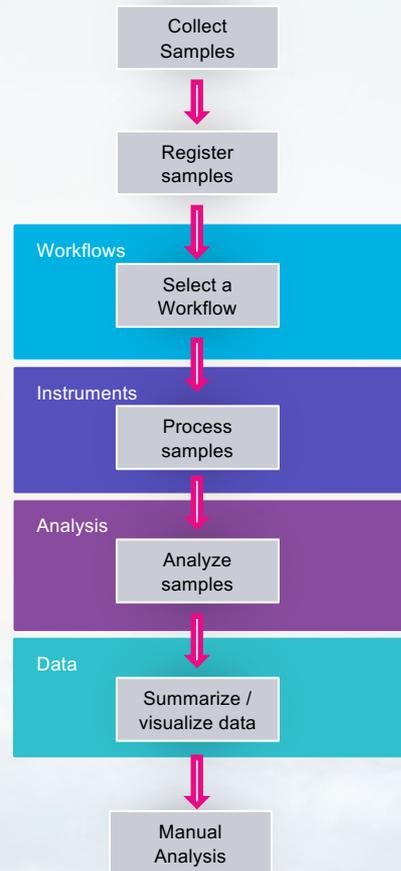
FARMERS

LAB TECHS

SCIENTISTS

IT

## PROCESS



# Technology

## Sample Management

- Animal Database for field data collection
- Lab7 ESP Sample Registry for full sample lifecycle tracking

## Workflow Management

- “Small” workflows managed in ESP are chained to form more complex workflows, allowing careful process tracking while maintaining operational flexibility

## Instrument Integration

- Mix of API calls, file system monitoring, and run sheet generation ties workflows directly to instruments, with a goal of no manual data entry
- Vendor support is key to seamless integration

## Analysis

- Standard analysis integrated directly into workflows when possible (e.g., assembly and annotation pipeline)
- Data made available for off-line analysis
- CLI tools enable data provenance outside of ESP

### PEOPLE

FARMERS

SCIENTISTS

LAB TECHS

IT

### PROCESS

Collect Samples

Register samples

Workflows

Select a Workflow

Instruments

Process samples

Analysis

Analyze samples

Data

Summarize / visualize data

Manual Analysis

### TECHNOLOGY

Field Database

Sample Registry

Workflow Library

Instrument APIs

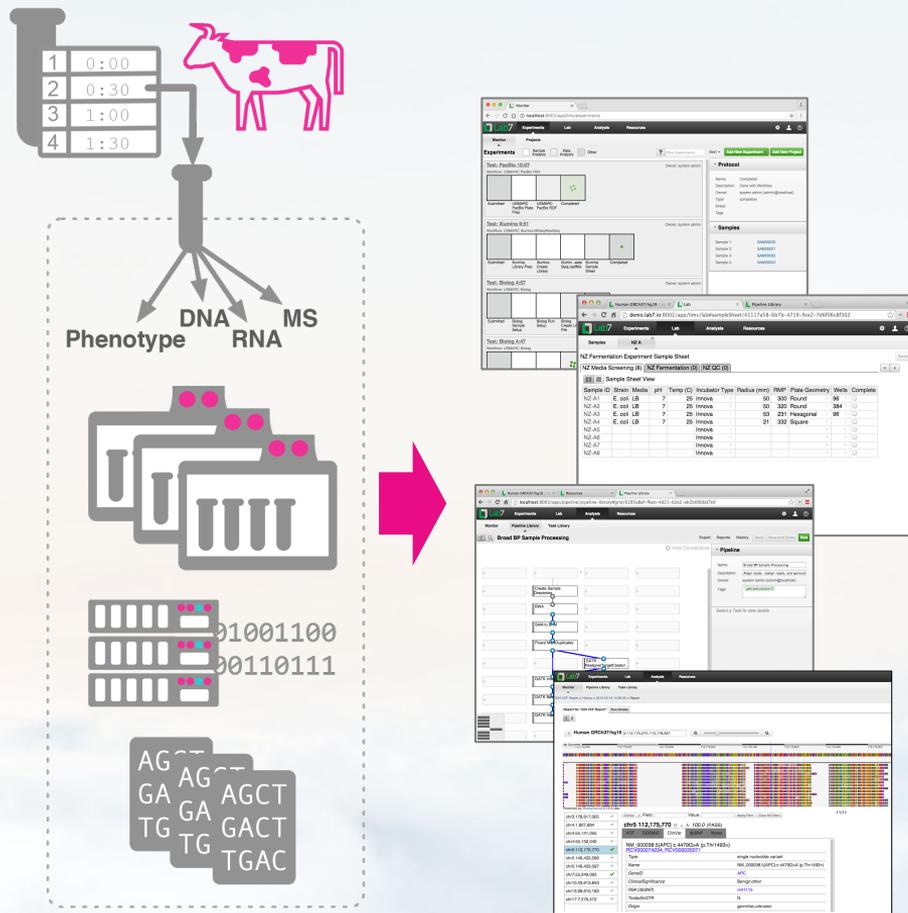
Pipeline Engine(s)

File System(s)

Analysis Tools

# Results

- Sample-to-answer workflow management
- Support stand-alone workflows and complex, integrated workflows
- Agnostic towards assays, instruments, compute resources, and analysis methods
- Leverages technology appropriate for the target users



# Case Study: USMARC Summary

## Demand for ESP:

### Multiomic Workflows

- Mammalian and microbial genomics on wide range of food producing species
- NGS, microarrays, mass spectrometry
- Constantly evolving new methods
- Informatics system manages and track everything

### Disparate Stakeholders

- Farmers
- Scientists
- Lab Techs
- IT

### Outdated LIMS Just Not Keeping Up

- Database at capacity
- Not designed for data-intensive workflows
- Difficult to configure for new technologies
- Not integrated with bioinformatic analysis

## Solution: connect people + processes + technology with ESP

### Sample Management

- Animal Database for field data collection
- ESP Sample Registry for full sample lifecycle tracking

### Workflow Management

- "Small" workflows managed in ESP are chained to form more complex workflows, allowing careful process tracking while maintaining operational flexibility

### Instrument Integration

- Mix of API calls, file system monitoring, and run sheet generation ties workflows directly to instruments, with a goal of no manual data entry
- NGS Sequencers: *Illumina NextSeq & MiSeq, PacBio RSII & Sequel*
- Other equipment: *Biolog OmniLog, Waters Mass Spec*

### Analysis

- Standard analysis integrated directly into workflows when possible (e.g., assembly and annotation pipeline)
- Data made available for off-line analysis
- CLI tools enable data provenance outside of ESP

## PEOPLE

FARMERS

LAB TECHS

SCIENTISTS

IT



## PROCESS + TECHNOLOGY

