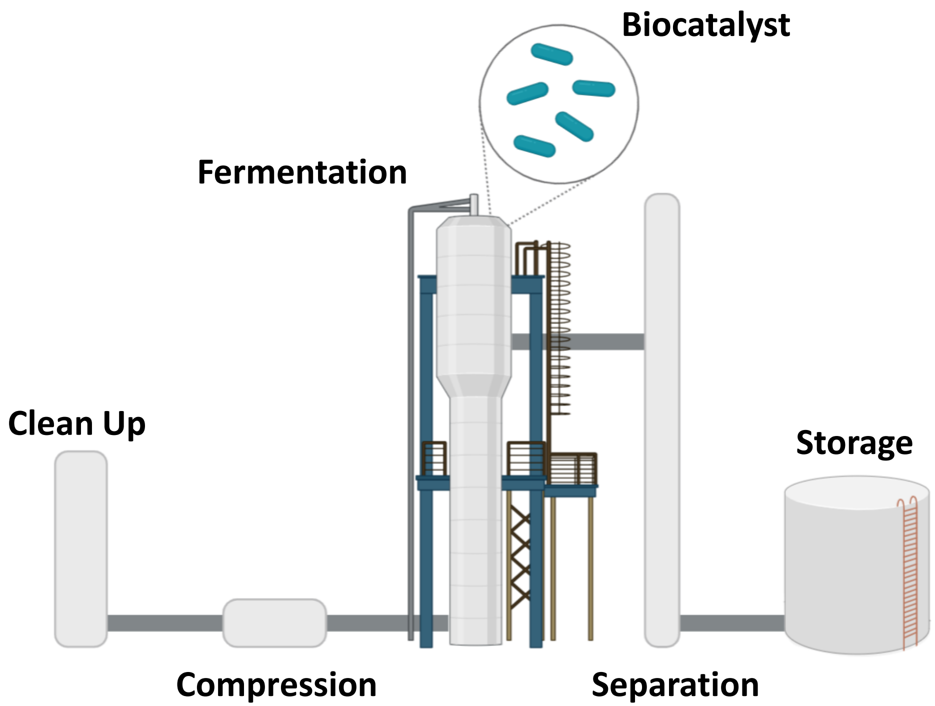
DOE Bioenergy Technologies Office (BETO) 2021 Project Peer Review

Data Integration and Deep Learning for Continuous Gas Fermentation Process Optimization

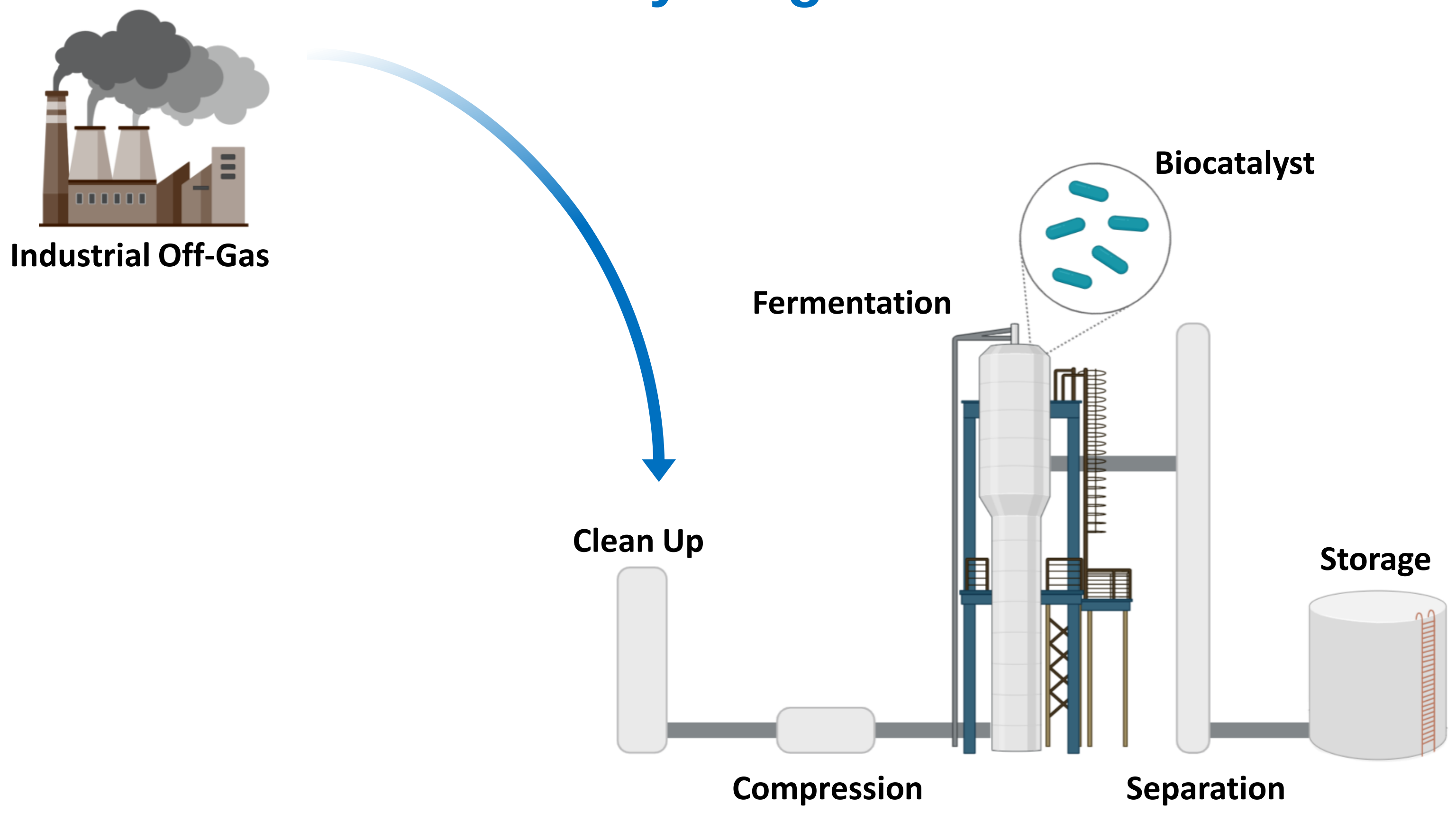
3.10.2021

Agile BioFoundry
Wayne Mitchell
Lanzatech

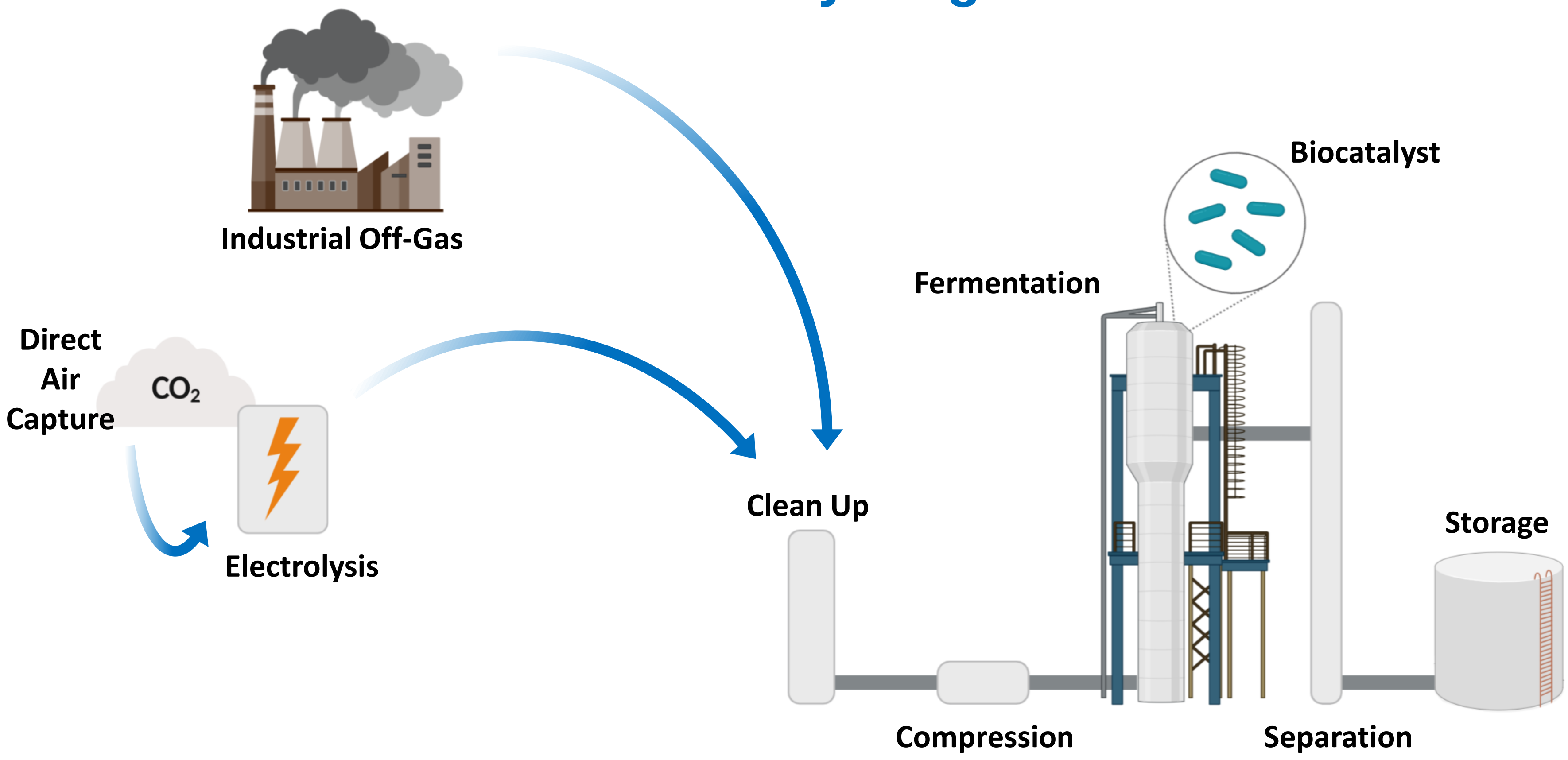
Recycling Waste Carbon



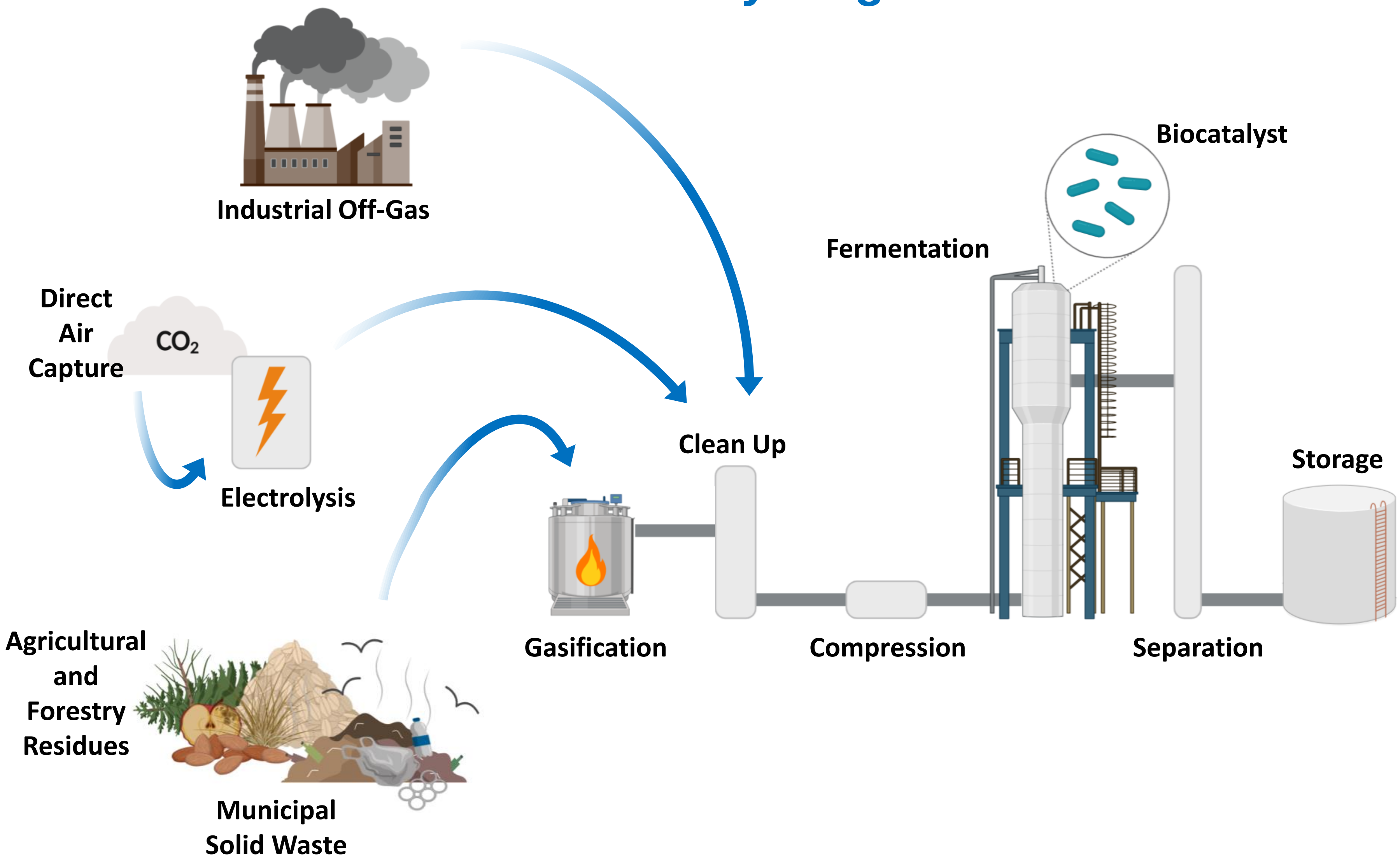
Recycling Waste Carbon



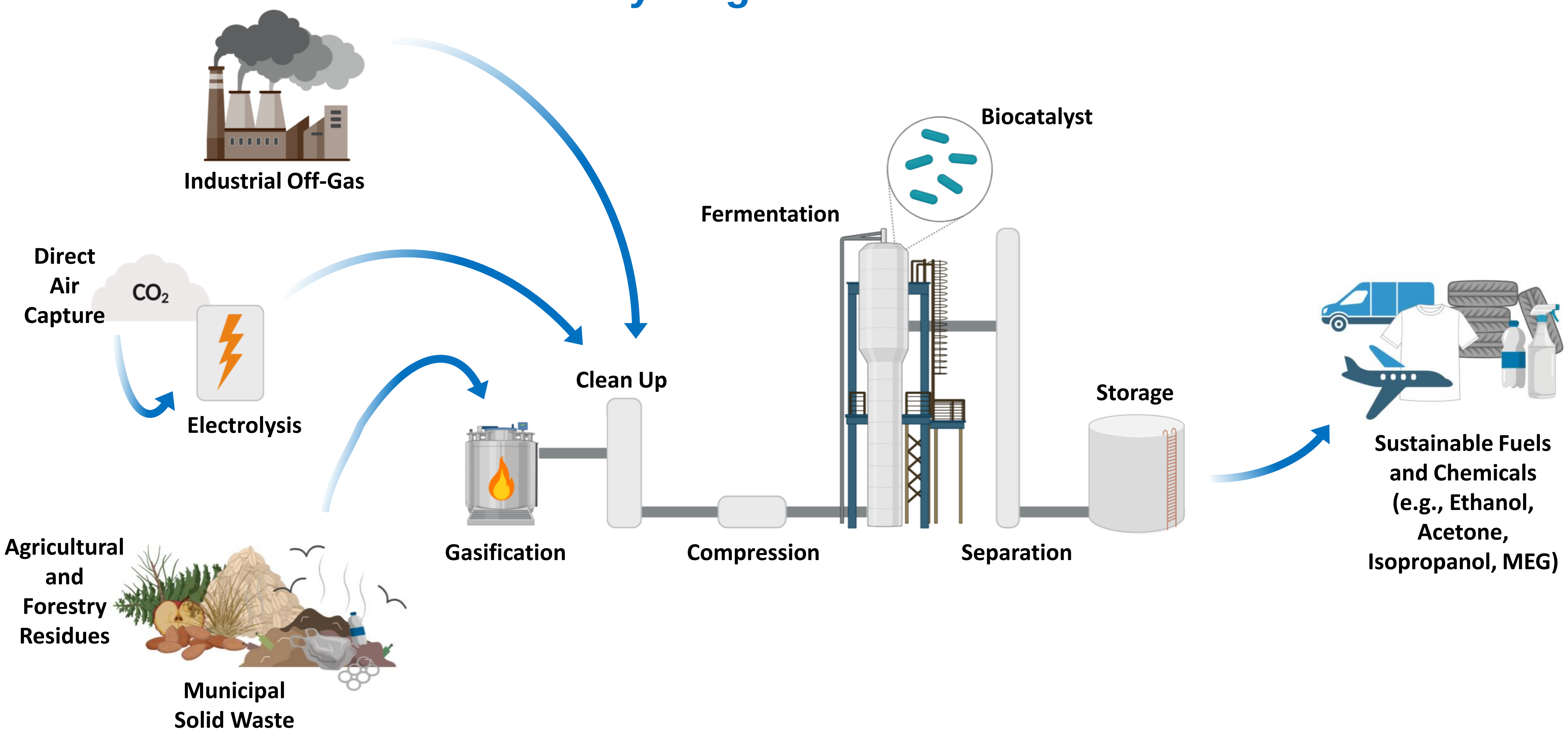
Recycling Waste Carbon



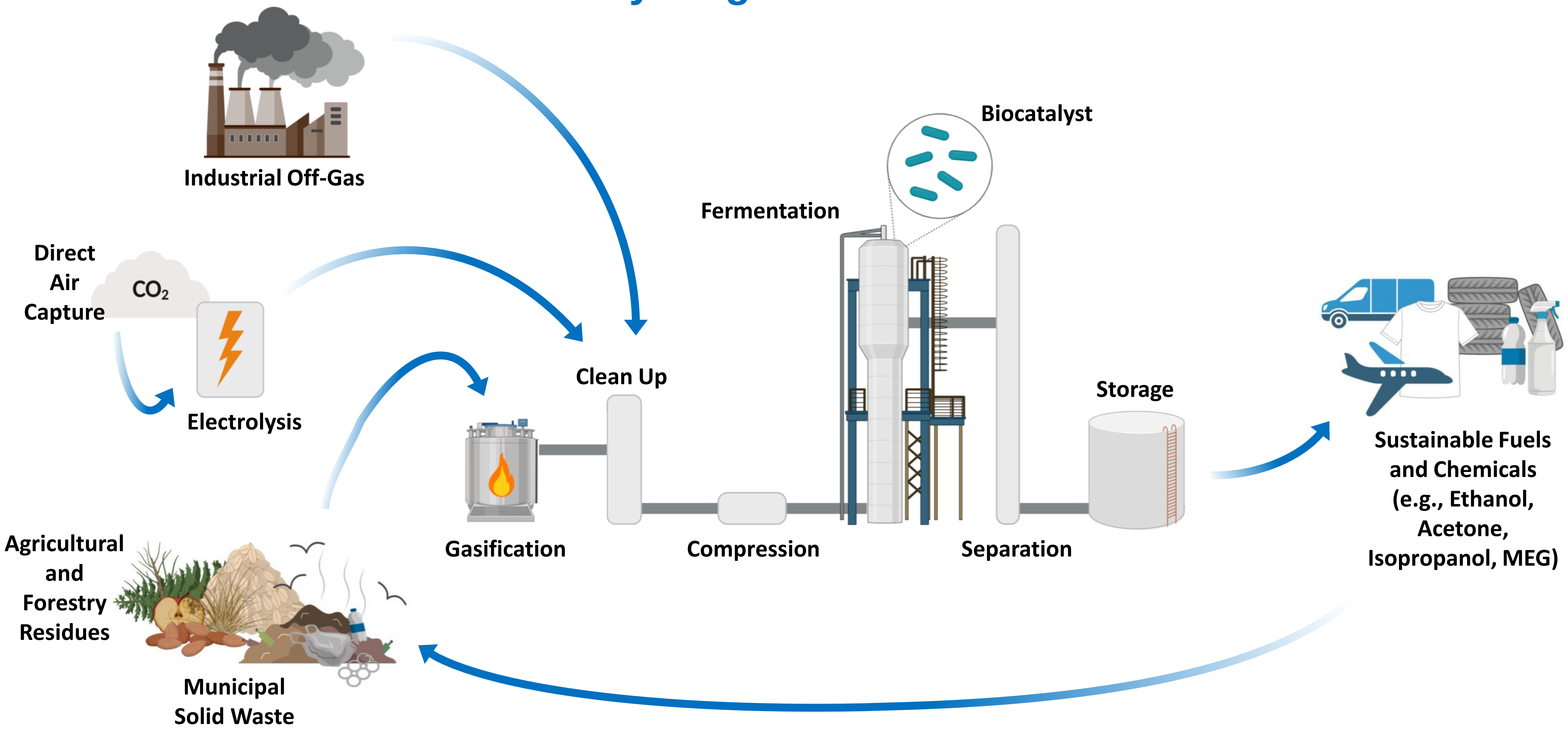
Recycling Waste Carbon



Recycling Waste Carbon



Recycling Waste Carbon



Versatile platform for converting any waste carbon to products

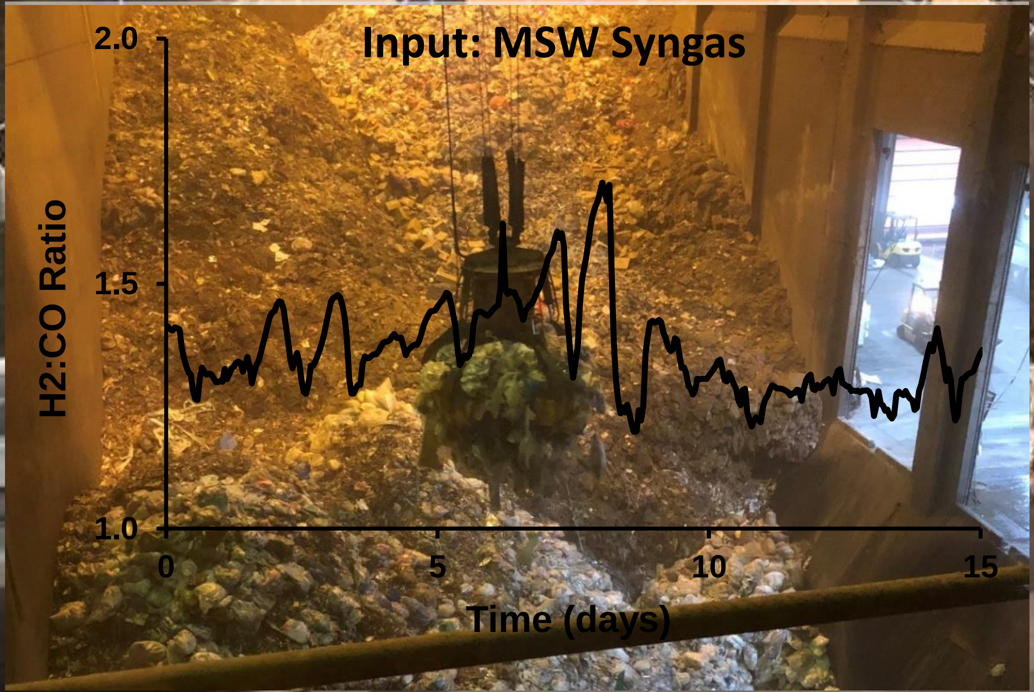
Operating Commercially Today

>20M Gallons Ethanol Produced

>100,000 tonnes CO_2 Avoided

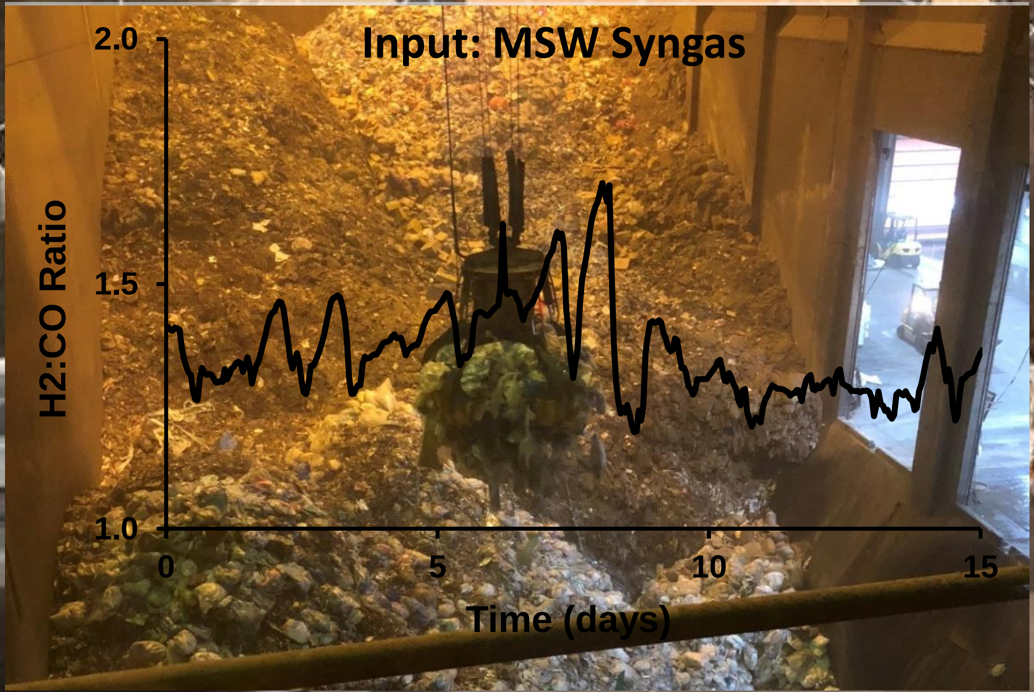
Biology Is Capable Of Processing Chaotic Inputs

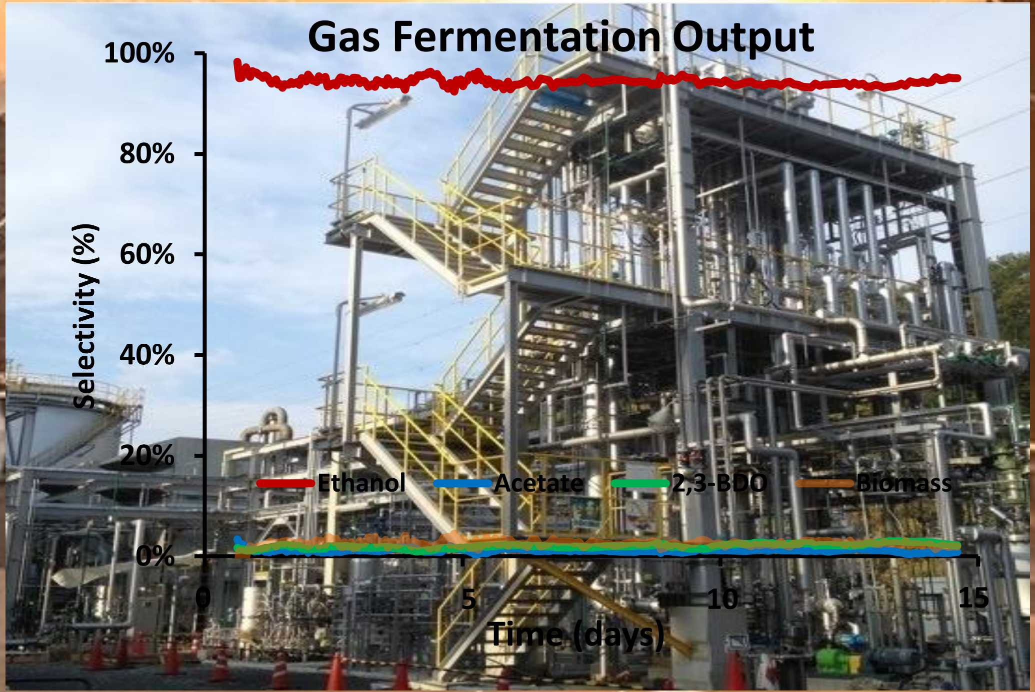
SEKISUI



Biology Is Capable Of Processing Chaotic Inputs

SEKISUI





CarbonSmart™

Köpke & Simpson (2020) *Curr Opin Biotechnol* 65: 180 189.

LanzaTech

1 – Management

Dr. Wayne Mitchell
Director,
Computational Biology

Dr. Phil Laible
Principal Investigator,
Biophysics

Dr. Gregg Beckham
Coordinating National Lab PI,
Group Leader

Asela Dassanyake
Manager,
Scientific Computing

Dr. Peter Larsen
Computational Biologist

Dr. Violeta Sanchez i Nogue
Researcher
Chemical Engineering

Project Goals are Are Well-aligned with Agile BioFoundry's Mission

Agile BioFoundry

√ • cut bioprocess scale-up time in half.

√ • develop and deploy technologies that enable commercially relevant biomanufacturing of a wide range of bioproducts by both new and established industrial hosts.

√ • adopt new biomanufacturing methods … in industrially relevant host microbes for production of commodity chemicals and biofuels.

√ • increase U.S. industrial competitiveness and creates new opportunities for private sector growth and jobs

https://agilebiofoundry.org/about/

Project Tasks

Task	Name	Status
1	Data collection/formatting	Complete
2	Metabolic Model Output	Complete
3	Artificial intelligence models generated/integrated	Complete
4	Delivery of functioning artificial intelligence models	Complete
5	Define and refine key questions that can be addressed	Complete
6	Downselection of bioreactor variables for statistical models	Complete
7	Biological-insight-driven variable selection	Complete
8	Changes to metabolic and environmental sampling	80%
9	Fulfill reporting requirements	Complete
10	Deep Learning Training	Complete
11	Modeling data uploaded into ABF EDD	50%
12	Open source decisioning	50%
13	Final Report	0%

Single-phase project without go / no-go check points.

Project Risk

Risk

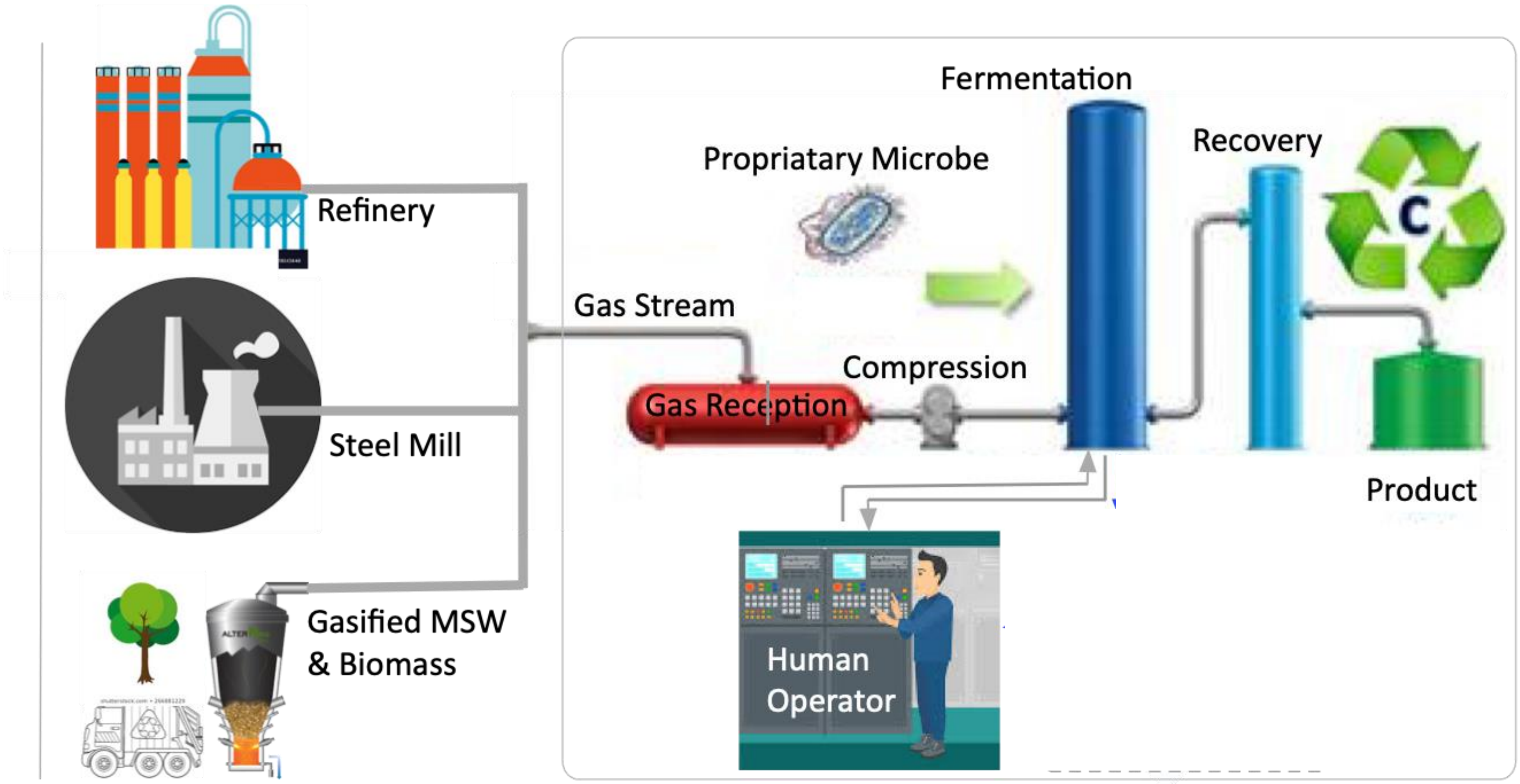
- Unable to reorganize the large transactional databases populated over many years and process scales into schema amenable to deep learning.
- Historical data may prove intractable
- Maybe insufficient data after curation

Mitigation

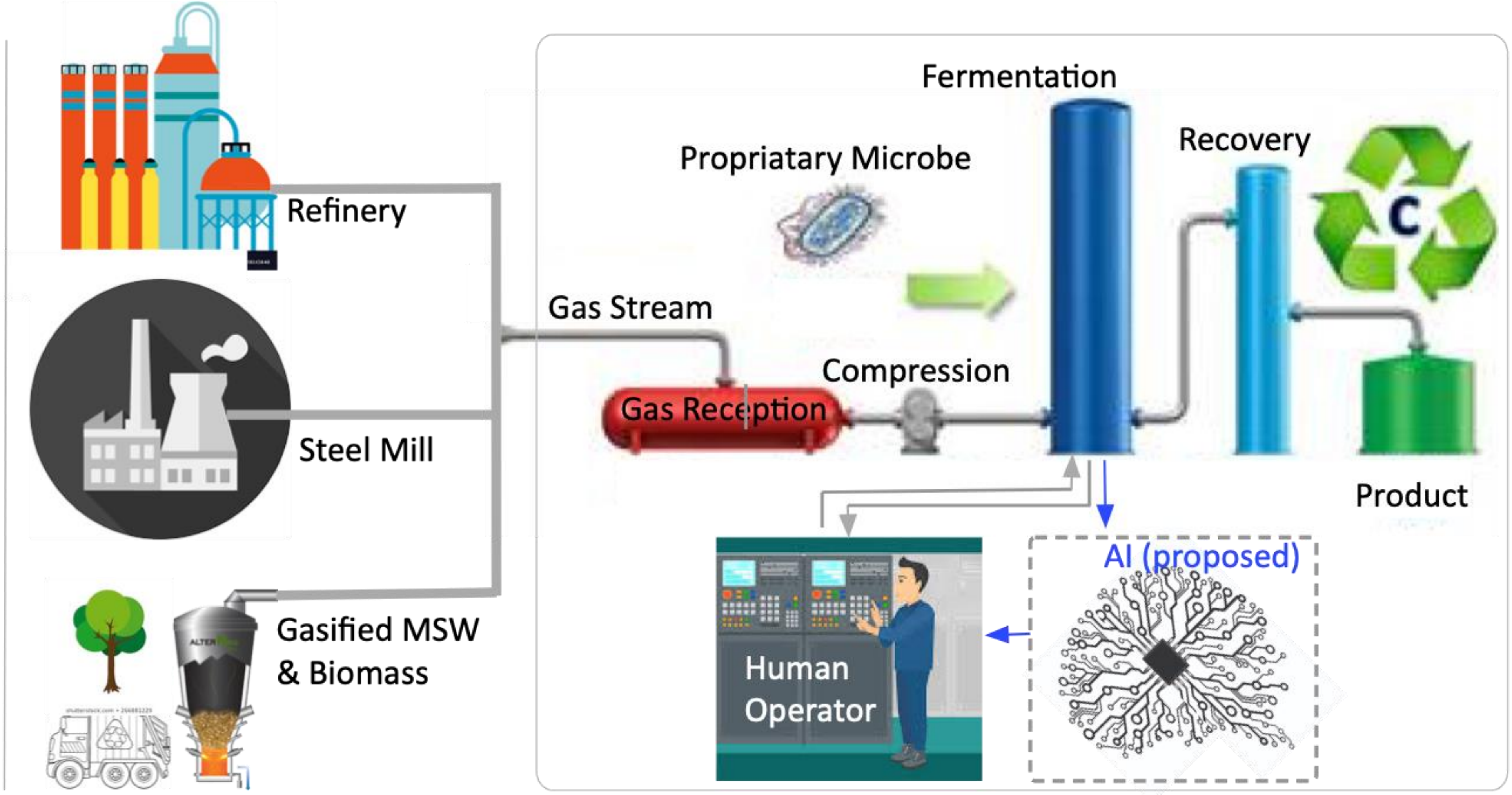
- Overcome through regular discussions and face-to-face meetings with database specialists within LanzaTech and modelers at ANL.
- Use high volume of new data acquisition to reconfigure data schema and collection to enable machine learning
- Enhance data sets by imputation and by injecting anonymized data from Genome Scale Models.

2 – Approach

Real-time AI Can Provide Fermentation Surveillance and Operator Guidance

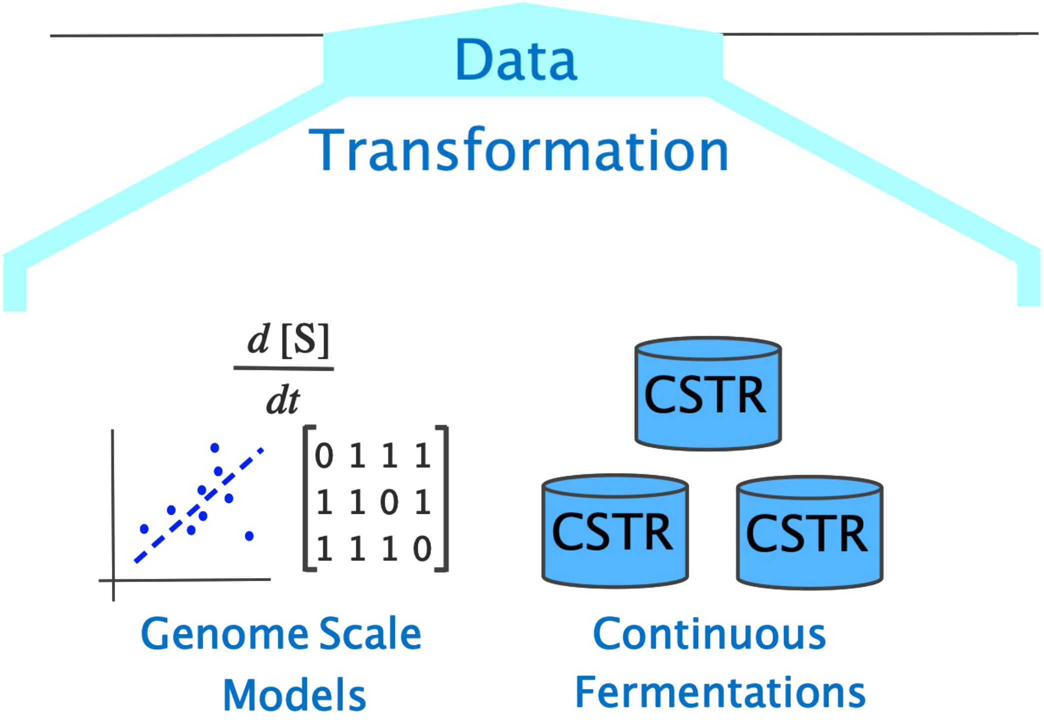


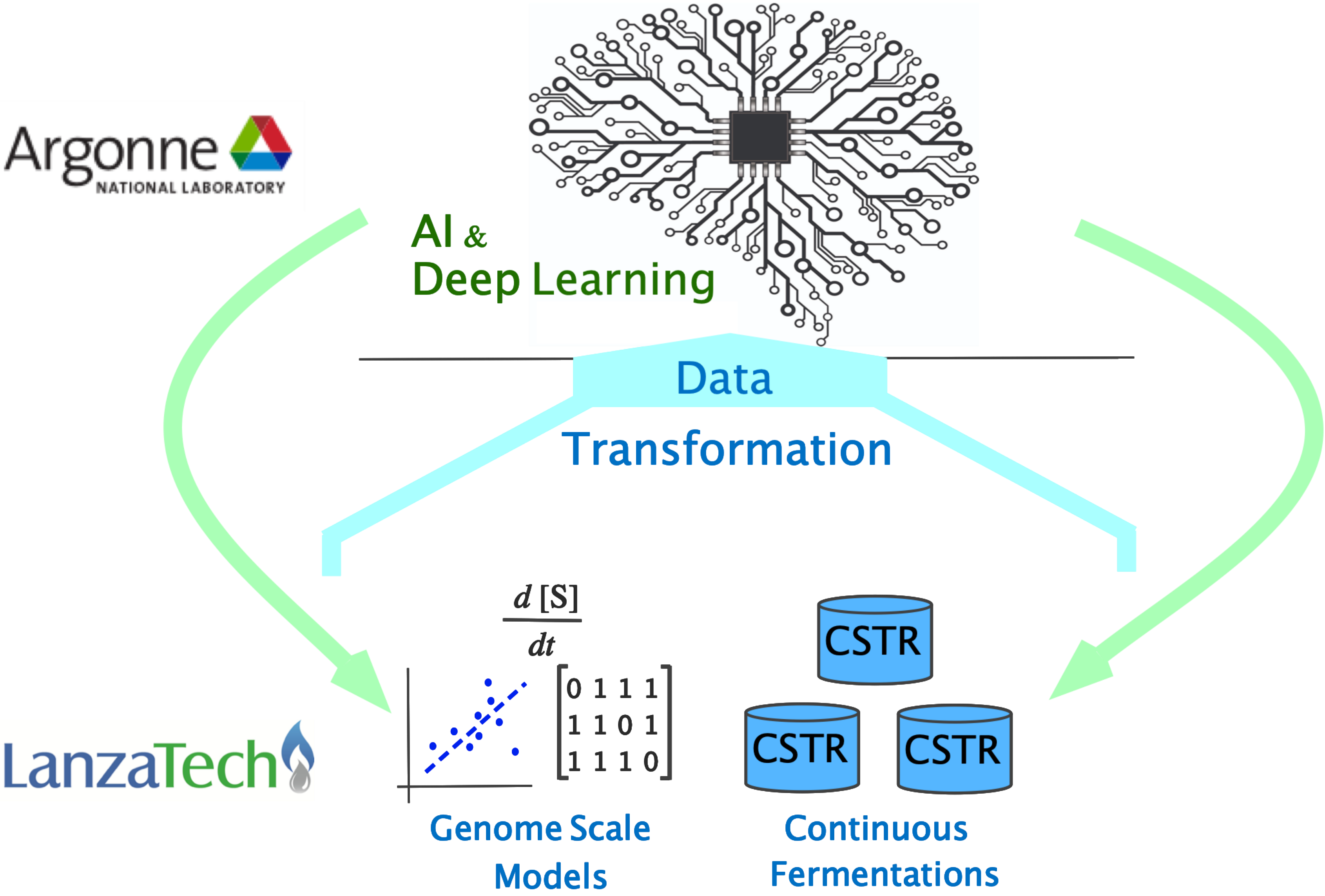
Real-time AI Can Provide Fermentation Surveillance and Operator Guidance



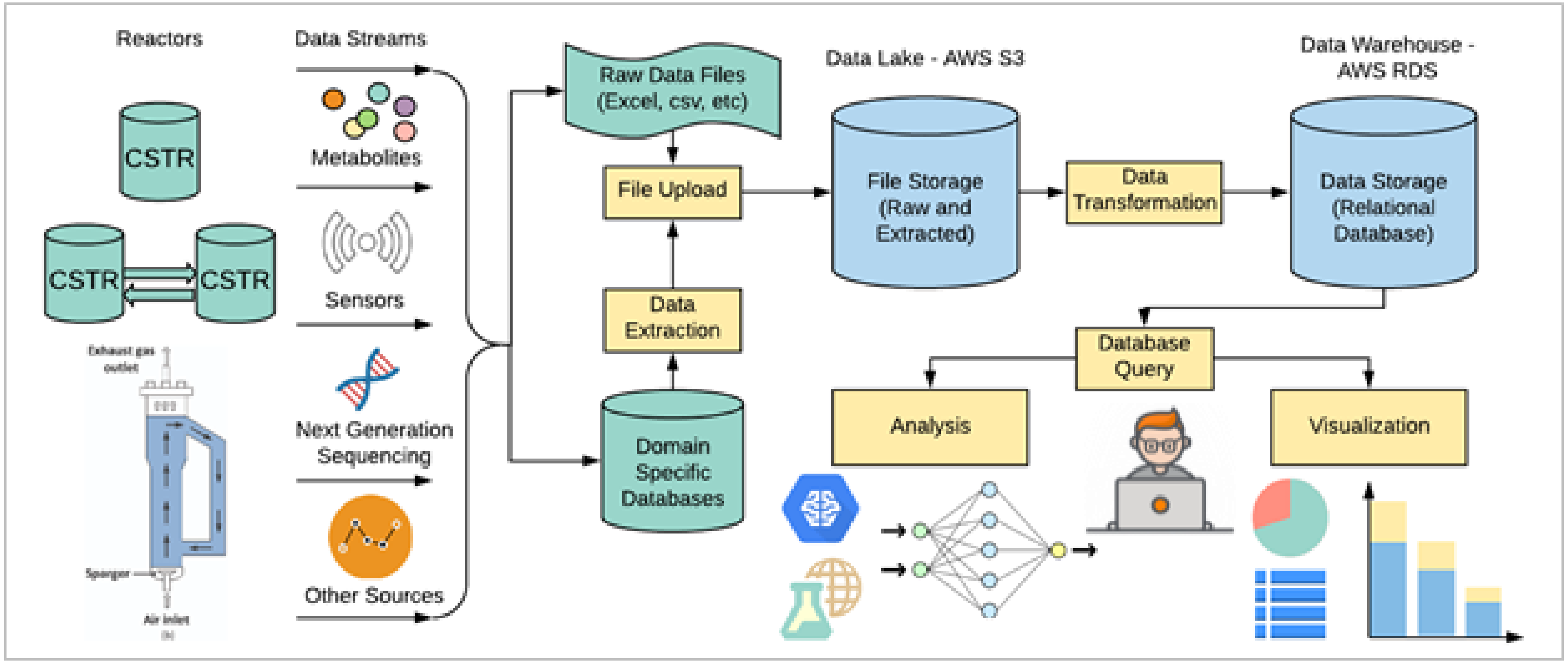
Argonne NATIONAL LABORATORY

LanzaTech

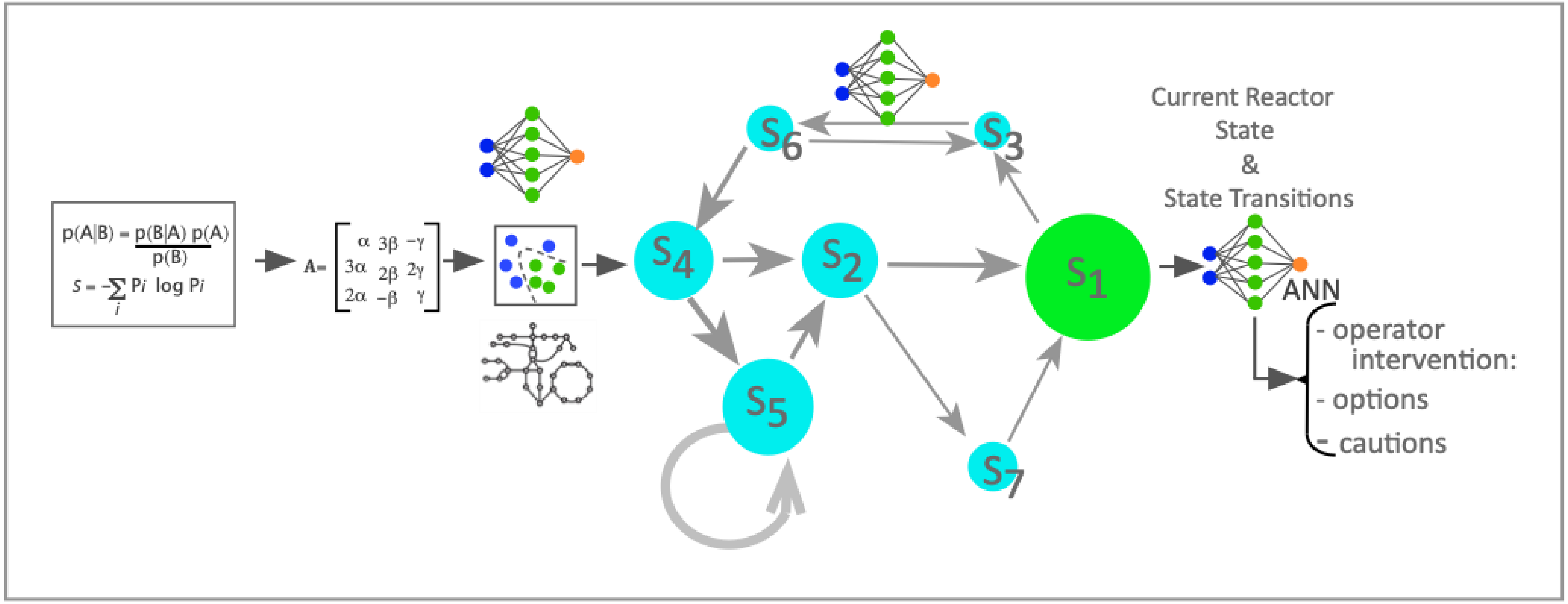



Argonne NATIONAL LABORATORY
AI & Deep Learning
Data
Transformation
$\frac{d\,[S]}{dt}$
$\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}$
CSTR
CSTR
CSTR
LanzaTech
Genome Scale Models
Continuous Fermentations

Data Capture, Curation, GEM Modelling
&
Steady State Annotation



Data imputation, Parameter tuning
&
Deep Learning



3 – Impact

Impact: Technical Innovations

- Analysis of unique pan-scale datasets with new, advanced methodologies brings the opportunity to make predictive models that can balance the multiple – and potentially competing – fermentation outcomes in a single, multi-layered model

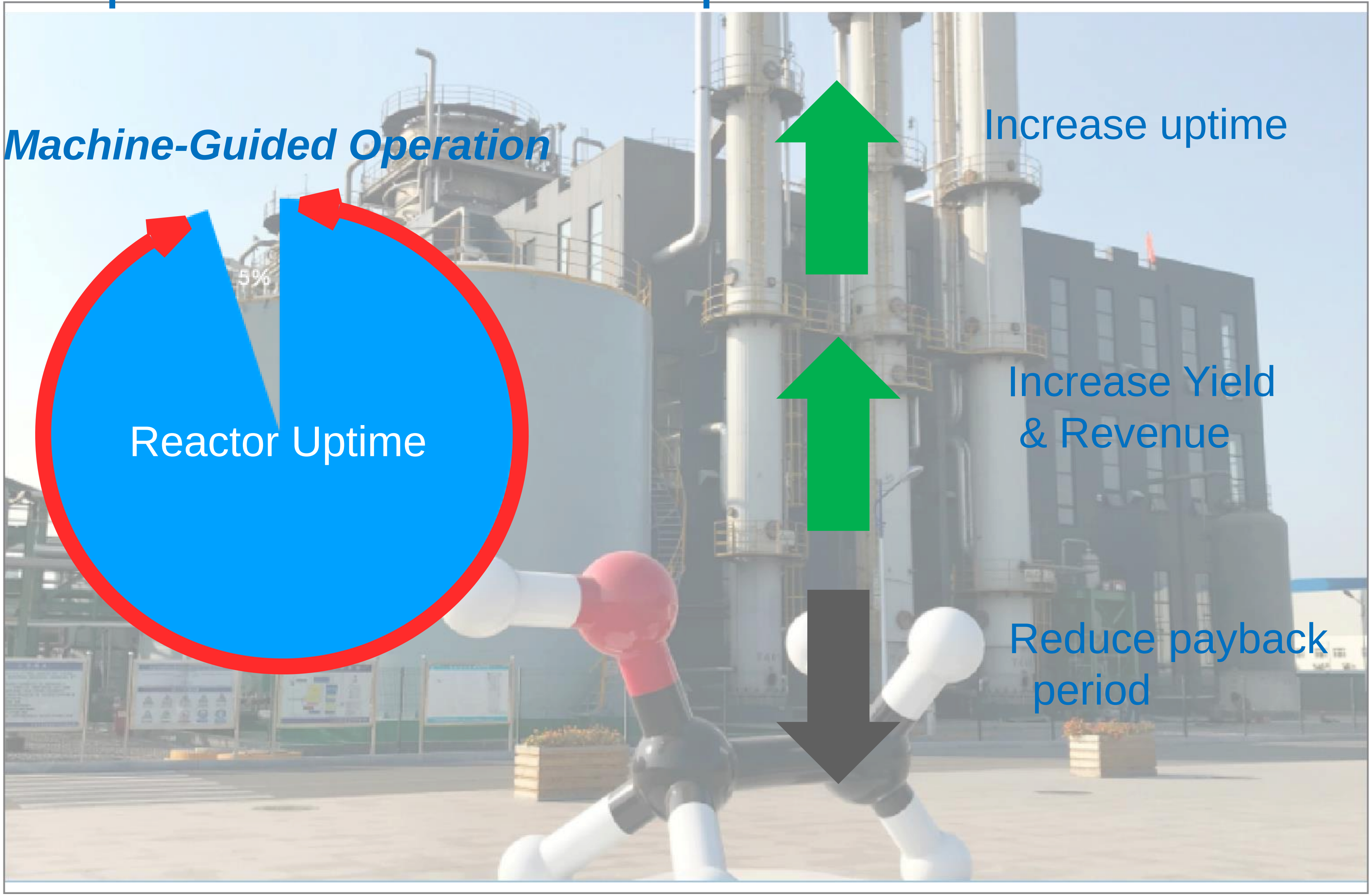
- Development of a tool to direct hypothesis-driven biological experiments and actively improve fermentation yields.

- No other system of this kind is in existence, to our knowledge.

  - AI models anonymized and saved as tables of data observations and model outputs (to **EDD**)
  - Deposited models (as R code) to **GitHub** and **JupyterHub** of the ABF

Potential Impact: Extend Run Time for Improved Economics

Potential Impact: Extend Run Time for Improved Economics



Potential Impact: Reduce Minimum Requisite Operator Skillset

The Present

Human Expert

Potential Impact: Reduce Minimum Requisite Operator Skillset

The Present

Human Expert

The Goal

AI assisted
Human Technician

Potential Impact: Reduce Minimum Requisite Operator Skillset

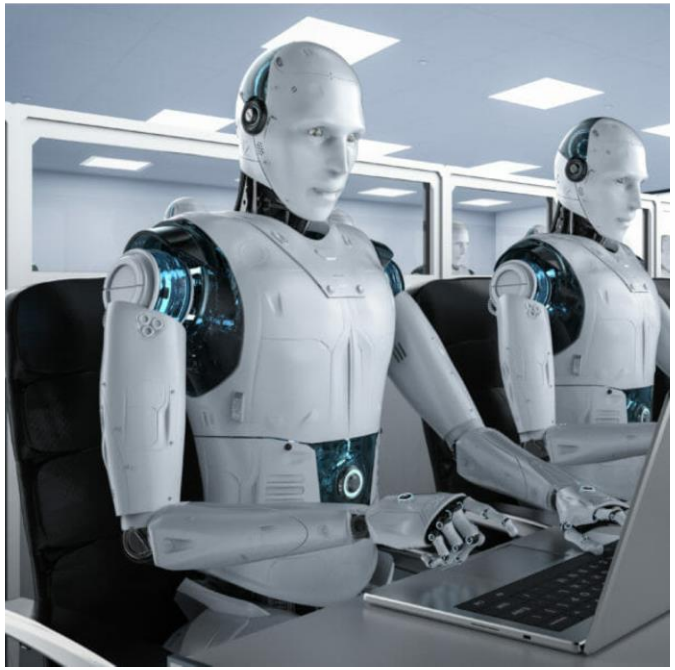
The Present

Human Expert

The Goal

AI assisted
Human Technician

The Future ?



Autonomous AI

Potential Impact: Simulation-based Technician Training

Potential Impact: Simulation-based Technician Training

4 – Progress and Outcomes

Progress and Outcomes: Industry Data Preparation and Transfer

- Built out the data catchment and warehouse
- Established methods for data curation
- ***Transferred curated data to Argonne***
- Established method to create anonymized Genome Scale Model output
- ***Transferred anonymized Genome Scale Model output to Argonne***

Data preparation, curation, and anonymization is essential for industry-laboratory collaboration

Progress and Outcomes: Adaptation of Data

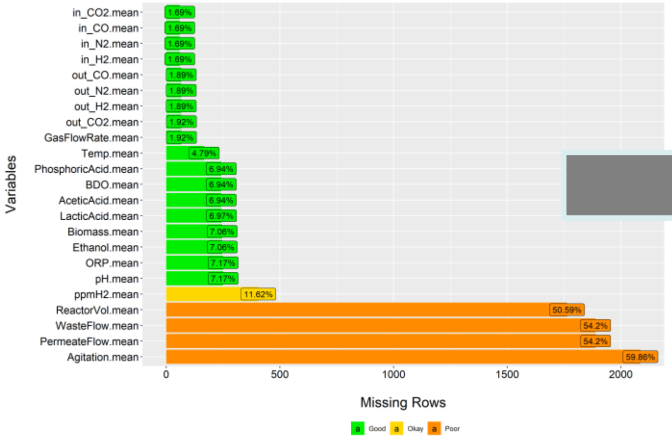
Adapt LT Data for ANL AI Approaches

Challenge: Available data, collected across multiple instruments and experiments needed to be re-formatted. Multiple periodicities of data collected (from seconds to days) and missing values

Solutions:

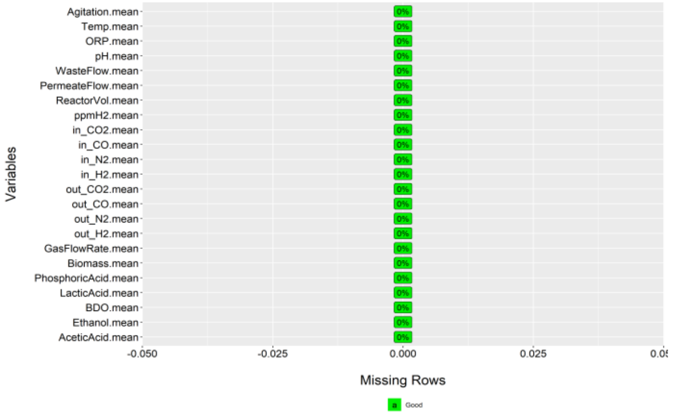
- Used a ‘rolling window’ approach to collect all data into uniform frequency.
- Applied statistical tools to interpolate missing data

Before Imputing



Kalman Smoothing

After Imputing

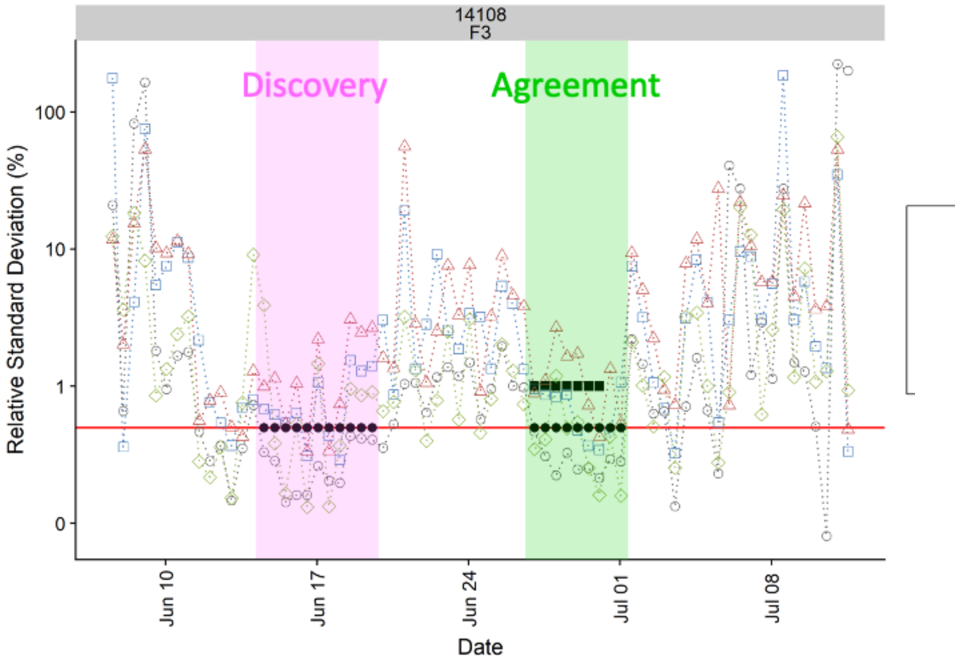


Progress and Outcomes: Algorithmic Tool to Identify ‘Steady State’ Conditions

Challenge: At inception, desired ‘Steady State’ conditions only could be identified by expert curation.

Solutions:

- Developed algorithmic tool for scanning many thousands of bioreactor observations for steady state conditions
- Tool correctly identifies regions indicated by expert curation (‘Agreement’) as well as proposed regions of steady state missed during curation (‘Discovery’)

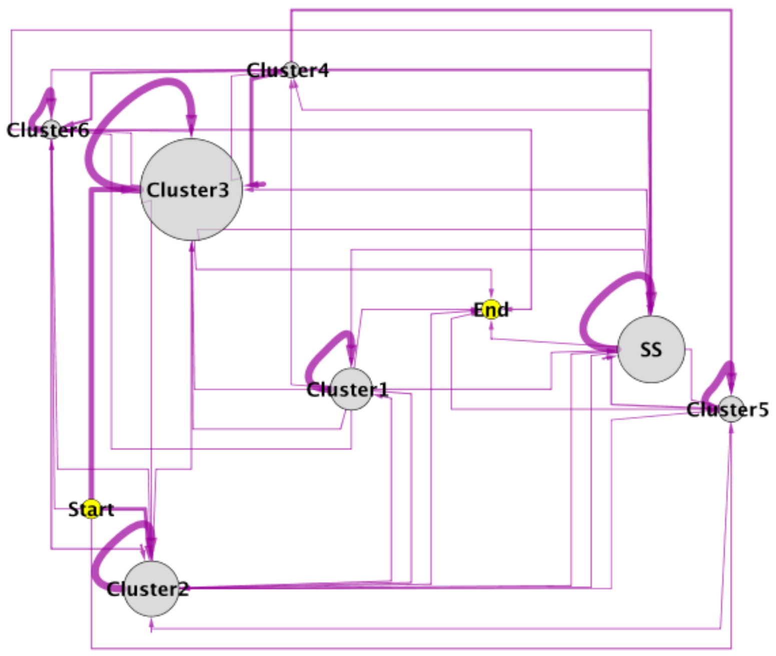


Progress and Outcomes: Data Reduction

Challenge: In data, there are many thousands of observations with dozens of data points per observation.

Solutions:

- Significantly reduced complexity of bioreactor data through statistical clustering
- Discovered discrete 'meta-stable' states, in addition to steady state,
- Provides a mechanism by which complex interactions of dozens of features in bioreactor can be simplified to navigating transitions between meta-stable states to achieve shortest possible distance to steady state condition



- Nodes = Cluster state for bioreactor run. Size of node is proportionate to frequency of cluster in total bioreactor data
- Edges = Observed transitions between cluster states. Weight of edge is proportionate to frequency of transition

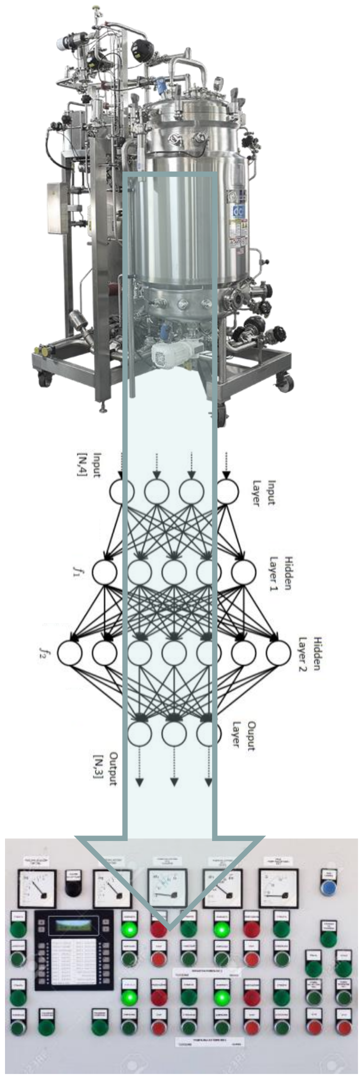
Progress and Outcomes: Training

Used Reduced Data to Train AI Models for Optimized Production

Challenge: Given a current ‘meta-stable’ state, what are the operational parameters required to achieve a desired state, including steady state?

Solutions:

- Use collected data to train Artificial Neural Network
- Model Parameters
  - Input =
    - current bioreactor sensor readings
    - desired bioreactor metastable state (a change in state or maintenance of current state)
  - Output
    - bioreactor process settings required to achieve desired state
- Training model on thousands of state-transition observations across multiple bioreactor runs



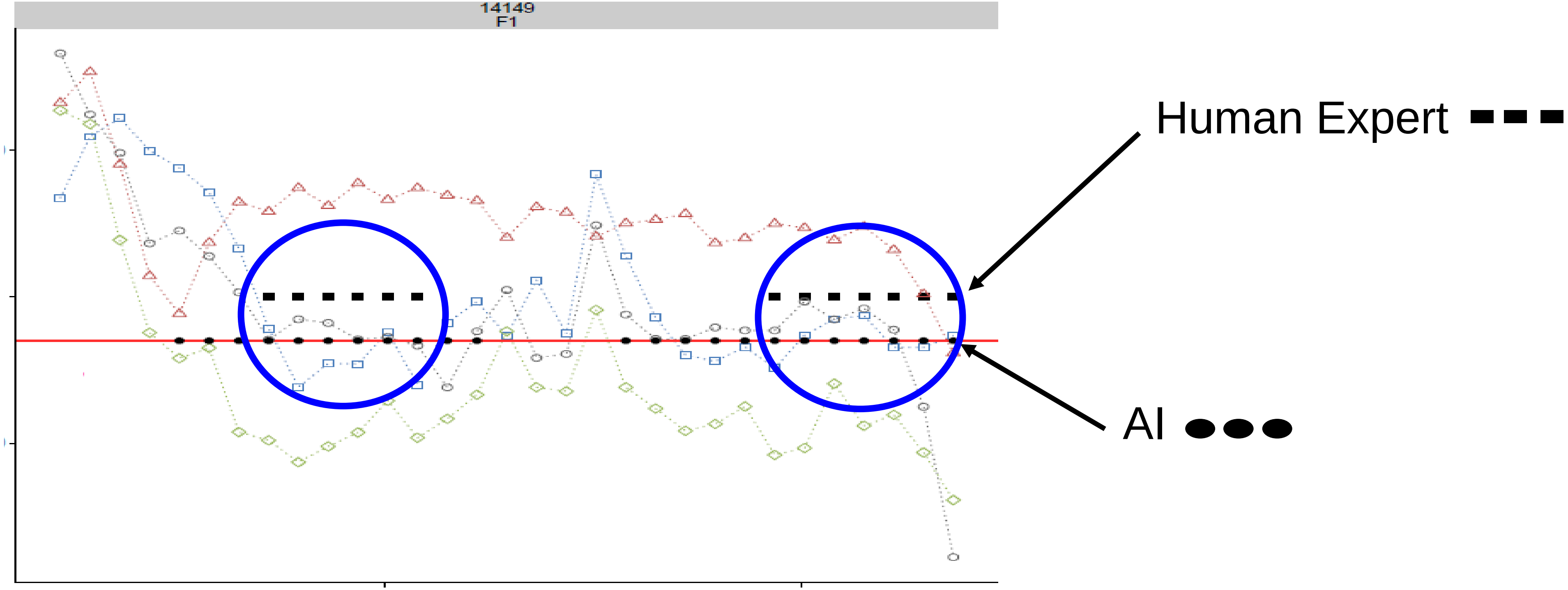
Progress and Outcomes: Validation and Model-Sharing

Challenge: ANN models must be shared and leveraged to maximize utility within ABF and current and future industry partners

Solutions:

- Save models, as ANN file and anonymized tables of data observations and model outputs, to **ABF Experimental Data Depot**
- Deposit approach (as R code) to **GitHub** or **JupyterHub** of the ABF

Progress and Outcomes: AI Model Recapitulates Human Experts

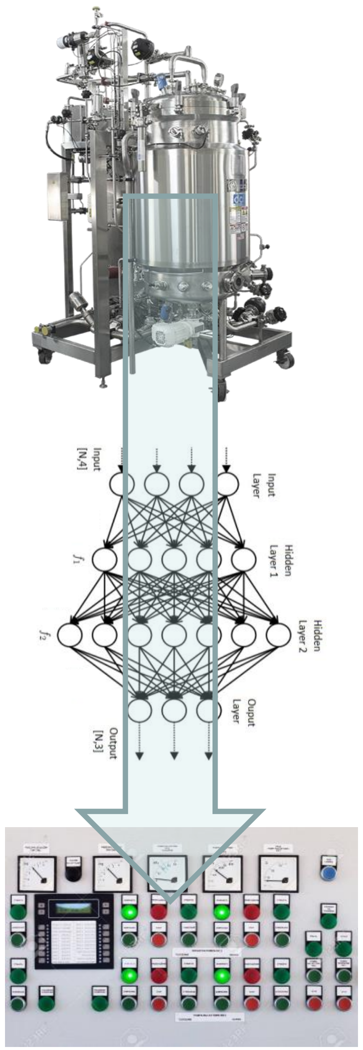


5 – Summary

Challenge: Given a current set of bioreactor sensor inputs, state, what are the operational parameters required to achieve a desired state, including steady state?

Solutions:

- Reduce data complexity by identifying one of several ‘meta-stable’ states from 1000’s of collected bioreactor sensor observations
- Use reduced-complexity data to train Artificial Neural Network
- Model Parameters
    - Input to model =
        - current bioreactor sensor readings
        - desired bioreactor metastable state (a change in state or maintenance of current state)
    - Output of model =
        - bioreactor process settings required to achieve desired state
- Model trained and cross-validated on extensive LT bioreactor datasets to propose unique computational approach and discovery of transition networks in data



Quad Chart Overview

Timeline
- 5/2/2018
- 5/1/2021

	FY20 Costed	Total Award
DOE Funding		$ 500,000

Project Partner
- Argonne National Lab
- NREL

Project Goal
Leverage LT's extensive database of bioreactor data to train Artificial Intelligence (**AI**) modules that will monitor bioreactors in real-time for continuous process modifications that maximize fermentation production, robustness, and stability

End of Project Milestone
Through analysis of multi-omics datasets identify 10 new engineering targets that would lead to increased process stability, reproducibility , and profitability and expansion into new product space.

Funding Mechanism
Agile Biofoundry

*Only fill out if applicable.

Publications, Patents, Presentations, Awards, and Commercialization

- No Publications, Patents, Public Presentations