



# How to Make Your Lab AI-ready

Philip Michaels, Solutions Architect

# Pistoia Alliance

Wed Nov 15 2023

- Momentum in AI/ML Adoption
- Challenges in AI/ML Adoption
- Use Case Enablement w/ TetraScience

# AI in the Life Sciences

## The Investment

**Spending for AI** in Life Sciences are expected to climb from \$1.87 billion in 2023 to \$6.70 billion in 2030

AI investments of **R&D budgets** will grow from 1.5 % in 2023 to 4 % in 2030

## The Promise

AI drug discovery to bring an additional **50 therapies** to market over the next 10 years is generating more than **\$50 billion** in revenue.

Generative AI could unlock value equivalent to **2.6 to 4.5 % of annual revenue (\$60 billion to \$100 billion)** in the pharmaceutical / medical-product industries.

## The Vision

"Our ambition is to become the first pharma company **powered by artificial intelligence at scale**, giving our people tools and technologies that focus on insights and allow them to **make better everyday decisions.**"

*Paul Hudson, CEO, Sanofi*

"[AI] is really set to **transform our industry** and at GSK it has been at the forefront of this, making significant and early investments in [these areas] with the ambition of **doubling our R&D success rates.**"

*Tony Wood, Chief Scientific Officer, GSK*

# The Potential of **Scientific AI** from Molecule to Market

- Target identification
- Candidate prediction
- Virtual screening
- *De novo* design
- Biosimulation
- Formulation optimization
- Safety prediction
- Efficacy prediction
- ADMET prediction
- Trial optimization
- Manufacturing optimization
- Digital quality control
- Predictive maintenance
- Knowledge reuse (LLM)
- Drug repurposing

## Poll Question

What is your Scientific AI maturity?

1. We are already using AI
2. We are planning to leverage AI in the near-future
3. AI is more of a buzz-word than reality right now

# AI-Readiness

## What's the problem?

“The first thing we’ve learned is the importance of **having outstanding data** to actually base your ML on. In our own shop, we’ve been working on a few big projects, and we’ve had to spend most of the time just cleaning the data sets before you can even run the algorithm. It’s taken us **years just to clean the data sets.**”

*Vas Narasimhan, M.D. , CEO, Novartis*



Data are  
siloes



Data are  
proprietary



Data are  
static



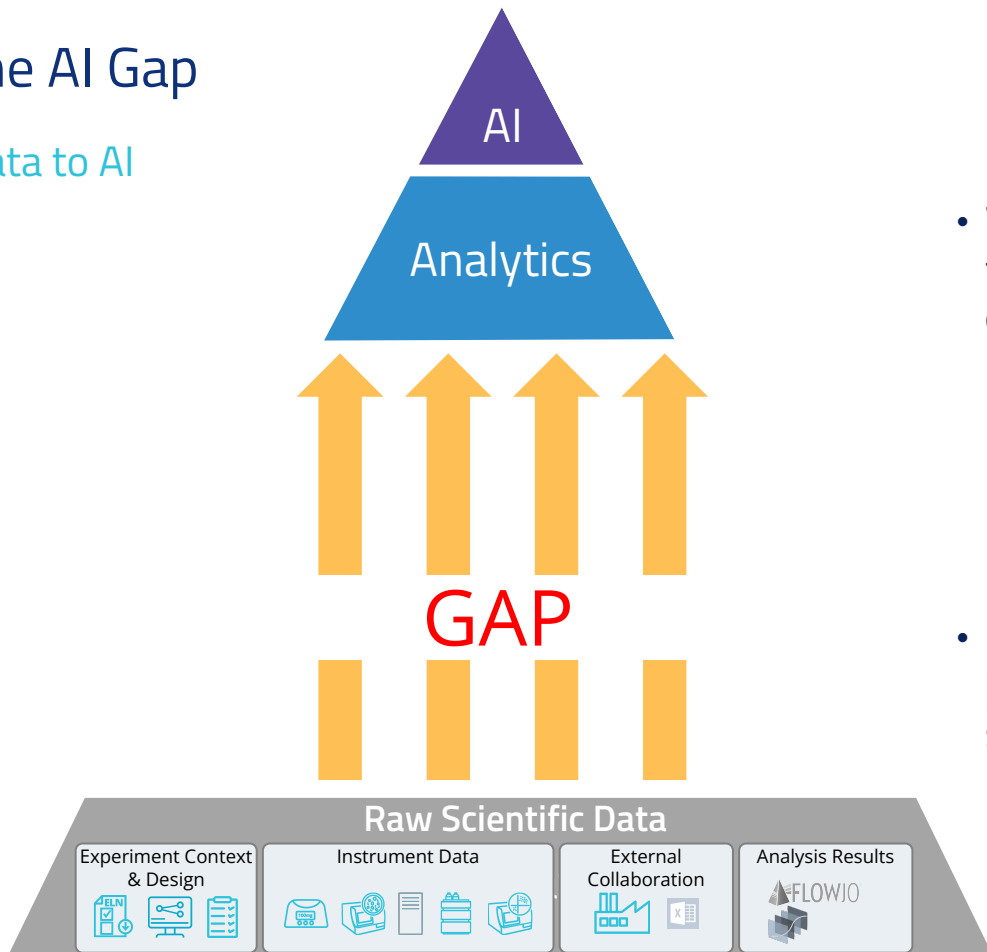
Data are  
subscale

“People should really recognize that the value of data is not in the tool or the method or the idea, **the value of the data is in the effort** that you put into making the data valuable.”

*Jeffrey Reid, Vice President, RGC Chief Data Officer, Regeneron*

# Closing the AI Gap

From raw data to AI



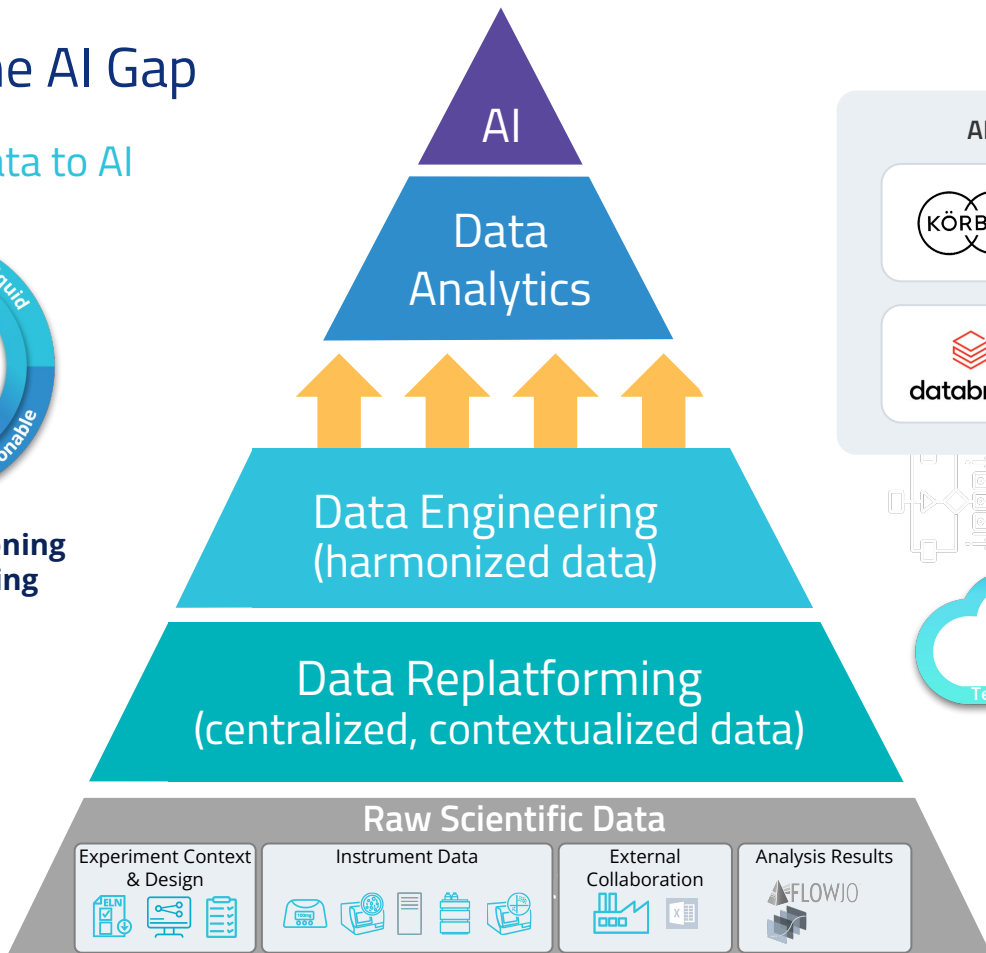
- We estimate that there are more than **10,000+ data silos in your organization**
  - Instruments, ELNs, informatics applications, CRO, CDMO, homegrown applications
- Resulting in **subscale, illiquid, proprietary, and unstructured** scientific data sets

# Closing the AI Gap

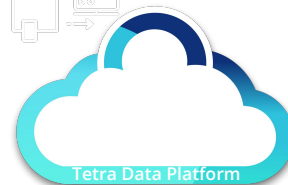
From raw data to AI



Schema Versioning  
& Reprocessing



ANALYTICS & TETRA DATA APP STORE



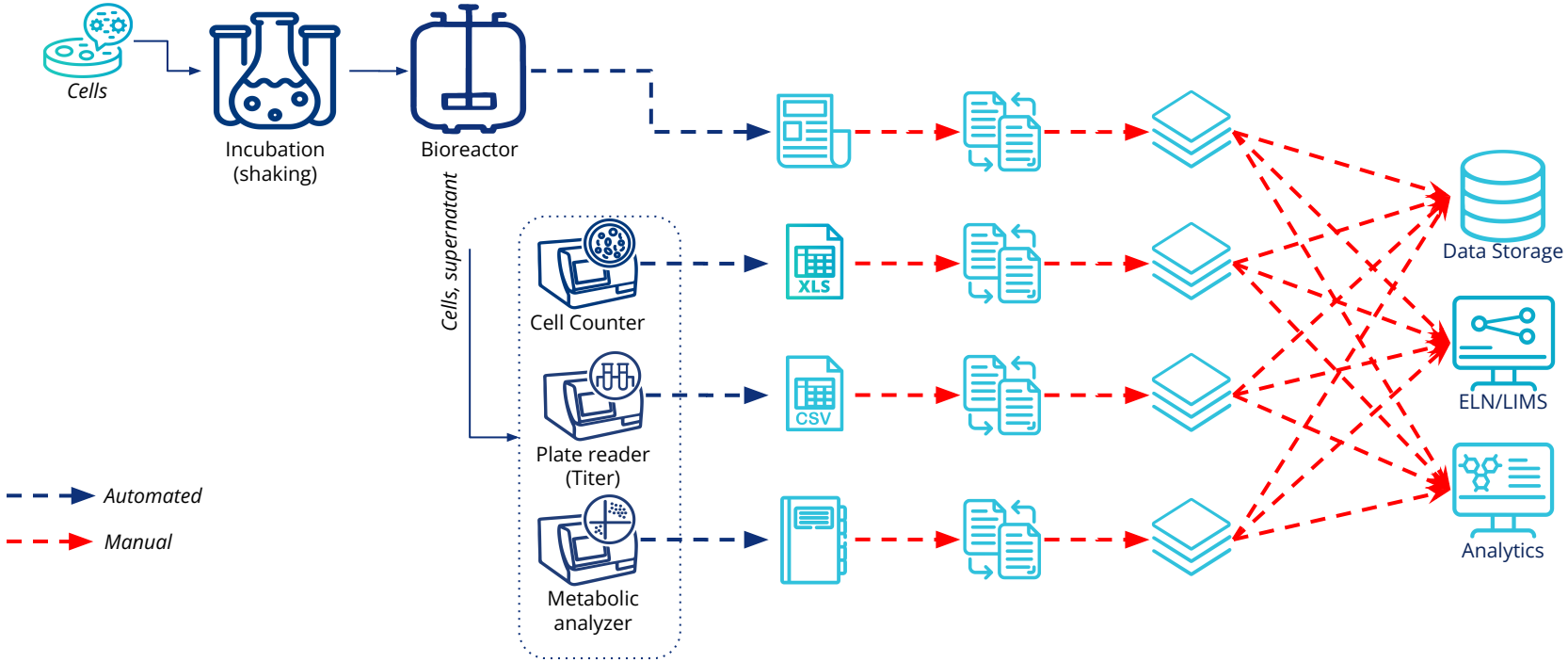
tetrascience  
Core software product



# Scientific Use Case with Tetra Scientific Data Cloud: **Upstream Bioprocessing**

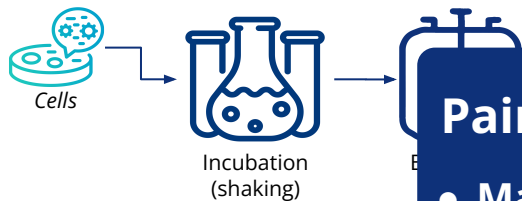
# Upstream Bioprocessing: Today

## Upstream Cell Culture Fermentation



# Upstream Bioprocessing: Today

## Upstream Cell Culture Fermentation



### Pain Points

- **Manual transfer/transcription** of data
- **No integration** between functions
- **Difficult to capture** experiment context
- **Limited** or no trending capabilities
- **Poor scalability**
- Little to no **data curation for AI/ML!**

- - -> Automated

- - -> Manual

Metabolic analyzer



Data Storage



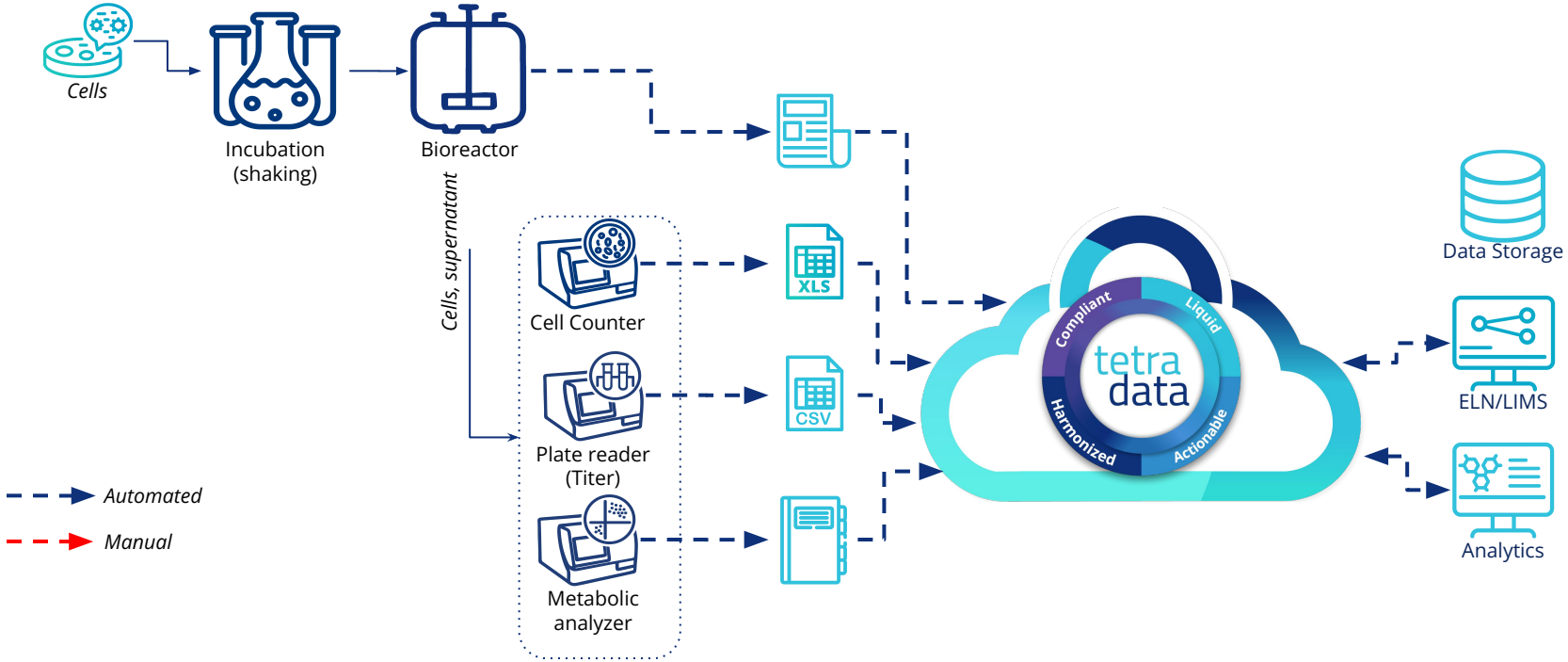
ELN/LIMS



Analytics

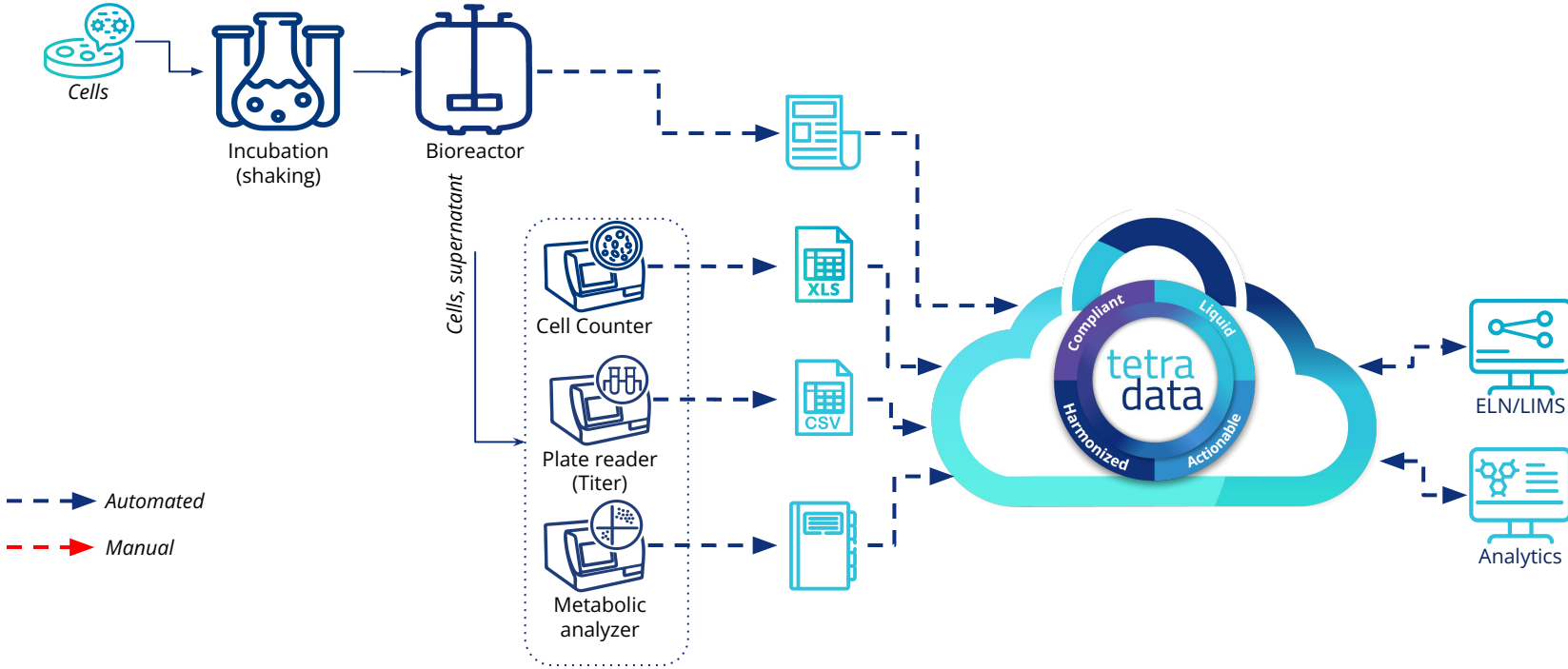
# Upstream Bioprocessing: With Scientific Data Cloud

## Upstream Cell Culture Fermentation



# Upstream Bioprocessing: With Scientific Data Cloud

## Upstream Cell Culture Fermentation



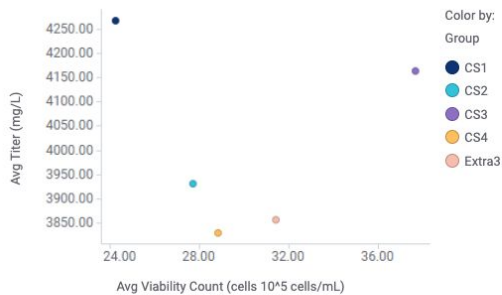
# Spotfire Dashboard



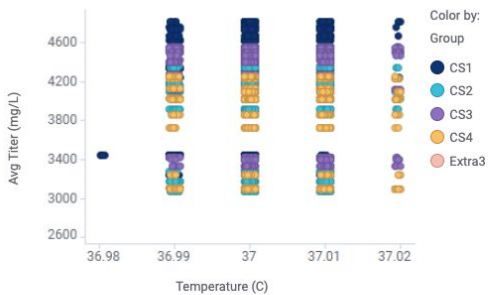
## Experimental Effects on Titer

Clear Marking

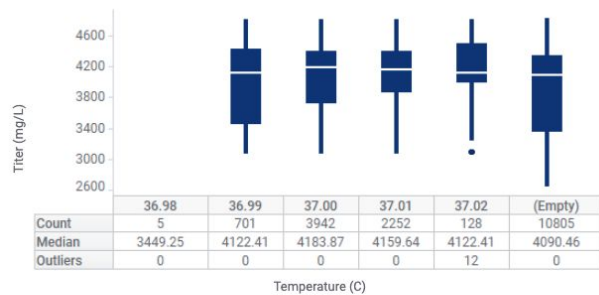
Titer vs Cell Viability at D21



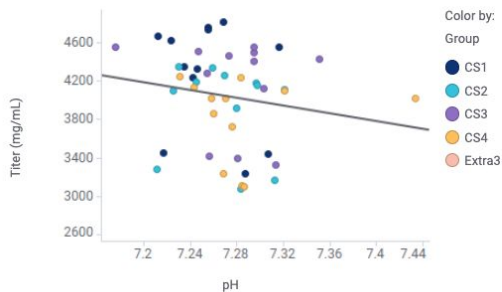
Titer (mg/mL) vs Temperature (C)



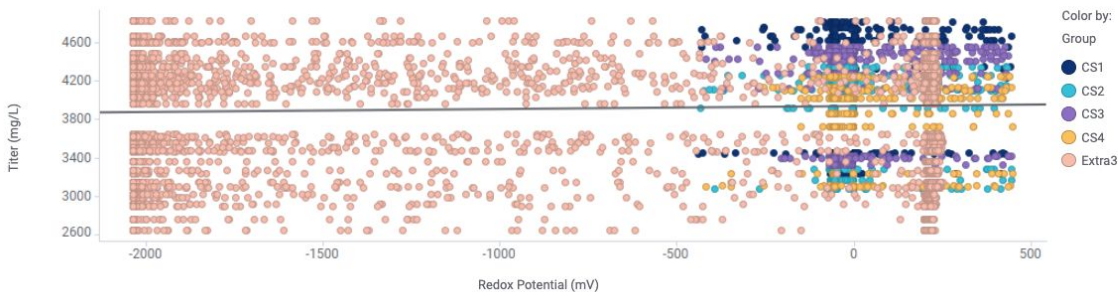
Box plot



Titer (mg/mL) vs. pH

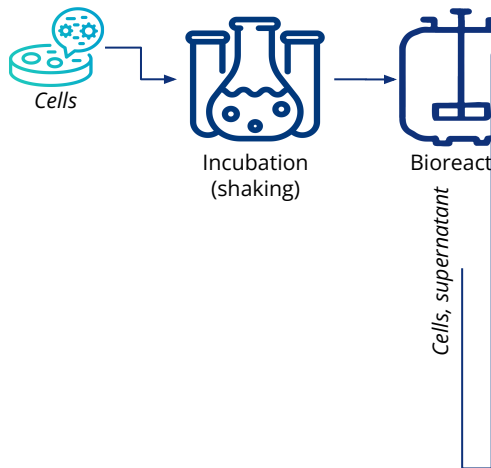


Titer (mg/L) vs Redox Potential (mV)



# Upstream Bioprocessing: With Scientific Data Cloud

## Upstream Cell Culture Fermentation

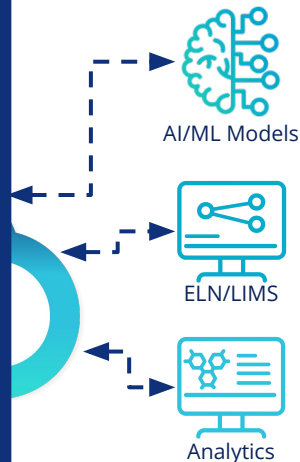


- - -> Automated

- - -> Manual

### Value:

- **Secure & Centralized** Data Storage
- **Reduce** scientific labor
- Ensure **Data Integrity**
- Facilitate **automation**
- **Search** and Query aggregate results
- API/SQL Endpoints for **Integration**
- Continuous feed for **AI Development!**



# Tetra Data Schema: selected schema as of Q2 2023\*

120+ different Tetra Data schema from instruments, control software, etc.



## Microscopes

Azure Biosystems, Hitachi, Leica, Nikon, Renishaw, Zeiss



## Flow Cytometers

BD Biosciences, Beckman Coulter, Cytex, Miltenyi Biotec, Sartorius, Sony, ThermoFisher, Generalized FCS



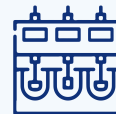
## Plate Readers

Agilent, BMG Labtech, Luminox, MSD, Molecular Devices, PerkinElmer, Sartorius, Tecan, ThermoFisher, Unchained Labs, and Wyatt



## Liquid Handlers

HighRes BioSolutions, Hamilton, Tecan, Beckman Coulter



## Dissolution Testers

Agilent, Sotax AG, Teledyne Hanson



## Chromatography Data Systems

Agilent Chemstation, Cytiva UNICORN, ThermoFisher Chromeleon, Empower



## Mass Spectrometers

ThermoFisher XCalibur, MassLynx/TargetLynx/MetaboLynx



## Cell Counters

Beckman Coulter, Chemometec, Generalized



## Time series

Ambr Bioreactors, time series data, aggregated time series

**\*Note: confirm modalities/techniques for specific scientific use cases with Tetra Catalysts or README documentation**

**Other Instrument Data Sources:** Particle Size Analyzer, PCR, Protein Analyzer, Blood Gas Analyzer, Imaging System, Particle Counter, Spectrophotometer, XRD, Bioanalyzer, Bioreactor, CDS-MS, Cell Analyzer, Compaction Simulator, Differential Scanning Calorimeter, Electromechanical Tester, Endotoxin Tester, Flow Imaging Microscope, Gas Pycnometer, Hardness Tester, Incubator, KFC, Liquid Scintillation Counter, MALS, Moisture Analyzer, NMR, Osmometer, Oxygen Analyzer, Peptide Synthesizer, Powder Flow Tester, Sanger Sequencing, Slope Spectroscopy, Spectrofluorometer, Sperm Analyzer, SPR, TGA, Thermocycler, Total Organic Carbon Analyzer, Vapor Sorption Analyzer



# Questions & More Information

- Come talk to us at our booth!
- More **information** at
  - [www.tetrascience.com](http://www.tetrascience.com)
  - [info@tetrascience.com](mailto:info@tetrascience.com)

# Appendix



# What is Tetra Data?

## Standardized JSON Schemas aka Intermediate Data Schema (IDS)



**LC-UV Empower IDS v4**

The following describes the additions and changes made to the LC-UV Empower IDS v3

- Removed `results[x].peaks[x].suitability`, and the data moved to the `results[x].peaks[x]` level
  - This has led to the removal of `results[x].peaks[x].suitability.custom_fields`, as R is identical to `results[x].peaks[x].custom_fields`:

```
    "usp_tailing":null,
    "usp_resolution":null,
    "usp_signal_to_noise_ratio":null,
    "usp_plate_count":null,
    "m_gravim":null,
```
- Added `sample.set.custom_fields` and `sample.set.method.custom_fields` to IDS:

```
    "sample":{
      "type":"Control",
      "name":"Q640555",
      "dilution":1,
      "set":{
        "name":"SampleName",
        "method":{
          "name":"SampleNameSHRILL",
          "id":"2315",
          "created_at":"2011-03-15T08:28:36.000Z",
          "custom_fields":{
            "key":"Analysis_Number",
            "value":"M_A",
            "string_value":"M_A",
            "value_data_type":"string"
          }
        },
        "custom_fields":{
          {
            "key":"Analysis_Number",
            "value":"M_A",
            "string_value":"M_A",
            "value_data_type":"string"
          }
        }
      }
    }
  }
}
```
- Removed `method.instrument.columns[x]` and `method.instrument.setups[x]` as nested



### SQL tables generated from IDS

LCuv Empower V 3 Databases

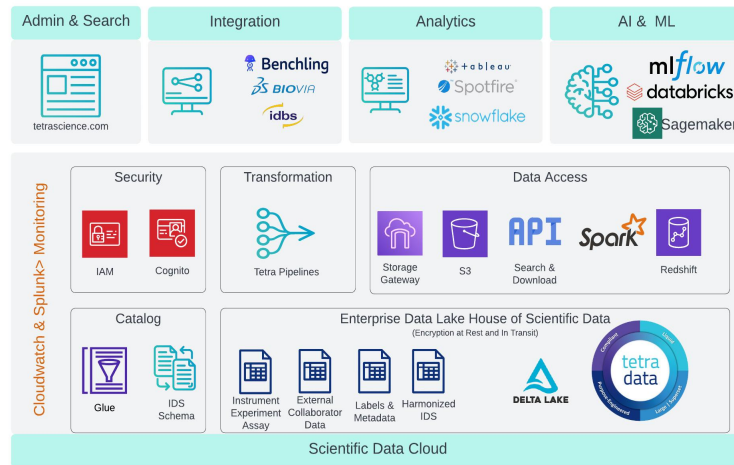
SQL query editor  
Need help writing your query? Check out the [SQL syntax documentation](#).

```
SELECT * FROM "lcuv_empower_v3_results_peaks_custom_fields" LIMIT 100;
```

Results  
Took 1.139 seconds, scanned 6.93 kb.

	uuid	parent_uuid	key
1	e693c772- bdba	2a5c1d99- 56a2	Percent_ww
2	42e201cf- dea6	36f2a10a- 4082	Percent_ww
3	bd755cc9- 066a	fe7344fe- e849	Percent_ww
4	624a5332- 1e0d	b64bc65b- 43ce	Percent_ww
5	4c78b866- 0ade	23942aff- 853f	Percent_ww
6	5fa5f97c- 41ee	5f485da8- daf6	Percent_ww
7	7238e462- ba53	10b383e8- a438	Percent_ww
8	0be1d7db- 6a80	fe7344fe- e849	Percent_ww
9	35a262b0- 9f3e	b64bc65b- 43ce	Percent_ww
10	45a45f9a- 57b6	23942aff- 853f	Percent_ww
11	49328c1c- 154a	5f485da8- daf6	Percent_ww
12	16eae59- b098	10b383e8- a438	Percent_ww
13	2aa28b16- 4950	fe7344fe- e849	Percent_ww
14	9866197f- 7877	b64bc65b- 43ce	Percent_ww
15	b044edb6- a20d	23942aff- 853f	Percent_ww

# Redshift/Snowflake Support and Evolving to LakeHouse



Native support of Redshift and Snowflake supporting federated data access to fit seamless with customer's existing data infrastructure

Releasing scientific data lakehouse in mid 2024