Max Petersen

Practice Manager, Lab Data Automation

Data Continuity in the Connected Lab:

Addressing Data Orchestration Challenges in Complex Multi-Stage Experiments

SiEE, 4/9/24

Lab Data Automation @ Zifo

Challenge

- Diverse: Multitude of instrument types
- Inhomogeneous: Same type, but from multiple vendors
- Long-lived: 4-20x original OS
- Typical @ large orgs: 2000+ instruments
  - 100+ manufacturers
  - 300+ models
  - 500+ applications / tools / software systems
  - 20+ departments
  - And growing every year ~30-35%

Data Acquisition

Zifo Position in the Field

Zifo is a scientific service provider

- 2500+ members
- Global
- Partner focused

Lab Data Automation

- 8+ years experience
- 50+ team
- SMEs, data/cloud architects, BPMN experts,

Diverse portfolio of complimentary services

- Lab compute (instrument networking)
- Managed services (cloud infrastructure)
- Validation services (GxP environments)

Impact

Custom implementations & pre-build tech

- Cloud architecture
- Guaranteed delivery
- Scalability

Instrument connectors

- Skilled team for all connection types (OPC, RS232, dll, API, …)
- Ability to implement new connectors quickly

Partnerships & platform interoperability

- Working with many players in the field: TetraSciences, Sapio, Benchling,
- Successful augmentation of bespoke systemes

Implications of the Connected Lab on FAIR Data Practices

Findable

- Data model/ontology: Consistent data representation
  - Sample ID, experiment ID,
- Metadata: Correct values in data model/ontology
  - Scattered in sample mgmt., ELN, LIMS, …
  - Auto-generated IDs by robotic systems
- Data/results accessible in the systems scientists use to get their work done
  - Key for efficiency gains and business process improvements

Accessible

- Simplified data curation
  - High overhead (~80% of total effort)
  - Data FAIRification/data quality challenges
- AI/ML applications
  - Data quality and inconsistency challenges
  - Efficiency (continuous learning)
- (Near) Real time analytics
  - Time-critical applications e.g. bioprocess control
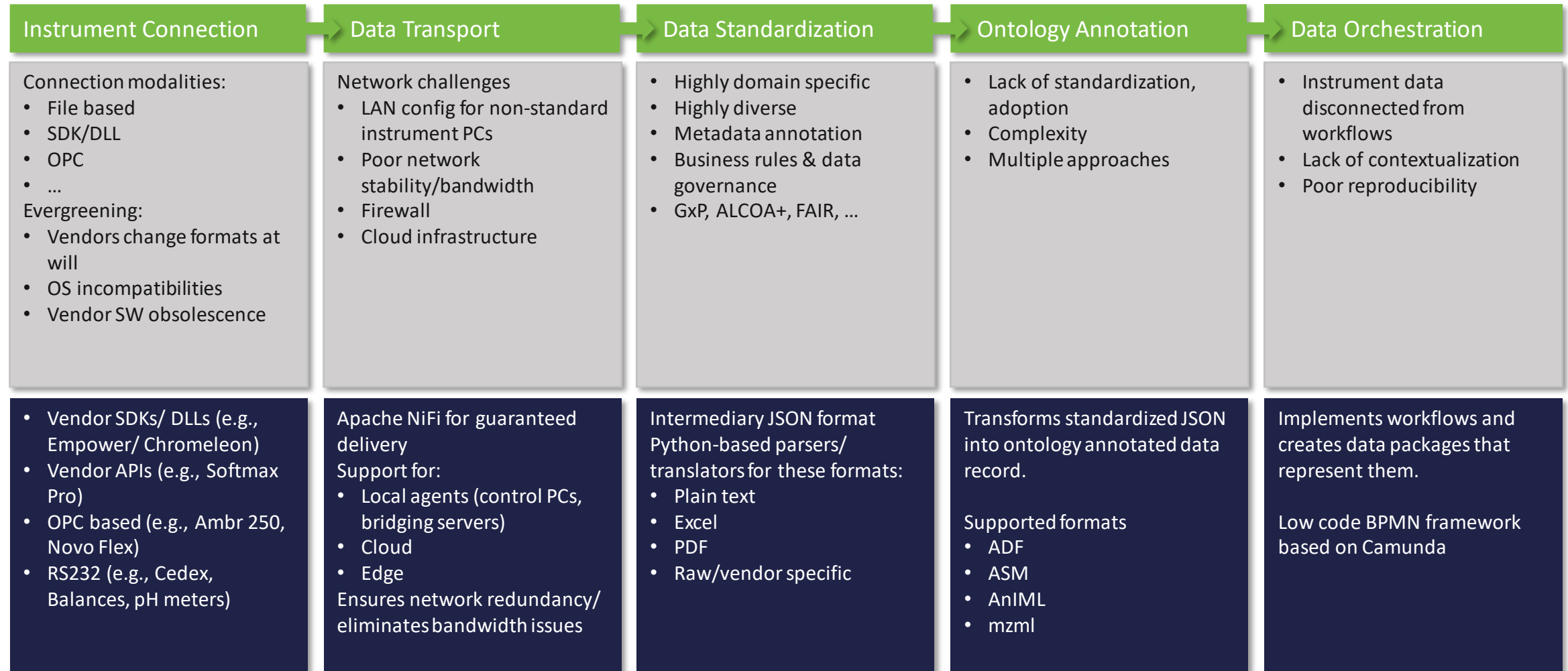 - Scale-up challenges caused by non-standardized data

Interoperable

- Across instruments
  - Data standardization
  - Machine and human readable
- Across systems
  - Bi-directional data flow synchronizes metadata
  - Data and results available in ELN, LIMS, SDMS, LES, MES,
- Across platforms
  - Application independent, data product

Reusable

- The correct experimental context
  - Data orchestration organizes Data packets alongside workflows
- SOPs and workflows encoded into lab data automation
  - Data orchestration provides a repeatable framework

Tech Challenges: Form Instrument to Cloud

Instrument Connection	Data Transport	Data Standardization	Ontology Annotation	Data Orchestration
Connection modalities: • File based • SDK/DLL • OPC • … Evergreening: • Vendors change formats at will • OS incompatibilities • Vendor SW obsolescence	Network challenges • LAN config for non-standard instrument PCs • Poor network stability/bandwidth • Firewall • Cloud infrastructure	• Highly domain specific • Highly diverse • Metadata annotation • Business rules & data governance • GxP, ALCOA+, FAIR, …	• Lack of standardization, adoption • Complexity • Multiple approaches	• Instrument data disconnected from workflows • Lack of contextualization • Poor reproducibility
• Vendor SDKs/ DLLs (e.g., Empower/ Chromeleon) • Vendor APIs (e.g., Softmax Pro) • OPC based (e.g., Ambr 250, Novo Flex) • RS232 (e.g., Cedex, Balances, pH meters)	Apache NiFi for guaranteed delivery Support for: • Local agents (control PCs, bridging servers) • Cloud • Edge Ensures network redundancy/ eliminates bandwidth issues	Intermediary JSON format Python-based parsers/ translators for these formats: • Plain text • Excel • PDF • Raw/vendor specific	Transforms standardized JSON into ontology annotated data record. Supported formats • ADF • ASM • AnIML • mzml	Implements workflows and creates data packages that represent them. Low code BPMN framework based on Camunda

Parser and Connectors – The Story So Far (250+)

Instrument Connectors & Parsers on the Web

Lab Data Automation:

Workflow Support & Data Orchestration
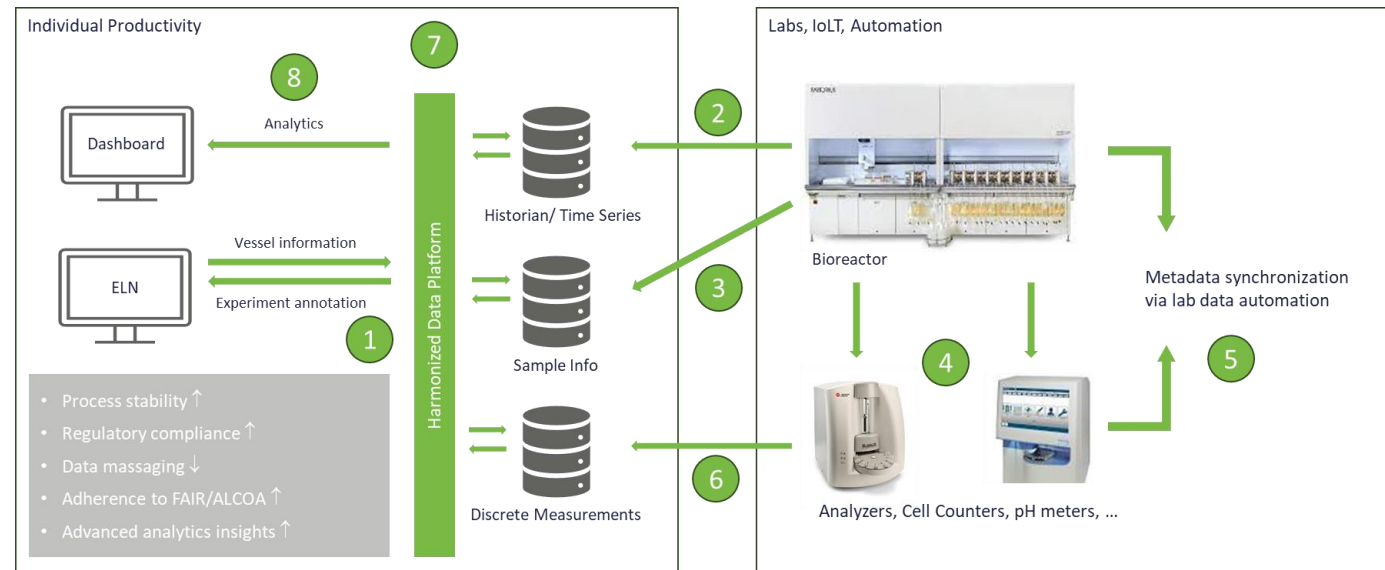
1. **ELN/LIMS Integration (Pull)**
   - Existing connectors for dotmatics, IDBS, Labware, SampleManager, Sapio, Signals, ...

2. **Discrete Timeseries**
   - Existing OSI-Pi connectors

3. **Bioreactor Integration**
   - Existing connectors for Akta, Ambr
   - Lab compute support

4. **Bioanalyzer Integrations**
   - Existing connectors for Flex2, Vicell, Cedex, ...
   - Data orchestration
   - Lab compute support

5. **Metadata Synchronization**
   - Sample ID (bioreactor)
   - Vessel info (ELN)

6. **ELN/LIMS Integration (Push)**
   - Existing connectors for dotmatics, IDBS, Labware, SampleManager, Sapio, Signals, ...

7. **Cloud Data Platform Integration**
   - Existing connectors to common cloud providers
   - Implementation/consulting services
   - Managed services

8. **Analytics/Dashboards**
   - Implementation services
   - Data sciences services

CMC Data Capture Automation

Client

Large global Bio-Pharma company specializing in vaccine development and specialty medicines.

Problem

Explosion of data volume and a drive to accelerate and bring in cost efficiencies in the Biologics CMC space warranted a need to build a Digital future roadmap, Data Analytics on top of encompassing Instrument and experiment Data Automation as a priority use-case.

Solution

The goal was to implement data warehousing system on top of CMC's Experiment and Instrument in alignment with FAIR and Data fabric architecture to data analytics and ML workloads.

Technology

- Python
- ADF
- NET
- Azure

- Databricks
- Kubernetes
- Denodo
- ADLSV2

- Azure DevOps
- GitHub

Success

Implemented data pipelines from various instrument and experiment data, make it align with Data models, Data Standards which can serve as the source for other applications such as ELNs, LIMS also dashboarding and analytical tools

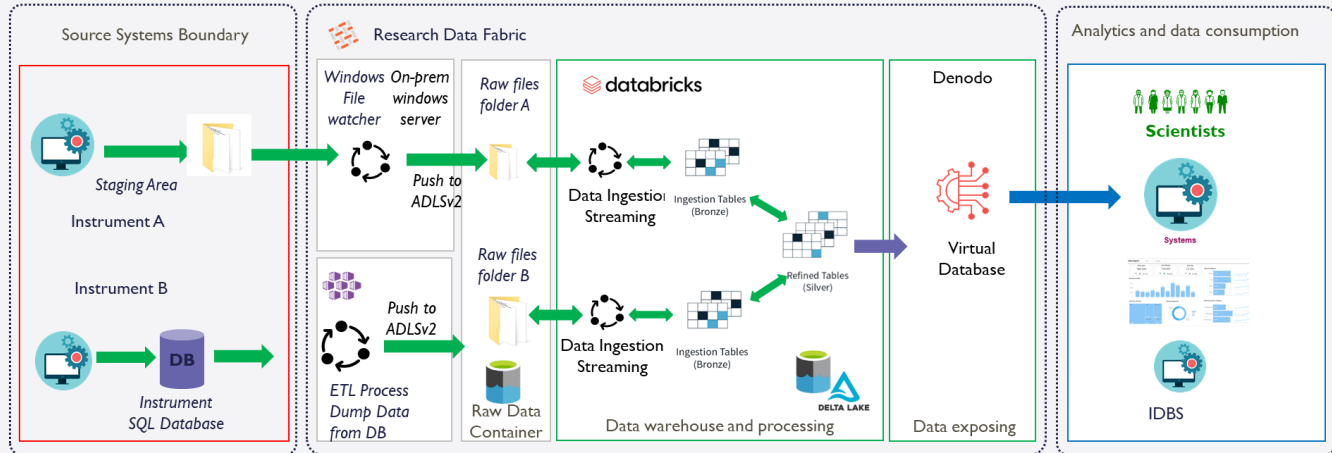## Technical Architecture Diagram – Experiment Systems
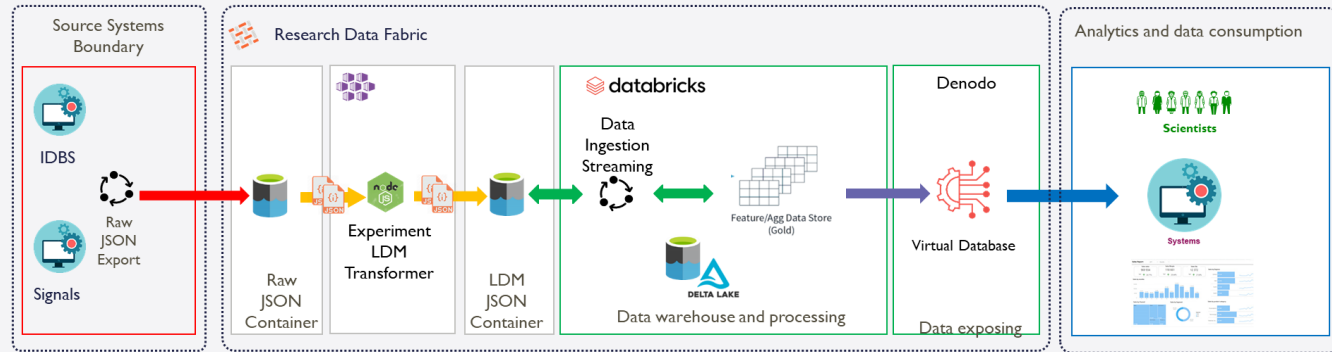
Technical Architecture Diagram – Instruments

TO FIND OUT MORE VISIT US AT
ZIFORND.COM