Creating The Proper Infrastructure for AI-Driven Development

Thursday, September 07, 2023

Presented by

Noel Hollingsworth | CEO & Co-Founder | **Uncountable**

Agenda

Featured Speaker: Noel Hollingsworth

Noel Hollingsworth is Co-Founder and CEO at Uncountable. In his role, he works closely with Uncountable's customers to implement next-generation data management systems.

Prior to his work at Uncountable, Noel led data teams at startups and was awarded Forbes 30 under 30 for his work with machine learning and artificial intelligence.

uncountable.

About Uncountable

Founded in 2016
with offices in San Francisco, New York City, and Munich

90+ customers across industries
including paints, rubbers, 3D printing, foams, cosmetics, alternative foods, and more!

One-of-a-kind platform
that centralizes R&D data and helps reduce new product development timelines

Proven domain expertise
began as a data science company helping Fortune 500 materials companies accelerate development of new projects.

uncountable.

Uncountable Proudly Supports Clients

That Span Across a Variety of Industries

BEHR
INX A SAKATA INX COMPANY
CooperStandard
CABOT
PENN COLOR A WORLD OF COLOR
Sika
codelpa COLORES DEL PACIFICO
LA·CO INDUSTRIES, INC.
SunChemical a member of the DIC group
LESAFFRE
Beiersdorf
AGC
SCG
NATURAL FIBER WELDING
6K
ST. JUDE CHILDREN'S RESEARCH HOSPITAL
Carbon
SIMULATE®
ripple
ADVANO
pixelligent
AeroShield
BlueNalu
AirBoss Rubber Solutions
TotalEnergies
3D SYSTEMS
SUMMIT NANOTECH
NOHBO
nanoprecision medical
SOPREMA
REPSOL
tesa
ZEON

Scan QR Code To Read
Customer Case Studies

Or visit:
www.uncountable.com/case-studies

The Current Landscape: AI in R&D

- **Modern AI in R&D**
  - There's a common misconception that rebranding as an AI company is as simple as having some off-the-shelf models or working with an AI vendor
  - AI is complex, high-risk, expensive, and often requires significant business transformation to collect the data necessary
  - 2018 reports: Implementing AI demands an ultra-specialized talent pool that only 22,000 PhD-level experts worldwide

- **Goals of AI***
  https://link.springer.com/chapter/10.1007/978-3-030-50344-4_18
  - More objective identification of user requirements to drive enterprise innovation
  - More precise exploration of market trends
  - Higher efficiency in product design
  - Less risks in R&D process
  - Improved knowledge sharing ability

Uncountable The Core Issue Many Fortune 500 R&D Teams Face is Unstructured, Decentralized Data

Raw Materials

- Physical & Chemical Attributes
- Batch History
- Prices
- Regulatory & Compliance

Formulations

- Compositions
- Calculations
- Multi-part systems
- Order of Addition
- Dispersions & Intermediates

Processing

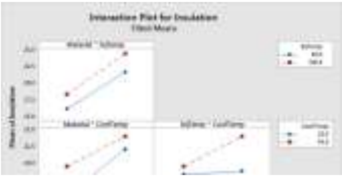
- Mixing Conditions
- Curing Parameters
- Metadata

Formula Properties

- Physical Properties
- Analytical Measurements
- Observations
- Images
- Curves, Reports

Product Properties

- Application Testing
- Customer Feedback
- Exposure Testing



uncountable.

Leveraging AI in R&D: Achieving Competitive Advantage

- **Future prospects and trends in AI-driven R&D**
  - Importance of creating a sustainable foundation for “future-proofing” AI prior to prematurely trying to leverage or implement AI-driven technology into R&D efforts
  - How the future of R&D will look with AI vs. how it looks today

- **Ways AI can enhance R&D performance and competitiveness**
  - Short-term
  - Long-term

Data Systems: Structured vs. Unstructured

Unstructured

Examples

- Spreadsheets
- Word Documents
- PDFs
- Lab Journals/ELNs
- SharePoint/Shared Drive

Advantages

- Free
- Unrestricted entry of information
- Known/second nature “habitual”

Disadvantages

- Limited scope & scalability for application of info
- Ctrl+F keyword searching
- Limited collaboration
- Inability to innovate efficiently & at market-rate

Structured

Examples

- Databases
- LIMS
- Inventory Systems
- Uncountable

Advantages

- Instant access to specific information/data
- Shareable & scalable information
- Intelligent insights & reporting

Disadvantages

- Requires intentional/deliberate entry of information
- Change management
- Migration of historical data into new system
- Disciplined use

Top 3 Problems Deploying AI Without Structured Data

- **Why Excel & Unstructured Data System Are Insufficient**

1. **Volume of Data**
    - A small data set with the best AI model in the world is worse than both expert scientists and simpler AI models applied to “big data”
        - The most important aspect of any AI model is its underlying data - both size and cleanliness

2. **Relevancy to Problems**
    - Will create desire to squeeze square peg in round hole - When we do have some data, we must apply AI, even if it’s not a fit
        - AI is not a fit for all use cases!

3. **Scientist Trust**
    - Desire to be an AI first company without gathering appropriate data results in scientist trust being lost
        - AI ends up being applied to projects that aren’t good fits, or only to high priority projects that carry substantial failure risk - when there are issues, team loses faith in the process
        - Sufficient Data is important, but not the only prerequisite

Common Data Systems: LIMS, ELN, and ERPs

- **Laboratory Information Management System "LIMS"**
  - Sample management / metadata
  - Output Capture
  - Task management
- **Electronic Laboratory Notebooks "ELNs"**
  - Experimental capture in real time / Collaborative
  - Needs "spreadsheet" support (formulation/analysis)
  - Unstructured
- **Enterprise Resource Planning "ERP"s**
  - Inventory systems
  - Equipment management
  - BOMs
  - Information store / system of record

uncountable.

Importance of Structuring Lab Data for AI

- **Why Structured Data is Important for AI**
  - Standardization/Consistency
  - Reliability
  - Feature Engineering
  - Scalability
  - Data Integrity
  - Interoperability
  - Ground Truth & Labeling
  - Reduction in Noise
  - Easier Analysis & Debugging
  - Data Governance

uncountable.

Importance of Structuring Lab Data for AI: Example of Brookfield Viscosity

- **Standard Way Data Gets Recorded In Spreadsheets and Notebooks:**
  - Viscosity, 7D = 3000
  - Brookfield Visc. Sp #4 = 5500
  - BV, ON = 1800

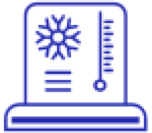
- **Best Practices for Structuring Lab Data for AI:**
  - Brookfield Viscosity = 5000
    - Liquid Aging Time + Temperature: 7D at 23°C
    - Spindle #4
    - RPM: 150
    - Test Temperature: °23
      - Exact temperature and time
      - Machine SN, Operator

uncountable.

Types of Data Systems: R&D Organizations

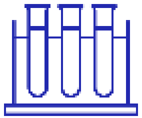
Spreadsheets

ELN/Lab Journals

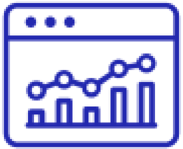


LIMS

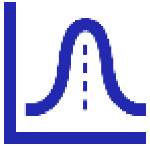
Inventory



Visualizations / Analysis

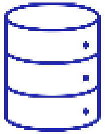


Statistical Tools



Predictive tools

Other Internal Databases



Considerations: Setting The Right Expectations

- **Too big of a search space**
  - 100s of ingredients, but limited data points
    - Either from collection, cleanliness, or standardization

- **Moonshot objectives**
  - What are you trying to achieve in this project vs long term goals
  - What are more reasonable targets that would allow you to claim “progress”

- **Perception of perfection**
  - Why would model suggest such a thing?
  - Why isn’t model more accurate?
  - Can it model pictures of exposure ratings?

Defining & Creating an AI Roadmap

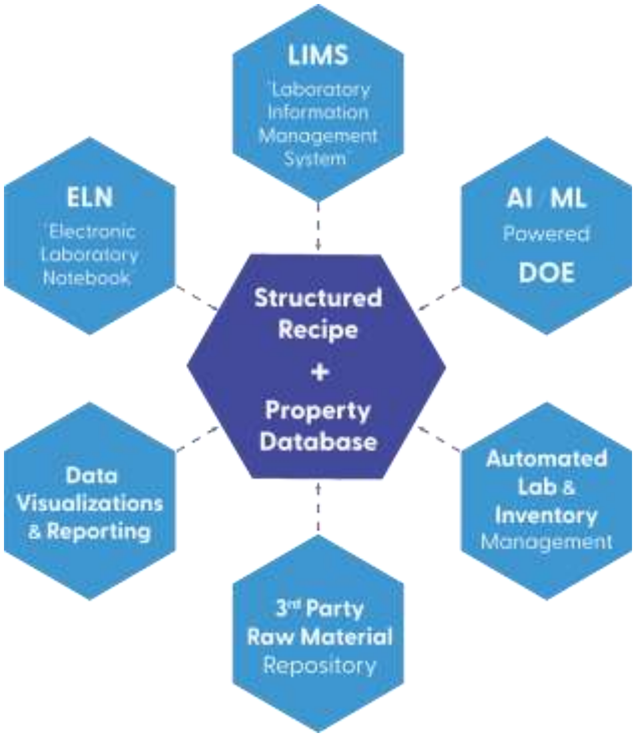
1. **Before (Preparation)**
   - Ensure structured data system in place
   - Verify all scientist work is being captured in a way fit for AI
     - All data points and all aspects of data
       - Example: Viscosity centipoise, temperature, spindle, rpm…
   - Utilize in-house expertise to understand/validate vendor and partner "claims"

2. **During (Deployment)**
   - Identify appropriate targets for AI - Example Criteria:
     - Large Amounts of Data
     - Known Success Criteria
     - Consistent Output Results
   - Ensure AI is embedded into daily workflows
     - Not judged off success in a project where majority of results are out of scientists control

3. **After (Maintenance)**
   - Identify areas where data capture is insufficient
   - Deploy systems and/or recurring procedures to collect data

All-In-One Structured Data Platform

We Created The Our Platform *To* ***Centralize, Connect, And Structure All Types Of R&D Data.***



uncountable.

uncountable

Example 1

Platform-Wide Data

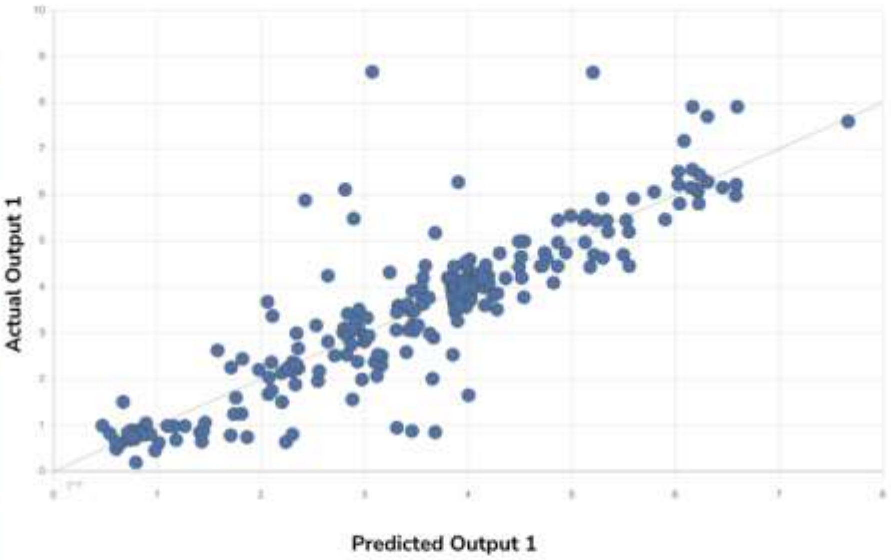
Example 1: Output Fits

Training Accuracy

Numeric Outputs

Plot	Name	Summary Statistics (Training Data)			Model Predictions		
		# Samples	Mean μ	Std. Dev. σ	RMSE	r^2 Score	Explained Error %
	Output 1	1360	172	145	54.4	0.86	62.6%
	Output 2	319	8.17	10.3	3.39	0.891	67.1%
	Output 3	254	16.4	7.38	5.03	0.533	31.8%
	Output 4	541	65.2	71.4	17.7	0.939	75.2%
	Output 5	359	8.02	5.4	2.96	0.698	45.1%
	Output 6	397	41.6	31.9	8.23	0.933	74.2%
	Output 7	1511	7520	12200	6070	0.752	50.2%
	Output 8	251	3.51	1.7	0.816	0.767	51.8%
	Output 9	895	7.25	4.03	2.12	0.725	47.6%
	Output 10	1475	8.26	6.71	2.8	0.813	56.8%
	Output 11	1497	0.00279	0.00112	0.000767	0.533	31.7%

High Predictive Accuracy   Low Predictive Accuracy   Add Comparison Mode



uncountable.

Example 1: Linear Coefficients

uncountable.

Example 2

Targeted Experiments

Example 2: Suggested Experiments

Suggested Formulations

Recipe Name		Recipe 1	Recipe 2	Recipe 3	Recipe 4	Recipe 5	Recipe 6
Import Recipe?							
Ingredient 1	= 2.5	2.5	2.5	2.5	2.5	2.5	2.5
Ingredient 2	= 0.202	0.2021	0.2021	0.2021	0.2021	0.2021	0.2021
Ingredient 3	= 0.0121	0.01209	0.01209	0.01209	0.01209	0.01209	0.01209
Ingredient 4	[6.52, 10.2]		9.709	10.11	8.072	8.411	9.5
Ingredient 5	[10.3, 25]	22.71	22.71	17.27	18	18.82	18.69
Ingredient 6							
Ingredient 7	[4.78, 10]	9.658	8.146		6.247	6.307	
Ingredient 8	[5.84, 10.1]	6.138	6.347	6.328	6.235	5.906	5.877
Ingredient 9	[5.2, 50]	27.86	39.03	32.46	30.69	46.62	49.81
Ingredient 10	[5.1, 14.3]	10.26	11.35	9.108	10.64	11.42	13.41
Ingredient 11	[14.8, 32]	20.67		22	17.4		
		100	100	100	100	100	100
Calculation 1	[2, 4]	2.25	3.83	2.09	2.39	2.98	2.78
Calculation 2	[0.8, 1.2]	0.804	1.11	1.02	1.08	1.13	1.08
Calculation 3		3.15	3.87	2.89	3.21	3.01	2.81
Calculation 4		1.91	9.82	0.595	3.04	6.36	5.41
	Goal						
Predicted Output 1	≥ 300	221 ± 111	236 ± 111	224 ± 109	216 ± 107	210 ± 108	215 ± 106
Predicted Output 2	≥ 6	7.5 ± 2.75	5.82 ± 3.11	7.71 ± 2.88	7.39 ± 2.84	5.71 ± 3.08	5.94 ± 3.11
Predicted Output 3	≤ 4000	2100 ± 1390	1860 ± 1830	2140 ± 1580	2050 ± 1460	1810 ± 1780	1900 ± 1840
Predicted Output 4	≥ 2	2.5 ± 1.07	2.68 ± 1.15	1.94 ± 1.12	2.56 ± 1.07	2.46 ± 1.1	3.05 ± 1.16

uncountable.

uncountable.

Q&A

Just a few of many

Uncountable Long-Term Customers

Beiersdorf

CABOT

Carbon

AGC

SCG

BEHR

Sika®

uncountable®

Thank You!

Want a Demo?

Scan The QR Code

Questions?

Email: info@uncountable.com
Inquiries: www.uncountable.com/contact-us

www.uncountable.com