



Traversing the pyramid: Data → Information → Knowledge → Wisdom

Richa Sarin

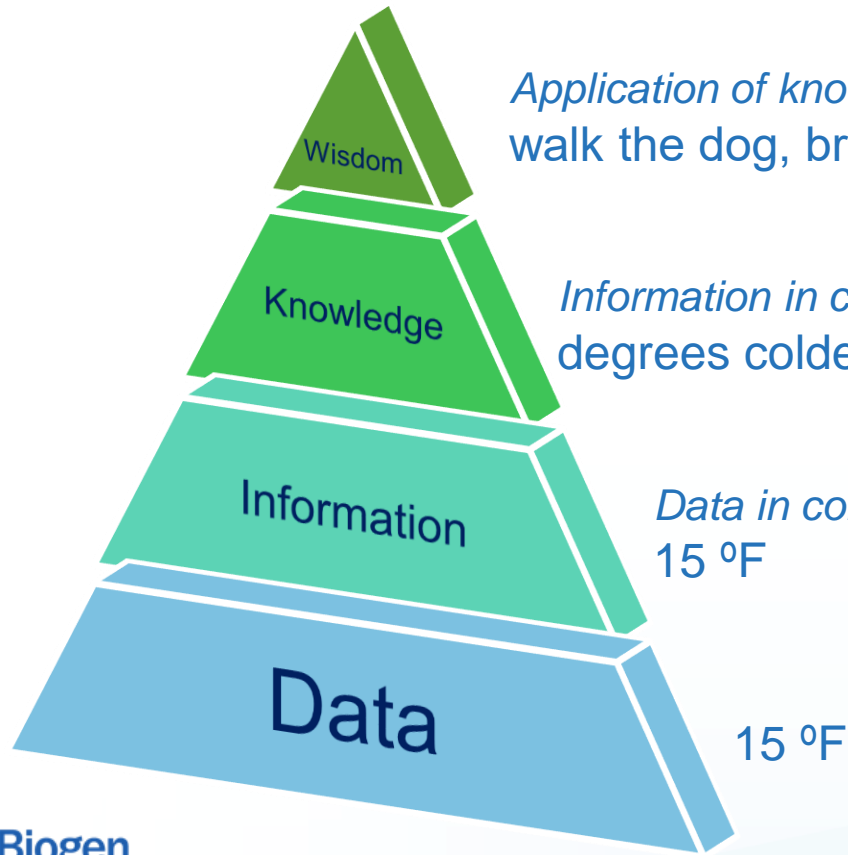
Scientist, Analytical Development

Biogen



Putting Prior Knowledge to Immediate Use: Knowledge Management and Digital Tools to Accelerate Development

The Pyramid



Application of knowledge in context: I need to wear my winter coat when I walk the dog, brrrr.....

Information in context: The next two days of January are 5 degrees colder compared to 2019

Data in context: The high in Boston for the next two days is 15 °F

15 °F

Agenda

Overview & background

Vision

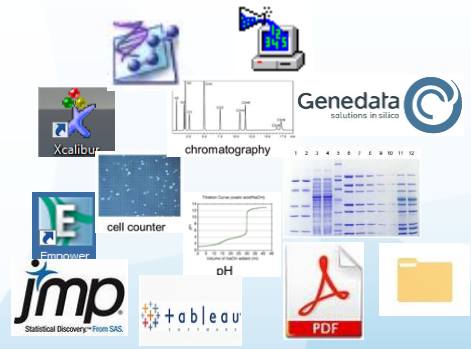
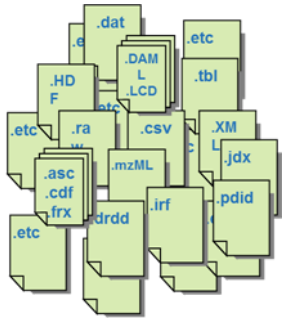
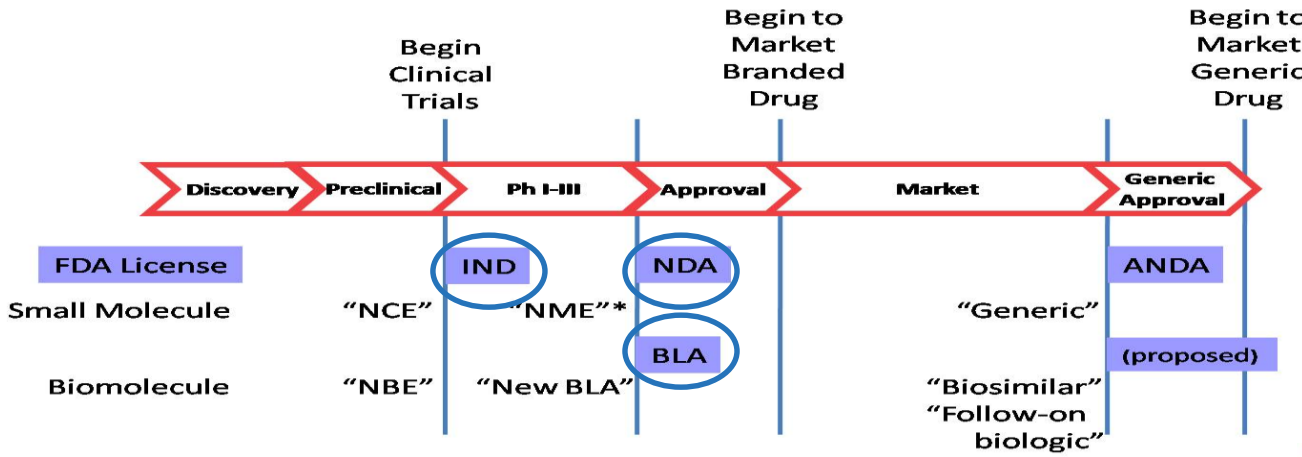
Case Studies

- Structured authoring: Recipe driven method development and qualification
- Data analytics: Optimizing the Clone Selection Process

Conclusion & Future Work

Acknowledgements

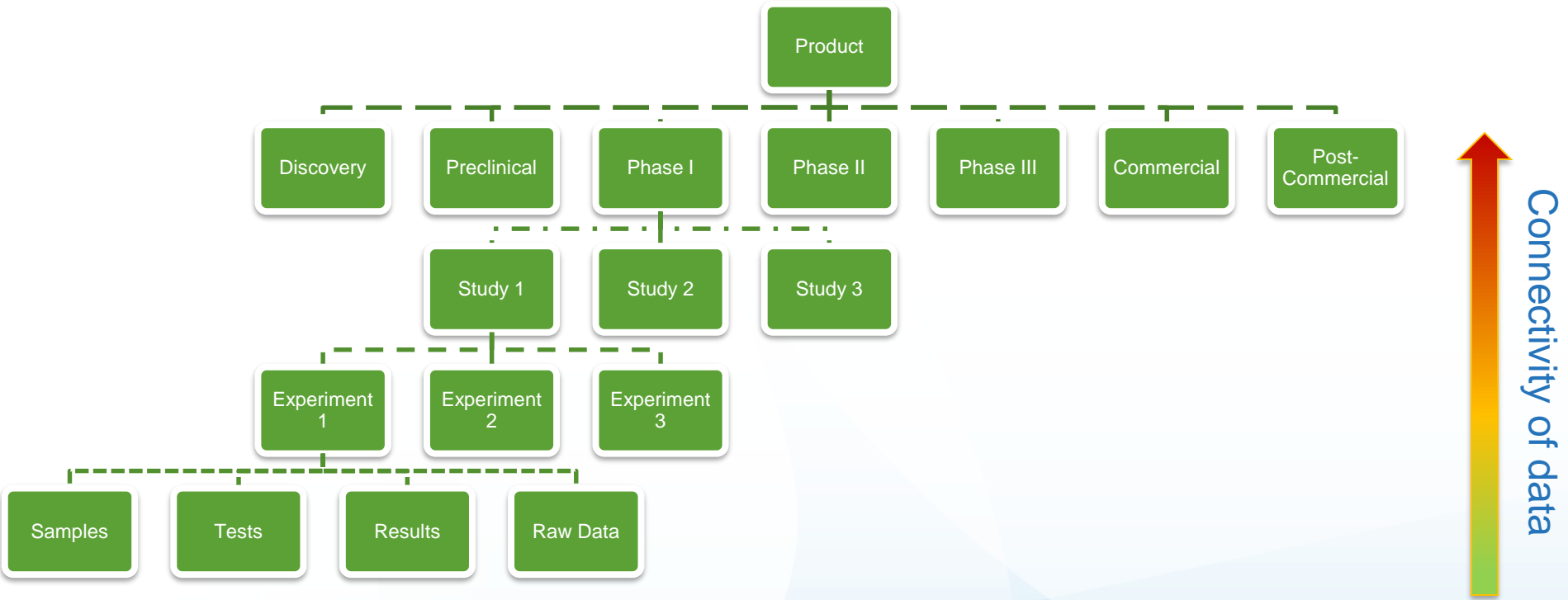
Overview



Challenges



Data challenges increase as you go up...



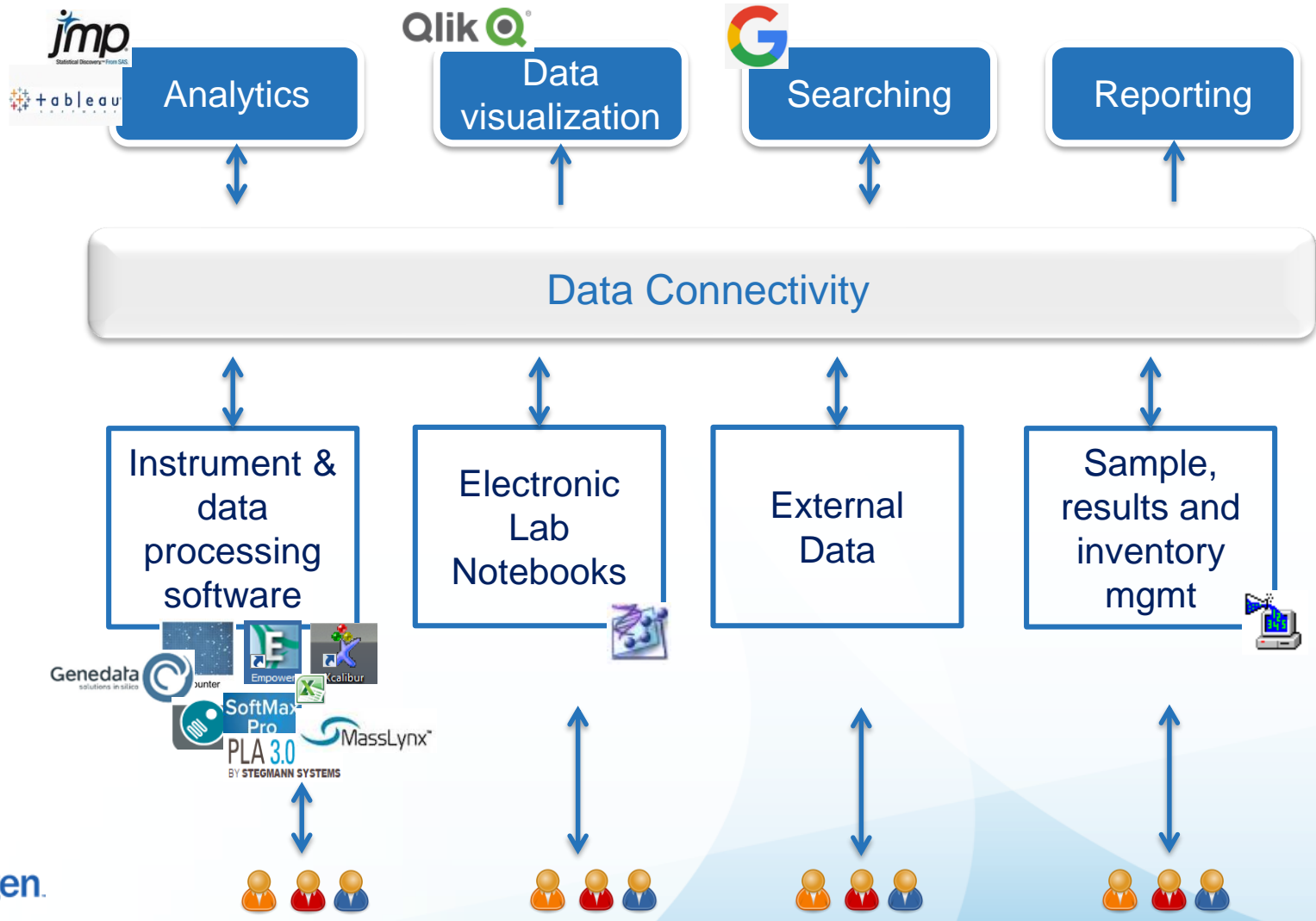
Data is our fundamental asset..

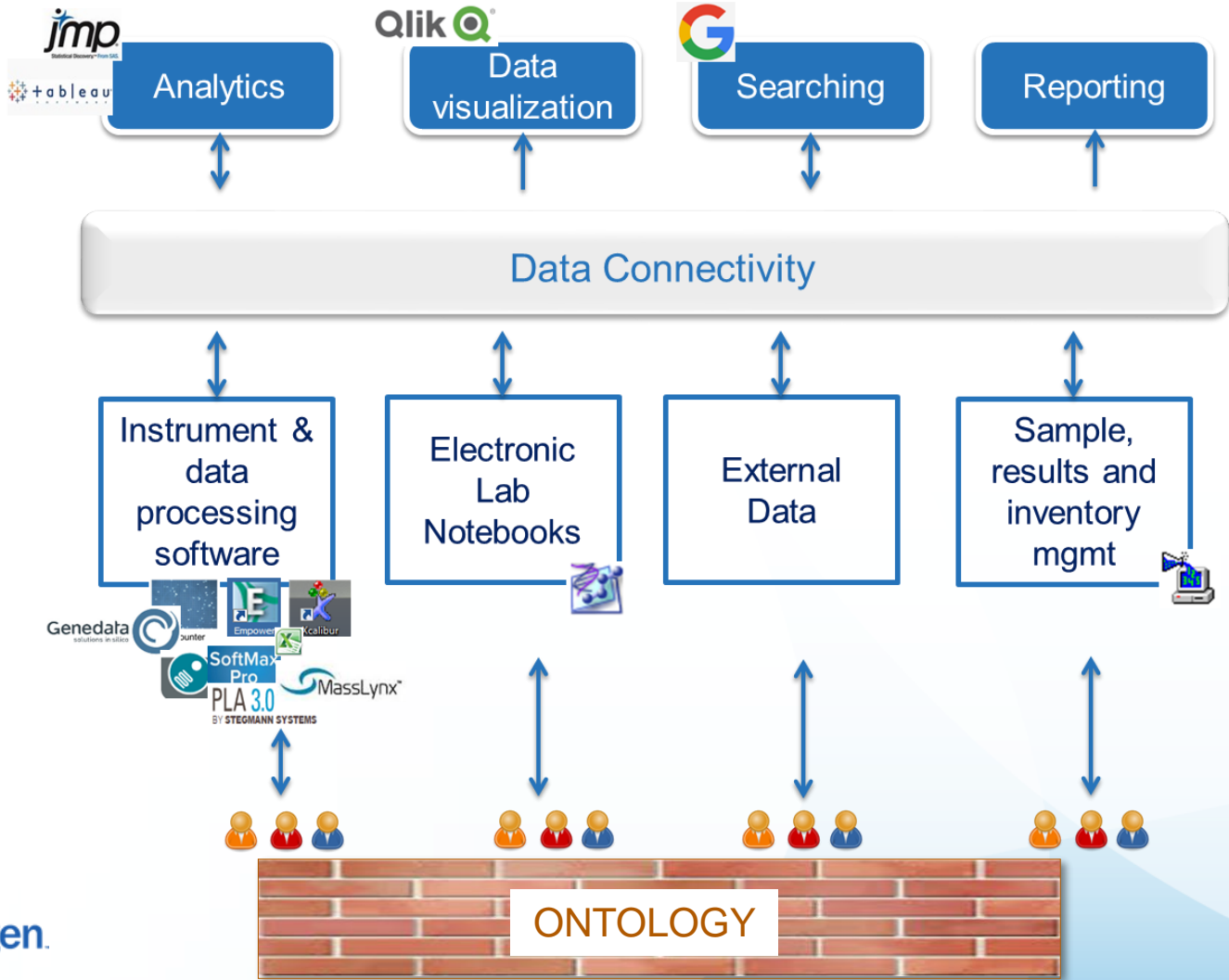
...We must utilize it to make informed decisions

But first...

....We need it to be easily accessible and consumable

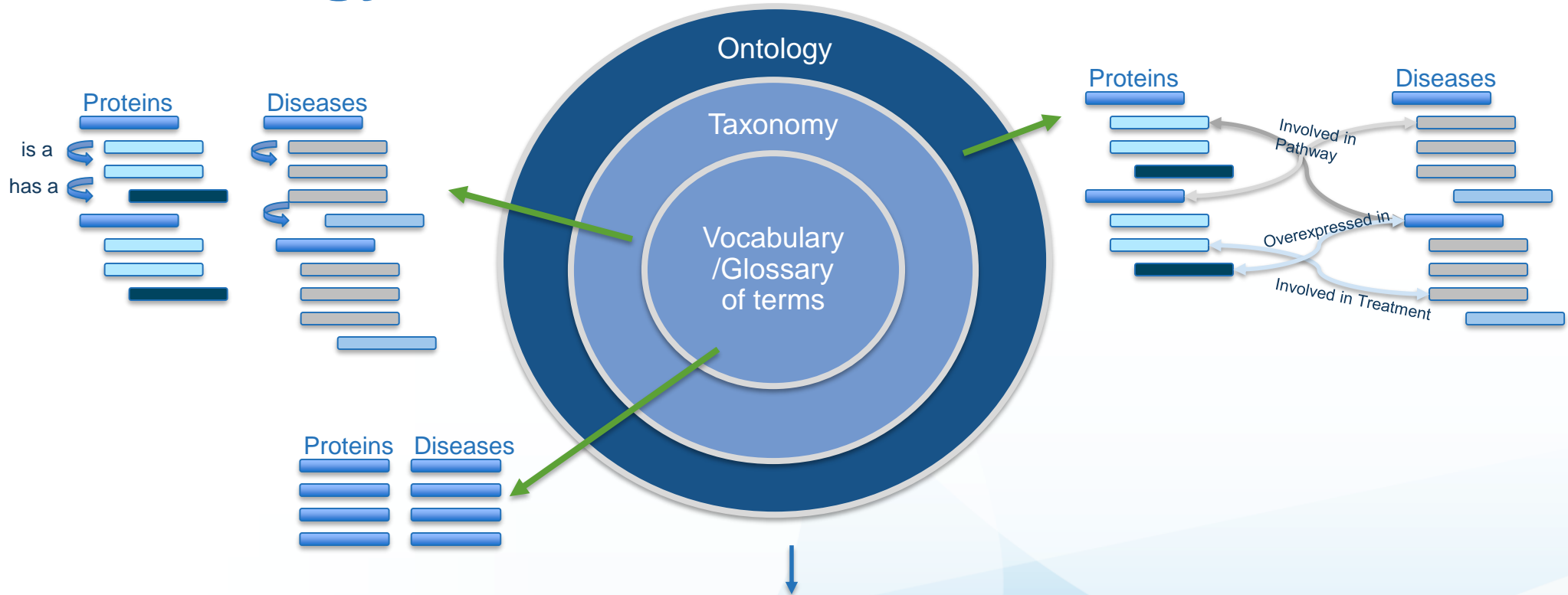






To build a successful data management strategy, we must understand our processes, data inputs/outputs, and systems.

An established ontology is the foundation of any data strategy



Machine readable & Findable Accessible Interoperable Reusable

Structured Authoring: Recipe Driven Development & Qualification (RDQ)



Opportunity



- Disparate experiment, data and reporting systems
- Lack of standardized approach to qualification
- No easy access to qualification data
- Manual execution of method steps



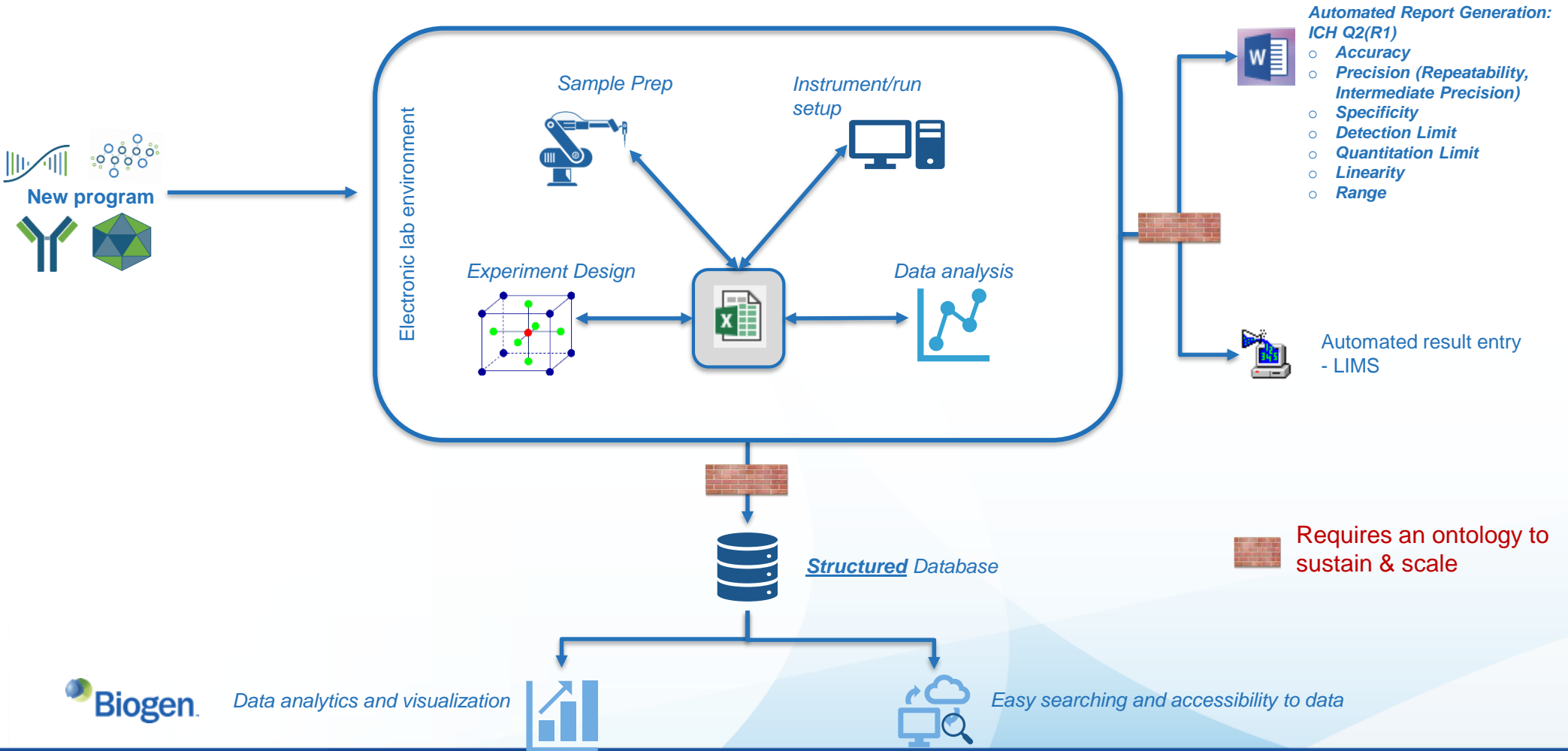
Vision: Recipe driven approach



Overview: Capillary Electrophoresis - Sodium Dodecyl Sulfate (CE-SDS)

- Purity & heterogeneity
- Separation based on molecular weight and hydrodynamic radius
- Platform release method

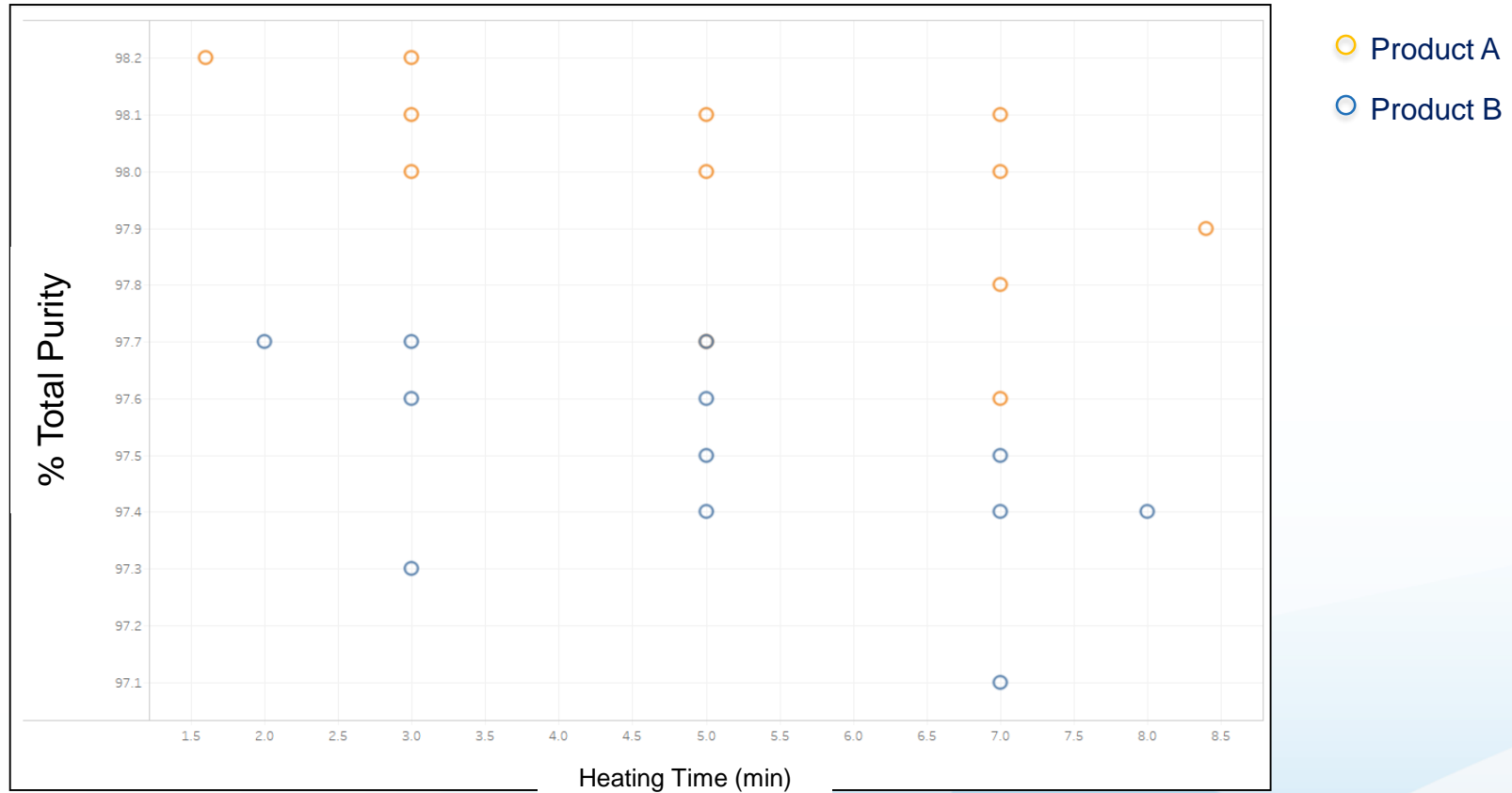
The RDQ Framework



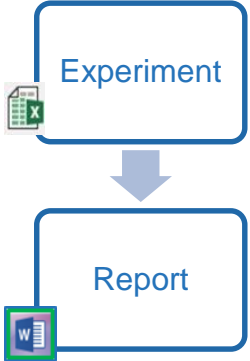
Impact of the RDQ framework

- ✓ This framework enables a **systematic & standardized** approach to qualification while optimizing user interaction and method execution
- ✓ Successful proof-of-concept with CE-SDS helped us realize efficiency gains of up to 66%

Monitor key attributes from multiple qualification experiments across multiple programs



Automated Report Generation



File Home Insert Draw Page Layout Formulas Data Review View Developer Help CE-SDS Qualification Share Comments

Study Empower Clear Study | DOE Empower DesignExpert Clear Study | Performance Empower Clear Study | Generate Clear Report All

Buffer and Stability Study | Robustness DOE | Performance Study | Results Generation

16

A B C D E F G H I J K L M N O P Q R S T U V W X

1 **Assay Performance (Intermediate Precision, Repeatability, and LOQ)**

2 **Experiment Description**

3 The purpose of this experiment is to use predetermined assay parameters to confirm assay performance. Assay performance is determined by intermediate precision, repeatability and determination of assay LOQ. This experiment executed repeatability and LOQ experiments while collecting data for stress sample evaluation and intermediate precision.

7 **Repeatability:**

Sample ID	% Area (Pre-Peak)	% Area (Intact Protein)
TEST_pH9_FormBuffer_Repeat 1	1.51	97.51
TEST_pH9_FormBuffer_Repeat 2	1.79	95.33
TEST_pH9_FormBuffer_Repeat 3	1.51	97.51
TEST_pH9_FormBuffer_Repeat 4	1.79	95.33
TEST_pH9_FormBuffer_Repeat 5	1.51	97.51
TEST_pH9_FormBuffer_Repeat 6	1.79	95.33
Average	1.65	96.42
Standard Deviation	0.154	1.198
% RSD	9.3	1.2

19 **Intermediate Precision:**

Sample ID	Analyst/ Capillary 1		Analyst/ Capillary 2		QC Analyst/ Capillary 3		Overall	
	% Area (Pre-Peak)	% Area (Intact Protein)	% Area (Pre-Peak)	% Area (Intact Protein)	% Area (Pre-Peak)	% Area (Intact Protein)	% Area (Pre-Peak)	% Area (Intact Protein)
TEST_IP_1	1.51	97.51						
TEST_IP_2	1.79	95.33						
TEST_IP_3	1.51	97.51						
TEST_IP_4	1.79	95.33						
TEST_IP_5	1.51	97.51						
TEST_IP_6	1.79	95.33						
Average	1.65	96.42	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	1.65	96.42
Standard Deviation	0.154	1.198	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	0.154	1.198
% RSD	9.3	1.2	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	9.3	1.2

32 **LOQ Determination:**

Sample ID	Concentration	Area (Intact Protein)	* Add Signal/Noise	LOQ Summary Calculations	DF 125 (0.8%)	DF 250 (0.4%)
TEST_pH9_FormBuffer_LOQ_125DF 1-1	0.8%	6070		Average	50963.00	41219.00
TEST_pH9_FormBuffer_LOQ_125DF 1-2	0.8%	31475		Standard	16877.10	16877.10
TEST_pH9_FormBuffer_LOQ_125DF 1-3	0.8%	6070		% RSD	33.12	40.94
TEST_pH9_FormBuffer_LOQ_250DF 1-1	0.4%	31475		S/N * (Average)		
TEST_pH9_FormBuffer_LOQ_250DF 1-2	0.4%	6070		Assay LOQ	0.40%	
TEST_pH9_FormBuffer_LOQ_250DF 1-3	0.4%	31475				

41 **Stress Sample Evaluation:**

Impurity	Control, Unstressed Sample				Stressed Sample, ABC, 49C for 3 Weeks			
	Peak Name	Migration Time (m)	Rel. Migration Time	% Area (Intact Protein)	Peak Name	Migration Time (m)	Rel. Migration Time	% Area (Intact Protein)
1	Internal Marker	12.37			Internal Marker	12.37		
2		15.26	1.23	0.06		15.26	1.23	0.06
3		18.42	1.49	0.05		18.42	1.49	0.05

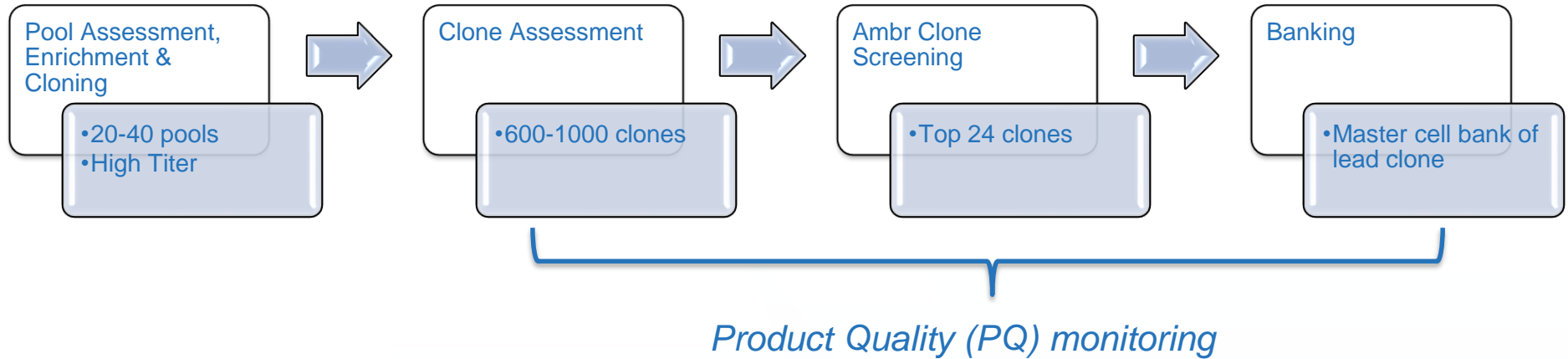
Method and Materials | Experiment_Buffer and Storage | Experiment_Robustness DOE | **Experiment_Assay Performance** | Config



Utilizing data analytics to enable swift and informed decisions: Optimizing the Clone Selection process

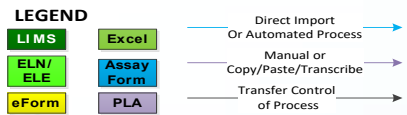
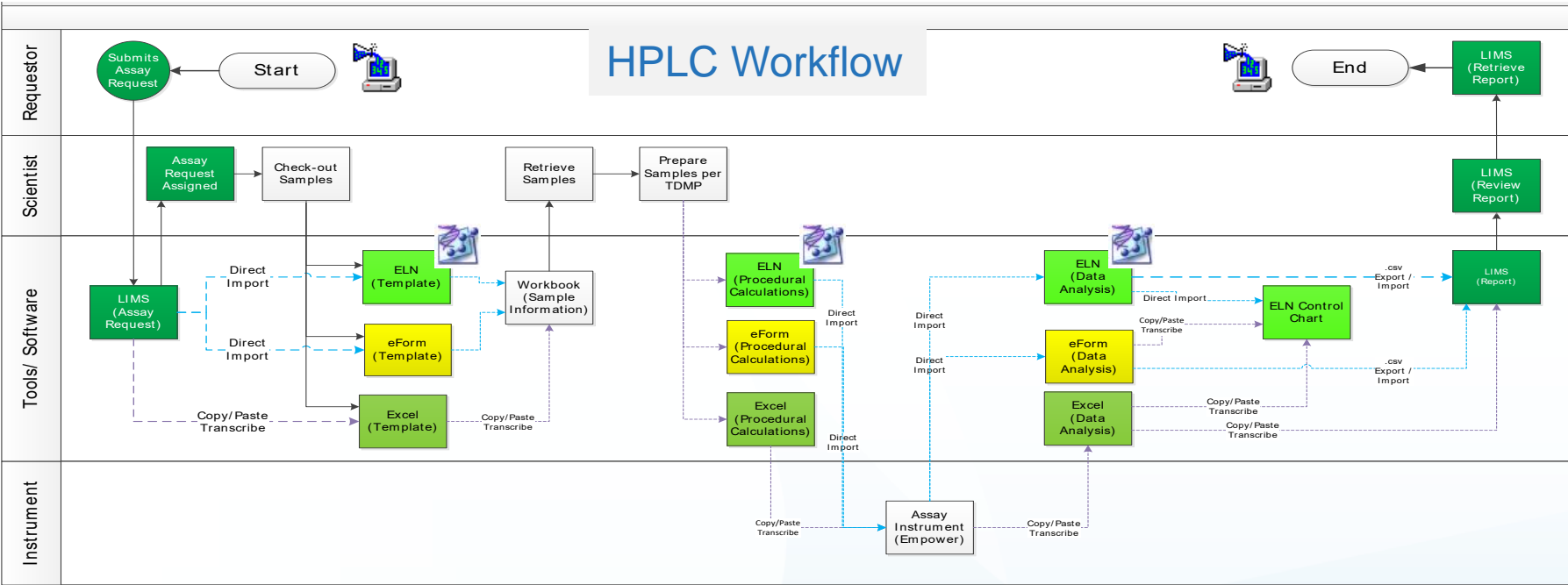


Cell Line Development & Analytical Testing Strategy

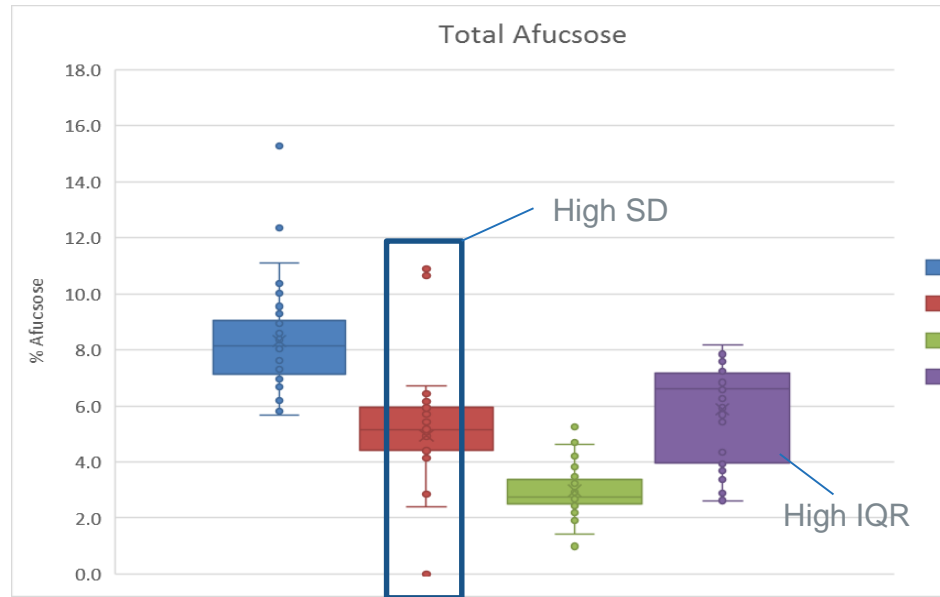


- Sequence variants
 - Glycoforms
 - Post-translational modifications
 - Charge variants
 - Purity & impurities
-
- Molecule-dependent
 - Cell line-dependent
 - Process-dependent

Understanding and Mapping the Process



Historical Data for 4 mAbs were Analyzed



Measures of Dispersion

$$\text{IQR} = Q_3 - Q_1$$

$$\text{MADM} = \text{median}(|x_i - \text{median}(x)|)$$

Standard Deviation (σ)

$$\text{CV} = \frac{\sigma}{\mu}$$

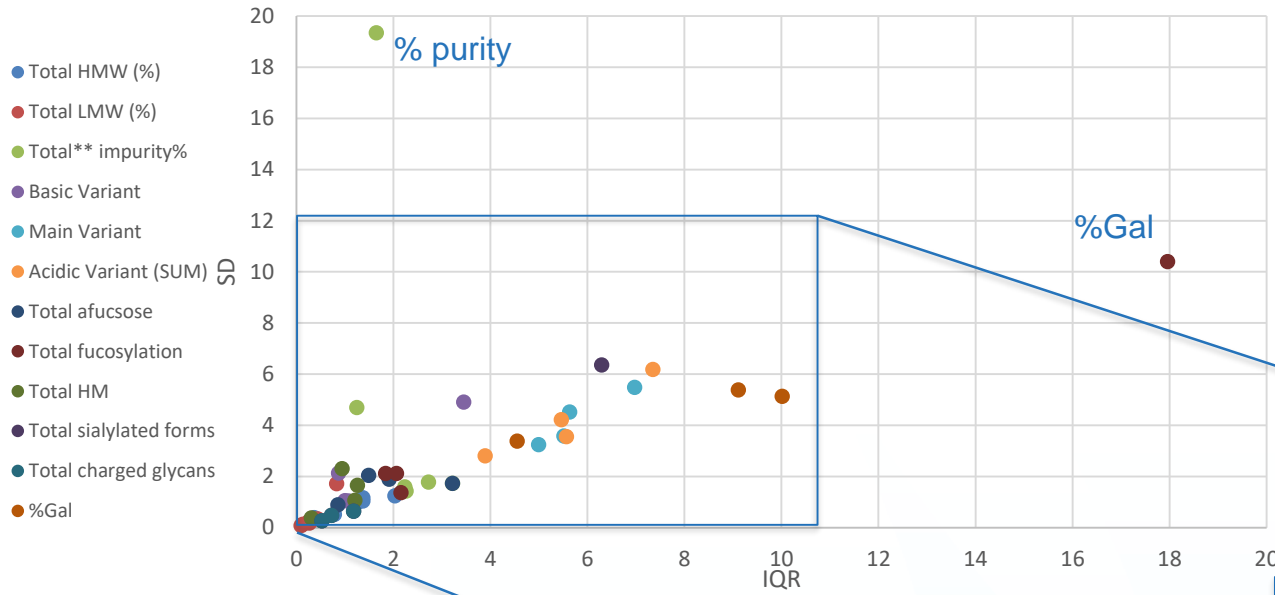
IQR correlated to MADM ($R^2 > 0.8$)

CV is scale invariant and skews high for sets with lower mean values

IQR, SD were used for analyses

- Assay data with higher IQR or SD indicate greater variability of the quality attribute (QA) tested (provided that assay itself has low variability)
- Higher variability QAs are valuable as they provide stronger differentiation of cell lines during screening

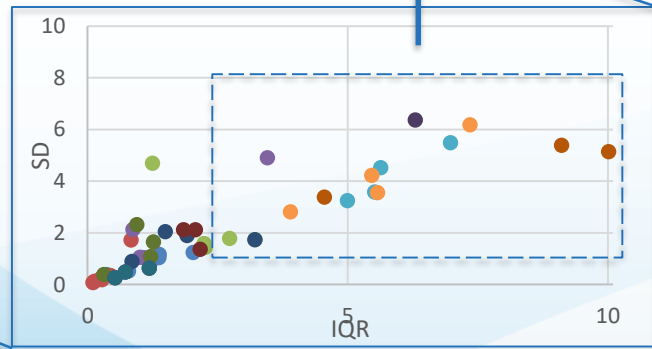
Assessment of Product Quality Results





PQs with the highest IQR or SD:

1. %Gal
2. %Acidic variant
3. %Main variant
4. %Purity
5. %Afucosylation
6. %Sialylation

- Data from top 36 clones were analyzed from 4 mAbs
- Assays located in the top right quadrant (N-glycan, GXII charge variant) provide more data for clone rejection than those in the bottom left (e.g. SEC)



Outcome

- ✓ Eliminated 2 of 6 assays needed to inform clone selection
- ✓ Enabled quicker selection of optimal clones
 - Automate selection process 
 - Predict top clones 

Conclusion

- Process mapping and understanding is a key first step in establishing a data & knowledge management strategy
- Ontologies are critical in ensuring a sustainable or scalable strategy
- Agile and well scoped projects allow us to easily demonstrate value of harnessing our data, paving the way for an enterprise level initiative

Preparing for the Future

- Establish and apply an ontology across our drug development process
- Utilize our data to continue growing our process knowledge and understanding
- Apply the recipe driven approach to our analytical methods and other outputs
- Align with initiatives within industry (E.g.: Knowledge-aided Assessment and Structured Applications (KASA) led by the FDA, Industry 4.0)



Recipe-driven filings & marketing applications!

Acknowledgements

Biogen

Analytical Development

Maxime Usdin, *Associate Scientist*
Naveed Hedayat, *Associate Scientist*
Andy Blum, *Scientist*
Devin Bailey, *Associate Scientist*
Dr. Peter Bryngelson, *Sr. Scientist*
Dr. Len Blackwell, *Principal Scientist*
Dr. Lou Coury, *Associate Director*
Dr. Bernice Yeung, *Director*
Dr. Brian Fahie, *Sr. Director*

Cell Line Development

James Lambropoulos, *Sr Engineer I*
Dr. Christina Alves, *Principal Scientist*

Global Process Analytics

Dr. Saly Romero-Torres, *Associate Director*
Dr. Rob Guenard, *Sr. Director*

Regulatory

Bem Atsma, *Associate Director*



THANK YOU!

QUESTIONS?

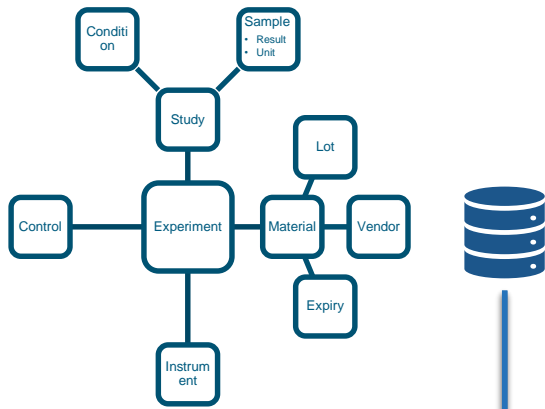




BACKUP



Database



Standardized and automated capture of critical aspects of the experiments

- ❖ Sample
- ❖ Results
- ❖ Materials
- ❖ Controls
- ❖ Instrument parameters



Track attributes across phases of development & multiple programs enabling **data-driven insights and decisions**

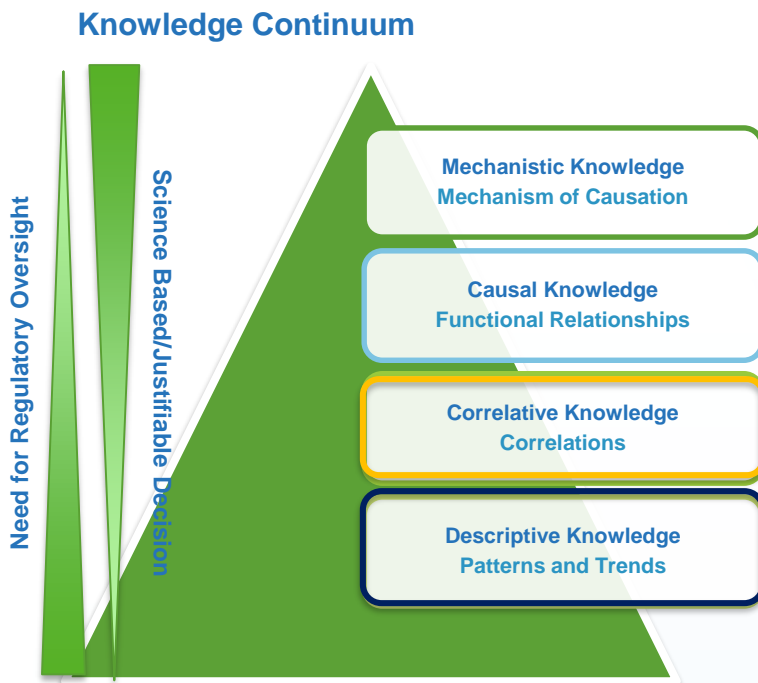


Cloud-based development of robust search capabilities enabling **easy accessibility** to data & **easy sharing** of data with our stakeholders (Quality etc.)

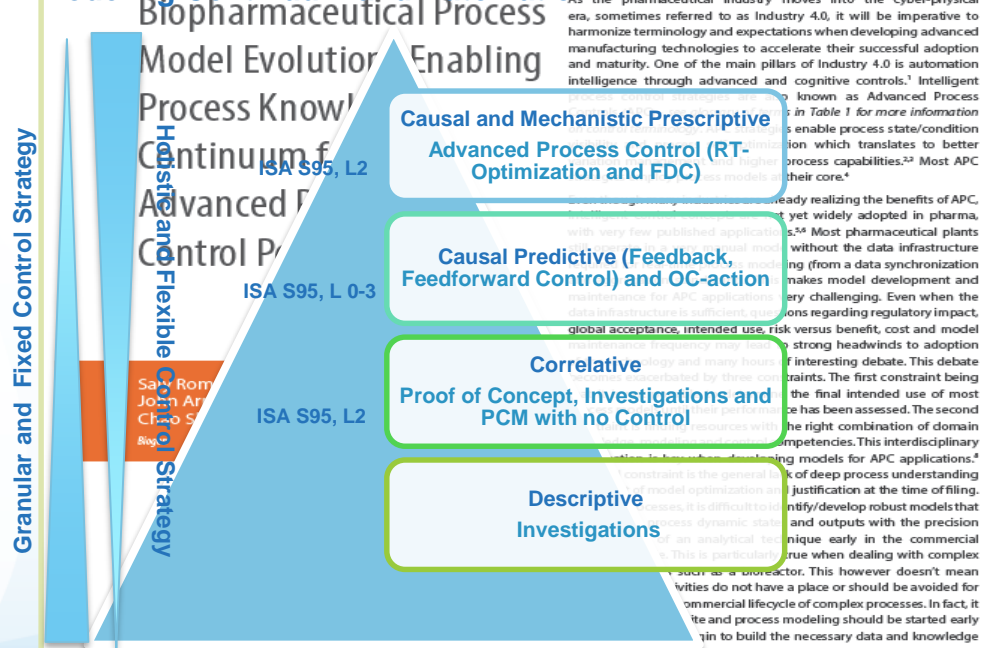


Modeling Philosophy

Modeling activities are tied to the knowledge continuum, hence just like knowledge, process models and applicability within (and outside) the automation architecture will evolve with time and experience.



Modeling Continuum and Automation



Adapted from Dr. Ajaz S. Hussain, 5th EGA Symposium on Biosimilars, London

Biopharmaceutical Process Model Evolution- Enabling Process Knowledge Continuum from an Advanced Process Control Perspective
Romero-Torres et.al.

American Pharmaceutical Review

Modeling Regulatory Expectations

FDA PAT Guidance

“Comprehensive statistical and risk analyses of the process are generally necessary to assess the reliability of predictive mathematical relationships. Based on the estimated risk, a simple correlation function may need further support or justification, such as a mechanistic explanation of causal links among the process, material measurements, and target quality specifications.”

ICH Q11

“A scientifically justified model can enable a prediction of quality, and can be used to support the extrapolation of operating conditions across multiple scales and equipment.”

USP 40 1039

“The underlying rationale for variable selection in chemometrics is twofold. First, certain predictor variables could be irrelevant to the intended purpose, and ideally, these variables should be minimized or removed from the model. For example, only specific wavelengths of the whole range generated by the spectrometer may bear information relevant to the variation in levels of the response variable. Variable selection for this purpose should be based on first principles and experience. Inclusion of unrelated predictors, such as irrelevant spectral regions, could potentially degrade performance of the model. A smaller number of variables from the preprocessed data can be used to achieve superior performance, such as accuracy, precision, and robustness.”

- ✓ A data-driven approach alone (without any scientific knowledge) is not apt for process control or quality related decisions



Objectives



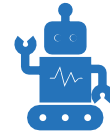
Establish..

templated platforms
standardizing our
practices **enabling
faster cycle times** for
new programs without
compromising data
quality (especially
partner programs)



Enable..

easy accessibility to
development and
qualification data



Automate...

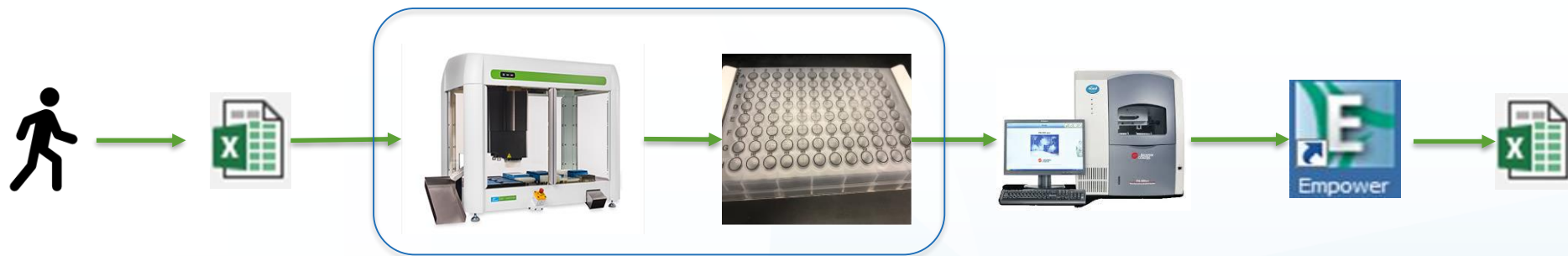
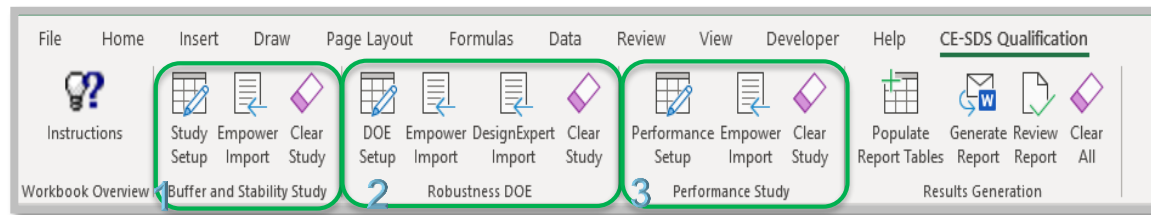
generation of
qualification reports
reducing analyst time by
~70%



Optimize...

execution of methods via
automation tools
promoting effective
transfers to results
generation

CE-SDS Qualification Workflow: Experiment Design and Execution



 Automation of reagent, sample prep & sample transfer reduced overall manual tasks by **~50%**