July 2018
Dial-a-Molecule Annual Meeting
i-Hub Imperial College
H. Dubina

Enabling Data-Rich Experimentation and Associated Data Analysis

METTLER TOLEDO

Agenda

1	**Needs and Challenges – Chemical Development**
2	**Industry trends in the area of Knowledge Management**
3	**Data Analytics – during the experiment, single experiment and multiple experiments**

Every year we try to engage with scientists in chemical development to more fully understand their challenges and needs

1 Roundtable

8 Info Days

4 Online Seminars

~7200 Meetings

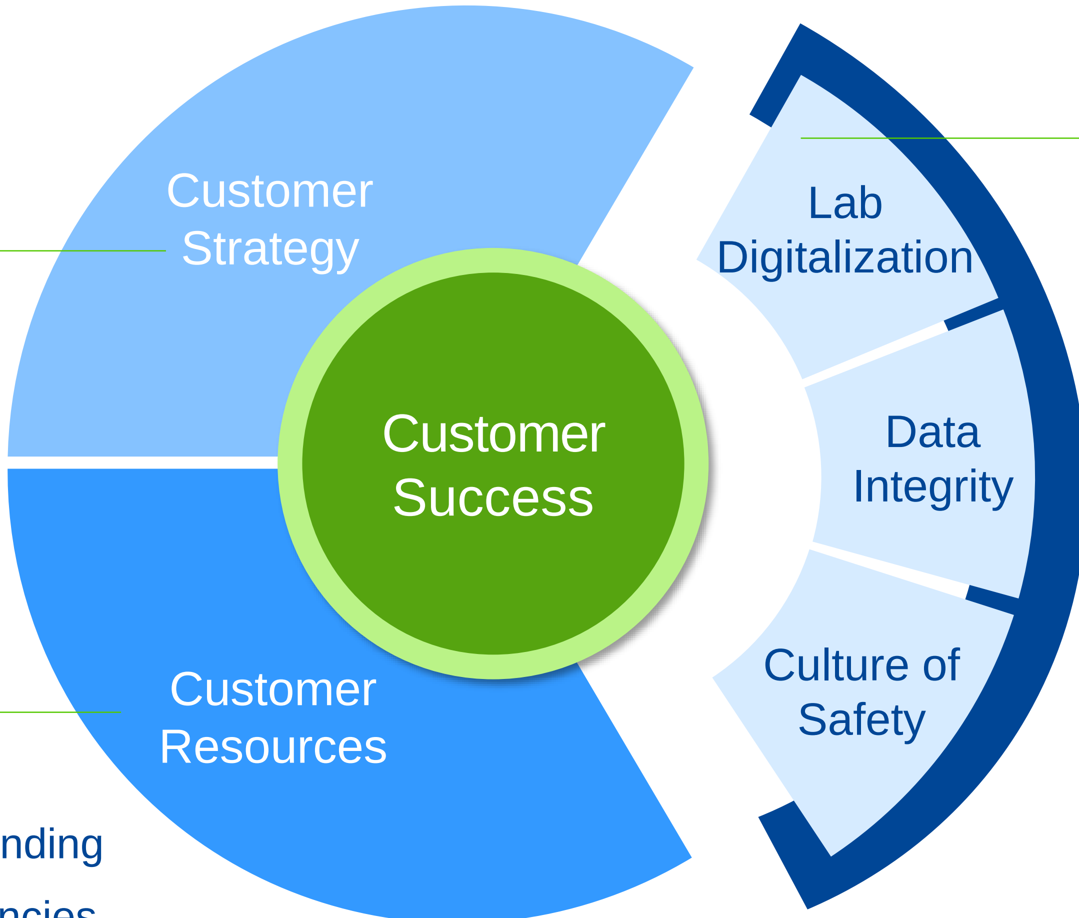
Summarized Common Objectives

- Increase **efficiency** and **effectiveness** of every chemist and engineer
- Develop **well understood** processes that are **profitable**, **green**, and **safe**
- **Increase velocity** through the development pipeline
- Launch products with the **best science** at the **lowest cost**

Needs and Challenges

- Department Objectives
- Key Initiatives
- Project Goals

- Human Capital
- Budget and Funding
- Core Competencies



Needs and Challenges

- Description of a critical need or challenge that can significantly impact scientific or business success

Common Needs and Challenges

Modernize Synthesis

Insight for Every Reaction

Take Control of Particles

Transform Organizational Productivity

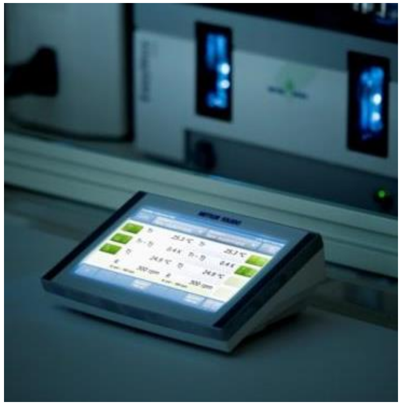
Develop a Culture of Safety

Deliver from Lab to Plant

Provide Key Core Competencies

Improve usability and lab safety

- More time for investigations
- Fast adoption and high utilization
- Increase user safety in lab

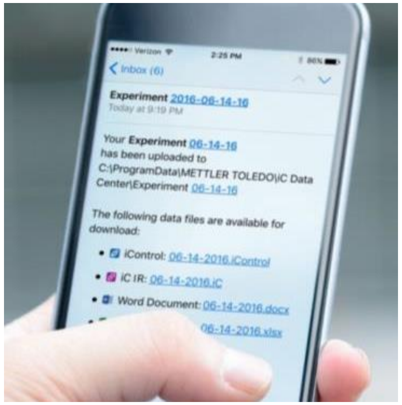


Enable personal productivity

- 100% capture of relevant data
- 24/7 experimental planning
- Reproducible recipes and results

Deliver information with every experiment

- Data-rich experimentation
- Improved process understanding
- Support for scale up and tech transfer



Support knowledge management

- Optimized workflow to ELN
- Shared results for every project
- Searchable and standardized data

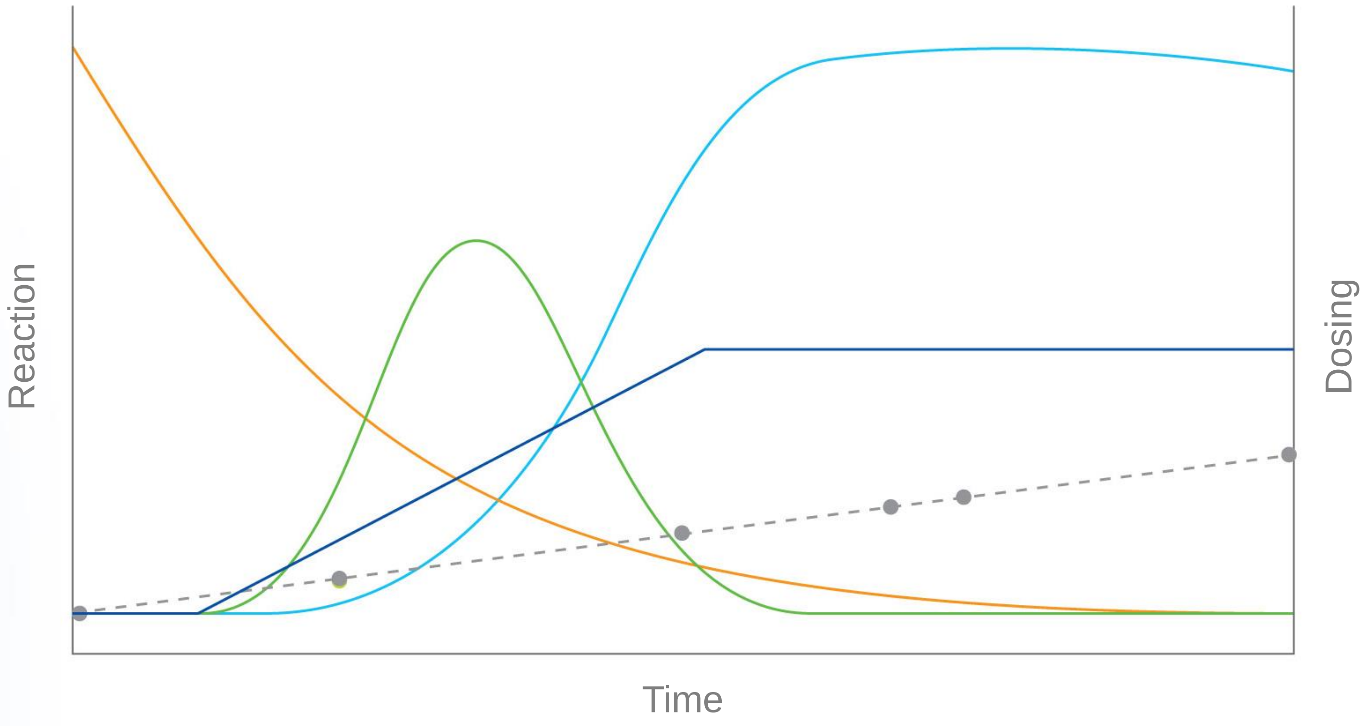
Drivers for increasing utilization of PAT tools

- **Ease of Use**
  - Connection and setup
  - Utility burden
  - Footprint
  - Laptop requirement
  - Manual interventions

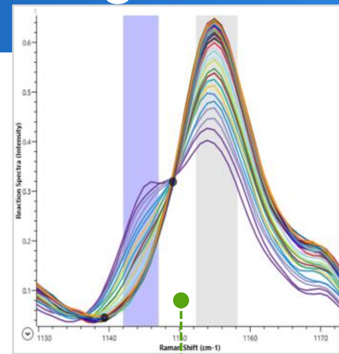
- **Reliability**
  - Uptime
  - Calibration (transfer)
  - Verification
  - Lab-to-plant comparison
  - Process oriented sensors
  - Cleaning
  - Fundamental Robustness
  - Lower TOC

- **Value of Information**
  - CQA or CPP
  - Accuracy
  - Sensitivity
  - Resolution
  - Fit for Purpose
  - Data vs. Information

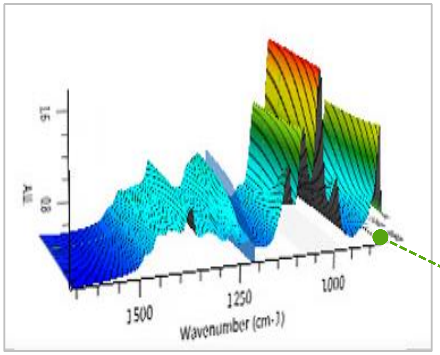
One software platform to easily extract, share, and store key findings



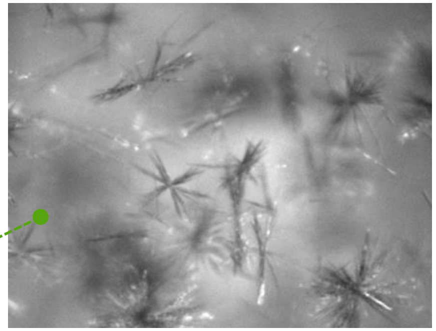
MT provides a complete integrated workflow



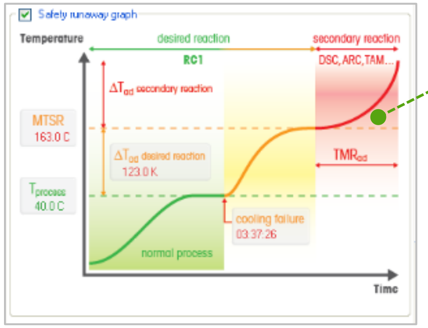
In Situ Raman Analysis



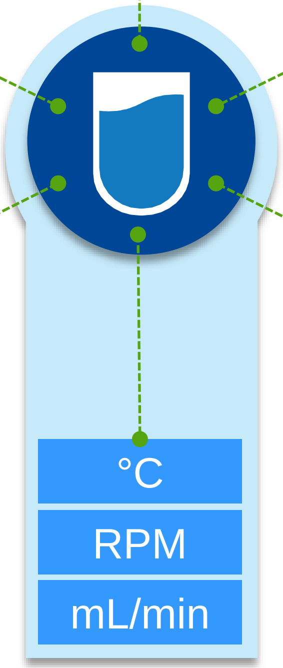
In Situ FTIR Analysis



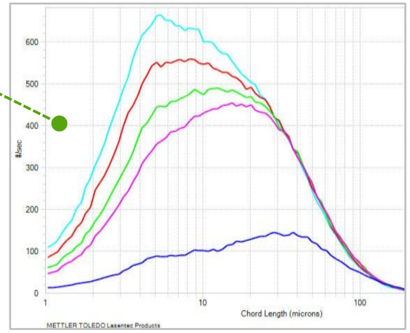
Inline Particle Microscopy



Reaction Calorimetry



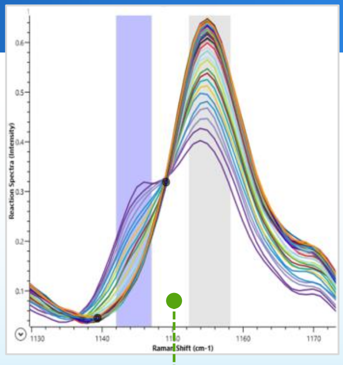
Automated Synthesis Reactors

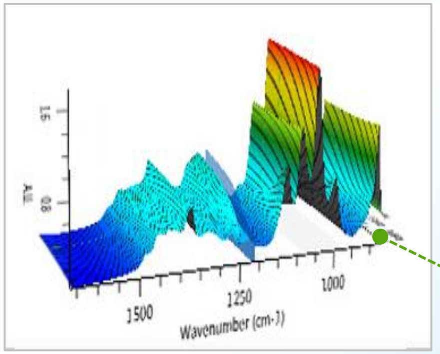


Inline Particle Size and Count

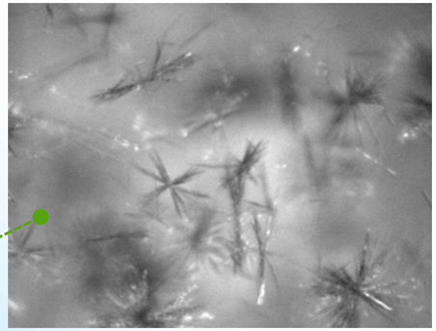
iC suite integrated everything into a single experiment

In Situ Raman Analysis

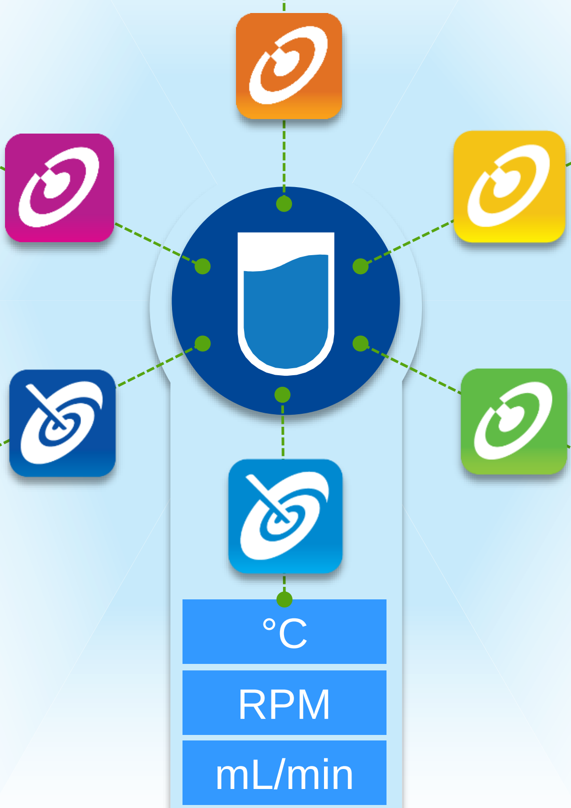


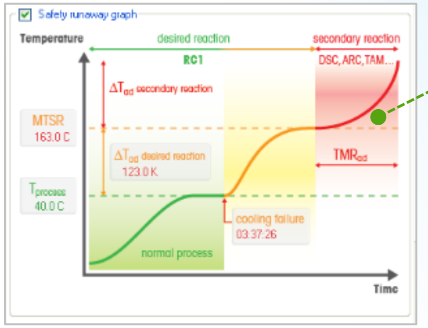


In Situ FTIR Analysis

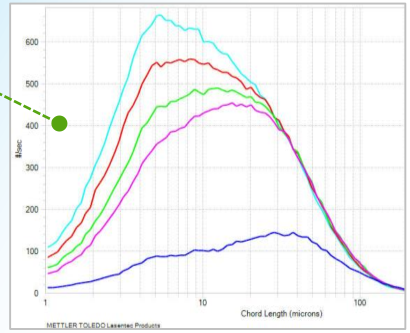


Inline Particle Microscopy





Reaction Calorimetry



Inline Particle Size and Count

Automated Synthesis Reactors

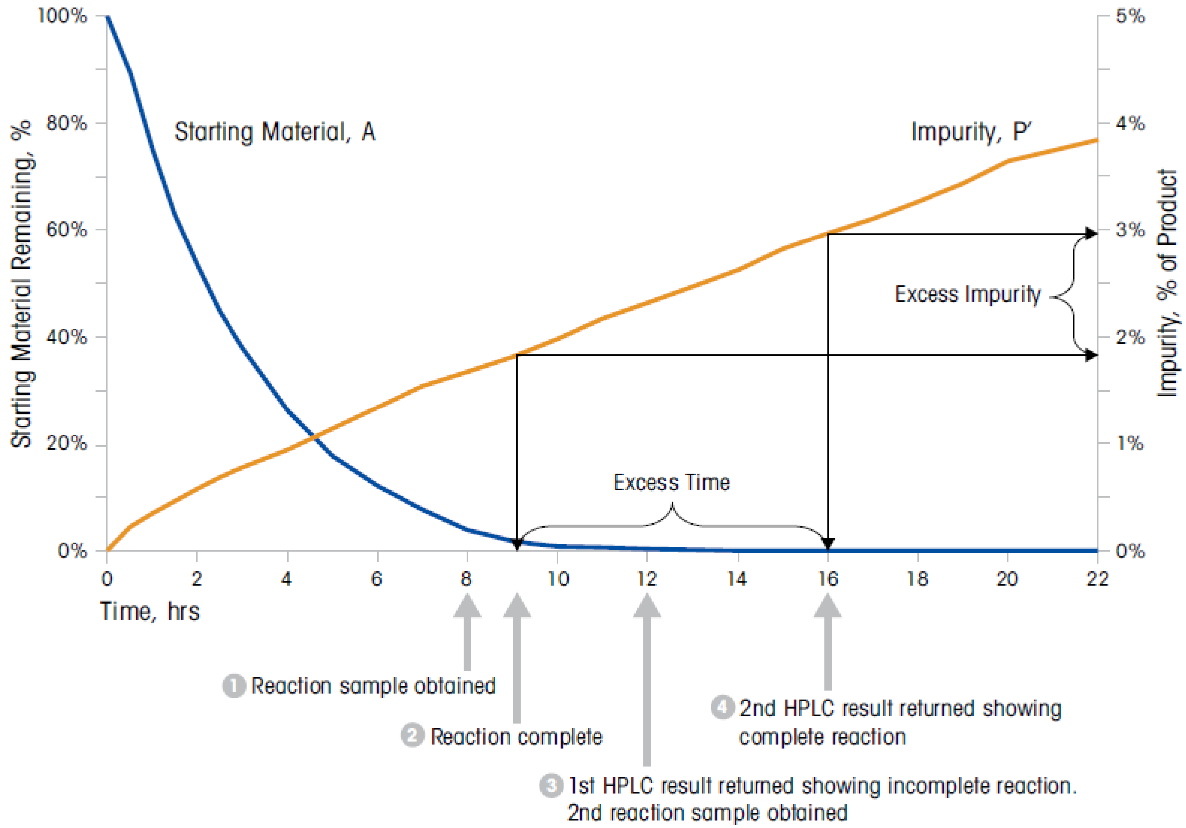
The use of Complementary Data Streams provide High Quality Insights

HPLC/IR Can be used together to highlight reaction events, or enhance each other

- MIR and HPLC/UPLC data are highly complementary
  - Together they cover a full dynamic range of sensitivity across an entire chemical reaction
  - Combines standard offline analytical technique and the standard online PAT technique
- Various use cases, but these are now common
  - Use online MIR as the trigger for sampling events – then confirm using HPLC/UPLC
  - Use MIR reaction profiles to pinpoint key reaction events – then collect samples around the critical points – enable data density in the right places
  - Use the offline analytical result to calibrate the MIR trend and get concentration data across an entire experiment

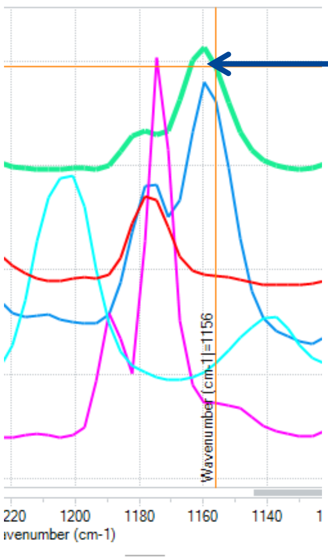
Online MIR used to target the correct sampling time for offline analysis

- Under a GMP process, can take significant turnaround time to receive analytical results
- If sample is mistimed, resulting delay can cause excess impurity to form in the reaction



Arani Chanda, Adrian M. Daly, David A. Foley, Mark A. LaPack, Samrat Mukherjee, John D. Orr, George L. Reid, III, Duncan R. Thompson, and Howard W. Ward, II, Industry Perspectives on Process Analytical Technology: Tools and Applications in API Development, Org. Process Res. Dev. 2015, 19, 63−83

Understanding the Formation of an Impurity

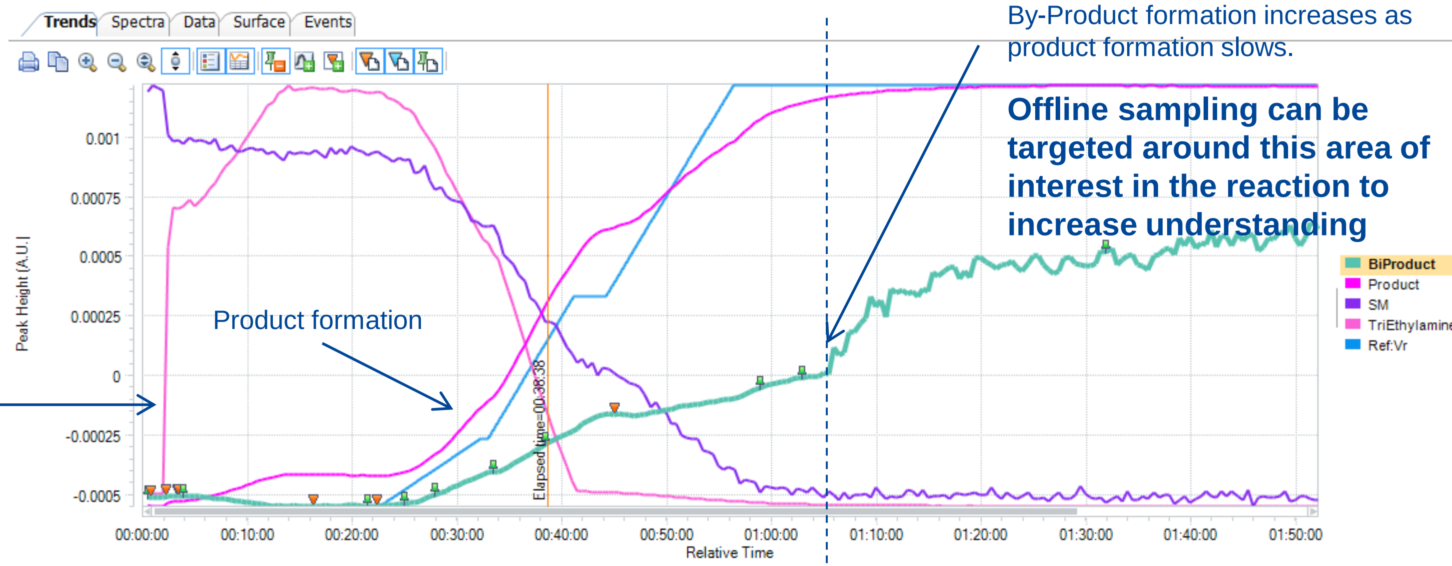


Impurity is seen in reaction spectra at 1163cm^{-1}

Profiling this peak provides a real time profile of impurity formation combined with other reaction profiles

By-Product formation increases as product formation slows.

Offline sampling can be targeted around this area of interest in the reaction to increase understanding

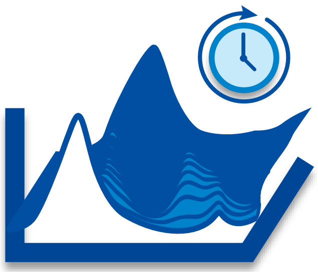


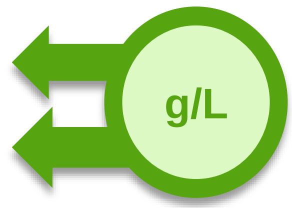
Product formation

Starting Material Added

Profiles suggest having no hold time after 1:05 would make reaction cleaner – real time end point detection

Simple Method to Transform Trends to Numbers





Continuous Reaction Monitoring

Trend Concentration Over Time

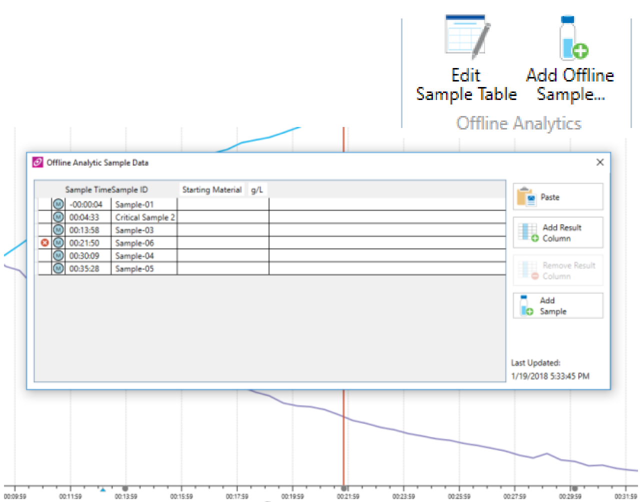
Quantify Trends

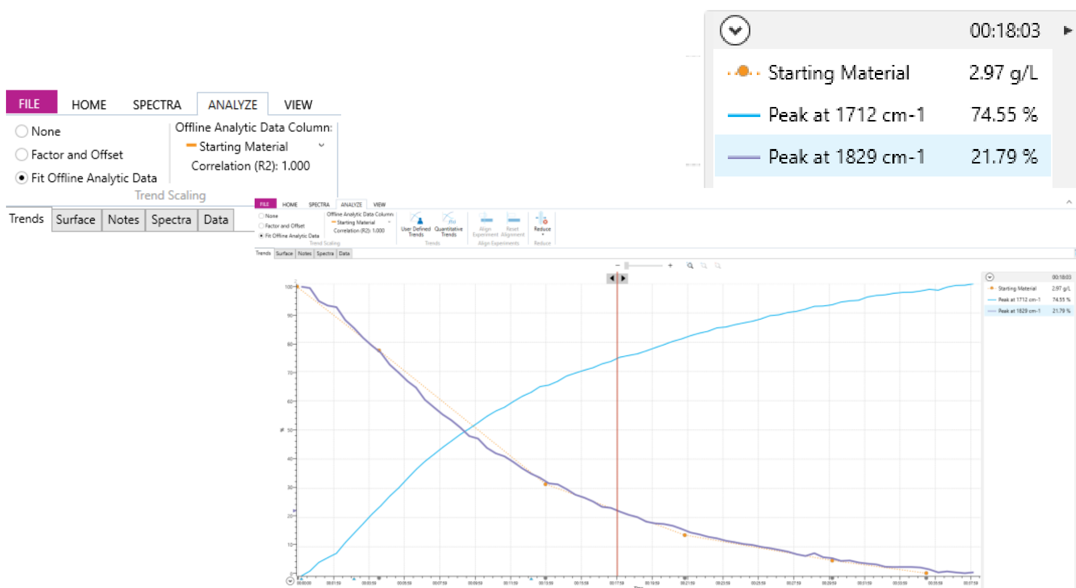
Beer's Law:

$$A = \varepsilon L C$$

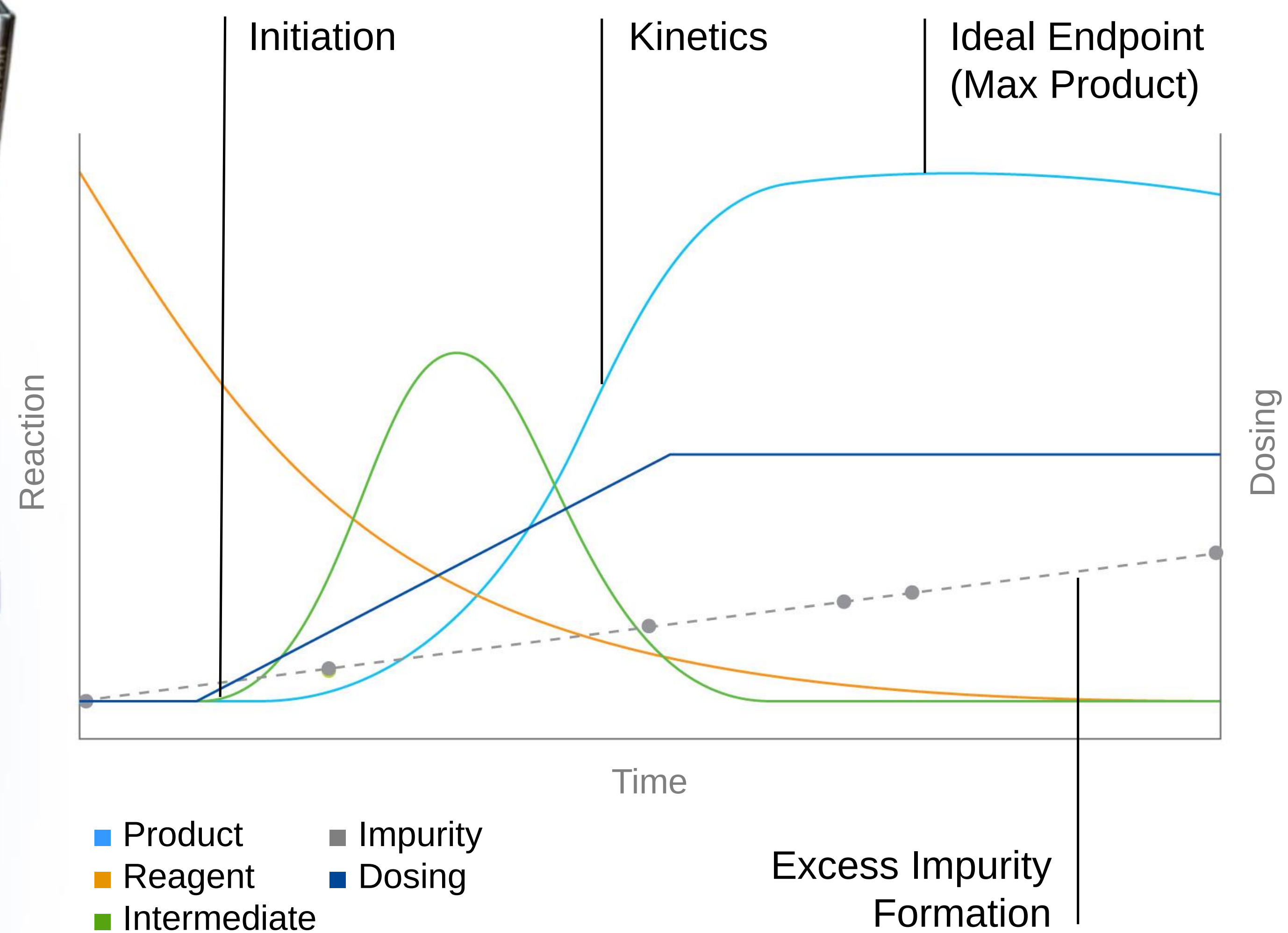
Therefore:

$$A \propto C$$

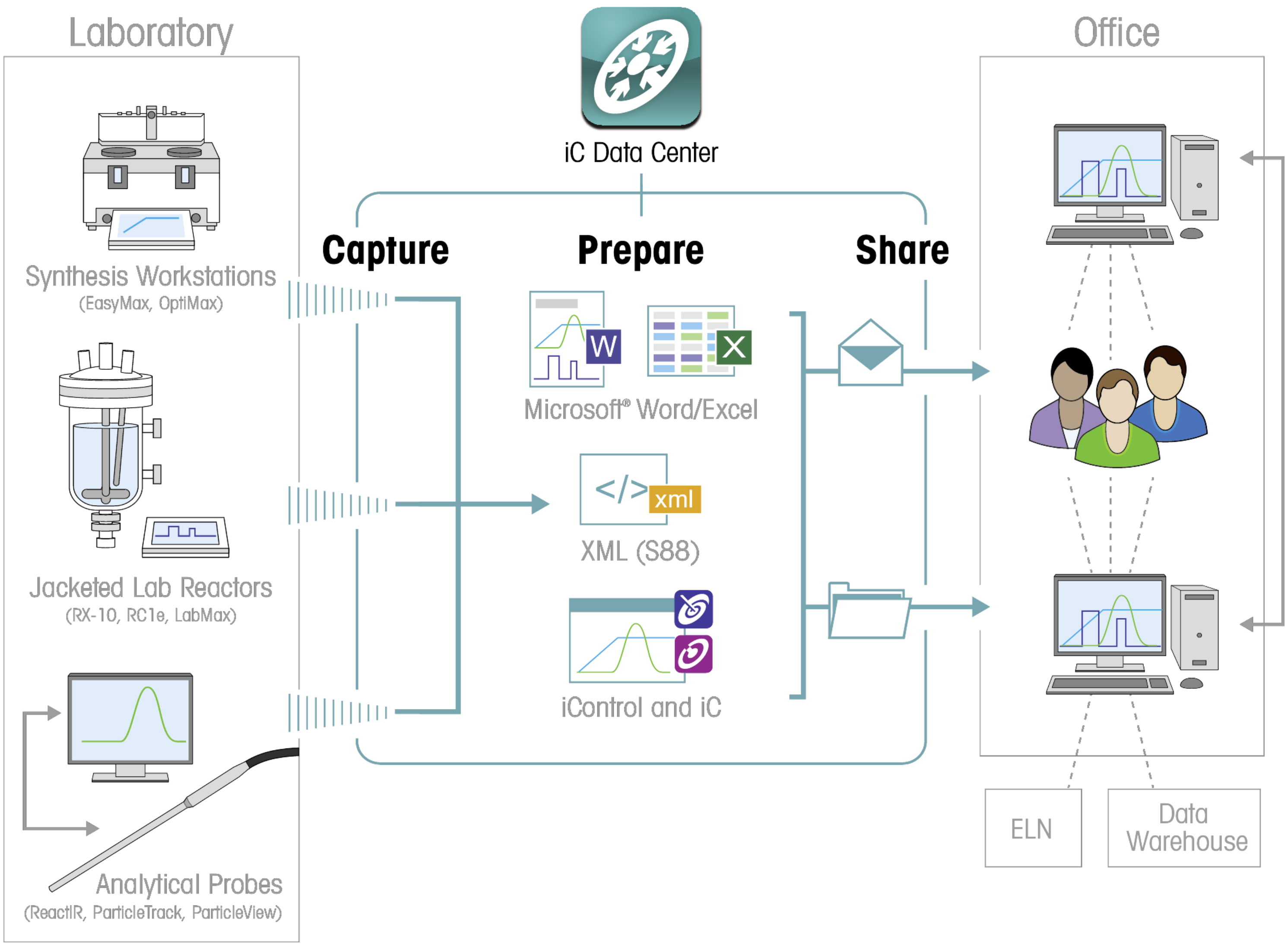




One software platform to easily extract, share, and store key findings



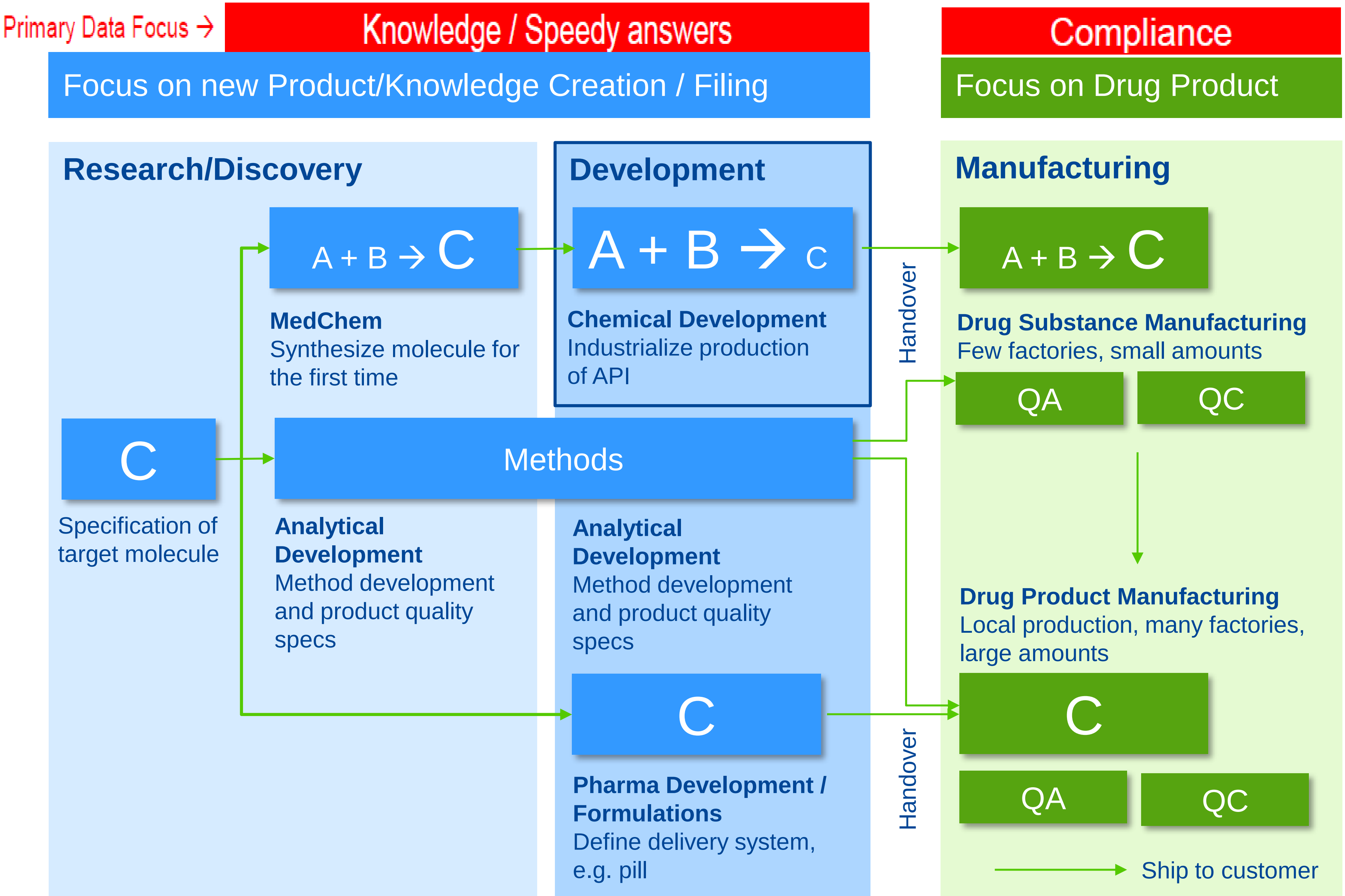
Collect data, generate reports, track utilization, and distribute learning



Agenda

1	**Needs and Challenges – Chemical Development**
2	**Industry trends in the area of Knowledge Management**
3	**Data Analytics – during the experiment, single experiment and multiple experiments**

Technical Drug Development Process (Small Molecules)

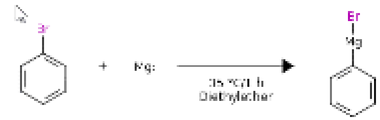


Workflow Focus

100% data and recipe capturing

Office

ELN: Experiment creation / design



Lab

Prepare reactor: fill in starting materials

A → 10g
B → 2g

Reaction: conversion of reagents to target API

A + B → C + D + E

Workup / Isolation: separate API from other materials

C + ~~D~~ + E

Purification: further increase purity of target API

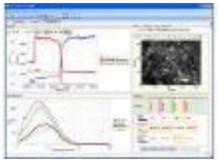
C + E

Final product: document yield (out of weight and quantitative analytical result, e.g. 20g * 90%)

C → 81%
E → 19%

Office

Analysis: extract needed data from iC SW products



Office

ELN
Document conclusion
Plan next step

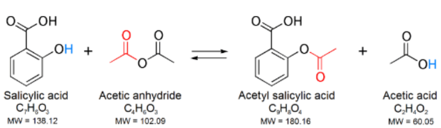
Merge of Process and Analytical data
Screen for optimal ..
- ... product conversion profile (IR, HPLC, ..)
- ... Impurity formation profile (HPLC, ..)

Determine endpoints based on above profiles

Compare conversions of multiple experiments
e.g. with Spotfire, etc.

Data generated in synthesis lab

Used Materials

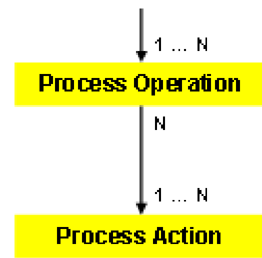


- Material names
- Actual amounts
- Lot no.
- Purity

Used Equipment

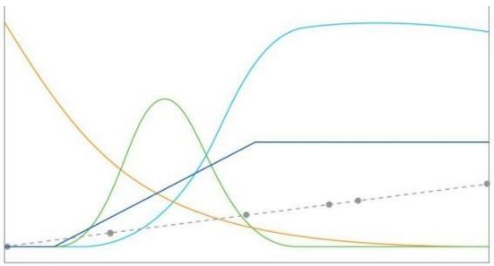
- Instrument serial #
- Reactor size
- Sensors and Actors: Sampling, Dosing, PAT, …

Executed Recipe



- Recipe steps: Heat, Dose, Filter, Wash, Dry, ..
- Parameters

Measured Data



- Process data: Temp, stirring, …
- Online analytics: IR, FBRM, Raman
- Offline analytics: HPLC, MS, NMR

Good Data and Knowledge Management

PROCESS & ANALYTICAL DATA

ANALYSIS & REPORT

- Better insights
- Speed up time to decision
- Limit non value activities

EFFICIENT FLOW

- Avoid experimental repetition
- Limit non-value activities
- Decrease human errors

STRUCTURED

- Efficient drug filing
- Efficient experiment recipe sharing
- Efficient use of predictive models
- Data mining/ analytics

The Six Levels of Good Data Management in Synthesis Labs

1 – 100 % of measured values captured

2 – Data captured in high quality

3 – Data captured in consistent data formats and with proper meta data

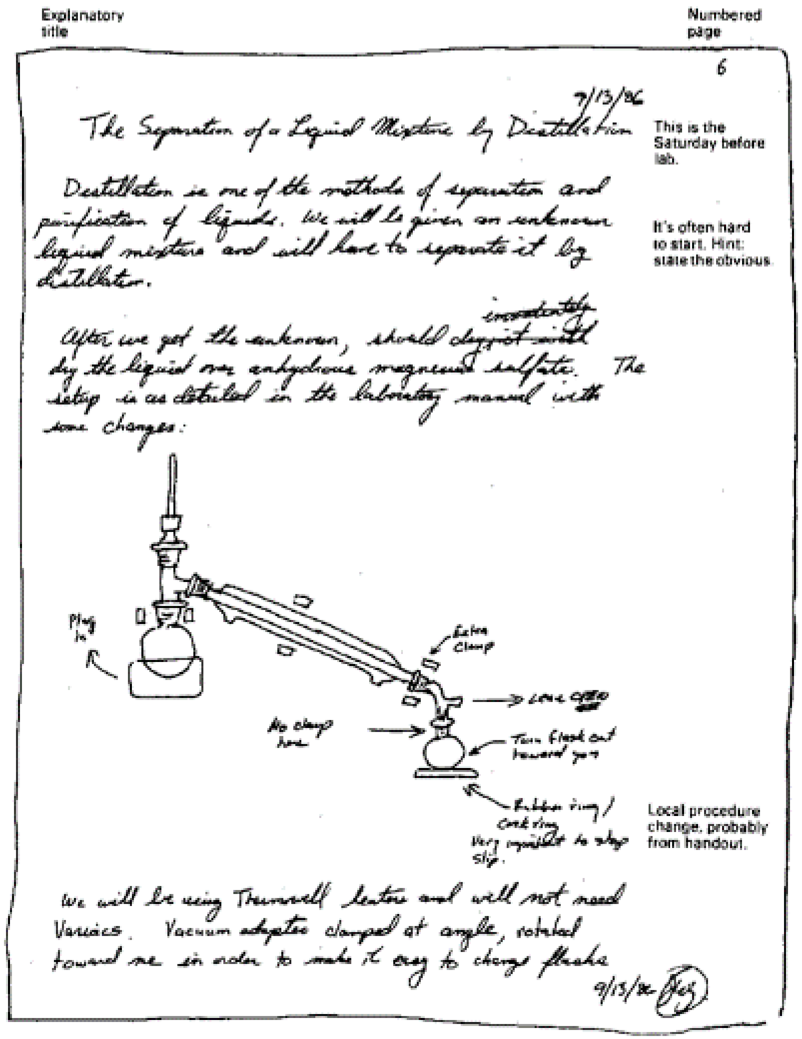
4 – Searchable data storage accessible for full team

Highest level observed in Pharma

5 – Complete and structured recipe data captured

6 – Data synchronization between systems in lab (LES) and office (ELN)

Challenge 1: All Measured Values Captured



Potential Issues If Not Present

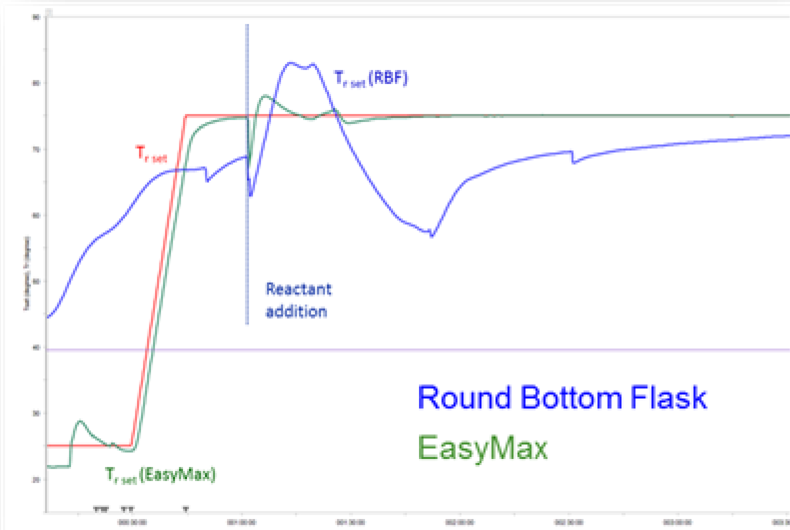
Manual data acquisition and reporting is unreliable and leads to low data quality.

Good Data Management

Information driven decisions based on visible parameter interdependencies.

Solution: Digital data acquisition system for every sensor

Challenge 2: Data Captured needs to be of High Quality



Potential Issues If Not Present

- Bad experimental reproducibility
- Need for multiple repetitions
- Risk of wrong conclusions

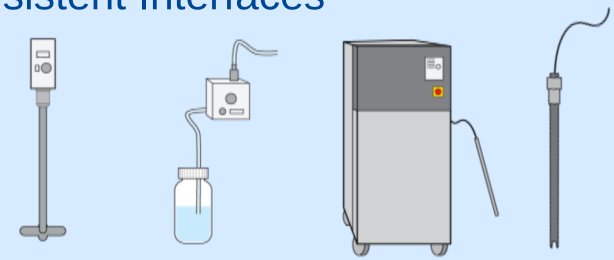
Good Data Management

Best use of statistical methods (DoE) for predictive modeling - Trust in Data

Solution: For temperature data → use of optimal temperature control

Challenge 3: Different Interfaces – Consistent Formats

Inconsistent Interfaces



Potential Issues If Not Present

- Wasted FTE time for data workup
- Experiment repetitions due to non-retrievable data.

Consistent Formats

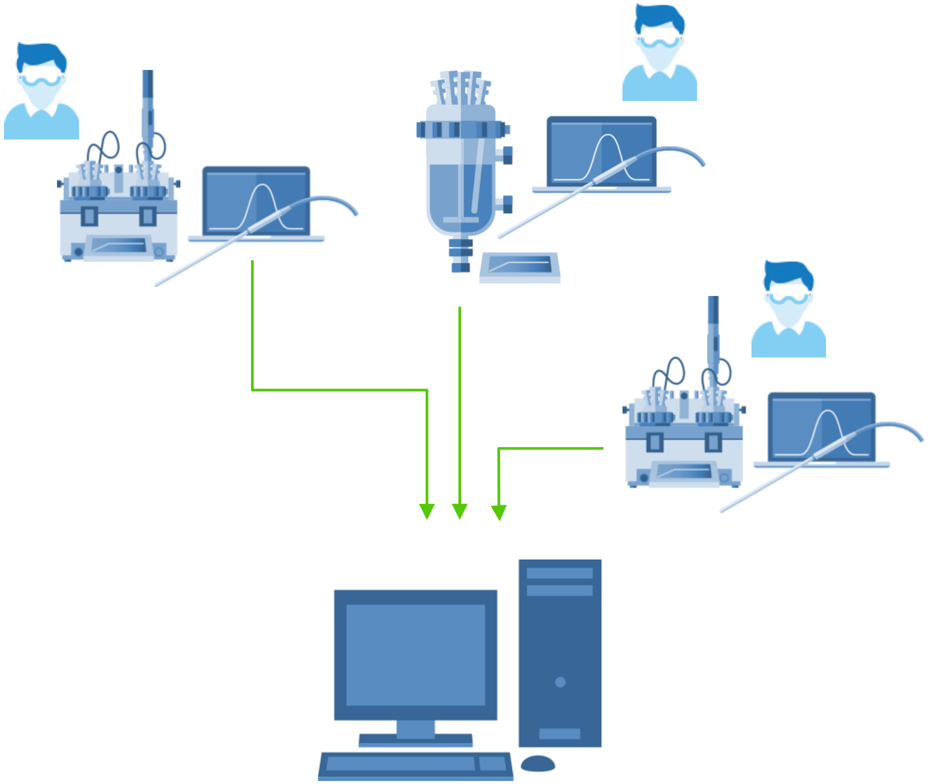
Microsoft® Word®/Excel®

Good Data Management

- Consistent data formats
- Simple storage and quick retrieval
- Improved searchability and comparability

Solution: Central data acquisition system with meta data enforcement

Challenge 4: Accessibility for Full Team



Potential Issues If Not Present

- Value of past experiments not leveraged
- Risk for lost experiments

Good Data Management

Creation of institutional knowledge

Solution: System to gather data at central location automatically

Challenge 5: Structured Recipe Captured

Stage

Crystallization
Decomposition
Drying of solid
Drying of solution
Extraction
Isolation
Reaction
Solution Preparation
Solvent Removal
Solvent Switch
Washing
Filtration

Operation

Agitate
Charge
Charge at Rate
Charge to pH
Distill
Dry
Filter
Homogenize
Inert
Isolate
Mill
Operator Message
Reflux Sample
Separate
Settle
Sieve
Temperature Adjust
Wash

Potential Issues If Not Present

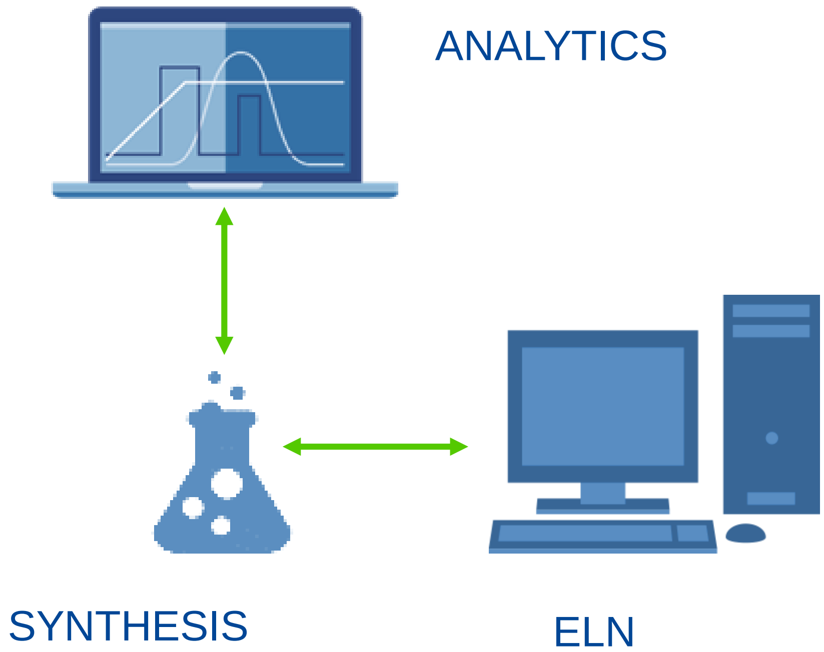
- Tedious recipe transfer processes
- Large obstacle for data mining

Good Data Management

- Easy recipe exchange
- Complete, accurate and reproducible recipe documentation

Solution: Recipe capturing for all stages and operations in your process

Challenge 6: ELN to Lab Data Synchronization



Potential Issues If Not Present

- Low trust in data stored in ELN.
- Cumbersome process to create reports for filing

Good Data Management

- No manual data transcription: FTE time saved / trustworthy data
- Many data mining opportunities

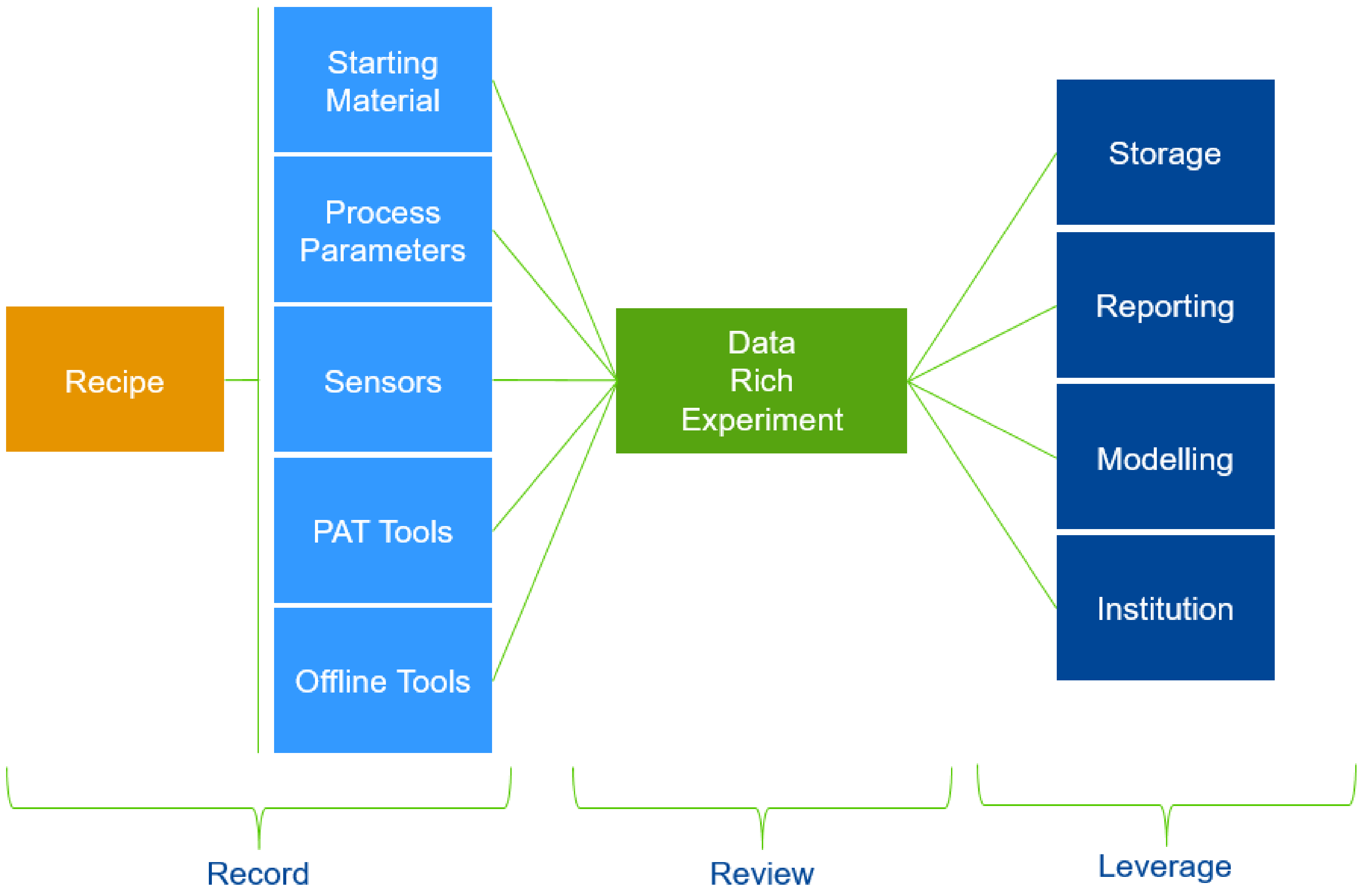
Solution: Implement interface between SW at the office and lab

Our organization is learning and advancing just because the data is so readily available the instruments are being used more and understood better. Funny how some see this as an “unanticipated” benefit.

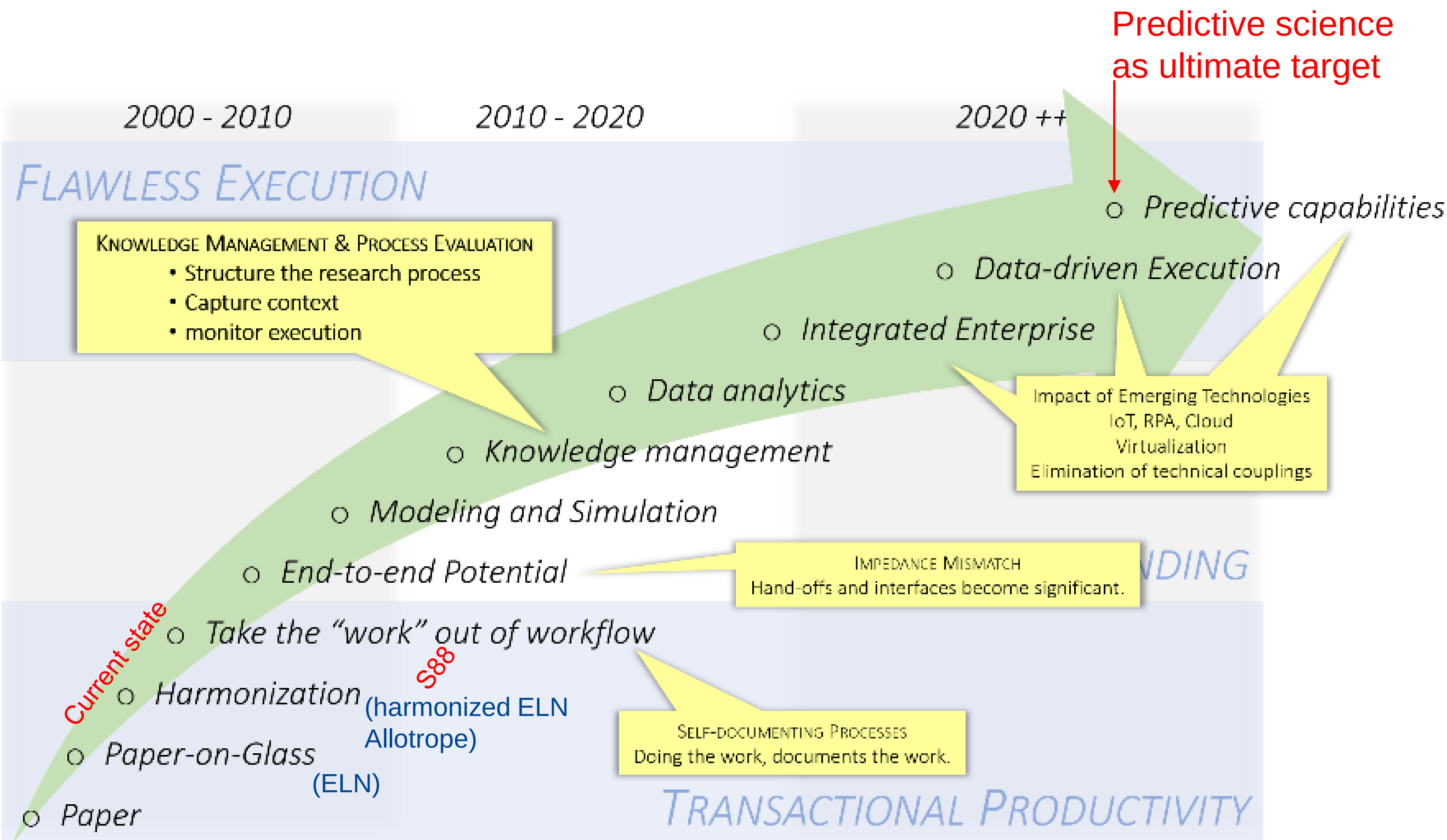
Impact of Good Data Management

	Achievement	Enables …	Problems if Not Present
1	100 % of measured values captured	Information driven decisions based on visible parameter interdependencies.	More experiments needed. Low data quality due to manual data capturing.
2	Data captured in high quality (reproducible)	Use of statistical models (DoE, ..), trust in data	Risk for wrong conclusions, bad repeatability
3	Data captured in consistent data format including meta data	No tedious and time consuming data workup.	Wasted FTE time and less information driven decisions.
4	Auto data storage at location accessible for full team	No lost experiments. Creation of institutional knowledge.	Value of past experiments not leveraged.
5	Complete and structured recipe data captured	Easy recipe exchange. Complete, accurate and reproducible recipe documentation.	Bad repeatability/Batch failures. Tedious recipe transfer processes.
6	Data synchronization between systems in lab (LES) and office (ELN)	No data transcription. No room for human errors.	Low data quality. Low trust in data stored in ELN.

Lab Digitalization Platform Approach

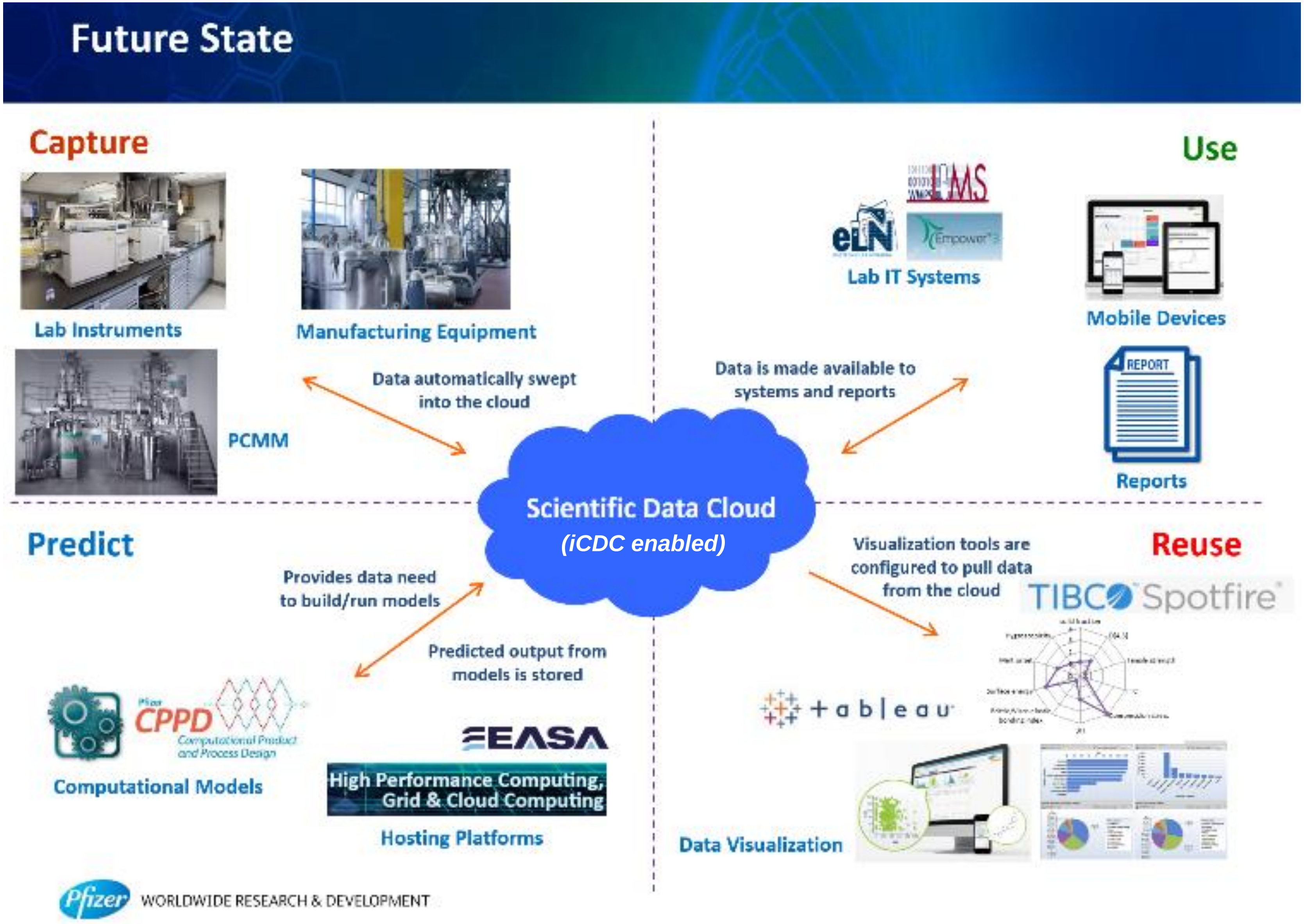


Example: Bristol Myers Squibb (BMS)



Source: Bristol-Myers Squibb at METTLER TOLEDO round table meeting, October 2017

Pfizer: Data Cloud - State of the Art in 2025?



Pfizer: Data Cloud

Why now?

Data Integrity / Regulatory Compliance

- Regulatory agencies are asking for **original raw data files** that are currently spread across multiple file storage systems

MHRA

- Regulatory inspectors have stated that **"quality is of concern"** if data cannot be found within minutes

Speed

- **High throughput** data analysis will enable simultaneous acceleration of multiple products through the pipeline

- Rapid access to high quality data will increase use of **predictive models** resulting in faster progression of projects

Scientific Insights

- **New insights** are needed to advance projects. SDC will enable **on-demand analysis** and **visualization** on relationships between datasets like solubility, API particle size and compression stress

Pfizer: Data Cloud – Data Visualization

Welcome - SDC Reaction Analytics | Reaction Graphs | Minimum and Maximum Temperatures | Time Difference | Raw Data Table

IDF
Scientific Data Cloud

Welcome to Reaction Analytics from SDC!

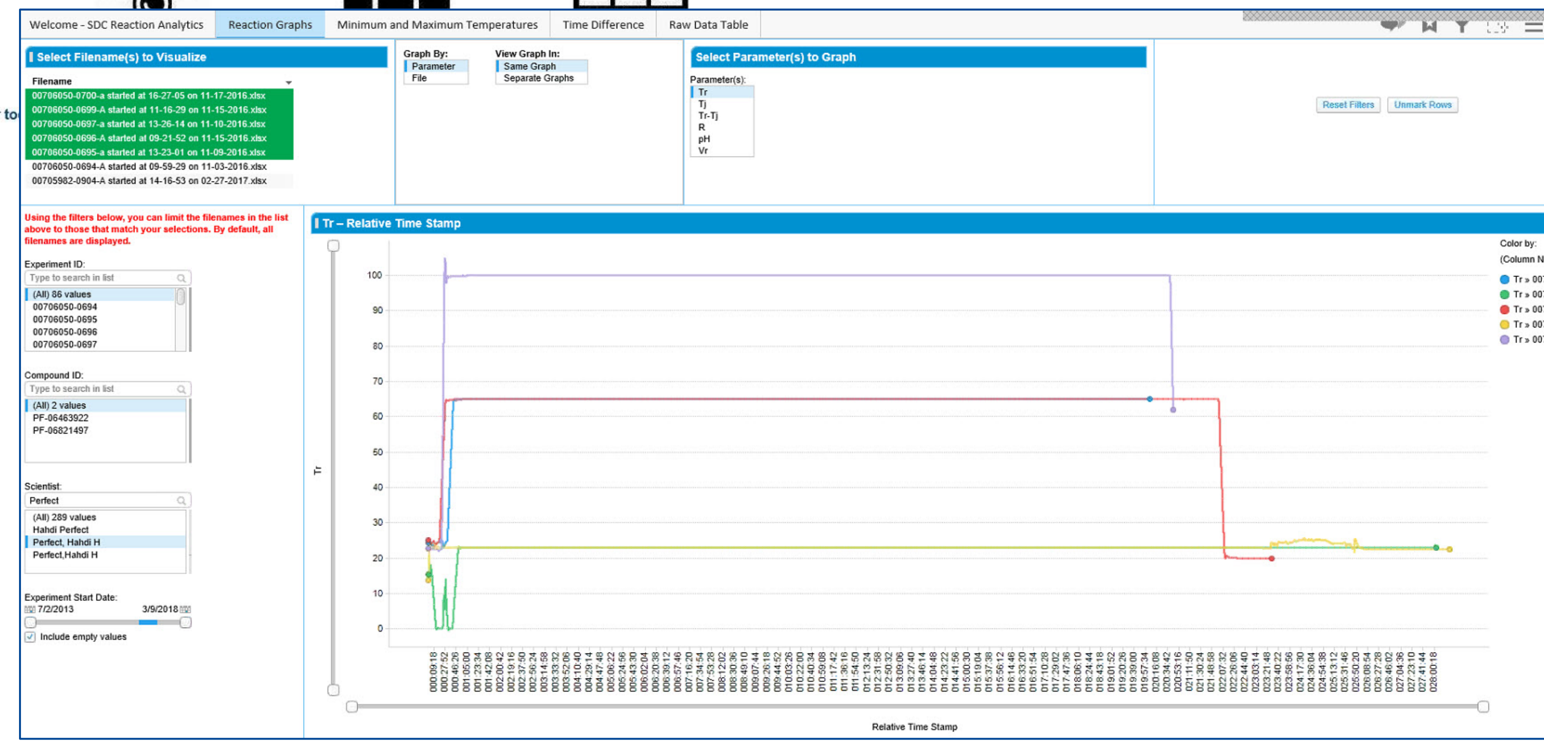
These data visualizations allow you to review Reaction Analytics data in novel ways. There are a few different visualizations that allow you to focus on one experiment or to compare and contrast multiple experiments. These are Reaction Graphs and Minimum and Maximum Temperature Graphs. With Distillation Time Graph, you will need to select a single experiment and from visual inspection, determine the start and end of the distillation process and the tool will calculate the length of that process. With the Data Table, you can download the filtered data for further analysis if so desired. [Note: The filtering on Reaction Graphs and Minimum and Maximum Temperature Graphs will determine the rows of data included in the Data Table.]

Reaction Graphs

For additional help in using the Spotfire Web Player to

- Introduction to Spotfire Web Player
- Spotfire Web Player Basics

Not to be used for Regulatory Purposes



Agenda

1	**Needs and Challenges – Chemical Development**
2	**Industry trends in the area of Knowledge Management**
3	**Data Analytics – during the experiment, single experiment and multiple experiments**

Why Find Trends?

Save large amounts of time – even for experts

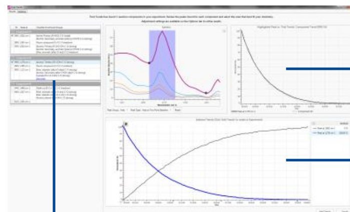
- It is too hard to get sensible trends out of time resolved MIR data
  - Good quality trends are critical for reaction analysis
  - Cannot do anything if you do not get the right trends
- Find Trends helps identify the right peaks to profile by finding the best isolated ones
  - Seems simple but this is really an expert level skill
- By comparing a peak picking model to an independently generated chemometrics model, result confidence is greatly increased
- The time taken to analyze the reaction is dramatically reduced
  - **What used to take 2 hours now takes 2 minutes**
- Leverage
  - Approach can be taken for any expert level data analysis skill
    - Difficult or complex data set analysis now in reach of less skilled users
  - Building of orthogonal models greatly increases result confidence
  - Experts like to be in control and do not like 'black-box' solutions
    - This approach gives them the choice to accept of the result or not
  - Faster results, greater result confidence (in line with LAB 5S)

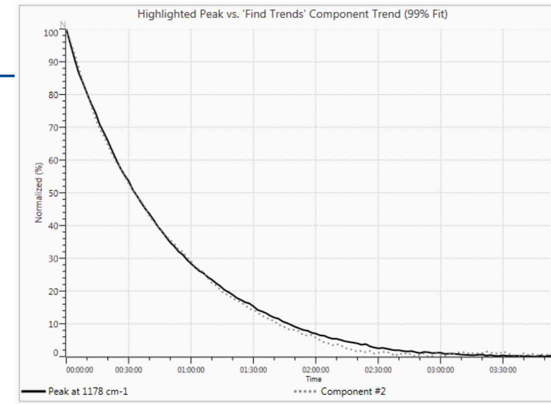
Find Trends – One Click™ Reaction Profiling

Objective:
Utilize expert knowledge to create a tool for fast profiling of high-quality trends

Results:
Development of **Find Trends**: Quickly provides useful trends for both expert and non-expert scientists

Simple Operation: Click **Find Trends**





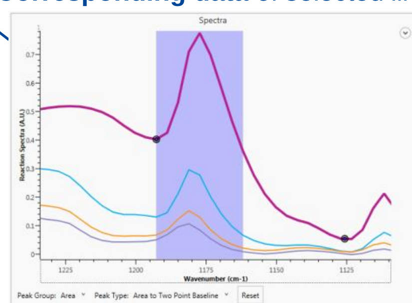
Fit	Peak at	Possible Functional Groups
Component #1		
☑ 99%	1062 cm-1	Amine, Primary (R-NH2) C-N (weak) Alcohol, Secondary, aromatic (phenyl-CHOH) C-O (strong)
☐ 98%	1367 cm-1	Fluoro compound (C-F) C-F (medium)
☐ 93%	1034 cm-1	Alcohol, Primary (R-CH2-OH) C-O (strong) Alcohol, Secondary, aromatic (phenyl-CHOH) C-O (strong) Ether, aromatic (alkyl-O-aryl) C-O (medium)
Component #2		
☑ 99%	1178 cm-1	Alcohol, Tertiary (R3-COH) C-O (strong)
☐ 99%	1365 cm-1	Fluoro compound (C-F) C-F (medium)
☐ 99%	1112 cm-1	Ether, aliphatic (alkyl-O-alkyl) C-O (strong) Alcohol, Secondary (alkyl-CHOH-alkyl) C-O (strong) Cyanate (R-C-O-C≡N) C-O (strong) Show More
Component #3		
☐ 94%	1448 cm-1	Olefin cis (R-C=C-R) C=C (medium)
☐ 84%	1247 cm-1	Ether, aromatic (alkyl-O-aryl) C-O (strong) Ester, aliphatic (alkyl-O-C=O) C-O (strong) Alcohol, phenol (aryl-OH) C-O (strong)
☐ 81%	2191 cm-1	

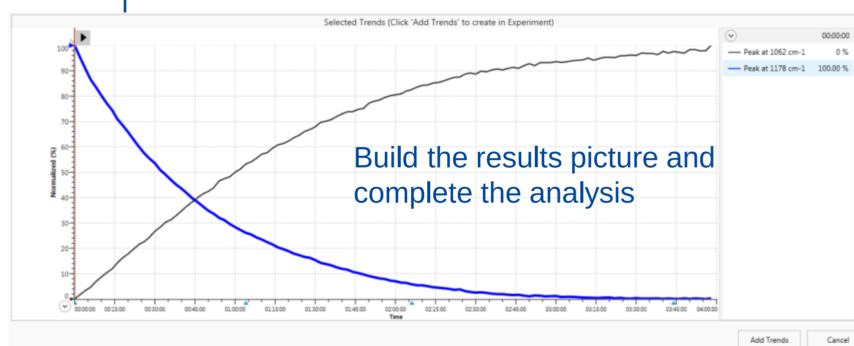
Software automatically **selects the best peaks** to trend the reaction.

Toggle through **best results to select the functional group** that fits the chemistry.

Check correlation between two independent, orthogonal models

Corresponding data of selected line





Build the results picture and complete the analysis

Verify results using correlation table, makes any adjustments to calculations and add to result set

Safe Results:
Find Trends generates two independent models. **One model is chemometrically generated** over a spectral region, the other model is generated from **isolated peak picking and trending**.

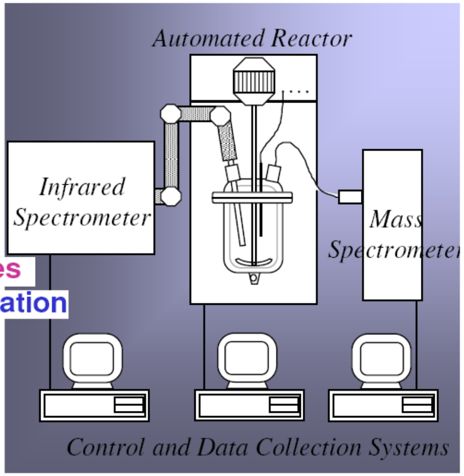
Comparison of these two independently generated profiles gives confidence the selected profile is correct.

Integrating complementary technologies

Process Technology Laboratories

Knowledge Obtained from Data-Rich Experiments

- Kinetic Models & Rate Constants
- Heat of Reaction
- Mass Balance
- Reaction Endpoints
- Liquid Composition
- Gas Composition
- Reaction Pathways
- Solubilities
- Viscosities
- Transient Intermediates
- Reactive Chem Information
- pH & Conductivity
- Particle Size
- Turbidity
- Polymorphs



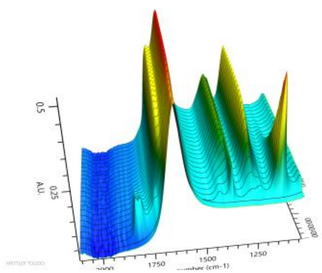
- 10X more data relative to other approaches
- Difficult or impossible to obtain otherwise

Lilly
Answers That Matter.

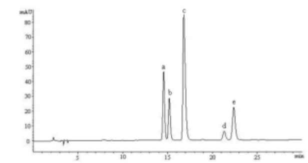
Mark LaPack, Eli Lilly & Company, 13th International Process Development Conference, 2006

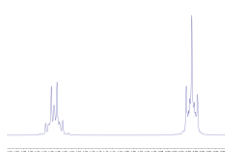
Quantitative intuition: Making smarter decisions

- Taking data from multiple sources and integrating them together
- Data-driven decision as the information content is much higher
- Better understanding of design space for more efficient work processes

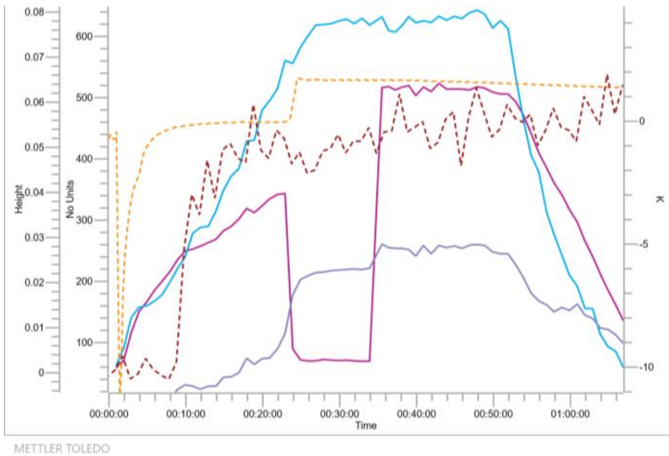


Online:
- HPLC
- NMR
- MS

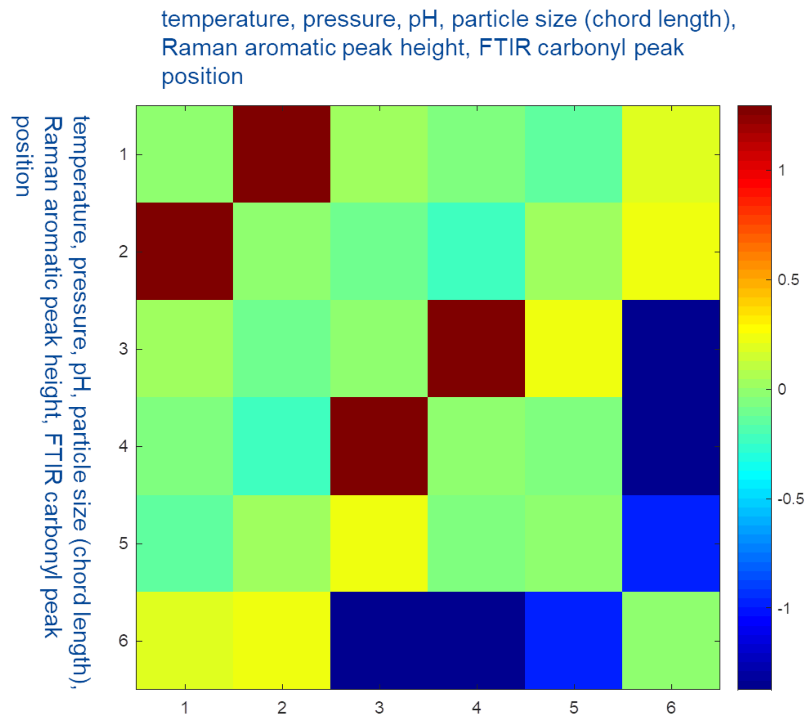




Focus on most important data for enhanced process understanding

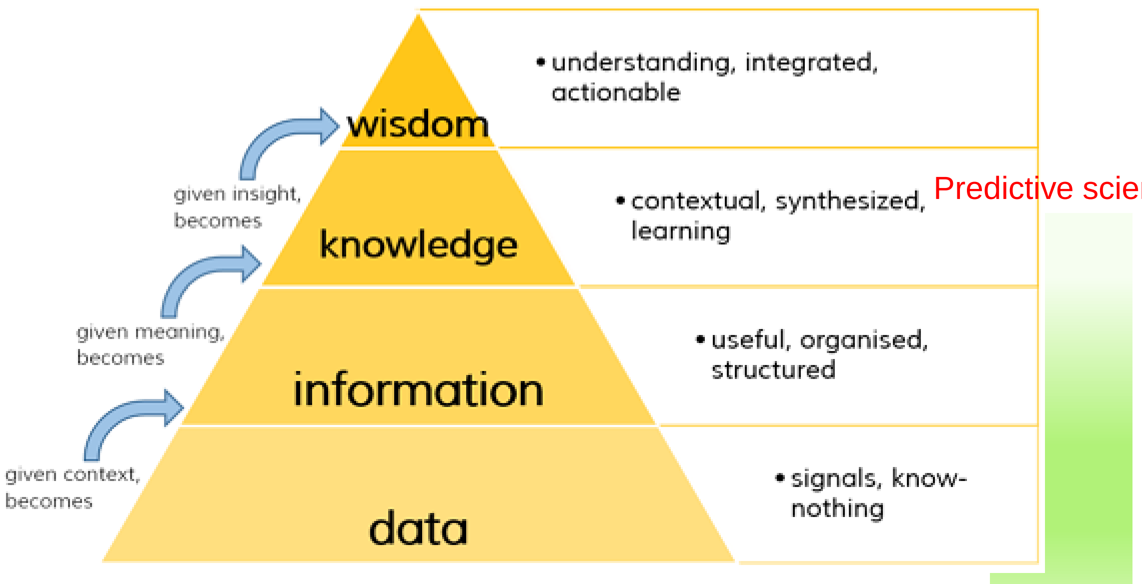


Trends from multiple sources



Correlate trends with heat map

Our Strategy in the Knowledge Pyramid



Predictive science

AC focus

- Automated decision taking based on experience from 1mio past experiments
- Decision on next experiment run based on gathered information and experience from last experiments in project
- heat flow trend
- IR peak trend
- Summary table on project parameters
- Process data mixed with core analytics
- Tr-Tj trend
- Raw IR spectra

AU - Focus Data-to-Information activities on value that can be created DURING experiment.