

De stand van zaken van data management software in het laboratorium

Peter Boogaard

Industrial Lab Automation

17 November 2017 – Wageningen – Nederland

peterboogaard@industriallabautomation.com

Peter Boogaard

- Independent consulting to address harmonization, integration and consolidation in science based processes
- Degree in Analytical Chemistry
- Introduced to Lab automation since 1981
- 30+ years @ international companies
- Founder of the annual Paperless Lab Academy
- International publications
- Active ISPE member (GAMP DI SIG & QbD/PAT)
- After work... Sailing & Music



www.industrialabautomation.com
peterboogaard@industrialabautomation.com
cell +31 654 263516



QbD & PAT
The Scientist Is No Longer In The Laboratory,
But Integrated In The Overall Quality Process

Considerations for
Software Expansions and Upgrades

Informatics:
The Glue to Build Enterprise Knowledge

Is less more in a
paperless laboratory?

JOINING UP
THE LABORATORY

BioPharma
Asia

WHEN DATA
COMES TOGETHER

Facing Cross-Industry Challenges
in the Food and Pharma Industries

Paperless Lab Academy® 2018
Milan, March 20-21

HOW TO IMPROVE DATA INTEGRITY
LABORATORY INFORMATICS
GUIDE 2017 ROADMAP TO DIGITAL
CONVERGENCE

GMP Journal

ENGINEERING PHARMACEUTICAL INNOVATION
ISPE
GAMP® Data Integrity
Special Interest Group

Empowering the eData Life Cycle

Delivering scientific evidence to data consumers



Paperless Lab Academy® 2018

Milan, March 20-21



"eConnect, eManage, eDecide, eArchive"

"eConnect"

- Effective workflow based self-documenting **data capture strategies**
- New data capture approaches to embrace **Internet of Lab Things** IoLT
- Models to embed intelligent software in instrumentation and sensors
- Integrating external collaborators and scientific literature knowledge

"eManage"

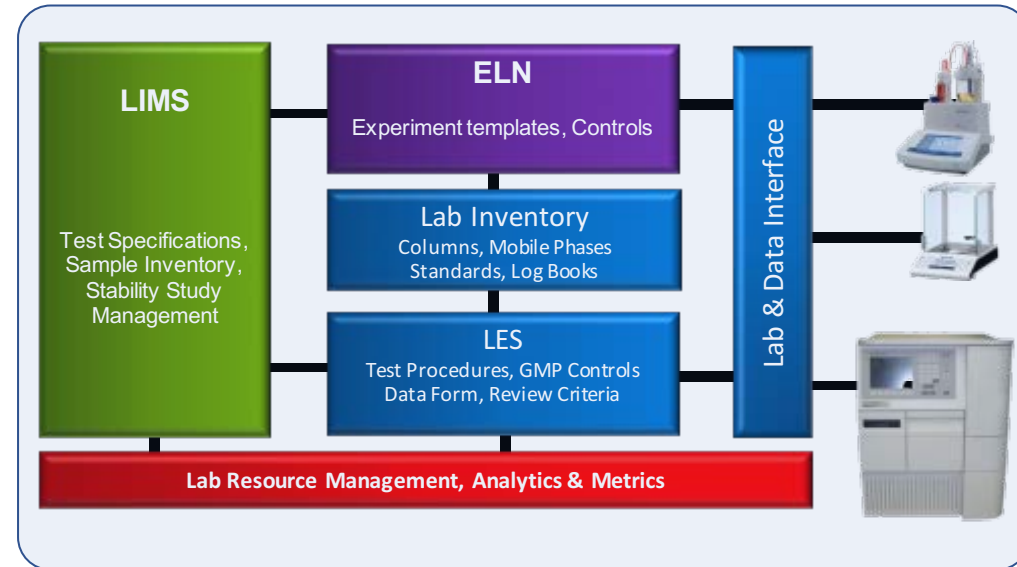
- Innovative methodologies to manage scientific master and meta data
- Concepts to prepare for new regulatory and data integrity challenges
- Strategy to harmonize / reconcile taxonomies and ontologies to boost insight

"eDecide"

- Adopting data analytics and visualization tools to materialize scientific knowledge
- Strategies to ease access and reduce complexity to non-scientific data consumers
- Enabling the power of Industry 4.0 to the laboratory

"eArchive"

- How to reduce the struggle finding the right data in the growing digital universe
- Essentials to secure long-term multi-departmental archiving
- Approaches to manage cost, sustainability and infrastructure processes



www.paperlesslabacademy.com

PUBLICATIONS Peter Boogaard - Industrial Lab Automation

Free Download @ www.industriallabautomation.com/Publications.php

ROADMAP TO DIGITAL CONVERGENCE

Data intensive science is becoming mainstream, and new technology will change the dynamics of how scientists will work together. Many cross-industry best practices can be used to enable cross-functional collaboration between internal information silos to transform scientific information into actionable insights. What will the Internet of Lab Things (IoLT) bring in the [future](#)?

GMP Journal

Trends In Laboratory Informatics. Just like almost anywhere else, informatics is the normality in lab operations. Data intensive analytics create gap-less knowledge management systems also facilitating the cross-sectoral collaboration.

Electronic Laboratory Notebooks

ELN means many things to many People. ELNs mean many things to many people. This article starts with the end user in mind and look at the application from a user centric perspective. What is an Electronic Laboratory Notebook (ELN)? What function does it serve? Where does it fit within my laboratory informatics strategy? Do we need an ELN, and if so what would be best for my company's needs? When should I use an ELN, or a LIMS or both?

HOW TO IMPROVE DATA INTEGRITY

Data integrity is currently one of the highest cited areas in regulatory observations. Yet, data integrity is not a new requirement. In this article I will highlight the how can reduce data integrity inconsistencies

Linking an Instrument to a Tablet

For how long do we need to be professor to link an instrument to a tablet in the laboratory? Why can we connect almost any smartphone using Bluetooth in our personal lives ourselves and why do we need a professor to transfer simple results in the laboratory to our ELN, LES or LIMS



In this article, I will share experiences and observations how the scientific high-tech community, can benefit from adopting paperless processes in the laboratory. Is it because paper doesn't require any significant investment budget, or is it the low barrier to access, since paper even works without power or the need to have access to an information infrastructure, or is it just simply that the "what's in it for me" question hasn't been answered satisfactorily for the scientists?

MANAGING CHANGE IN THE LABORATORY TO DELIVER MORE VALUE

New **scientific data consumers** are increasing the value of laboratory. Research, manufacturing and regulatory procedures have been unchanged for years. There is an urgency to revisit these. The need to integrate the legacy silo based departments is becoming a top priority agenda item in many boardrooms.

JOINING UP THE LABORATORY

It is **pure waste** to perform labour-intensive hunting for information across multi-vendor, multi-technique databases, manual transcription checking and to manually create reports. What are these challenges to create value for the consumers of the laboratory data?

WHEN DATA COMES TOGETHER

When considering data integration, we must first **stop thinking 'technology'**. Integration is not just about instruments or other software platforms. Instead, it is about integrating processes, accelerating ideas and facilitating mandatory compliance requirements more economically.

Less is More: Adopting a Self-documenting Paperless Mindset

The power of a Paperless Laboratory is the ability to enable organizations to implement **self-documenting processes** that produces GxP-compliant documentation to support corporate Cost of Goods Sold (COGS) optimization

Considerations for Software Expansions and Upgrades

Headaches about **upgrading** your software? Considerations for software expansions and upgrades. Before you decide to rock the boat, several key decision-making steps can help to ensure a smooth and successful upgrade. The last thing to do is to start is a project to change a working enterprise application

Informatics: The Glue to Build Enterprise Knowledge

It is time to put more emphasis on the preventing facet and re-order the sequence of the **CAPA** abbreviation **into PACA**. It is proven that his theory to adopt continuous improvement strategies to decrease variability resulted in significantly better products and financial performance.

Facing Cross-Industry Challenges in the Food and Pharma Industries

Food companies are becoming pharma companies. In this publication several overlaps and differences are discussed to spur a discussion on how both pharma and food industries may benefit from adopting the best practices.

Publication Download pages

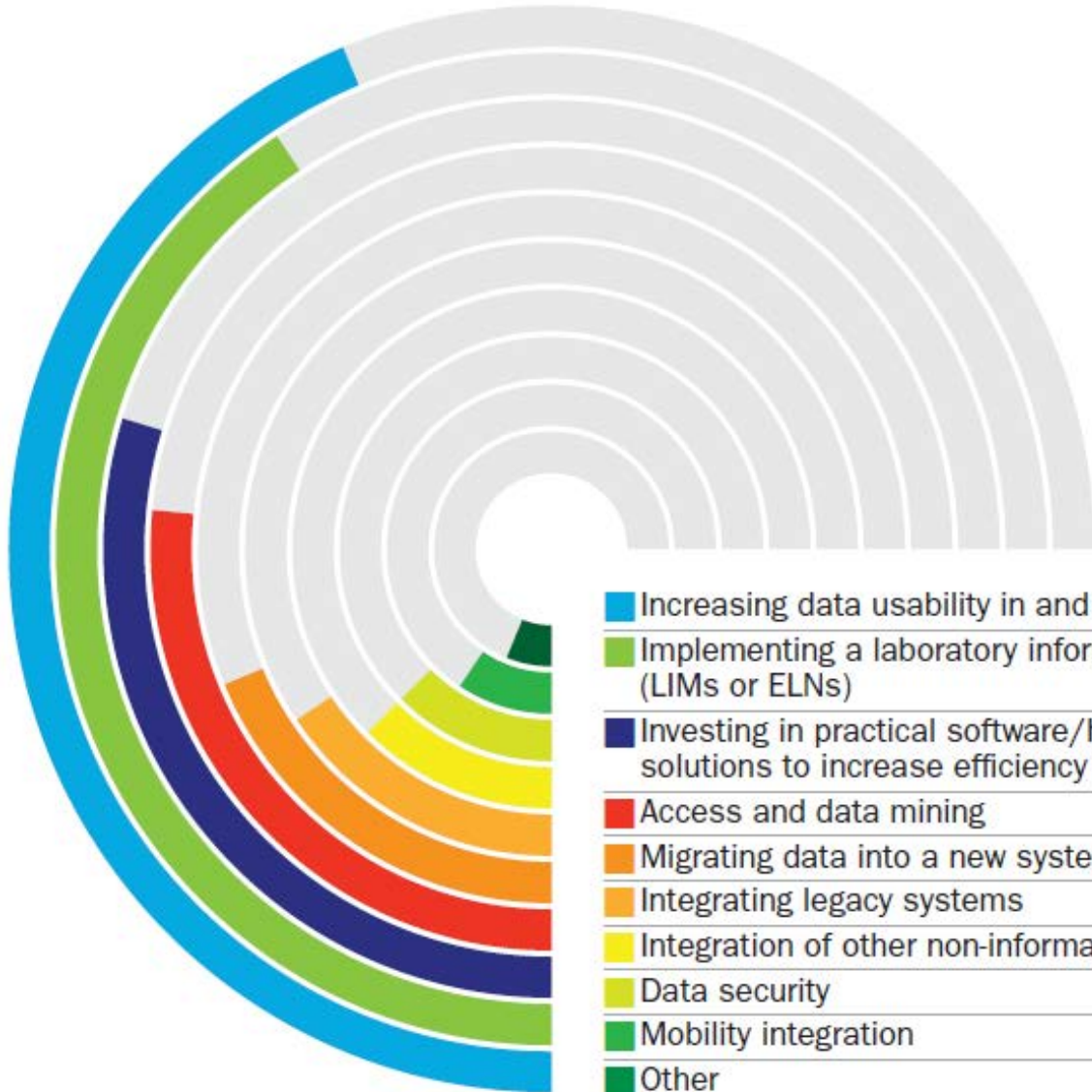
www.industriallabautomation.com/publications

What general technology
needs to be **improved**
significantly, to accelerate
the acceptance to work
electronic?

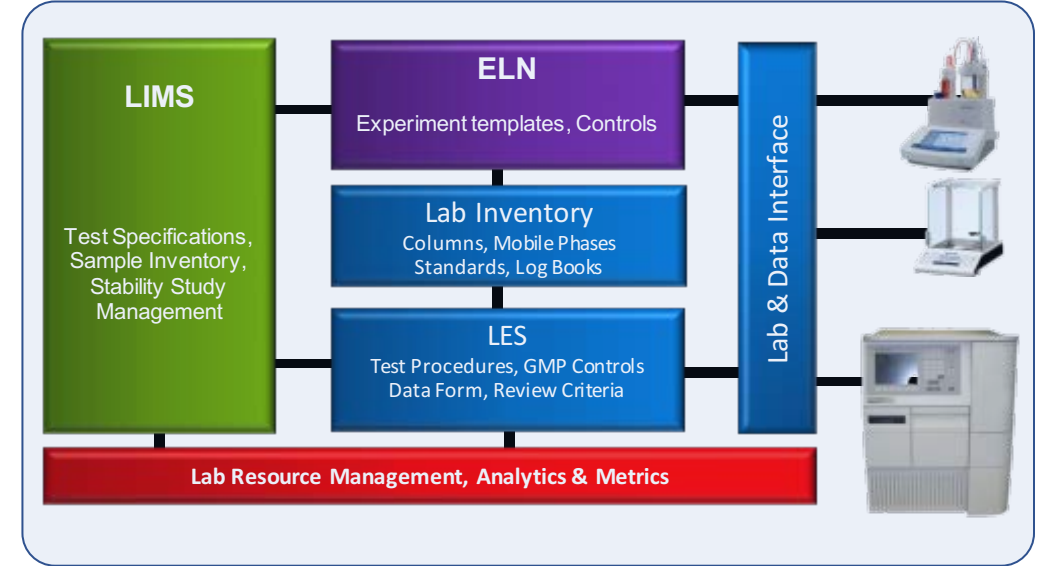
Batt%



Lab informatics expectations (2016)



Increasing data usability in and out of the laboratory	58.3%
Implementing a laboratory informatics system (LIMs or ELNs)	54.2%
Investing in practical software/hardware solutions to increase efficiency of the workflow	39.6%
Access and data mining	35.4%
Migrating data into a new system	25%
Integrating legacy systems	20.8%
Integration of other non-informatics software systems	16.7%
Data security	16.7%
Mobility integration	12.5%
Other	8.3%

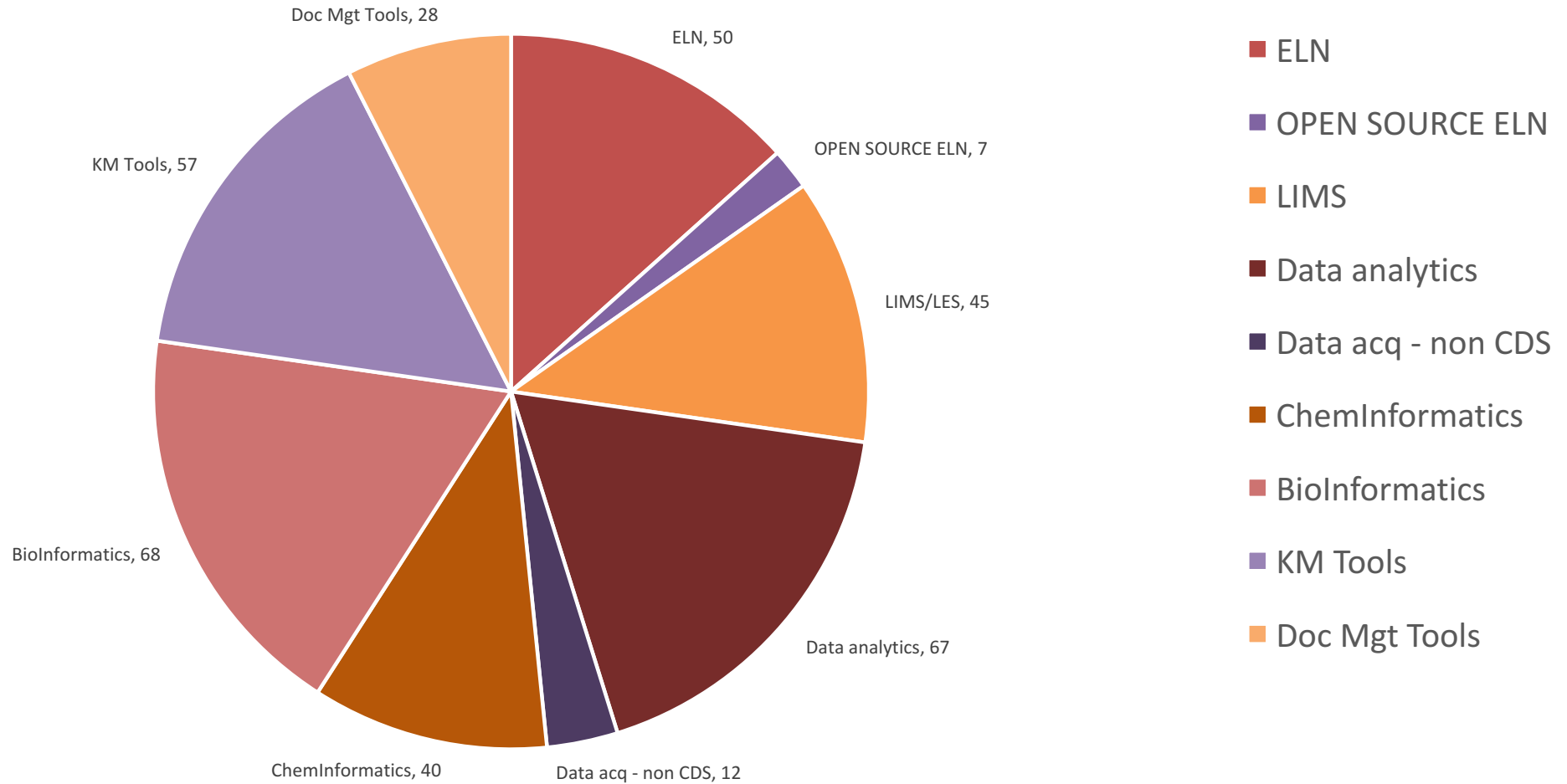


Data-intensive
science is
becoming
mainstream

Where to Start?

A laboratory setup featuring various pieces of glassware and equipment. In the foreground, there is a round-bottom flask on a stand, a condenser, and a receiver flask. The background shows more laboratory equipment, including a large spherical flask and a condenser. The entire scene is illuminated with a blue light, creating a professional and scientific atmosphere.

Scientific Application Landscape (total 374)



Source Atrium Research website 2017

Scientific Software landscape

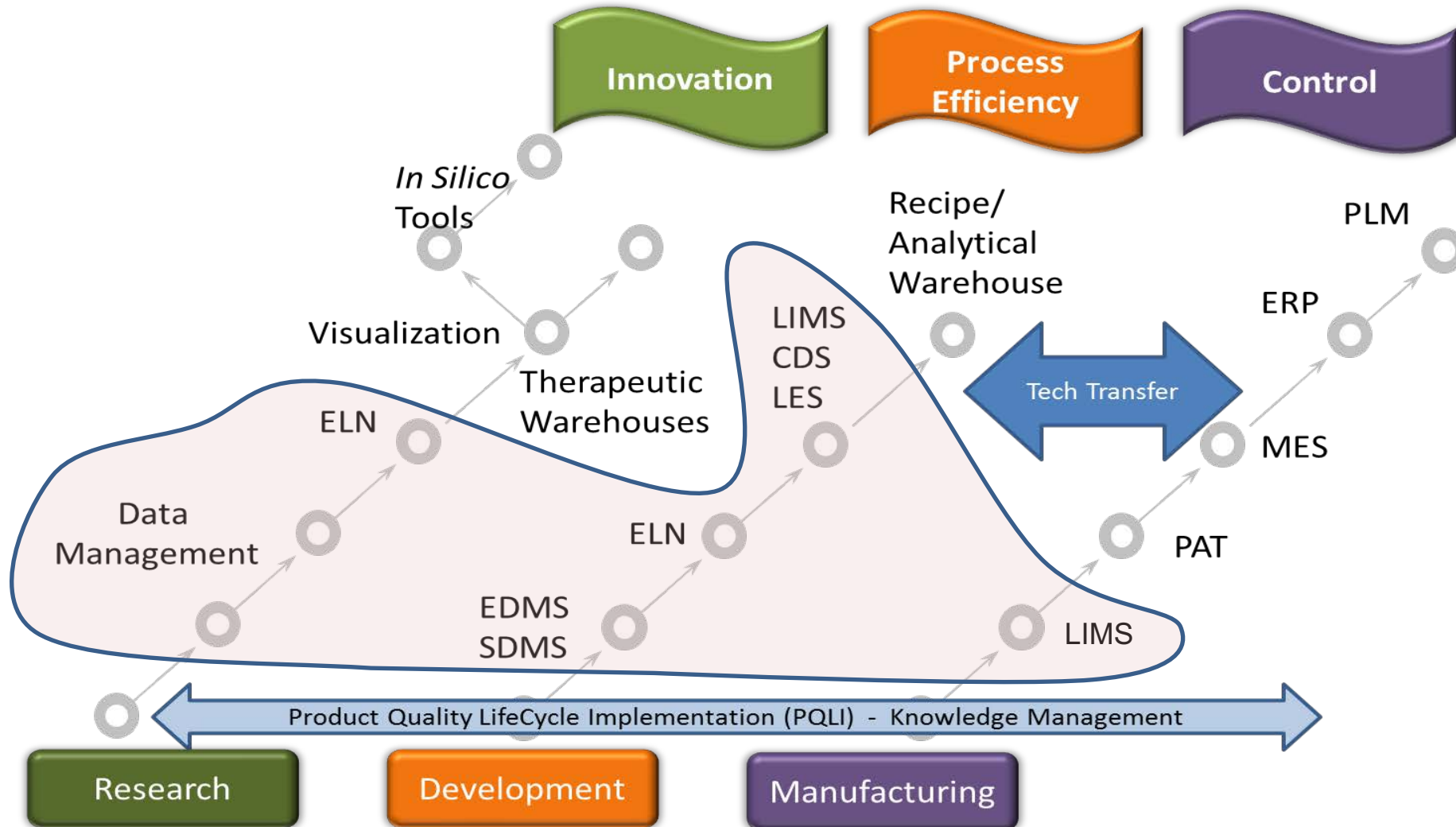


Research

Development

Manufacturing

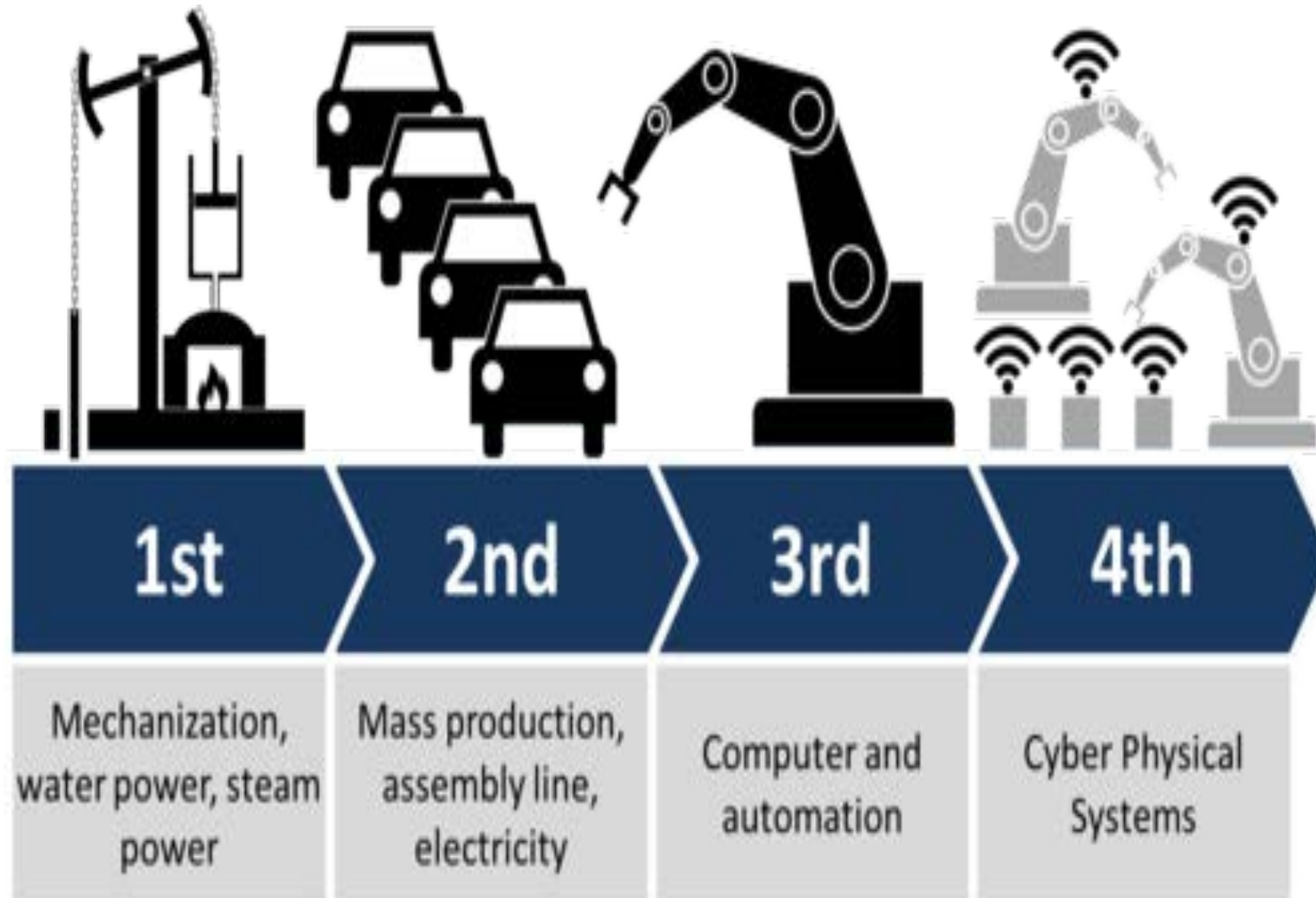
Scientific Software landscape



Source: Accelrys

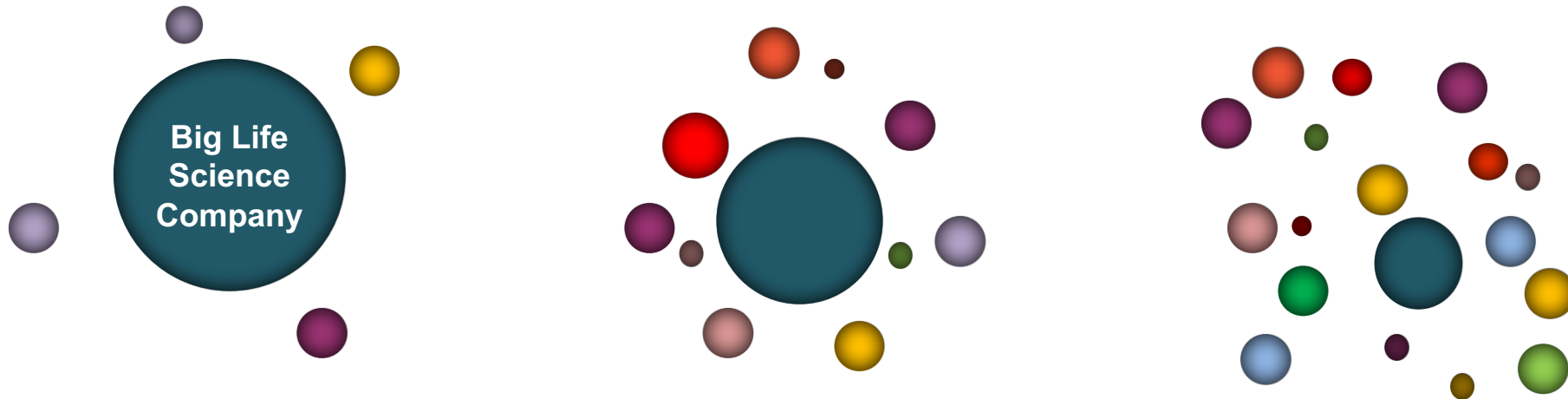
Industry 4.0

The Digital Revolution in Manufacturing



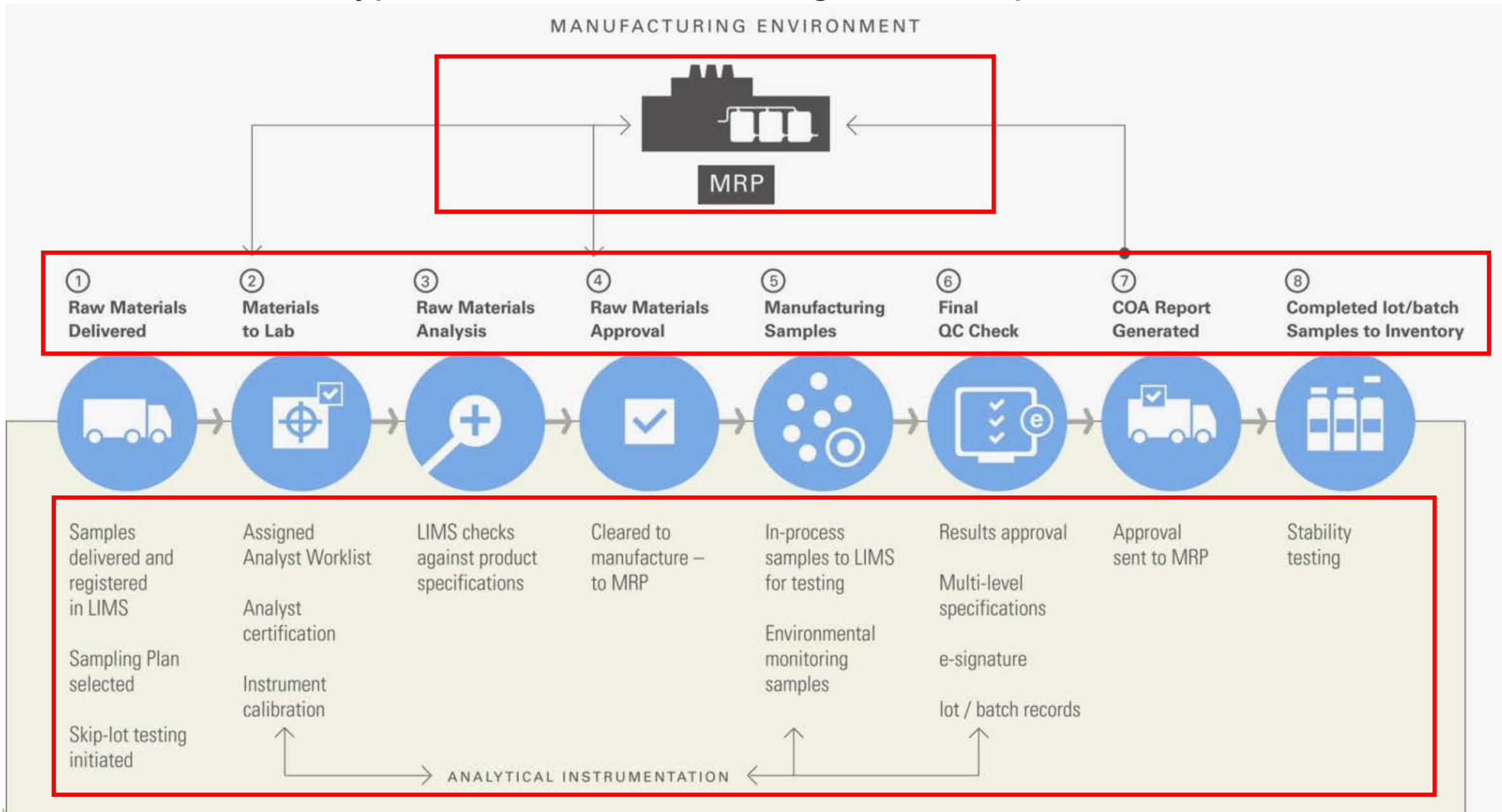
Scientific Information Landscape

A rapidly evolving ecosystem

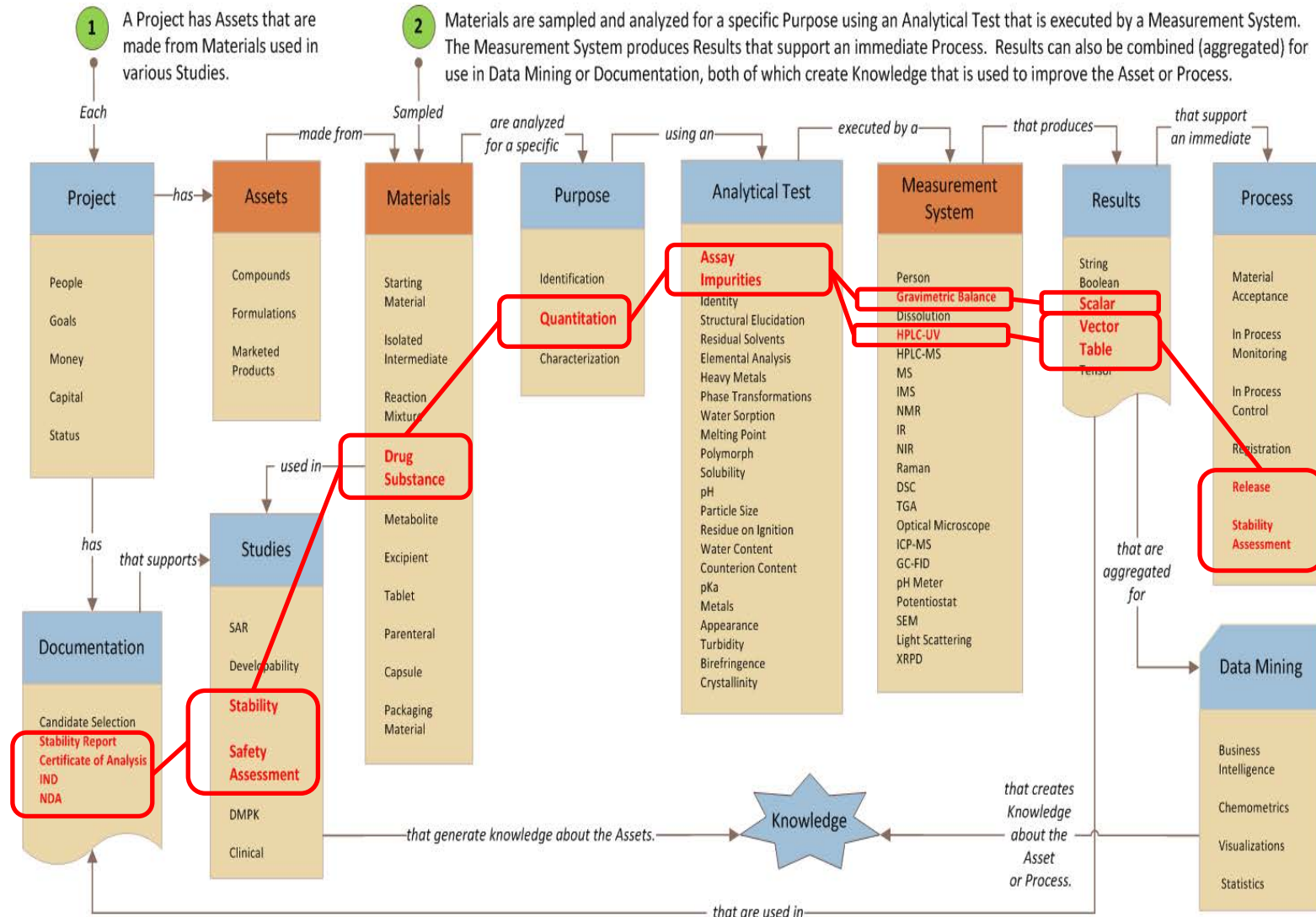


	Yesterday	Today	Tomorrow
Innovation Model	Innovation inside	Searching for Innovation	Heterogeneity of collaborations. Part of the wider ecosystem
IT	Internal apps & data	Struggling with change Security and Trust	Cloud/Services
Data	Mostly inside	In and Out	Distributed
Portfolio	Internally driven and owned	Partially shared	Shared portfolio

Typical LIMS manufacturing workflow process



Example of Laboratory Data Life Cycle model



#1 CHALLENGE

UNDERSTAND DATA
CONSUMER NEEDS
& THEIR OBJECTIVES

“We believe that the largest drug is the food that you eat three times a day, every day of your life”

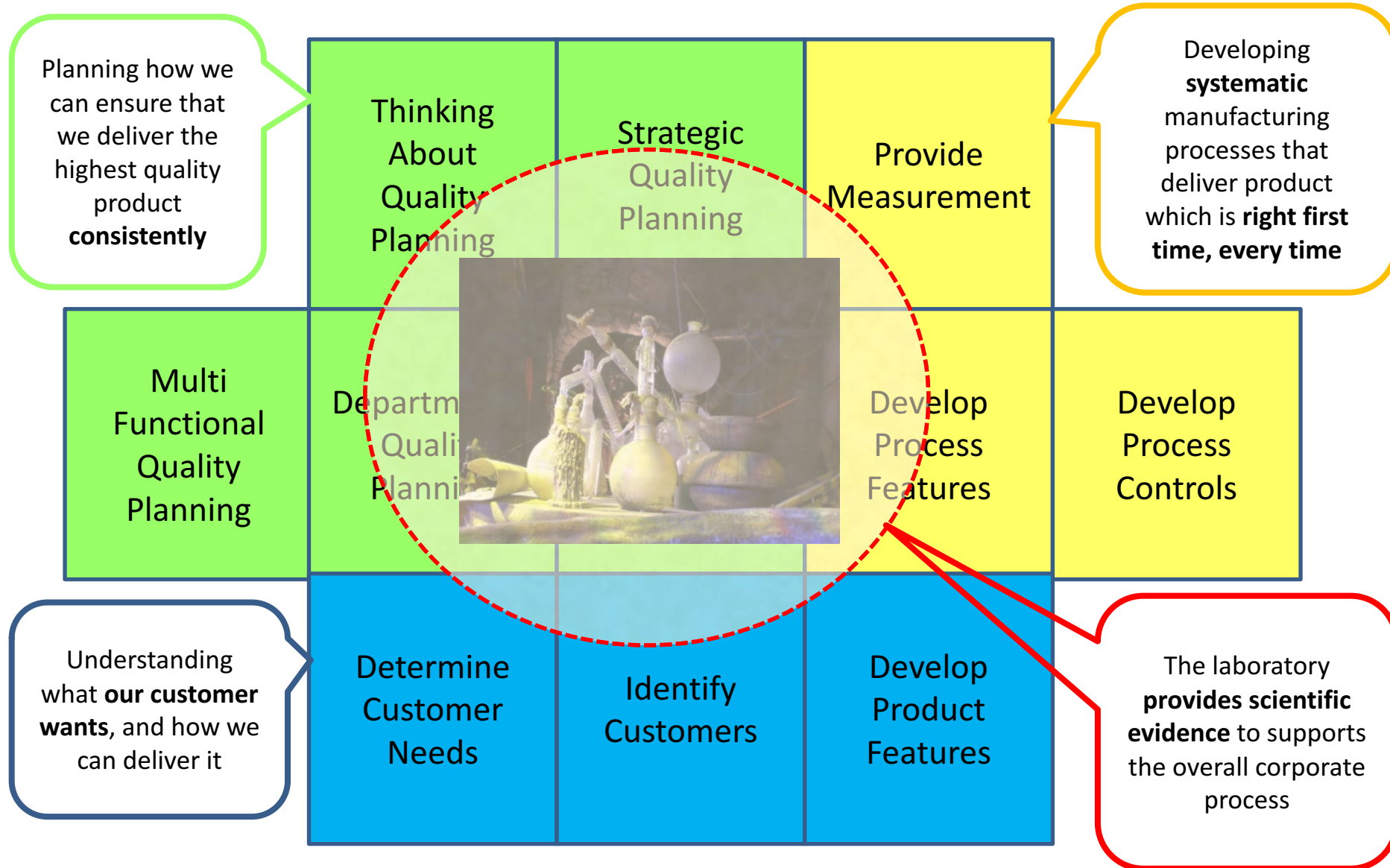


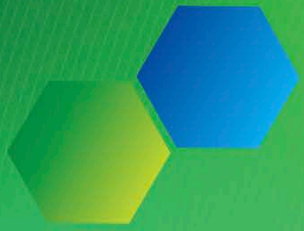
[Read full article](#)

Luis Cantarell, President and CEO of Nestlé Health Science - Brett Gundlock for The Globe and Mail Sept 6, 2012



Defining the holistic process





Paperless Lab
Academy

Gran Hotel Rey Don Jaime



FINDING the SPEED to INNOVATE

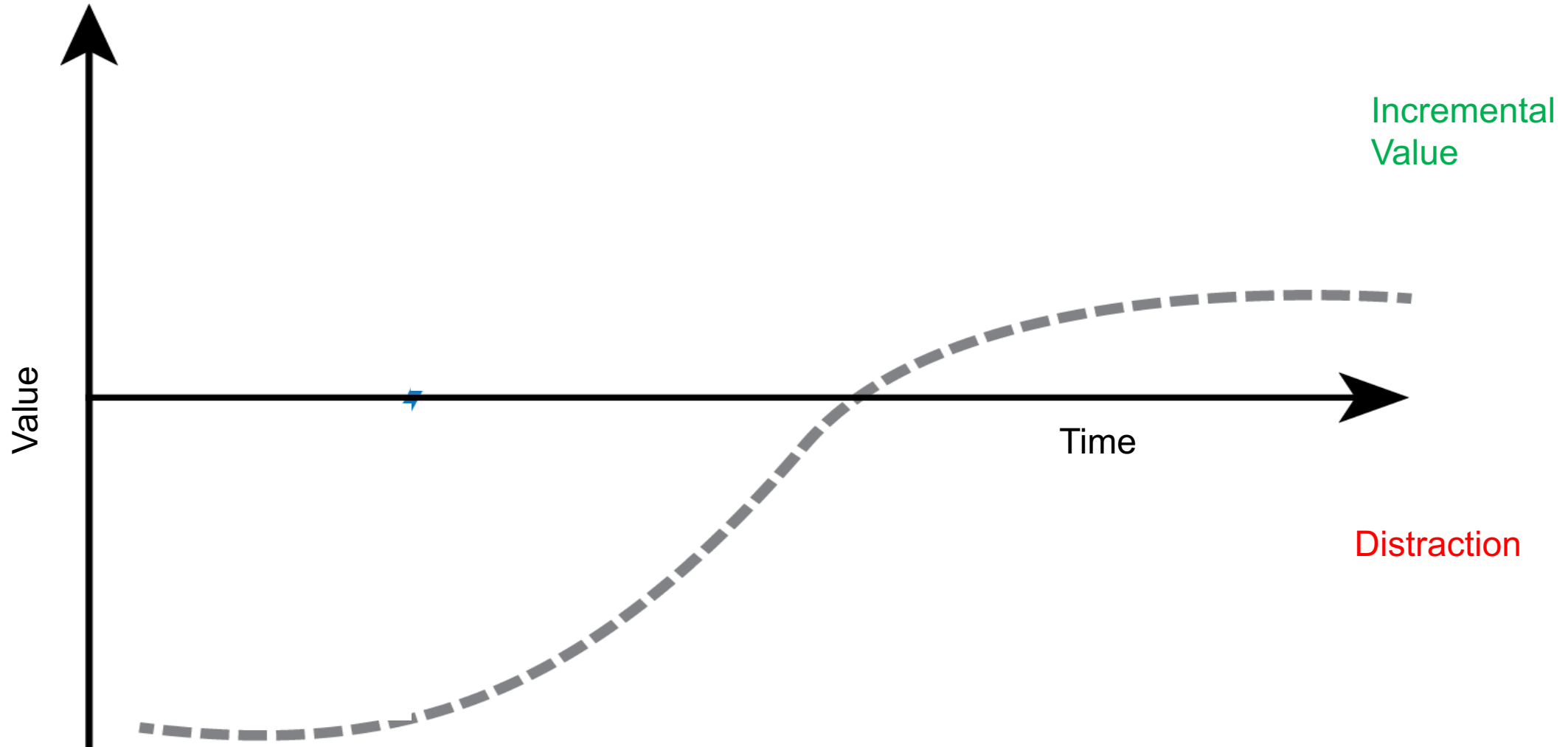
Strategies to effectively materialize laboratory knowledge

Create “Strategic Speed”

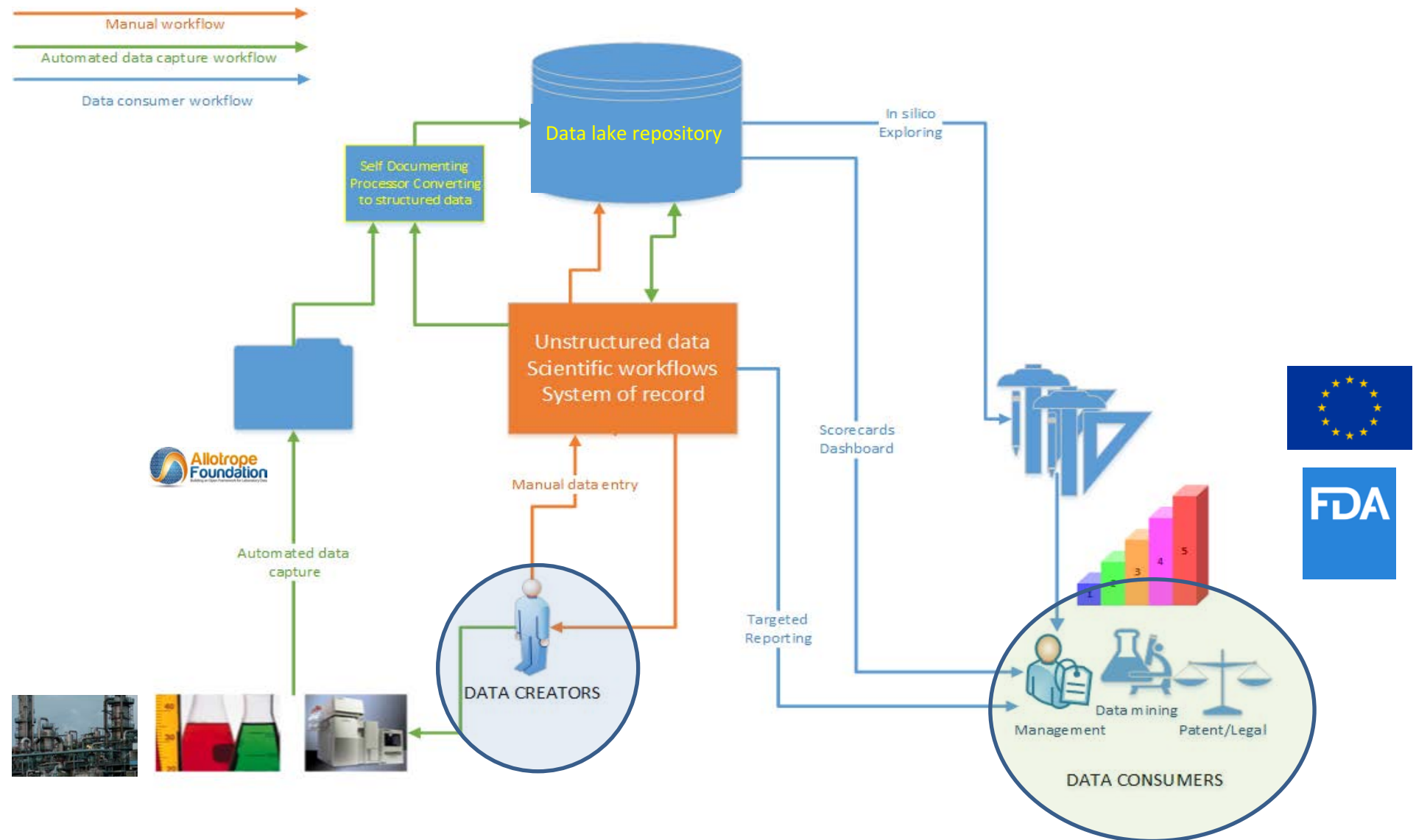
Minimize Time to Value; Maximize Value Over Time

Create "Strategic Speed"

Minimize Time to Value; Maximize Value Over Time

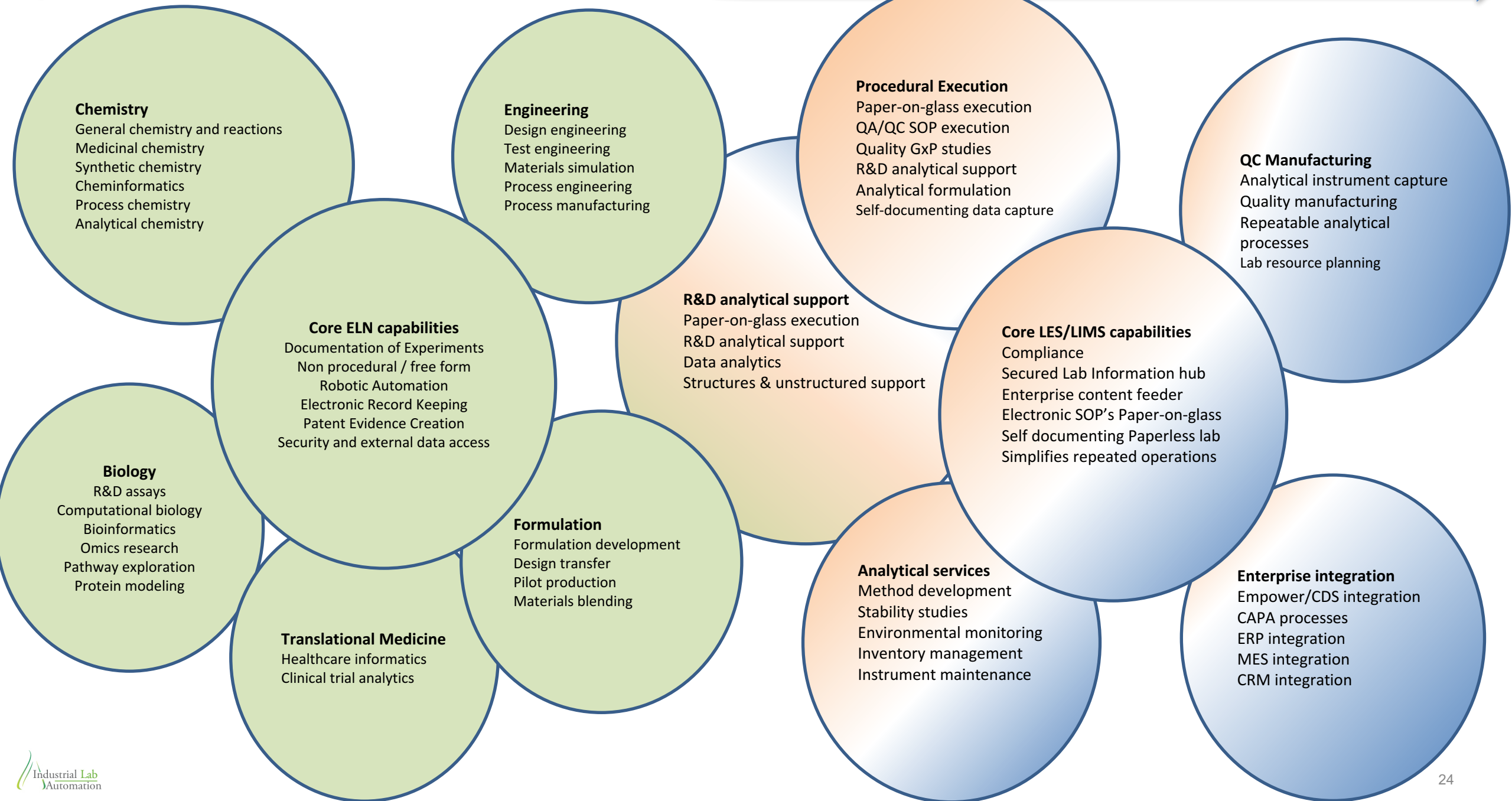


Value of lab data will be expanded by understanding the data consumer need

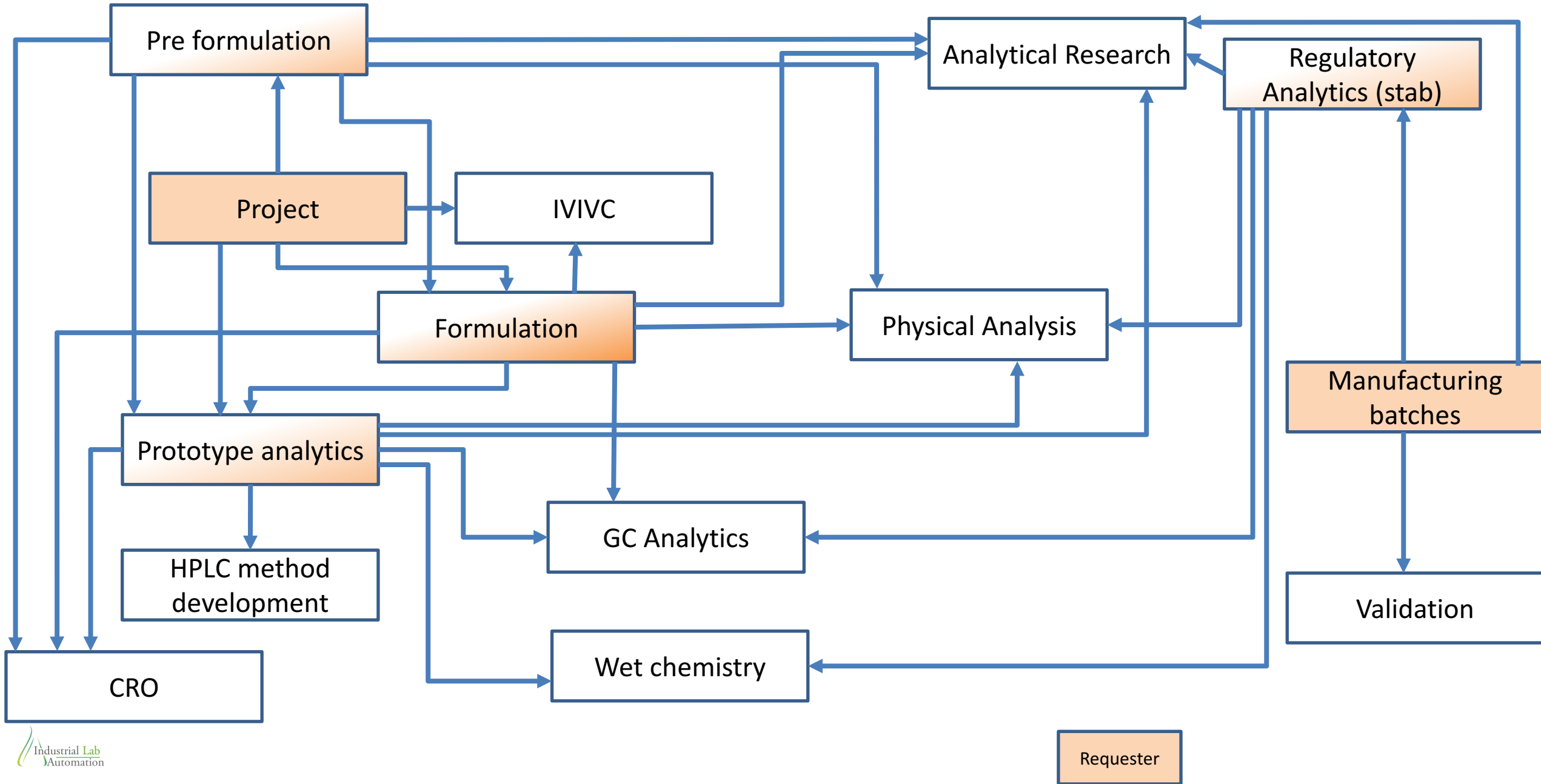


#2 CHALLENGE

USING THE RIGHT
TOOLS



Requester workflow

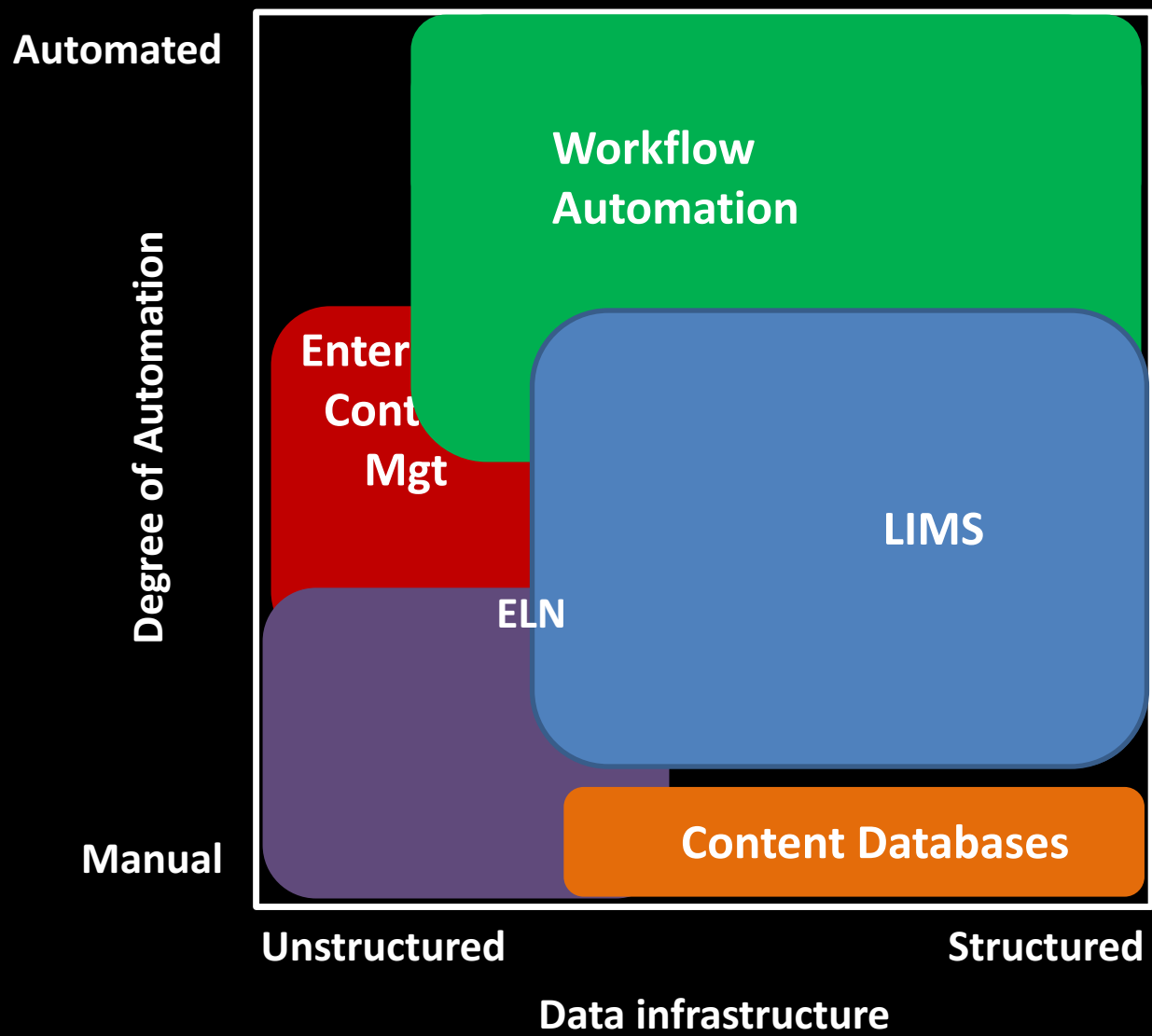


In its simplest form, an Electronic Laboratory Notebook can be thought of as for an electronic embodiment of what is currently being done in a paper laboratory notebook.

It is a tool that facilitates workflows that play out in laboratories.

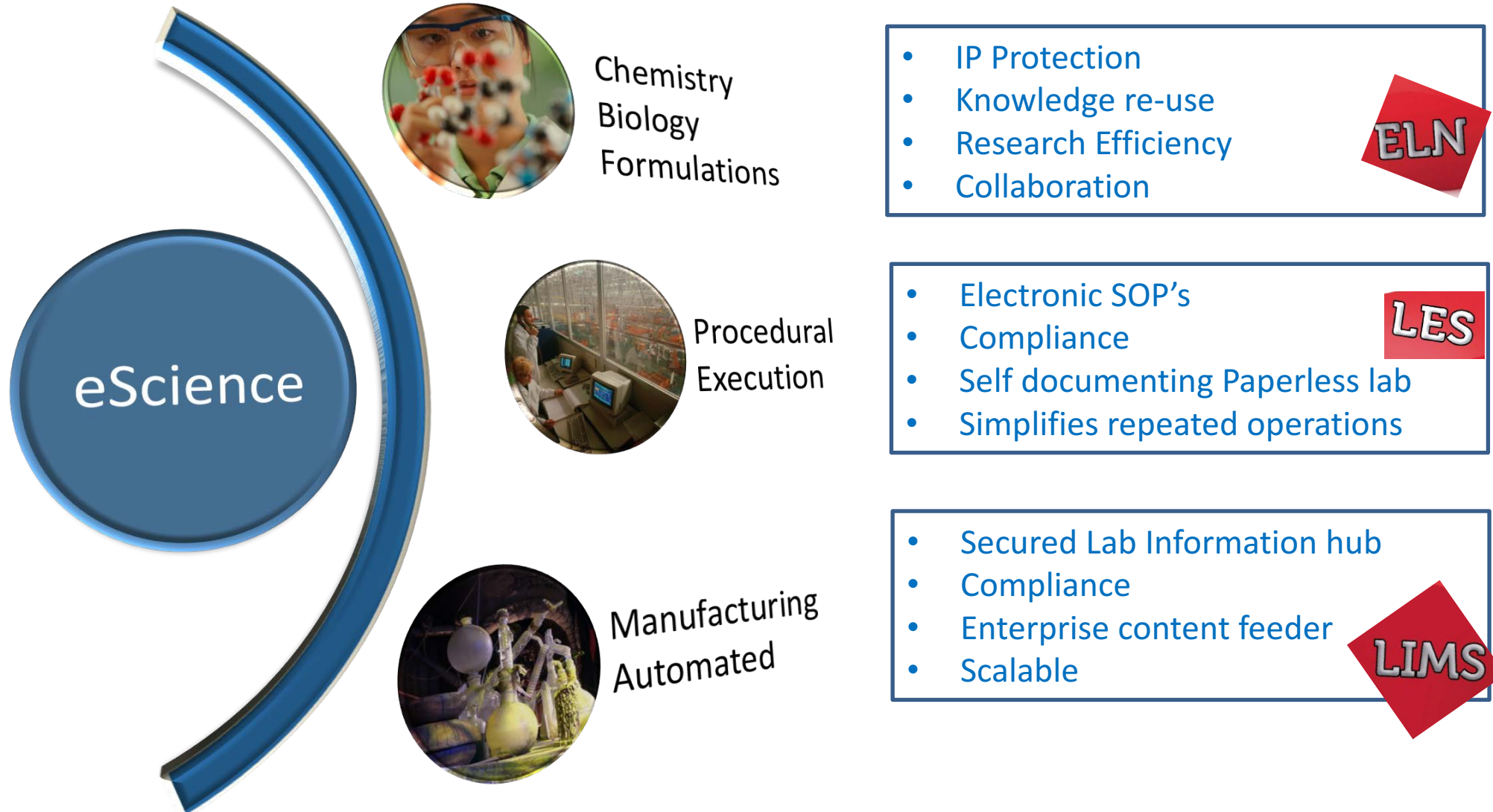
Having said that, Laboratory Information Management System (**LIMS**), Electronic Laboratory Notebook (**ELN**) and Lab Execution System (**LES**) applications **all support this basic definition**, to a greater or lesser extent.

So why bother?

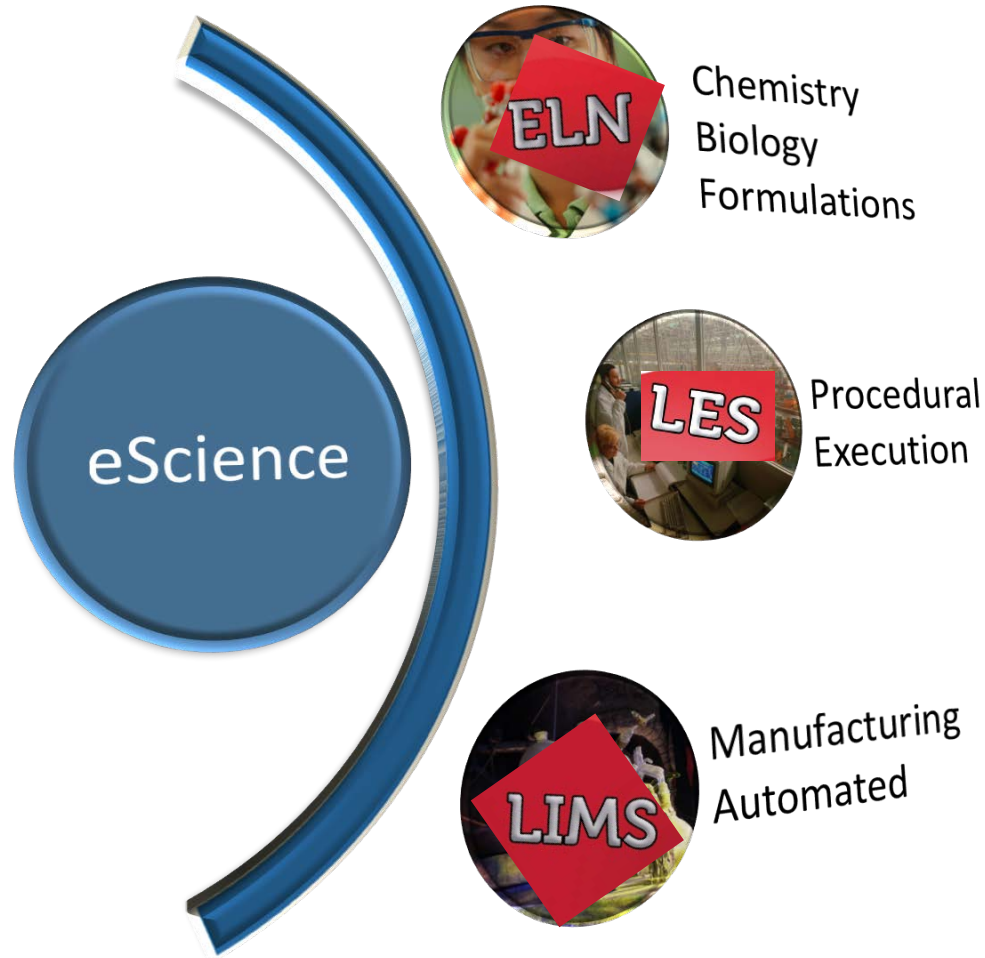


Courtesy Atrium Research

The role of Lab Data management software



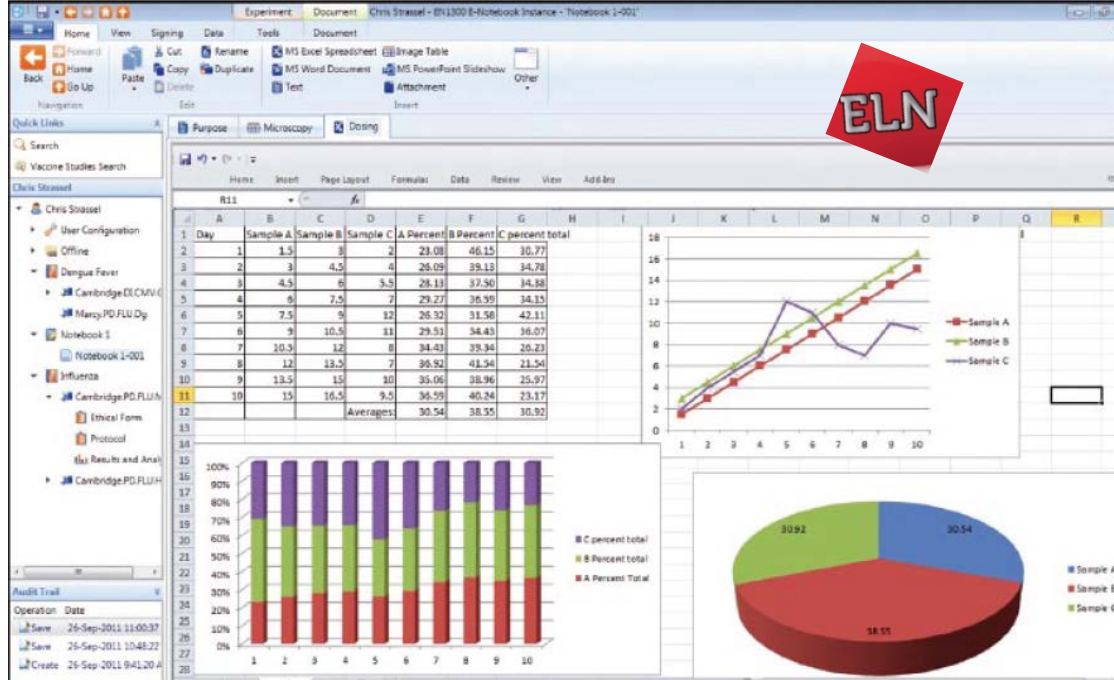
Summary - Behavior of Lab Data management software



- Experiment centric
- User centric / Adaptable free form
- Scientific databases, Instruments
- Legal & Scientific communities

- SOP workflow centric
- User centric / Natural Language
- Act as paper-on-glass application
- Strong Quality Assurance aspect

- Sample centric
- Organization centric / System defined
- Repeatable standard process workflow
- Regulated manufacturing



The LIMS interface shows a 'Study' form with the following fields: Study (SAMPLE GROUP), Study Datagroup (GLOBAL), Submission ID (100000246), and Log Priority (4). Below the form is a 'Sample Plans' table listing various sample types and their details.

Sample Plan	Material Name	Sample ID	User Sample ID	Priority	Qty	Location
<input checked="" type="checkbox"/> AIR PARTICLE COUNT	WATER	200000928	My Sample ID 1	6	3	
<input checked="" type="checkbox"/> RINSEWATER	WATER	200000929	My Sample ID 2	6	2	
<input checked="" type="checkbox"/> MIXER SWAB	WATER	200000930	My Sample ID 3	3	5	
<input checked="" type="checkbox"/> WALL SWAB	WATER	200000931	My Sample ID 4	1	8	
<input checked="" type="checkbox"/> FLOOR SWAB	WATER	200000932	My Sample ID 5	8	7	
<input checked="" type="checkbox"/> WATER PURITY	WATER	200000933	My Sample ID 6	6	1	

Titration Worksheet

LES

Product:	PROPOFOL	Lot Code:	
----------	----------	-----------	--

SOP	20070219_rev05	Current?	
Balance ID		Calibrated?	
pH Meter ID		Calibrated?	

Titants

NaOH	Normality	Expiration	
H2SO4	Normality	Expiration	

Blank

Blank	mL	Titrant	mL
NaOH		H2SO4	

Sample + 20 mL 1 N NaOH	Weight (mg)	Titrant 1 N H2SO4 in mL
Sample 1		
Sample 2		
Sample 3		

Calculation: ((Blank – Sample)(180.2)(N H2SO4/ N NaOH) = mg Product
 %RSD = (SD/average)*100

Sample	mg
Sample 1	
Sample 2	
Sample 3	
Average	
%RSD	

Performed by: _____ Date: _____

SOP: 2007.0219_rev05

Typical Objectives for LIMS/LES

- Validated source for regulatory compliance
- Automate manufacturing workflows
- Electronic Batch Record (EBR) support
- Facilitates self Documenting processes
- Enables cross functional KPIs
- CAPA / PACA evidence
- Complement manufacturing/ERP processes
- Achieve data integrity quality data consistency
- Stability studies
- Environmental monitoring
- Complements CDS systems
- Resource to enabling workflow optimization

Typical Objectives for ELN

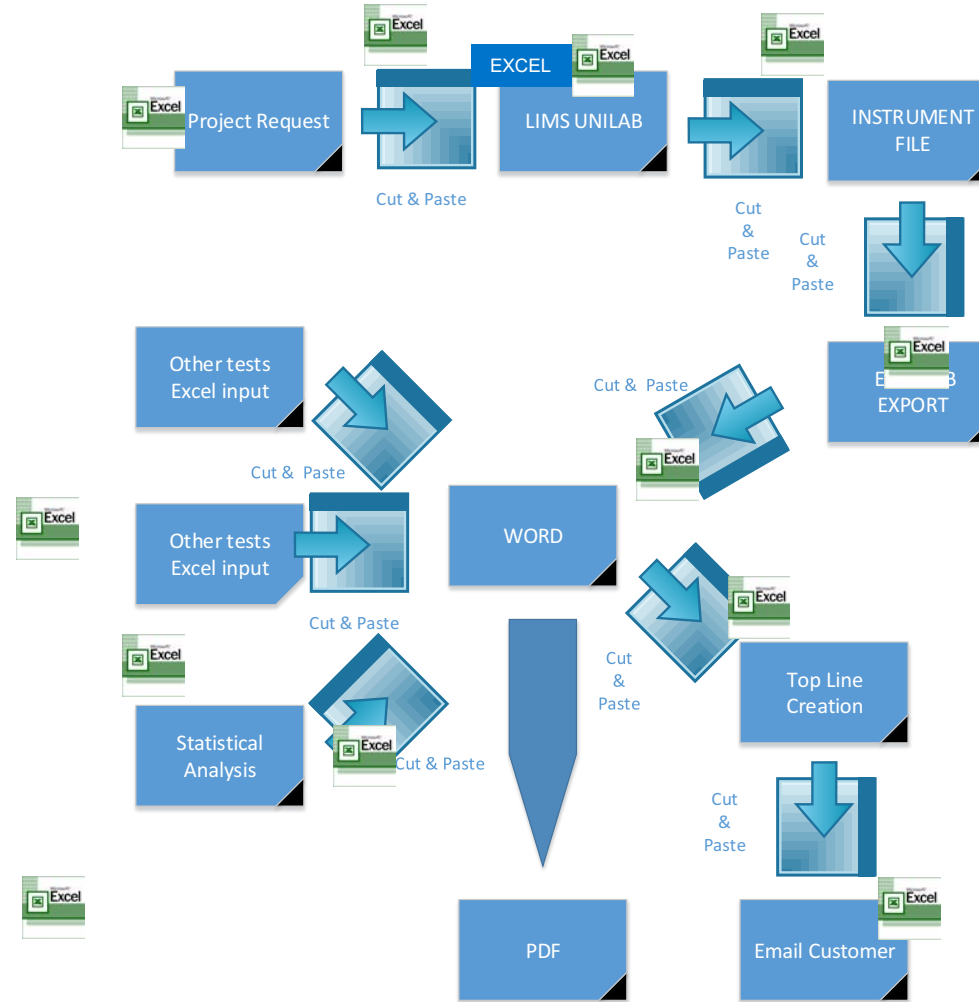
- Enables to re-use scientific insight
- Facilitates prior knowledge sharing
- Resource for data analytics and formulation optimization
- Self Documenting data acquisition processes
- Exposing KPIs
- Integrates external partner collaboration
- Achieve sustainability goals - reduce paper volume
- Support migration from hybrid (paper) legacy processes
- Achieve data integrity quality data consistency
- IP protection (legal)
- Simplifies tech transfer processes
- To attract new scientists

Data Integrity starts @ the source



Data Integrity nightmare

COPY/PASTE Madness



Metadata, why important?

- Without Meta data this is just a photo.



Meta data:

Speed, when, where, photo ID, last calibration, ...

Self-documenting processes

Reduces the value of data integrity at the source



Source: How to Improve Data Integrity
Laboratory Informatics Guide (LIG 2016) / Scientific Computing World

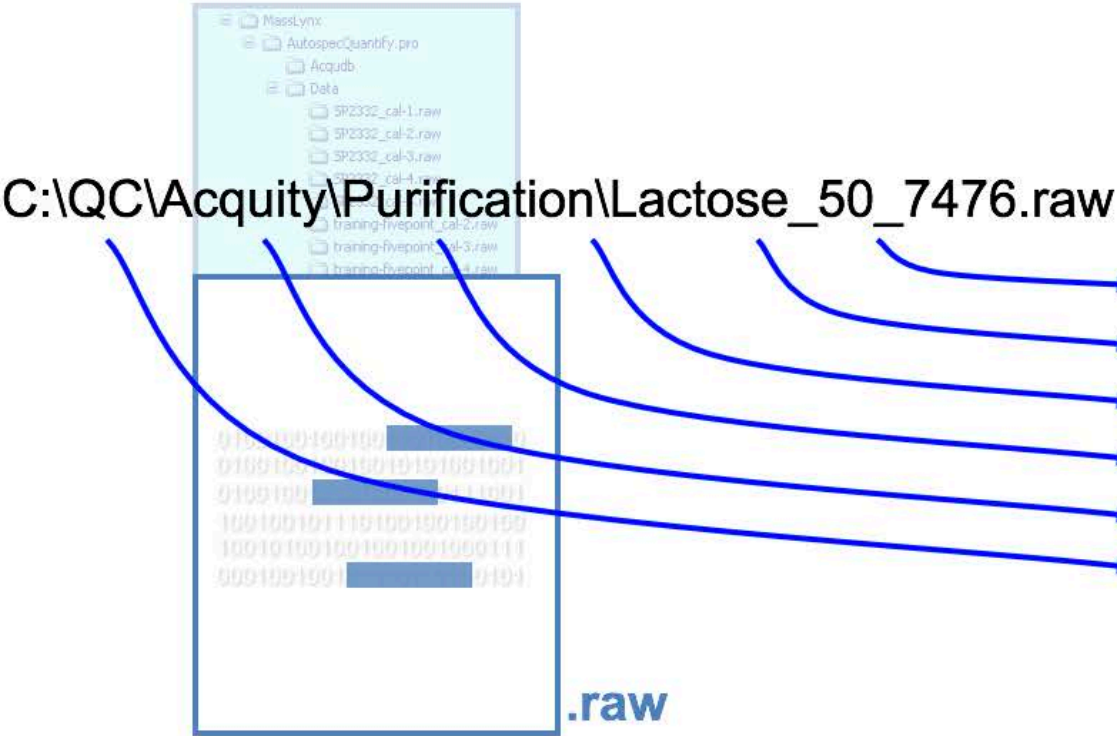
Self-documenting processes

Reduces the value of data integrity at the source



Example of Self documenting processes in the laboratory

SDMS automated acquisition extracts system and fixed metadata
 plus metadata from file name and folder structure
 and metadata from file content



File	Lactose_50_7476.raw
Unique ID	236F 2536 4509 ...
Date/Time	01.01.2006/18:25:47
Source	PC2222
Batch	7476
Dosage	50
Compound	Lactose
Project	Purification
Instrument	Acquity
Department	QC
Channel	1143
Vial	B04
Method	3/116

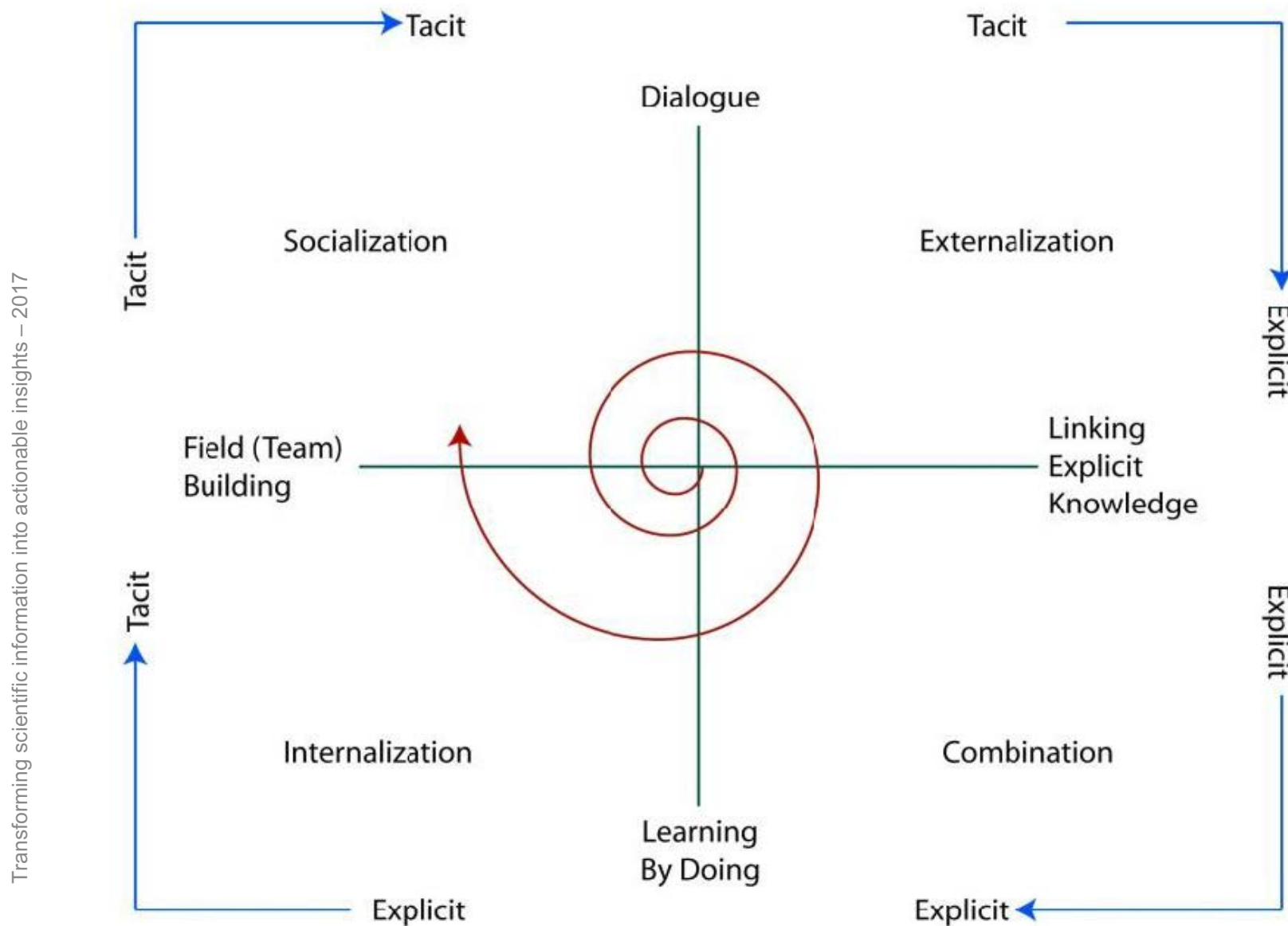
Source: WATERS Corp

#3 CHALLENGE

CREATING THE
RIGHT CONTENT

Knowledge is of
no value unless
you put it into
practice

Nonaka's *Spiral of Knowledge* framework for learning



Do we speak the same language?



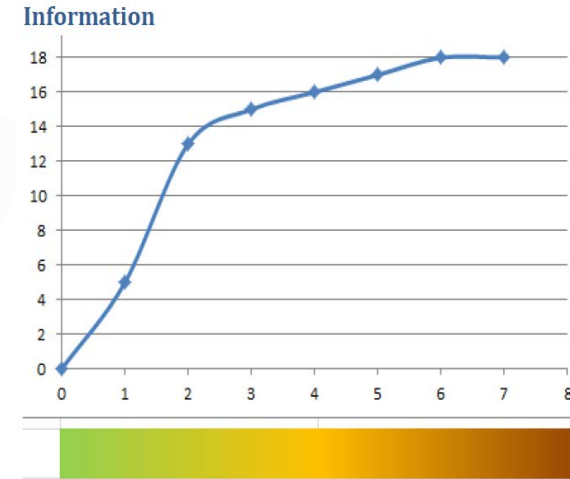
DATA

To tell the color
To tell what object
represents



INFORMATION

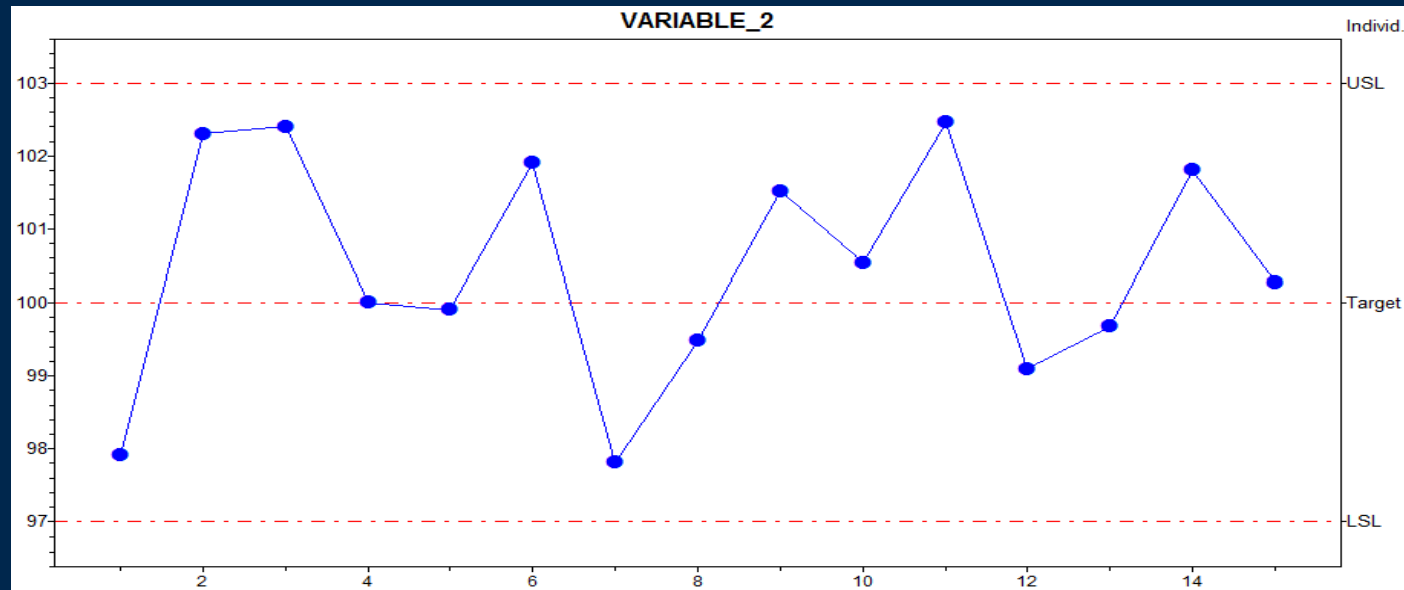
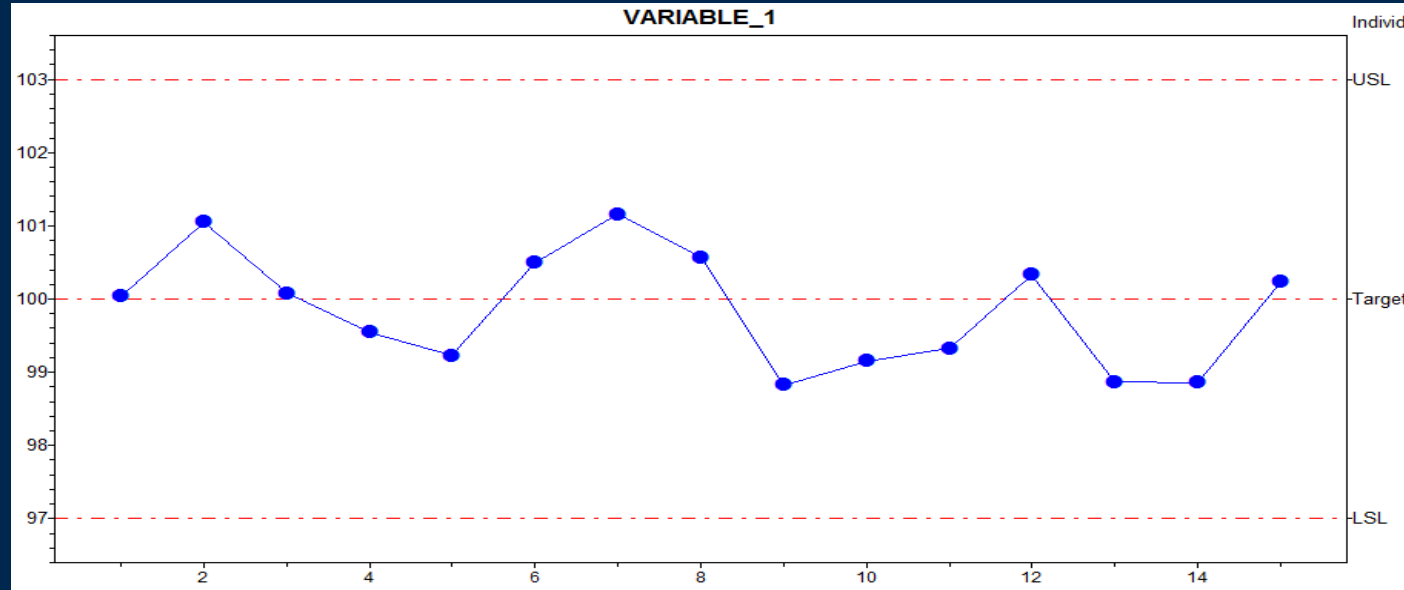
From color
Physical attribute
establish
sugar content
Chemical attribute



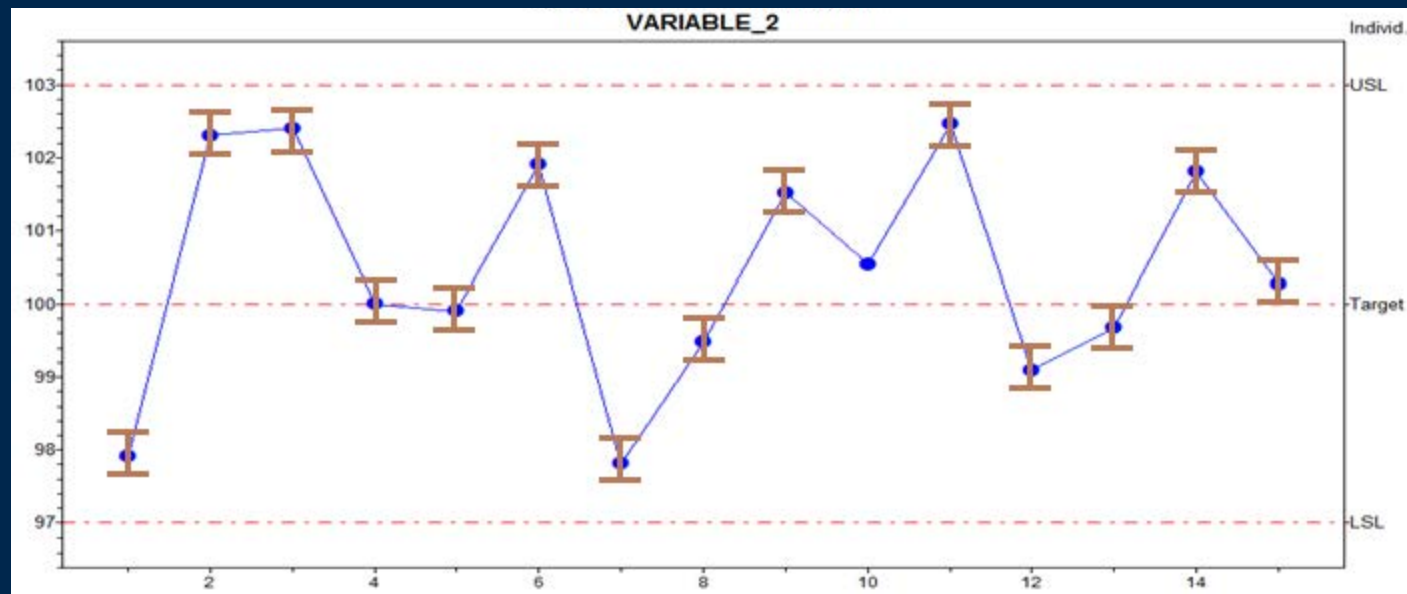
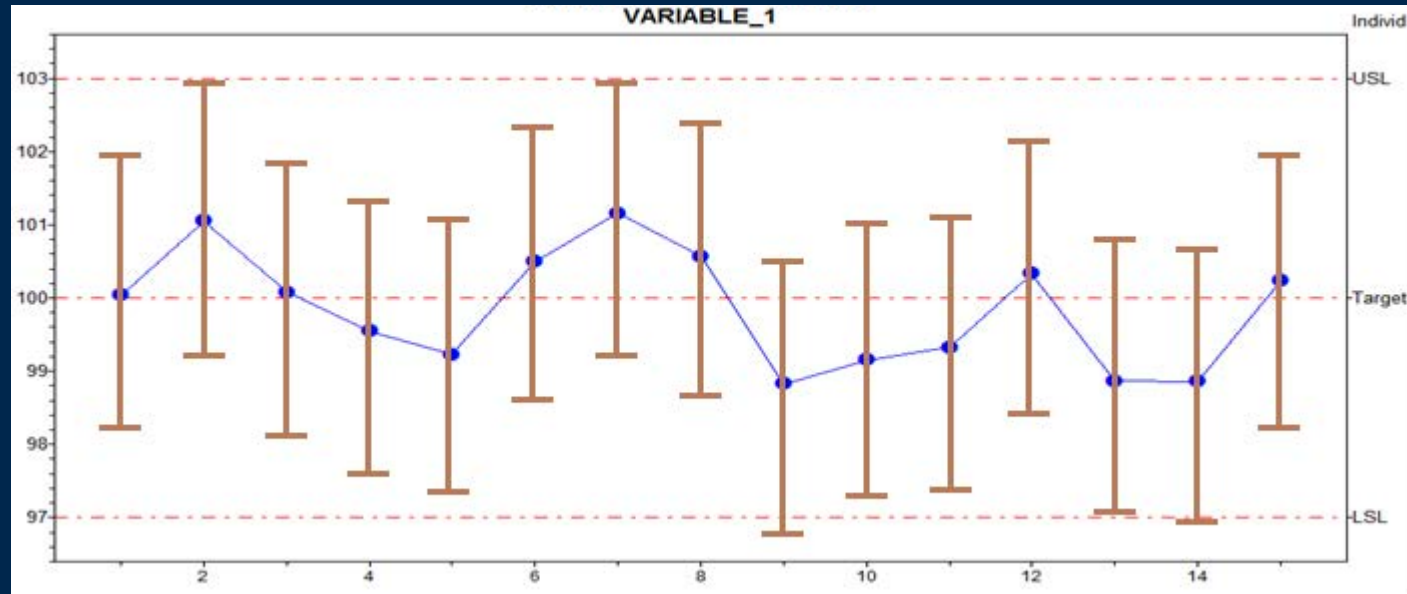
KNOWLEDGE / INSIGHT

What to buy
Make decision by **linking**
Information
to the objective

Measurement Uncertainty – which results do you like?



Measurement Uncertainty – which results do you like!



Graph Technology gains traction

Numbers

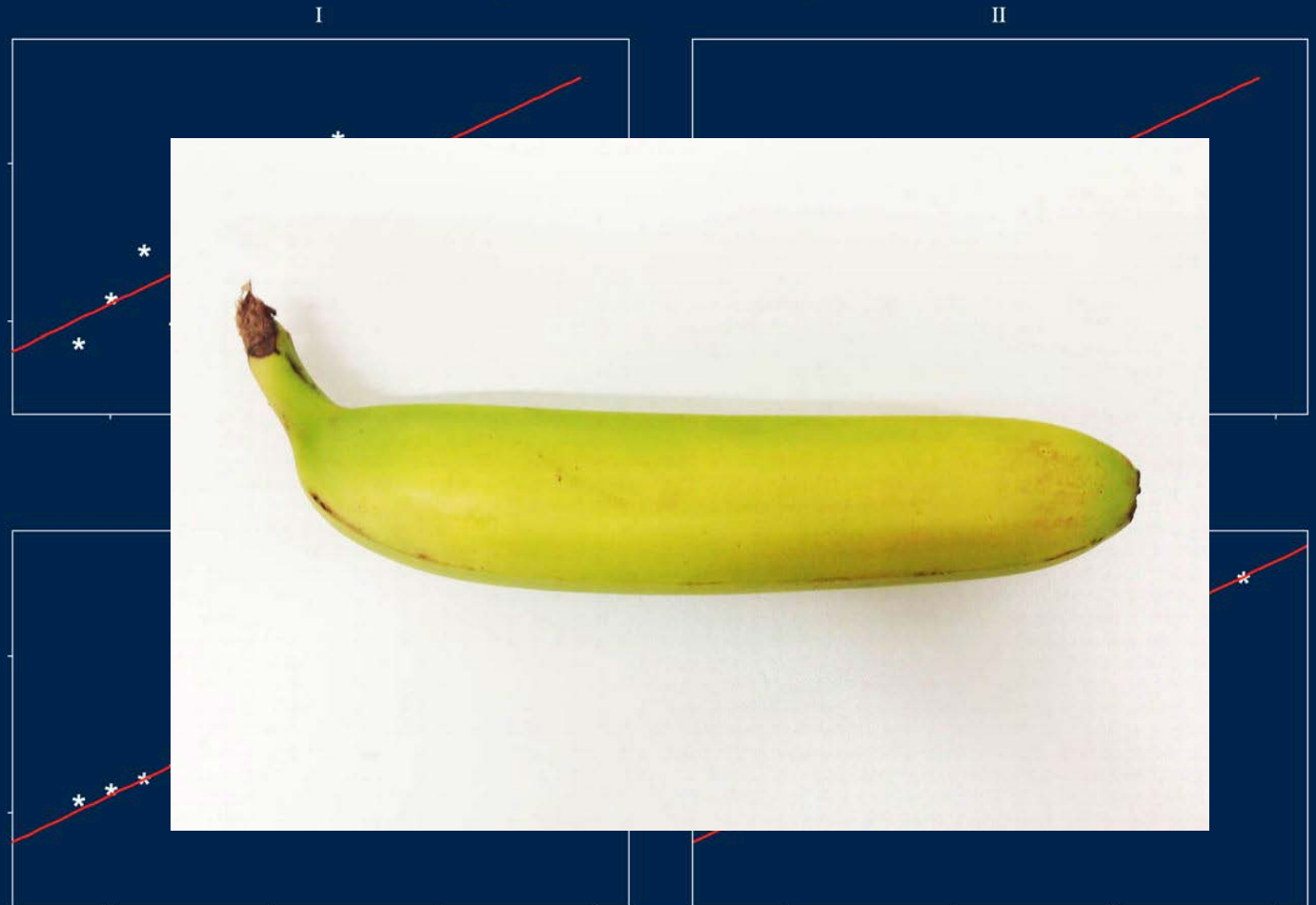
I		II		III		IV	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

F. J. Anscombe "Graphs in Statistical Analysis", American Statistician, 27 (February 1973), 17-21

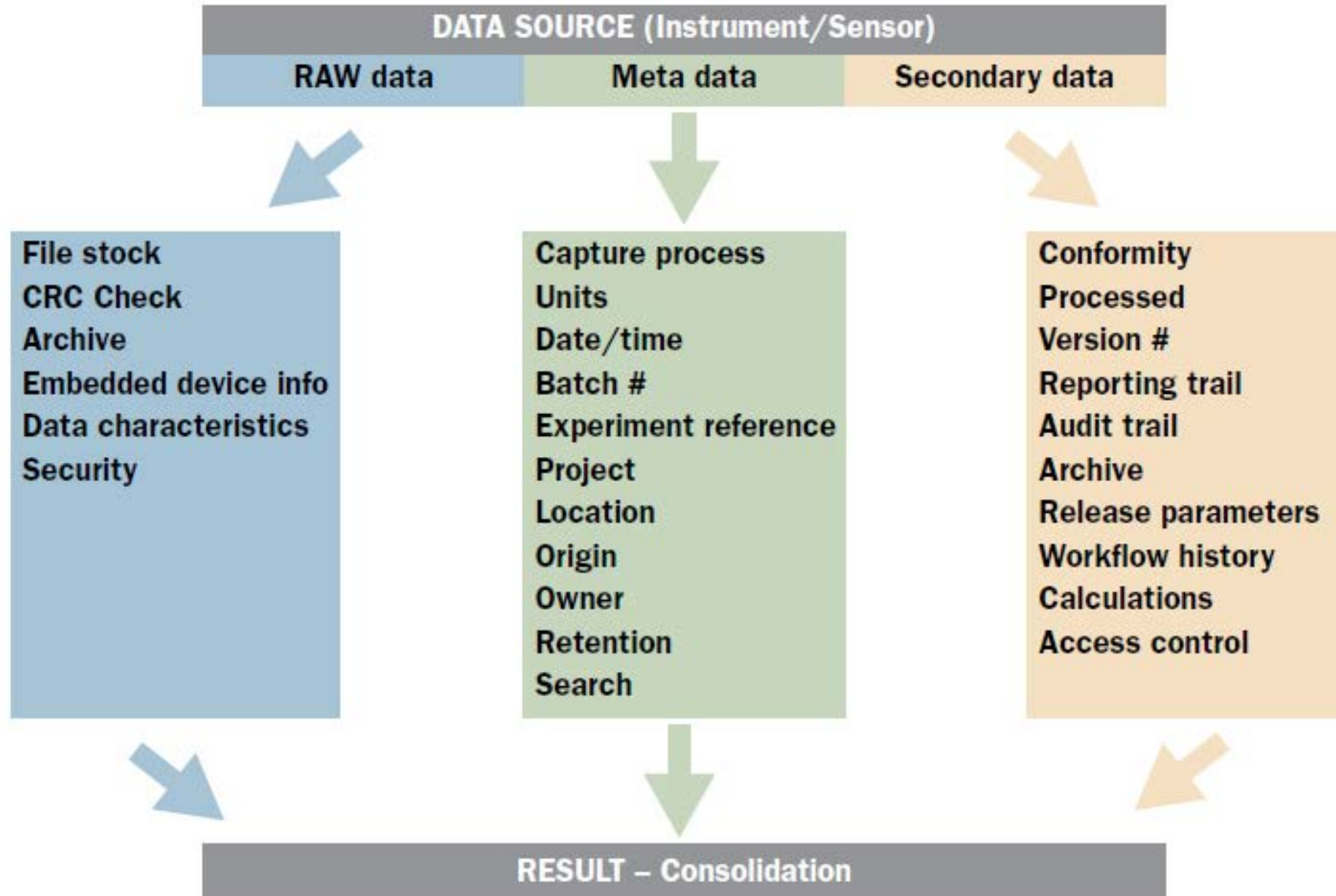
Statistics

Data pairs = 11
mean of X's = 9.0
mean of Y's = 7.5
standard deviation of X's = 3.32
standard deviation of Y's = 2.03
correlation = .82
regression line: $Y = 3 + 0.5 X$
r-squared = .67
slope t-statistic = 4.24
standard error of regression = 1.24

Graphical Analysis

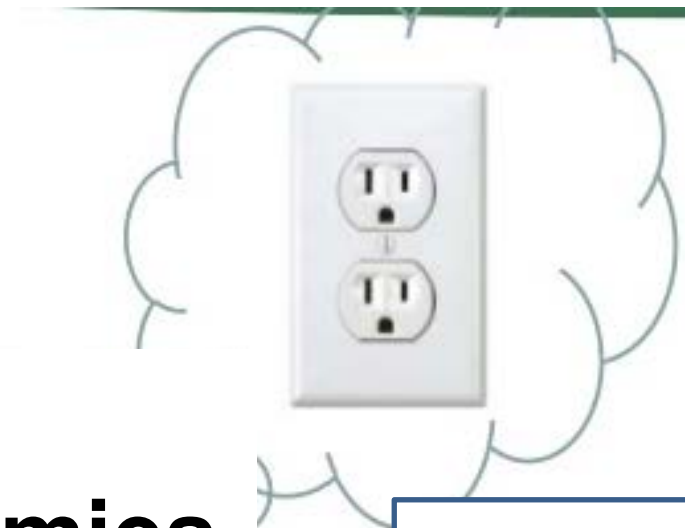


Scientific Data Sources



Importance of standards

Yes I have power.....



Apply consistent
ontologies and **taxonomies**
to assure **finding** the right
data



#4 CHALLENGE

EXPLORE NEW
CAPABILITIES

8.500.000.000 IoT (2016)



Internet of Things (IoT)



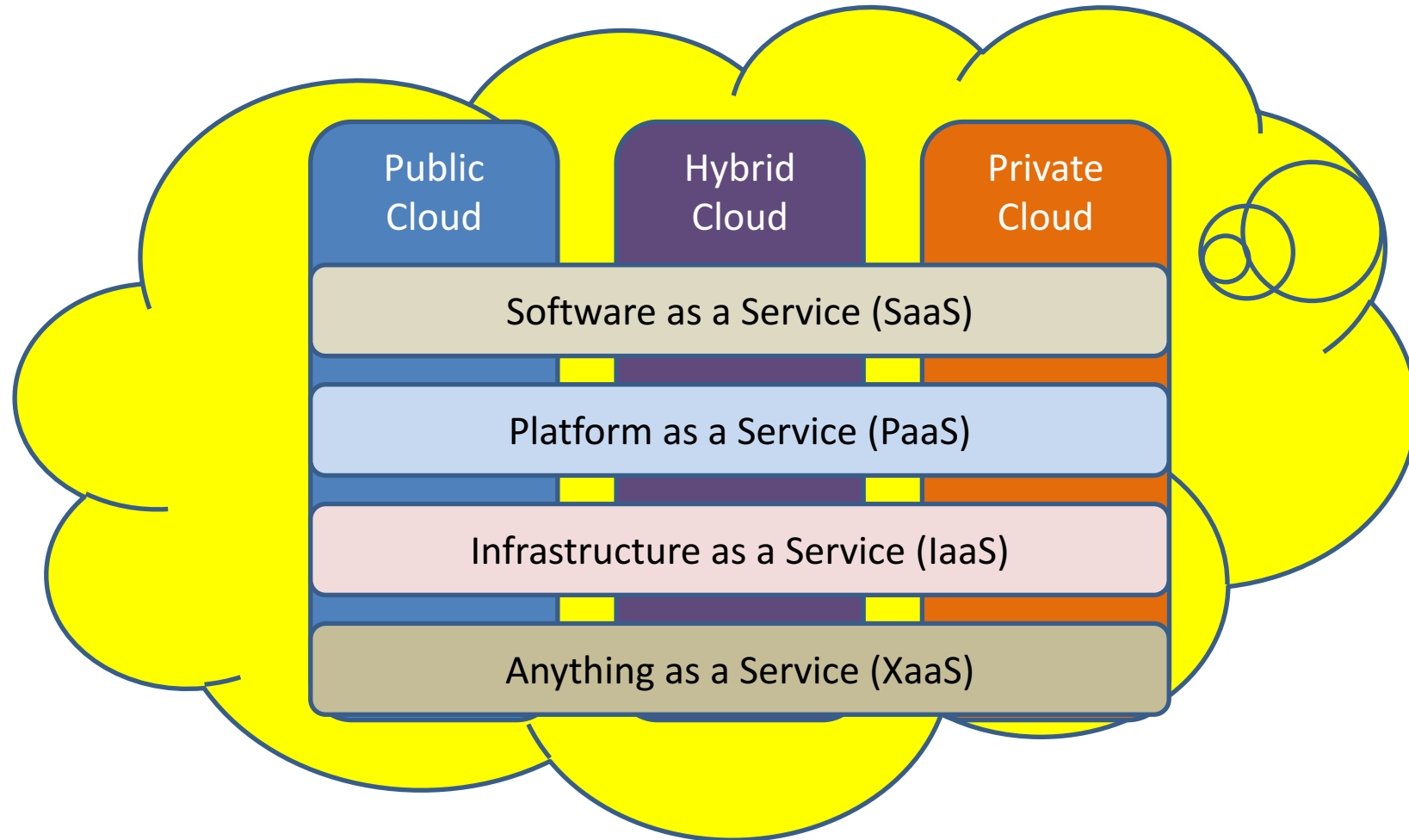
Transforming scientific information into actionable insights - 2017



Zelfs tell taes communiceren al draadloos

Zeilen - April 2017

Cloud related acronyms



SaaS ≠ Cloud

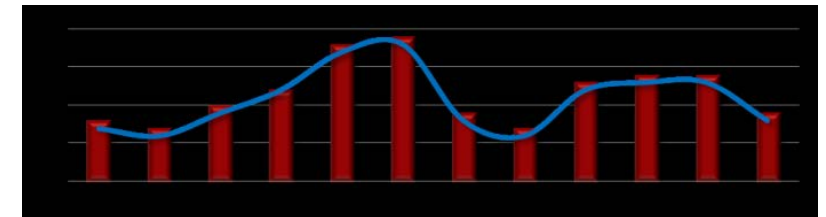
Cloud

- Service Level Agreement
- Where is my data?
- Emergency & escalation plan
- Back-up services
- On-demand scalability
- On-demand capacity
- Ownership

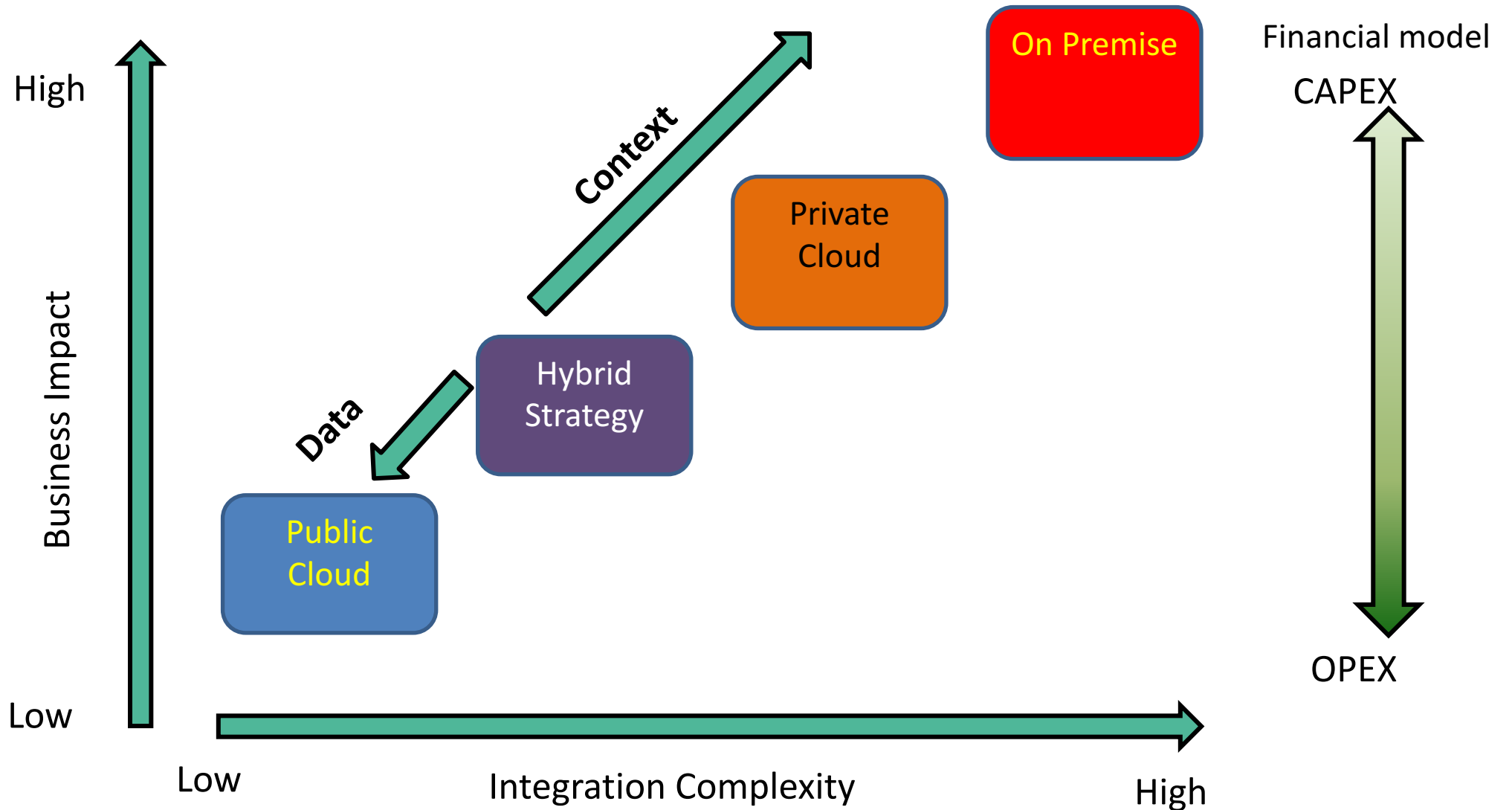


SaaS

- Subscription
- Free and paid services
- Application consultancy
- Application support
- Automatic updates
- No investment in software
- Minimal hardware investment

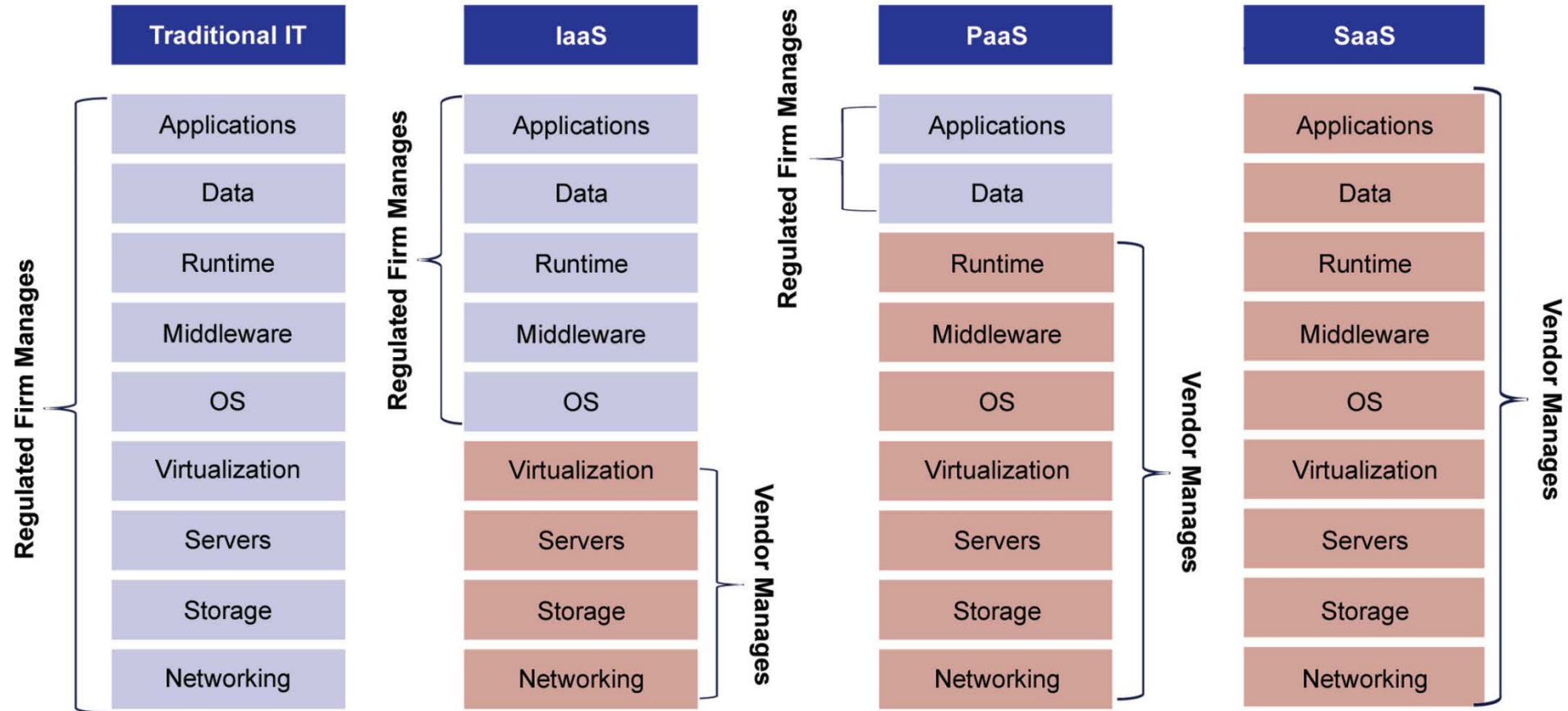


Accepted cloud strategies



The Route to clarity

GAMP Cloud Computing Special Interest Group (SIG)



LAB {SaaS + Cloud} candidates for the laboratory

- When **prototyping** requiring a lot of IT infrastructure
- Applications that **require mobile access**
- Applications **outside of the core** business
- Applications that require a **high upfront investment** (CAPEX vs OPEX)
- Projects requiring **intensive (external) collaboration**
- When **fast access to increased scalability** is a required
- Applications where **21CFR is not a direct requirement**

Industry 4.0 IoT Potential for Laboratories

Real-time Data Capture	Self-documenting process to include meta data and content in single process
Predictive maintenance	Predict equipment failures before they happen, and systematically prevent them
Remote monitoring	Remote monitoring solutions collect live data from assets , and use that data to trigger automatic alerts and operational actions based on current conditions, such as remote diagnostics and maintenance requests.
Deep learning data analytics	Ability to include big-data in R&D projects
Instrument interaction	Auto registers material and equipment that you are using. E.g. predict reagent consumption and initiate action when reagent is low with supplier
Customer/Consumer experience	Will provide contextualized and personalized C2C experience. Focus on customer experience, including physical safety and security
Enables multi technology validation	Ability to automatically compare real-time test results from multiple technologies on a large scale
Automate to transform	Integrate with existing Lab applications and explore to re-use existing dead-data silo's
Location tracking	Biometrically knows who you are. Automates user authentication process and reduces keystrokes. Simplifies application experience

New areas of scientific expertise

More QA and less QC

JANUVIA™ RTRT Control Strategy

Traditional Release Approach

PAT Testing Approach

API Blend

Lubricant Blend

Compress

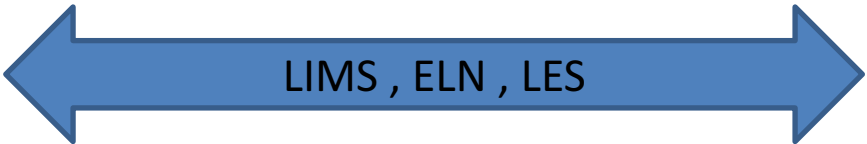
Film coat

Laboratory

Manufacturing Floor
NIR Composite Assay
On-line Dosage Uniformity by Weight

Manufacturing Floor
Disintegration
NIR ID
Appearance

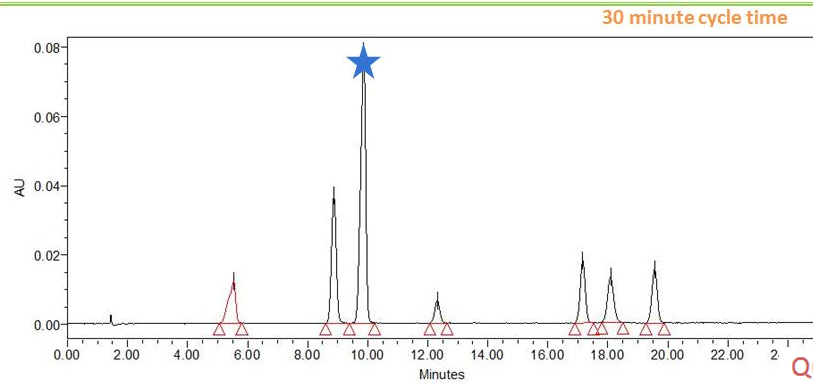
Lab Tests
HPLC
Content Uniformity
Composite Assay
Degrades
ID
Dissolution
Appearance



Transforming scientific information into actionable insights – 2017

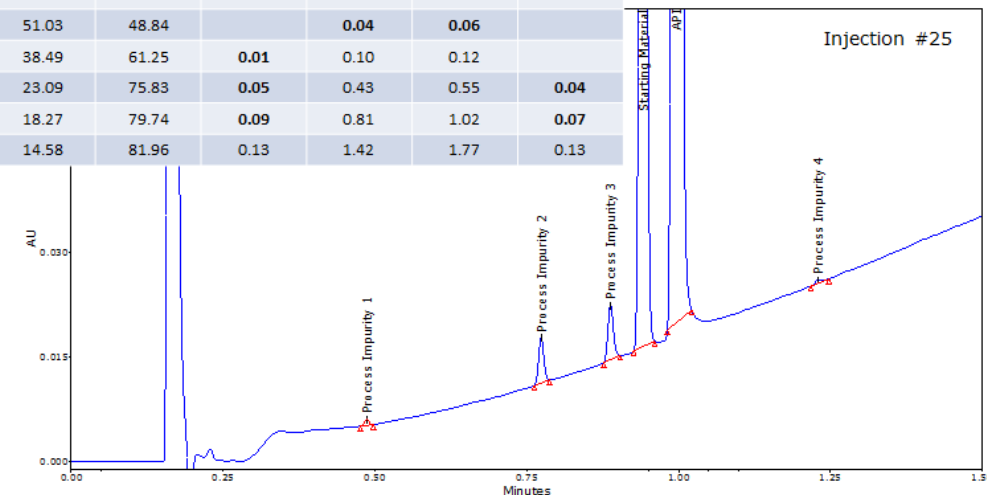
The scientist no longer just in the laboratory

Original HPLC Method
TOO SLOW for Online Analysis



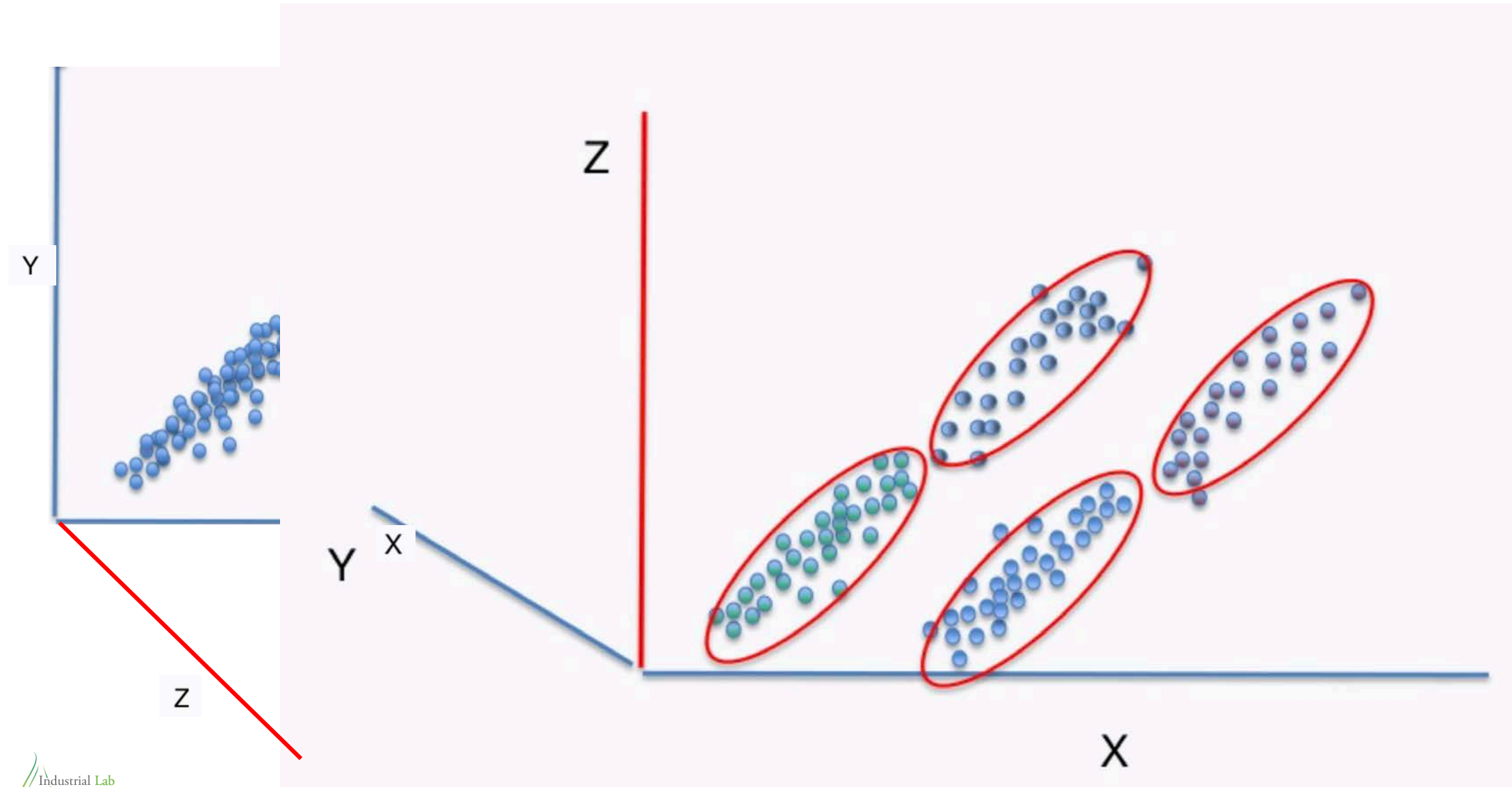
Quantify low level components in the presence of a high concentration API

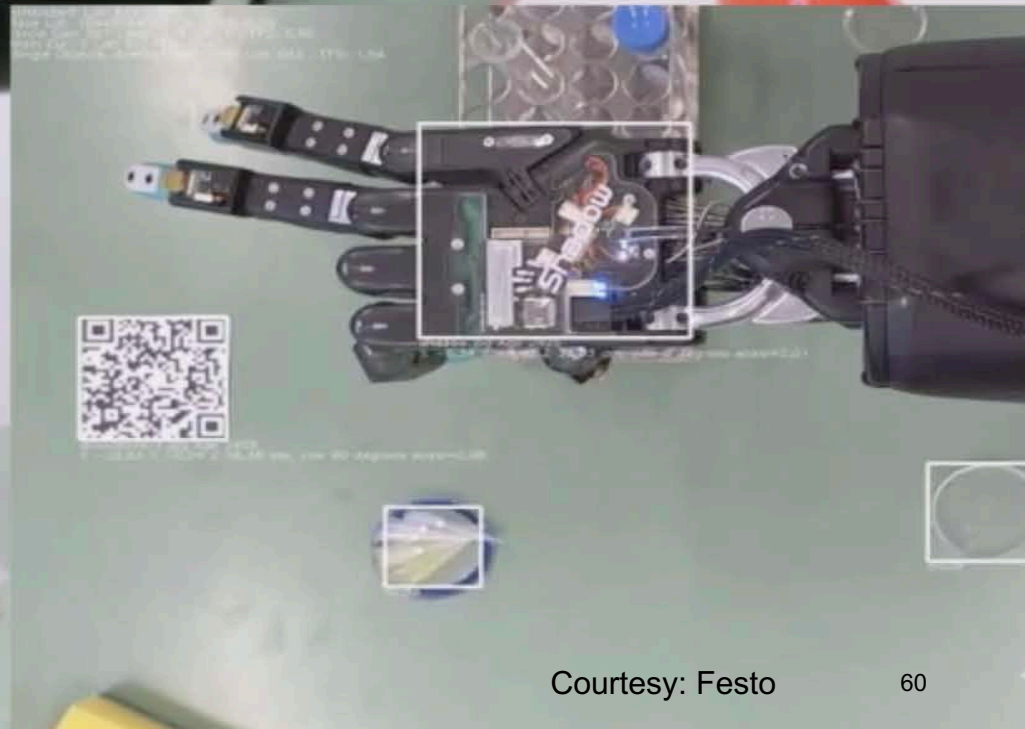
Reaction Aliquot	Starting Material	API	Process Imp#1	Process Imp#2	Process Imp#3	Process Imp#4
1	98.56	1.38				
5	89.54	10.43				
13	51.03	48.84		0.04	0.06	
17	38.49	61.25	0.01	0.10	0.12	
25	23.09	75.83	0.05	0.43	0.55	0.04
29	18.27	79.74	0.09	0.81	1.02	0.07
33	14.58	81.96	0.13	1.42	1.77	0.13



Courtesy WATERS Corp.

Multivariate analysis can reveal a change in correlation structure not visible with univariate analysis





XXXL

Big Data

Massive amounts of information derived from dry /wet lab investigations, feasibility studies and clinical trials.

Big Science

Research silos are evaporating with the merging of scientific methods. Traditional hypothesis-testing studies will couple with data-driven research.

Fast Data

Fast Data is a range of approaches that process data that might or might not be stored.

Big Collaboration

As evidence accumulates, personalized medicine will become a reality, and patient-specific disease interventions will become available. Teams of disease specialists, researchers and bioinformaticians working in concert in a virtual frontier.





BIG DATA LANDSCAPE 2017



INFRASTRUCTURE

HADOOP ON-PREMISE
cloudera Hortonworks
MAPR Pivotal
IBM InfoSphere
bluedata jethro

HADOOP IN THE CLOUD
amazon Microsoft Azure
Google Cloud Platform
IBM InfoSphere
treasure data
ibm aldiscale
CAZENA CenturyLink

STREAMING / IN-MEMORY
amazon databricks
confluent Strim
GridGain METAMARKETS
DATA TORRENT dataArtisans
ORACLE hazelcast TERRACOTTA

NOSQL DATABASES
Google Cloud Platform
ORACLE Amazon DynamoDB
Microsoft Azure MarkLogic
mongoDB DATASTAX
KERO SPIKE Couchbase
redislabs influxdata

NEWSQL DATABASES
SAP Clustrix
Cockroach LABS
mssql splic
MarkLogic
citusdata Trifolium
dooqdb paradigm4

GRAPH DBS
neo4j
IBM ORACLE
Microsoft Azure
Pivotal
snowflake
Infoworks

MPP DBS
TERADATA
VERTICA
NETERZA
Cobcon
kognitio
SQL
dremio

CLOUD EDW
amazon
Google Cloud Platform
Microsoft Azure
Pivotal
snowflake
Infoworks

DATA TRANSFORMATION
talend pentaho
alteryx TRIFACTA
tamr Paxata
StreamSets UNIFI

DATA INTEGRATION
informatica
MuleSoft
segment
TEALIUM
enigma
ALOMA
ZALONI
splenty
import
Stitch

DATA GOVERNANCE
informatica
IBM
skyhigh
collibra
Alation
Waterline

MGMT / MONITORING
amazon New Relic
APFDYNAMICS actio
WAVEFRONT
unravel
splunk
Yrcona
pagerduty
Numerator

STORAGE
amazon
Google Cloud Platform
Microsoft Azure
ALLUXIO
nimbora
Qumulo COHO
parasas

CLUSTER SERVICES
amazon
Google Cloud Platform
Microsoft Azure
ALLUXIO
nimbora
Qumulo COHO
parasas

APP DEV
lightbend
rainforest
CRSK

CROWDSOURCING
amazon
mechanicalturk
upwork
WorkFusion
CrowdSource

HARDWARE
Google TPU ARM
PERVIRA
Graphcore
MYTHIC
NVIDIA
Movidius
SCORTEX

ANALYTICS

DATA ANALYST PLATFORMS
Microsoft pentaho alteryx
Digital Processing guavus AYASDI
WATTIVO Datameer Quid
ClearStory OrigamiLogic interana
Bottlenose ARIMO ENDOR MODE

DATA SCIENCE PLATFORMS
IBM KNIME data
iku
DOMINO yhat rapidminer
CONTINUUM ANALYTICS
Alpine Angoss
ALGORITHMIA

BI PLATFORMS
Microsoft
amazon
Google Cloud Platform
looker Wave Analytics
ARCADA DATA
GoodData

VISUALIZATION
tableau SAP
amazon
Google Cloud Platform
qlik CELONIS
Periscope ZEPH
CHARTIO
Kognitio

VERTICAL ANALYTICS
PREDIX
G3 IOT CAPE
UPTAKE
Orbit Insight
TACHYUS
Alluvium
databrama

STATISTICAL COMPUTING
sas
SPSS
MATLAB

DATA SERVICES
Palantir
Palantir
DATA SCIENCE
kaggle
DataKind

MACHINE LEARNING
amazon
Google Cloud Platform
H2O
DataRobot
context relevant
VIZENZE
bonsai
deepinsight
Lutanian

HORIZONTAL AI
IBM Watson Cortana
Facel
sentient
Voyager
clarifai
Affective
senseME
Numenta
OSARO
CURIOUS AI
VISION

SPEECH & NLP
twilio
amazon
NarrativeScience
semantic
nuance
ARRIA
IDIBON
snips
Soundhound Inc.

SEARCH
elastic
Autonomy
ORACLE
Lucidworks
swiftype
alphasense
Searchkit
SINEOLIA

LOG ANALYTICS
sumologic
loggly
logz.io

SOCIAL ANALYTICS
Hootsuite
NETBASE
DATA SIFT
synthesio
bitly
predata

WEB / MOBILE / COMMERCE ANALYTICS
Google Analytics
mixpanel
sumall
retention
SIGOPT
granify
custora

APPLICATIONS - ENTERPRISE

SALES
CHORUS
INSIDESALES.COM
conversica
clari AVISO TACT
fuse|machines TROOPS

MARKETING - B2B
RADIUS App Annie
EVERSTRING Lattice
infer MINTIGO
sense tubular
DataFox ZENAGGIO

MARKETING - B2C
Zeta bloomreach
blueyonder [PERSADO]
ACTIONIQ
Kahuna BLUECORE
SAILTHRU QUANTIFIND
mparticle Ampero

CUSTOMER SERVICE
MEDALLIA zendesk
CLARABRIDGE NGDATA
CLICKFOX
DigitalGenius
AUTOMAT
msgd INTERCOM

HUMAN CAPITAL
HireVue entelo
hiQ GIGSTER
Rextia RESTLESS HANDB
Wade&Wendy
Clustro Stella
pymetrics

LEGAL
RAVEL
Seal
Everlaw
JUDICIALIA
Brevia
R S S
casetext

FINANCE
anaplan
Zuora
tidemark
S4HANA
TRADESHIFT

ENTERPRISE PRODUCTIVITY
slack
ORACLE
Lumina diffbot
butter ai KASIST

BACK OFFICE AUTOMATION
HyperScience
apricity
AppZen

SECURITY
TANIUM
CYLANCE StackPath
DARKTRACE Illumio CODE42
VECTRA Thecat Metrix DataGravity
cyberason ANTRIA CephCloud Guardian
SCANIFY ANOMALI sift science
BlueTalon Recorded Future
Clara tallia
SECURE IDEAL
FRONTSCALE
SiftLogic

CROSS-INFRASTRUCTURE/ANALYTICS

amazon Google Cloud Platform Microsoft IBM SAP Hewlett Packard Enterprise sas 101 data vmware TIBCO TERADATA ORACLE NetApp

APPLICATIONS - INDUSTRY

ADVERTISING
AppNexus
criteo XAD
rocketHub
theTradeDesk
drawbridge
TAPAD DataXu
Oppier

EDUCATION
KNEWTON
Clever
Cleara
kidaptive
MOAT
Algorithms
Liventent
gumgum
Databases
Yieldmo

GOVERNMENT
Socrata
OPENGOV
mark43
FiscalNote
OpenDataSoft

FINANCE - LENDING
OnDeck Affirm
Kreditech AVANT
INSIKT
MoneyLion
TrueAccord
trocly
cignifi
aire
Active AI

FINANCE - INVESTING
Dataminr
KENSIC
Quantopian
NUMERAL
ISENTUMY
claritymoney
ALGORIZ AEDVA
PavenPack

REAL ESTATE
Opendoor
vts
CREDIC
reonomy
COMPSTAK

INSURANCE
Lemonade
CYENCE
SHR Technology
Tractable

HEALTHCARE
FLATIRON
HealthTop
Gingerio Glow
COTA zebra
img
entric
Qventus
AETHER
BAYLARS
IMAGEN

LIFE SCIENCES
23andMe color
zymergen
BenevolentAI
ZEPHYR
Clear Labs
Citrine
twoLAR
Atomwise
sag
genomics

TRANSPORTATION
UBER
TESLA
CLEARPATH
drive.ai
AICure
PILTOAI
OTTO
PTIMUS
riexar
comma.ai
Civil Maps
netrodyne
NIO

AGRICULTURE
FARMERS
FarmersEdge
FarmLogs
BLUE RIVER
mavrx
terrotron
prospero

COMMERCE
instacart
STITCH FIX
RetailNext
HowGood

OTHER
eHarmony stem
nathan
robotics
BOSS
RACINE
hoppin
BOXEVER
select
UN
Unbabel
Second Spectrum
duetto
Apacch

OPEN SOURCE

FRAMEWORK
Hadoop
Flink
YARN
TEZ
SPARK
MESOS
CDAP

QUERY / DATA FLOW
Spark SQL
presto
SLAMBATA
DRILL
Google Cloud Dataflow

DATA ACCESS
cassandra
nifi mongoDB
CouchDB
SciDB
CouchDB
CouchDB
riak
HBASE
Spanner
accumulo

COORDINATION
talend
Apache Zookeeper
Apache Ambari

STREAMING
Spark
Flink
kafka
druid
STORM

STAT TOOLS
python
ScalaLab
SciPy

AI / MACHINE LEARNING / DEEP LEARNING
theano
Caffe
TensorFlow
Apache KINCA
OpenPI
neon FeatureFu
DSSTNE
mlib
DL4J
DEEPLARNING4J
MxNet
DM TK
Keras
VELES
DIMSUM
Aerosolve

SEARCH
elasticsearch
Solr
Elasticsearch

LOG ANALYSIS
kibana
logstash

VISUALIZATION
BEAKER
Rodeo

COLLABORATION
jupyter
ANACONDA

SECURITY
Apache Ranger
KNOX
Sentry

DATA SOURCES & APIS

HEALTH
JAWBONE
VALIDIC
practicefusion
fitbit GARMIN
Human API kinsa

IOT
GE Digital
UPTAKE ThingWorx
helium
samsara

FINANCIAL & ECONOMIC DATA
Bloomberg THOMSON REUTERS
DOW JONES
S&P CAPITAL IQ CBRIGHTS xignite
quandl
YDLEE
PREMISE
estimize
Eagle Alpha
StockTwits
PLAID
mattermark
Thinknum

AIR / SPACE / SEA
PLANET
Airware
spire
SKYWATCH
AIRBOTICS
AEROSTORY
WINDWARDS
TELLUSLABS
DroneDeploy
MarsTraffic

PEOPLE / ENTITIES
acxiom
Experian
EPSILON
InsideView
Crimson Hexagon
BASIS
quantcast
SAFE GRAPH

LOCATION INTELLIGENCE
FOURSQUARE
Sense
PlaceIQ
esri
CARTO
Mapillary
STREETLINE

OTHER
qualtrics
DATA.GOV
dataworld
panjiva
enigma

DATA RESOURCES

INCUBATORS & SCHOOLS
PLURAL SIGHT
DataCamp
DataElite
INSIGHT
The Data Incubator
galvanize
METIS

RESEARCH
facebook research
OpenAI
MIRI
ESALL
ALLEN INSTITUTE
ARTIFICIAL INTELLIGENCE

#5 CHALLENGE

MAKE IT EASY

ACCESABLE

Barriers

Technology

Almost disappeared

Globally accessible & available

People

Challenge

Requires a mindset change

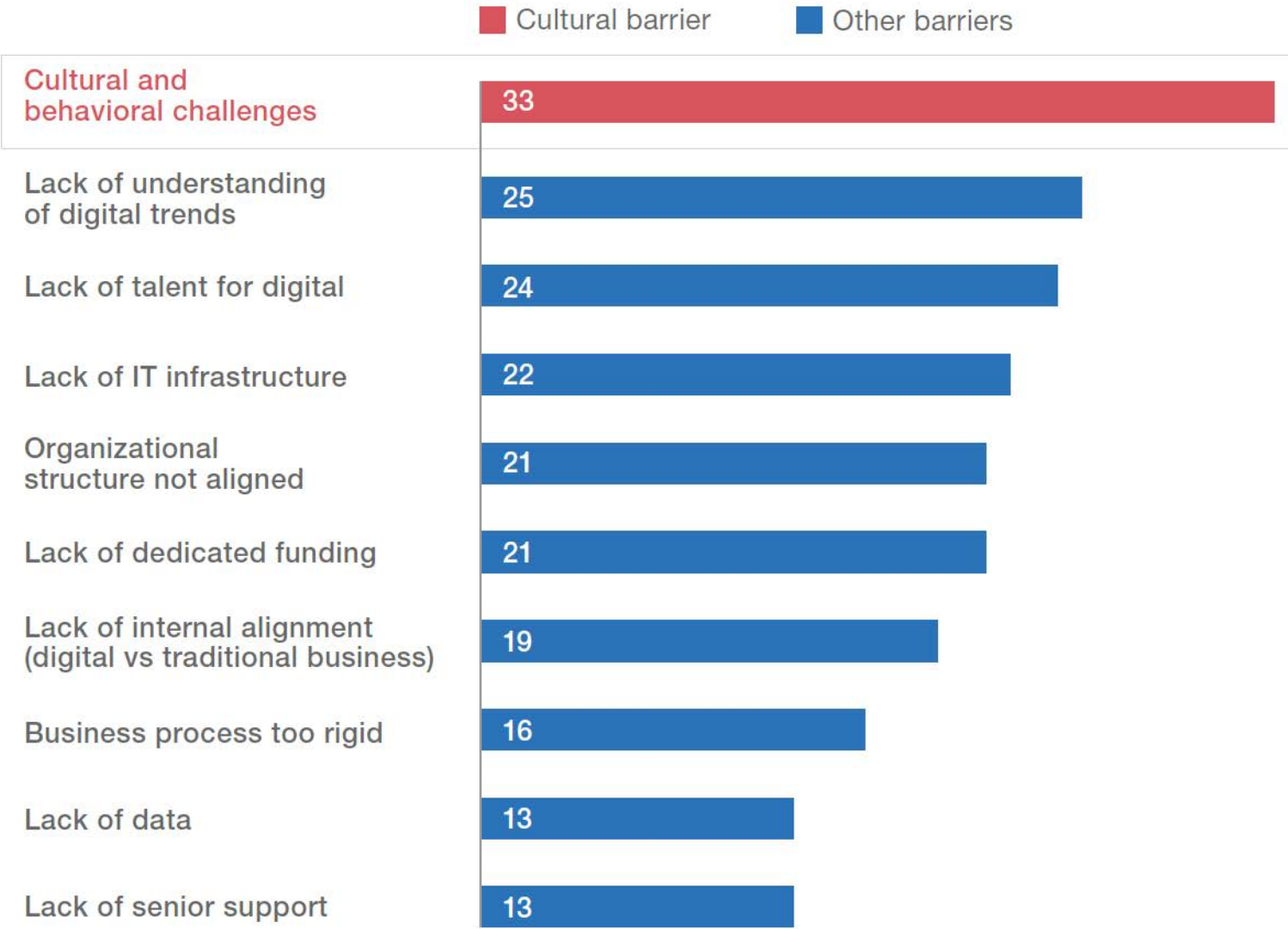
COMMENTARY LETTERS

Heng
CAI/NYTS

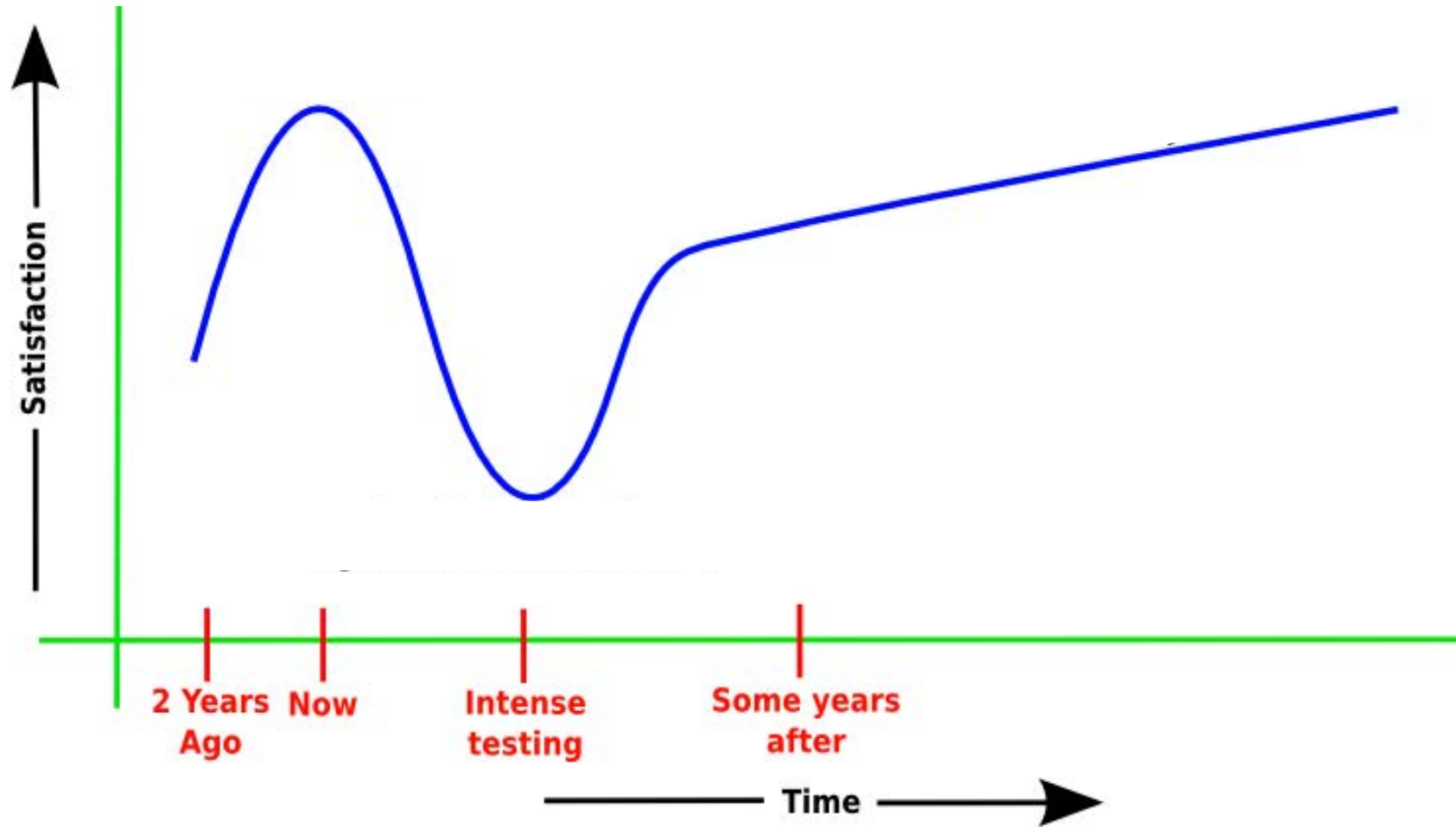
BANK



Significant challenges to meeting digital priorities

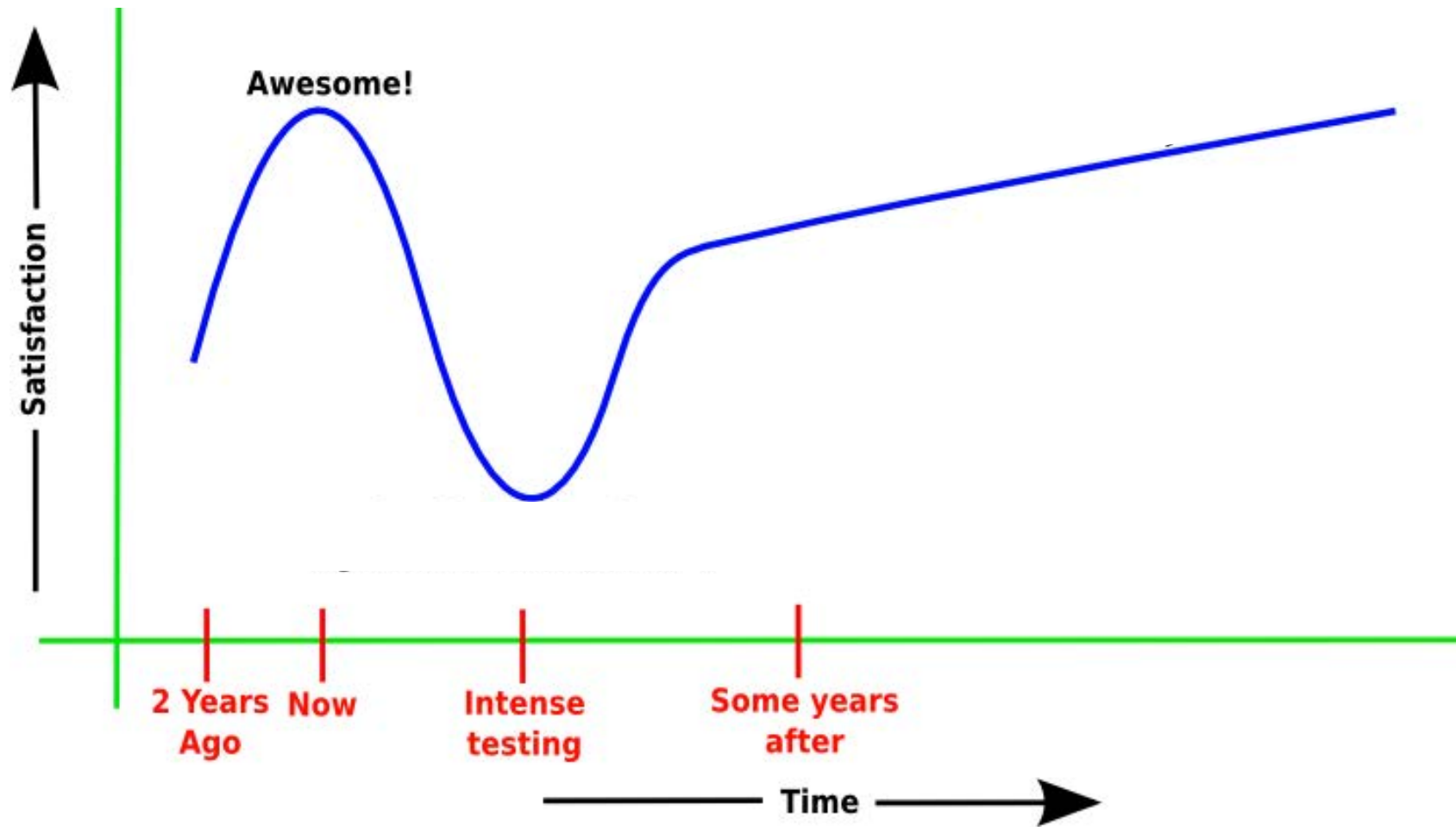


Adoption curve



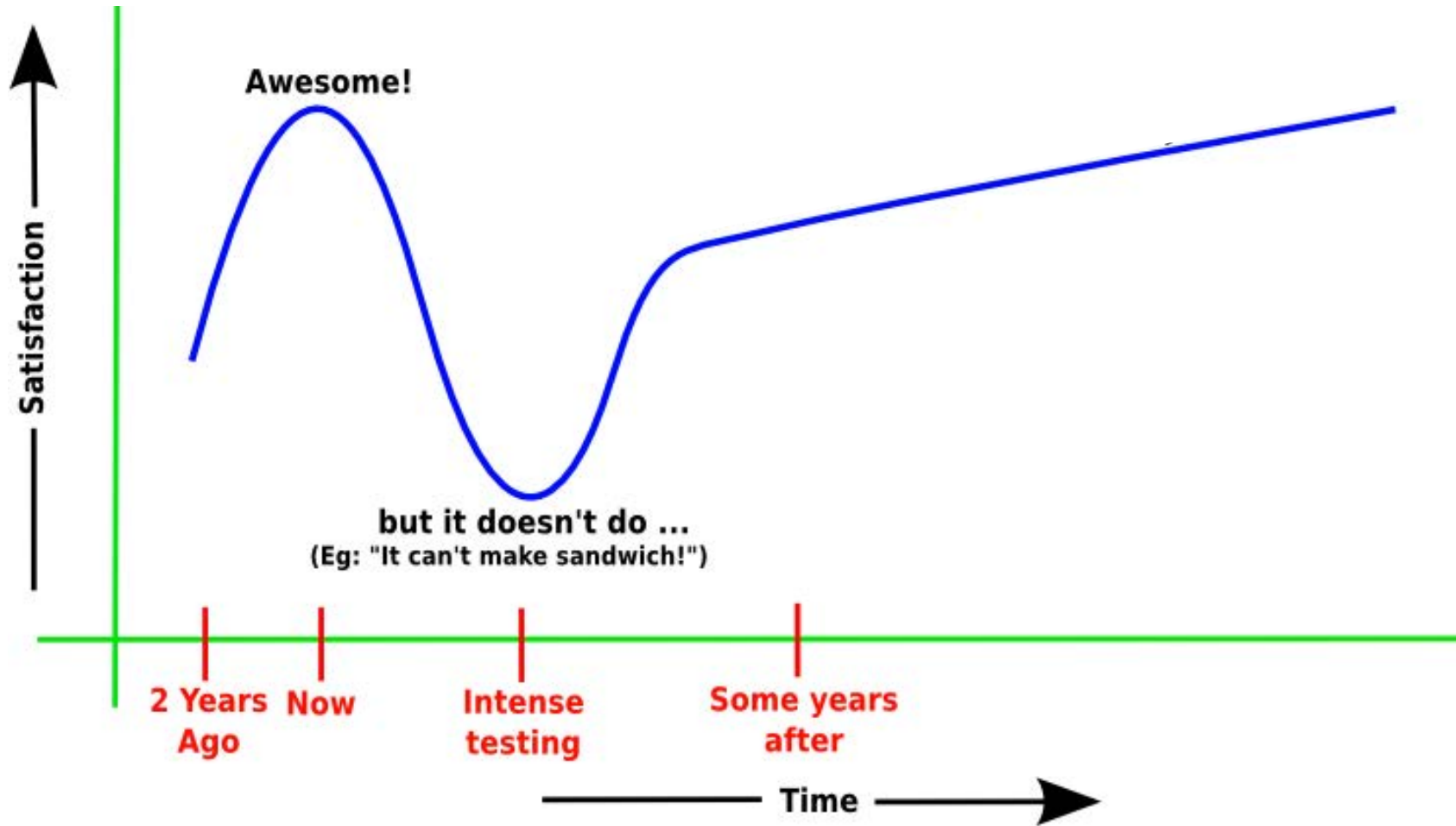
Source: tech.puredanger.com

Adoption curve



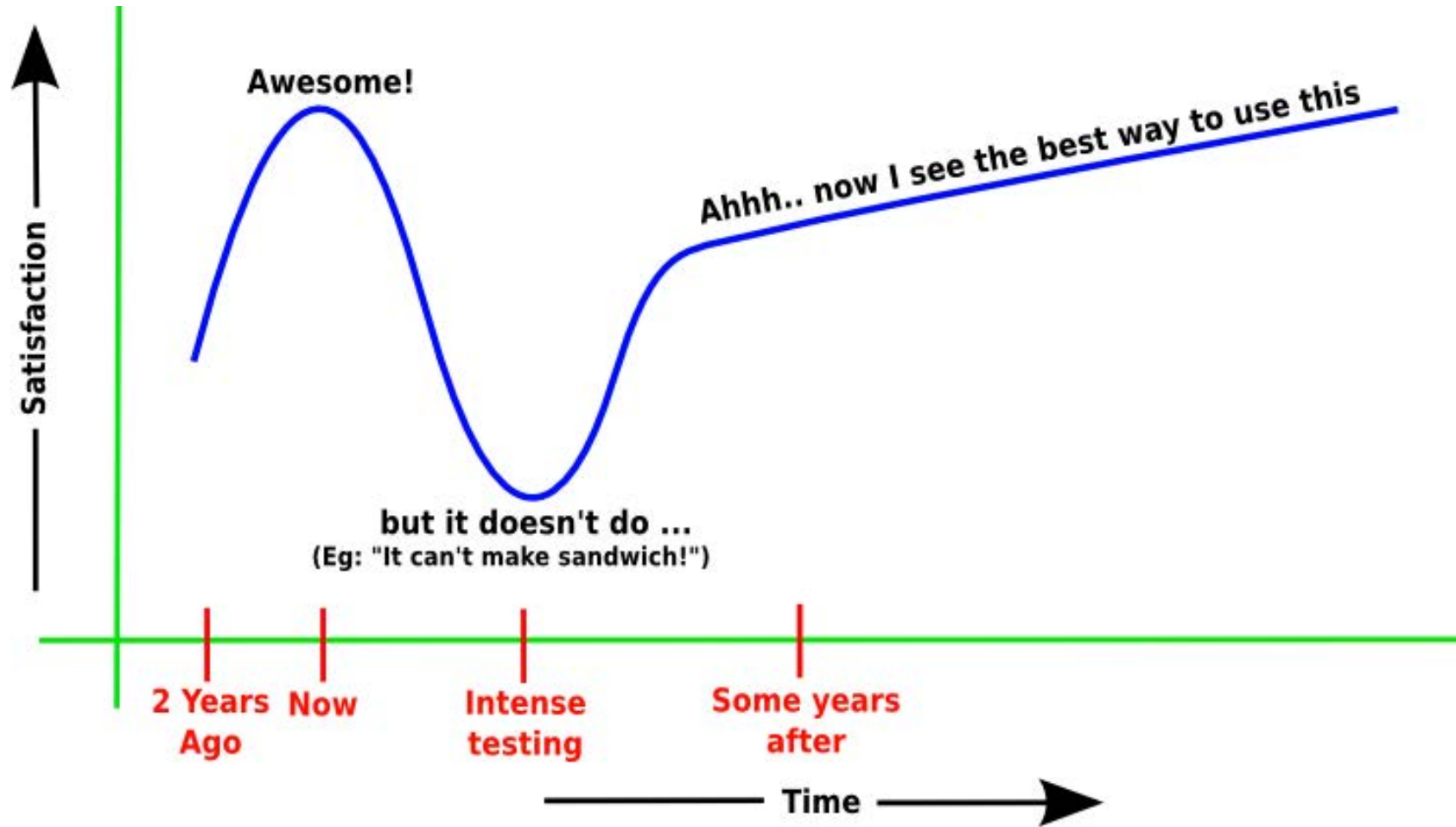
Source: tech.puredanger.com

Adoption curve



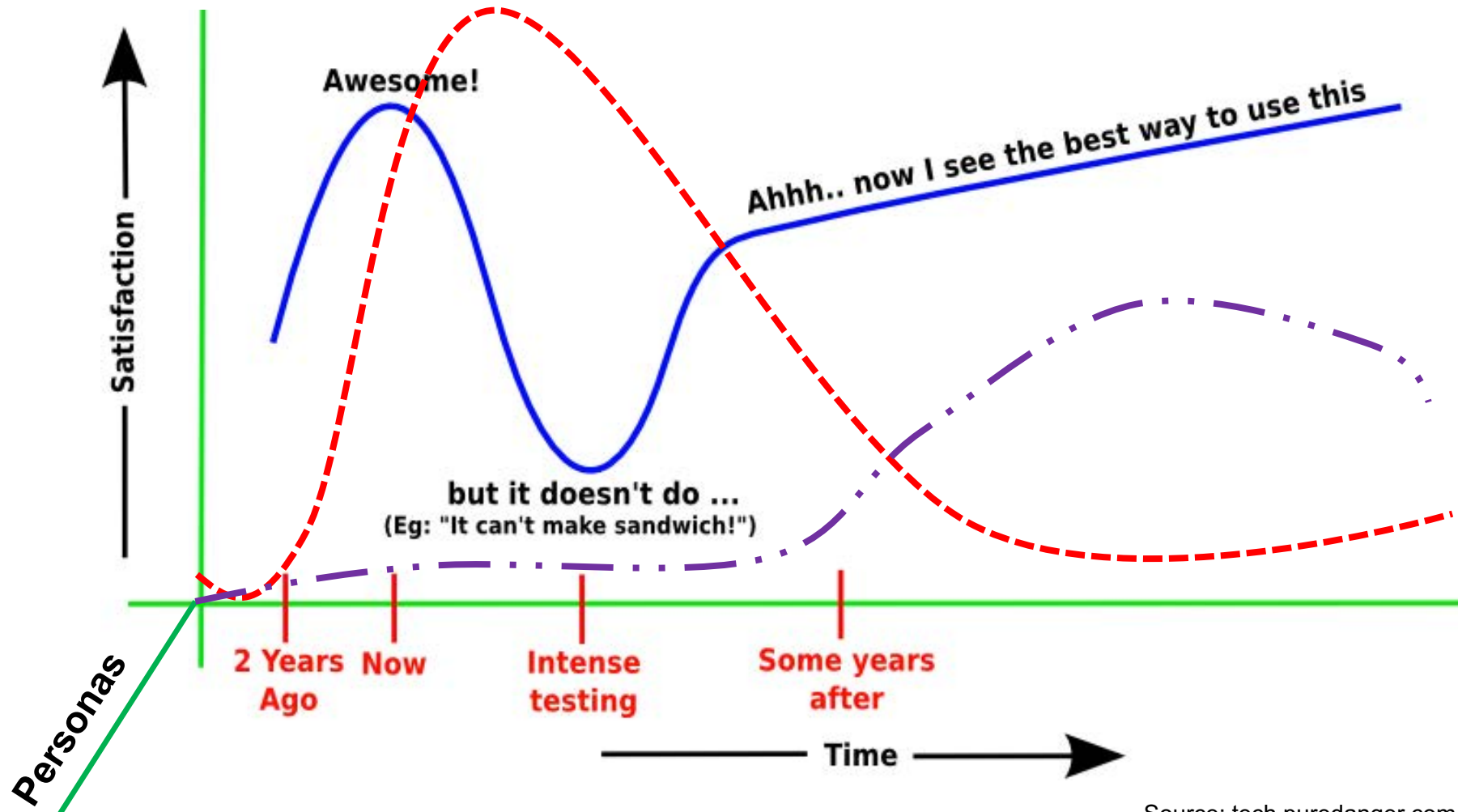
Source: tech.puredanger.com

Adoption curve



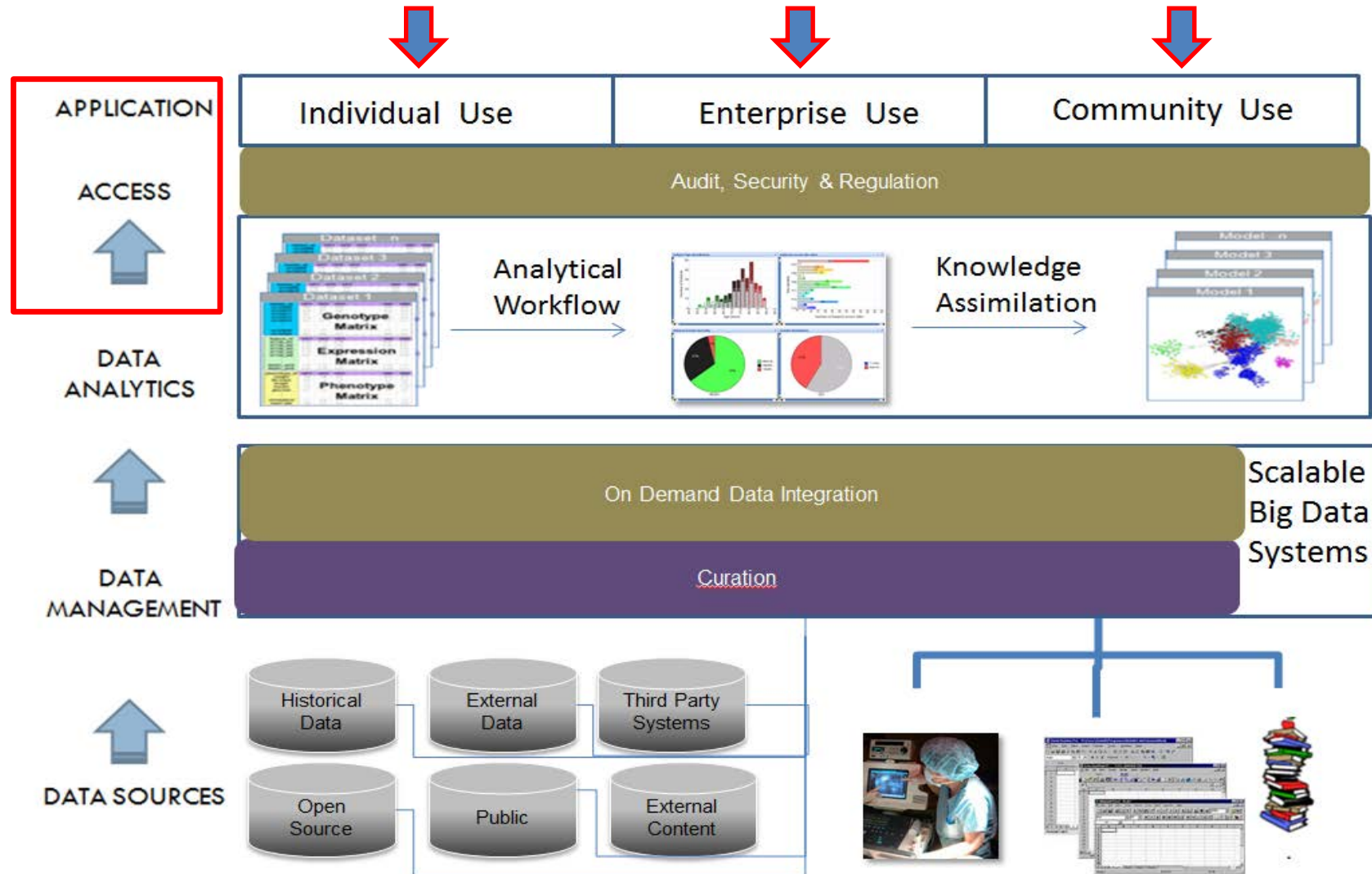
Source: tech.puredanger.com

Adoption curve



Source: tech.puredanger.com

Models of communication processes

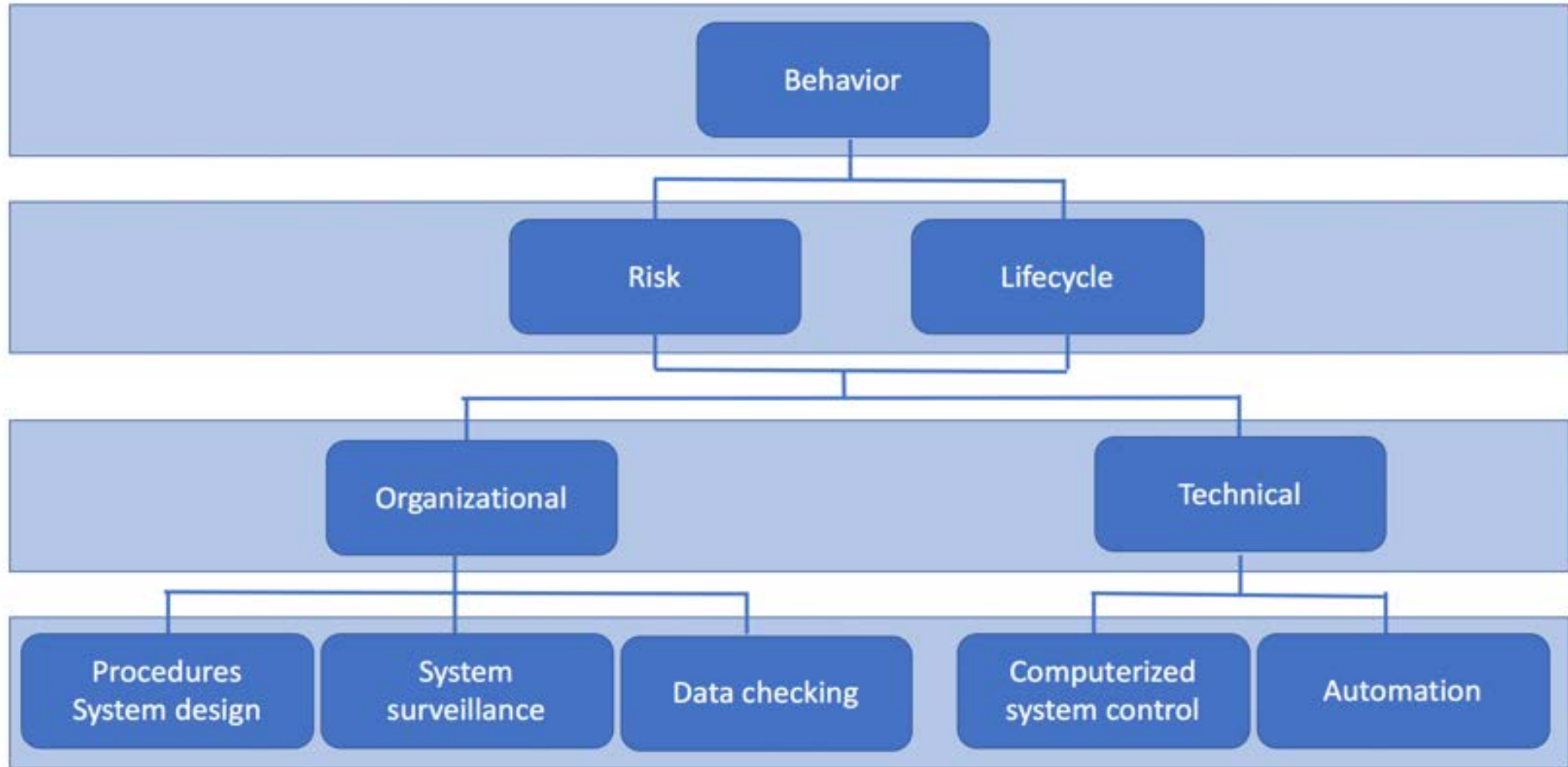


#6 CHALLENGE

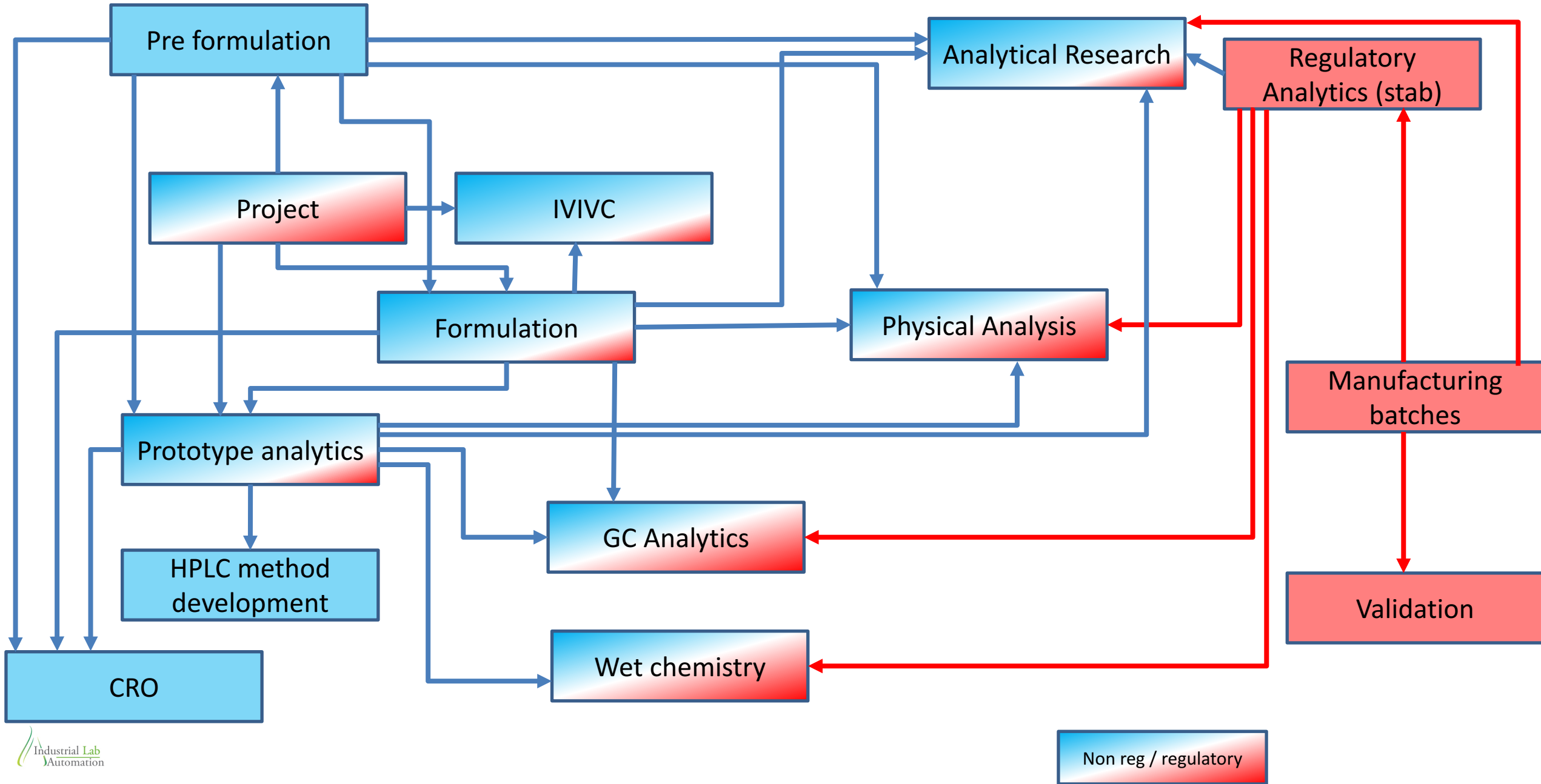
ASSURE DATA
GOVERNANCE

Data integrity refers to
maintaining and assuring
the accuracy and
consistency of data over
its entire life-cycle

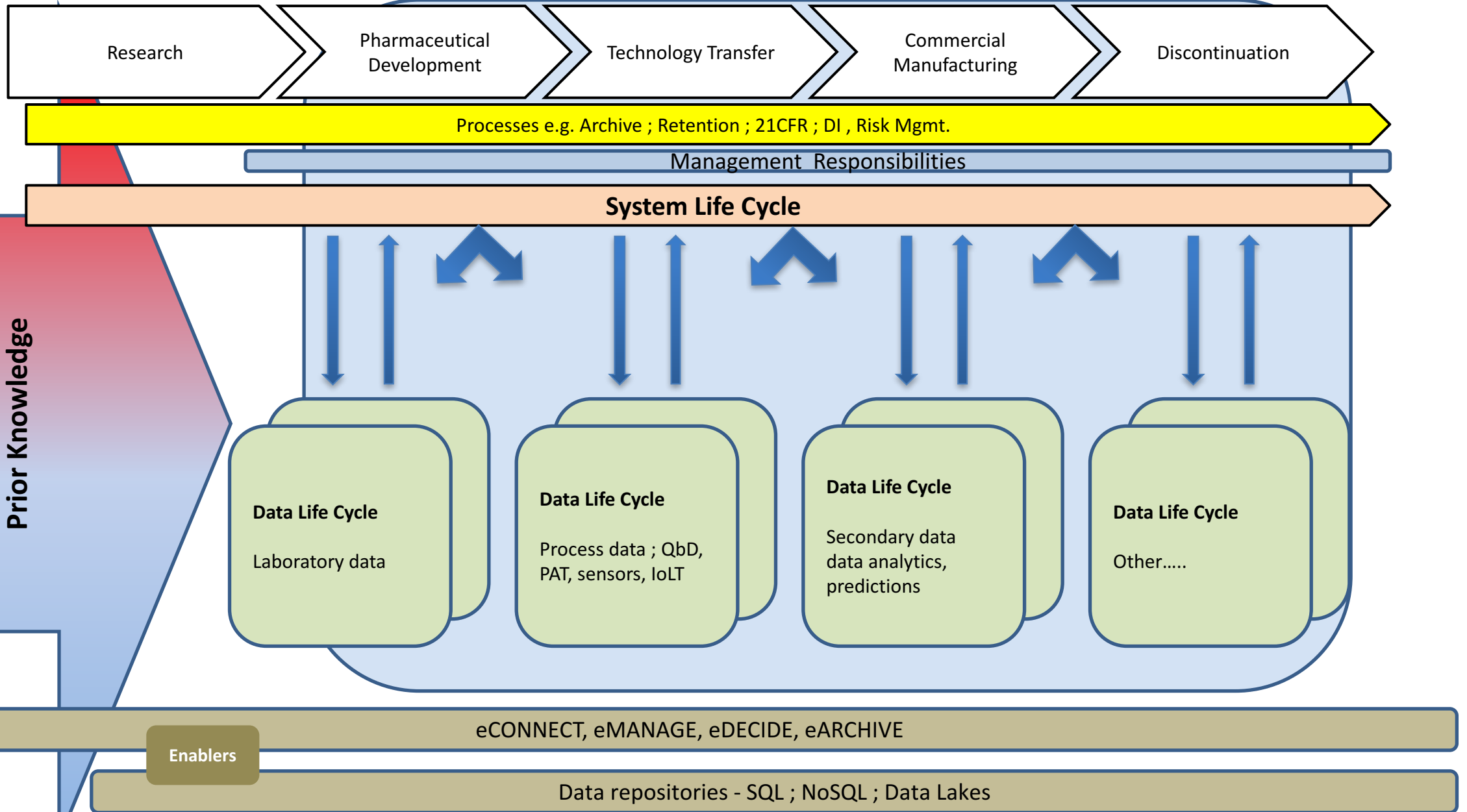
Change management requires a data lifecycle based approach



Cross departmental regulatory requirements

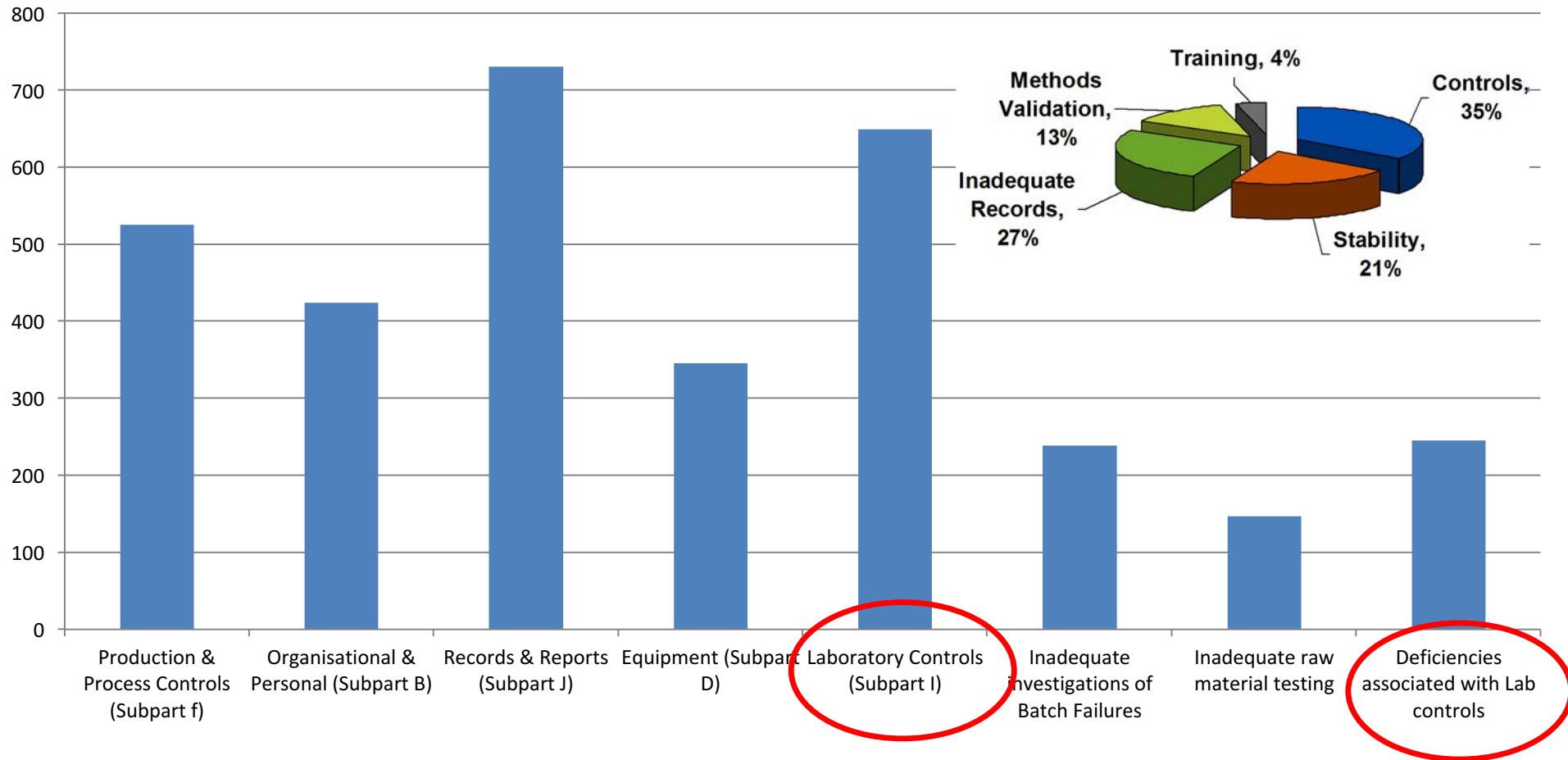


Empowering the e-DATA LIFE CYCLE



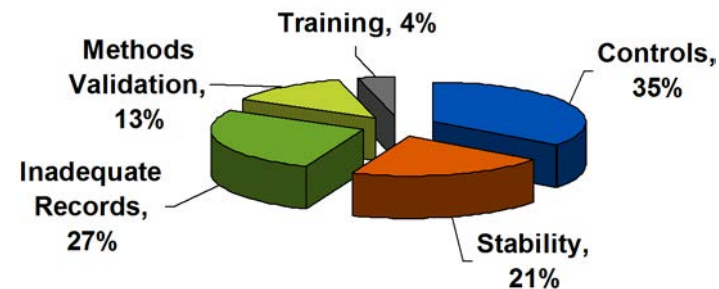
US FDA Observations Summary

21 CFR Part 211 Observations CURRENT GOOD MANUFACTURING PRACTICE FOR FINISHED PHARMACEUTICALS



Laboratory data integrity observations

- **Alteration** of raw, original data and records
- Multiple analyses of assay with the same sample **without adequate justification**
- Manipulation of a poorly defined **analytical procedure** and associated data analysis in order to obtain passing results
- **Backdating** stability test results to meet the required commitments
- Creating acceptable test results **without performing** the test
- Using test results from previous batches to **substitute testing** for another batch



Source: FDA

Attention areas: data governance

- Data security protocols
- Master Data management
- Each interface is a potential Data Integrity challenge
- Long-term data access control
- Privacy regulations incl. new European GDPR compliance
- Enforcement of data and ontology standards

Transforming scientific information into actionable insights

Take away message

- People
 - Data consumer vs data creator mindset change
 - Think data lifecycle
 - Think capabilities vs technology first
 - Avoid applying previous excuses
- Technology
 - Apply industry data access & security processes
 - Apply cross industry standard technologies
 - Apply visualization to data analytics
 - Utilize 24x7 global IT technology to reduce TOC
 - Include roadmap to address upcoming technology obsolesce
 - Go mainstream
- Processes
 - Apply as much self documenting context & meta data in all processes
 - Apply consistent standards (incl. ontologies and taxonomies) to assure finding the right data
 - Involve management to assist breaking internal silo barriers to address data life cycle
 - Define long-term ownership of master data

De stand van zaken van data management software in het laboratorium

We expect too much in 1 year And not enough in 10



Peter Boogaard

Industrial Lab Automation

17 November 2017 – Wageningen – Nederland

peterboogaard@industriallabautomation.com

