



illuminate[®]



Deploying Machine Learning in Data Management

PHUSE EU Connect 2020

Sheila Rocchio
eClinical Solutions

Agenda

1. Industry Trends Interviewing Data Review
2. AI Opportunities to Accelerate Data Review
3. Three Use Cases for AI Enabled Review
4. Summary

Industry Trends: Modernize Data Management

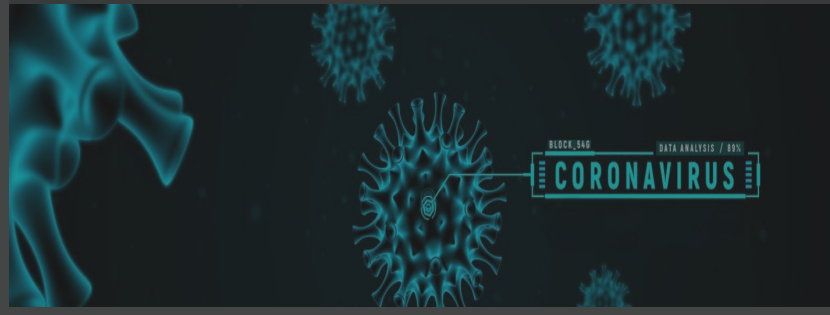
Transformational shift in data management with technology and regulatory forces providing traction



Modernize clinical trials and investments for patient-centric decentralized trials

Manage and integrate more and disparate non-CRF data

Adoption of modern technology infrastructure and technology advancements like AI/ML



Significant pressure to conduct and accelerate research and deliver vaccines and treatments to patients faster

Emergency regulatory guidance from FDA and EMEA for remote approaches to monitoring, e-consent, decentralized or virtual trials, and shipping investigational products direct to patients

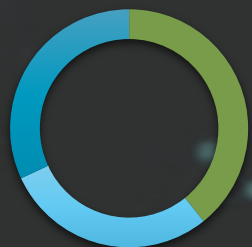


Industry convergence to accelerate and advance research, modernize clinical trials and deliver new and complex therapies

Digital and analytics-based modern data management is now an imperative to evolve and adapt in the new normal

Tufts-eClinical Solutions Data Strategies and Transformation Study

Summary of key findings and takeaways from the survey



Increased cycle time LPLV to DBL with >5 data sources – no data strategy



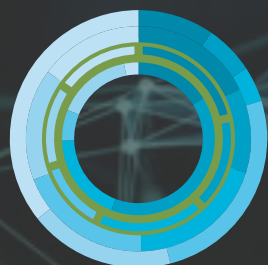
Labor-intensive or time-consuming to manage the entire data lifecycle



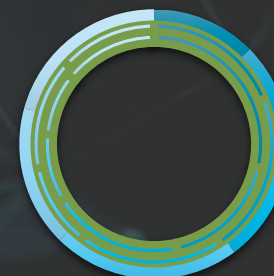
Reduction in cycle time LPLV to DBL with robust data strategy



Formal data strategy advances Data Science Technology Maturity

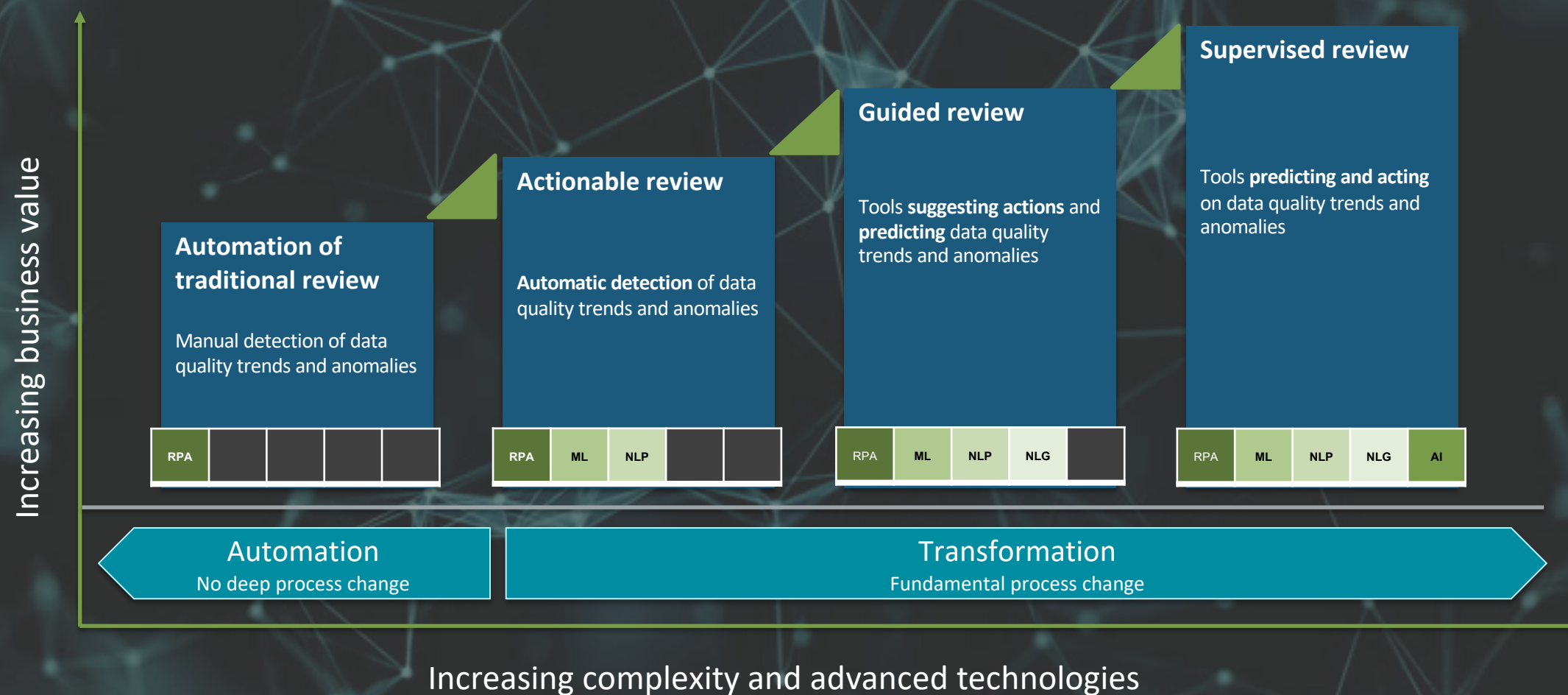


Dependent on SAS to integrate and organize data



Dependent on SAS and Excel to consume, review, and analyze data

Evolution of Data Review Powered by Data Science



Automation of Traditional Data Review Leveraging RPA

Automate manual, repetitive tasks to reduce time and associated costs



Operational Analytics Automation

Ability to automate the process for downloading streaming reports from EDC and uploading to clinical data platform to publish analytics and metrics for Data Management oversight.



Medical Coding Automation

Auto-generate files of verbatim terms and synonyms and load into Coder to reduce manual coding and increased coding consistencies.

Loosely coupled technology easy to configure by a business user

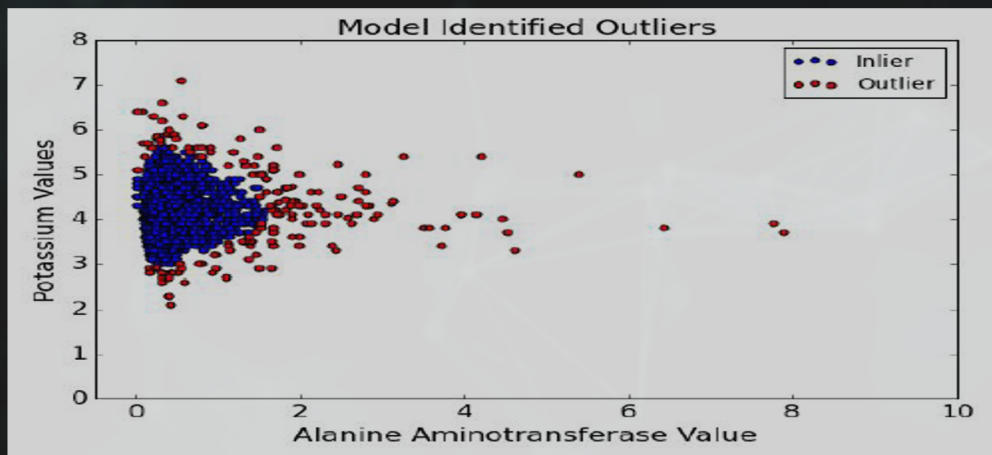
Actionable Data Review Leveraging AI & ML

Outlier detection model for lab and vital sign data

Identifying data quality issues within SDTM data

- Data mishandled at the site level
- Patient developing abnormal data trends

Outlier detection in lab tests simultaneously considers common lab tests to identify suspicious data and the ML model outputs labels “Inlier” and “Outlier”

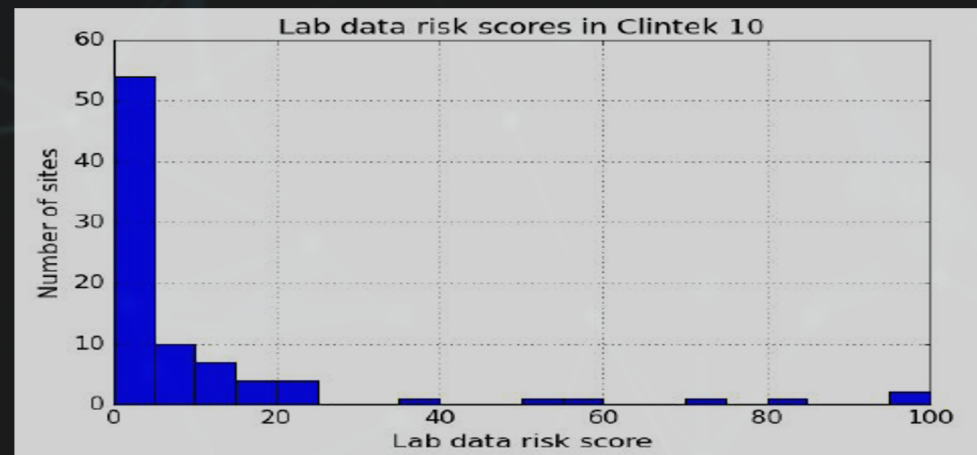


Risk scoring algorithm for lab data outliers

In order to assess data risk for insights into site performance, we calculate a Risk score: Ratio of “Outlier” labels and total number of data points aggregated at patient and site level

High data risk values suggest that a site is improperly managing data

Action: Highest risk outlier patients are assigned to a reviewer



Guided Data Review Leveraging AI & ML

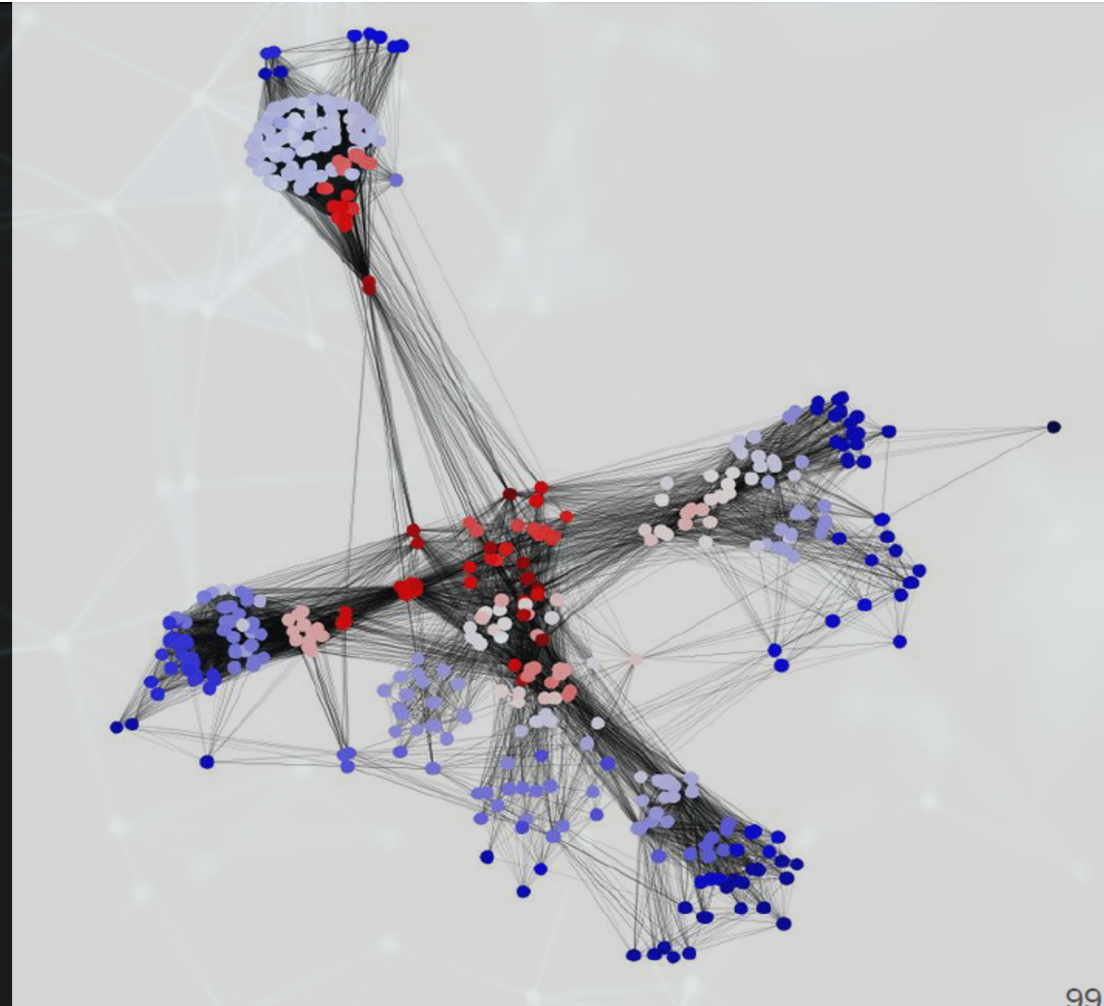
Data Similarity Networks

Assessment of data quality before review can be based on similarity to historical data that has been reviewed and approved or rejected

Using ML algorithms and advanced NLP techniques to construct data similarity networks where “data tables” are represented as graphical databases

- High quality data is highly connected in this network, outlier data is weakly connected
- Data credibility can be established across the data network using similarity scores

Action: For “rows” with lower scores or less connected data can be passed to a data reviewer



99

Scalable & Flexible Technologies for Deploying New Models

Ready to Use Tools
(ready to run, no need to develop anything)

Data Pipelines
(SQL, Spark)

Languages
(C, Python)

Packages
(TensorFlow)

Workspaces
(Jupyter)

Visualizations
(Bokeh, Shiny)

Advanced Environments
(develop advanced IP, manage your org needs)

Build & Package

- Container Repository
- Package Repository

Development

- CI / CD
- Version Control
- Test Driven Development

Deployment

- Micro-Services
- Large Compute (clusters, private clouds, TPUs, etc.)

Data Hub
(access Data)

Data Platform

Common Infrastructure
(access Compute)

Modern cloud-based Infrastructure and Platforms

Key Takeaways

1

Technology modernization

Invest in modern technology infrastructure to manage data proliferation, achieve analytics maturity and transform data review

2

Build AI roadmap

Build an AI and automation roadmap for data review use cases in conjunction with a robust data strategy; upskill with new technology stacks to gain experience

3

The opportunity is now

Advanced techniques can be easily applied to scale and support the evolution of data management to clinical data science

Q & A

