

Data Engineering Project, "Educating for the Future" PhUSE Working Group

Vijay Pasapula, Gilead Sciences, Inc., Foster City, CA; Vijay.Pasapula@gilead.com
Parag Shiralkar, Sumptuous Data Sciences, LLC, NJ; Parag.shiralkar@sumptuous-ds.com



June 9th-11th
Silver Spring Civic Center
Silver Spring, Maryland, USA

US 2019

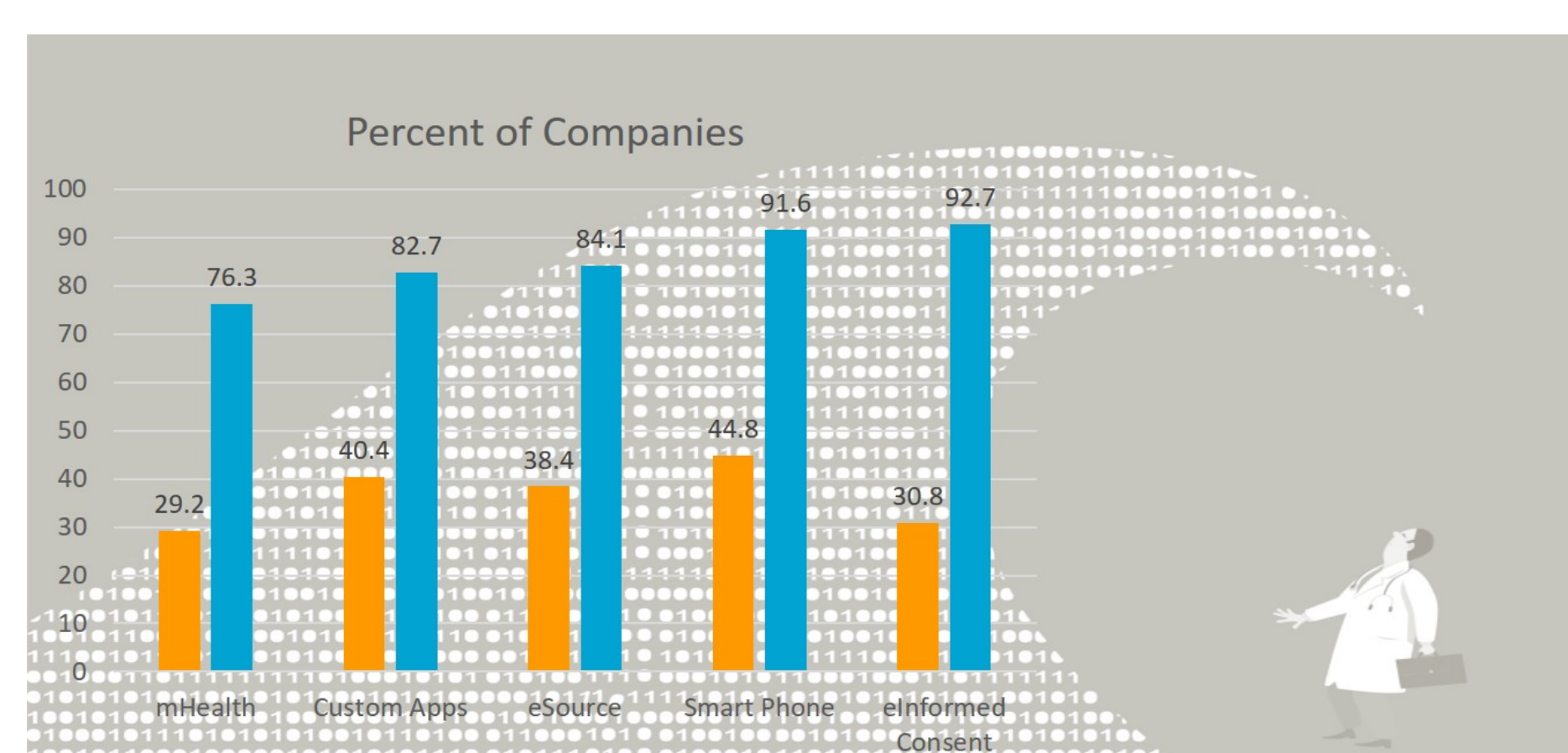
The Computational
Science Symposium



With an expected doubling the data collected over the next 3 years, of data from non-EDC sources (such as smartphones, wearables and custom apps) the traditional methods of managing data for clinical trials presents executives with a resourcing headache. As such, many companies are looking for lower cost strategies to mitigate this resourcing challenge. However, citing case studies from other industries, there are new methodologies/technologies in data engineering which could enable automation of much of the “heavy-lifting” currently practiced in clinical data management and statistical programming. This poster discusses the Data Engineering Project within the PhUSE Computational Science (CS) Working Group, Educating for the Future, focusing on the Vision/Mission/Goals of this project, Sub-topics in the project, such as Data Engineering Overview, Data Engineering vs. Data Science, Other Industries (e.g. Transportation, Agriculture) Use Cases, Automation (Software and tools), Culture, Benefits & Challenges, Education Methods etc., progress since project inception and future work.

Data Engineering encompasses the many considerations that need to be taken into account to optimally curate, transform, secure and disseminate data suitable for analysis. **Data Science** is the ability to turn data into a narrative. Data Engineer role is Identification, Automation, and Curation of data. Where as Data Scientist role is Analysis, Predications, and Serendipitous discovery

Data Sources - Current and Projected Use in 3 years



Source: Tufts – Veeva 2017 EClinical Landscape Study

Industry Standards



Clinical Data Interchange Standards Consortium

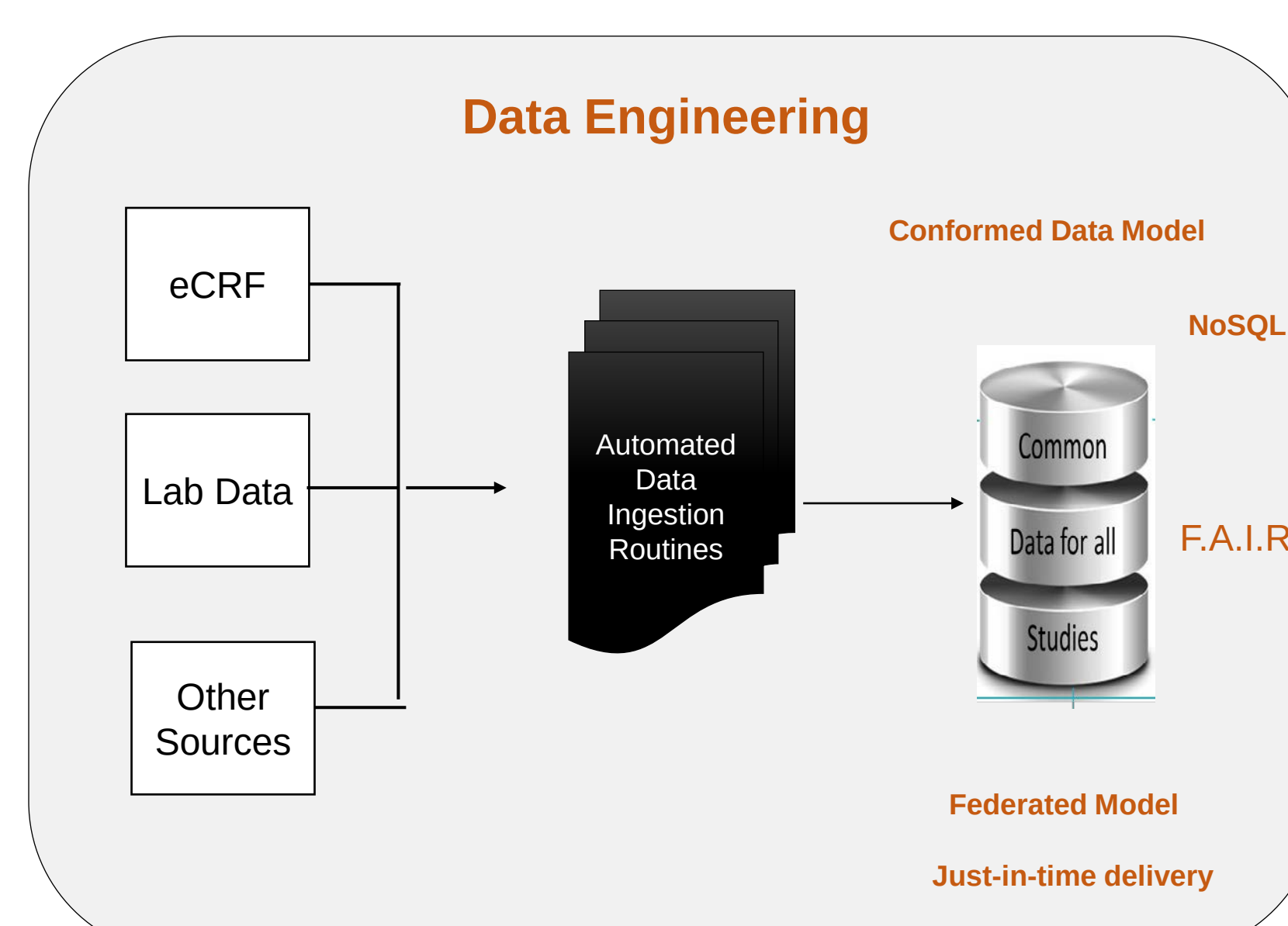


- Sets a core framework
- Advantage to the Pharmaceutical industry
- Data types becoming more diverse

Automation Software



Vision/Mission/Goal of the project: To explore how data engineering techniques, successfully deployed in other industries, could be utilised in the pharmaceutical industry, with a goal to facilitate the education of the pharmaceutical industry on these techniques.



* F.A.I.R. = Findability, Accessibility, Interoperability, Reusability

Culture



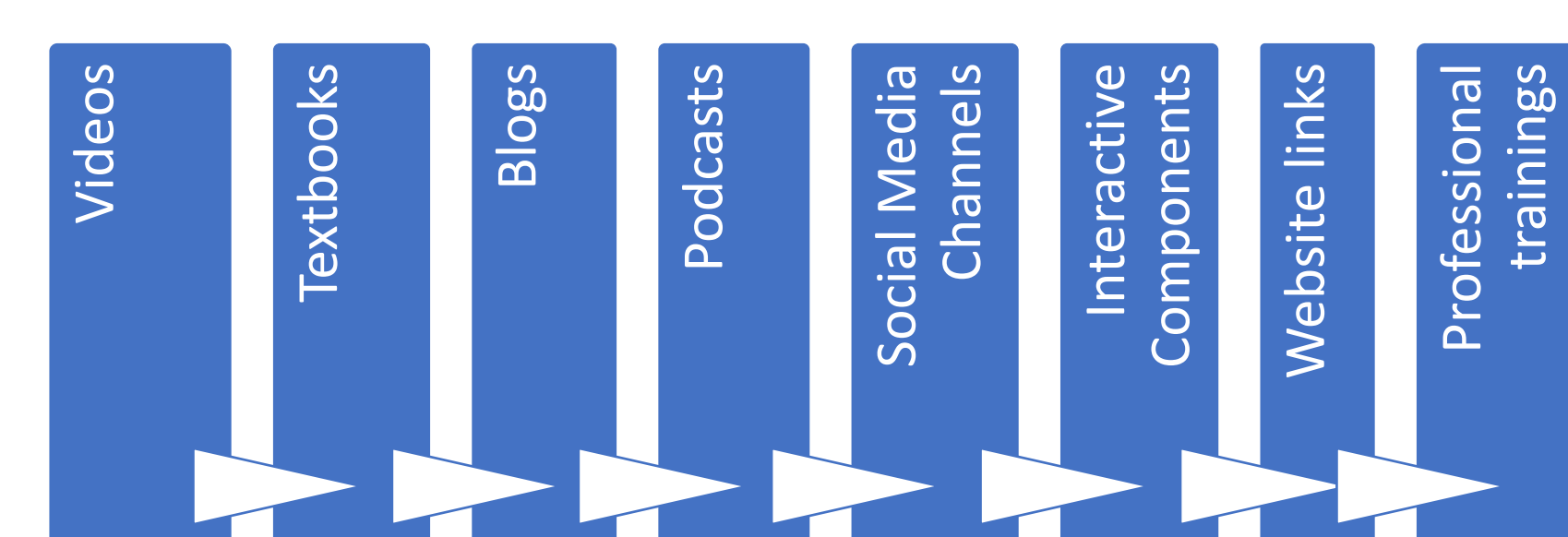
Other Industry Use Cases



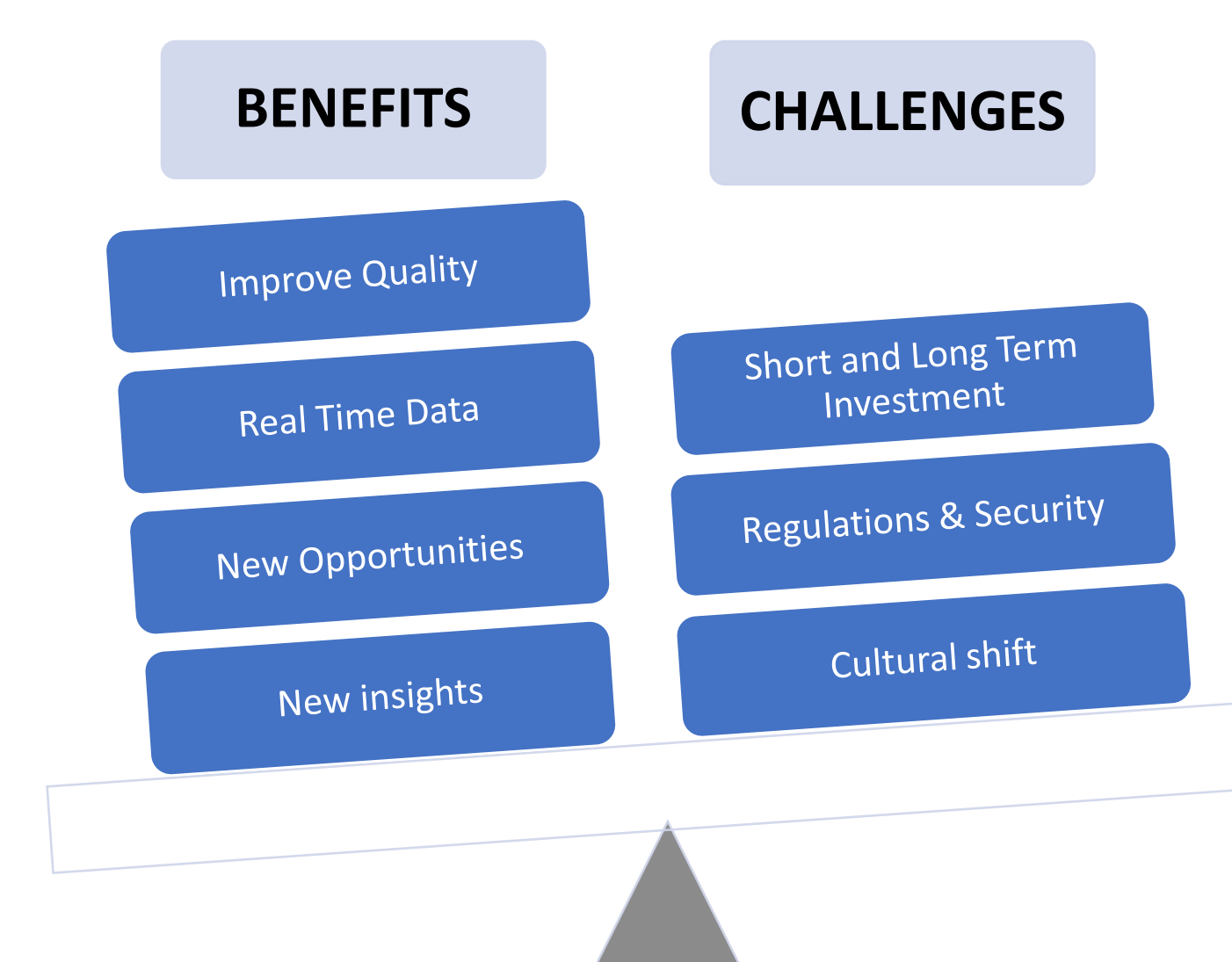
Agriculture: Weather, Growing season, Crop Yield prediction using tools HARVIST, AgrometShell, Crop Yield Predictor, Descrettes Labs

Sub-Topics: Industry Standards, Automation, Other Industry Use Cases, Culture, Benefits & Challenges, Education Methods.

Education Methods



Benefits & Challenges



Progress: Learned techniques, tools and software employed in Data Engineering space, Initiated SquareSpace page, Presented update in PhUSE EU Connect (2018) and PhUSE US Connect (2019)

What's Next: Further exploration of other industry use cases, Automation Software, Multiple Data Sources, Data Engineering Techniques, SquareSpace, build Learning pathways

References:

- “Data Engineering project- (Educating for the future PhUSE working group), paper SP05, PhUSE Connect 2019, by Guy Garrett and Beverly Hayes
- Tufts – Veeva 2017 EClinical Landscape Study. Tufts University, 2018, pp. 11–13, Tufts –Veeva 2017 EClinical Landscape Study.

Interested in joining the Data Engineering Project?

Contact:

BHayes2@its.jnj.com
Guy.Garrett@AchieveIntelligence.com
wendy@phuse.eu



<https://education.phuse.eu/eftf/data-engineering/>

Acknowledgements:

Project Leads: Guy Garrett, Bev Hayes, Wendy Dobson; **Data Engineering Project Members:** Allison Covucci, Amy Gillespie, Berber Snoeijer, Beate Hientzsch, Jatin Patel, Karnika Dalal, Mark Bynens, Mike Carniello, Mohit Juneja, Paul Slagle, Ralf Goetzelmann, Renu Shukla, Rohit Banga, Sagar Jain, Shaaz Ansari;
Educating for the Future Working Group Leads: Ian Fleming and James McDermott

Brand and product names are trademarks of their respective companies. No Copyright infringement intended.