ORACLE
Health Sciences

for the Analysis of Clinical Data

Greg Jones
Oracle Health Sciences

Jun 2018

ORACLE

Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

R for the Analysis of Clinical Data

R - Agenda

1. R: Open, Available & Extensible, Scalable
2. Clean, Transform, Aggregate Data for R Analysis
3. Use Case 1: CERN (the European Organization for Nuclear Research)
4. Use Case 2: NHS Business Services Authority (the UK National Health Service)
5. Use Case 3: Cross Study Analysis
6. Use Case 4: Healthcare Analysis
7. Use Case 5: R for Machine Learning Analysis
8. R: Regulatory Considerations
9. R In Commercial Applications
10. R: The Future

R for the Analysis of Clinical Data

R - Open, Available

The growth in the range of inter-connected devices across healthcare represents an exponential growth in the volume of data collected in ever more elaborate Clinical Trials

This growth in the volume of data presents new challenges for Clinical Data Scientists and requires new solutions and new tools for cross-study analysis

To meet these demands, Clinical Data Scientists are increasingly choosing open source solutions to leverage the active open source communities of experienced developers and statisticians

The R scripting language is increasingly popular and supports big data, predictive analytics, and offers the potential to leverage machine learning and artificial intelligence

R - Extensible, Scalable

R is used by a growing number of data analysts inside corporations and academia, whether being used to set ad prices, find new drugs more quickly or fine-tune financial models

It is also free. R is free for anyone to use and modify so statisticians, engineers and data scientists can improve the software's code or write variations for specific tasks

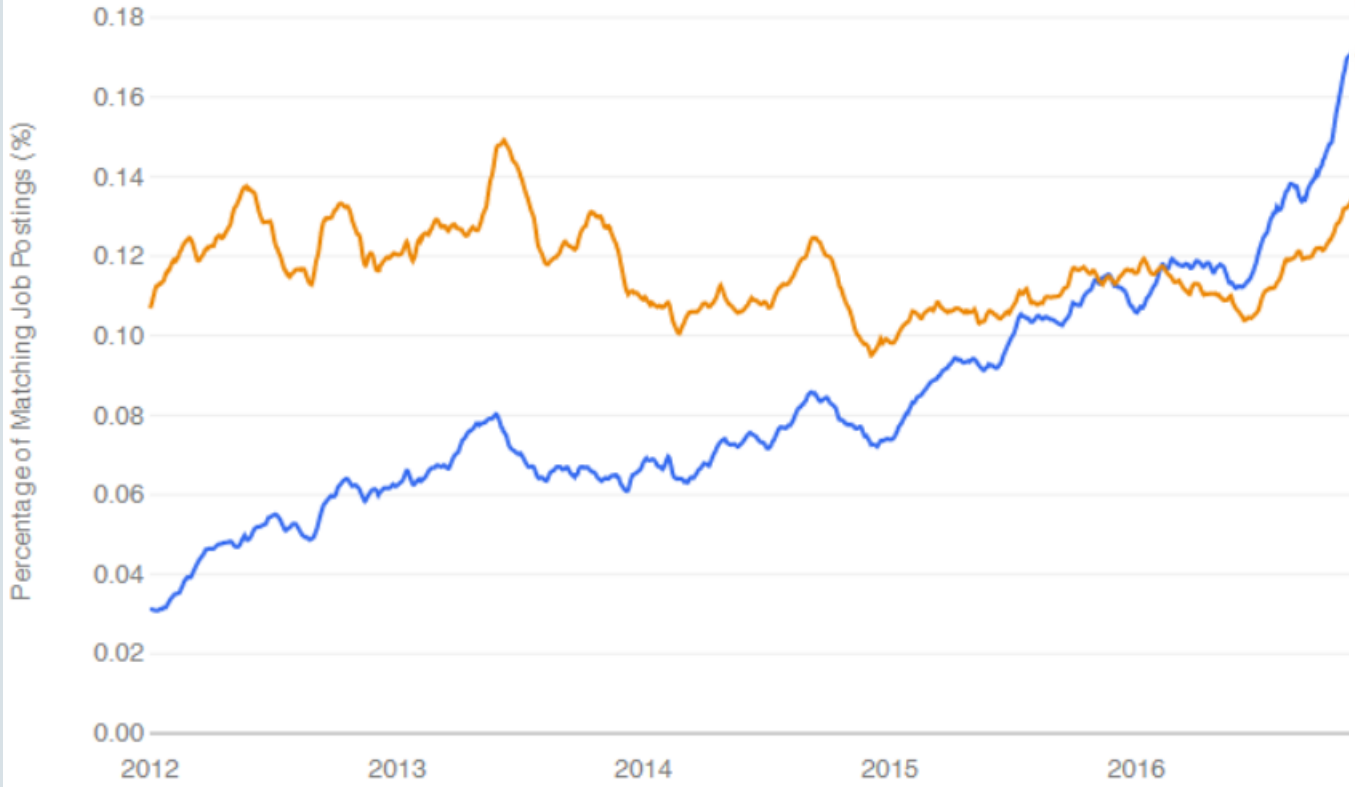
Packages written for R add advanced algorithms, richly coloured and textured graphs and mining techniques to dig deeper into databases, object stores, data lakes and big data sources

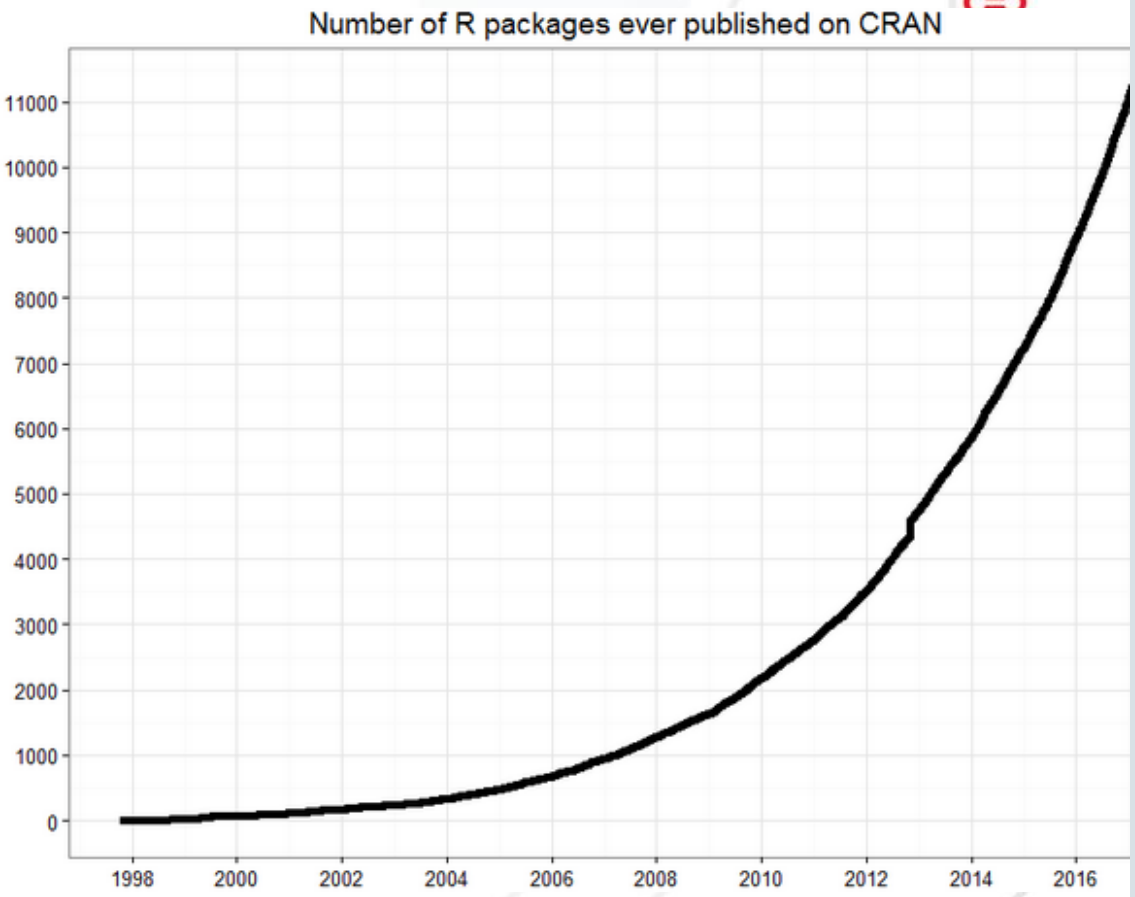
Pharma companies have created customized packages for R to let scientists manipulate their own data during nonclinical drug studies rather than send the information off to a statistician

R for the Analysis of Clinical Data

R is Gaining Popularity

Data Science job trends for R (blue) and SAS (orange)





Life Sciences Warehouse Cloud Platform

Structured Data

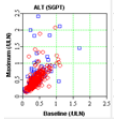
ePRO Data

CTMS Data

EDC Data

PK/PD Data

Safety Data



ECG Data

IxRS Data

EHR Data

Devices/ Wearables

Labs & LIMS Data

Unstructured Data

Continuous Streaming Device Data

Unstructured EHR Data

Scientific Publication Data

Let's start with the traditional Data Sources for a Clinical Trial.

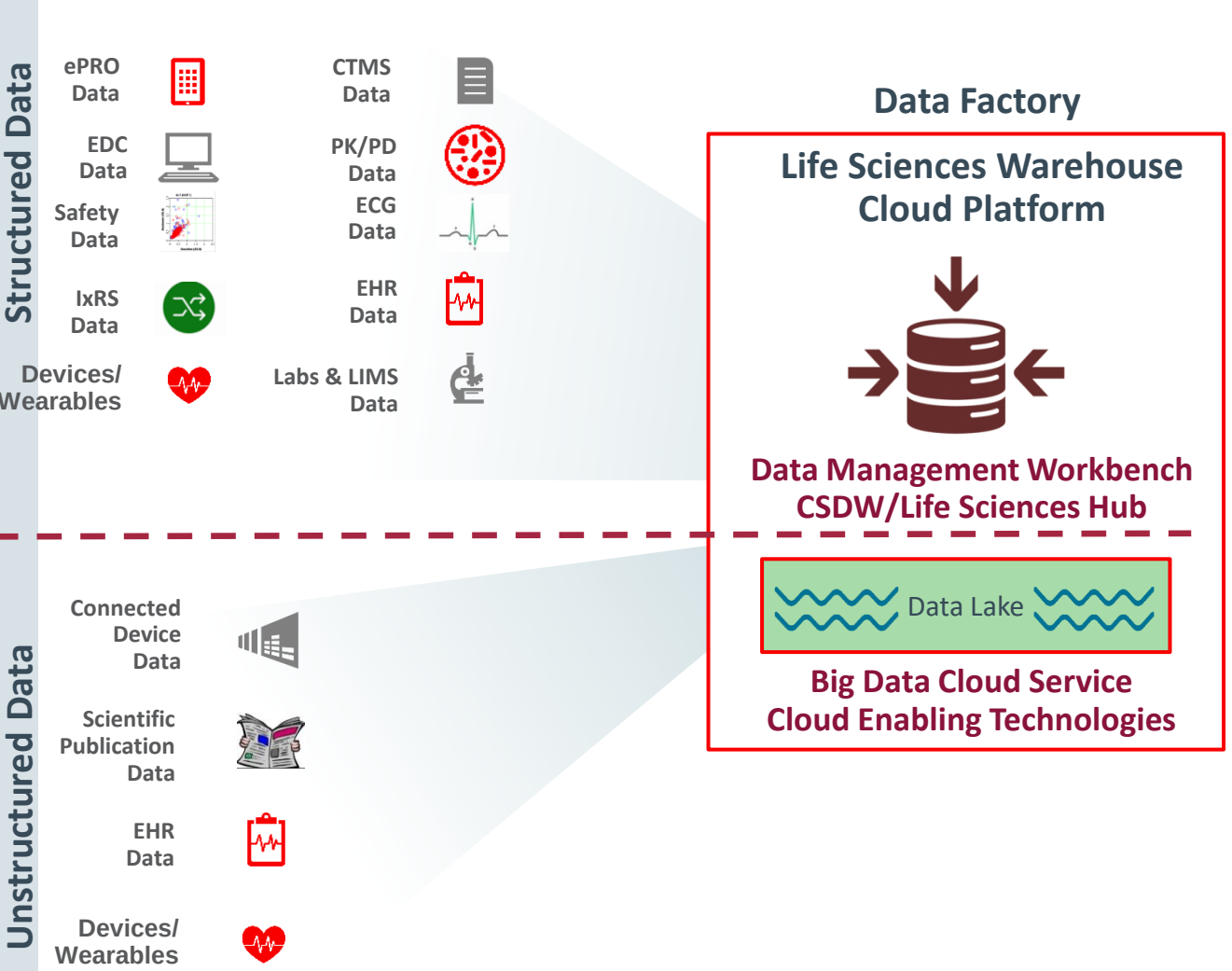
There are traditional Data Sources that are typically Structured in nature. These Data Sources represent data in tables and columns format.

There are also relatively newer, non-traditional Data Sources that can be Unstructured in nature. These Data Sources typically provide data in non table and column format. For example, large text files that come from Scientific Publications.

Some of these Data Sources – like EHR Data – are Real World Data Sources added to the more traditional Data Sources that come from running Clinical Trials. By combining these RWD Data Sources together we have the opportunity to accelerate Clinical Research and do it more efficiently.

Also, as you can see EHR and Device/Wearable Data Sources can be both Structured and Unstructured in nature. Therefore, flexibility is critical.

Life Sciences Warehouse Cloud Platform



Data Factory component of LSW Cloud Platform

- Highly optimized to drive low cost, efficient Submission Preparation Processes

Data Management Workbench

Works closely together with LSH to support the full lifecycle for Structured Data Source Management.

Enables automation of critical processes for highly productive loading, cleaning, transforming and aggregating Data to prepare for multiple types of downstream analysis by the Reviewer community

Several Oracle products in the Data Factory component

First is Life Sciences Hub. This product is the Structured Data Source Data Warehouse.

Big Data Cloud Service is the last major piece.

This component is optional and can be added to the Data Factory at any time when there's a need to support non-traditional Unstructured Data Sources such as Unstructured EHR Data.

Typical Unstructured Data Source Use Cases result in "processing" the Unstructured Data Source and driving a reduced or summarized result into DMW for inclusion in the Submission Preparaion Processes.

Combining the Big Data Cloud Service with DMW and LSH gives you a federated environment where we "use the right tools for the right job" to control and manage your Structured and Unstructured Data Sources together for rapidly cleaning, transforming, and aggregating all the Data Sources in preparing them for Analysis.

Life Sciences Warehouse Cloud Platform

Structured Data

- ePRO Data
- EDC Data
- Safety Data
- IxRS Data
- Devices/ Wearables
- CTMS Data
- PK/PD Data
- ECG Data
- EHR Data
- Labs & LIMS Data

Unstructured Data

- Connected Device Data
- Scientific Publication Data
- EHR Data
- Devices/ Wearables

Data Factory

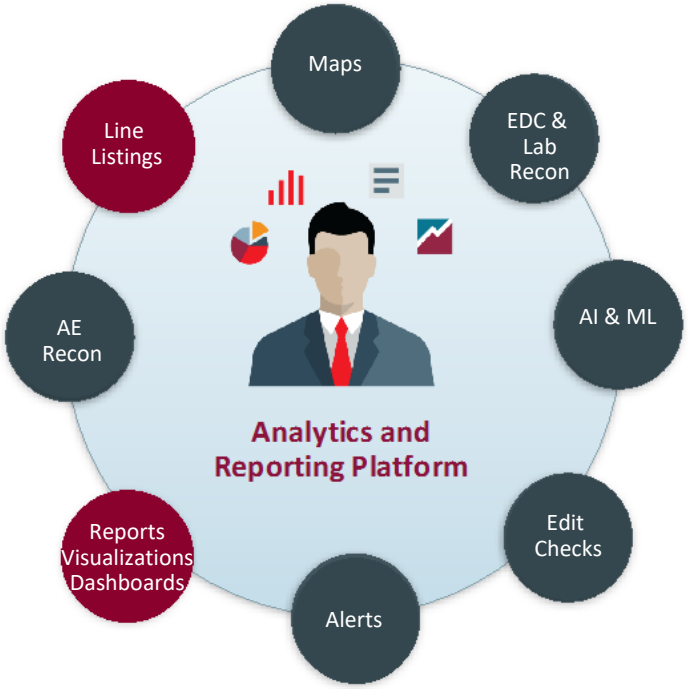
Life Sciences Warehouse Cloud Platform

Data Management Workbench
Life Sciences Hub

Data Lake

Big Data Cloud Service
Cloud Enabling Technologies

- SDTM Datasets
- REVIEW Models
- Visualization connector



Analytics and Reporting Platform

Perform the Data Review Activities across the Clinical Development Organization.

Support the various roles across the Organization including: Data Managers, Safety Analysts, Medical Reviewers, etc.

Use Oracle Analytics Cloud Service and LSW Cloud Platform to build and deliver the various Reports, Visualizations, and Dashboards

ORACLE

Life Sciences Warehouse Cloud Platform

Structured Data

- ePRO Data
- CTMS Data
- EDC Data
- PK/PD Data
- Safety Data
- ECG Data
- IxRS Data
- EHR Data
- Devices/ Wearables
- Labs & LIMS Data

Unstructured Data

- Connected Device Data
- Scientific Publication Data
- EHR Data
- Devices/ Wearables

Data Factory

Life Sciences Warehouse Cloud Platform

Data Management Workbench
Life Sciences Hub

Data Lake

Big Data Cloud Service
Cloud Enabling Technologies

- SDTM Datasets
- REVIEW Models
- Visualization connector

Discovery Output

Discovery Lab

AI Platform

Data Visualizer

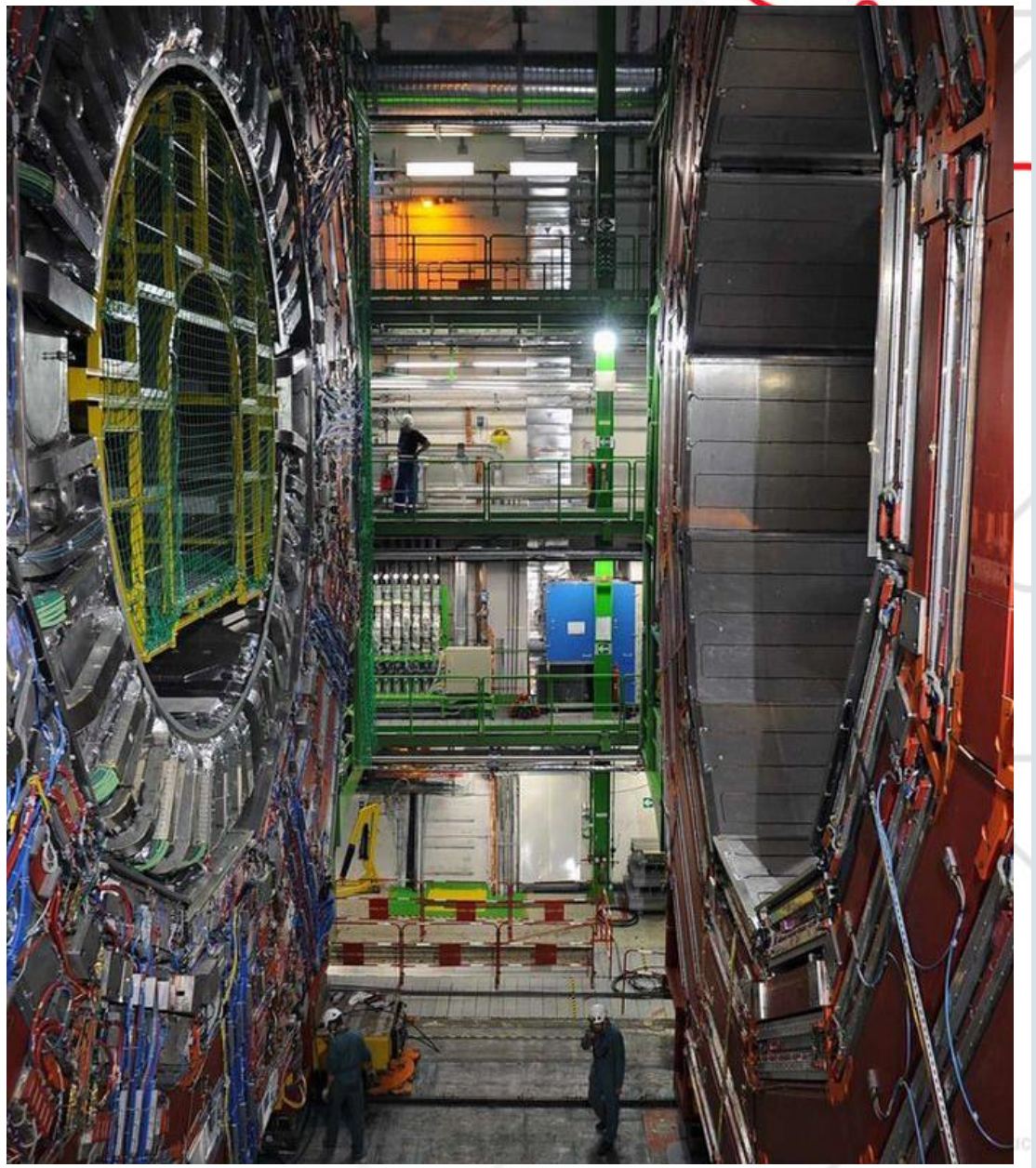
The Discovery Lab - generally a new capability in a modern Analytics Platform

Enables exploration of massive volumes of historical Clinical Trial & Real World Data Sources by Data Scientists using various tools for Visualization, Machine Learning, etc. to derive insights for future trial design.

R for the Analysis of Clinical Data

Use Case 1 - CERN

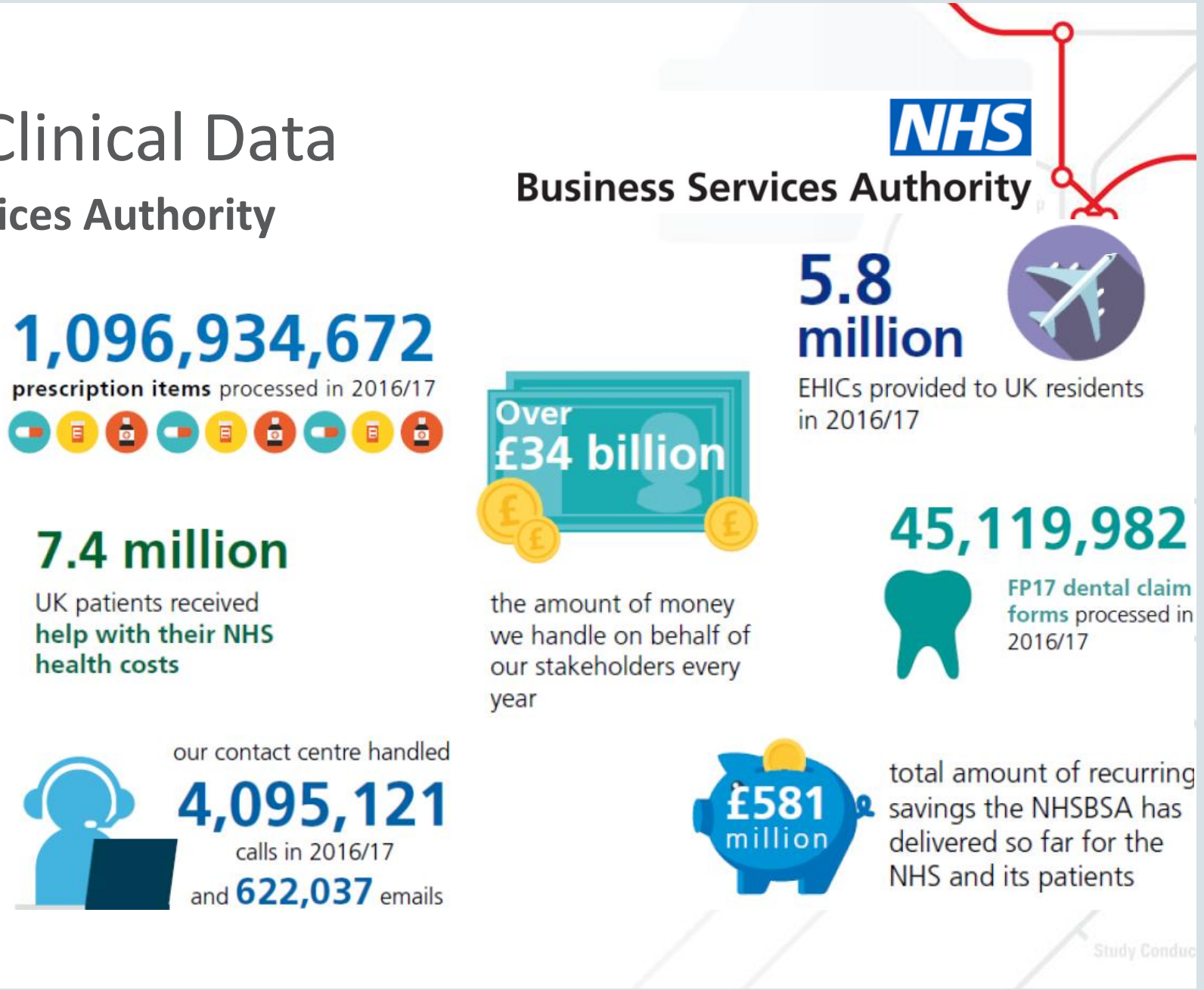
- Established in 1954, CERN (the European Organization for Nuclear Research) is the largest particle-physics laboratory in the world
- CERN uses big data, cloud computing, and analytics to help researchers unravel the mysteries of the universe
- The CERN team is building Machine Learning models to predict potential failures
- These models use R and run in the Oracle Database



R for the Analysis of Clinical Data

Use Case 2 - NHS Business Services Authority

- The NHS (UK National Health Service) is the largest and oldest single-payer healthcare system in the world
- The NHS Business Services Authority learned to make the most of its data thanks to analytics tools, and identified huge potential savings

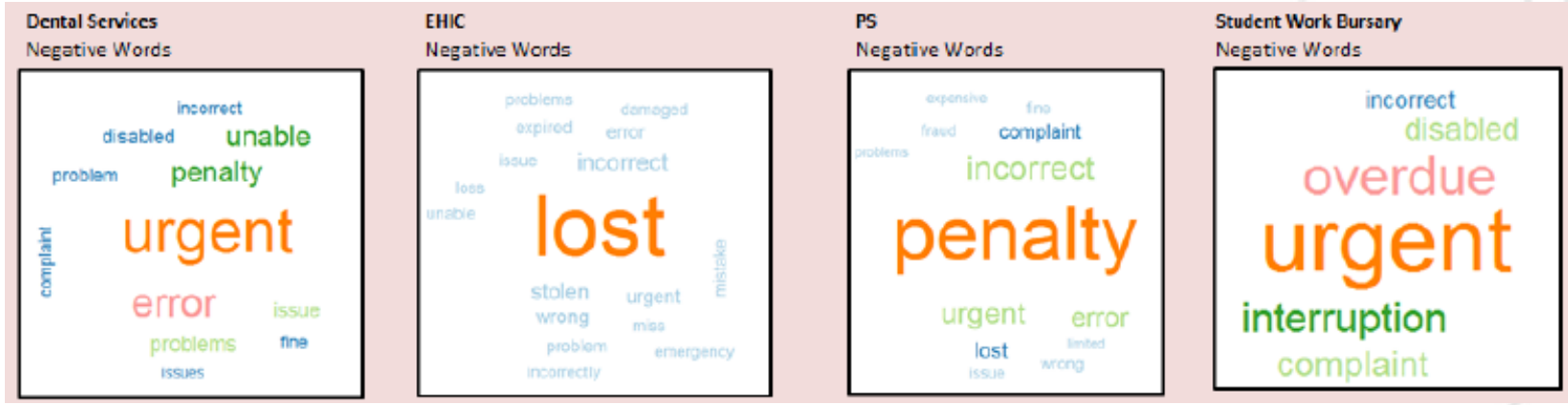


R for the Analysis of Clinical Data

Use Case 2 - NHS Business Services Authority

NHS Business Services Authority

- Data analysis using R:



- The NHS was able to identify potential savings of over GB£ 1 billion
- By providing accurate, reliable data back to clinicians and policy makers it has enabled antibiotic prescribing to be reduced by 7%.

"The overall solution is very fast, and our investment very quickly provided value. We can now do so much more with our data, resulting in significant savings for the NHS as a whole."

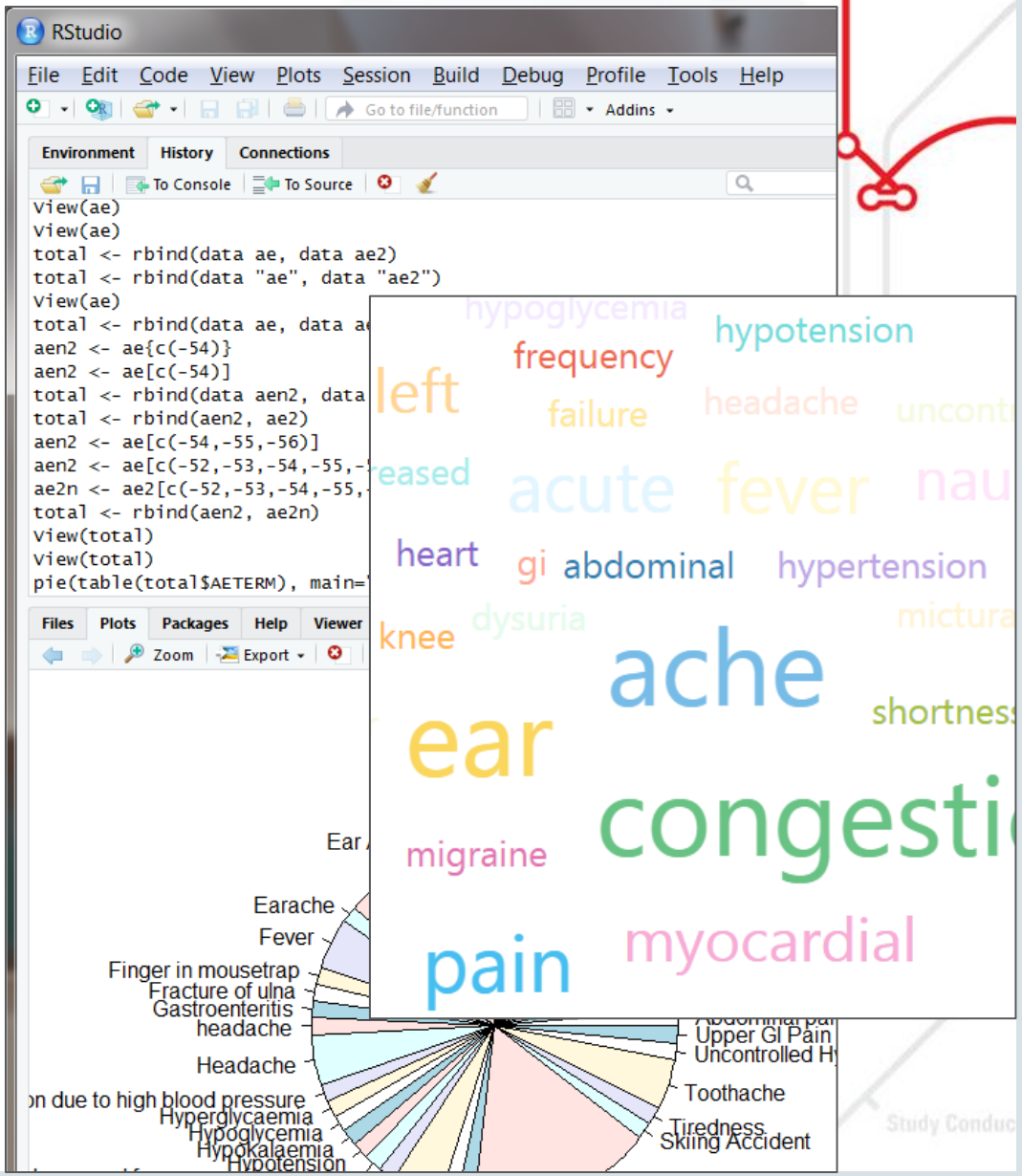
Nina Monckton, NHS Business Services Authority

ORACLE

R for the Analysis of Clinical Data

Use Case 3 - Cross Study Analysis

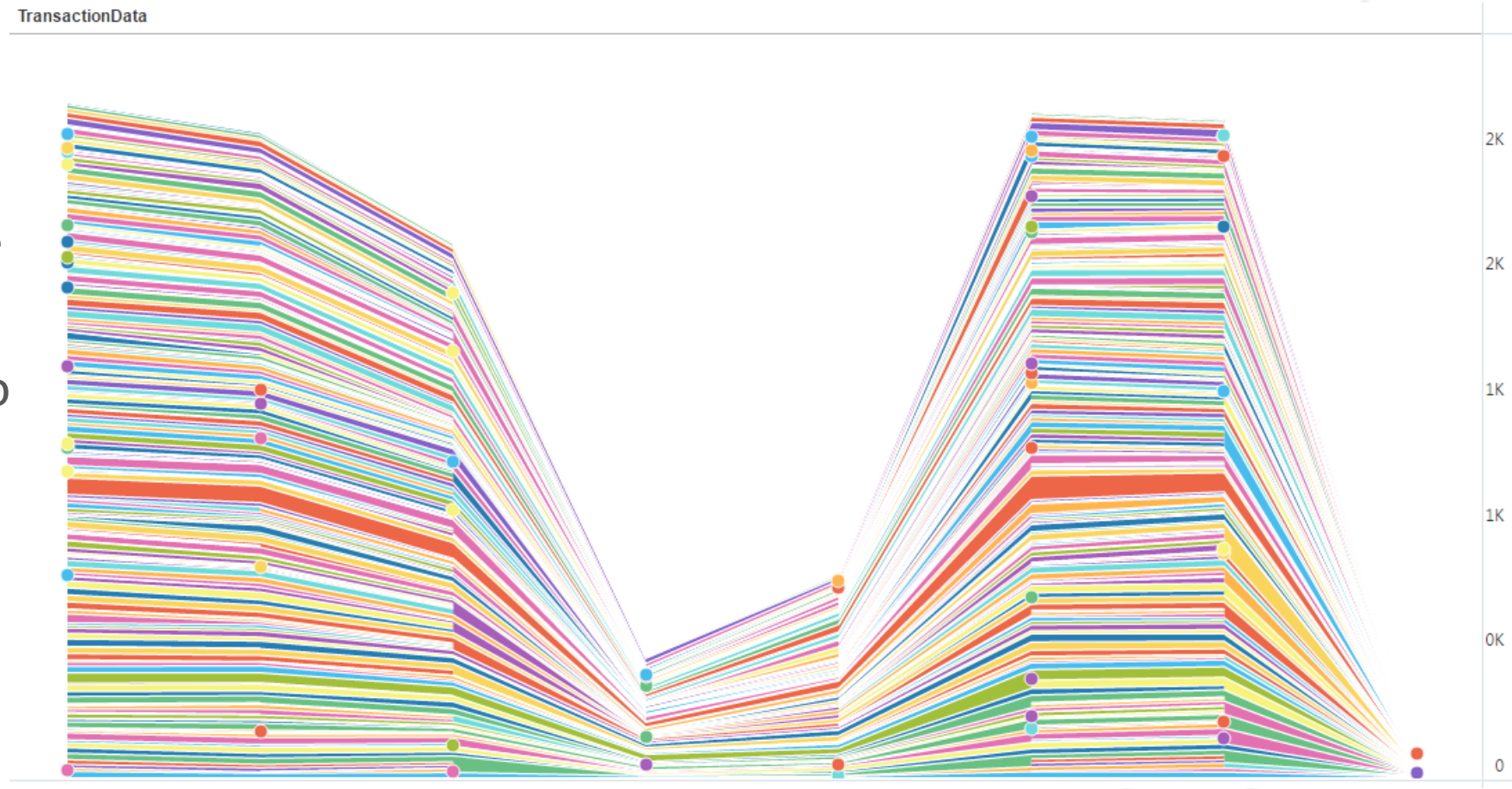
- The team used the Oracle R Distribution 3.1.1 and RStudio to prepare the analysis:
  1. Connect to standardized DMW data
  2. Combine data across multiple studies using rbind
  3. Train predictive analytics algorithms
  4. Create complex visualizations in R
  5. Apply term analysis
  6. Export to SAS V5 xpt



R for the Analysis of Clinical Data

Use Case 4 – Transaction Analysis

- Historical observation of study performance
- Time series algorithms to uncover trends
- Predictive analysis of system load



R for the Analysis of Clinical Data

Use Case 5 – Healthcare Predictions: Re-admission Rates, Heart Disease Likelihood

- Machine Learning can predict hospital re-admission rates
- Machine Learning can predict likelihood of heart disease

Model - Heart disease likelyhood Training - CART

General | Quality | Related

F1 Value Model Accuracy 47%

Actual Values \ Predicted Values	Absent	Less Likely	Likely	Others	Total
Absent	22	0	0	15	37 (62%)
Less Likely	4	0	0	4	8 (13%)
Likely	2	0	0	6	8 (13%)
Others	1	0	0	6	7 (12%)
Total	29 (48%)	0 (0%)	0 (0%)	31 (52%)	60 (100%)

PredictedValue, ID, Sex, Age, PredictionConfidencePercentage, Prediction

ID	Sex	Age	PredictedValue	PredictionConfidencePercentage
202	FEMALE	50.00	Absent	0.52
203	FEMALE	64.00	Absent	0.52
204	MALE	57.00	Absent	0.52
205	FEMALE	64.00	Absent	0.52
206	MALE	43.00	Absent	0.52
207	MALE	45.00	Absent	0.52
208	MALE	58.00	Present	0.28
209	MALE	50.00	Present	0.28
210	MALE	55.00	Absent	0.52
211	FEMALE	62.00	Present	0.28
212	FEMALE	37.00	Absent	0.52
213	MALE	38.00	Present	0.28
214	MALE	41.00	Present	0.28

https://www.youtube.com/watch?v=IichF5pBt_U

https://www.oracle.com/solutions/business-analytics/data-visualization/library.html

R for the Analysis of Clinical Data

Regulatory Considerations

- The FDA's Statistical Software Clarifying Statement declares that any suitable software can be used in a regulatory submission
  - XPT file format is an open standard, not restricted to SAS
  - XPT files can be read into R with the read.xport function, and data can be exported with the write.xport function in the SASxport package
  - RStudio, a popular editor for R, uses the Haven package to import SAS datasets
- The R Foundation also provides guidance on how R complies with other FDA regulations
  - Regulatory Compliance and Validation Issues - A Guidance Document for the Use of R in Regulated Clinical Trial Environments.

R for the Analysis of Clinical Data

R in Commercial Applications

- Oracle Analytics Cloud and Data Visualization Desktop use R for their Advanced Analytics and Machine Learning functions, allowing users to leverage existing R packages and upload their own to power their analyses
- Oracle R Distribution - Oracle's supported redistribution of open source R, provided as a free download from Oracle, enhanced with high performance linear algebra libraries
- ROracle - An open source R package, maintained by Oracle and enhanced to use the Oracle Call Interface (OCI) libraries to handle database connections - providing a high-performance interface to Oracle Database
- Oracle R Enterprise – ORE makes the open source R statistical programming language and environment ready for the enterprise with scalability, performance, and ease of deployment
- Oracle R Advanced Analytics for Hadoop - High performance native access to the Hadoop Distributed File System (HDFS) and MapReduce programming framework for R users

R for the Analysis of Clinical Data

The Future

- R use is clearly growing across many industries and it is seen as one of the key tools for today's Clinical Data Scientist
- R is embedded in many leading industry solutions
- R can power Machine Learning and Artificial Intelligence
- The availability of a commercial distribution of R can re-assure users in even highly regulated industries
- Confirmation from the FDA that it can be used to analyse clinical studies leaves no barriers to R adoption across the clinical trial lifecycle and beyond

Integrated Cloud

Applications & Platform Services


ORACLE®