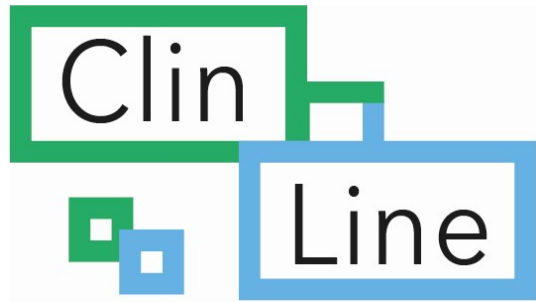




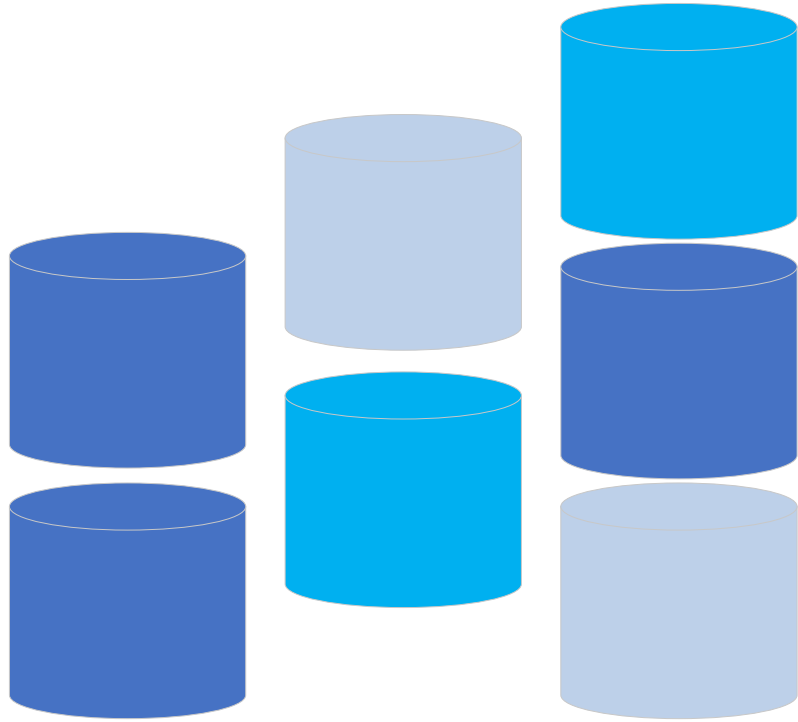
Processing Big Data from Multiple Sources

PhUSE data engineering project

Berber Snoeijer, ClinLine

13 November 2019

Processing Big Data from Multiple Sources



Content

- Introduction
- Classification of data sources
- Introducing FAIR
- FAIR in relation to ALCOA+
- Use case CHDR MORE[®]

Introduction



- Educating for the future
 - Project Data engineering
- Emerging trends and technologies
 - Real World Evidence

Classification of data sources



- Why is it important?
- What is the perspective?

Classification of data sources

Clinical Trial Domain

EDC CRF

Trial Lab data

Trial ePRO

Health Care Domain

EHR

Claims

Lab data

Registries

Pharmacy
data

PGHD Domain

Social Media

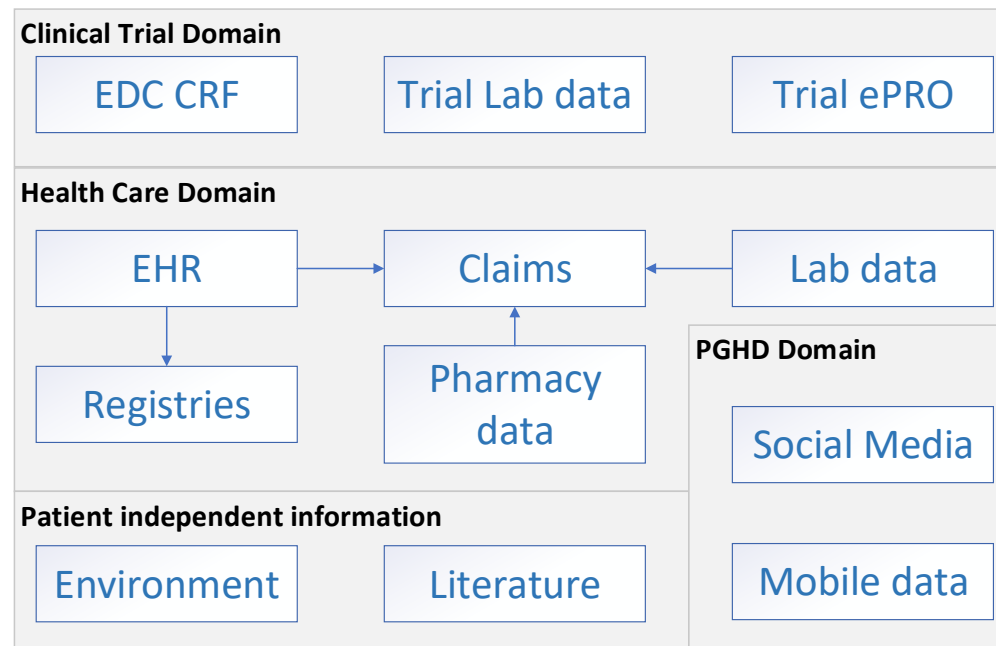
Patient independent information

Environment

Literature

Mobile data

Classification of data sources



FDA

- Use of electronic health record data in clinical investigations, July 2018
 - Data from Health Care Domain
 - Source data verification must be possible
 - Ownership
 - No Part II compliance required for source

Classification of data sources



FDA

- Non-interoperable data sources
- Interoperable data sources
- Fully integrated data sources

Introducing FAIR

2016: FAIR Guiding Principles for scientific data management and stewardship

- Findable
- Accesible
- Interoperable
- Reusable

Sci. Data 3:160018 doi:
10.1038/sdata.2016.18 (2016).



Introducing FAIR

Findable

- F1: (Meta)data are assigned a globally unique and persistent identifier
- F2: Data are described with rich metadata
- F3: Metadata clearly and explicitly include the identifier of the data they describe
- F4: (Meta)data are registered or indexed in a searchable resource



Introducing FAIR

Accessible

- A1: (Meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 The protocol is open, free, and universally implementable
 - A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2: Metadata are accessible, even when the data are no longer available



Introducing FAIR

Interoperable

- I1 (Meta)data use a formal, accessible, shared and broadly applicable language for knowledge representation
- I2 (Meta)data use vocabularies that follow FAIR principles
- I3 (Meta)data include qualified references to other (meta)data



Introducing FAIR

Reusable

- R1 (Meta)data are richly described with a plurality of accurate and relevant attributes
 - R1.1 (Meta)data are released with a clear and accessible data usage license
 - R1.2 (Meta)data are associated with detailed provenance
 - R1.3 (Meta)data meet domain-relevant community standards



FAIR versus ALCOA+



Introducing
FAIR

- Attributable
- Legible
- Contemporaneous
- Accurate
- Complete
- Consistent
- Enduring
- Available

Introducing
FAIR

FAIR in addition to ALCOA+

- Standards like CDISC
- Data transparency
- Use of Real World Data for clinical trials
- Linkage of data



CHDR
Centre for Human Drug Research

Use case:
CHDR
MORE[®]

Process

Use case
CHDR
MORE®

- Collection:
 - Withings Health Platform
 - Questionnaires
 - Phone usage
- RADAR-base
 - IMI project
 - Open source
 - Compliant to FAIR principles
 - Standardization
 - Privacy
 - visualizations



Process



Use case
CHDR
MORE®

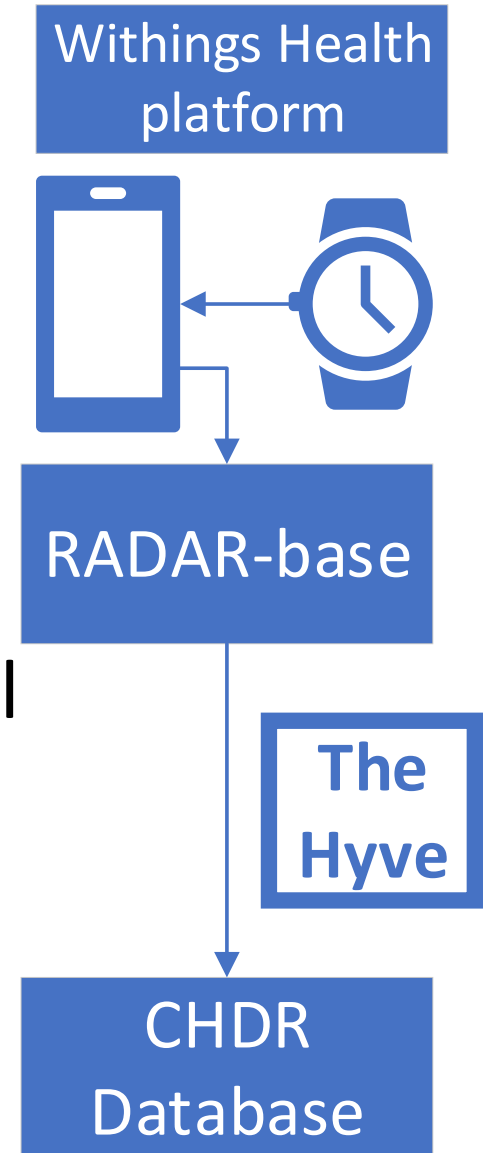
- Deployment and integration
 - Output requirements
 - Privacy requirements
 - Mapping
 - Data validation requirements
 - Data discontinuation detection
 - ...

Use case
CHDR
MORE®

Process

Smooth process of
data exchange:

- FAIR compliant
- Consented for
- Defined in study protocol
- Anonymized: QR codes

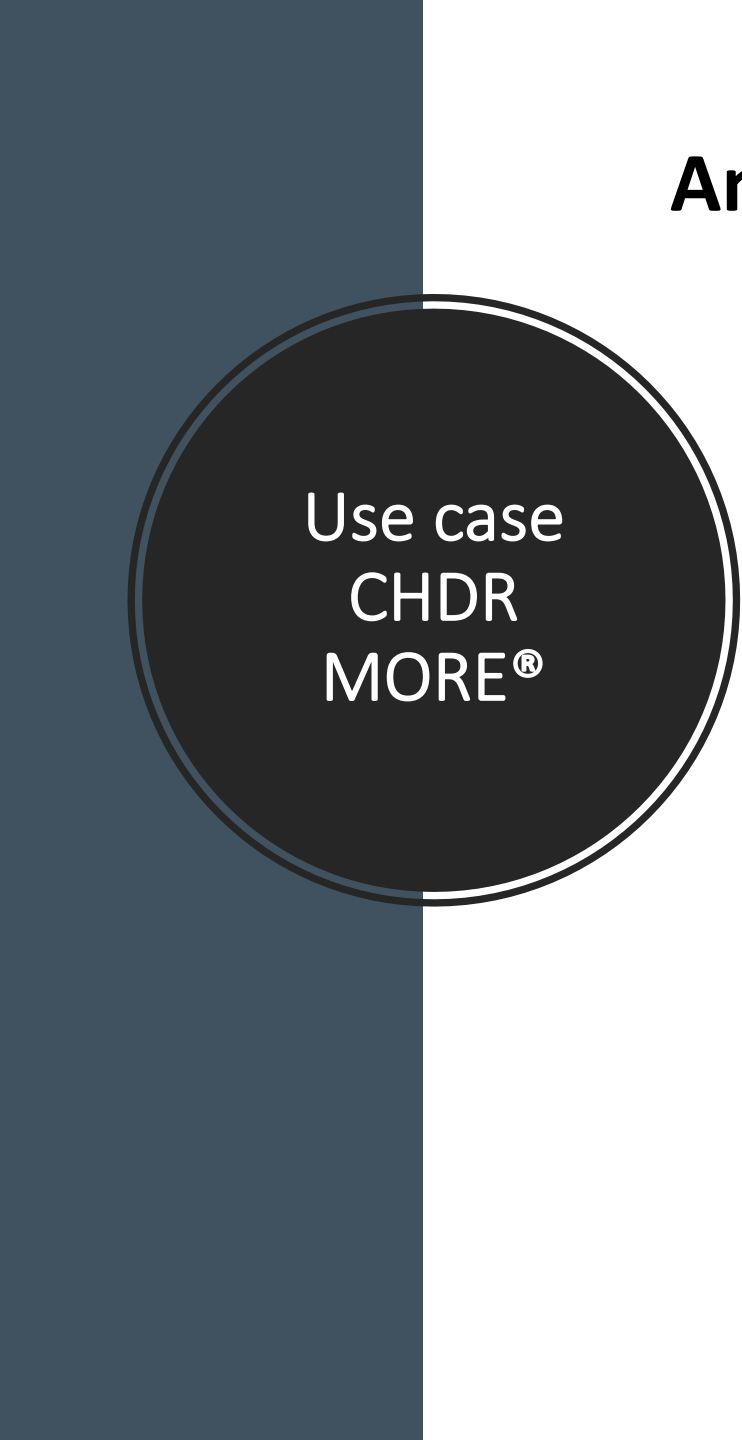


Analysis

Use case
CHDR
MORE®

- Continuous monitoring = BIG data
 - Movements, Blood pressures, ...
- Python / PySpark
- Statistical models and machine learning
- Studies
 - Pediatrics,
 - Dermatology
 - Neurology

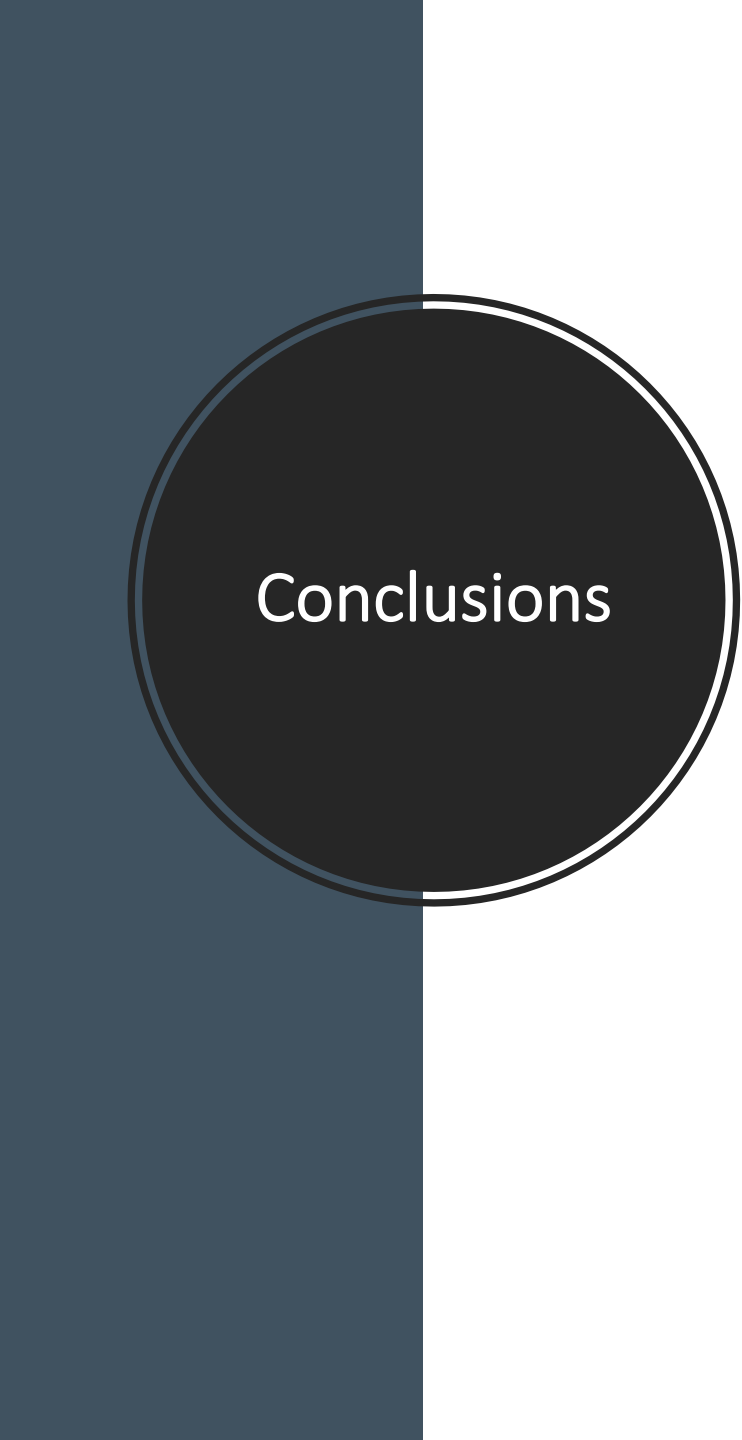




Use case
CHDR
MORE®

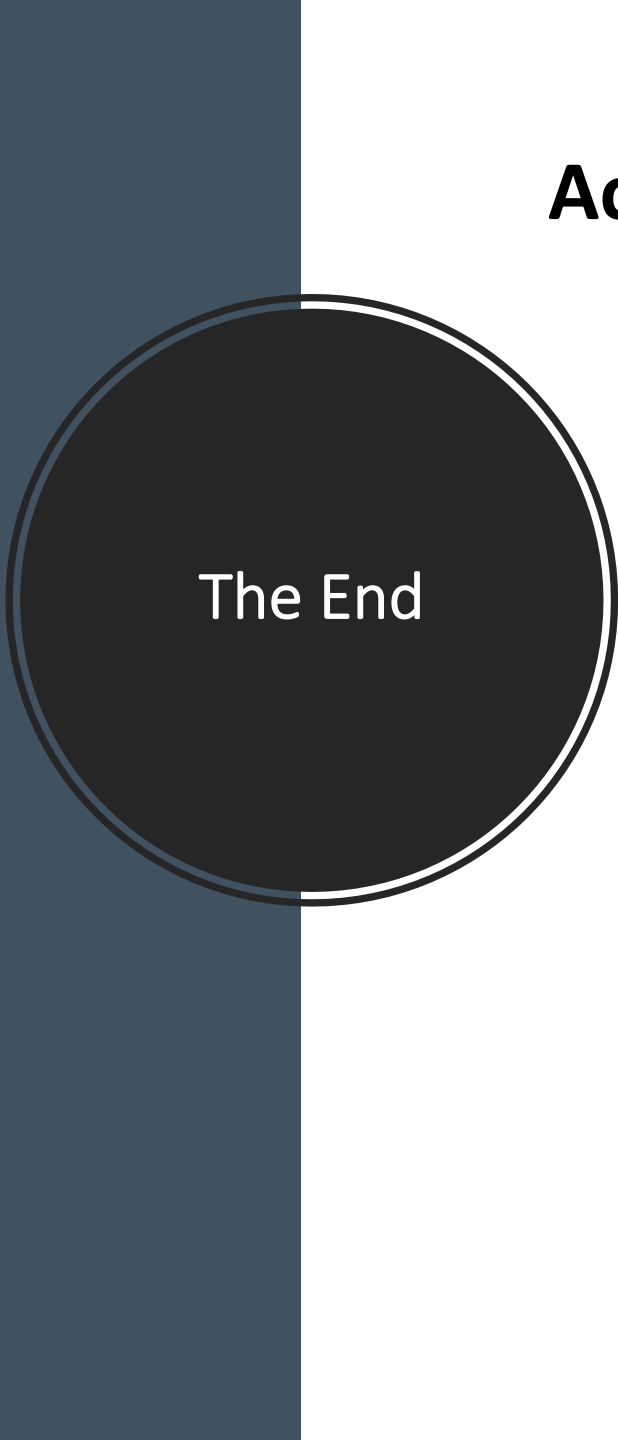
Analysis - Depression

- Difference between depressed and healthy subjects based on measured activity
- Defining a standard?
 - Validated against HAMD depression scale
 - Usage of Phones and wearables rapidly changing
 - Software updates out of control
 - retirement and emerge of new apps out of control
- Comparing activity of groups within same settings is however valid.



Conclusions

- **Classification of data sources** important for defining what guidelines apply and how data can be utilized in clinical studies.
- **FAIR principles** are an extension not a replacement of ALCOA+
- **FAIR principles** become more and more important with the introduction of Real World Data
- **Implementing new data sources** is a promising new step in clinical data development



The End

Acknowledgements

- Subgroup data sources:
 - Vijay Pasapula, Gilead Sciences
 - Allison Covucci, BMS
 - Susan Olson, EDJ Analytics
 - Beverly Hayes, J&J
- CHDR:
 - Ghobad Maleki
 - Ahnjili Zhuparris
- The Hyve:
 - Rosa Bianco Gallo
 - Julia Kurps

Contact:

Berber Snoeijer, ClinLine

b.Snoeijer@ClinLine.eu

+31 6 21197621