Analyses of Data Quality Frameworks

"Data Quality Framework for medicines regulation" workshop 7 April 2022

Presenter: Paolo Alcini, Data Analytics and Methods Task Force, EMA

Objective for Conducting the Analysis

EMA is looking to develop a Data Quality Framework to be used within the European regulatory network for medicine regulation, setting out the principles to assess and measure the data quality across multiple use cases applicable to the use of data within medicine regulation in European network.

Europe's healthcare sector is generating increasing amounts of raw data contained in electronic systems from ***primary, specialist and hospital care, drug and disease registries, clinical trials, genetic and genomic testing, manufacturing processes, spontaneous adverse event reports and mobile health systems***.

Methodology for the Analysis

Approach:

- The world wide web is saturated with articles, ideas or abstract approaches to data quality in almost all data spaces and industries. We needed a way to sieve out the ones that are applicable to EMA's need and addresses the multiple use cases mentioned.
- Our approach was to point out those data quality frameworks that have been tested and tried in practice and address the need for regulatory use and cover as many data spaces as possible.

Challenge:

- Some existing frameworks are mainly focused on standardisation of the terminology using an ontology and not on the quality of the data after the collection occurs.
- Most "frameworks" dealing with data quality are abstract concepts vs an actual usable framework
- Overlooked Frameworks... time constraint!

Criteria for the DQF (Data Quality Framework)

- **Formalised Data Quality Framework:**
  - There is sufficient documentation and supporting evidence that the framework has been applied against data and achieved the intended results.

- **Regulatory Compliance:**
  - The framework has proved to meet the need of regulatory bodies and their need for medicine regulation

- **Experience:**
  - There is evidence of other organisations and/or communities applying the data quality framework against their Real-World Data

- **Extensible Framework:**
  - The framework can be extended to include other data types or use cases than it was originally intended to be used for.

- **Product or Tool:**
  - There is some sort of technical solution provided with the framework that can be applied against the data to determine the "readiness" of the data for the intended purpose

Our Use Cases and Data Sources

- **Drug Research**
  - Omics
  - Bioanalytical
  - Microbiome
  - Evidence Based Veterinary Medicine
- **Veterinary**
  - EHR
  - Observational
  - Genotyping
  - Disease Outbreaks
  - Innovative Clinical Endpoints
- **Drug Development (pre-clinical)**
  - Omics
  - Bioanalytical
- **Drug Development (clinical)**
  - Clinical Trial Data
- **Manufacturing**
  - Batch Execution
  - Purity
- **Drug Safety Surveillance**
  - Spontaneous ADR
  - Observational
  - EHR
  - Claims

Our Use Cases and Data Sources cont...

- **Epidemiology**
  - EHR
  - Observational
  - Claims
  - Social Media
  - Mobile Health
- **Health Economics and Outcomes**
  - EHR
  - Observational
  - Claims
  - Social Media
  - Mobile health
- **Drug and Healthcare Utilisation**
  - EHR
  - Claims

Our Discoveries – Primary Use Data

- **Drug Research**
  - ISO 15189
  - GLP
  - GMP
- **Drug Development (pre-clinical)**
  - CDISC (SEND)
  - GCLP
- **Drug Development (clinical)**
  - ISO 14155
  - Euro Phase II of Policy 0070
  - CDISC (SEND)
- **Drug/Device Manufacturing**
  - ANSI/ISA-88/95
  - IDMP
  - 21 CFR part 210/211/212
  - CGMP
  - Sentinel
  - Advarra
  - AHIMA
  - NIH

Our Discoveries – Primary/Secondary Use Data

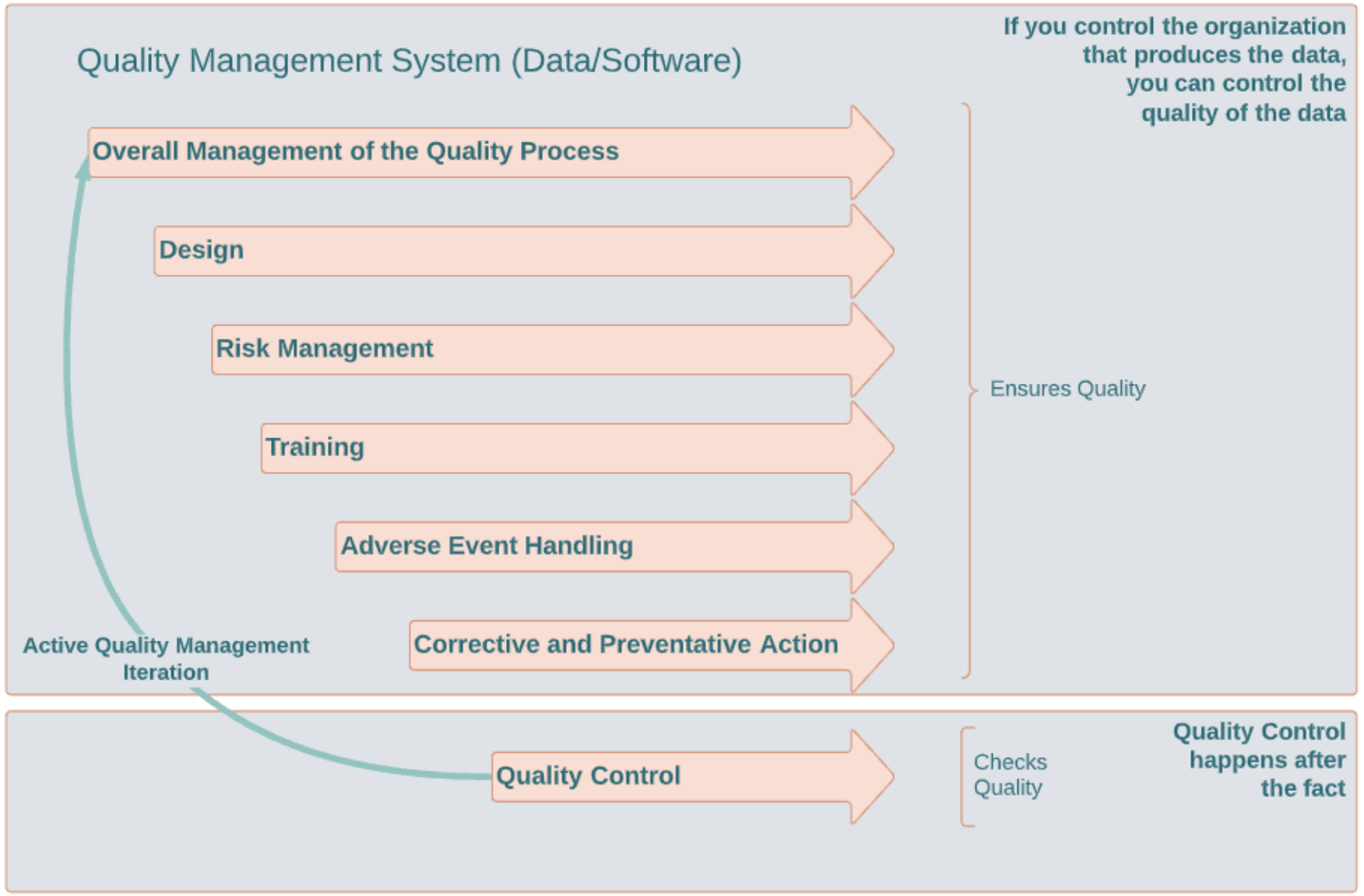
- **Drug Safety Surveillance**
  - OMOP
  - SNOMED
  - RxNorm
  - ISO/IEC 25012
  - ISO 8000
  - ISO 900x
  - Khan's
  - BMC
  - NESTcc
  - Sentinel
- **Epidemiology & Health Economics/Outcomes**
  - OMOP
  - SNOMED
  - RxNorm
  - ISO/IEC 25012
  - ISO 8000
  - ISO 900x
  - Khan's
  - BMC
  - NESTcc
  - Sentinel
  - TEHDAS
  - GS1
  - JAMIA
  - PCORnet
  - Duke-Margolis

Our Discoveries – Primary/Secondary Use Data cont…

- **Drug & Health Utilisation**
  - NDC
  - CPT
  - ICD
  - Khan's
  - TEHDAS
  - BMC
  - NESTcc
  - Sentinel
  - PCORnet
- **Veterinary**
  - SNOMED CT
  - Empres-i
  - CDISC (SEND)
  - UPD
  - TRACES
  - EudraVigilance
  - Boland-Casal-Kraus Framework
  - CORE
  - VanderWaal-Morrison-Neuhauser-Vilata-Perez Framework

Existing Regulations vs Need for Control

- We understand that there are a plethora of regulatory guidelines and control around the collection, quality and reporting of Primary Use Data or "Controlled Data". There was no reason to "recreate the wheel" so to say.
- The in-depth analysis of data quality frameworks was focused on Secondary Use Data or "Uncontrolled Data". Data that the regulatory body has no control over but uses to make regulatory decisions.
  - **Controlled Data:** These are produced in a process that controls the entire life cycle, including feedback loops to detect and correct errors and prevent their future occurrence. In most cases, these data are collected for the purpose of supporting regulatory decisions and therefore are governed by existing regulations and guidelines from international regulatory bodies.
  - **Uncontrolled Data:** Data that are collected for a different purpose, but then exported and reused for purposes of regulatory decision making (secondary use). The production process may or may not be regulated regarding its primary use, but even if it does, the quality of data generated in such process and the “fitness for purpose” of secondary use must be considered uncontrolled.

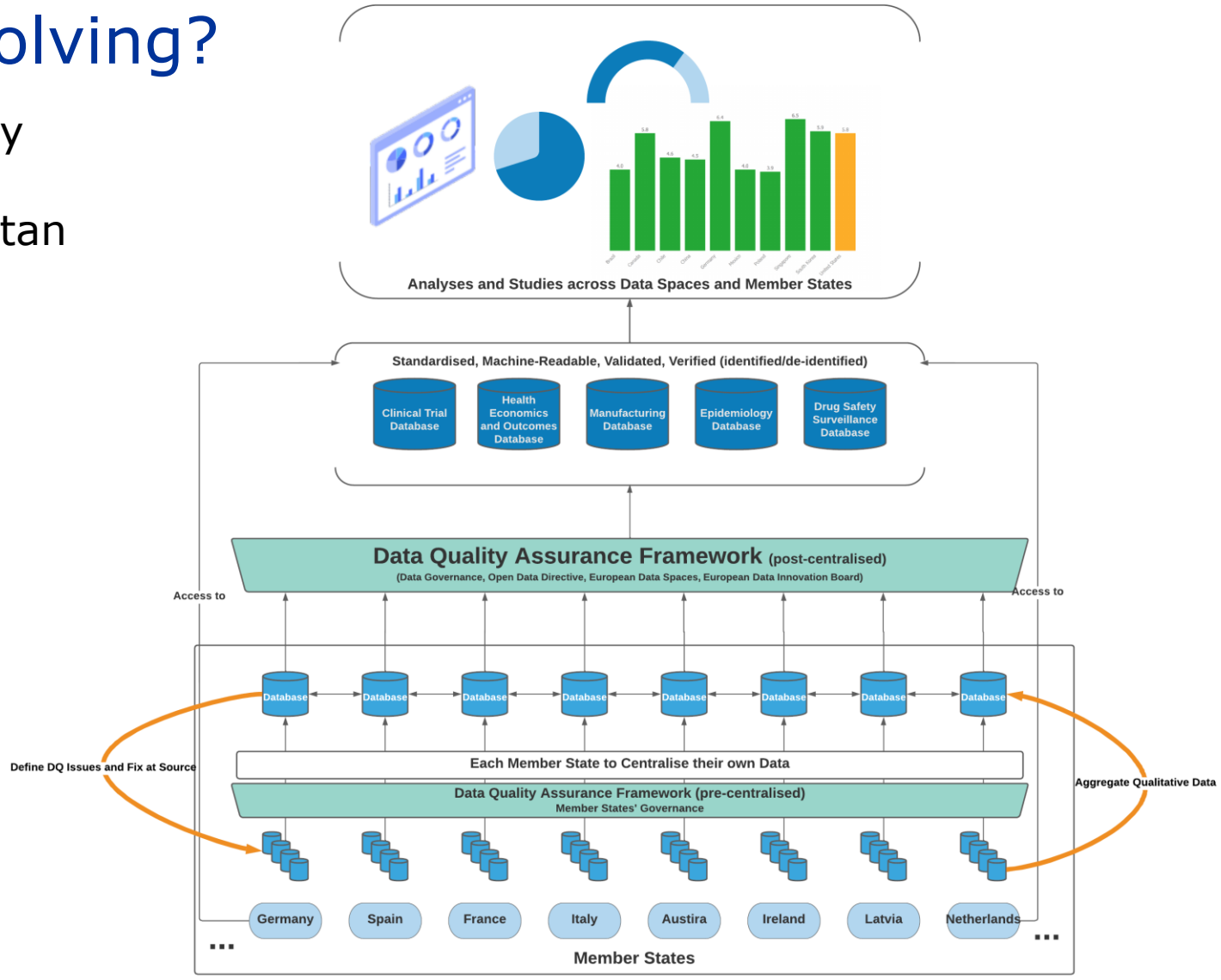


The data governed by the QMS (Quality Management System) is already **controlled** by regulatory bodies through extensive guidelines and standards.

The **uncontrolled** data that is received "*after the fact*" still needs to be standardised, checked and prepared for use.

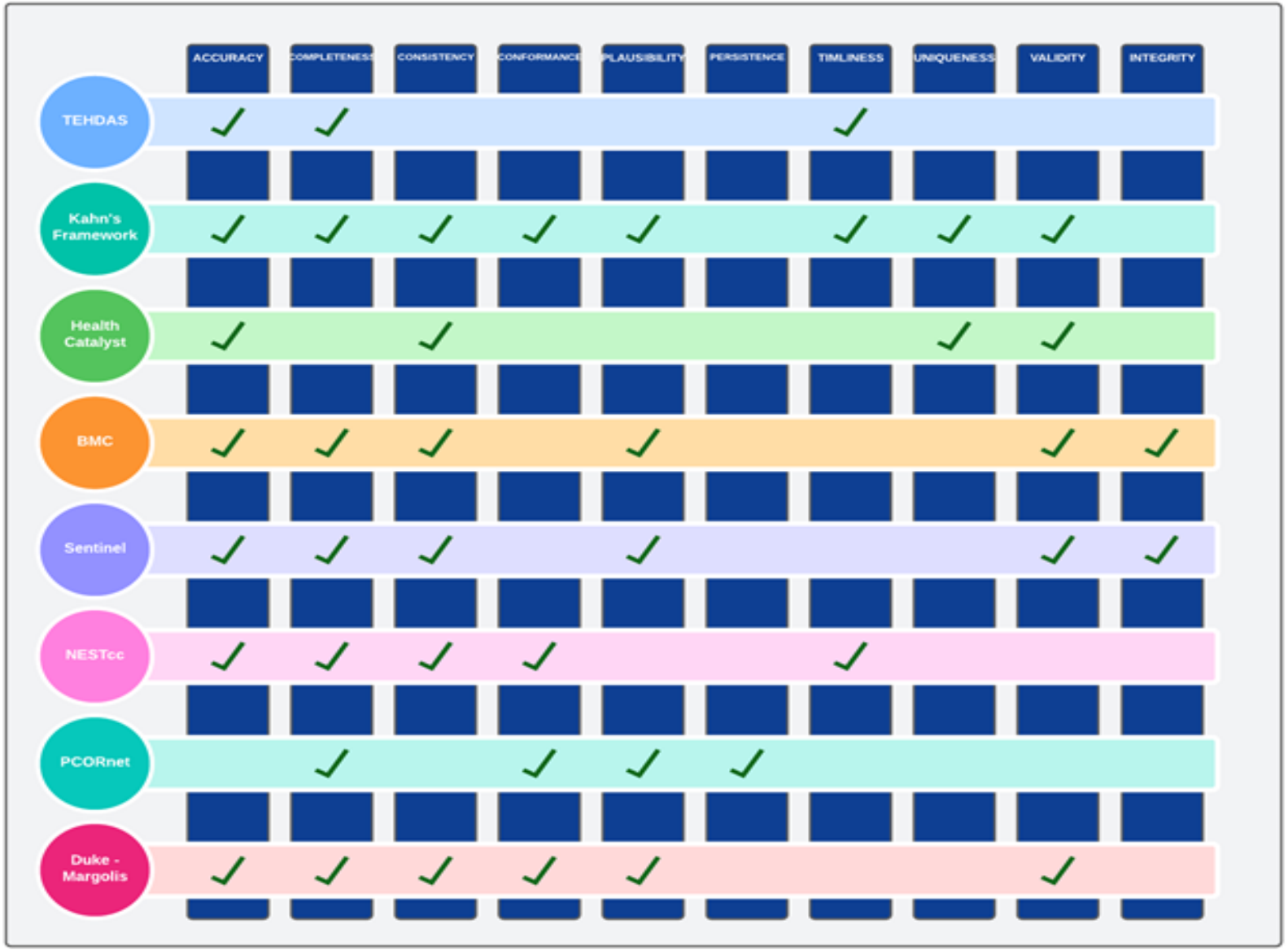
What are we solving?

We need a Data Quality Framework for Non-Standardised and/or Stan dardised Data



Secondary Use Data Quality Frameworks Analysed for Data Quality Dimensions

Each DQF concentrates on certain data quality dimensions. The dimensions are used to determine if the analysed data is fit-for-regulatory purpose for a particular use case. The more dimensions the framework covers, the more extensible the framework is for other data types.



Thank you!

For any question on this presentation, please contact: dataqualityframework@ema.europa.eu
Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000