



health **RI**

enabling data driven health

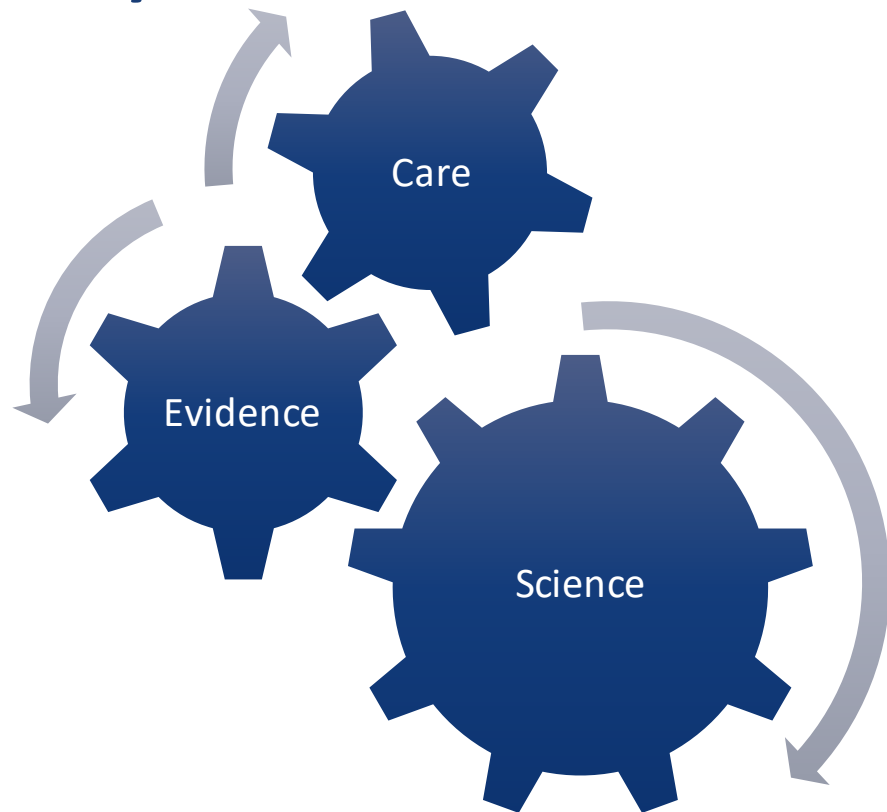
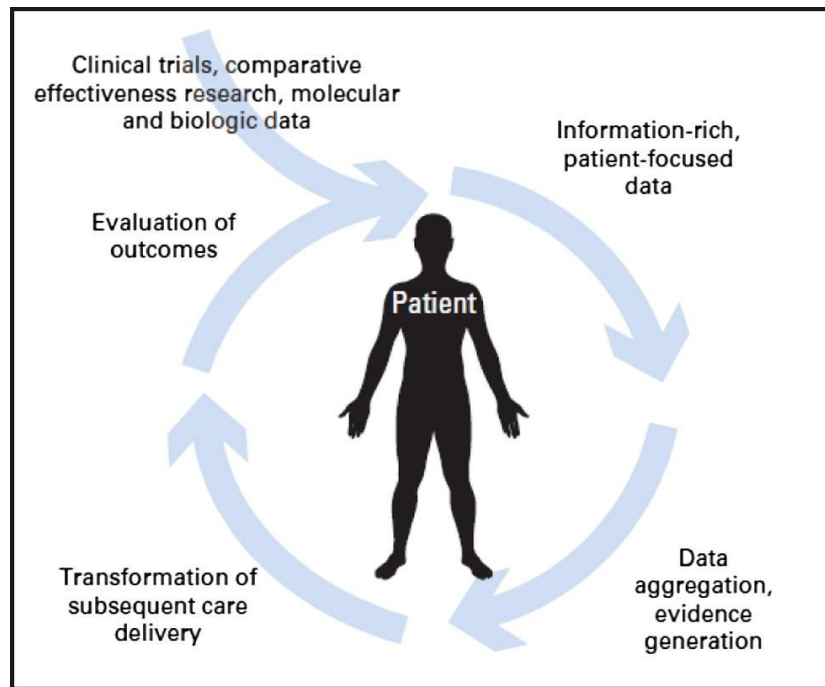
Data standardization – researcher perspective

EMA workshop; 18 May 2021

Jan-Willem Boiten (janwillem.boiten@lygature.org)



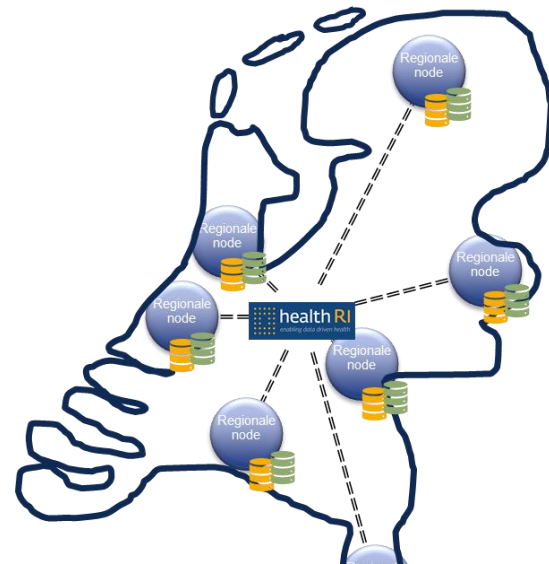
The ambition: a learning health system



Health-RI national data infrastructure: the hub & nodes model

Key principles

- Network of existing collaborations
- Reuse what's already there: registries, biobanks, cohorts, etc
- Leave data at the source
- Data protection & governance primary design consideration
- Resource for AI & machine learning
- Focus on data quality



← News

Dutch Government has pledged 69 million euros investment to Health-RI

 News | 9 April 2021

Health-RI data infrastructure: data quality is critical

Key principles

- Network of existing collaborations
- Reuse what's already there: registries, biobanks, cohorts, etc
- Leave data at the source
- Data protection & governance primary consideration
- Resource for AI & machine learning
- Focus on data quality



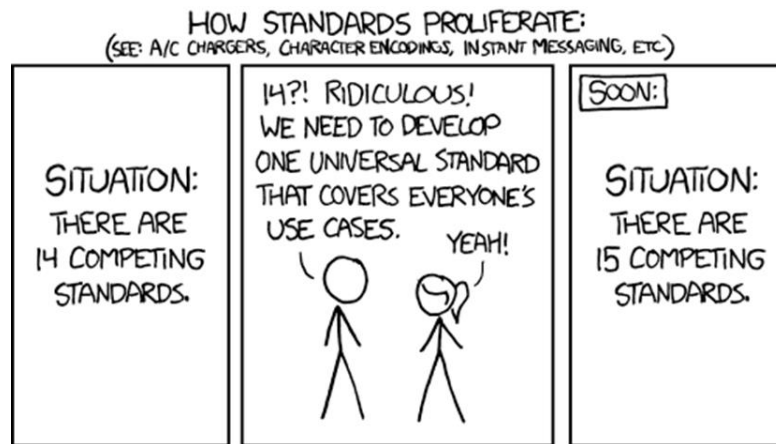
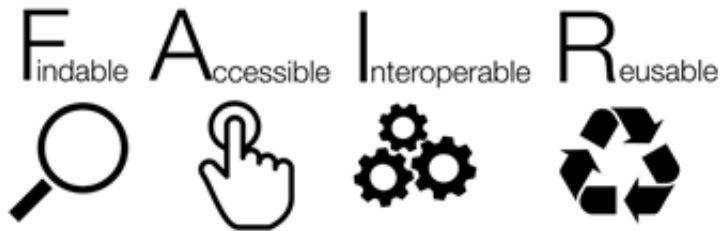
How IBM Watson Overpromised and Underdelivered on AI Health Care

After its triumph on *Jeopardy!*, IBM's AI seemed poised to revolutionize medicine. Doctors are still waiting

Standards are not that important – it is all about FAIR data

What is important: **FAIR data**!

(however, standards & tools help to enforce FAIR data policies)



FAIRsharing.org: repository of standards



1069 Standards

Terminology Artifact	708
Model/Format	241
Reporting Guideline	120

[View all](#)



1001 Databases

Life Science	746
Biomedical Science	184
General Purpose	10

[View all](#)



99 Policies

Funder	22
Journal	69
Society	4

[View all](#)



Our Community of 69 Journals, Funders, and Databases Includes



Previous work: HANDS – Handbook for Natural Data Stewardship

High-level guidance embraced by all Dutch University Medical Centers



Key principles in HANDS:

- Researcher is the data steward
- UMC is responsible for the data
- FAIR principles
- Reuse of data whenever possible
- Privacy protection
- Involve the patient/participant
- Be 'in control'
- Use available expertise
- Reuse existing infrastructure



Implementation supported by network of data stewards

Slide: Paula Jansen; UMCU

A practical data harmonisation use case - COVID-19 observational data portal integrated across medical centres

- Simple approach demonstrating the practical complexities around data standardization projects



WHO-CRF codebook
per hospital/study/..



WHO-CRF codebook
harmonised to ISARIC



Simple website; central only metadata and
aggregated numbers of limited set of data items

High-level timeline

First week July 2020

October 2020

January 2021

Snapshots Mapping (impression)

readme_spreadsheet

File Edit View Insert Format Data Tools Help

Deviations from WHO

Differences between hospitals

It is very hard to make data FAIR/interoperable afterwards:

- We need to use a data standard from the start
- We need to supply sufficient meta data for others to interpret our data

We need to be FAIR-by-design

CEDAR webforms and associated BioPortal for metadata capture

Project Title (1..N)

<< 1 >> 🔗 ☰

Title of the project*

ErasmusMC COVID cohort

Language Tag

en

Project Acronym

ErasmusMC COVID cohort

ZonMW Project Identifier

ZonMW Project Identifier*

0000000000

Project GUPRI

Project Website

Funder Information (1..N)

<< 1 >> 🔗 ☰

Funder Name*

ErasmusMC

Funder GRID reference

Start date*

Jan 1, 2020

End date*



CEDAR

ID from admin template*

<https://repo.metadatacenter.org/template-instances/f3431b2d-39bc-47f0-9dde-45f9a772b83c>

Date*

Dec 11, 2020

Number of patients

316

Number of patients with COVID diagnosis

291

Number of patients in IC

143

Number of deceased patients

57

Number of males

199

Number of females

117

Number of patients with other gender

0

Number of patients younger than 20 years

10

Number of patients between 20 and 40 years

31

Number of patients between 40 and 60 years

103

Result: automatically filled high-level catalogue of COVID studies

Project overview

Lead institution



ZonMw



The impact of COVID-19 on children and adolescents with Autism Spectrum Disorder (ASD) and their families

Last changed
01/03/2021

[ZonMw](#) [Erasmus University Rotterdam](#) [Netherlands](#)

TEST M4M by Jeroen (updt 1+MG survey accessible genomes, dd 12-02-2021)

Last changed
12/02/2021

TEST_CEDAR_M4M_2_MVP2

[ZonMw](#) [VU University Medical Center](#) [Netherlands](#)

Clinical prediction models for Covid-19: development, international validation and use

Last changed
14/12/2020

COPE

[ZonMw](#) [Erasmus MC](#) [Netherlands](#)

Effecten van Covid-19 op huisartsbezoek van patiënten in achterstandswijken

Last changed
02/12/2020

[ZonMw](#) [Erasmus University Rotterdam](#) [Netherlands](#)

SARSLIVA: utility of saliva in COVID-19 diagnosis - a household study

Last changed
27/11/2020

SARSLIVA

[ZonMw](#) [Spaarne Gasthuis - Spaarne Ziekenhuis](#) [Netherlands](#)

Barriers to large-scale sharing of data

[..] the problem is not really technical [...]. Rather, the problems are **ethical, political, and administrative**

Lancet Oncol 2011;12:933

1. Administrative (I don't have the time)
2. Political (I don't want to)
3. Ethical (I am not allowed to)

.....

4. Technical (I can't)



Slide: André Dekker

Lessons learned

- Data standardization initiatives always start too ambitious
 - Be realistic and pragmatic (**80/20 rule**)
- Data capturing at the source (clinical care) is the holy grail, but very complex
 - Lack of structured data and high percentage of missing data
 - **Incentives** for data collectors are often insufficiently considered
 - Language needs to be considered when reusing care data
- Critical **meta data** to interpret the data is often missing
- **Legal basis** for data sharing is often different in care and research
 - Data is available, nicely standardized but cannot be used for research purposes
- Fear for potential data protection issues is obstructing data sharing
 - A crystal-clear **data governance** is a requirement for successful reuse of data