Using EHRs for clinical research: Quality matters

Prof Dipak Kalra
President of i~HD

The European Institute For
Innovation Thr~ugh Health Data

The Value of Health

Political and policy context

Economic context:

- Legacy of the crisis: high debts and deficits

- Continued increases in public health spending anticipated

- Concerns about how this will be paid for (sustainability of public finances)
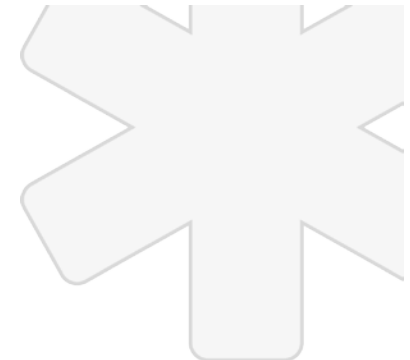
Population health:

- Ageing and rising levels of chronic disease and comorbidity

- Public health problems and inequalities
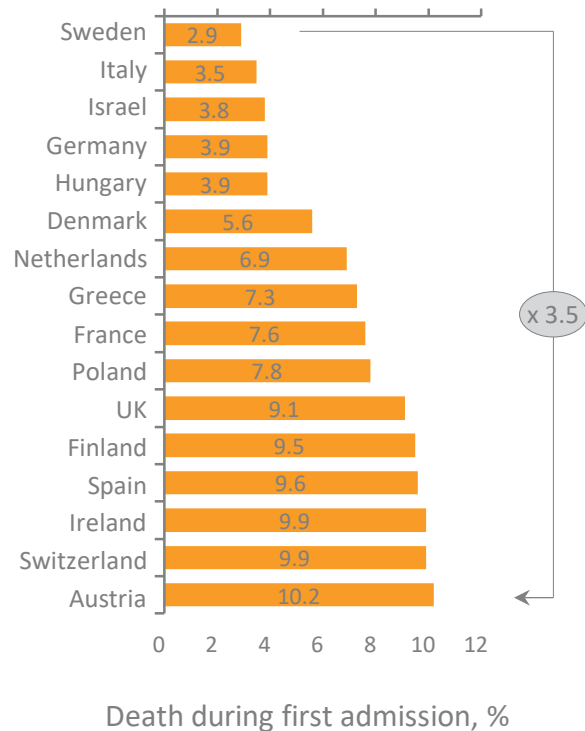
Health systems:

- Challenge of responding to changing population needs

- Need for structural reforms – e.g. integrated care, eHealth

- Evidence of marked variation in clinical practices and significant levels of 'waste'
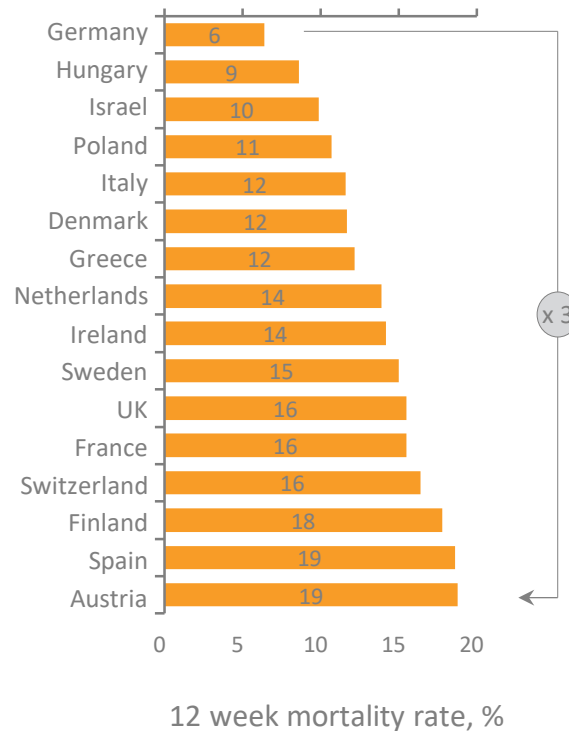
European Commission

ISSN 2443-8014 (online)

Joint Report
on Health Care and
Long-Term Care Systems
& Fiscal Sustainability

Volume 1

INSTITUTIONAL PAPER 037 | OCTOBER 2016

EUROPEAN ECONOMY

Economic and
Financial Affairs
Economic Policy
Committee

Slide courtesy of Daniel Furby, FIPRA International

Variation of heart failure outcomes across European countries

> 3x variation in death during first HF admission

Country	Death during first admission, %
Sweden	2.9
Italy	3.5
Israel	3.8
Germany	3.9
Hungary	3.9
Denmark	5.6
Netherlands	6.9
Greece	7.3
France	7.6
Poland	7.8
UK	9.1
Finland	9.5
Spain	9.6
Ireland	9.9
Switzerland	9.9
Austria	10.2

x 3.5

Death during first admission, %

3x variation in 12 week mortality rate for HF

Country	12 week mortality rate, %
Germany	6
Hungary	9
Israel	10
Poland	11
Italy	12
Denmark	12
Greece	12
Netherlands	14
Ireland	14
Sweden	15
UK	16
France	16
Switzerland	16
Finland	18
Spain	19
Austria	19

x 3

12 week mortality rate, %

>4x variation in 12 week readmission rate for HF

Country	12 week readmission rate, %
Poland	12
Greece	17
Hungary	21
Italy	21
Spain	26
Switzerland	28
Israel	28
Denmark	29
UK	31
Ireland	36
Germany	36
Netherlands	36
France	37
Sweden	41
Austria	41
Finland	54

x 4.5

12 week readmission rate, %
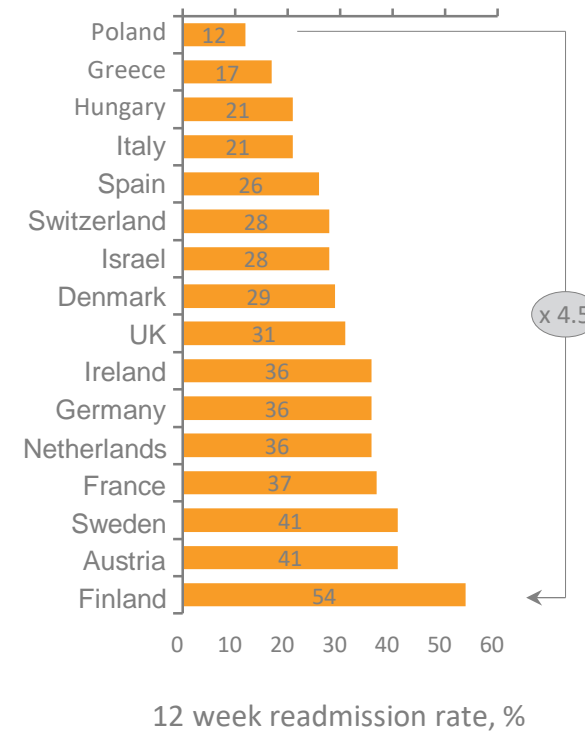
Slide courtesy of Thomas Allvin, EFPIA

efpia

3

Growing international adoption of electronic health record systems

Stage	Description
Stage 7	Paperless environment enabling data continuity to deliver and manage patient care for all hospital services and supporting coordinated care delivery → Optimized quality of care and patient safety → Optimized clinical operational and business process efficiency
Stage 6	Optimized diagnostic and care delivery efficiency by standard treatment protocols; Further reduction / elimination of medical errors → improved patient safety; Reduced / eliminated paper-based documentation and dictation and transcription costs; Improved billing and coding
Stage 5	Elimination of film reduces costs and need for storage space, image sharing works faster and more efficiently, redundant imaging can be avoided
Stage 4	Improved care delivery efficiency and reduced medical errors with CPOE
Stage 3	Optimised collection of nursing and clinical information; Standardized care delivery → first improvements of care quality
Stage 2	Optimised access to diagnostic results, sharing of patient-information
Stage 1	Improved key dept. operations and access to diagnostic information
Stage 0	Low clinical value; mainly patient administration and hospital management

Cross Country EMRAM Score Distribution# (2015 Q2)

Stage	United States*	Canada*	Austria	Netherlands	Germany*	Italy*	Spain*
Stage 7	3.7%	0.2%	0.0%	0.0%	0.6%	0.0%	0.4%
Stage 6	22.2%	0.8%	0.0%	9.5%	0.0%	1.4%	3.9%
Stage 5	30.8%	0.9%	35.7%	38.1%	11.6%	19.4%	42.4%
Stage 4	13.6%	3.3%	2.4%	3.2%	6.7%	0.9%	5.2%
Stage 3	19.7%	31.4%	0.0%	1.6%	4.9%	4.7%	1.7%
Stage 2	4.3%	30.6%	38.1%	46.0%	23.8%	40.3%	26.2%
Stage 1	2.2%	14.2%	2.4%	1.6%	0.6%	22.3%	6.6%
Stage 0	3.5%	18.7%	21.4%	0.0%	51.8%	10.9%	13.5%
	N = 5462	N = 641	N = 42	N= 63	N = 164	N = 211	N = 229

Source: HIMSS Analytics® Database

i~HD

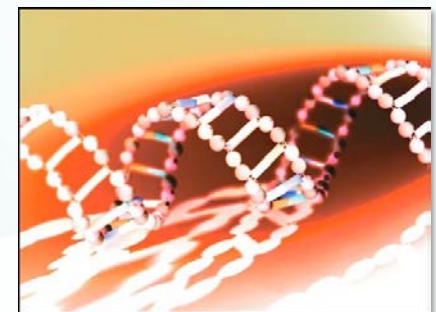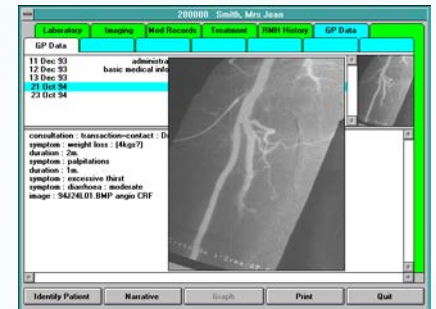# Physicians using the EHR for routine clinical documentation

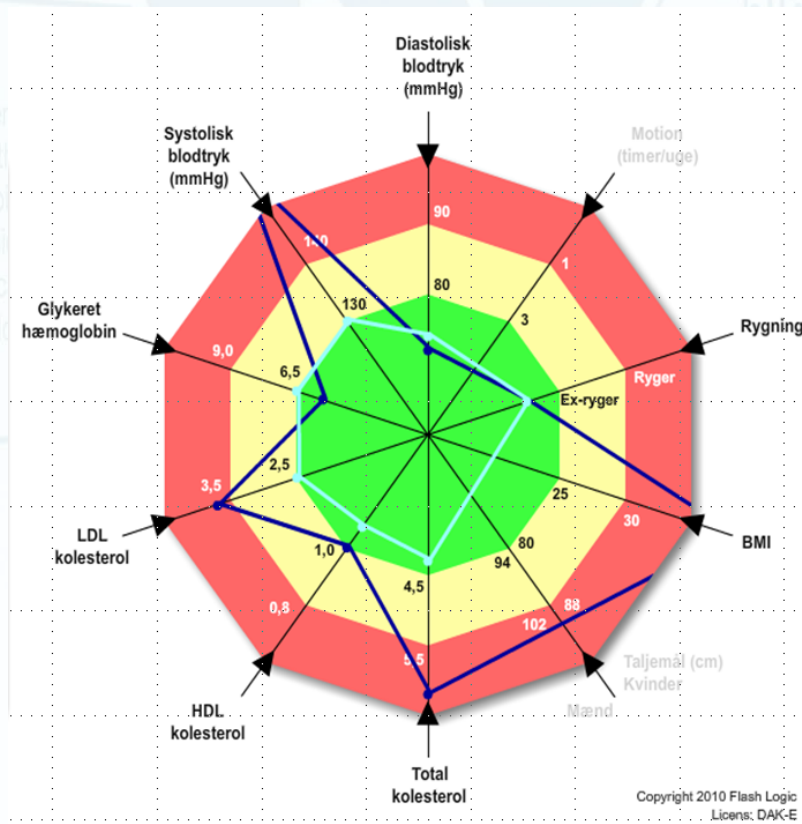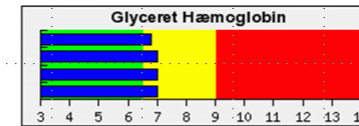| Spain	Germany	US	France	Singapore	Australia	UK	Canada
95%	93%	93%	92%	87%	83%	82%	76%
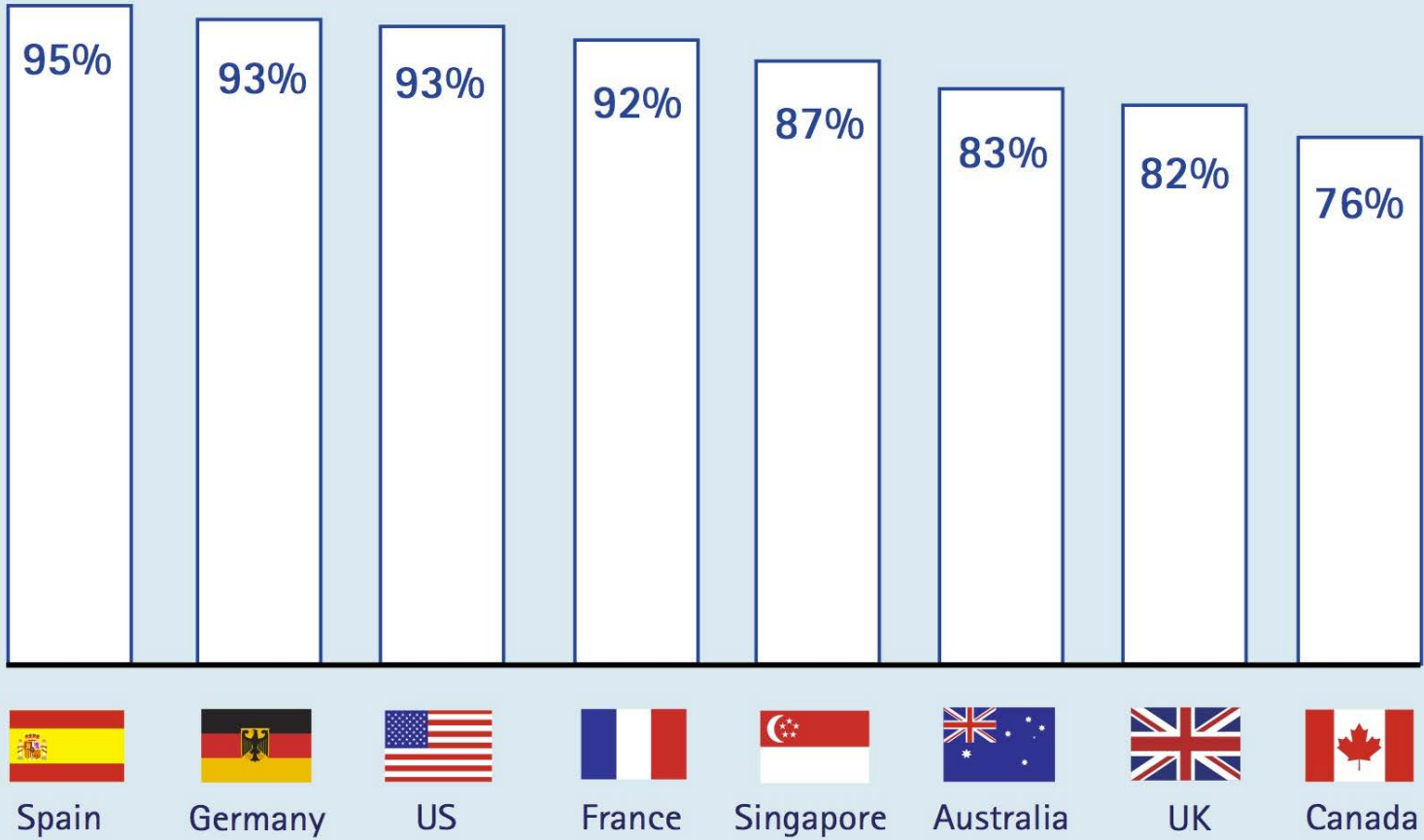
Source: Accenture eight-country survey of 3,700 doctors, Nov-Dec, 2012

EHR trends

- Patient-centred (gatekeeper?), life long records
- Multi-disciplinary / multi-professional
- Transmural, distributed and virtual
- Structured and coded (cf. semantic interoperability)
- More metadata and coding at a granular level!
- Intelligent (cf. decision support), clinical pathways…
- Predictive (e.g. genetic data, physiological models)
- More sensitive content (privacy protection)
- Integrative
- Personalised

i~HD

The patients individual plan and goals
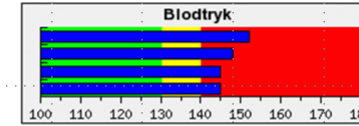
Large scale clinical data repositories

- Kaiser Permanente is able to track outcomes and develop data driven algorithms using the EHRs of its 9 million patients

  - HIV death rate half of national average

  - Decrease in coronary heart disease death rate by a third

  - Decrease in pressure ulcers by two thirds

  - Death due to sepsis reduced by > 50% (for all USA, would save 72,000 lives p.a.)

Some examples of new research evidence from big data

- Validating >200 novel biomarkers predicting cardiovascular risk

- Investigating variation of 174,000 observed national prescribing patterns to national guidelines for COPD

- Comparing ~8,000 treatment outcomes for leukaemia by age: uncovering a major unmet treatment need

- Developing new cancer risk stratification algorithms by mining >700 million records

Singh G, Schulthess D, Hughes N, Vannieuwenhuyse B, Kalra D. Real world big data for clinical research and drug development.
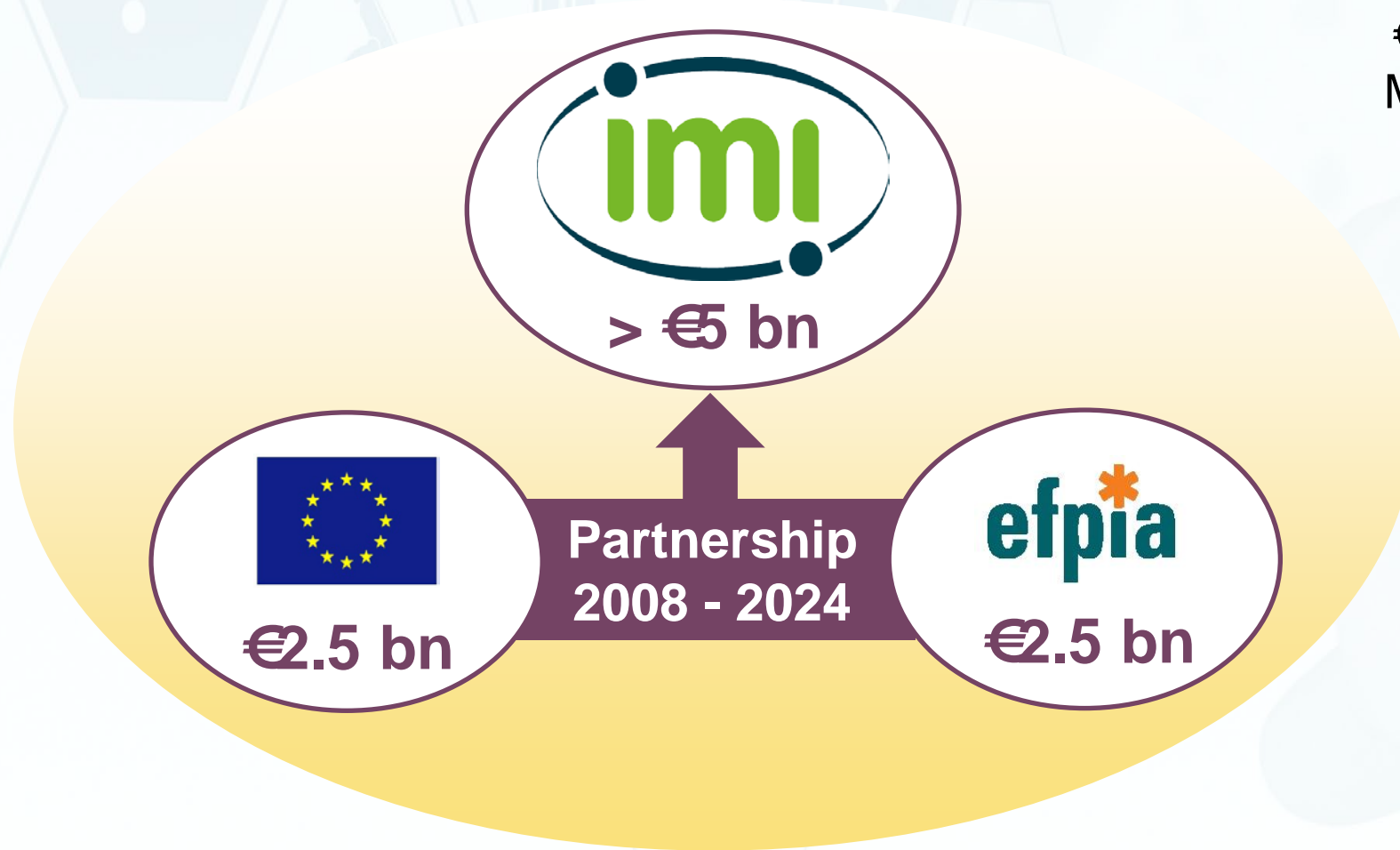*Drug Discov Today. 2017*
https://doi.org/10.1016/j.drudis.2017.12.002

i~HD

IMI – Europe's partnership for health

IMI1: 2008-2013
€2 bn budget
59 projects

IMI2: 2014-2024
€3.3 bn budget
More ambitious
More open
Greater scope

> €5 bn

€2.5 bn

Partnership 2008 - 2024

€2.5 bn

Slide courtesy of Pierre Meulien, Executive Director of IMI

The EHR4CR project

- EHR4CR – Electronic Health Records for Clinical Research

  - 4+1 year project (2011-2016), 35 partners, budget >17M€

- Objectives & Scope

  - Provide a platform for **trustworthy re-use of EHR data** to support innovation in clinical research and healthcare operations.

  - Securely reusing **health data** for optimising clinical trials

  - **7 pilot sites across Europe**

- Status

  - Extended into 2016 for making the transition to a sustainable platform

  - Initiated a **Champion Programme**, connecting hospitals to an operational platform, building up experience with pharma

  - The **European Institute for Innovation through Health Data** – an independent governance body
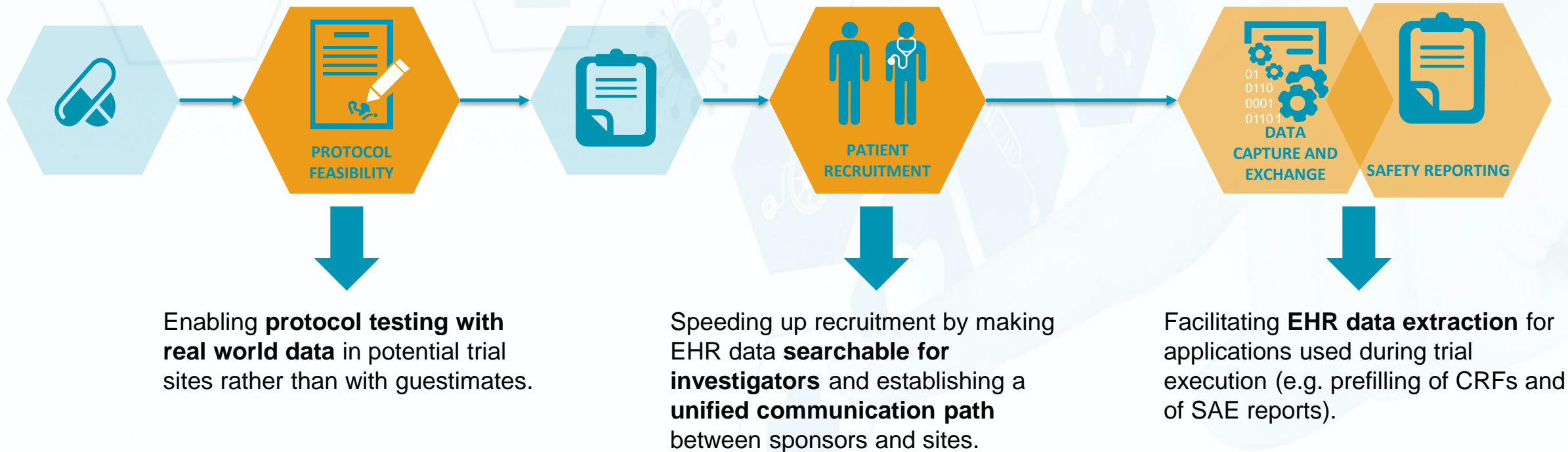
The EHR4CR objective

- Research and develop a **trustworthy service platform able to unlock clinical information stored in EHRs** for improving clinical research
  - Clear focus on three (3) relevant use cases

PROTOCOL FEASIBILITY

PATIENT RECRUITMENT

DATA CAPTURE AND EXCHANGE

SAFETY REPORTING

Enabling **protocol testing with real world data** in potential trial sites rather than with guestimates.

Speeding up recruitment by making EHR data **searchable for investigators** and establishing a **unified communication path** between sponsors and sites.

Facilitating **EHR data extraction** for applications used during trial execution (e.g. prefilling of CRFs and of SAE reports).

Confirming public acceptance

- High percentage of respondents were in favour of re-using EHR data for research
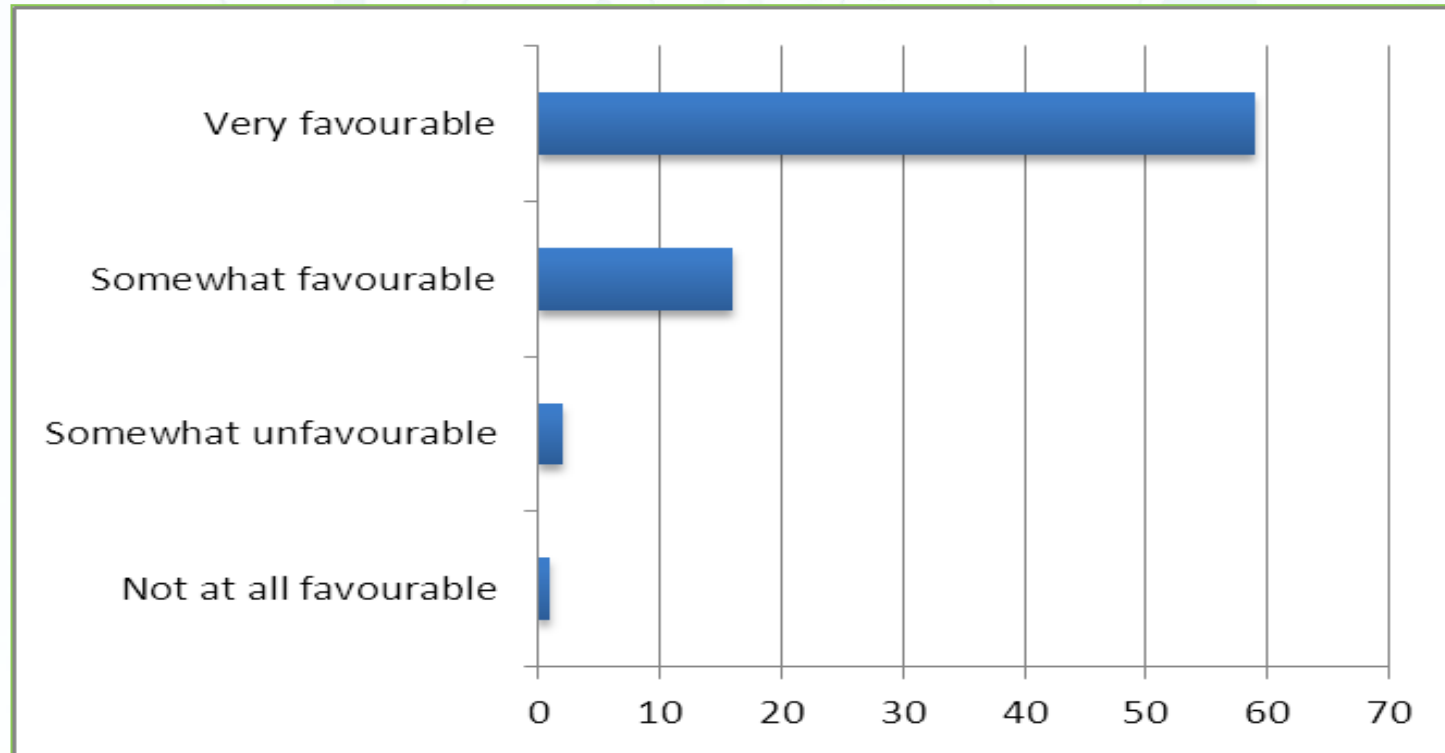
The EHR4CR results

- An innovative platform to enable the trustworthy reuse of health data for research

- The platform can connect securely to the data within multiple hospital EHR systems and clinical data warehouses across Europe

- It enables trial sponsors (e.g. pharma) to
  - predict the number of eligible patients for a candidate clinical trial protocol
  - assess its feasibility and to locate the most relevant hospital sites

- It enables connected hospitals to
  - efficiently identify and contact the patients who may be eligible for particular clinical trials

- Now being deployed commercially: the InSite Platform, by Custodix

Confirming data availability

A European inventory of common electronic health record data elements for clinical trial feasibility

Justin Doods, Florence Botteri, Martin Dugas, Fleur Fritz and on behalf of EHR4CR WP7

Trials 2014, 15:18
http://www.trialsjournal.com/content/15/1/18

Data Group	Data Item		
Demographics	Gender	Medication	Medication start date
Demographics	Case Status	Medication	Verbatim Drug name
Demographics	Date of Birth	Findings	Date / Time of Finding
Demographics	Admission date	Laboratory Findings	Neutrophils Blood
Diagnosis	Diagnosis Text	Laboratory Findings	TSH in serum
Diagnosis	Diagnosis Code	Medication	Dosage
Demographics	Discharge date	Findings	Weight
Diagnosis	Diagnosis Date	Laboratory Findings	GFR
Laboratory Findings	Potassium in serum	Medical History	currently pregnant
Laboratory Findings	Sodium in Serum	Medical History	menopausal status
Laboratory Findings	Platelets Blood	Findings	Height
Laboratory Findings	SGPT (ALT) in serum	Medical History	Allergies and Hypersensitivity Reactions
Laboratory Findings	Total Protein in serum	Laboratory Findings	biPTH
Laboratory Findings	Total Bilirubin in serum	Medical History	Smoking Status
Procedure	Procedure Code	Medication	Route
Laboratory Findings	Creatinine in serum	Laboratory Findings	HbA1c Blood
Laboratory Findings	Glucose in serum	Medical History	Alcohol Abuse
Laboratory Findings	SGOT (AST) in serum	Laboratory Findings	Blood Urea Nitrogen [BUN]
Laboratory Findings	Alkaline Phosphatase	Medication	Medication Code
Laboratory Findings	Total Cholesterol in serum	Findings	Pulse
Laboratory Findings	Erythrocytes	Laboratory Findings	PSA
Laboratory Findings	Haemoglobin Blood	Laboratory Findings	NTproBNP
Laboratory Findings	Albumin	Diagnosis	Histologically confirmed diagnosis
Laboratory Findings	Calcium in serum	Laboratory Findings	Beta HCG in serum
Laboratory Findings	Leukocytes	laboratory findings	HER2 status
Procedure	Procedure Text	Laboratory Findings	Ca x P
Laboratory Findings	Sampling Date / Time of Laboratory Finding	Medication	Drug class
Laboratory Findings	Triglycerides	Laboratory Findings	Cardiac troponin T
Laboratory Findings	CRP in serum	Medical History	pregnancy number
Laboratory Findings	HDL in serum	Medication	Medication end date
Laboratory Findings	INR Blood	Findings	Temperature
Laboratory Findings	Haematocrit Blood	Laboratory Findings	Direct Bilirubin in serum
Procedure	Procedure Date	Medical History	Diet
Laboratory Findings	Eosinophils Blood	Medical History	Substance Abuse
Laboratory Findings	Lymphocytes Blood	Laboratory Findings	BNP
Laboratory Findings	PTT Blood	laboratory finding	MAGE-A3 status
Laboratory Findings	GGT	Medical History	Lactation
Findings	Blood pressure systolic	Scores or Classification	GRID-HAMD
Findings	Blood pressure diastolic	Scores or Classification	Hoehn and Yahr
Laboratory Findings	LDL in serum	Scores or Classification	MMSE
		Scores or Classification	UPDRS Section 1

Figure 3 Heat map of the data exports from the data inventory current version. The first two columns describe the ISO 11179 data element concept (data group/data item). The third column shows the average usage of the data element over all sites while the following columns (site 1 to site 9) display the frequency at the individual sites. The Data Inventory is ordered by the average usage sorted in descending order from most available to least. The frequency ranges from 100% (dark green) to 0% (dark red). Data elements that are not available at a site are shown as Not Available (NA) (black).

InSite – technical overview, for protocol feasibility

Custodix provides
expertise and tools to
support local sites with
mappings

Only aggregated data
(patient counts) leave
the hospital,
only on approval

EHR

ETL

CDW

InSite
Local Install

InSite
Central
Platform

Secure access
for researchers

Local InSite
applications to
support recruitment

Protected by Privacy
Enhancing Techniques
e.g. suppression of small
counts
Full audit trail inside hospital
External governance by i~HD

InSite – Protocol feasibility query

RWE through the lifecycle

Dr Peter Arlett, EMA, 2016
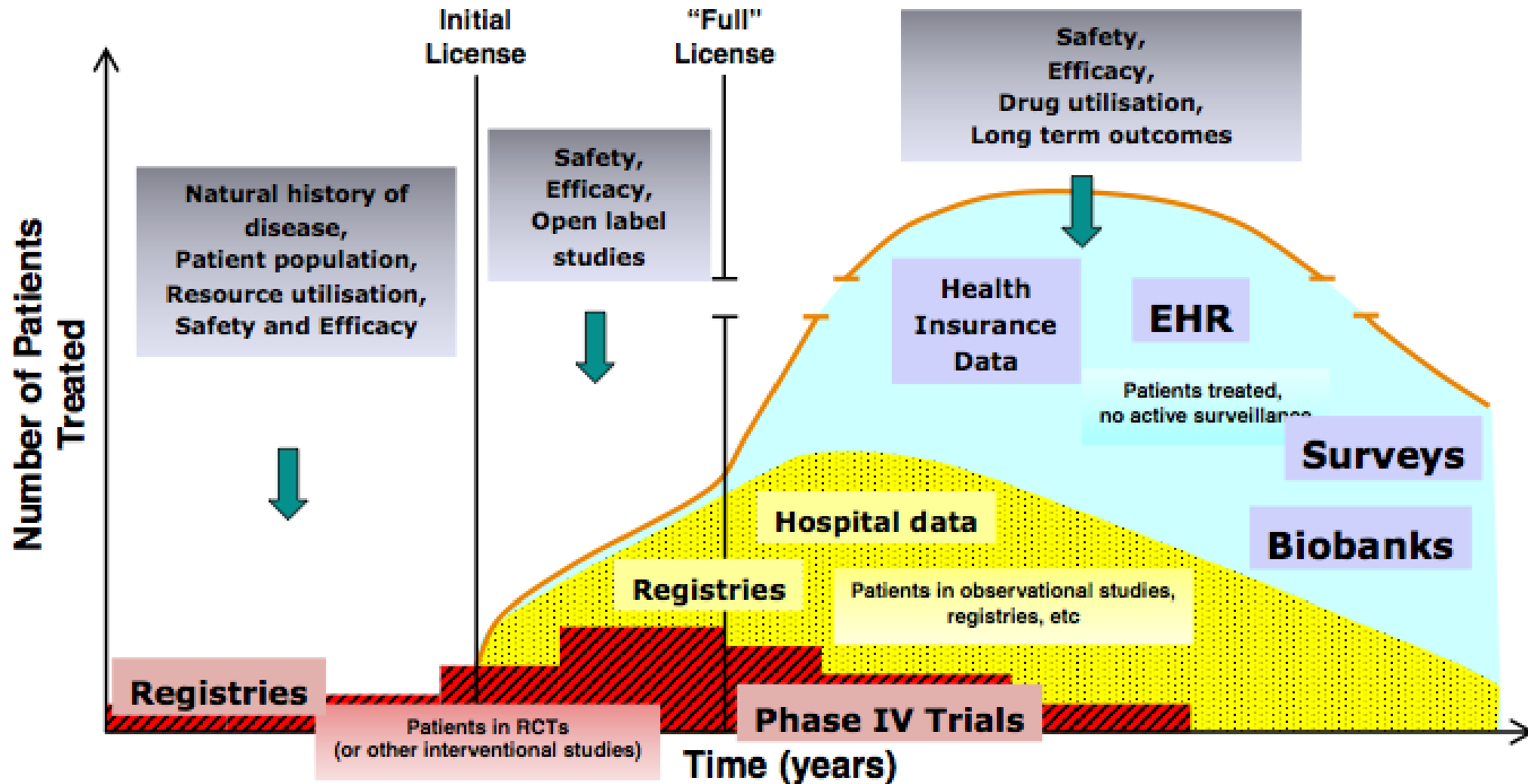
EMIF vision

To become the trusted European hub for health care data intelligence, enabling new insights into diseases and treatments

Discover

Reuse

Assess

Slide courtesy of Bart Vannieuwenhuyse, Janssen
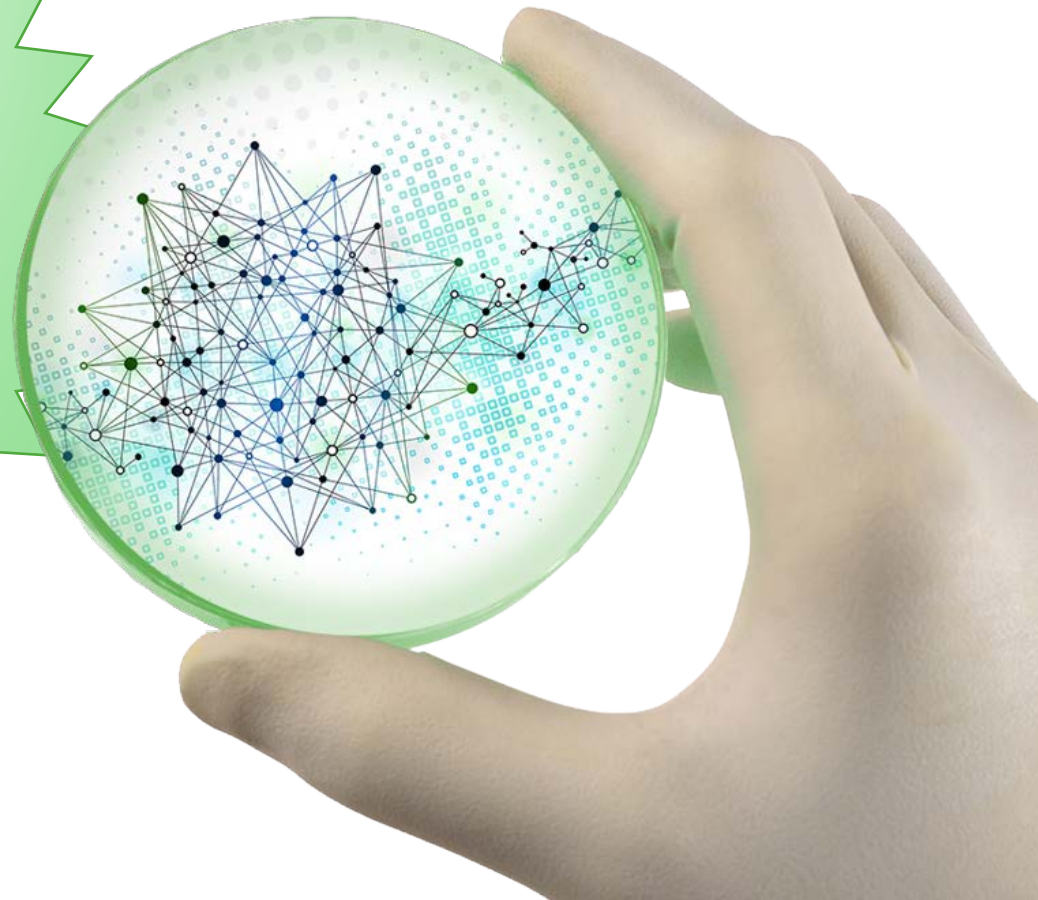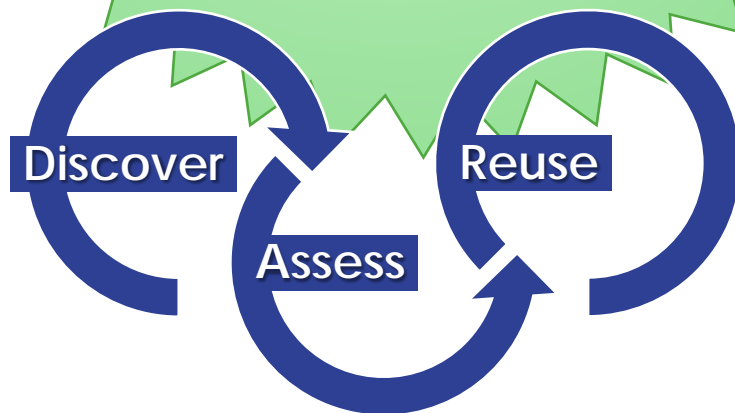
Data Discovery → Data Access → Data Reuse
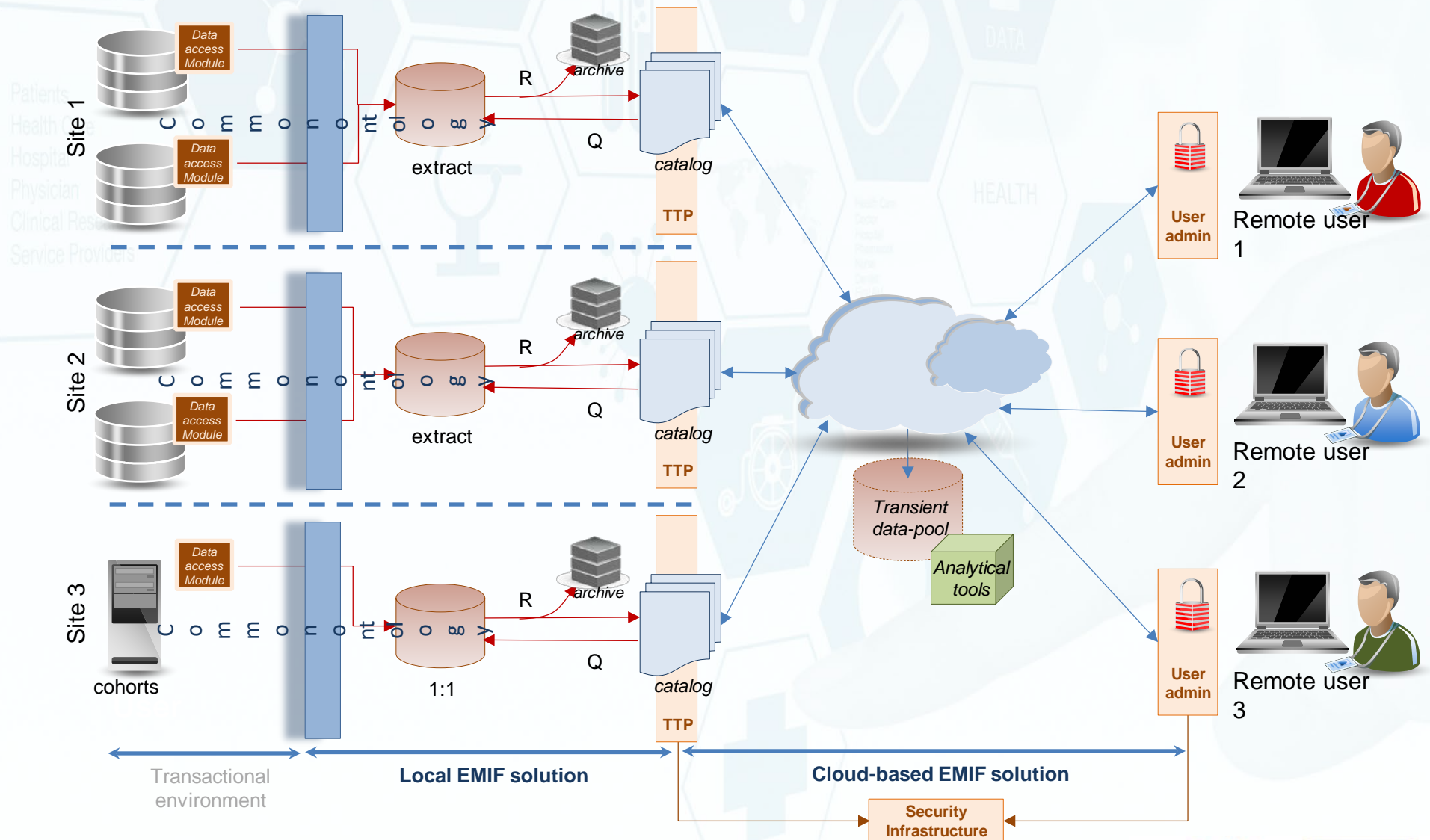
EMIF catalogue: meta-data of available datasets (emif catalogue.eu)

Catalogue – suitability

Source: Rudi Verbeeck - Janssen

The OMOP Common Data Model

Standardized clinical data

- Person
- Observation_period
- Specimen
- Death
- Visit_occurrence
- Procedure_occurrence
- Drug_exposure
- Device_exposure
- Condition_occurrence
- Measurement
- Note
- Observation
- Fact_relationship

Standardized health system data
- Location
- Care_site
- Provider

Standardized health economics
- Payer_plan_period
- Cost

Standardized derived elements
- Cohort
- Cohort_attribute
- Condition_era
- Drug_era
- Dose_era

Standardized meta-data
- CDM_source

Standardized vocabularies
- Concept
- Vocabulary
- Domain
- Concept_class
- Concept_relationship
- Relationship
- Concept_synonym
- Concept_ancestor
- Source_to_concept_map
- Drug_strength
- Cohort_definition
- Attribute_definition

23

TranSMART data load using Webprotégé
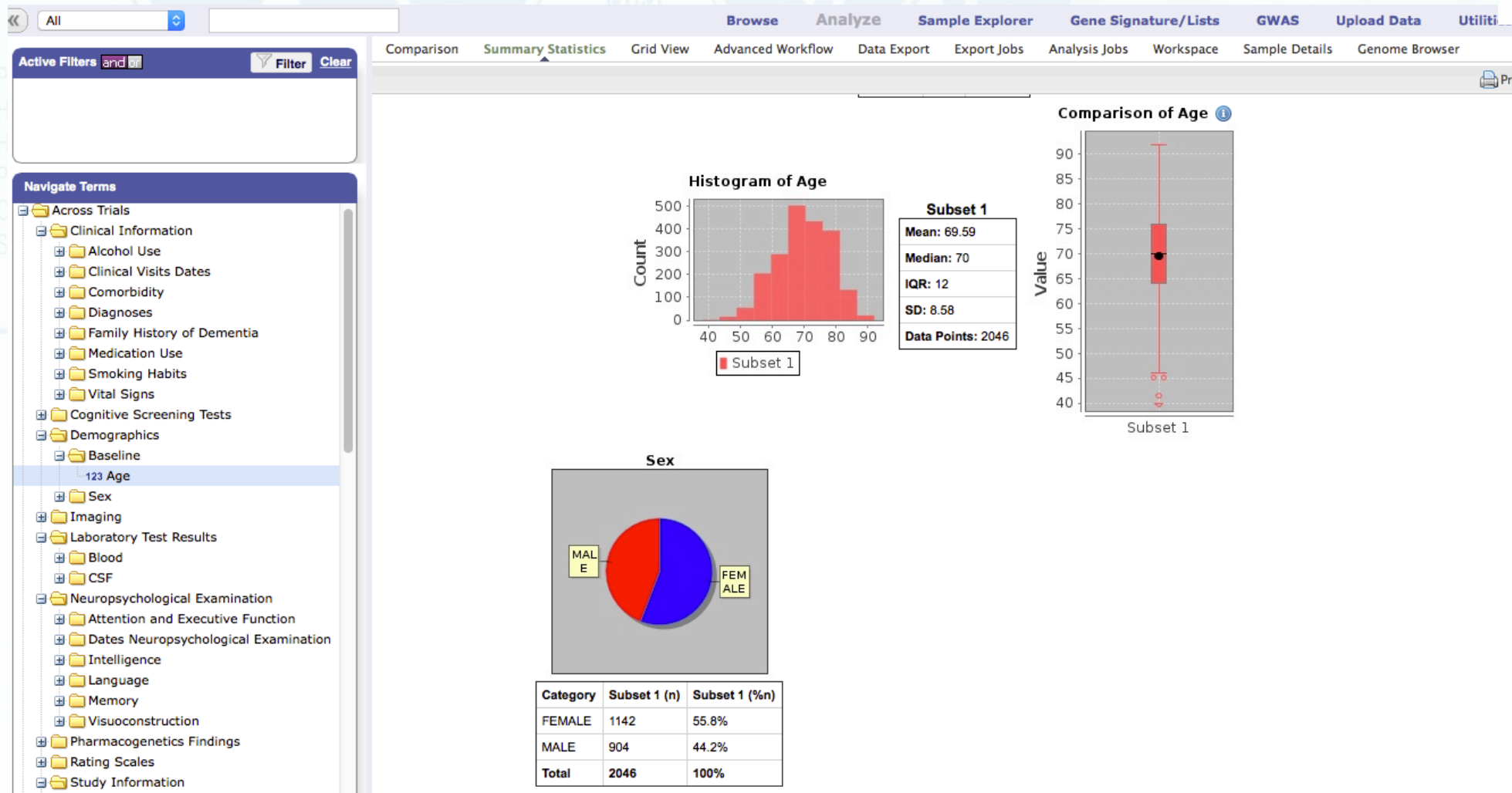

Webprotégé
- Collaborative environment
- Define global concepts
- Owned by EMIF AD
- Specify mappings to local concepts in free text
- Implement mappings in tranSMART ETL

EMIF-AD tranSMART data platform

A convergence of opportunity

Clinical Research

Optimise clinical research processes
- achieve faster and more accurate patient identification
- identify sites that have access to the most suitable patients
- reduce protocol amendments

Enhance access to Real World Data
- study the use of new medicines in real populations
- conduct comparative effectiveness studies
- monitor long term safety
- build evidence for adaptive pathways

Healthcare

Improve quality and safety of care
- enhance care co-ordination
- increase adherence to clinical evidence
- reduce medical errors and treatment delays

Support patients in self-care and health maintenance

Improve efficiency of care
- optimise care pathways to improve outcomes
- collate evidence for public health strategy and decision-making

**Need to improve access to combined
health data from diverse sources across Europe**

Common challenges to the use of health data for person centred care, and the re-use of health data for clinical research

Privacy protection, ethics and security

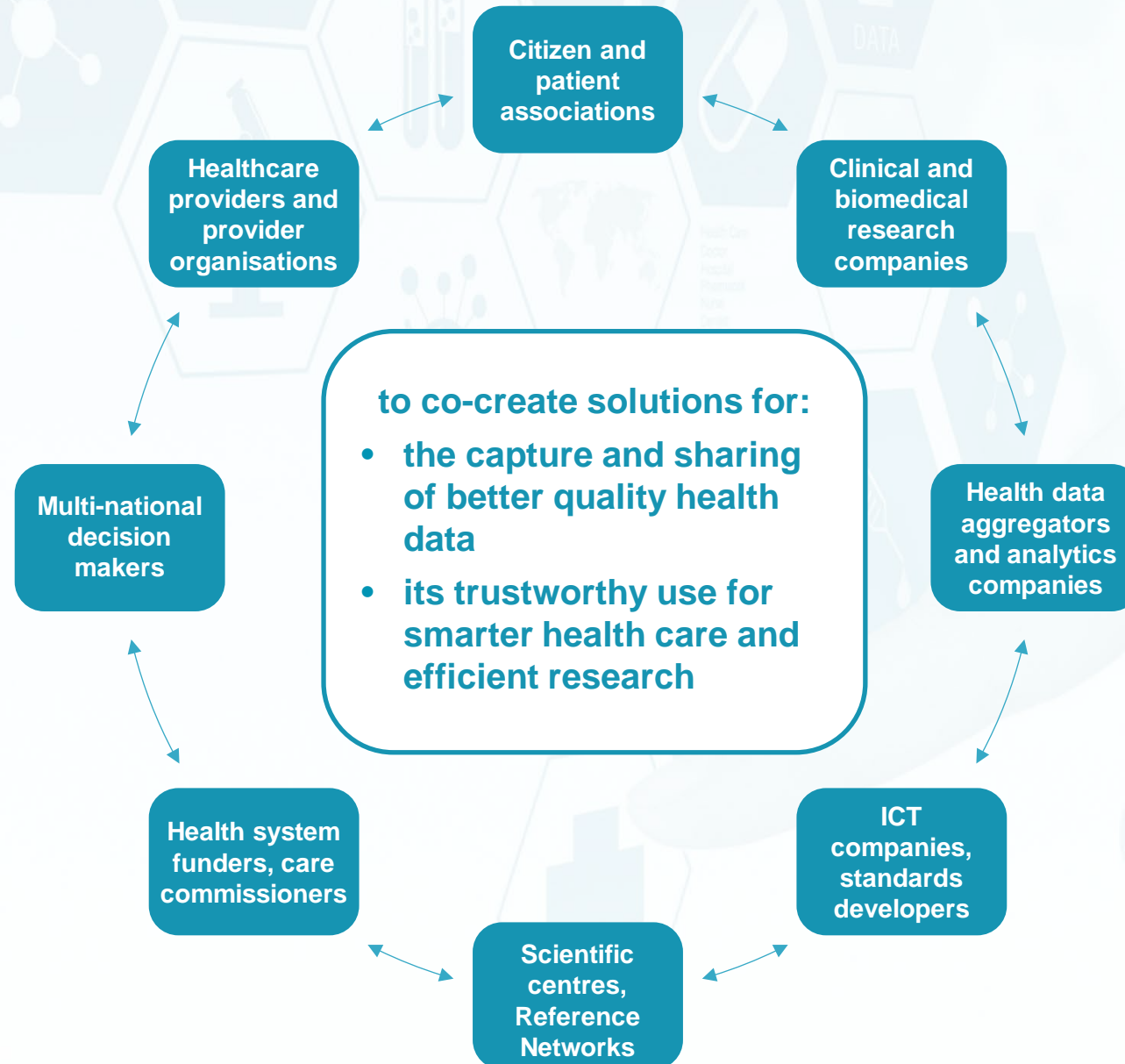*accessing data*

Quality and interoperability of health data

learning from the data

Demonstrating value

transforming healthcare

i~HD

The European Institute for Innovation through Health Data (i~HD): a neutral body, bringing stakeholders together

Citizen and patient associations

Clinical and biomedical research companies

Healthcare providers and provider organisations

Health data aggregators and analytics companies

to co-create solutions for:

- **the capture and sharing of better quality health data**

- **its trustworthy use for smarter health care and efficient research**

Multi-national decision makers

ICT companies, standards developers

Health system funders, care commissioners

Scientific centres, Reference Networks

Goal of i~HD: assuring public trust when reusing EHRs for research
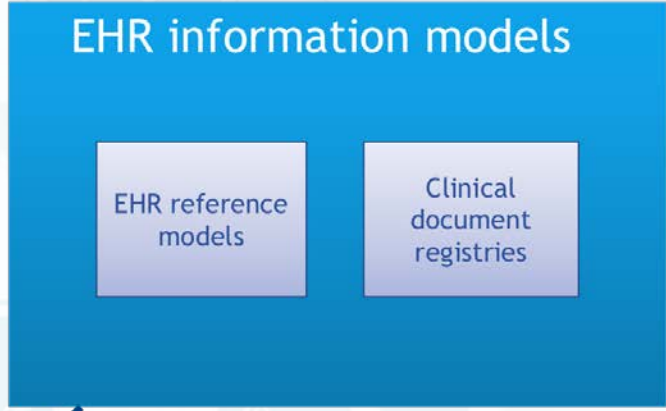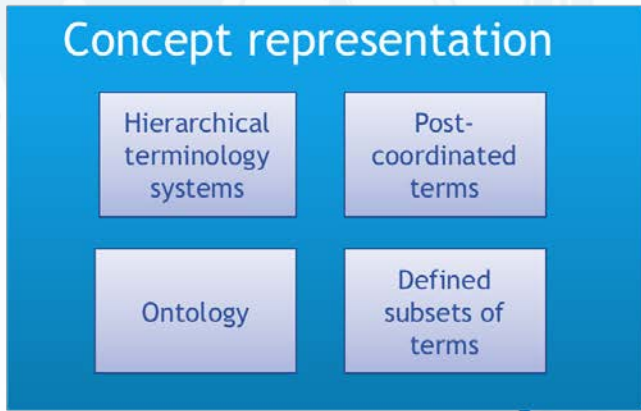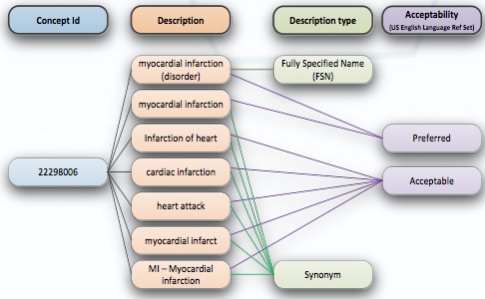
- Compliance with data protection legislation, at a European level and across all European Member States

- Consistent information governance practices and expectations across Europe
  - Societally acceptable good practices for governing many uses of health data
  - Reflect state of the art in privacy protection and information security
  - Greater confidence and reduced risk for those providing data for research use e.g. hospitals, GPs, patients
  - Greater confidence and reduced risk for those performing the research, managing the data or sponsoring the research

- Greater societal endorsement of public health and research uses of health data

- A scaling up of learning from health data, leading to more rapid innovation in treatments, and accelerated health system transformation towards better health outcomes
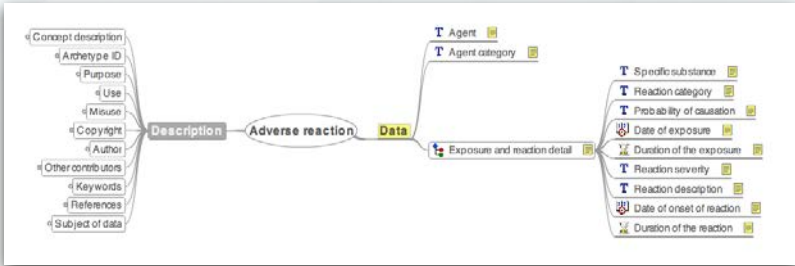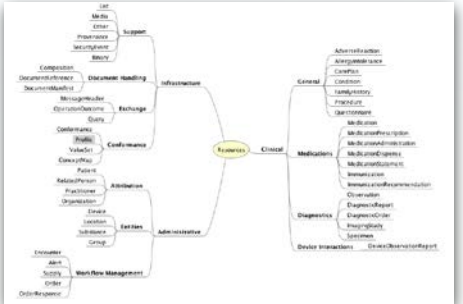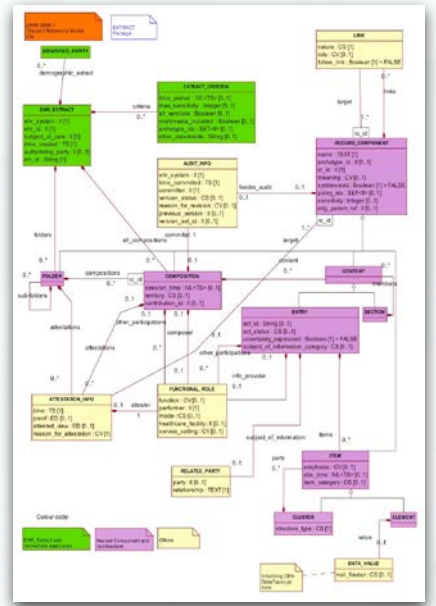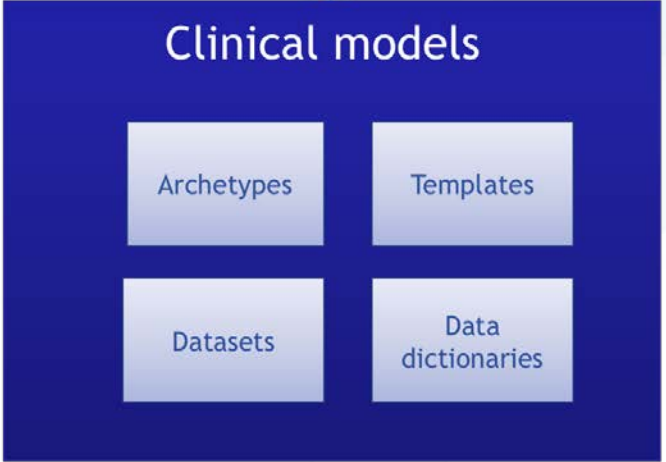
Essential needs for interoperability

- Guideline and decision support systems, notification and alerting components, and analytic tools need to process <u>integrated health data</u> drawn from <u>multiple EHR systems</u> in a <u>consistent manner</u>

- <u>Intelligent personal health guidelines <u>interoperating with PHRs</u> <u>and EHRs</u> need to support the centring of care on patients

- Health services, insurers and public health bodies need <u>fine grained activity and outcome data</u> to inform service planning, commissioning and prevention/wellness programmes

- New generation personalised medicine, underpinned by 'omics sciences and translational research such as the VPH, needs to <u>integrate EHRs with data from research</u>: fundamental biomedical science, clinical and population health research, and clinical trials

*i~*HD

Overview of assets used to represent clinical meaning

Concept representation
- Hierarchical terminology systems
- Post-coordinated terms
- Ontology
- Defined subsets of terms

EHR information models
- EHR reference models
- Clinical document registries

Use Case Driven

Clinical models
- Archetypes
- Templates
- Datasets
- Data dictionaries

Developing good practices in the collaborative development of standards

Clinical specifications

Technical specifications

combine modelling, terminology, ontology & workflow representations to produce harmonised semantic resources

Clinician and patient involvement

Research and public health involvement

Slide courtesy of Robert Vander Stichele
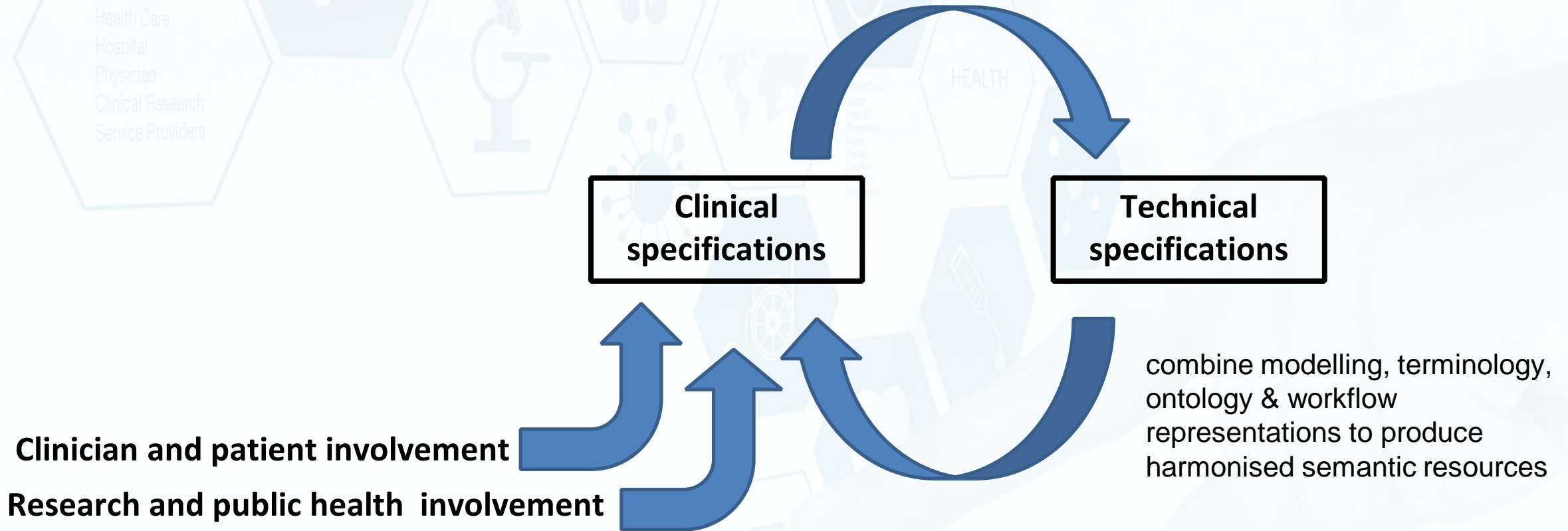
i~HD semantic interoperability priorities

- Fostering the co-design of semantic interoperability assets by clinicians, patients, research

- Designing quality processes for clinical information models and terminology value sets

- Quality labelling and providing a public directory of interoperability assets

- Promoting adoption of interoperability standards and profiles

- Forging an alliance amongst standards developers, implementers, purchasers and users, on future EHR data quality and standards

i~HD Data Quality Taskforce aims

- Develop data quality assessment methods, tools and improvement strategies to maximise quality of health data

- Promote the importance of data quality

- Guidance in assessing and improving data quality

- Scale up a multi-stakeholder understanding and commitment to increase data quality

➔ Focus on three areas:
  - ➔ Healthcare
  - ➔ Clinical trials
  - ➔ Big data
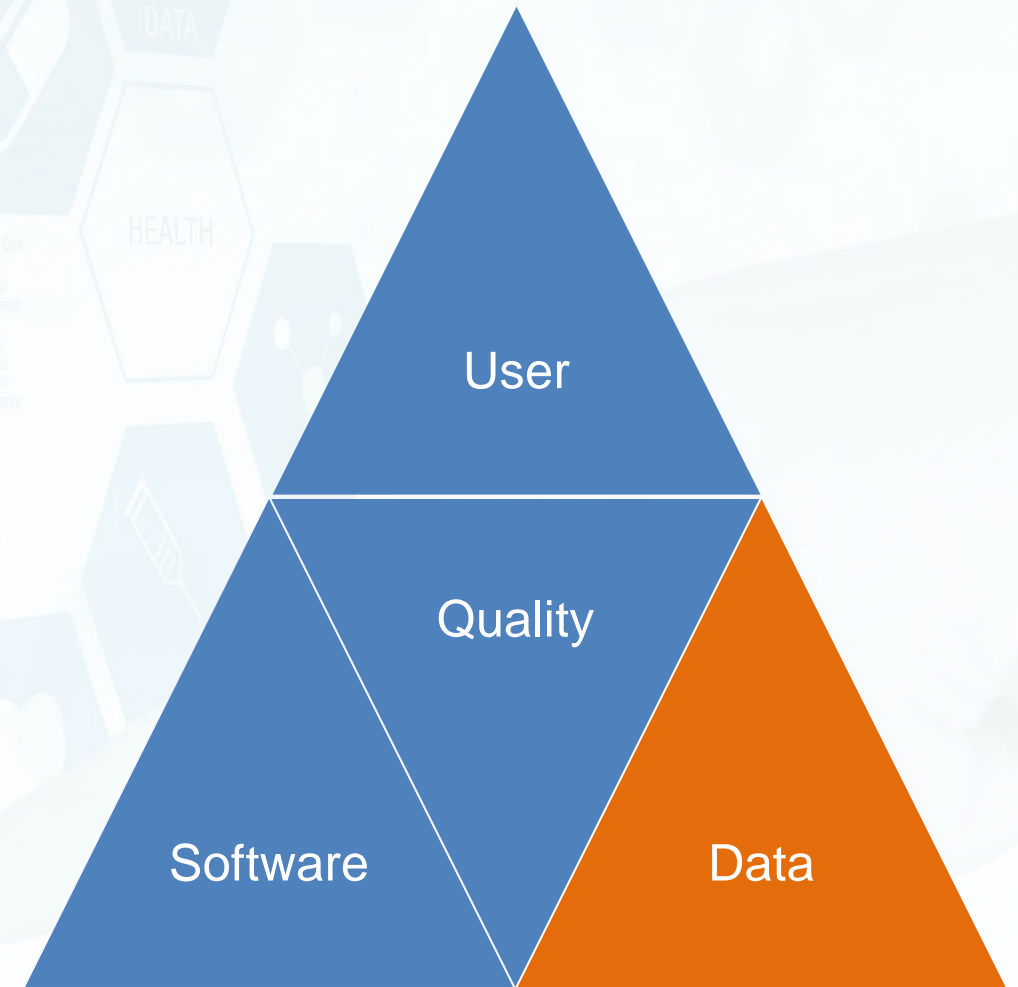
Slide courtesy of Pascal Coorevits, Ghent University & EuroRec and Carlos Sáez, Universitat Politècnica de València

User

Quality

Software

Data

Data quality dimensions

Name	Definition
Completeness	Data values are **present**
Consistency	Data satisfy **constraints** (format, allowable ranges and values, domain rules, relations)
Correctness	Values are **true and unbiased** with respect to their real-world state
Uniqueness	Records representing a single patient are **not replicated**

Name	Definition
Timeliness	Data is **up-to-date** to their real world state for the task at hand
Stability	Data inherent concepts and statistics are **comparable among sources** (hospitals, professionals, etc) **and over time**
Relevance	Data are **useful** for their task
Contextualization	Data are annotated with the acquisition context, their **meaning and semantics**
Trustworthiness	Data can be trusted based on the **reputation** of the stakeholders involved in their acquisition

Slide courtesy of Pascal Coorevits, Ghent University & EuroRec
and Carlos Sáez, Universitat Politècnica de València

i~HD

Population registries,
Clinical trials databases

Care pathways,
decision support,
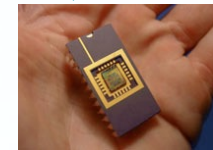trends and alerts

Genomic data

Environmental data

Bio-sensors

Social networks

Mobile devices
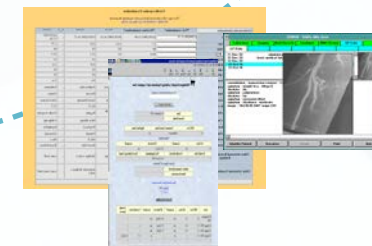
Clinical
applications

i~HD

36

Demonstrating value from the use of health data

Value to healthcare

Value to patients and to society

Value to research

Grow a Network of Excellence

Evaluate the benefits of using health data on a large scale:

- outcomes evidence to improve care
- faster and more efficient clinical research
- stronger evidence for public health strategy
- patient empowerment for better health

Promote our mission and collaborate globally