EU-STANDS4PM
standards for in silico models for personalised medicine

A European standardization framework for data integration and data-driven *in-silico* models for personalised medicine

Annual meeting 2022 - Introduction

Marc Kirschner, coordinator on behalf of the EU-STANDS4PM consortium
m.kirschner@fz-juelich.de

© EU-STANDS4PM - 2022

At a glance

- **Type:** Horizon2020 Coordination and support action
- **Project duration:** 4 years (Jan 2019 - Dec 2022)
- **Consortium:** 16 partners, 8 EU countries
- **Budget:** 2.0 Mio €
- **Coordinator:** Forschungszentrum Jülich GmbH, Project Management Jülich

Major objective: To establish a forum to develop recommendations, guidelines and standards for in silico models in personalized medicine.

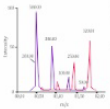
© EU-STANDS4PM - 2022

Data and models for personalized medicine

Models addressing clinical questions

Response to therapy
Risk prediction
Pharmacodynamics
Pharmacokinetics
In silico clinical trials

Possible data collections



Molecular data
-Omics
-Serology
- ...

Clinical data
-Medical records
-Medical imaging
- ...

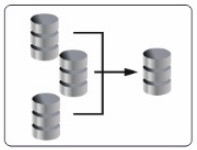
Lifestyle data
-Questionnaires
-Wearable devices
- ...

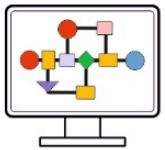
Registry data
-National registies
-Quality registries
- ...

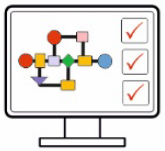
Reference data
-Sequences
-Phenomes
- ...

Pre-examination data
-Specimen collection/storage
-Post-collection modifications
- ...

Data integration → **Model simulation** ⇄ **Model validation** → **Clinical applicability**







Requirements for standardized procedures

Requirements for standardized procedures

© EU-STANDS4PM - 2022

Project structure and output

pan-European expert forum

WP3-
Legal&ethical frame

WP1-
Data sources

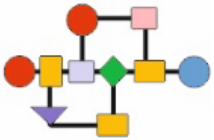
WP2-
In silico models

WP4-
Data access& governance (pilot)

WP4-
Awareness

Models for personalized medicine

Response to therapy
Risk prediction
Pharmacodynamics
Pharmacokinetics
In silico clinical trials



Recommendations and standards for

- ⇨ Data integration and model validation
- ⇨ Ethico-legal issues and needs
- ⇨ Data governance

Target communities

- ⇨ European collaborative research
- ⇨ Funding organizations

© EU-STANDS4PM - 2022

18 May 2022 – EU-STANDS4PM annual meeting part I

Welcome

10:00 – 10:15 **Introduction**
Marc Kirschner, Forschungszentrum Jülich, Germany

Presentations of project output by work package and task leaders

10:15 – 10:45 **A catalogue of personalized medicine reference data**
Niklas Blomberg et al., EMBL-EBI ELIXIR, United Kingdom

10:45 – 11:00 **Innovative data governance for collaborative research projects: A new harmonized data access agreement for controlled access data.**
Stephan Beck, University College London, United Kingdom

11:00 – 12:00 **The legal and ethical framework necessary to develop lawful and ethical data integration in a fair and transparent way, respecting patients' rights.**
Mette Hartlev, University of Copenhagen, Denmark
Miranda Mourby, University of Oxford, United Kingdom
Eugenijus Gefenas, Vilnius University, Lithuania

12:00 – 12:30 **Virtual coffee break**

12:30 – 13:30 **Development of community-based guidelines and normative documents: Recommendations and requirements for predictive computational models in personalised medicine**
Lars Küpfer, RWTH Achen University, Germany
Catherine B. Collin University of Copenhagen, Denmark
Martin Golebiewski, HITS gGmbH, Germany

13:30 – 14:15 **Moderated discussion**: Q&A session to EU-STANDS4PM project output
Catherine B. Collin, University of Copenhagen, Denmark

ca. 14:00 Closure

19 May 2022 – EU-STANDS4PM annual meeting part II

Welcome

10:00 – 10:05 Introduction
Marc Kirschner, Forschungszentrum Jülich, Germany

Presentations by invited speakers

10:05 – 10:30 **-The SimCardioTest project and verification & validation of in-silico models**
-Standards in VPHi
Liesbet Geris (University of Liège, Belgium)

10:30 – 10:45 **EU4CHILD – A crowdsourced ecosystem to fight childhood cancer**
Alberto Tozzi (Ospedale Pediatrico Bambino Gesù, Rome, Italy)

10:45 – 11:15 **Challenges for clinically driven research: Guidelines for implementing computational models in clinical integrated decision support systems**
Tito Poli (University of Parma, Italy)

11:15 – 11:30 **Coffee break**

11:30 – 12:00 **The role of standards in EU funded collaborative research projects**
Gergely Tardos (European Commission, DG Research & Innovation, Belgium)

12:00 – 12:30 **Computational modeling and simulation in medicine – challenges and needs for standardization**
Regina Geierhofer (Siemens Healthineers, Technical Regulations & Standards, Secretary IEC TC62, SC62B und SC62C, Germany)
Charlott Danielson (Fraunhofer Research Institution for Individualized and Cell-Based Medical Engineering (IMTE), Germany)

12:30 – 13:00 **Lunch break**

13:00 – 13:30 **ISO standards relevant for personalized medicine**
Heike Moser (DIN, Germany)
Martin Golebiewski (HITS gGmbH, Germany)

13:30 – 14:15 **Moderated discussion:** Implementation strategies for the standard documents developed by EU-STANDS4PM
Martin Golebiewski (HITS gGmbH, Germany)

ca. 14:15 Closure

The players and their key tasks

H2020 core projects/ data governance pilot

UCL — Stephan Beck

Università di Parma, BD2Decide — Tito Poli

QIAGEN, SPIDIA4P — Uwe Ölmüller

CAU Christian-Albrechts-Universität zu Kiel, SYSCID — Philip Rosenstiel

Rotterdam University, LifeCycle — Vincent Jaddoe

Karolinska Institutet, MultipleMS — Ingrid Kockum

Standardization

DIN — Heike Moser

HITS — Martin Golebiewski

QIAGEN — Kalle Günther

Data and models

elixir — Niklas Blomberg

University of Copenhagen — Catherine Collin

Universität Rostock — Olaf Wolkenhauer

RWTH Aachen University — Lars Küpfer

Legal/ethical frame

University of Copenhagen — Mette Hartlev

University of Oxford — Miranda Mourby

Vilniaus Universitetas 1579 — Eugenijus Gefenas

Regulatory support

famhp federal agency for medicines and health products — Flora M. Tschinanu

Coordination

JÜLICH

ptj Projektträger Jülich Forschungszentrum Jülich

Marc Kirschner
Sylvia Krobitsch

© EU-STANDS4PM - 2022

Acknowledgements

EU-STANDS4PM is funded by the European Union Horizon2020 framework programme of the European Commission, Directorate-General for Research and Innovation under Grant Agreement # 825843.

© EU-STANDS4PM - 2022

EU-STANDS4PM
standards for in silico models for personalised medicine

Thank you for joining the EU-STANDS4PM annual meeting 2022!

© EU-STANDS4PM - 2022

EU-STANDS4PM
standards for in silico models for personalised medicine

A European standardization framework for data integration and data-driven *in-silico* models for personalised medicine

WP1: Data sources and standards for predictions in personalized medicine

Niklas Blomberg, ELIXIR, niklas.blomberg@elixir-europe.org
Ingrid Skelton Kockum, KI,
Martin Golebiewski, HITS, martin.golebiewski@h-its.org

Partners: HITS, KI, EMBL, ELIXIR, DIN, UCPH, Bayer, QIAGEN, ERASMUS

© EU-STANDS4PM - 2022

WP1 - Data sources and standards for predictions in personalized medicine

- Objective 1: To establish a pan-European standardization framework for in silico methodologies applied in personalized medicine
- Objective 2: To develop pan-European, generally admitted recommendations for standardization guidelines for in silico methodologies applied in personalized medicine

A central goal of WP1 is to produce an overview of existing relevant data sources and domain-specific data standards already in use by various scientific research communities.

The proposed EU-wide mapping process thus focuses on

1. **European data-bases, collections and registries,**
2. **current personalized/systems medicine projects**
3. **good examples/successful case studies for integrating phenotype and large-scale data.**

© EU-STANDS4PM - 2022

Tasks

Task 1.1 Survey of European and international reference databases, collections and registries for personalized medicine *(Lead: Ingrid Kockum, KI)*

Task 1.2 Harmonizing data and metadata standards, minimal information guidelines for reporting and international standardization efforts *(Lead: Martin Golebiewski, HITS)*

Task 1.4 Preparations for the development of European level standardization documents with regard to interoperability of health-related data *(Heike Moser, DIN) -> Operationally merged with Task 1.2*

Task 1.3 Assembling successful case-studies — examples and good practice for integrating phenotype and large scale data

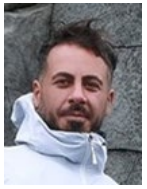
© EU-STANDS4PM - 2022

Task 1.1 Survey of European and international reference databases, collections and registries for personalized medicine

Aim: Survey existing data resources, collections and registries relevant for personalized medicine

Deliverable 1.1: Survey of available relevant data sources

- Survey was open for 6 months (November 2019 - April 2020)
- Targeted to key contacts such as research infrastructures, national cohorts, consortiums
- Online survey with 92 questions
  - 5 general
  - 52 datasources and standards
  - 32 modeling methods and standards
  - 3 data access consent
- 71 respondents
  - 11 EU countries, UK and US


Ali Manouchehrinia
Karolinska Insitutet

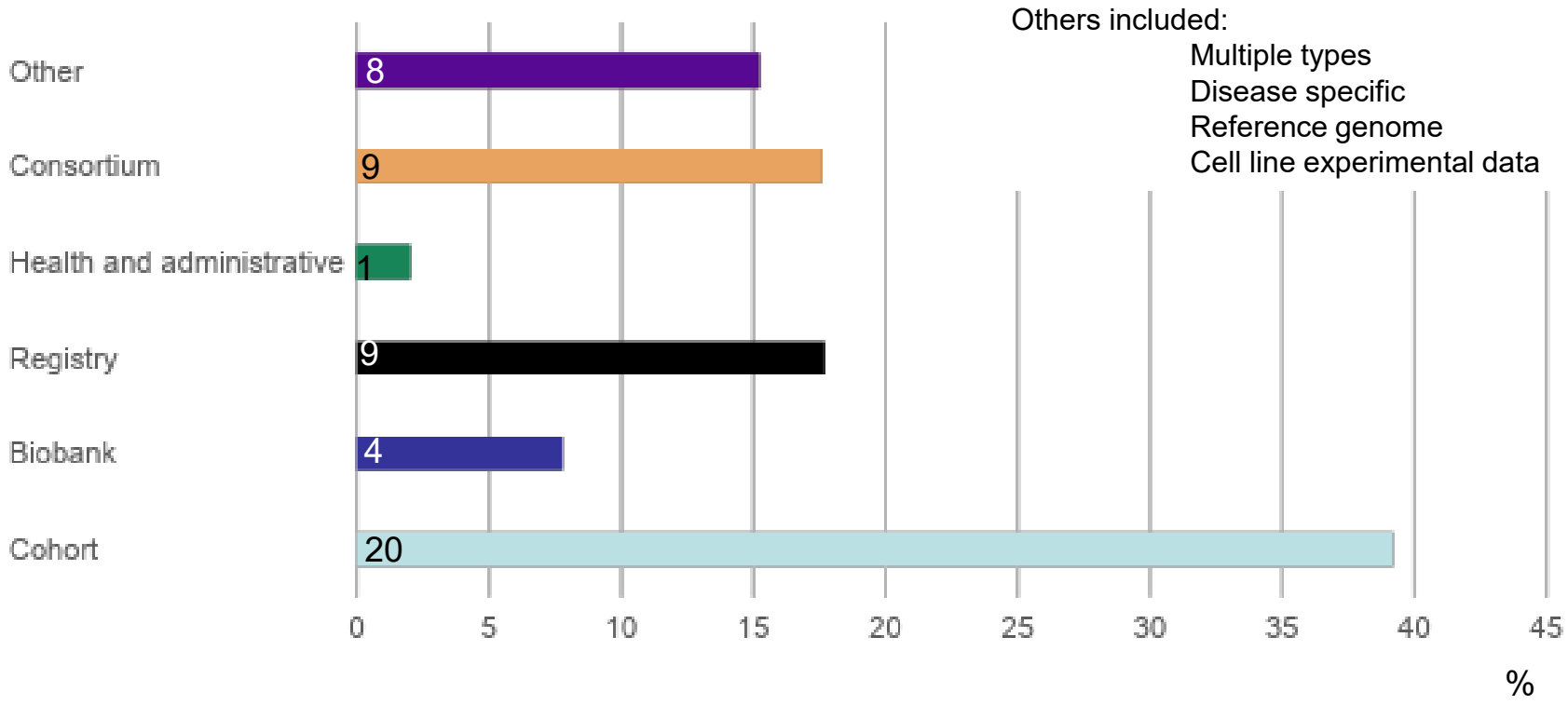

Arshiya Merchant
Elixir


Niklas Blomberg
Elixir


Ingrid Kockum
Karolinska Insitutet

© EU-STANDS4PM - 2022

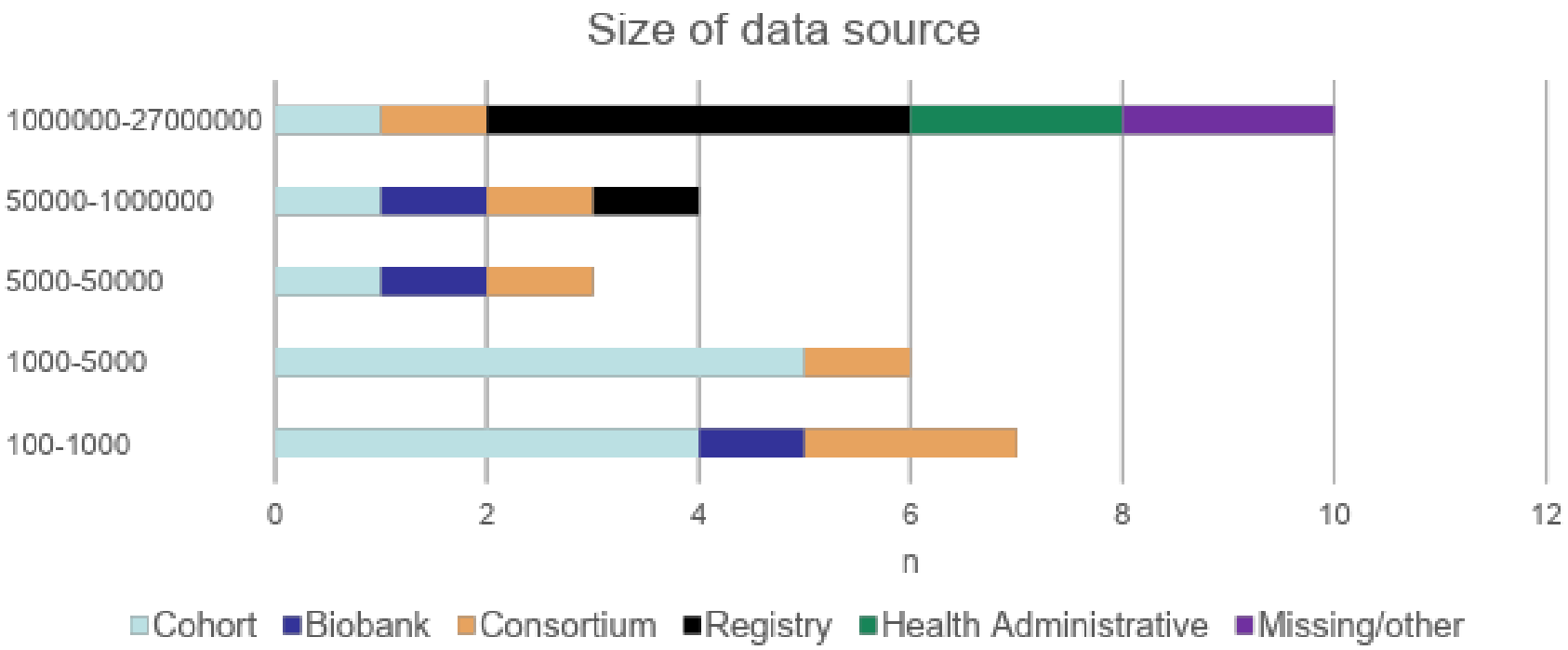
What is the type of dataset?



Most common type of study was cohort study which also includes case-contol studies

© EU-STANDS4PM - 2022

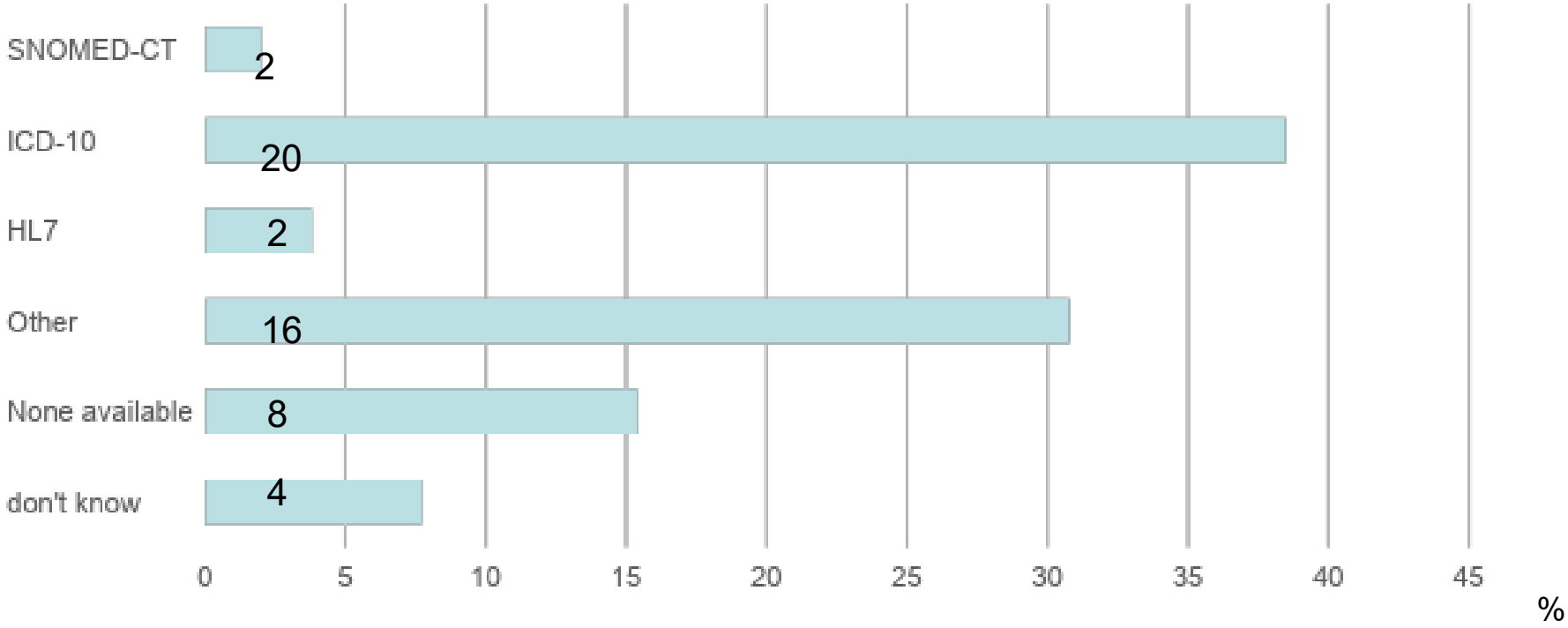
What is the number of individuals in the dataset?



Big variation in the size of the data set covered

© EU-STANDS4PM - 2022

What are the existing standard formats, format guidelines, ontologies etc?

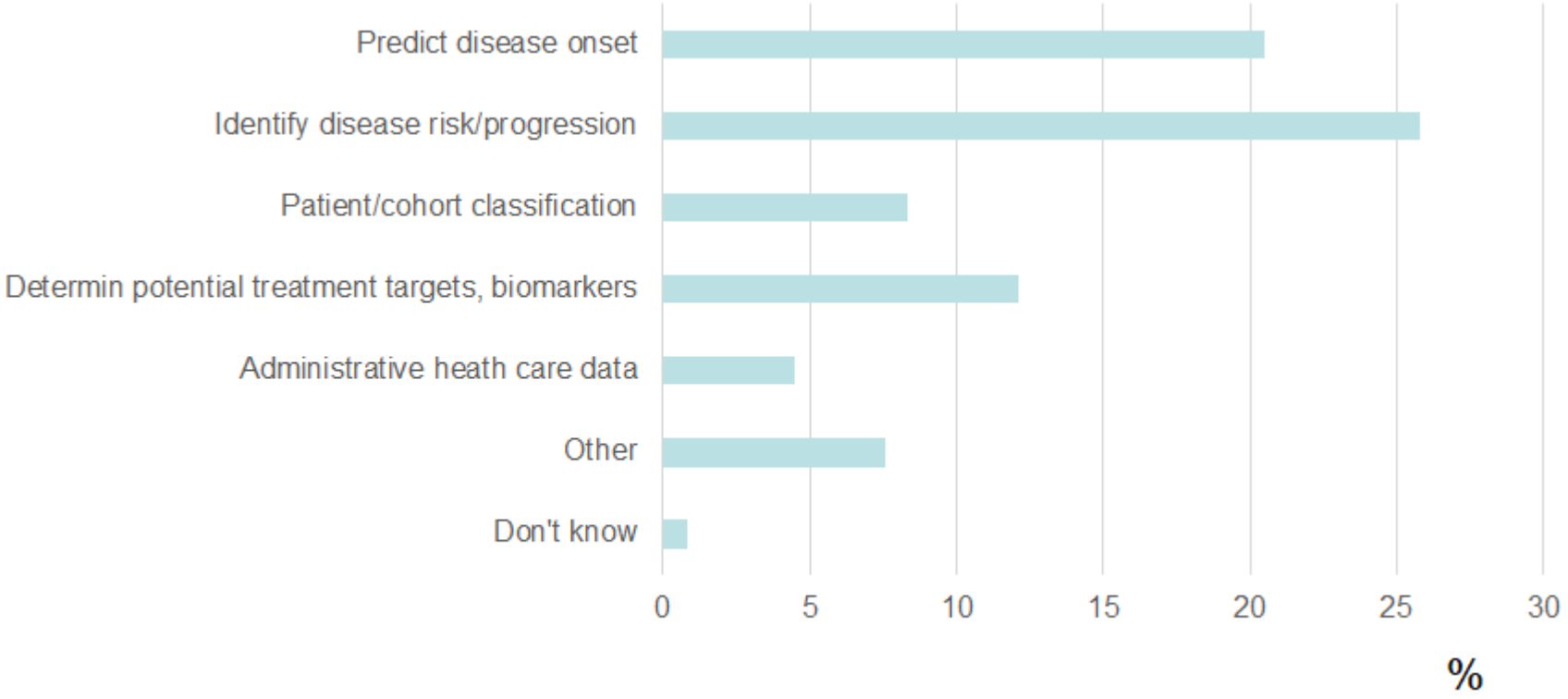


Others include:
Open EHR, Plink, SNPtest, RedCap, ICD9, READ v1-3CT, DM6D, NHS A&E, CLOSER ontology (consortium specific), ATC and NUP codes, bam, vcf, DICOM + contact details to person who knows

Most commonly used standard is ICD10/9. Many do not use any standard.

© EU-STANDS4PM - 2022

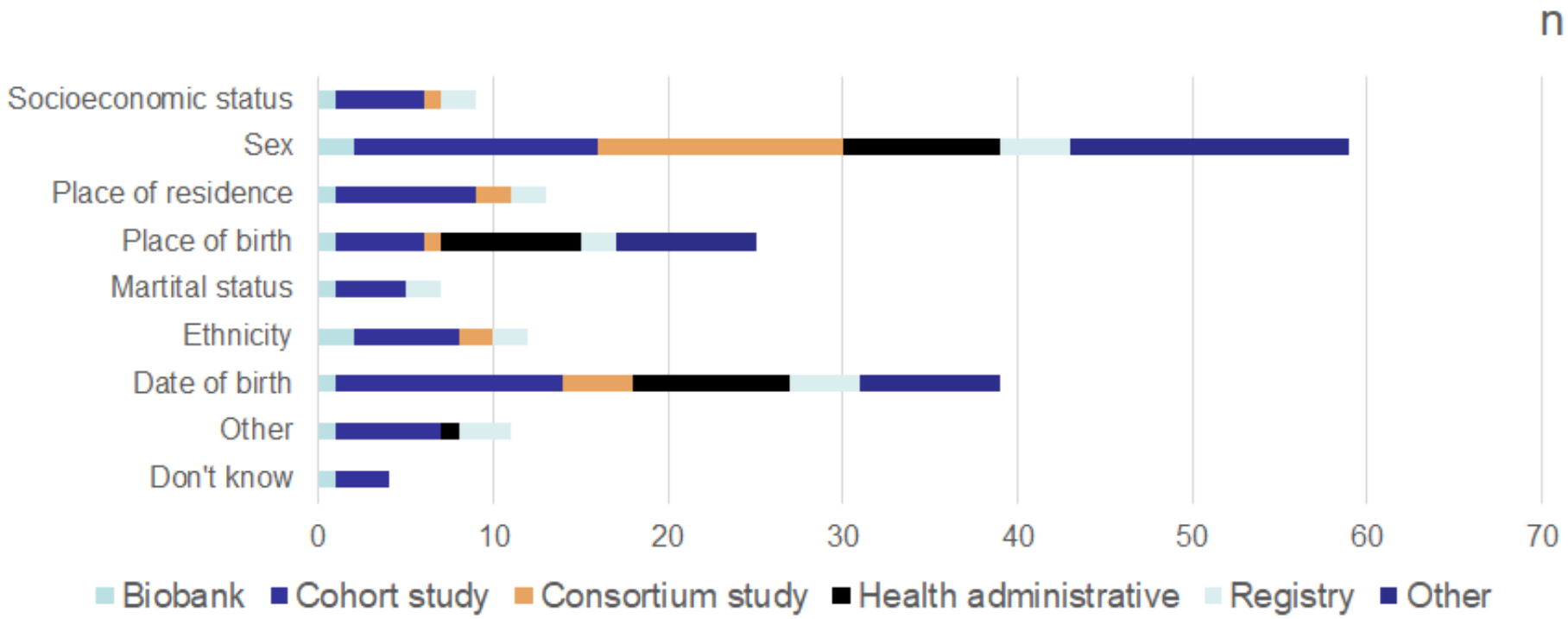
What is the main research question/aim?



Most commonly addressed research question is identification disease risk/progression and disease onset

© EU-STANDS4PM - 2022

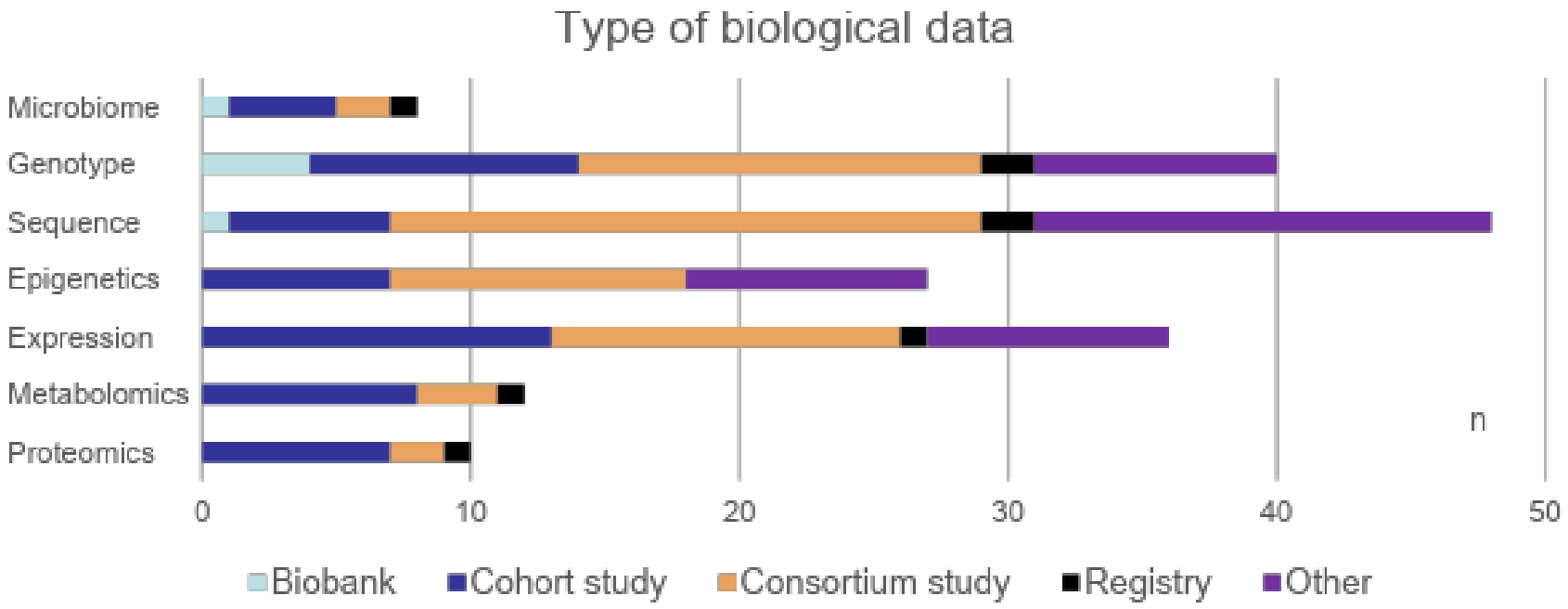
What demographic data is being collected?



Sex is collcetd in almost all datasets, date of birth is also common

© EU-STANDS4PM - 2022

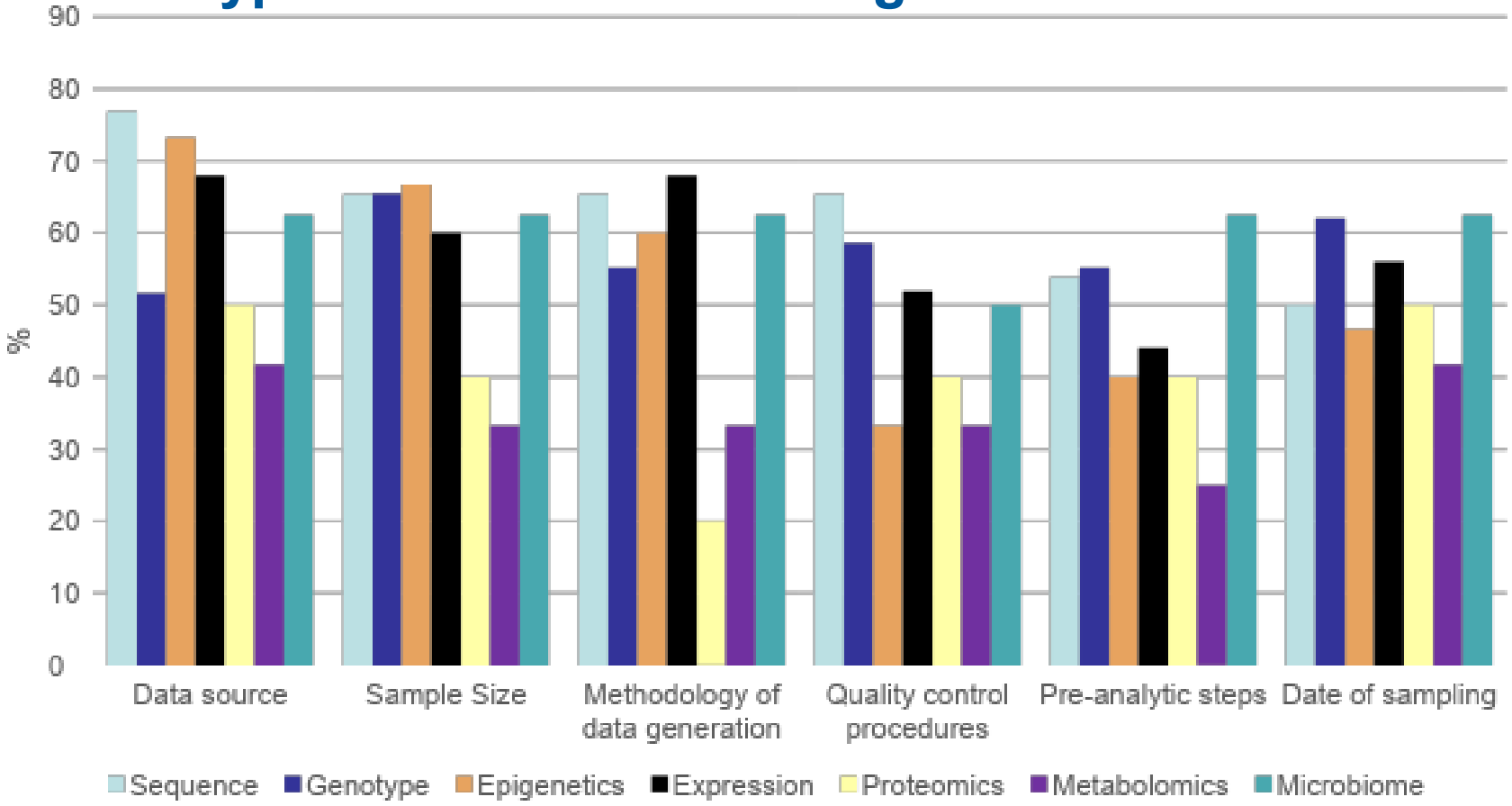
Is biological data being collected? What type?



Genotypes and sequence data most common type of biological data followed by expression and epigenetics
Microbiome, metabolomics and proteomics rarely included

© EU-STANDS4PM - 2022

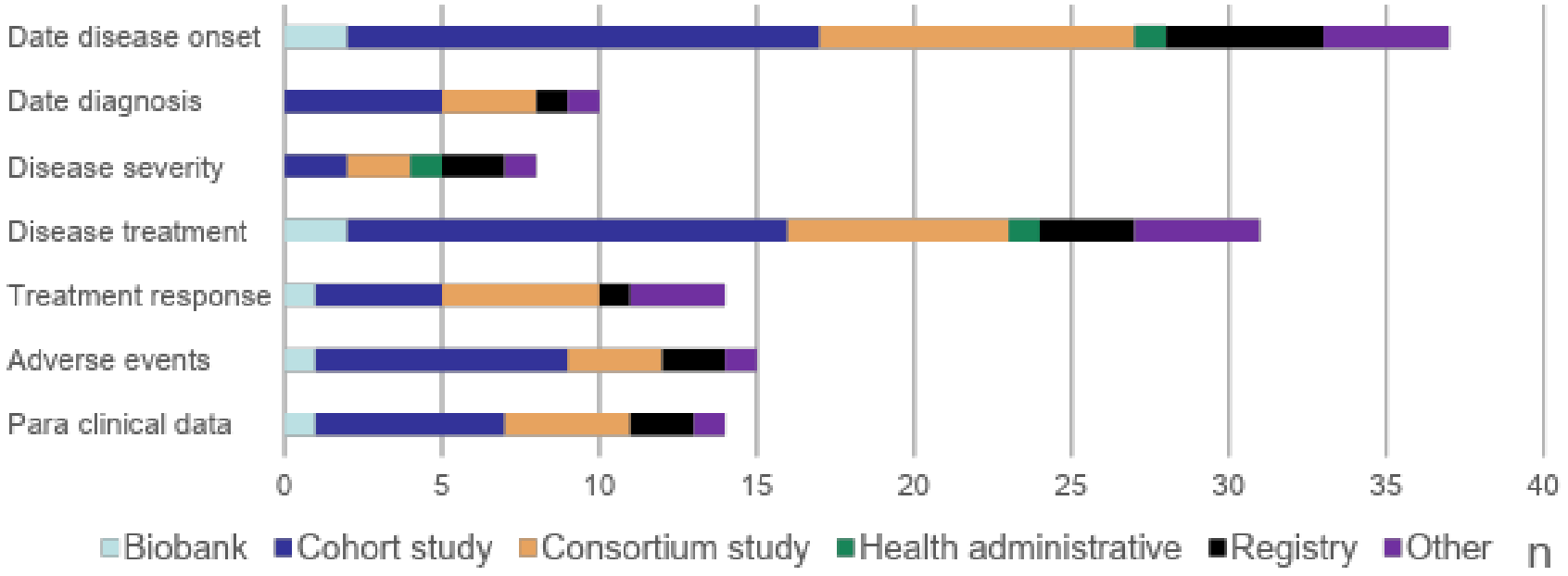
What type of meta data is being collected?



Meta data is not always collected. Date of sampling and pre-analytic steps seems to be especially badly captured

© EU-STANDS4PM - 2022

Disease specific data



Disease onset and treatment fairly frequently captured. Treatment response and adverse events rarely captured

© EU-STANDS4PM - 2022

Conclusions data types and standards

- Large variety in type of datasets/studies
  - Design, size, country of origin
- ICD10 & ICD9 most commonly used standards
- Sex and date of birth most common demographic data
- Metadata for biological data often missing
- Medication not that often captured, ATC standard not that commonly used
- ICD not so often used for hospitalization and co-morbidity data
- Central data for personalised medicine such as response to treatment and adverse events are not capture so often
- **Room for standardization**

© EU-STANDS4PM - 2022

D1.1 EU-wide mapping report with focus on international databases collections and registries

This report will describe and develop:

- A comprehensive catalogue of data resources relevant to personalized medicine.
- Leverage data catalogues developed within the European Open Science Cloud (EOSC) and the related and developing European Health and Innovation Cloud (HRIC).
- A forward looking action plan with recommendations, especially to policymakers and funders that will support responsible and secure data sharing and access to reference data across borders.

© EU-STANDS4PM - 2022

Report: Type of data sources reviewed

- Data Repositories
- Reference Databases
- Case-control data
- Cohort Data
- Biobank Data
- Patient Registries
  - Research or Clinical-Led Registries
  - Patient Powered Registries
- Clinical Trial Registries
- Administrative Health Data
- Adverse events database
- Results from survey

© EU-STANDS4PM - 2022

Report: Identified gaps

- Distinct lack of metadata & pre-processing data that comes along with "core data“
- Little information tracking the treatment regimens patients’ were prescribed, and tracking of their response to these treatments.
- There is a lack of generic ‘terminology’ for the collection, analysis, and interpretation of data.

© EU-STANDS4PM - 2022

Report: Recommendations

- Metadata on pre-processing of data needs to be captured
- Standard terminology should be used in funders documentation to ensure a similar level of understanding and awareness for both policymakers and researchers
- Accessibility to health data from different jurisdictions will promote acceleration of PM research. Clear role for federated EGA for data deposition and access.

© EU-STANDS4PM - 2022

EU-STANDS4PM
standards for in silico models for personalised medicine

1.2. Harmonizing data and metadata standards, minimal information guidelines for reporting and international standardization efforts

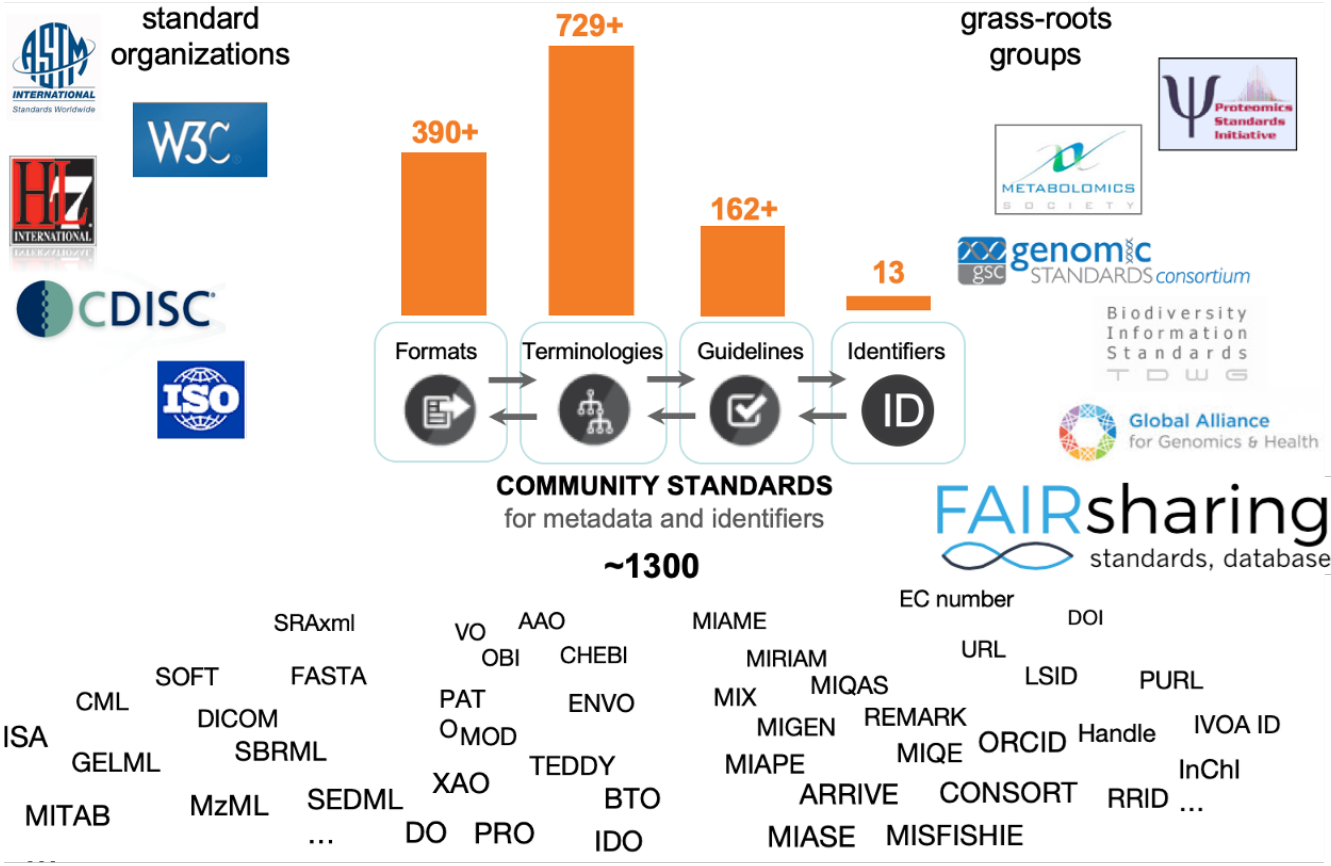
Martin Golebiewski
HITS gGmbH (Heidelberg, Germany)
martin.golebiewski@h-its.org

1.4. Preparations for the development of European level standardization documents with regard to interoperability of health-related data

Heike Moser
DIN German Institute for Standardization
heike.moser@din.de

© EU-STANDS4PM - 2022

The Forest of Standards in Life Sciences



Source: Susanna-Assunta Sansone (University of Oxford, UK)

For modelling standards see review:

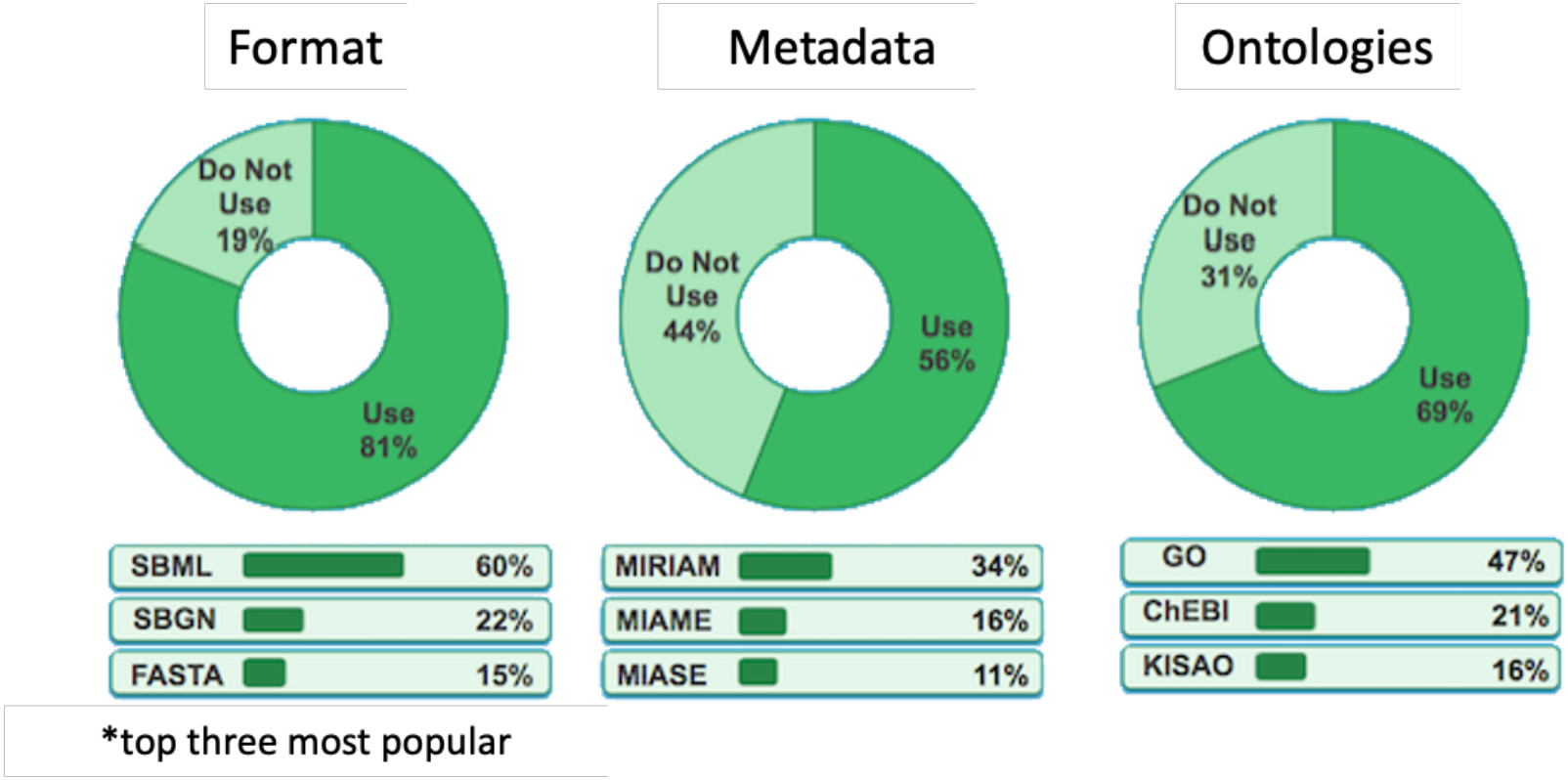
Golebiewski M: **Data Formats for Systems Biology and Quantitative Modeling**,
In: Ranganathan S., Nakai K., Schönbach C. and Gribskov M. (eds.) **Encyclopedia of Bioinformatics and Computational Biology**, Volume 2, 2019, Pages 884-893

Registry for Modeling Standards:

https://normsys.h-its.org

© EU-STANDS4PM - 2022

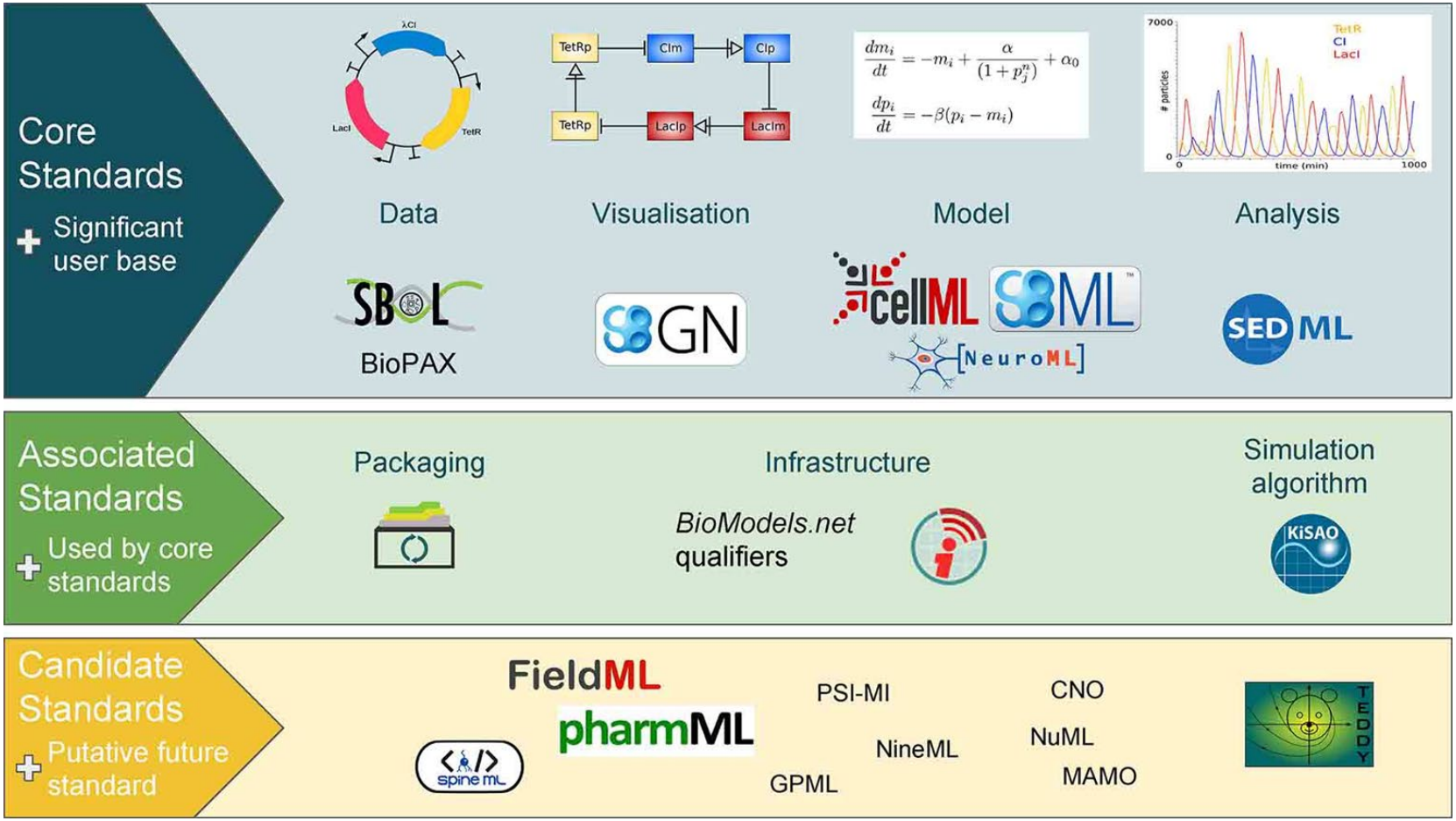
Researchers do not always use data standards



The evolution of standards and data management practices in systems biology (2015). Stanford, Wolstencroft, Golebiewski, *et al.*, **Molecular Systems Biology, 11(12):851**

© EU-STANDS4PM - 2022

COMBINE Community Standards for Computational Modelling in Biology



© EU-STANDS4PM - 2022

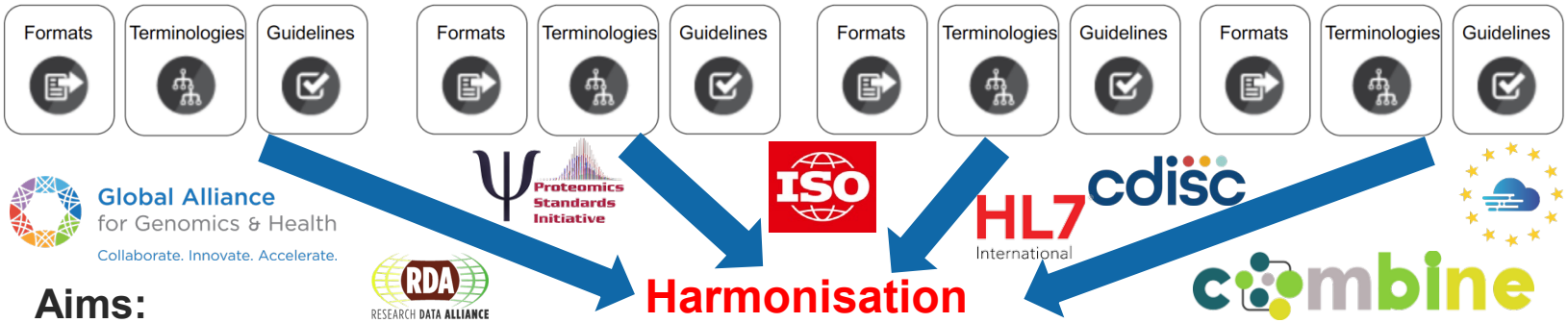
Schreiber F, Sommer B, Czauderna T, Golebiewski M, Gorochowski TE, Hucka M, Keating SM, König M, Myers C, Nickerson D, Waltemath D: **Specifications of standards in systems and synthetic biology: status and developments in 2020.** J Integr Bioinform. (2020) 17(2-3): 20200022. doi: 10.1515/jib-2020-0022

EU-STANDS4PM
standards for in silico models for personalised medicine

T1.2 / T1.4

SYSCID BD2DECIDE LifeCycle MultipleMS SPIDIA4P ... other projects

Analysis: Used formats | Metadata formats | Terminologies | Traceability



Aims:

- Capture, survey and support of bottom-up “grass-roots” standards
- Ensure interoperability for cross-domain and cross-technology data integration
- Guidelines & recommendations for using (meta-)data standards in personalised medicine
- Interface with corresponding international standardisation communities
- Promote the long-term sustainability of the standards together with the European Open Science Cloud (EOSC), ESFRI infrastructures and relevant standardisation committees (ISO/TC 276/WG 5 Data processing and integration & ISO/TC 215 Health informatics, …)

© EU-STANDS4PM - 2022

Collecting requirements from and engaging in relevant scientific standardization communities (T1.2)

- **Collaborations with standardization communities initiated:**
  - GA4GH (Global Alliance for Genomics and Health)
  - HL7 (Health Level Seven International)
  - CDISC (Clinical Data Interchange Standards Consortium)
  - COMBINE (Computational Modeling in Biology Network)
  - ModeleXchange community (is currently forming)
- **Contact to FAIR data communities established:**
  - FAIRsharing.org (resource on data standards, databases and policies)
  - Identifiers.org (registration and resolution service for persistent identifiers)
- **Contribution to COVID-19 Guidelines and Recommendations from RDA (Research Data Alliance):**
  https://doi.org/10.15497/rda00052
  - Recommendations for the sharing of COVID-19 clinical data
  - Reference to EU-STANDS4PM harmonized data access agreements (WP4)
  - Reference to EU-STANDS4PM legal framework document (WP3)

RDA RESEARCH DATA ALLIANCE

© EU-STANDS4PM - 2022

Collecting requirements from and engaging in relevant scientific standardization communities (T1.2)

- **EU-STANDS4PM stakeholder workshop at COMBINE2019** on July 18, 2019 at HITS in Heidelberg (Germany)
  - Co-located with COMBINE 2019 with **>100 attendees from 18 countries**
  - Workshop report published in JIB in 2020 (special issue)
  - COMBINE meeting report also published in JIB in 2020

© EU-STANDS4PM - 2022

D1.2 EU-wide mapping report on good practice examples for integrating phenotype and large scale data

Report will provide:

- **Collection of examples and good practice** of current personalised medicine and systems medicine projects applying data standards
- Overview on relevant **data standards, terminology standards and minimal information guidelines** (with a domain-specific FAIRSharing collection)
- Needs and gaps for **standardizing data input** and **quality of modelling results** as major focus (model itself often is a “black box”)

Report will be based on:

- Outcomes from **EU-STANDS4PM workshop at COMBINE 2019**
- Outcomes from **EU-STANDS4PM workshop: Using patient derived data for in silico modelling in personalized medicine** (February 2020)
- **Deliverable D1.1** and **WP2 white paper**
- **Input from grass-root standardization communities** (e.g. COMBINE, GA4GH Phenopackets, ModeleXchange community, etc.)
- **Experience from SDOs** (HL7, CDISC, ISO/TC 215 and ISO/TC 276, etc.)

© EU-STANDS4PM - 2022

Community Activities in T1.2/T1.4

- Pafticipation in a **EC stakeholder workshop on 'Human Digital Twin'** on November 6th, 2020 (with a short presentation by M. Golebiewski on "Establishing and harmonizing technical standards, and (potential) key performance indicators" that highlighted EU-STANDS4PM recommendations)
- From the key take away messages of the workshop:
  "There is a need for establishing an inclusive ecosystem around Digital Twins for Healthcare, comprising relevant actors to share knowledge, foster collaboration and bring together diverse groups of stakeholders, including patient representatives, academia, research organisations, industry representatives, regulators and evaluators (i.e. HTA institutions, regulatory bodies) and health care payer organisations, involving clinicians in all steps of the development and clinical implementation. The ecosystem's goals are:
  1. to build on and integrate (existing) technologies and disciplines, develop terminologies, establish and set standards;
  2. to facilitate the development of adapted evaluation/assessment tools for developers and regulators including a clear benchmarking framework to integrate digital twins and modelbased approaches for regulatory purposes paving hereby the way for more targeted therapeutics;
  3. to bridge the 'valley of death' for the commercial uptake of digital twins or other in silico approaches;
  4. to ensure proper representation of the end-users' perspective, including regulators, patients, healthcare professionals and payers."

© EU-STANDS4PM - 2022

Community Activities in T1.2/T1.4

- Pafticipation in the **Putting Science into Standards (PSIS) 2021 Workshop 'Organ on Chip: Towards Standardization'** organized by the Joint Research Centre of the European Commission (JRC) and the European Standardization Organizations CEN and CENELEC on April 28th-29th, 2021
- focused on **Organ-on-Chip (OoC) or Micro Physiological Systems (MPS)**, innovative devices that emulate human/animal biology and can reproduce one or more aspects of an organ's functionality
- Session "Standards for Data acquisition and management" chaired by M. Golebiewski" (with introductory presentation, including an overview about relevant EU-STANDS4PM activities)
- From the key messages:
  - OoC/MPS and in silico modelling should go hand-in-hand
  - standardizing workflows, data and models is crucial

© EU-STANDS4PM - 2022

Progress in “formal” standardization (see also WP2)

- EU-STANDS4PM **Strategy document “Development of formal standard documents”** (access via the EU-STANDS4PM intranet)
- **EU-STANDS4PM recognised as liaison organization for ISO/TC 276 Biotechnology and ISO/TC 215 Health Informatics**
  - Easier access to standard drafts and possibility to influence the drafts
- WP1 experts contributed to the **drafting of ISO 20691 “Requirements for data formatting and description in the life sciences for downstream data processing and integration workflows”** (see intranet for contributions)
  - “framework” or “Hub” standard developed by ISO/TC 276/WG5
  - refers to domain-specific standards (in the informative annex)
  - gives researchers a guideline for applying the standards in complex workflows (also for modelling in personalised medicine and related fields)
- **ISO 20691 will be submitted next week by ISO/TC 276** as ISO Draft International Standard (DIS)
- **EU-STANDS4PM draft for an ISO Technical Specification (TS)** submitted this week to ISO/TC 276/WG 5 Data Processing and Integration (see WP2)

© EU-STANDS4PM - 2022

ISO 20691
Requirements for data formatting and description in the life sciences

Foreword

Introduction

1 Scope

2 Normative references

3 Terms and definitions

4 Criteria for formats and identifiers

5 Technical criteria and requirements

6 Semantic criteria and requirements

7 Requirements for ontologies suitable for annotation of biological data

8 Requirements for domain specific data standards

9 Requirements for data repositories for biological data

Annex A (informative) Recommended formats for life science data

Annex B (informative) Minimal reporting standards for data, models and metadata

ISO/TC 276 Biotechnology WG 5 (Data Processing and Integration) works on a draft for a new ISO guideline standard for data in the life sciences:

Reference framework („hub“) standard for (non-ISO) community standards

- Requirements and rules for the concerted application of community standards for formatting, description and documentation of datatypes in the life sciences
- Catalogue of criteria and requirements for interoperable life science data formats and semantic data description standards

› **will be submitted next week by ISO/TC 276**
as ISO Draft International Standard (DIS)

© EU-STANDS4PM - 2022

Summary

© EU-STANDS4PM - 2022

›Relevant data landscape is highly complex - many actors, wide range of standards and access models. Opportunity: large datasets avaliable for reuse.

> Use of networking for describing requirements (e.g. experiences from H2020 projects, co-located workshop at COMBINE 2019 in Heidelberg, community survey, etc.)

›Analysis of existing standards based on the requirements defined by experts from WP 1 (applicable, changed applicable, not available)

›Contact with standardization bodies (ISO, CEN, DIN...) and SDOs (GA4GH, HL7, CDISC, COMBINE) when standards need to be changed in order to adapt them

›Help to increase interoperability of data (and data standards) for complex workflows in personalized medicine

›Check whether there are bodies that can standardize the new standardization topics - if necessary, initiate new Working groups

›Create drafts (or contribute to drafting) for new or changing standards

›Publishing and advertising of the new/changed standards using the existing network

© EU-STANDS4PM - 2022

Acknowledgements

EU-STANDS4PM is funded by the European Union Horizon2020 framework programme of the European Commission, Directorate-General for Research and Innovation under Grant Agreement # 825843.

© EU-STANDS4PM - 2022

EU-STANDS4PM
standards for in silico models for personalised medicine

A European standardization framework for data integration and data-driven *in-silico* models for personalised medicine

Benefits of Workflow Standardization in IVD Testing

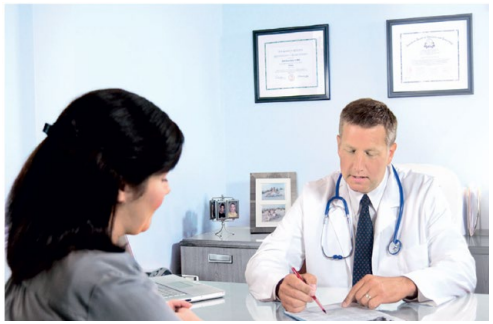
Dr. Kalle Günther
QIAGEN GmbH
kalle.guenther@qiagen.com

- 150,000 papers documenting thousands of claimed biomarkers, but fewer than 100 have been validated for routine clinical practice

Bring on the biomarkers, George Poste, Nature 2011

- Diagnostic errors cause about 10% of all patient deaths and about 17% of adverse events

Institute of Medicine (IOM) Report Sept. 2015



- The pre-analytical phase accounts for 46% to 68% of errors observed during the total testing process

Medical Laboratory Observer, May 2014

- Unnecessary expenditure caused by pre-analytical errors in a typical U.S. hospital (~ 650 beds) of ~ $1.2 million per year

Green SF. Clin Biochem. 2013

- Irreproducible preclinical research exceeds 50%, US $28B / year spent on preclinical research that is not reproducible - in the US alone

Freedman LP, Cockburn IM, Simcoe TS (2015) PLoS Biol 13(6): e1002165.doi:10.1371/journal.pbio.1002165

❾ Demand for Improvements and Workflow Standardization

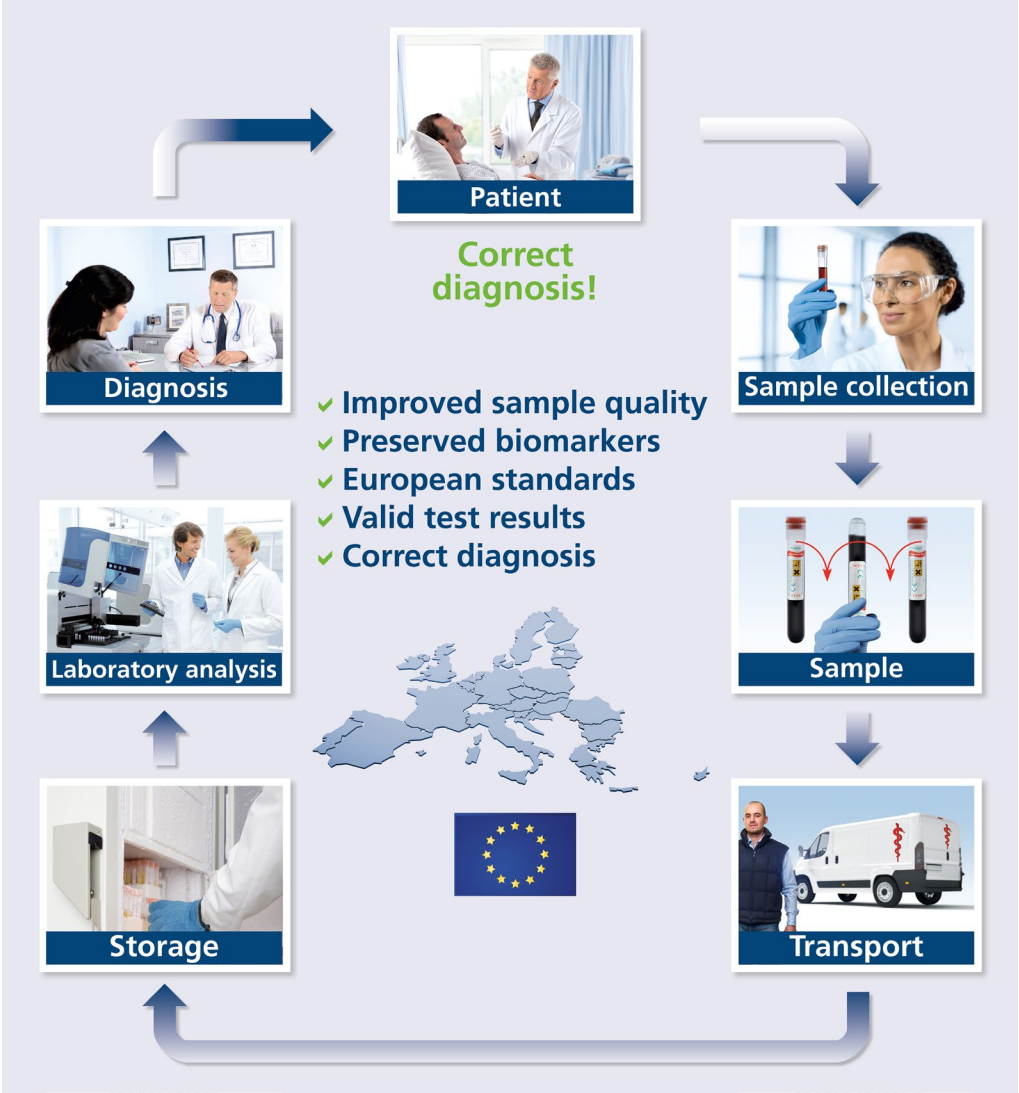
Deficiencies in Routine Healthcare and Research

Data collection:

- Storage temperature?
- Temperature cycling?
- Storage period?

❾ No information about important pre-analytical workflow parameters

An Analytical Test Result is the Result of an Entire Workflow



New In Vitro Diagnostic Regulations 2017 (IVDR) in Europe

- Pre-analytical workflow parameters in several sections
  - 6. PRODUCT VERIFICATION AND VALIDATION (Annex II)
  - 6.1. Information on analytical performance of the device
  - 6.1.1. Specimen type

This Section shall describe the different specimen types that can be analysed, including their stability such as storage, where applicable specimen transport conditions and, with a view to time-critical analysis methods, information on the timeframe between taking the specimen and its analysis and storage conditions
such as duration, temperature limits and freeze/thaw cycles

- ➔ European legislation requires to specify, develop, verify and validate the pre-examination phase for a dedicated examination development and regulatory approval
- ➔ Standardization of examination workflows is becoming now essential from that perspective

Role of Legislation, Standards and Technologies

New EU IVDR – in-vitro Diagnostic Device Regulation 2017 in Europe

Pre-analytical workflow parameters

22 new pre-analytical EN ISO & CEN Standards

SOPs

Technologies & Products

1) The SPIDIA4P project has received funding from the European Union´s Horizon 2020 research and innovation programme under grant agreement no. 733112.

Example: ISO 20186-1 Venous whole blood - Part 1: Isolated cellular RNA

RNA

Licensed to BBMRI-ERIC / Viridiana Beltran (viridiana.beltran.venegas@bbmri-eric.eu)
ISO Store Order: OP-370461 / Downloaded: 2019-05-27
Single user licence only, copying and networking prohibited.

ISO 20186-1:2019(E)

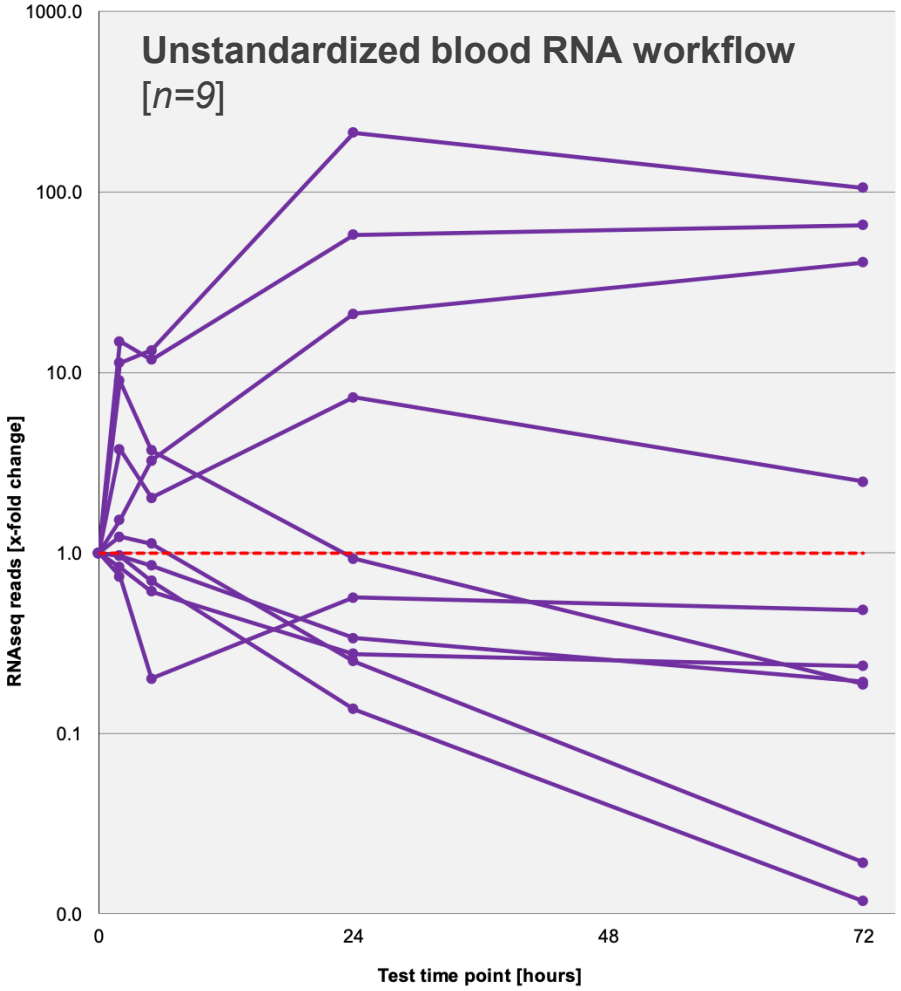
Contents Page

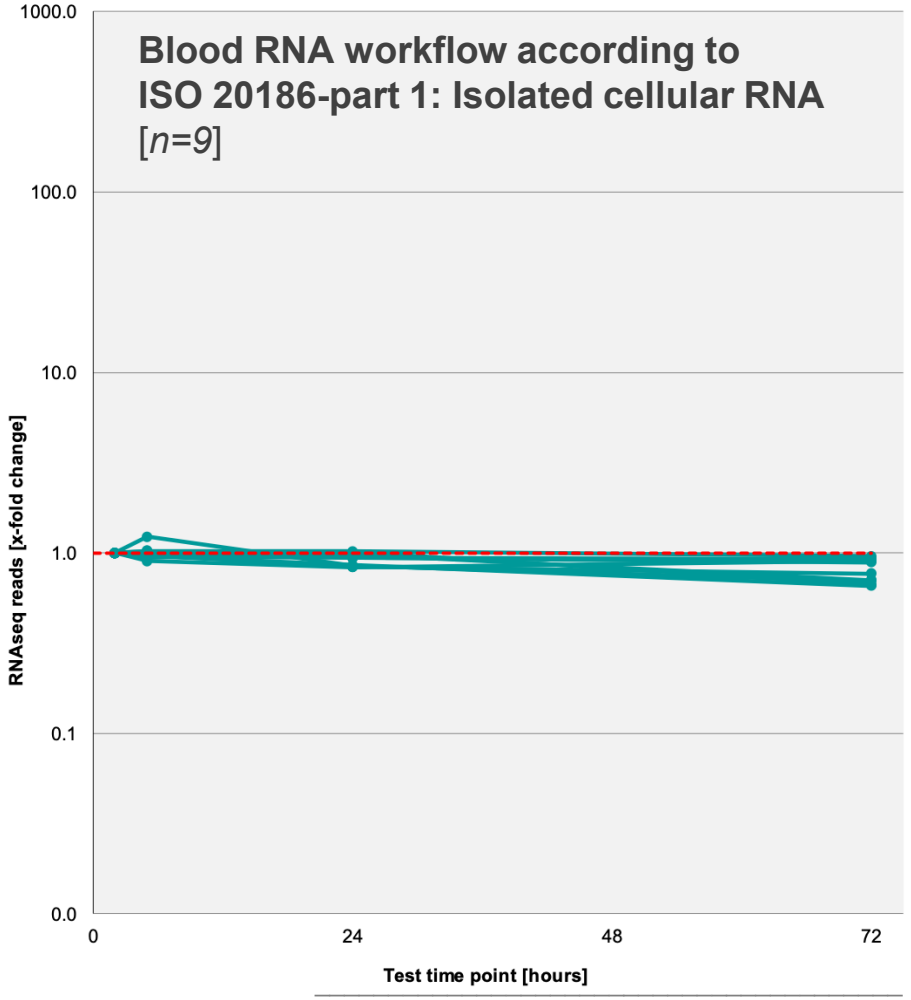
© ISO 2019 – All rights reserved

SOPs

Data collection at each pre-examination step!

Reliable results due to workflow standardization (10 mRNAs)





Unpublished data by K. Günther, QIAGEN GmbH

Acknowledgements

EU-STANDS4PM is funded by the European Union Horizon2020 framework programme of the European Commission, Directorate-General for Research and Innovation under Grant Agreement # 825843.

Work Package 2

Integrative data analysis and in silico models in personalized medicine

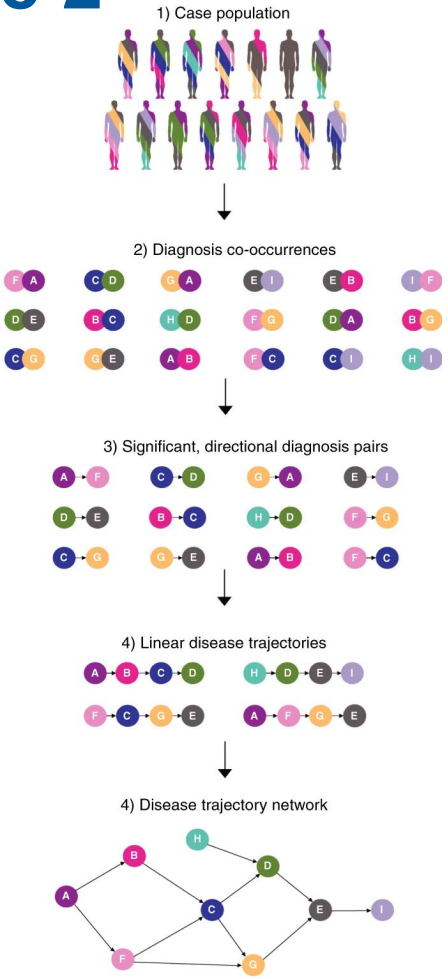
Review article

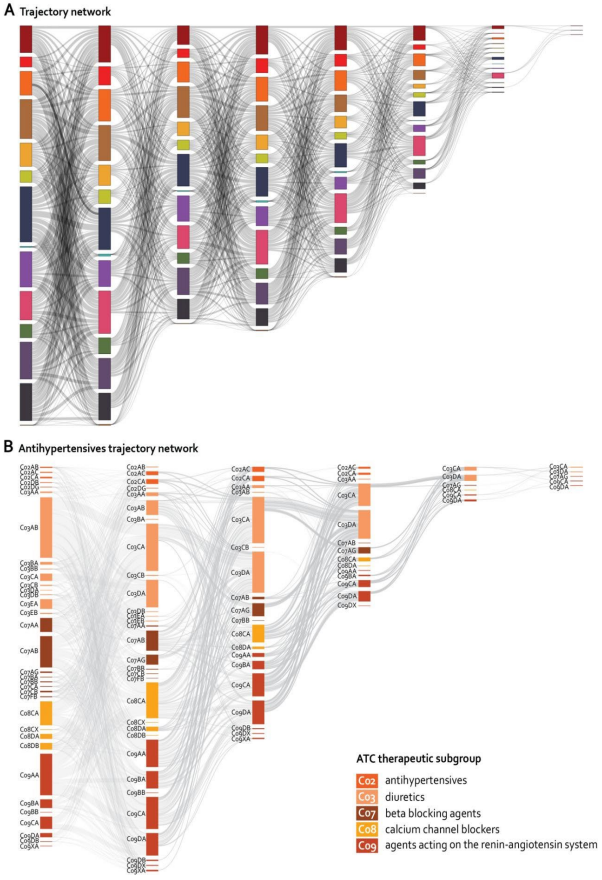
Computational Models for Clinical Applications in Personalized Medicine: Guidelines and Recommendations for Data Integration and Model Validation

J. Pers. Med. 2022, 12(2), 166; https://doi.org/10.3390/jpm12020166

© EU-STANDS4PM - 2022

Work Package 2
Trajectory





Aguayo-Orozco, A., Haue, A.D., Jørgensen, I.F. et al. Optimizing drug selection from a prescription trajectory of one patient. npj Digit. Med. 4, 150 (2021). https://doi.org/10.1038/s41746-021-00522-4

© EU-STANDS4PM - 2022

Siggaard, T., Reguant, R., Jørgensen, I.F. *et al.* Disease trajectory browser for exploring temporal, population-wide disease progression patterns in 7.2 million Danish patients. *Nat Commun* **11,** 4952 (2020). https://doi.org/10.1038/s41467-020-18682-4

WP 2 trajectory: Integrative data analysis and in silico models for personalised medicine

© EU-STANDS4PM - 2022

Work flow and process of WP2: in silico models

Interaction with the community:
Data collection:
Survey

Interaction with the community“Case Study” workshop on models for Personalised Medicine (HiTS) *Analysing models in use cases*

White Paper: Analysis & Recommendations
Requirements for data input and model validation

Review article Computational Models for Clinical Applications in Personalized Medicine—Guidelines and Recommendations for Data Integration and Model Validation

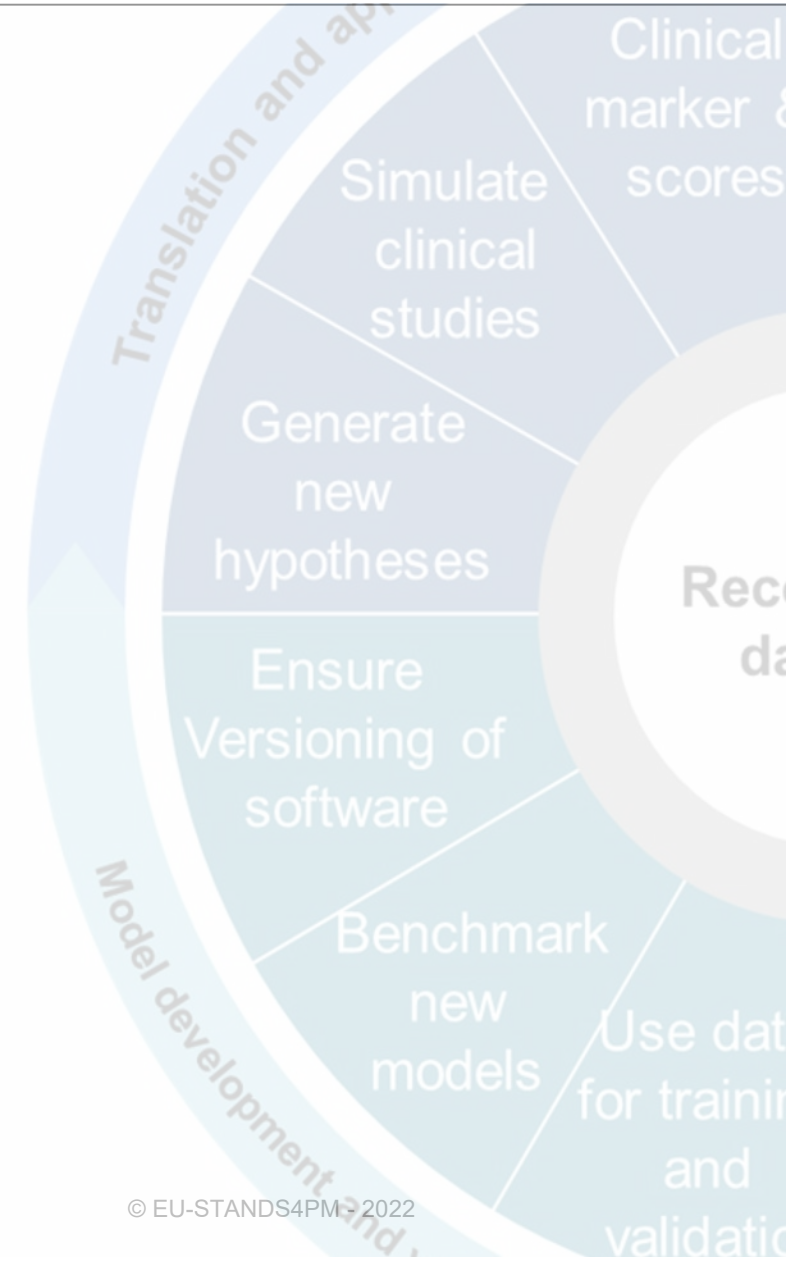
ISO/TS

Focus area - Interaction with the community“ - workshop on explainabilty in models for Personalised Medicine

Focus area - Article on XAI

© EU-STANDS4PM - 2022

EU-STANDS4PM
standards for in silico models for personalised medicine



© EU-STANDS4PM - 2022

Journal of *Personalized Medicine*

MDPI

Review

Computational Models for Clinical Applications in Personalized Medicine—Guidelines and Recommendations for Data Integration and Model Validation

Catherine Bjerre Collin [1], Tom Gebhardt [2], Martin Golebiewski [3], Tugce Karaderi [1,4], Maximilian Hillemanns [2], Faiz Muhammad Khan [2], Ali Salehzadeh-Yazdi [5], Marc Kirschner [6], Sylvia Krobitsch [6], EU-STANDS4PM consortium † and Lars Kuepfer [7,*]

1 Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 N Copenhagen, Denmark; catherine.bjerre.collin@cpr.ku.dk (C.B.C.); tugce.karaderi@sund.ku.dk (T.K.)
2 Department of Systems Biology and Bioinformatics, University of Rostock, 18057 Rostock, Germany; tom.gebhardt@uni-rostock.de (T.G.); maximilian.hillemanns@uni-rostock.de (M.H.); faiz.khan3@uni-rostock.de (F.M.K.)
3 Heidelberg Institute for Theoretical Studies gGmbH, 69118 Heidelberg, Germany; martin.golebiewski@h-its.org
4 Center for Health Data Science, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 N Copenhagen, Denmark
5 Max-Planck-Institute for Multidisciplinary Sciences, 37077 Göttingen, Germany; ali.salehzadeh-yazdi@mpinat.mpg.de
6 Forschungszentrum Jülich GmbH, Project Management Jülich, 52425 Jülich, Germany; m.kirschner@fz-juelich.de (M.K.); s.krobitsch@fz-juelich.de (S.K.)
7 Institute for Systems Medicine with Focus on Organ Interaction, University Hospital RWTH Aachen, 52074 Aachen, Germany
* Correspondence: lkuepfer@ukaachen.de; Tel.: +49-241-8085900
† Coordinator contact of EU-STANDS4PM; Membership of EU-STANDS4PM consortium is provided in the acknowledgements.

check for updates

Citation: Collin, C.B.; Gebhardt, T.; Golebiewski, M.; Karaderi, T.; Hillemanns, M.; Khan, F.M.; Salehzadeh-Yazdi, A.; Kirschner, M.; Krobitsch, S.; EU-STANDS4PM consortium; et al. Computational Models for Clinical Applications in Personalized Medicine—Guidelines and Recommendations for Data Integration and Model Validation. *J. Pers. Med.* **2022**, *12*, 166. https://doi.org/10.3390/jpm12020166

Academic Editor: William J. Duddy

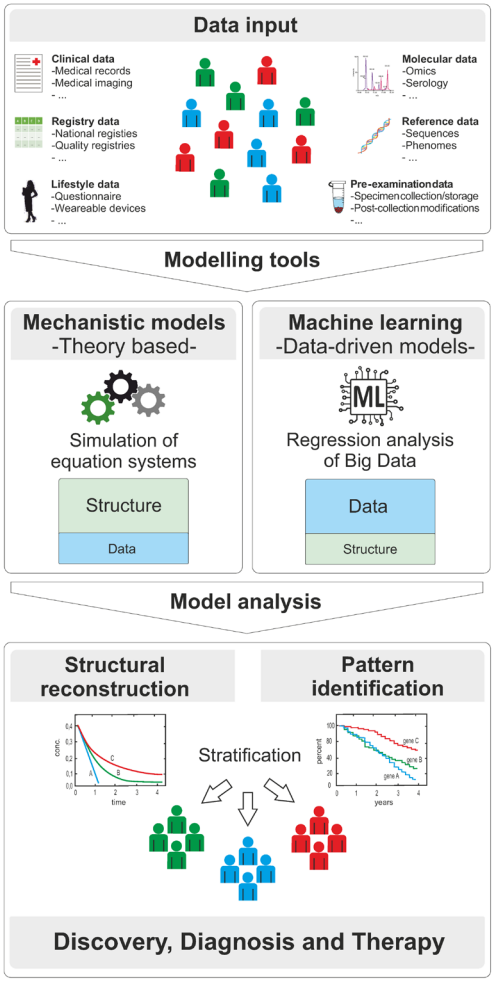
Received: 19 November 2021
Accepted: 20 January 2022
Published: 26 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: The future development of personalized medicine depends on a vast exchange of data from different sources, as well as harmonized integrative analysis of large-scale clinical health and sample data. Computational-modelling approaches play a key role in the analysis of the underlying molecular processes and pathways that characterize human biology, but they also lead to a more profound understanding of the mechanisms and factors that drive diseases; hence, they allow personalized treatment strategies that are guided by central clinical questions. However, despite the growing popularity of computational-modelling approaches in different stakeholder communities, there are still many hurdles to overcome for their clinical routine implementation in the future. Especially the integration of heterogeneous data from multiple sources and types are challenging tasks that require clear guidelines that also have to comply with high ethical and legal standards. Here, we discuss the most relevant computational models for personalized medicine in detail that can be considered as best-practice guidelines for application in clinical care. We define specific challenges and provide applicable guidelines and recommendations for study design, data acquisition, and operation as well as for model validation and clinical translation and other research areas.

Data-driven or theory-based

- Mechanistic models: represent governing physiological processes: functional understanding of underlying mechanisms
- Data-driven (ML, DL, AI) aim for knowledge discovery, require large data sets and do not require prior functional understanding – pattern observation

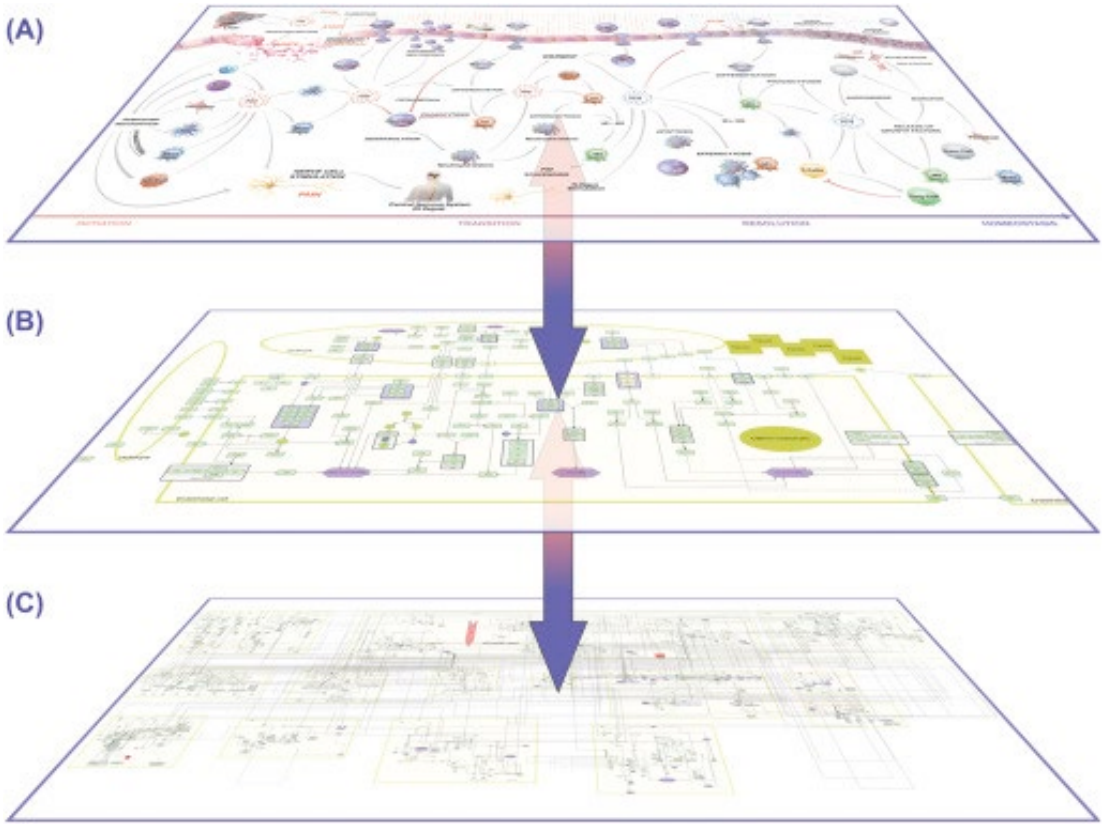


J. Pers. Med. 2022, 12(2), 166; https://doi.org/10.3390/jpm12020166

© EU-STANDS4PM - 2022

Discovery through mechanistic models

Example: MIM Molecular interaction map:
map on inflammation resolution provides functionality to visualize Omics data and allows making hypotheses on the role of connected molecules in a disease phenotype
structure the growing knowledge of the field in a comprehensible manner.



Fujita, K.A., Ostaszewski, M., Matsuoka, Y. et al. Integrating Pathways of Parkinson's Disease in a Molecular Interaction Map. Mol Neurobiol 49, 88–102 (2014)

© EU-STANDS4PM - 2022

Modelling Approaches for Clinical Applications in Personalized Medicine

Mechanistic Models
The aim of a mechanistic model is to functionally understand, examine, and predict the emergent properties of individual components of a biological system and the manner in which they are coupled.

Previously established concepts range from static molecular interaction maps and constraint based modelling to qualitative logic-based models to more detailed quantitative kinetic models. The choice of a model formalism depends on the availability of data, the type of research question and the size and structure of the system.

Molecular interaction maps (MIMs) are static models that depict the physical and causal interactions among biological species in the form of networks

Constraint-Based Models
Constraint-based models, such as GEnome-scale Metabolic models (GEM), provide a mathematical framework to gaining an understanding of metabolic capacities of a cell, enabling system-wide analysis of genetic perturbations, exploring metabolic diseases, and finding the essential enzymatic reactions as well as drug targets
Boolean Models
Boolean modelling (BM) is the simplest form of logic-based models where nodes (e.g., a gene, protein, a transcription factor, or microRNA, etc.) are described by one of two possible states

Quantitative Models
Quantitative modeling, such as the ordinary differential equations (ODEs)-based approach, quantitatively analyses the behavior of a biochemical reaction over time.

Pharmacokinetic models are a particular application of ODE models that describe the concentration of a drug in plasma or different tissues

Software Resources and Tools
In the following, we provide a list of widely used resources and tools for the construction, visualization, and simulation of MIMs, including qualitative and quantitative models and pharmacokinetic models

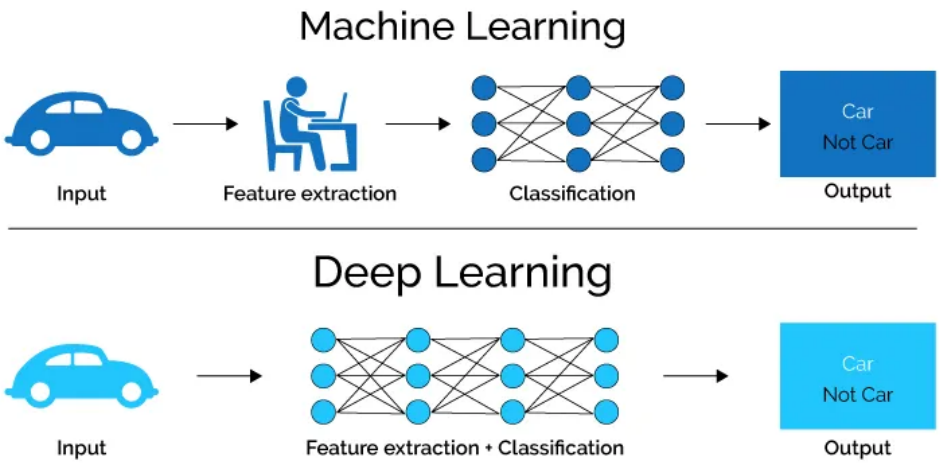
© EU-STANDS4PM - 2022

Machine learning

Data-driven approaches treat the causal mechanism as unknown and aim to model a function that operates on large-scale data input to predict the outcome, regardless of the unknown physiological processes.

'Machine learning' refers broadly to the process of fitting predictive models to data or of identifying informative groupings within data.

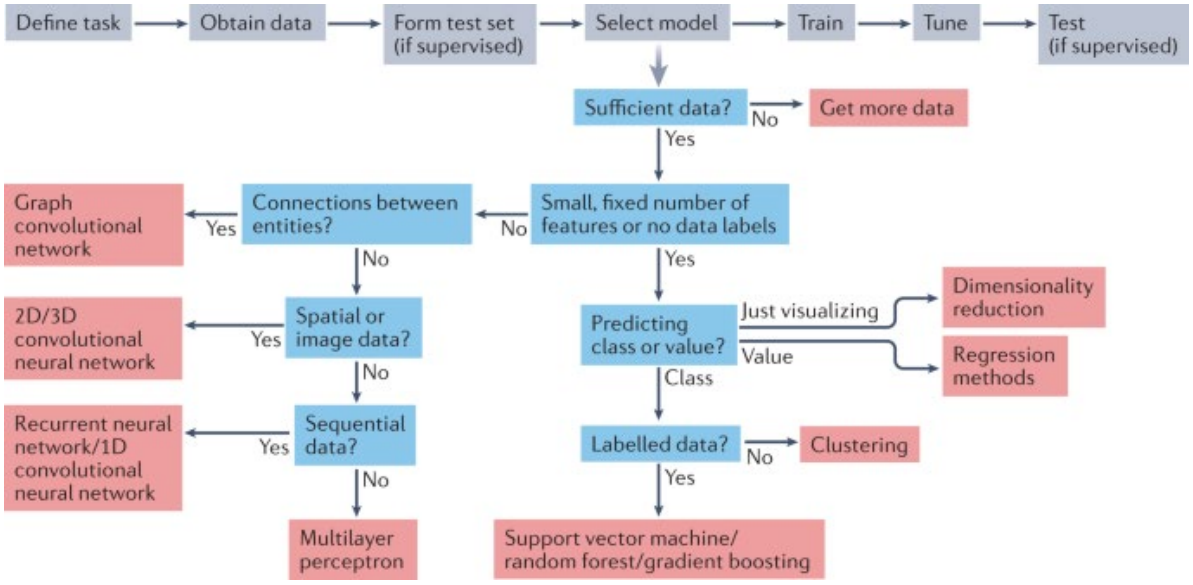
Machine learning is particularly useful when the dataset one wishes to analyse is too large (many individual data points) or too complex (contains a large number of features) for human analysis



© EU-STANDS4PM - 2022

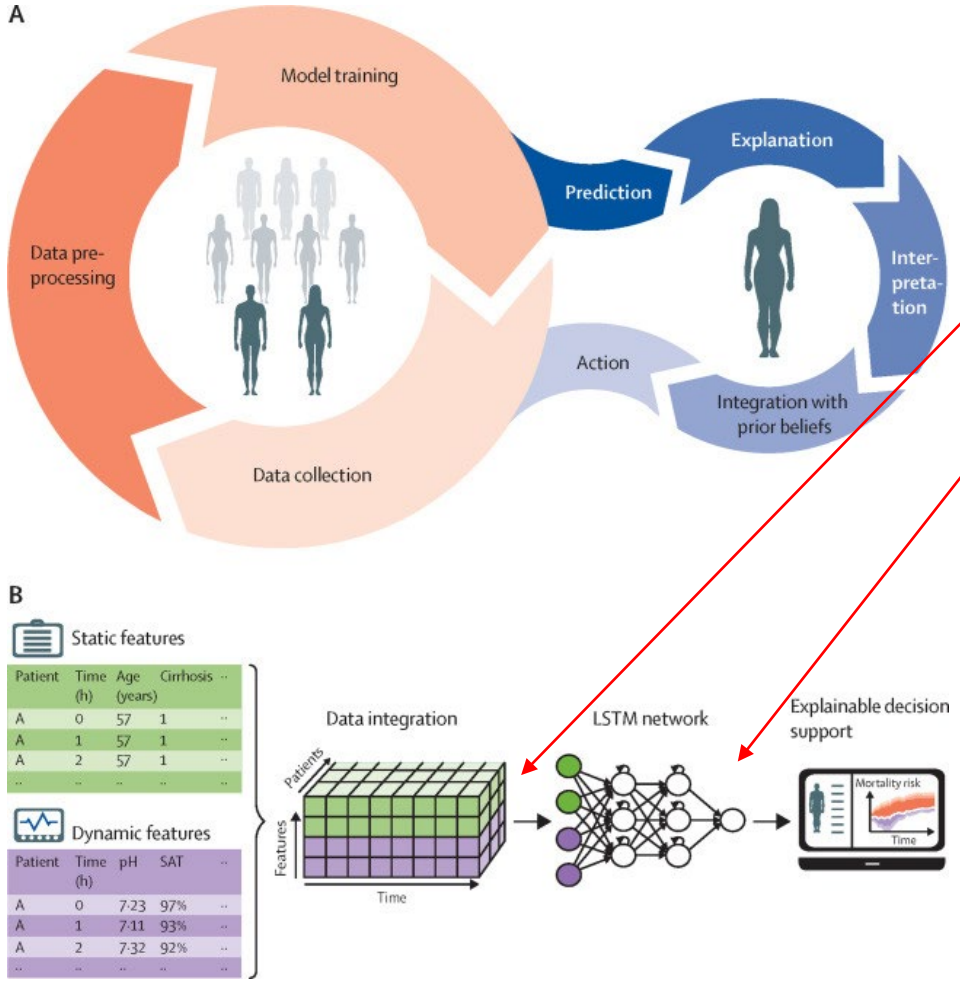
Traditional machine learning

- Deep learning
- Artificial neural networks
- Supervised learning – data is labelled
- Unsupervised – patterns in unlabelled data
- Semi-supervised – where labelled data are rare
- Classification, regression and clustering
- Traditional machine learning
- Deep learning
- Artificial neural networks
- Supervised learning – data is labelled
- Unsupervised – patterns in unlabelled data
- Semi-supervised – where labelled data are rare
- Classification, regression and clustering



© EU-STANDS4PM - 2022

WP1-2 Workshop: Using patient derived data for in silico modelling in personalized medicine: Outcome - Focus of standardization for modelling



Standardization of model input (data) and model output / validation

modified from:
Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records, Thorsen-Meyer, Brunak et al., Lancet Digital 2020

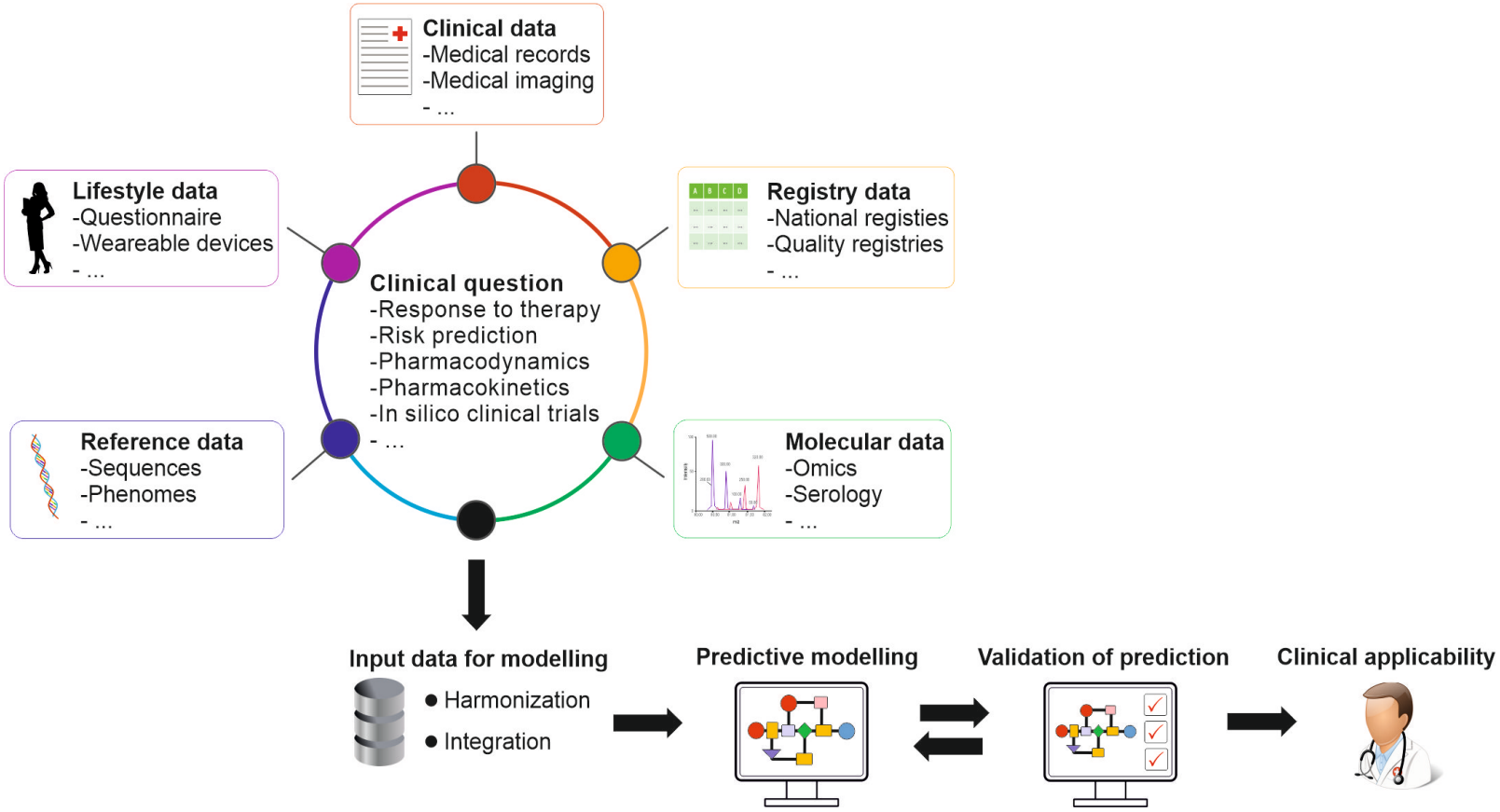
© EU-STANDS4PM - 2022

Review article: computational models for clinical applications in PM – Guidelines/recommendations
- data integration
- model validation

- Most relevant computational models for PM
- Best-practice guidelines
- Complex heterogeneous data
- Computational models functional understanding

© EU-STANDS4PM - 2022

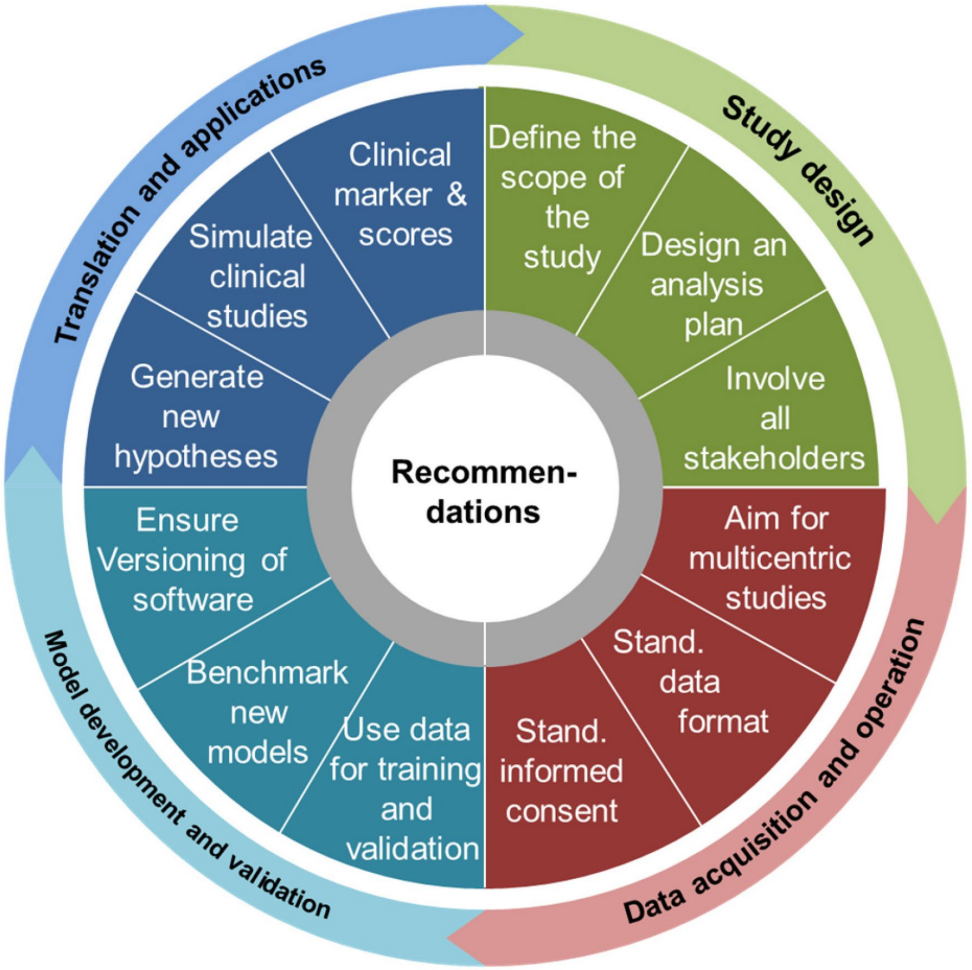
Modelling workflow: from and to the clinic



© Adapted courtesy of Marc Kirschner

© EU-STANDS4PM - 2022

Recommendations



© EU-STANDS4PM - 2022

Conclusions

There are many successful examples for the application of computational models in discovery, diagnosis, and therapy. However, several challenges remain to fully realize the possibilities of personalized data in clinical practice, in particular regarding data provision, model building, and model filing as well as legal issues and ethics.

To support successful study outcomes:

- Careful planning of study design
- Common standards for data sampling, data acquisition, and data operation
- Data harmonization
- Data should be divided for training and validation;
- Model documentation should be written according to best practice guidelines;
- It is important to openly communicate model assumptions and biases in the computational results;
- New patient data should be continuously used for benchmarking of the computational results.

© EU-STANDS4PM - 2022

Acknowledgements

EU-STANDS4PM is funded by the European Union Horizon2020 framework programme of the European Commission, Directorate-General for Research and Innovation under Grant Agreement # 825843.

© EU-STANDS4PM - 2022

EU-STANDS4PM
standards for in silico models for personalised medicine

A European standardization framework for data integration and data-driven *in-silico* models for personalised medicine

WP3: Guidance on the legal, ethical and policy considerations arising from the use of in silico modelling for personalized medicine

Miranda Mourby OU, Eugenijus Gefenas, Vilma Lukaseviciene & Jurate Lekstutiene VU,
Mette Hartlev & Katharina Ò Cathaoir UCPH,

© EU-STANDS4PM - 2022

AIM of WP3

Overall aim of WP3 is – in close collaboration with the other WPs:

- provide guidance on the legal, ethical and policy considerations arising from the use of in silico modelling for personalized medicine,
- identify and analyse legal and ethical issues
- ensure that recommendations for standardization are legally and ethically sound

© EU-STANDS4PM - 2022

WP3 team

Miranda Mourby,
Oxford University
Researcher in Law & barrister
Specialized in privacy and data protection law

Mette Hartlev
Copenhagen University
Professor of Health Law,
Chair of National Committee on Health Research Ethics
Specialized in patients rights and law, science and technology studies

Katharina Ò Cathaoir
Copenhagen University
Assistant professor of Health Law,
Specialized in patients rights & health and human rights law

Eugenijus Gefenas
Vilnius University
Professor and director of the *Centre for Health Ethics, Law and History*
Chair of European network of research ethics committees (EUREC)

Vilma Lukaseviciene
Vilnius University
Lecturer of Bioethics
Specialized in clinical trials regulation

Jurate Lekstutiene
Vilnius University
Lecturer of Bioethics
Specialized in clinical trials regulation

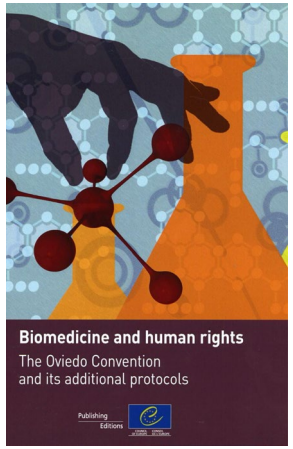
© EU-STANDS4PM - 2022

Data integration for development of in silico models - Legal and ethical aspects

- Privacy and data protection regulation – impact on data integration for the development of in silico models
- *Lead: Miranda Mourby*
- Clinical trials and research ethics regulation – impact on data integration and development of in silico models
- *Lead: Eugenijus Gefenas, Vilma Lukaseviciene & Jurate Lekstutiene*
- Patients' rights – impact on data integration for development and application of in silico models
- *Lead: Mette Hartlev and Katharina Ò Cathaoir*

GDPR.EU

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH



© EU-STANDS4PM - 2022

Deliverables

- D3.1: Report containing (i) a survey of the international and European data protection and clinical trials regulation as well as legal and ethical regulation relevant for protection of equal access to health, and right to information and self-determination relevant for harmonization and integration of data in in-silico modelling, and (ii) an assessment of the challenges and options this regulation provides for harmonization and integration of data in in-silico modelling. [M12]
- D3.2: Report outlining technically feasible and legally and ethically sustainable avenues for harmonization and integration of big data of relevance for personalized medicine into *in silico* modelling. [M30]
- D3.3: Report containing the final recommendations for legally and ethically sustainable transnational data harmonization, integration and *in silico* models for personalized medicine. [M40]

© EU-STANDS4PM - 2022

Legal fragmentation

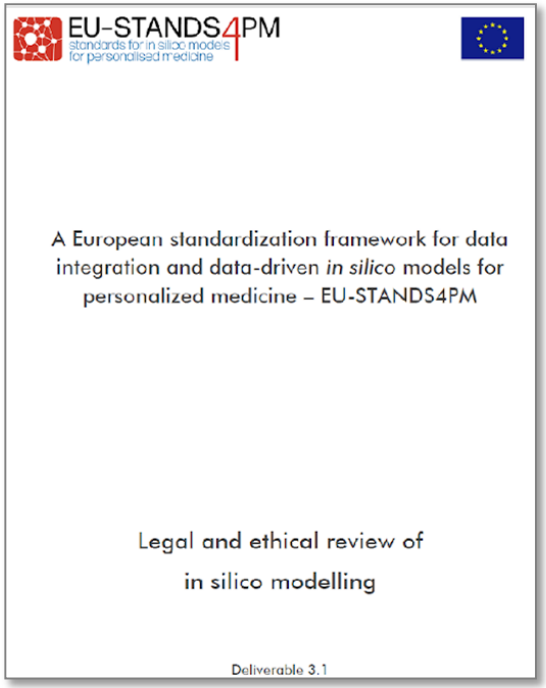
"…while the GDPR is a much appreciated piece of legislation, variation in interpretation of the law and national level legislation linked to its implementation have led to a fragmented approach which makes cross-border cooperation for care provision, healthcare system administration or research difficult."

European Commission, DG Health and Food Safety, 'Assessment of the EU Member States' rules on health data in the light of the GDPR' 12 February 2021, available from: 'https://ec.europa.eu/health/sites/health/files/ehealth/docs/ms_rules_health-data_en.pdf'

© EU-STANDS4PM - 2022

Main findings in first report

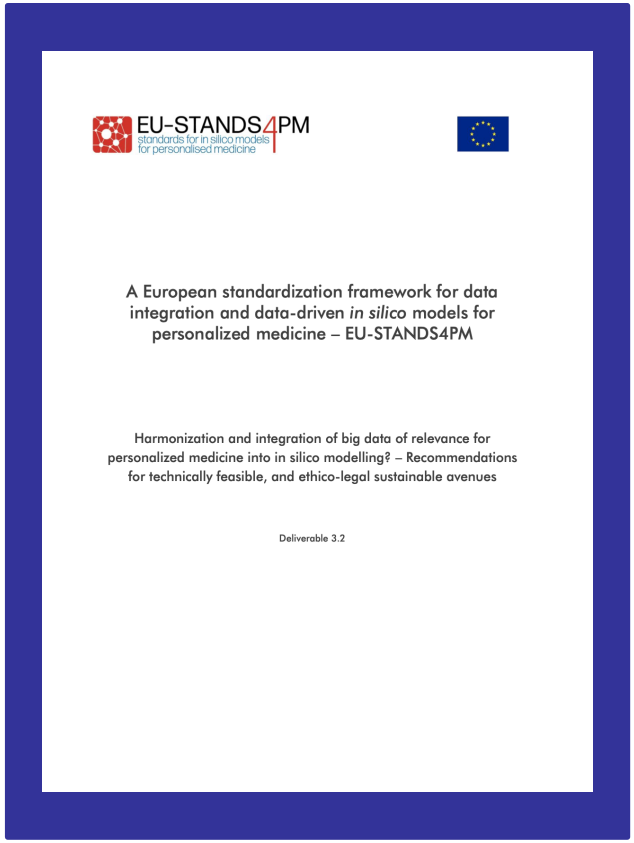
- Data protection regulation suffer from legal fragmentation
  - National laws differ both within and beyond EU
  - Uncertainty regarding legal interpretations and application of GDPR and national laws on data sharing practises
- Uncertainty regarding role and interaction of GDPR consent (*informational consent*) and research ethics consent (*interventional consent*)
- Lack of awareness of the role of patients' rights in developing in silico models



© EU-STANDS4PM - 2022

Second report

Recommendations for technically feasible, and ethico-legal sustainable avenues harmonization and integration of big data of relevance for personalized medicine in in silico modelling



© EU-STANDS4PM - 2022

Program – stakeholder workshop 19 April 2022

10 am: Welcome and introduction to the workshop: (15 minutes)

10.15: Data Protection & Technical Avenues for Data Integration (55 minutes)

- Presentation: Miranda Mourby (15 minutes)
- Expert Commentary: Catherine Bjerre Collin & Fruzsina Molnar-Gabor (10 minutes)
- Stakeholder Questions & Feedback (30 minutes)

11.10: Interaction between research ethics guidelines and data protection regulation (55 minutes)

- Presentation: Eugenijus Gefenas, Jurate Lekstutiene & Vilma Lukaseviciene (15 minutes)
- Expert Commentary – Barbara Prainsack & Dominique Sprumont (10 minutes)
- Stakeholder feedback (30 minutes)

12.05 Break (10 minutes)

12.15: Rights of Patients & Research Subjects (50 minutes)

- Presentation: Katharina Ó Cathaoir (10 minutes)
- Expert Commentary - Santa Slokenberga & Edward Dove (10 minutes)
- Stakeholder Questions & Feedback (30 minutes)

13.05: Wrapping up and way forward (10 minutes)

13.15: Close

© EU-STANDS4PM - 2022

Data Protection Conclusions

- The GDPR = not a perfectly harmonised framework.
- Federated + synthetic data generation lessens personal data burden.
- BUT real patient data needed in some cases?
- Two EU initiatives of particular relevance:
  - Data Governance Act
  - European Health Data Space
- Questions remain, relevant to final recommendations…

© EU-STANDS4PM - 2022

1) Data Altruism Consent

Question: broad or granular consent?

The European Commission will need to decide in designing their consent form.

Chapter IV facilitates data altruism (data voluntarily made available by individuals or companies for the common good). It establishes the possibility for organisations engaging in data altruism to register as a 'Data Altruism Organisation recognised in the EU' in order to increase trust in their operations. In addition, a common European data altruism consent form will be developed to lower the costs of collecting consent and to facilitate portability of the data (where the data to be made available is not held by the individual).

© EU-STANDS4PM - 2022

2) Data Cooperatives

- Potentially another way people can join together and ‘donate’ information.
- BUT they still have to exercise their own data subject rights.
- Do we need something more radical to facilitate research? Are rights within genomic data too complex for us to negotiate individually?

- For example: subject access request for a personal genome??

© EU-STANDS4PM - 2022

3) Data Minimisation

Option	Privacy Risk	Utility
Synthetic Data	Least?	Hypothesis generation only?
Federated Data Access	Lower	Promising
Remote/ Cloud Based Access	Medium	Still necessary to 'see' data directly?
Contractual Access	Higher	Necessary to physical download/ transfer data for analytical purposes?

Question: when is federated data querying insufficient for in silico modelling purposes?

© EU-STANDS4PM - 2022

Changing understanding of consent rule in the context of health data research

- *controversies between interpretation of* **consent rule** in **different normative frameworks:**

- ***research ethics*** *(Examples: WMA Declaration of Helsinki 2013; CIOMS Guidelines 2016; CoE Recommendation 2016):*
- Integrating **both**:
  - **interventional** consent (consent to intervention – very few exceptions)
  - **informational** consent (consent to process health data – includes modalities and exceptions)

- ***protection of personal data*** *(*GDPR (2016), national data protection laws*):*
  - **Only informational** consent

© EU-STANDS4PM - 2022

Misconceptions and controversies

Informational vs interventional "consent misconception" for participants, RECs and others

- *E.g., research participants can think that consent to participate in a research project also extends to the consent to process their personal data (Dove and Chen 2020) and therefore* ***mistakenly believe that s/he is still able to access the data, object to its further processing or erase it***

Controversies of informational consent

- two major areas of dispute:
  - Broad consent
  - Waiving of consent

© EU-STANDS4PM - 2022

Informational consent controversy I: the use of broad consent (e.g., in the context of biobanks)

Research ethics regulations

- broad consent is explicitly allowed - future research use can be "...**extending to a number of wholly or partially undefined studies**" (CIOMS, guideline 11)

Data protection regulations

- Sets a very strict standard for IC - it must be "clear, concise, specific and granular, freely given and revocable", GDPR, Art.7
- consent **to only "certain areas of scientific research"** (GDPR, Recital 33):

Which interpretation of broad consent is preferable in the context of biobanking?

© EU-STANDS4PM - 2022

Informational consent controversy II: Waiving of consent (e.g., secondary research of data)

Research ethics regulations

- waiving of consent is justified only as an **exception (**Declaration of Helsinki, 2013**) or**
- **only "where the attempt to contact the person concerned" has been made and proved to be unsuccessful** (Council of Europe. Recommendation 2016).

Data protection regulations

- GDPR, Art. 6 provides a "research-friendly approach" (Shabani and Borry, 2018) as it **does not give consent any predefined priority** for health data processing, &
- Art. 9(2)(j) so-called **'research condition' allows an alternative to re-consent** for the research use of previously collected health data and biological materials

© EU-STANDS4PM - 2022

Implications for researchers, research institutions, RECs and DPOs

- divergent interpretation of key concepts, such as “broad consent”, “public interest”, applicability of “research condition” **due to the institutional division**:
  - **Research Ethics Committees** (RECs) – mostly follow research ethics guidelines
  - **Data protection bodies** are primarily responsible for data protection issues in research following the GDPR

- **How should RECs and DPOs interact in the changing normative environment of health data research?**

© EU-STANDS4PM - 2022

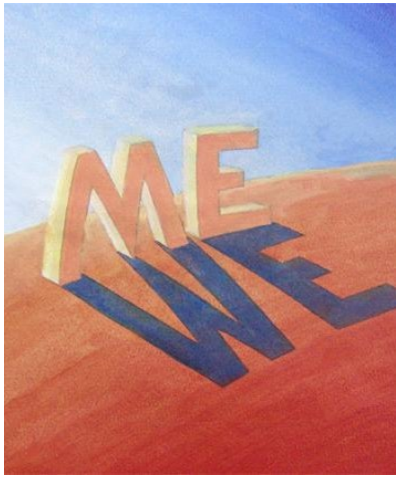
Rights of patients and research subjects - why are they important?

- Patients and research subjects are contributing with their date for the development of in silico models
- Clinical use of in silico models has an impact on patients and patient rights

© EU-STANDS4PM - 2022

Confidentiality and trust

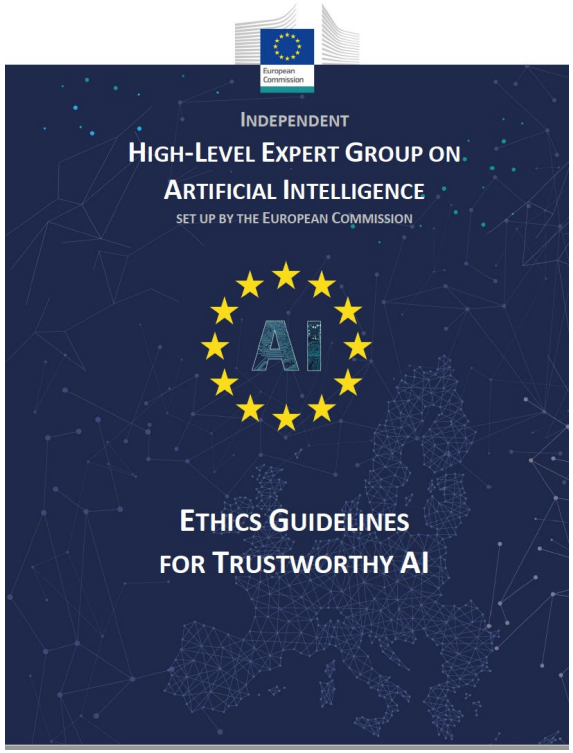
- What does confidentiality imply in a big data research environment?
- Balancing of interests
  - The individual v science and society?
- Confidence and trust in the health care services is also a societal interest
- How can confidence and trust be sustained?



© EU-STANDS4PM - 2022

Transparancy

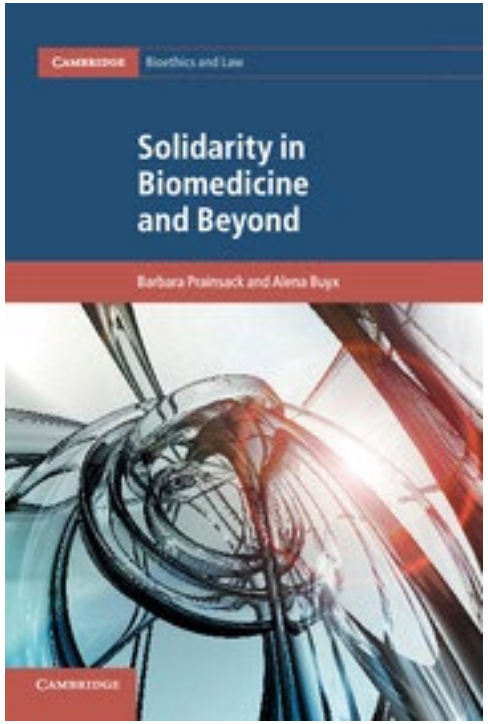
- About the **potential** use of data including data sharing across borders and with private business
  - When data are collected
- About the **actual** use of data for research or other purposes
  - Opportunity to follow data
  - Dynamic consent?
- When used for clinical purposes – transparancy about the logic of AI-based clinical advise
  - Explainability and patient right to information



© EU-STANDS4PM - 2022

A model for solidarity

Solidarity is enacted commitments to accept costs to assist others with whom a person/persons recognise similarity in a relevant respect
Barbarba Prainsack and Alena Buyx, Solidarity in Biomedicine and Beyond

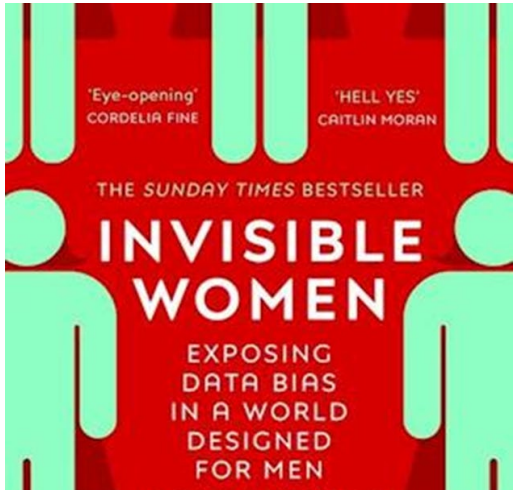


- Solidaristic practise always includes intentionality and decision-making
- Data and samples must be used in a way which creates social value
- Risk minimization

© EU-STANDS4PM - 2022

Bias

- Risk of bias is prevalent in AI-models – should be taken into consideration when developed to avoid discrimination
- Quality standards can facilitate the development of non-biased AI-models



Fairness, diversity and non-discrimination — Best possible efforts should be made to avoid unfair bias *(e.g. stemming from the used data sets or the ways the AI is developed)*. AI systems should be user-centric and whenever relevant, designed to be usable by different types of end-users with different abilities. AI systems should avoid functional bias by offering the same level of functionality and benefits to end-users with different abilities, beliefs, preferences and interests, to the extent possible. Inclusion and diversity must be enabled during the entire life cycle of the AI system. Use diverse design teams and ensure participation of affected stakeholders to ensure objectivity and inclusiveness of the developed systems/approaches.

© EU-STANDS4PM - 2022

Publications

- Mette Hartlev, 'Health Disparities and New Health Technologies – A Patients' and Human Rights Perspective', *European Journal of Health Law*, Vol 28 (2) (2020), p.142-164.
- Miranda Mourby, 'Leading by Science' through Covid-19: the GDPR & Automated Decision-Making', *International Journal of Population Data Science*, Vol 5 (4) (2020)
- Søren Brunak, Catherine Bjerre Collin, Katharina Eva Ó Cathaoir, Martin Golebiewski, Marc Kirschner, Ingrid Kockum, Heike Moser and Dagmar Waltemath, 'Towards standardization guidelines for *in silico* approaches in personalized medicine', *Journal of Integrative Bioinformatics,* Vol 17 (2-3) (2020)
- Miranda Mourby, Catherine Bjerre Colding and Katharina Ò Cathaoir, 'Explaining Machine-Learning in Healthcare – the GDPR and the *Montgomery* duty of disclosure', *Computer Law and Security Review,* vol 43 (2021)
- Eugenijus Gefenas, Jurate Lekstutieni,Vilma Lukaseviciene, Mette Hartlev, Miranda Mourby and Katharina Ó Cathaoir, 'Controversies between regulations of research ethics and protection of personal data: informed consent at a cross-road', *Journal of Medicine, Health Care and Philosophy,* Vol 25 (1) (2022), p. 23-30

© EU-STANDS4PM - 2022

Acknowledgements

EU-STANDS4PM is funded by the European Union Horizon2020 framework programme of the European Commission, Directorate-General for Research and Innovation under Grant Agreement # 825843.

© EU-STANDS4PM - 2022

EU-STANDS4PM
standards for in silico models for personalised medicine

A European standardization framework for data integration and data-driven *in-silico* models for personalised medicine

WP3: Innovative data governance for collaborative research projects: A new harmonized data access agreement for controlled access data

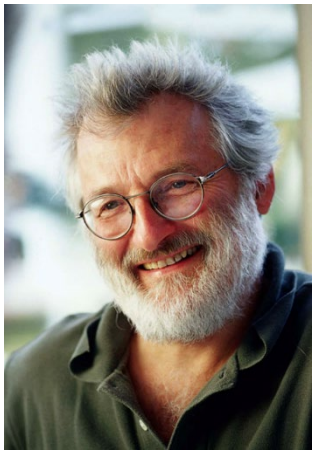
Stephan Beck, UCL
s.beck@ucl.ac.uk

© EU-STANDS4PM - 2022

Background

Bermuda Agreement (1996)

*Immediate **open access** release to accelerate research and prevent privileged exploitation of human DNA sequence*



Sir John Sulston

1942 - 2018

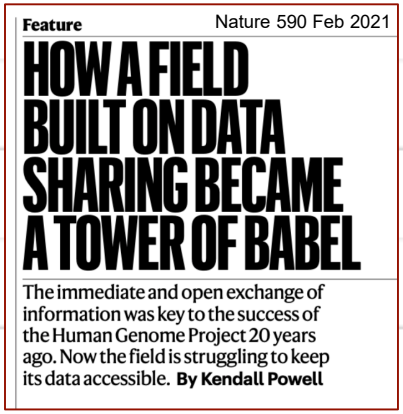
"His dedication to free access to scientific information was the basis of the open access movement and helped ensure that the reference human genome sequence was published openly for the benefit of all humanity"

Jeremy Farrar, Wellcome Trust

© EU-STANDS4PM - 2022

Controlled Access

- Controlled Access was introduced in 2007
- Controlled Access databases: **EGA** in EU and **dbGaP** in USA
- Requires completing of a **DAA** and approval by a **DAC**


Feature Nature 590 Feb 2021

HOW A FIELD BUILT ON DATA SHARING BECAME A TOWER OF BABEL

The immediate and open exchange of information was key to the success of the Human Genome Project 20 years ago. Now the field is struggling to keep its data accessible. By Kendall Powell

Source:

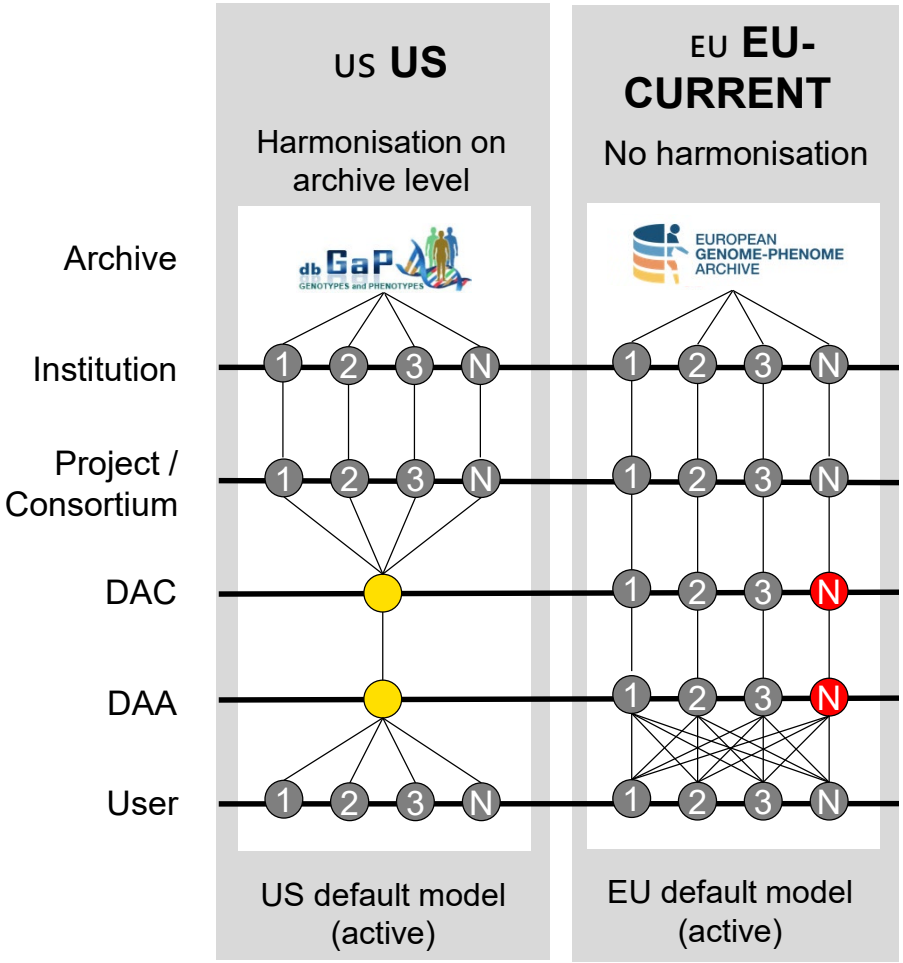
- **1932 CAPs** (Controlled Access Projects, Apr 2022) http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/summary.cgi?
- **1723 DACs** (Data Access Committees, Sep 2021) https://www.ebi.ac.uk/ega/dacs
- **1500+ DAAs** (Data Access Agreements, Sep 2021) estimate, no official statistics

© EU-STANDS4PM - 2022

➢ **Need for harmonisation**

https://gdpr-info.eu/

Controlled Access Models



- Harmonised Data Access Agreement
- GDPR compliant
- Wide consultation involving 17 projects (10 EU and 7 non-EU) and an international stakeholder workshop
- Onboarding of EGA and other stakeholders

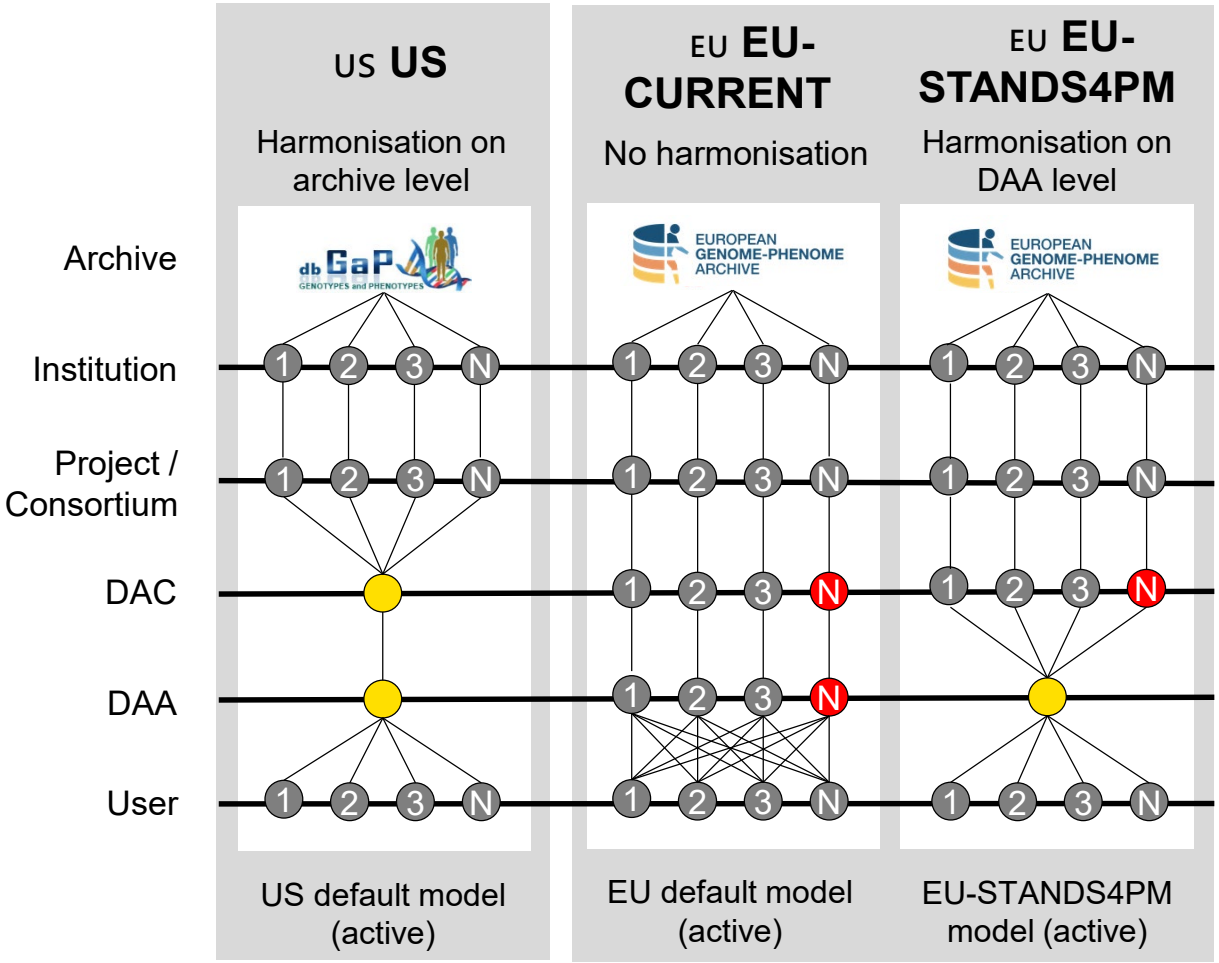
Harmonised

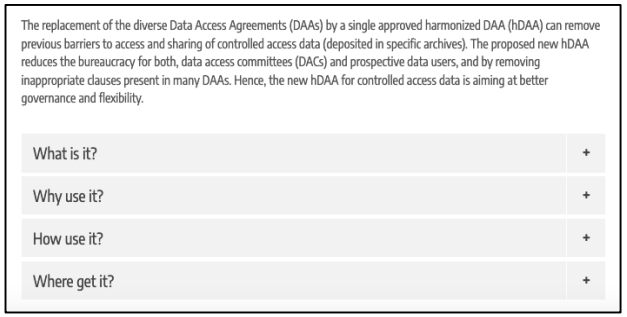
N Number of current EU DACs/DAAs = 1723 (Apr 2022)

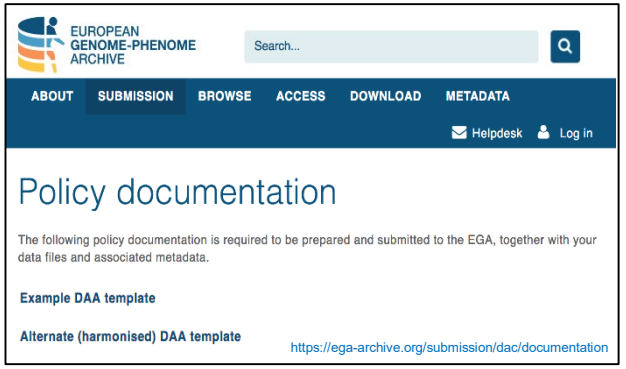
© EU-STANDS4PM - 2022

Controlled Access Models

Stamatina Liosi







Harmonised

N Number of current EU DACs/DAAs = 1723 (Apr 2022)

© EU-STANDS4PM - 2022

Implementation & uptake of hDAA

- **Implemented at EGA since October 2020**
- **Included in 2022 ERA-PerMed call**
  https://erapermed.isciii.es/joint-calls/joint-transnational-call-2022/
- **Used by 3 out of the 17 targeted projects (all non-EU)**

EpiMatch: EGAS00001006033
EPIC data (N=576)
https://www.ucl.ac.uk/cancer/research/centres-and-networks/blood-and-transplant-research-unit/research/improving-donor-selection

EpiHK: EGAD00001005486
IHEC reference epigenomes (N=19)
https://epihk.org/

BCP-ALL: EGAS00001004407
WGS (N=1007), RNA-seq (N=1186) and EPIC data (N=32)
https://www.nature.com/articles/s43018-021-00219-3

© EU-STANDS4PM - 2022

Next steps

- Onboard federated EGA nodes (how?)
- For how long will hDAA/DAC-based data access be relevant before the field moves to federated access?
- Publication of a paper describing the hDAA

© EU-STANDS4PM - 2022

Acknowledgements

EU-STANDS4PM

Rolf Apweiler
Stephan Beck
Catherine Bjerre Collin
Niklas Blomberg
Soren Brunak
Romy Gaillard
Eugenijus Gefenas
Martin Golebiewski
Mette Hartlev
Habbah Hurst
Vincent Jaddoe
Günther Kalle
Marc Kirschner
Ingrid Kockum
Sylvia Krobitsch
Florian Krumm
Lars Kuepfer
Stamatina Liosi
Vilma Lukaseviciene
Arshiya Merchant
Miranda Mourby
Neha Mishra
Heike Moser
Wolfgang Müller
Flora Musuamba Tshinanu
Katharina Eva Ó Cathaoir
Uwe Oelmüller
Tito Poli
Philip Rosenstiel
Dagmar Waltemath
Olaf Wolkenhauer

EU-STANDS4PM is funded by the European Union Horizon2020 framework programme of the European Commission, Directorate-General for Research and Innovation under Grant Agreement # 825843.

© EU-STANDS4PM - 2022