Breakout Session
Standards for human digital twins

and their implementability

Gerhard Mayer Martin Golebiewski
Heidelberg Institute for Theoretical Studies
SDBV (Scientific DataBases and Visualization) group

EDITH

EDITH project has received funding from the EU H2020 Research and Innovation Programme, under Grant Agreement n. 101083771

Agenda of Standards Break-out session

Presentation:

- Introduction and FAIRSharing collection (Martin)
- Implementation guide and standards document (Gerhard)
- Discussion(Questions, Remarks, Comments, Ideas, …)

Questions to the audience:

- … about Implementability of the presented standards on the EDITH simulation environment
- … about missing standards, terminologies or services

Implementation guide and standards document

- Standards document
  https://zenodo.org/uploads/10492796

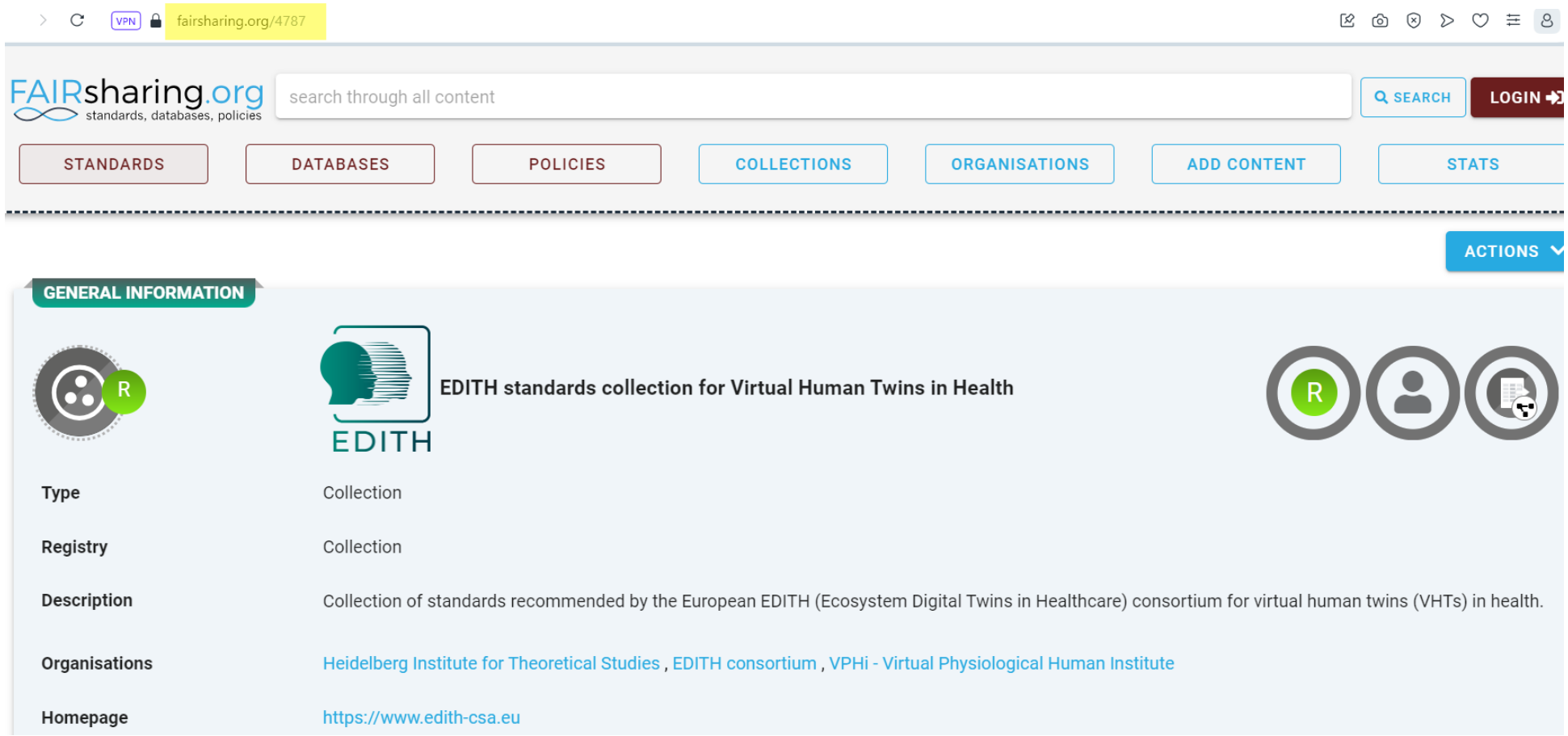
- Implementation guide
  https://zenodo.org/records/10524795

- FairSharing collection
  https://fairsharing.org/4787

FAIRSharing collection

FairSharing collection (https://fairsharing.org/4787)

currently 153 standards, taxonomies and guidelines



Implementation guide (IG) and standards document

EDITH

EDITH project has received funding from the EU H2020 Research and Innovation Programme, under Grant Agreement n. 101083771

Implementation guide - Introduction

Resources:

- EDITH-T2.3 document "Analysing current landscape of standards, identifying needs and gaps"
- EDITH-T4.2 document "EDITH standards implementation guide (IG))"
- EDITH Fairsharing collection of standards, terminologies and guidelines (https://fairsharing.org/4787)

Best practices:

- Toward Good Simulation Practice (**GSP**) guidelines of the Avicenna Alliance: risk-informed credibility assessment of *in silico* simulation models
- Modelling Good Research Practices defined by **ISPOR** and **SMDM**

Recommended standards for Europe:

- European Electronic Health Record Exchange Format (**EEHRxF**)
- IEEE P7003 Standard for **Algorithmic Bias** Considerations
- IEEE P7001/D4 Draft Standard for **Transparency** of Autonomous Systems
- IEEE 7000 Standard Model Process for Addressing **Ethical Concerns** during System Design
- …

Implementation guide – Data handling

Data preparation (ISO 20691:2022 and ISO/TS 9491-1:2023)

- **sampling** the data.
- **data formatting and harmonization,** e.g., lab value concentrations must have a **unit** associated with them. This unit can either be mass/volume or mol/volume. Therefore, the values shall be converted to a unique scale. For that the molecular weight of the analyte must be known.
- data description by **descriptive metadata,** describing for example the context of the datasets.
- **semantic annotation** of the data, e.g., by annotating genes and proteins with ontology terms.
- definition of a data **interoperability framework** (e.g. BRIDG, FHIR, …)
- **data integration,** either on the personal or on the variable level → use of persistent identifiers.
- adding data **provenance information**.
- defining who can **access** the data.
- **→ FAIReR data**

Implementation guide - Schemas

Input and output interfaces should avoid **unstructured** .csv or .txt data

- Use of **schemas** for describing the data formats, e.g.
  - JSON schema
  - XML schema (.xsd)
  - ObjTables for describing the content of .csv files
  - Own clearly documented "schema" for .csv files uniquely describing the used cells (values and units)

- Standardized **serialization formats** for binary data
  - Apache Avro
  - Google protobuf

Implementation guide – Data integration and metadata

- Use of persistent identifiers like compact Uniform Resource Identifiers (**CURIEs**). General form "**prefix:local unique identifier**", where the prefix encodes the resource; Bioregistry for resolving the CURIEs
- Data **interoperability frameworks** like e.g. Biomedical Research Integrated Domain Group (**BRIDG**)

Metadata standards:

- Dublin Core Metadata (**DC**): set of 15 basic metadata elements
- Data Catalog Vocabulary (**DCAT):** interoperability between catalogues
- MetaData Registry (**MDR, ISO/IEC 11179:2023)**: how to maintain database of metadata
- Open Archives Initiative Protocol for Metadata Harvesting (**OAI-PMH**): set of 6 HTML verbs

Use of ontologies:

- Systems Biology Ontology (**SBO**)
- Terminology for Description of Dynamics (**Teddy**)
- Kinetic Simulation Algorithm Ontology (**KiSAO**)
- Provenance, Authoring, and Versioning (**PAV**)
- …

Implementation guide – Metadata annotation

Metadata annotation:

- Triplet phrases: ***subject – predicate – object***"
- the recommended predicates are mostly '**is**' or '**isVersionOf**' (see Table 2 in **ISO 20691** for other predicates)
- done by embedding **RDF <annotation> elements** into XML-based data files

Example:

<rdf:Description rdf:about="./MyModel.xml#meta2">
<dcterms:description>Cardiomyocyte cytosolic ATP concentration</dcterms:description>
</rdf:Description>

Checks for data quality:

- Completeness (missing values)
- Plausibility (range checks, cross-reference checks)
- domain-specific quality formats, for instance .**mzQC** for proteomics; in genomics / sequencing the quality information is often part of the data file (**FASTQ** …)

Implementation guide – Executing models

Model parameterization:

- In SBML: ***'Parameter'*** and ***'Constraint'*** components
- SED-ML: **Parameter** class
- Other: tabular parameter estimation (**PETab**) format encoding the model parameter information

Model execution:

- readers for **SED-ML**, unzipper for **OMEX** , … must be present in the execution environment

Depending on the model type, a proper **model solver** for running the simulation must be available

- for **deterministic** SBML models: **RoadRunner**, **CellDesigner**, **Copasi**, **Morpheus**, the **SBMLToolbox** or the **systems biology simulation core algorithm**
- **Gillespie algorithm** for **stochastic** simulations, discrete-event simulations (**DES**) and multi-agent-based simulations (**MABS**).

Implementation guide – Executing environment

Execution environment:

- Workstation
- High-Performance Cluster (**HPC**)
- cloud (Amazon **AWS**, **Google Cloud** or **MS Azure**)

→a **Docker** resp. **Apptainer** container should be available

Runtime environments:

- C (libc, msvcrt.dll)
- CLR (Common Language Runtime)
- JRE (Java Runtime Environment)
- Julia, Jupyter, Mathematica, Matlab / GNU Octave, Python, R, …

Workflow execution engine:

- Support for Common Workflow Language **(.cwl)** files
- Support for workload manager Simple Linux Utility for Resource Management (**Slurm**)
- Examples are **Arvados**, **Toil**, **StreamFlow**, **Sapporo** and **yadage**

Implementation guide – Verification and validation

Proposal is to follow the procedure described by the **ASME V&V 40** standard, see the Toward Good Simulation Practice (**GSP**) book.

Table 1: Terminology used by the ASME V&V 40 standard for quality assessment

Quality Term	Description	Evidence Type
Verification	Did you solve the underlying mathematical model correctly?	Mathematical Evidence
Validation	Does the underlying mathematical model correctly represent the reality of interest?	Experimental Evidence
Uncertainty Quantification	What is the uncertainty in the inputs (e.g., parameters, initial conditions), and what uncertainty in the outputs results from that?	Statistical Evidence
Applicability	How relevant is the validation evidence to support using the model in the context of use?	Engineering Judgement
Credibility	Based on the available evidence, is there trust in the predictive capability of the computational model for the context of use (CoU)?	Engineering Judgement

Implementation guide – Verification and validation

- Definition of the scientific / medical **question of interest (QoI)** and the **context of use (CoU):** The CoU is a complete description of the planned modelling use and defines the role and scope of the model used to address the question of interest
- **model risk** (with its two components **model influence** and **decision consequence**) - the possibility that the results of the model simulation are wrong and lead to negative consequences for the patient - must be assessed
- **applicability** of a model is given by the evidence to support the use of the model in the defined CoU
- **risk-informed credibility assessment** is performed which encompasses the **three credibility factors**
  - model **verification** with the two factors **code verification** (source code or algorithmic errors) and **calculation verification** (discretization or iterative errors)
  - model **validation** asks if the model can correctly simulate reality, e.g., the correctness of the underlying model assumptions and approximations
  - **uncertainty quantification** (**UQ**): sensitivity analysis to determine how sensitive the model output reacts to uncertainties in the model assumptions and input parameters

Implementation guide – Reporting, visualization and archiving

- **Reporting:**
  - SBML models: simulation results should be stored in Systems Biology Results Markup Language (**SBMRL**) files.
  - Consolidated Standards of Reporting Trials (**CONSORT**) for clinical trials
  - Strengthening the Reporting of Observational Studies in epidemiology (**STROBE**)
  - Standards for the Reporting of Diagnostic accuracy studies (**STARD**)
  - Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (**TRIPOD**)

Analogous set of regulations for Artificial Intelligence reporting (suffix **–AI**)

- **Visualization:** visual representations of signaling, metabolic and gene regulatory pathways
  - Systems Biology Graphical Notation (**SBGN**)
  - Biological Pathway Exchange (**BioPAX**)
  - …
  - **Archiving:** all the data, models, parameters, and simulation results must be archived according to the General Data Protection Regulation (timeframe, consent, …)

Implementation guide – EHR data and CDS

Electronic Health Record (EHR) data:

- should follow the European Electronic Health Record Exchange Format (**EEHRxF**) recommendations; can be implemented as
  - HL7 Fast Healthcare Interoperability Resources (**FHIR**) profiles
  - Integrating Healthcare Enterprise (**IHE**) profiles
  - Open Electronic Health Record (**OpenEHR**) format
  - For integrating EHR data with research data, the Biomedical Research Integrated Domain Group (BRIDG) standard can be used.

Clinical decision support (CDS) systems:

- knowledge representation the **Arden syntax** 3.0
- encoding of clinical decision support logic by
  - Clinical Quality Language (**CQL**) using **clinical quality measures** (**CQM**): measures performance on population health in response to the delivery of health care services
  - Clinical Decision Support Hooks **(CDS Hooks)**
  - Substitutable Medical Applications and Reusable Technology for CDS (**SMART on FHIR**)

Implementation guide – Building an approved model

- The general **model building process** should be done according to the ISO/TS 9491-1:2023 ("Predictive computational models in personalized medicine research – Part 1: Constructing, verifying and validating models") standard respecting the model formatting rules described in ISO 20691:2022 ("Requirements for data formatting and description in the life sciences").
- Iterating the **model execution cycle** consisting of the steps
  - model building/adaption
  - parametrization
  - execution
  - validation, verification and uncertainty quantification (**VVUQ**) → credibility

  until the model credibility is high enough to get regulatory approval at an official **Health Technology Assessment (HTA)** admission office, e.g. **FDA** or **EMEA**.
- Execute the approved model on the targeted execution environment. Before execution, the used patient data must be prepared according to the description in *section "Data handling"*.

The END – Questions, Comments, Remarks, Ideas, …

EDITH - European Virtual Human Twin
http://www.edith-csa.eu

Deliverables available under the tab 'Dissemination/Materials':
https://www.edith-csa.eu/materials/

Indication of interest via the contact form on site
https://www.edith-csa.eu/contact

EDITH

EDITH project has received funding from the EU H2020 Research and Innovation Programme, under Grant Agreement n. 101083771

Question to the audience: Implementability

Do you see any problems for the **implementability** of the presented standards on the **EDITH simulation environment**, especially with respect to the

- Catalogue / Repository
- Simulation platform
- Workflow execution engines

Questions to the audience: Missing standards, terminologies or services

- How can SDOs and standardization initiatives help to build a standard based VHT infrastructure?
- Are there relevant standards existing, but missing in our standards document?
- Are there domains for which proper standards or terminologies are still not defined?
- Is there a need for basic services, e.g., for a dedicated terminology service for VHTs?