Strengths and deficits of CEN Workshop Agreements for data documentation in materials modelling and characterization

Björn Schembera (UStuttgart)
Martin T Horsch (NMBU)
Heinz A Preisig (NTNU)

Going beyond FAIR

Björn Schembera

Institute for Applied Analysis and Numerical Simulation

University of Stuttgart

Dark Data - Definition

- Dark Data[1] which is
  - Hidden
  - Unavailable
  - Unstructured
  - Undocumented/unannotated
  - Biased
  - Stemming from abandoned research
- Up to 80% of the global data is dark[2]
- Why should we care?
  - Economical: it costs
  - Ecological: “..annual global [..] footprints resulting from storing dark data might approach 5.26 million tons CO2 [..].”[3]
  - Responsibility problem
  - Epistemic problem (epistemic opaqueness)
  - Legal Implications

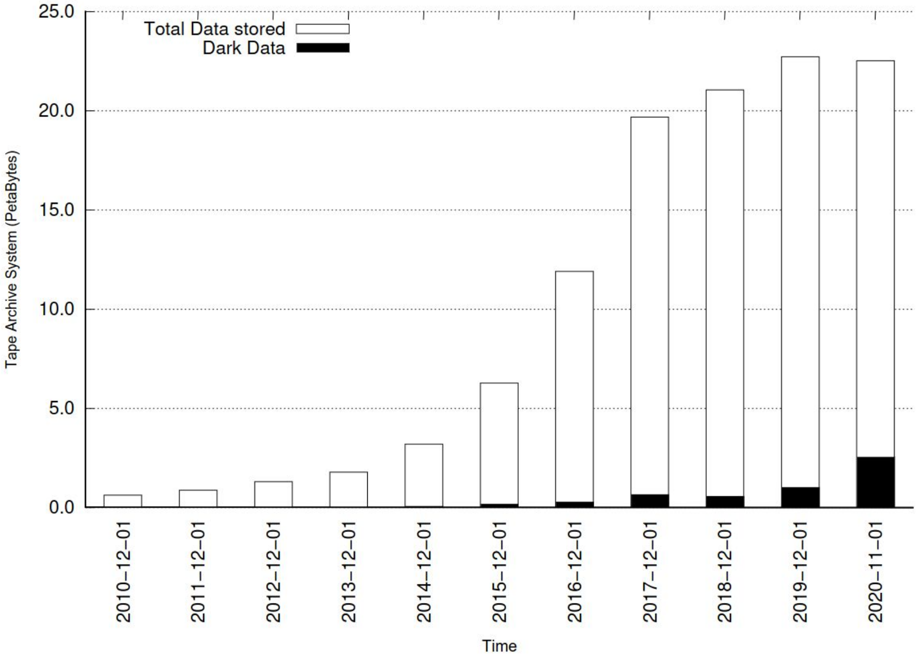
[1] Schembera, Björn, and Martin T. Horsch. Dark data and epistemic metadata in molecular modeling. In preparation

[2] Ahmad, Norita, Areeba Hamid, and Vian Ahmed. "Data Science: Hype and Reality."Computer 55.2 (2022): 95-101.

[3] Al Kez, Dlzar, et al. Exploring the sustainability challenges facing digitalization and internet data centers. J. of Cleaner Production 371 (2022).

Dark Data - Example 1: HPC Center

- Dark data at an HPC Center[4], > 11% dark data by 2020
  - Lots of data is dark due to orphaned accounts or missing metadata



[4] Schembera, Björn, and Juan M. Durán. "Dark data as the new challenge for big data science and the introduction of the scientific data officer."Philosophy & Technology 33.1 (2020): 93-115.

Dark Data - Example 2: Mathematics

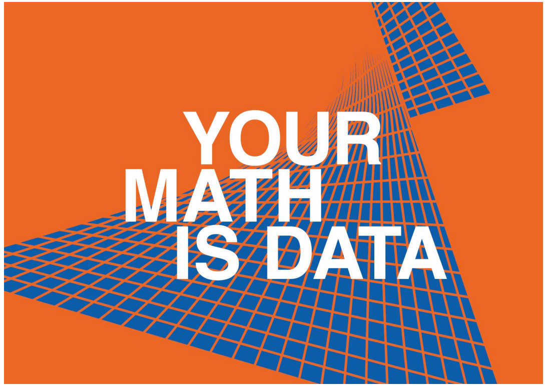
- Another example from the field of mathematics[5]

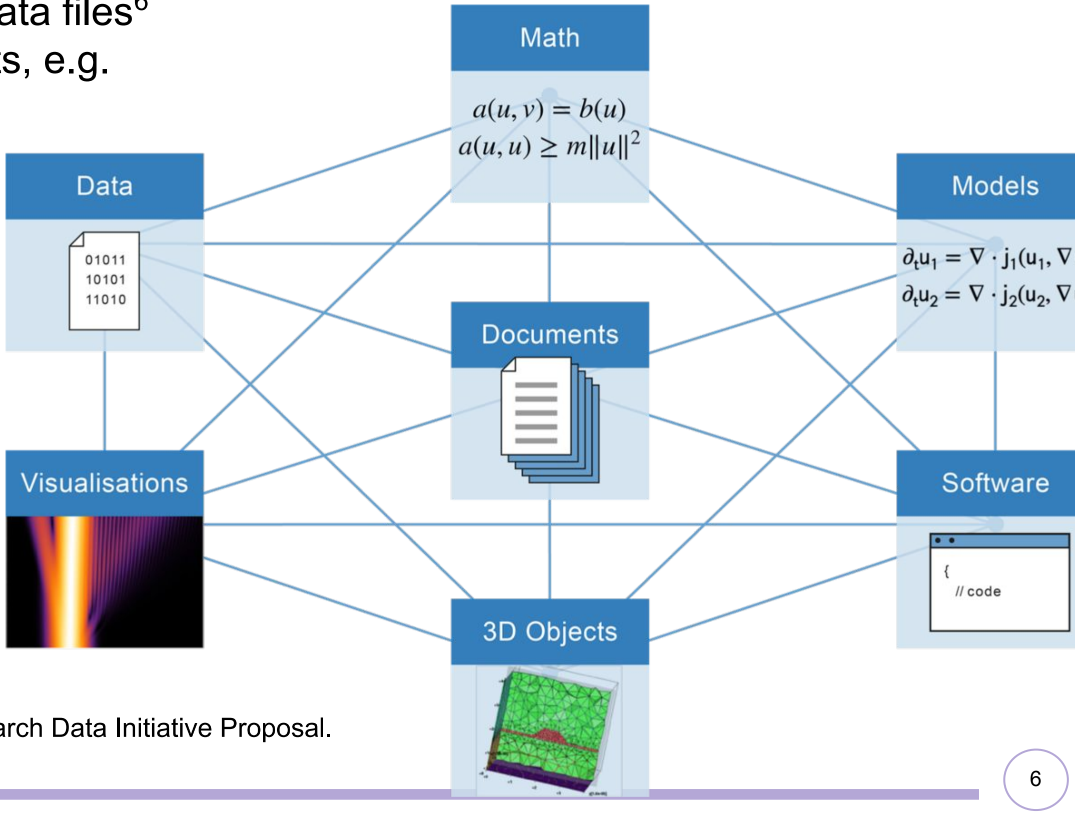
“[..] a classification of all conditional independence structures on up to four discrete random variables, originally published in a series of papers (Matus and Studeny, 1995; Matus, 1995, 1999). [..] Simcek (2006) digitized this result and left the field after his PhD in 2007. His **research data was deleted in 2021 from his former institute’s website** [..]. It was encoded in a packed binary **format which is hard to read, search, and reuse**. Some files supporting the correctness of the classification for binary distributions use an **unspecified, compiler-specific binary serialization format** for floating-point data. The programs used for the creation and inspection of the database were written in a **dialect of the Pascal** programming language which **has not been maintained since 2006**. The **sparse documentation is in Czech**.”

[5] Boege, Tobias, et al. Data Management Planning in the German Mathematical Community. arXiv preprint arXiv:2211.12071 (2022).

Your Math is Data!

- Research Data goes beyond pure data files[6]
- It's a spiderweb of information assets, e.g.
  - Models
  - Algorithms
  - …
- MaRDI@NFDI is working on this

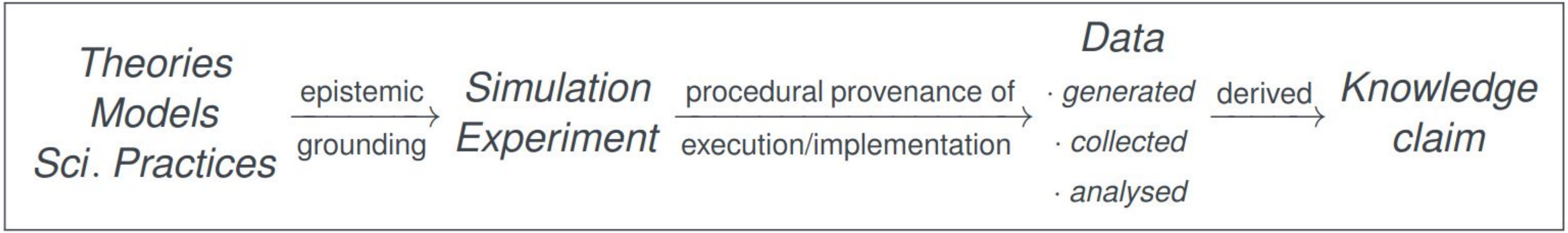




[6] The MaRDI consortium. (2022). MaRDI: Mathematical Research Data Initiative Proposal. https://doi.org/10.5281/zenodo.6552436l

Dark Data vs. FAIR Data

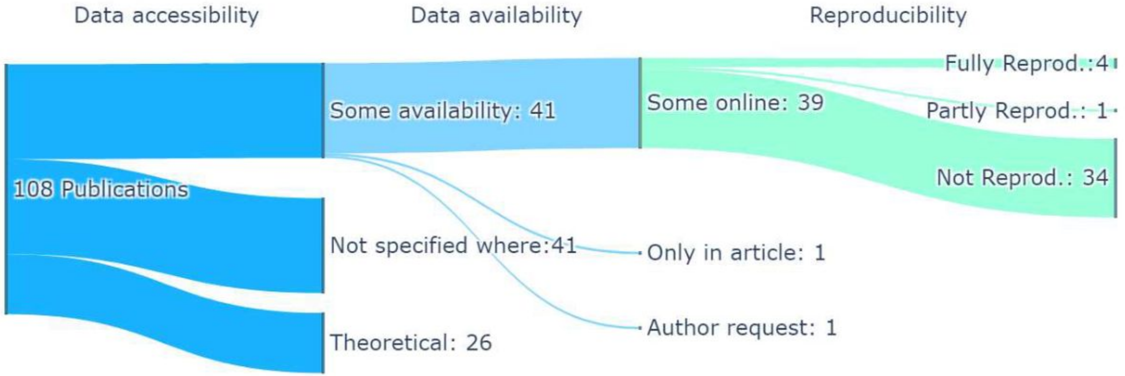
- FAIR has reached a quasi-standard paradigm in research data management
- Dark data is not FAIR -> Diminish dark data to achieve FAIRness
  - Organizational measures
    - Data Management Plans
    - Scientific Data Officers / Data curators
    - Incentives (extrinsic / intrinsic)
  - Technical measures
    - Semantic Technology
    - Data infrastructures
- Usually we limit ourselves to record pure data provenance (origin and genesis) when talking FAIR



Reproducibility and FAIR Data

- FAIR does not make statements about reproducibility, just about reusability

Example: A study that analyzed 108 publications for reproducibility:



} This is data FAIR, however not reproducible

Riedel, Christian, et al. "Including data management in research culture increases the reproducibility of scientific results." *INFORMATIK 2022* (2022).

…is FAIR really the End of the Line?

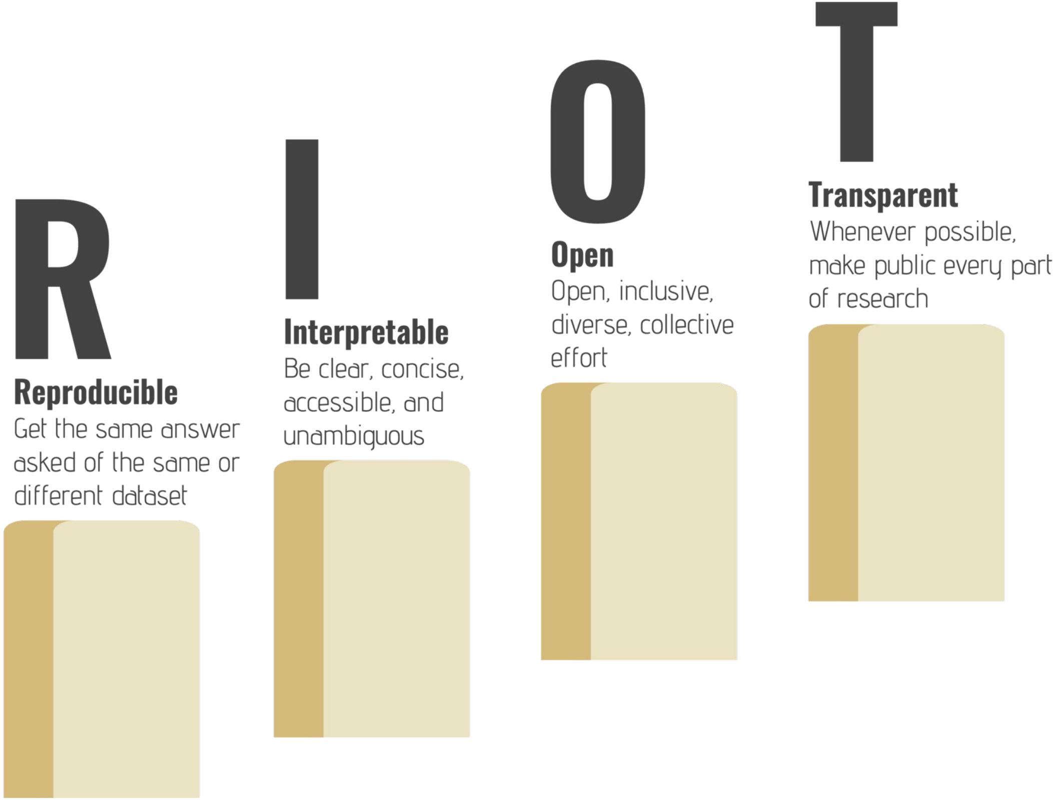
- FAIR formulates a minimum standard in data documentation
- Its ultimate goal is to optimize data for reusability with regards to data formats, licenses, retrievability, and provenance
- However it does not make statements about
  - To what extent the data is reproducible
  - What's the scientific ground for the data?
  - Why is it valid
  - Which claims have been formulated
  - Responsibility
- There is much more than just FAIR data
  - RIOT[7]: Reproducible, interpretable, open, transparent
  - CARE[8]: Collective benefit, authority to control, responsibility, ethics
  - XAIR[9]: Explainable AI ready

[8] E. Ganley et al., BMC Res. Notes 15: 51, doi:10.1186/s13104-022-05932-5, 2022

[9] S. Russo Carroll et al., Sci. Data 8: 108, doi:10.1038/s41597-021-00892-0, 2021

[10] Horsch, M., et al. "Epistemic metadata for computational engineering information systems." *Manuscript, to appear in Proc. FOIS* 2023 (2023).

RIOT



R

Reproducible
Get the same answer asked of the same or different dataset

I

Interpretable
Be clear, concise, accessible, and unambiguous

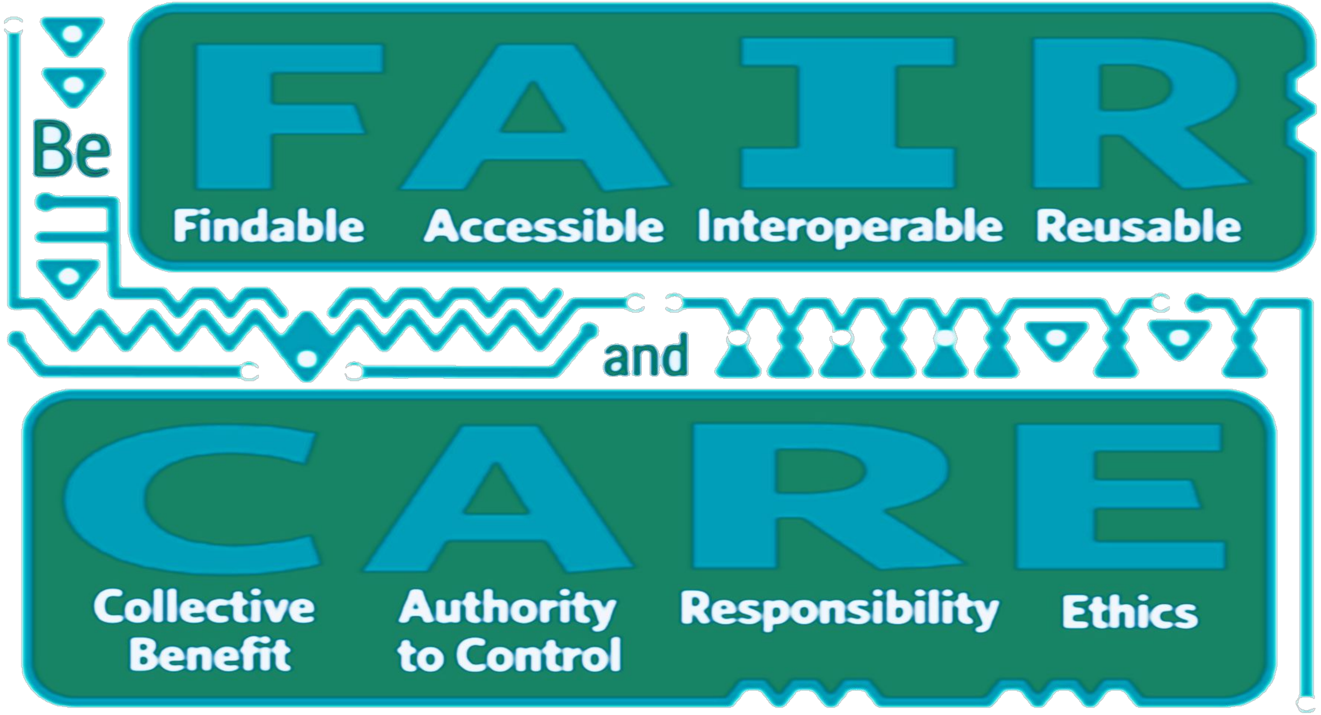
O

Open
Open, inclusive, diverse, collective effort

T

Transparent
Whenever possible, make public every part of research

CARE


Be
FAIR
Findable
Accessible
Interoperable
Reusable
and
CARE
Collective Benefit
Authority to Control
Responsibility
Ethics

Epistemic Metadata and XAIR

- Goal: Make the knowledge reusable and not only the data
- This can be accomplished by epistemic metadata
  - documenting the knowledge status of data
  - Research data must be stored and exchanged jointly with this metadata
- This makes the data XAIR (**Explainable** AI ready)
- All XAIR data is FAIR
- Not all FAIR data is XAIR
- Metadata standardization is utterly important

Personification of Episteme in Celsus Library in Ephesus, Source: Wikipedia

Toward data documentation standardization meeting regulatory and users' requirements

Martin T Horsch

Dept of Data Science

Norwegian University of Life Sciences

Responsible data documentation

European AI Act proposal: “To address the **opacity** that may make certain AI systems incomprehensible to or too complex for natural persons, a certain degree of transparency should be required for high-risk AI systems. [...] High-risk AI systems should therefore be accompanied by **relevant documentation**”

- "High-risk" includes energy safety, water, *etc.*, and all that affects fundamental rights.
- This is not in force yet, negotiations are taking place at least until end of the year.

Epistemic opacity (Humphreys, 2011): A cognitive “process is **epistemically opaque** relative to [...] agent *X* at time *t* [... if …] X does not know at t all of the epistemically relevant elements”

Tendency: Making data trustworthy through explanations becomes a **legal** requirement.

Responsible data documentation

European AI Act proposal: “To address the **opacity** that may make certain AI systems incomprehensible to or too complex for natural persons, a certain degree of transparency should be required for high-risk AI systems. [...] High-risk AI systems should therefore be accompanied by **relevant documentation**”

- "High-risk" includes energy safety, water, *etc.*, and all that affects fundamental rights.
- This is not in force yet, negotiations are taking place at least until end of the year.

Epistemic opacity (Humphreys, 2011): A cognitive “process is **epistemically opaque** relative to [...] agent *X* at time *t* [... if …] X does not know at t all of the epistemically relevant elements”

- The "epistemically relevant elements" from Humphreys are the same as the "relevant documentation" from the AI Act. We call them the **epistemic metadata**.

Tendency: Making data trustworthy through explanations becomes a **legal** requirement.

This means that **explainable-AI-ready** (XAIR) data cannot rely on *informal* metadata standardization. *Formal standardization* going through the official agencies becomes necessary.

From informal to formal standards

European AI Act proposal: “To address the **opacity** that may make certain AI systems incomprehensible to or too complex for natural persons, a certain degree of transparency should be required for high-risk AI systems. [...] High-risk AI systems should therefore be accompanied by **relevant documentation**”

- "High-risk" includes energy safety, water, *etc.*, and all that affects fundamental rights.
- This is not in force yet, negotiations are taking place at least until end of the year.

Epistemic opacity (Humphreys, 2011): A cognitive “process is **epistemically opaque** relative to [...] agent *X* at time *t* [... if …] X does not know at t all of the epistemically relevant elements”

- The "epistemically relevant elements" from Humphreys are the same as the "relevant documentation" from the AI Act. We call them the **epistemic metadata**.

Beginning with the EC's Battery Regulation, **digital product passports** will be mandatory first for batteries, later textiles, electronics, and successively more and more products.

- Characterizing the **knowledge status** becomes a priority.

DigiPass CSA

Molecular modelling case study

Epistemic metadata and their documentation were explored in molecular thermodynamics:

First stage report (10 cases), doi:10.5281/zenodo.7516532, **2023**.

Discussion of five papers each from two research groups (Berlin, London) without involving the papers' authors. Obtained a tentative **taxonomy for epistemic metadata** and explored the patterns of epistemic grounding.

Second stage report (12 claims), doi:10.5281/zenodo.7608074, **2023**.

Discussion of two claims each from six papers, involving the papers' authors, some of whom became co-authors of the present work. **Ontology of epistemic metadata**, except for epistemic grounding, implemented in PIMS-II.

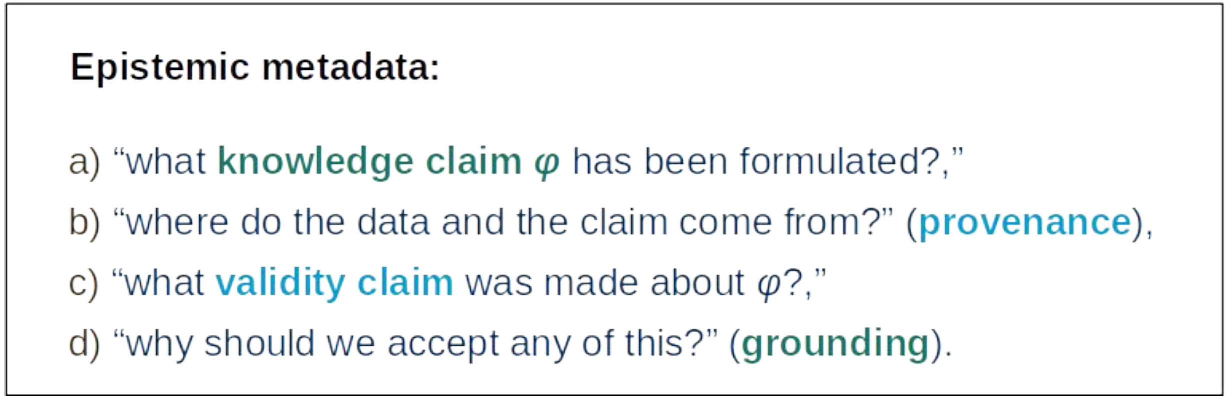
Good data documentation standards give researchers the freedom to say what they want to say. Ontologies should **provide a language, not micromanage** researchers' self-expression.

Epistemic metadata and reproducibility claims

Epistemic metadata

Metadata are “descriptive data about an object” (ISO 11179).

Epistemic metadata are metadata that support **characterizing the knowledge status** of data.



Epistemic metadata:

a) “what **knowledge claim** φ has been formulated?,”

b) “where do the data and the claim come from?” (**provenance**),

c) “what **validity claim** was made about φ?,”

d) “why should we accept any of this?” (**grounding**).

These concepts are implemented in the PIMS-II ontology.

Subject matter of research data

We understand subject matter of a knowledge claim and/or the associated research data as given by the research question that is being answered, or by the “equivalence relation over logical space” with respect to that question.[1]

Assertion: "*A* is the factually correct answer to question *Q*."

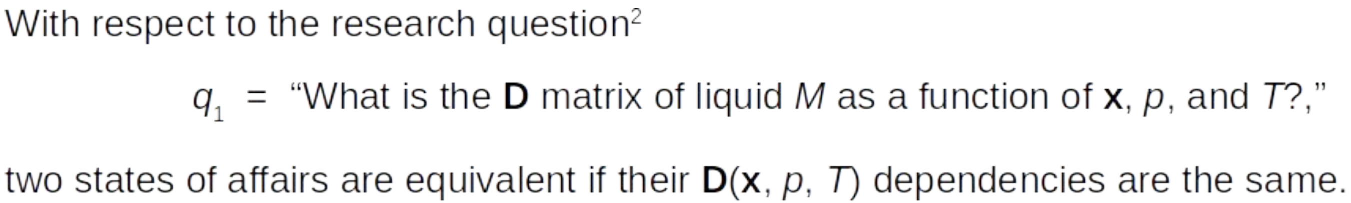
Subject matter of the assertion: Q.

Equivalence relation: Two states of affairs are equivalent if the answer to Q is the same for both.
Two knowledge bases are equivalent if they return equivalent tables for the respective SPARQL query.

[1] S. Yablo, *Aboutness*, Princeton Univ. Press (ISBN 978-0-691-14495-5), **2014**.

Subject matter of research data

We understand subject matter of a knowledge claim and/or the associated research data as given by the research question that is being answered, or by the “equivalence relation over logical space” with respect to that question.[1]



With respect to the research question[2]

$$q_1 \;=\; \text{“What is the } \mathbf{D} \text{ matrix of liquid } M \text{ as a function of } \mathbf{x},\ p, \text{ and } T?\text{,”}$$

two states of affairs are equivalent if their $\mathbf{D}(\mathbf{x}, p, T)$ dependencies are the same.

[1] S. Yablo, Aboutness, Princeton Univ. Press (ISBN 978-0-691-14495-5), **2014**.

[2] G. Guevara Carrión *et al.*, *J. Phys. Chem. B* **124**(22): 4527–4535, doi:10.1021/acs.jpcb.0c01625, **2020**.

Reproducibility claims

Common formulation and schema for reproducibility claims (RCs):

«Whenever research process *κ*’’ is carried out, it must lead to the outcome *φ*’’.»

1. Researcher *a* did *κ* and found *φ*.

2. Researcher *b* did *γ*, somehow similar to *κ*, and found something that is inconsistent with *φ*.

3. Now we think that *φ* has not been reproduced successfully, maybe it is "falsified." But why?

Reproducibility claims

Common formulation and schema for reproducibility claims (RCs):

«Whenever research process κ'' is carried out, it must lead to the outcome φ''.»

1. Researcher *a* did κ and found φ.

   Here, *a* also made the **positive reproducibility claim** $\psi = \Box(\varphi'' \mid \kappa'')$.

   - Read $\Box(\varphi'' \mid \kappa'')$ as «given that κ'' is done, necessarily a result consistent with φ'' is obtained.»

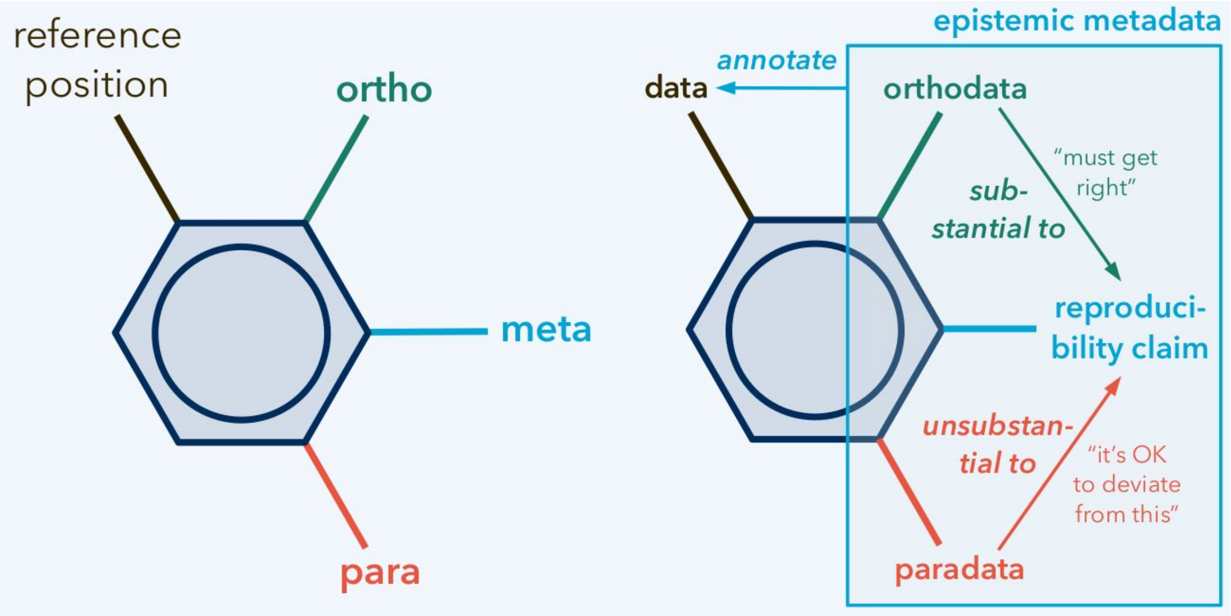
2. Researcher *b* did γ, consistent with κ'', and found something that is inconsistent with φ''.

   Here, *b* made the **negative reproducibility claim** $\Diamond(\neg\varphi'' \mid \kappa'') \equiv \neg\Box(\varphi'' \mid \kappa'') \equiv \neg\psi$.

   - Read $\Diamond(\neg\varphi'' \mid \kappa'')$ as «if κ'' is done, it can happen that a result consistent with $\neg\varphi''$ is obtained.»

3. What is relevant there is the contradiction between ψ and $\neg\psi$.

Reproducibility claims



provenance metadata *κ*

provenance paradata *κ*'

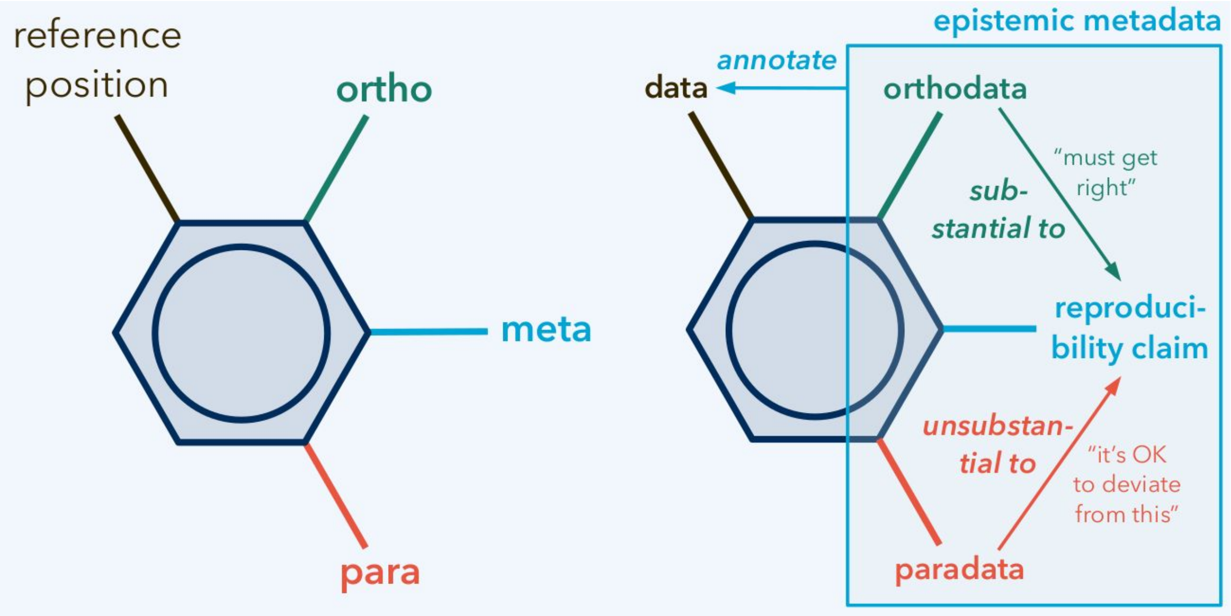
output metadata *φ*

output paradata *φ*'

«repeat *κ*, but no need to retain *κ*'»

«obtain *φ* again, except for *φ*' maybe»

Paradata and logical subtraction



provenance metadata κ

provenance paradata κ'

$\kappa'' = \kappa - \kappa'$

(provenance orthodata)

output metadata φ

output paradata φ'

$\varphi'' = \varphi - \varphi'$

(output orthodata)

«repeat κ, but no need to retain κ'»

«obtain φ again, except for φ' maybe»

Paradata and logical subtraction

Logical subtraction is a concept from analytic philosophy.

Its formalization is closely connected to the theory of **subject matter**.

Example from Yablo (*Aboutness*, **2014**): Someone who rejects ontological commitment to the existence of numbers is asked how many prime numbers there are greater than ten. "Infinitely many, of course, except that numbers don't exist."

Reproducibility in computational engineering:

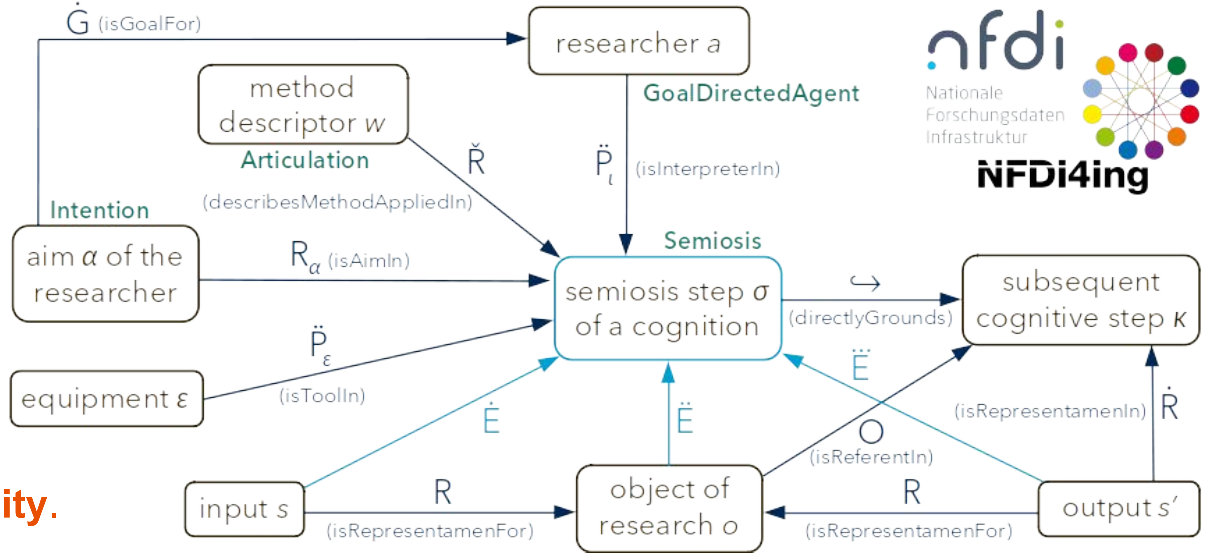
> Could you try to replicate my old simulation result? Just do the same as I did.
>
> Except that you of course log in with your user account, not mine.
>
> Your result was off by 0,5%? Don't worry, that is totally normal.

Connection to NFDI4Ing work

DORIS

- … will develop **standards for reproducibility**.
- … and **best-practice guidelines** for reproducibility.

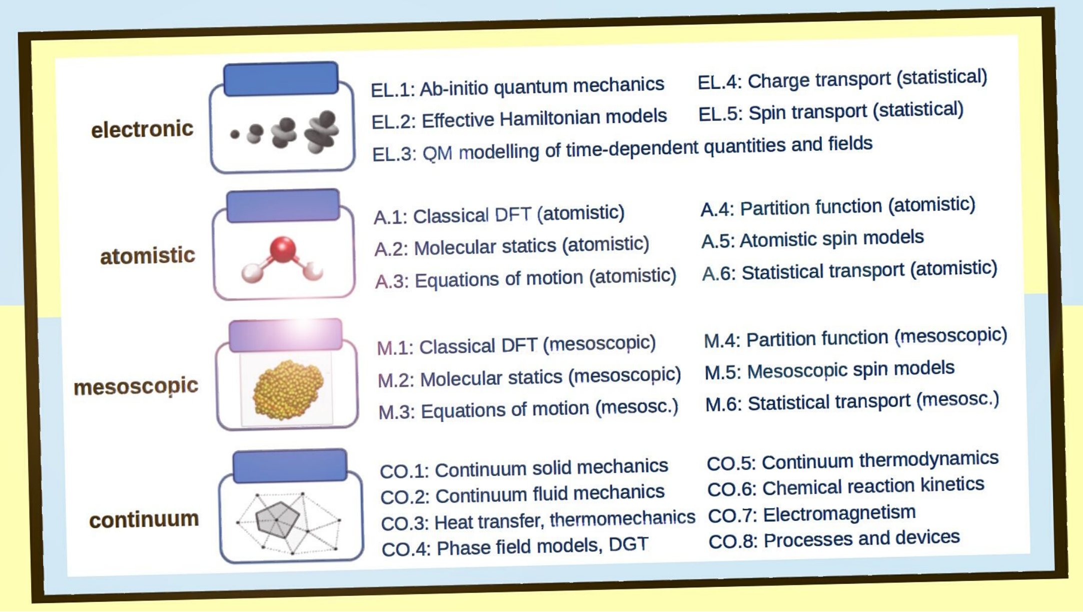
PIMS-II schema for a **semiosis** step, aligned with **processing step** from **m4i**.



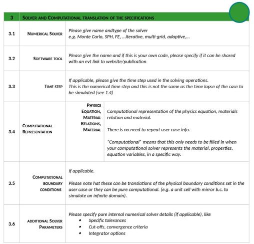
Previous work at the European level

CEN Workshop Agreements (CWAs)

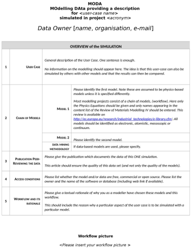
As an attempt at metadata standardization, **MODA** resulted in a closed epistemic space with a rigid categorization of modelling and simulation methodologies.



electronic	EL.1: Ab-initio quantum mechanics EL.2: Effective Hamiltonian models EL.3: QM modelling of time-dependent quantities and fields	EL.4: Charge transport (statistical) EL.5: Spin transport (statistical)
atomistic	A.1: Classical DFT (atomistic) A.2: Molecular statics (atomistic) A.3: Equations of motion (atomistic)	A.4: Partition function (atomistic) A.5: Atomistic spin models A.6: Statistical transport (atomistic)
mesoscopic	M.1: Classical DFT (mesoscopic) M.2: Molecular statics (mesoscopic) M.3: Equations of motion (mesosc.)	M.4: Partition function (mesoscopic) M.5: Mesoscopic spin models M.6: Statistical transport (mesosc.)
continuum	CO.1: Continuum solid mechanics CO.2: Continuum fluid mechanics CO.3: Heat transfer, thermomechanics CO.4: Phase field models, DGT	CO.5: Continuum thermodynamics CO.6: Chemical reaction kinetics CO.7: Electromagnetism CO.8: Processes and devices



3	Solver and Computational translation of the specifications		
3.1	Numerical Solver	*Please give name andtype of the solver* *e.g. Monte Carlo, SPH, FE, ...iterative, multi-grid, adaptive,...*	
3.2	Software tool	*Please give the name and if this is your own code, please specify if it can be shared with an evt link to website/publication.*	
3.3	Time step	*If applicable, please give the time step used in the solving operations.* *This is the numerical time step and this is not the same as the time lapse of the case to be simulated (see 1.4)*	
3.4	Computational Representation	Physics Equation, Material Relations, Material	*Computational representation of the physics equation, materials relation and material.* *There is no need to repeat user case info.* *"Computational" means that this only needs to be filled in when your computational solver represents the material, properties, equation variables, in a specific way.*
3.5	Computational boundary conditions	*If applicable.* *Please note hat these can be translations of the physical boundary conditions set in the user case or they can be pure computational. (e.g. a unit cell with mirror b.c. to simulate an infinite domain).*	
3.6	additional Solver Parameters	*Please specify pure internal numerical solver details (if applicable), like* • *Specific tolerances* • *Cut-offs, convergence criteria* • *Integrator options*	



MODA
MOdelling DAta providing a description
for <*user-case name*>
simulated in project <*acronym*>

Data Owner [*name, organisation, e-mail*]

OVERVIEW of the SIMULATION			
1	User Case	*General description of the User Case. One sentence is enough.* *No information on the modelling should appear here. The idea is that this user-case can also be simulated by others with other models and that the results can then be compared.*	
2	Chain of Models	Model 1	*Please identify the first model. Note these are assumed to be physics-based models unless it is specified differently.* *Most modelling projects consist of a chain of models, (workflow). Here only the Physics Equations should be given and only names appearing in the content list of the Review of Materials Modelling IV should be entered. This review is available on http://ec.europa.eu/research/industrial_technologies/e-library.cfm). All models should be identified as electronic, atomistic, mesoscopic or continuum.*
		Model 2	*Please identify the second model.*
		data mining methodology	If data-based models are used, please specify.
3	Publication Peer-Reviewing the data	*Please give the publication which documents the data of this ONE simulation.* *This article should ensure the quality of this data set (and not only the quality of the models).*	
4	Access conditions	*Please list whether the model and/or data are free, commercial or open source. Please list the owner and the name of the software or database (including web link if available).*	
5	Workflow and its rationale	*Please give a textual rationale of why you as a modeller have chosen these models and this workflow.* *This should include the reason why a particular aspect of the user case is to be simulated with a particular model.*	

Workflow picture

<Please insert your workflow picture >

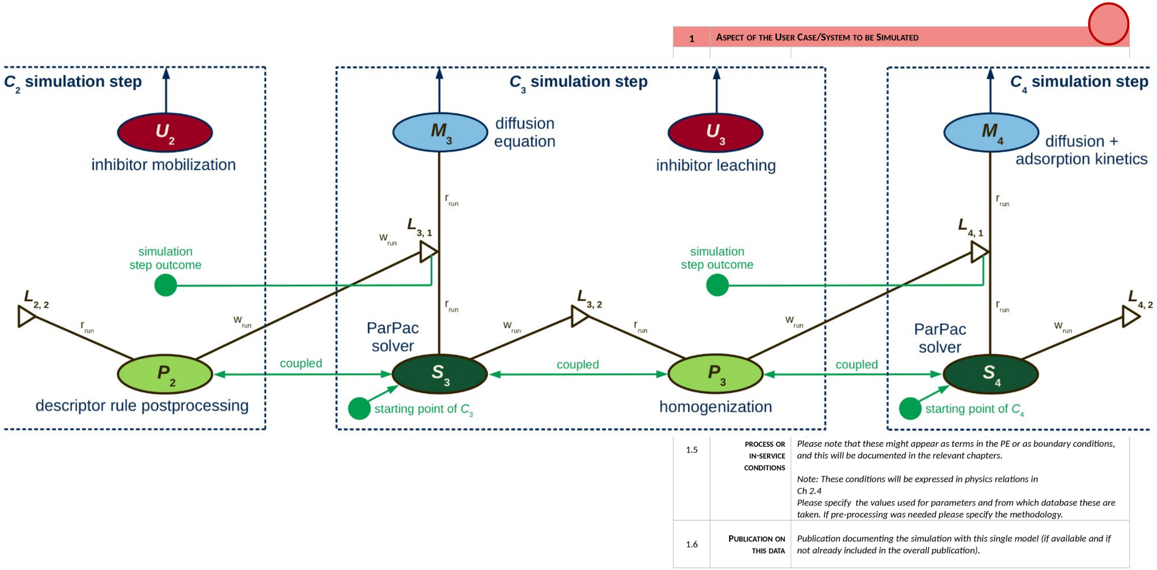
MODA: "Model Data"

CWA 17284:2018 E

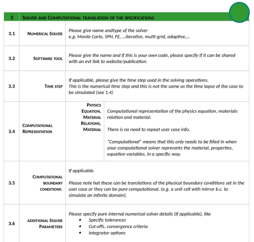
MODA

Criticism of the old CWAs

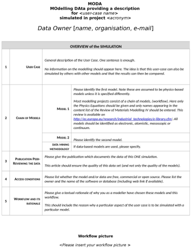
As an attempt at metadata standardization, MODA resulted in a closed epistemic space with a rigid categorization of modelling and simulation methodologies. Both **MODA** and **CHADA** documentations are **hard to create** and **hard to use** by humans, but **not machine-actionable**.



1	Aspect of the User Case/System to be Simulated	
1.5	Process or in-service conditions	*Please note that these might appear as terms in the PE or as boundary conditions, and this will be documented in the relevant chapters.* *Note: These conditions will be expressed in physics relations in Ch 2.4* *Please specify the values used for parameters and from which database these are taken. If pre-processing was needed please specify the methodology.*
1.6	Publication on this data	*Publication documenting the simulation with this single model (if available and if not already included in the overall publication).*



3	Solver and Computational translation of the specifications		
3.1	Numerical Solver	*Please give name andtype of the solver* *e.g. Monte Carlo, SPH, FE, ...iterative, multi-grid, adaptive,...*	
3.2	Software tool	*Please give the name and if this is your own code, please specify if it can be shared with an evt link to website/publication.*	
3.3	Time step	*If applicable, please give the time step used in the solving operations.* *This is the numerical time step and this is not the same as the time lapse of the case to be simulated (see 1.4)*	
3.4	Computational Representation	Physics Equation, Material Relations, Material	*Computational representation of the physics equation, materials relation and material.* *There is no need to repeat user case info.* *"Computational" means that this only needs to be filled in when your computational solver represents the material, properties, equation variables, in a specific way.*
3.5	Computational boundary conditions	*If applicable.* *Please note hat these can be translations of the physical boundary conditions set in the user case or they can be pure computational. (e.g. a unit cell with mirror b.c. to simulate an infinite domain).*	
3.6	additional Solver Parameters	*Please specify pure internal numerical solver details (if applicable), like* • *Specific tolerances* • *Cut-offs, convergence criteria* • *Integrator options*	



MODA
MOdelling DAta providing a description
for <*user-case name*>
simulated in project <*acronym*>

Data Owner [*name, organisation, e-mail*]

OVERVIEW of the SIMULATION			
1	User Case	*General description of the User Case. One sentence is enough.* *No information on the modelling should appear here. The idea is that this user-case can also be simulated by others with other models and that the results can then be compared.*	
2	Chain of Models	Model 1	*Please identify the first model. Note these are assumed to be physics-based models unless it is specified differently.* *Most modelling projects consist of a chain of models, (workflow). Here only the Physics Equations should be given and only names appearing in the content list of the Review of Materials Modelling IV should be entered. This review is available on http://ec.europa.eu/research/industrial_technologies/e-library.cfm). All models should be identified as electronic, atomistic, mesoscopic or continuum.*
		Model 2	*Please identify the second model.*
		data mining methodology	If data-based models are used, please specify.
3	Publication Peer-Reviewing the data	*Please give the publication which documents the data of this ONE simulation.* *This article should ensure the quality of this data set (and not only the quality of the models).*	
4	Access conditions	*Please list whether the model and/or data are free, commercial or open source. Please list the owner and the name of the software or database (including web link if available).*	
5	Workflow and its rationale	*Please give a textual rationale of why you as a modeller have chosen these models and this workflow.* *This should include the reason why a particular aspect of the user case is to be simulated with a particular model.*	

Workflow picture

<Please insert your workflow picture >

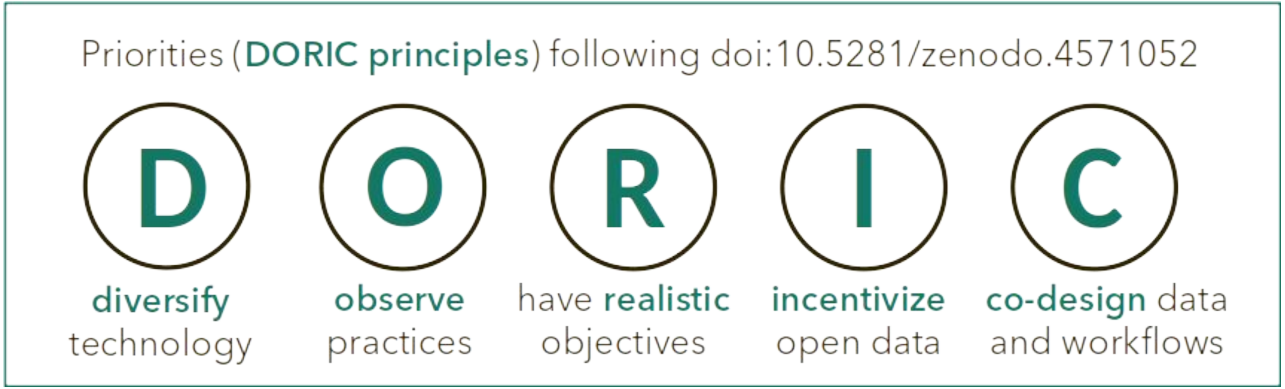
MODA: "Model Data"

CWA 17284:2018 E

CHADA: "Characterization Data"

CWA 17815:2021 E

Criticism of the old CWAs



Priorities (**DORIC principles**) following doi:10.5281/zenodo.4571052

D — **diversify** technology

O — **observe** practices

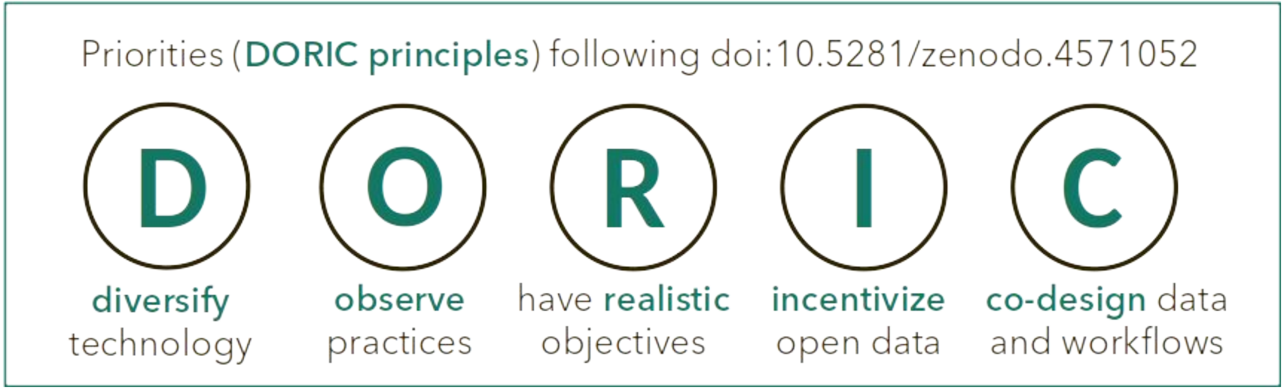
R — have **realistic** objectives

I — **incentivize** open data

C — **co-design** data and workflows

- **MODA** was a **closed epistemic space**, modelling methods had to be chosen from a small list.
- **MODA** imposed a **given level of detail** in data documentation; namely, **unrealistically** detailed.
- **MODA** documentations are **complicated**, and of **limited use** to all, including to humans.

Toward meeting user requirements



Priorities (**DORIC principles**) following doi:10.5281/zenodo.4571052

D — **diversify** technology

O — **observe** practices

R — have **realistic** objectives

I — **incentivize** open data

C — **co-design** data and workflows

- MODA was a closed epistemic space, modelling methods had to be chosen from a small list. **ModGra** gives the user a **highly expressive** graph **language** to describe their method.
- MODA imposed a given level of detail in data documentation; namely, unrealistically detailed. **ModGra** gives the user the choice to document the model at a **flexible level of detail**.
- MODA documentations are complicated, and of limited use to all, including to humans. **ModGra** specifies semantics at the level of physics and is **actionable** through ProMo tools.

CWA 17960:2022
ModGra: A graphical representation of physical process models

Heinz A Preisig

Dept of Chemical Engineering

Norwegian University of Science and Technology

Contents

1. How to go about generating a CWA
2. ModGra
   a. motivation
   b. approach
   c. basic components
   d. examples

CWA

CEN workshop agreement

step 1: decide on what shall be standardise -- aim at something minimal

step 2: contact people who may be interested

step 2: draft a working plan

step 3: get in contact with a standardisation organisation in a European country

step 4: provide workshop proposal to standardisation organisation

(form https://boss.cen.eu/media/BOSS%20CEN/formtemp/ws_proposal.docx)

step 5: form a committee -- aim at a wide spread in terms of expertise

step 6: standardisation organisation submits proposal to CEN and announce 30 days in advance a kick-off meeting

step 7: keep on meeting until standard is established -- standardisation organisation provides the secretary

step 8: the CWA is submitted to CEN

Instructions can be found on : https://boss.cen.eu/developingdeliverables/cwa/pages/ or google for *how to generate CEN workshop agreement*

Practicalities

- Aim at a small document
- Simple over complex
- A main body of a CWA document is structured like a contract
    - terms are defined
    - terms can only be used once defined or defined as commonly known
- Adding examples helps
- Simple language helps

Experience: terminology requires a lot of effort

ModGra

Approach: Reductionism

Break a process recursively down into smaller and smaller entities

How long?

- until a level of granulation is achieved that captures the essentials of the process
- the granules can be viewed as **simple systems, control volumes** characterised by
    - a time scale
    - a distribution property
        - intensive properties are a function of the location → **distributed** systems
        - intensive properties are NOT a function of the location → **lumped** systems

Approach

View a process model as a directed graph with:

- nodes being capacities, control volumes containing conserved extensive quantities
- arcs transferring conserved extensive quantities

Add information processing as directed graph with main application of control:

- nodes being input/output functions
- arcs transferring information -- signals

Abstraction: directed graph with tokens living in there, giving the graph context

similar to Petri's thinking

Foundation

A physical process contains

- mass,
- energy,
- momentum,
- charge

→ conserved quantities.

There are

- “holders”/”accumulators” and
- “transfers”

  of conserved quantities

Abstract to

tokens

being stored/accumulated and move

- **capacities** for tokens
- **transfer** of tokens

Foundation

Abstract to

tokens

being stored/accumulated and move

- **capacities** for tokens
- **transfer** of tokens

Abstract to

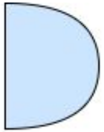
directed graph

with

- nodes being the capacities
- arcs the transport of tokens between capacities

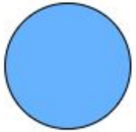
Capacities

constant



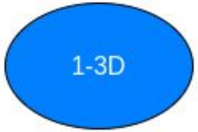
physical : temporal constant, spatial: uniform & infinitely large
information: temporal constant

dynamic lumped



physical : temporal dynamic, spatial: uniform (0D) & finite size
information: temporal dynamic

dynamic distributed

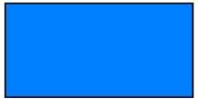


physical : temporal dynamic, spatial: not uniform (nD, n := 1,2,3) & finite size

dynamic point

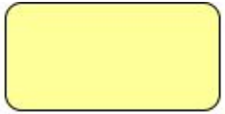
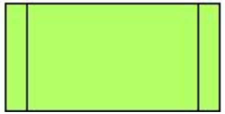
physical : temporal event-dynamic, spatial: infinite small
information: temporal event-dynamic

event-dynamic distributed



physical : temporal event-dynamic, spatial: not uniform & finite size
information: temporal event-dynamic

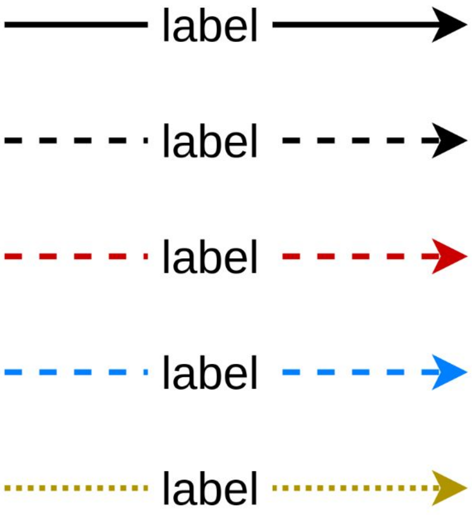
Extensions

composite		physical : information:	composite entity -- a subgraph ditto
surrogate		physical : information:	surrogate replacing typically a composite entity -- a subgraph ditto

Enables

- the construction of large systems
- empirical models

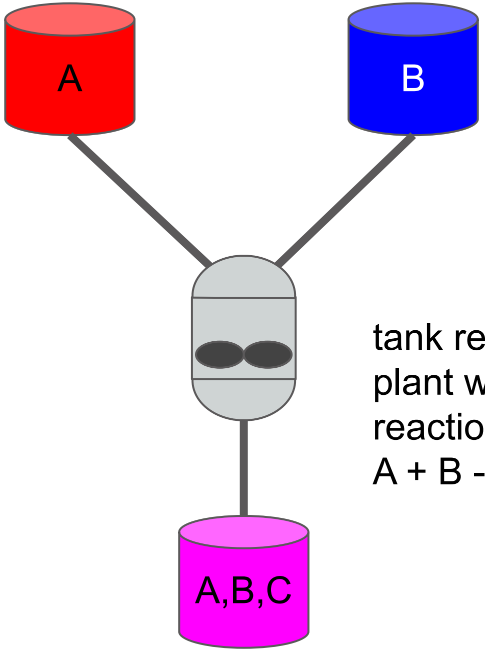
Transfers



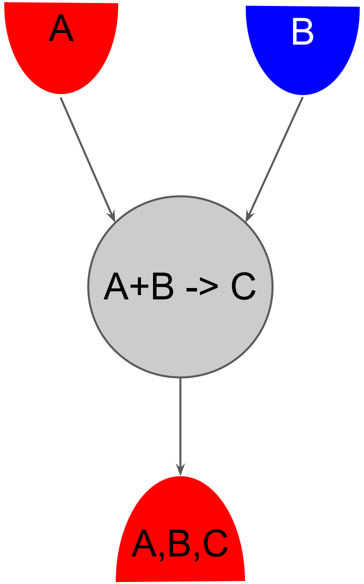
physical :	mass transfer
physical :	diffusional mass transfer
physical :	conductive heat transfer
physical :	work flow
information :	signal

Directionality defines a reference coordinate system for the respective flow

A simple example



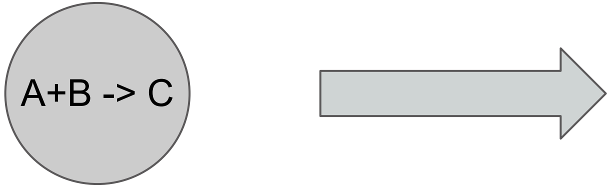
tank reactor plant with reaction
A + B -> C

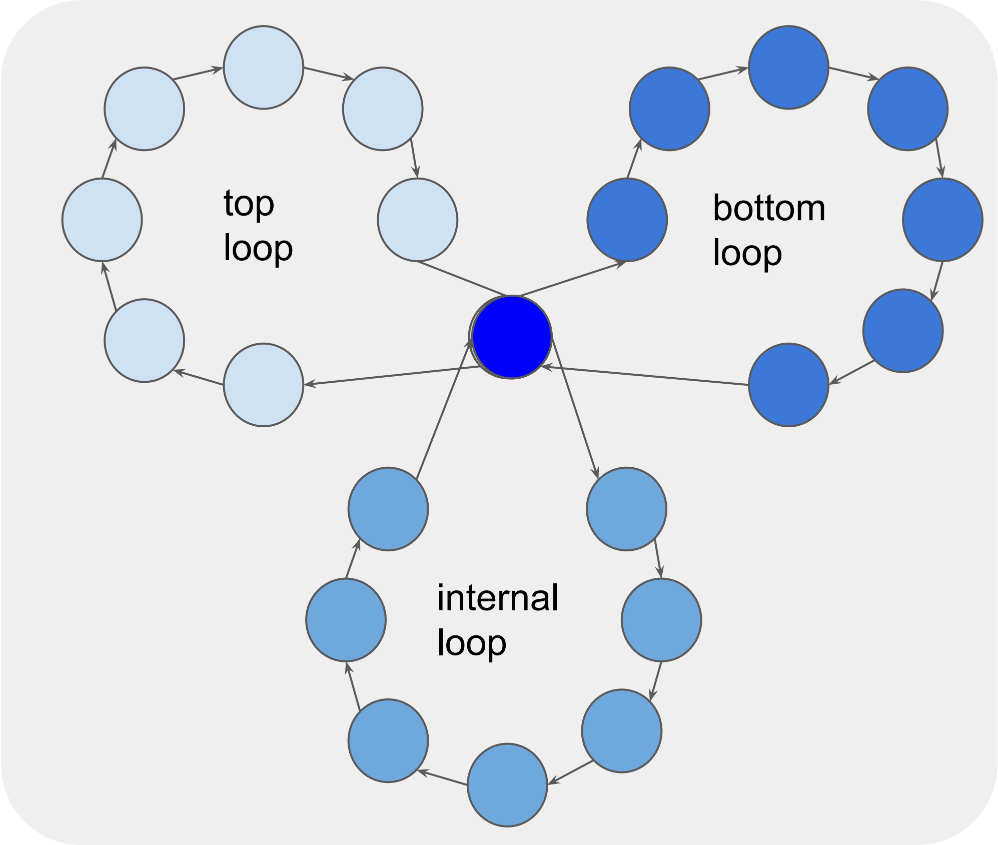


Simplest model

- reservoirs for supply and product
- lumped liquid volume in the tank
- no control

More complex mixing model

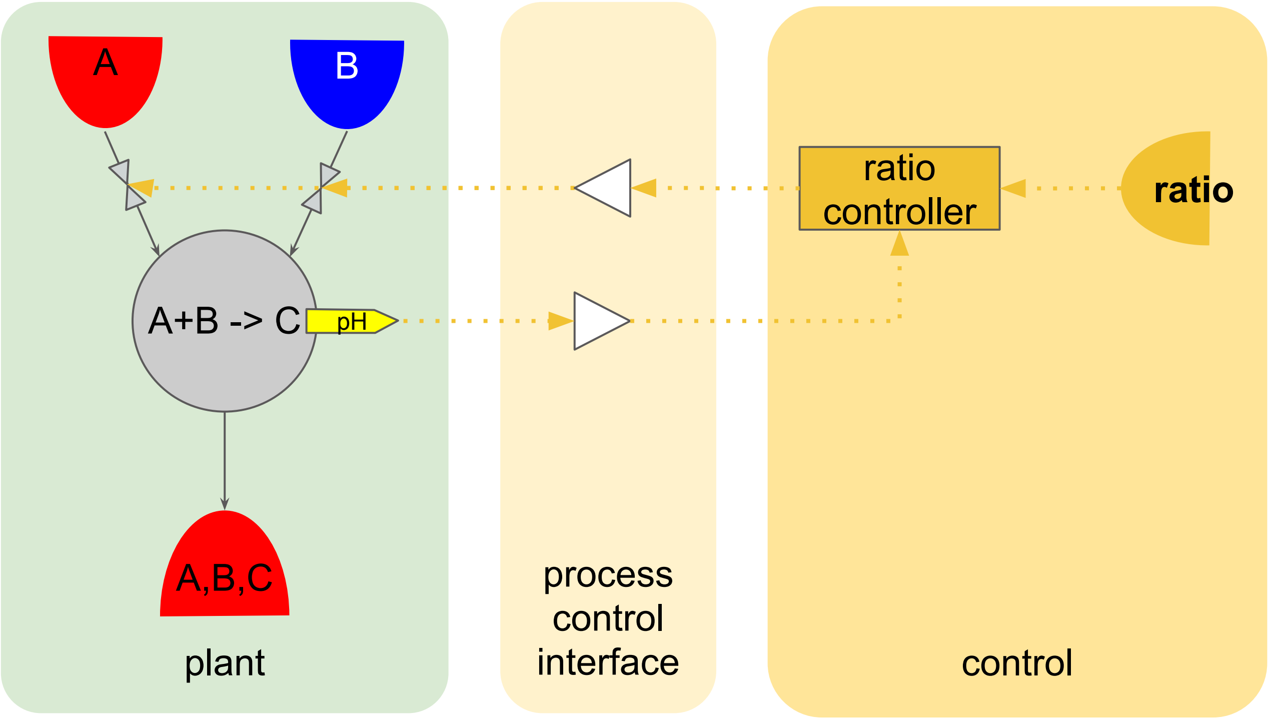




Only hydraulic shown
Would need definition

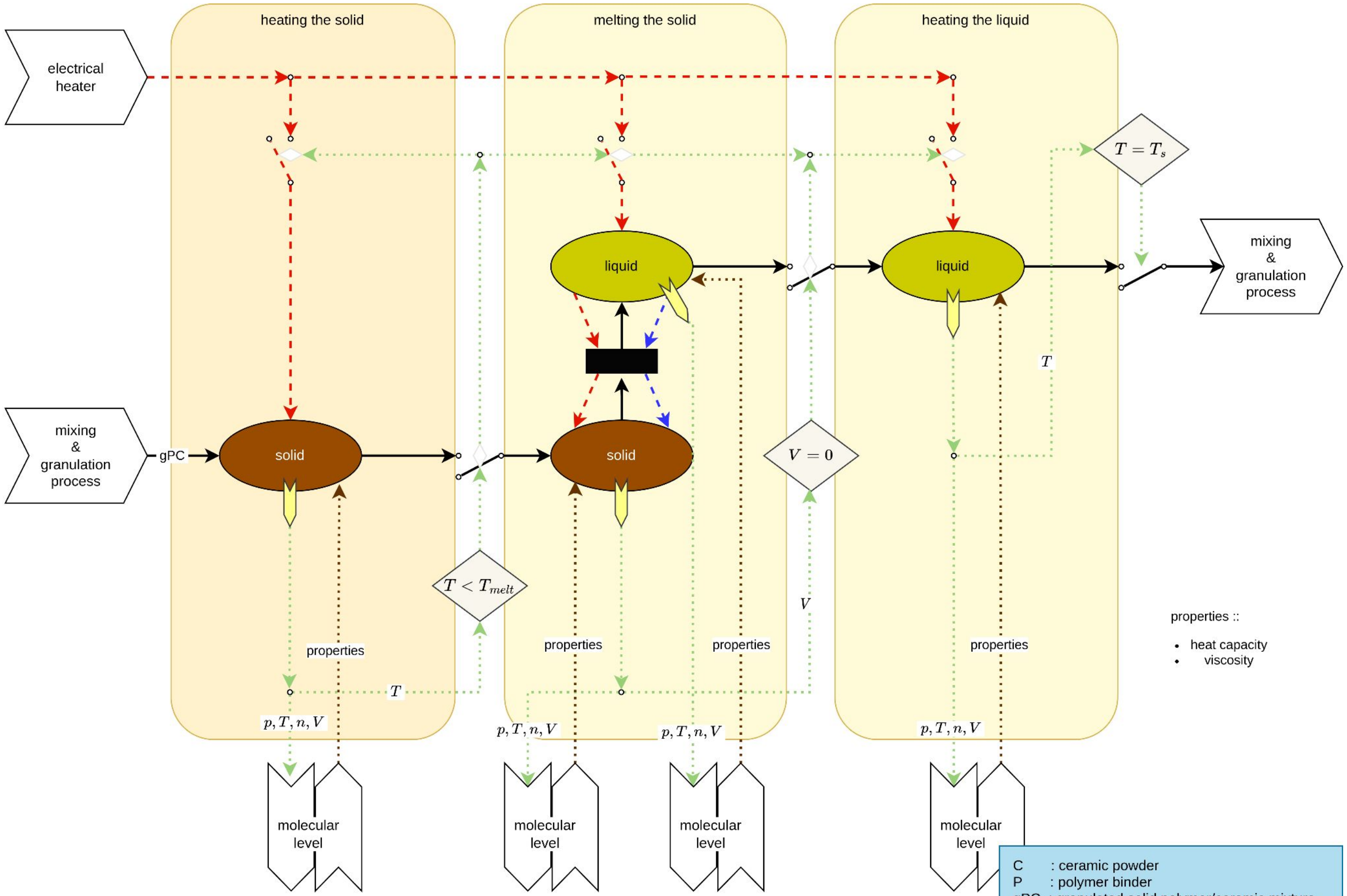
- where reaction takes place
- where A and B flows in
- where product is drawn

Adding control



Multi-scale model -> workflow

Melting process



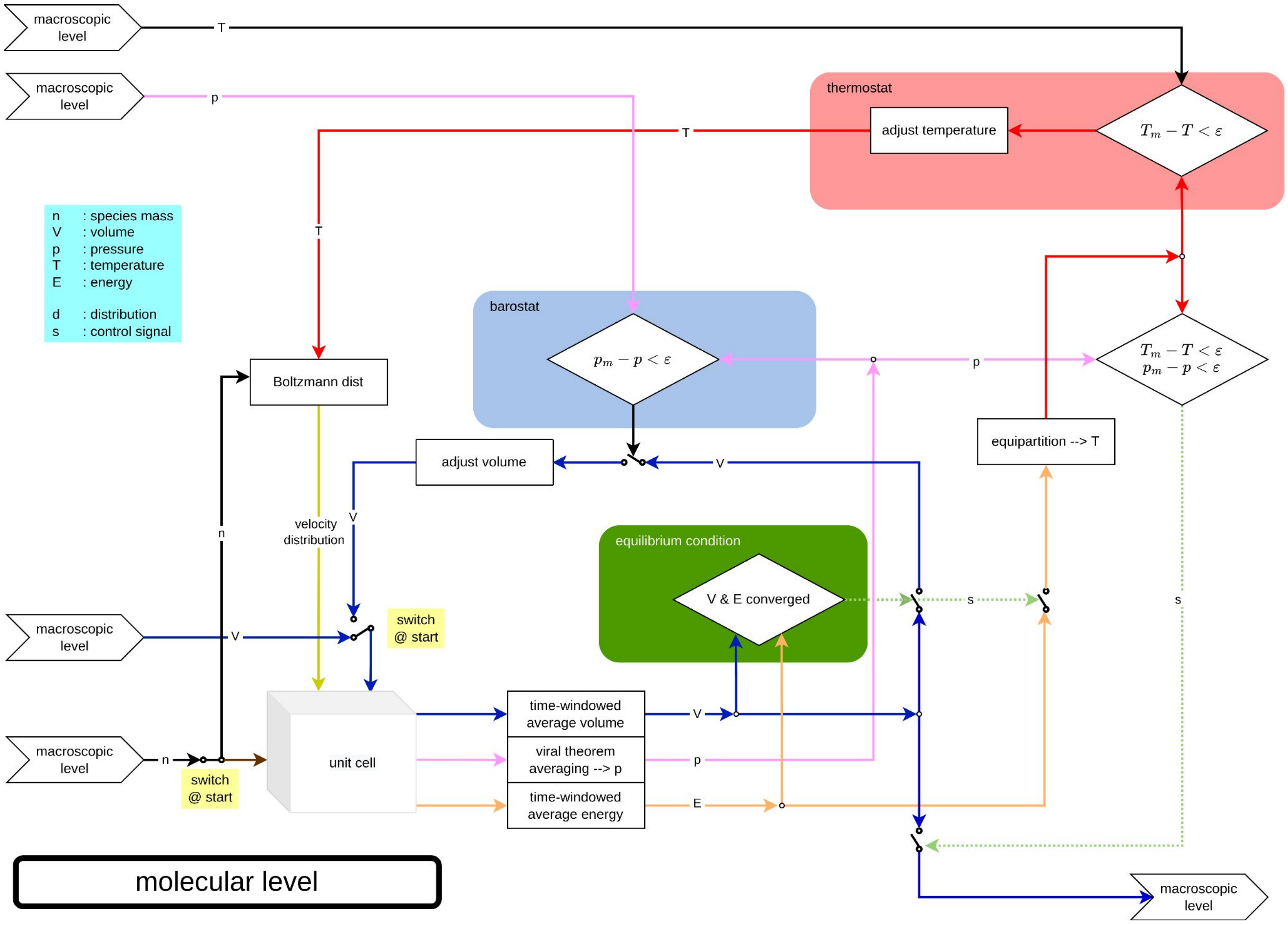
properties ::

- heat capacity
- viscosity

C : ceramic powder
P : polymer binder
gPC : granulated solid polymer/ceramic mixture
mPC : molten polymer with ceramic emulsified
part : moulded solid product

three staged melting process

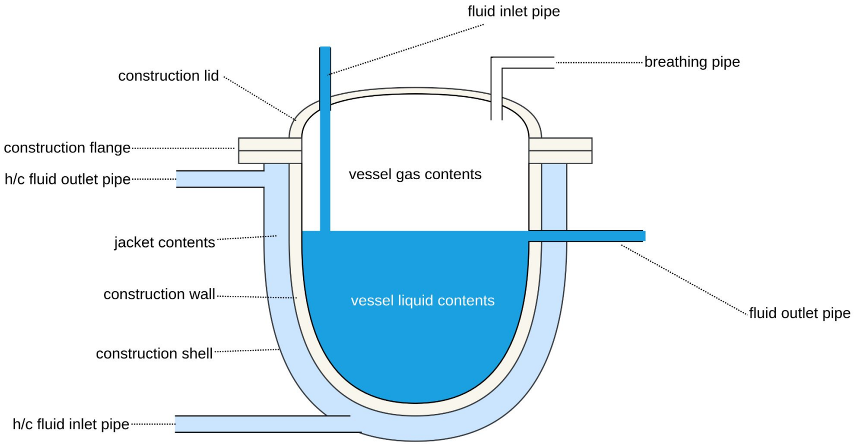
Molecular modelling



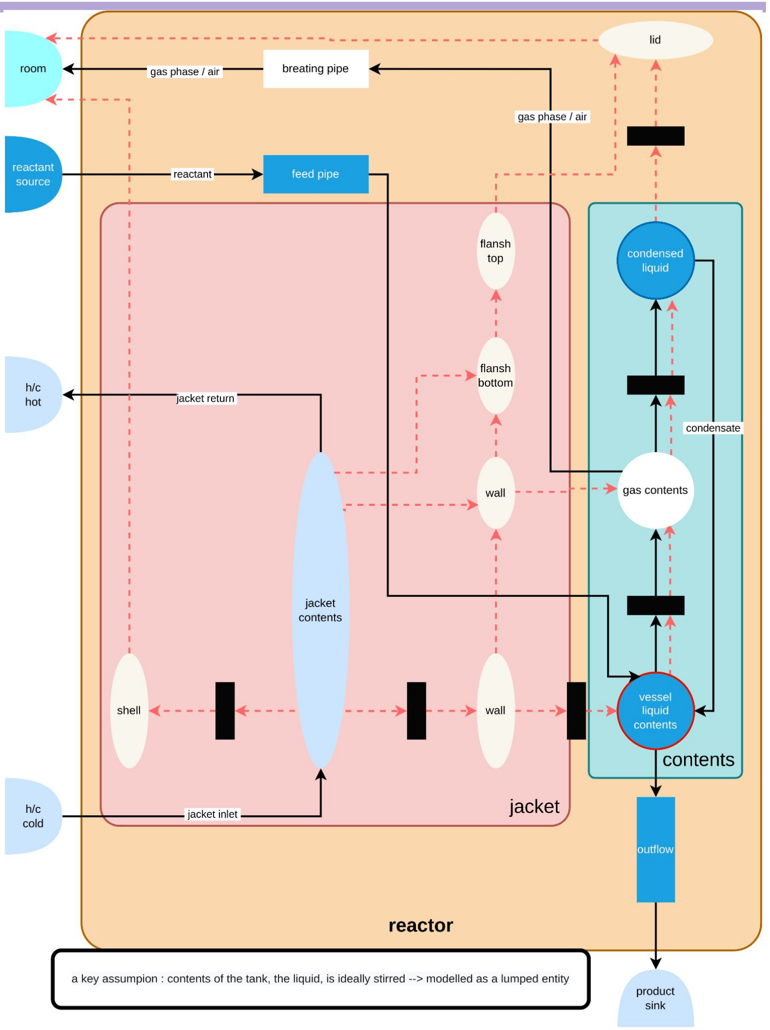
n	: species mass
V	: volume
p	: pressure
T	: temperature
E	: energy
d	: distribution
s	: control signal

Model simplification

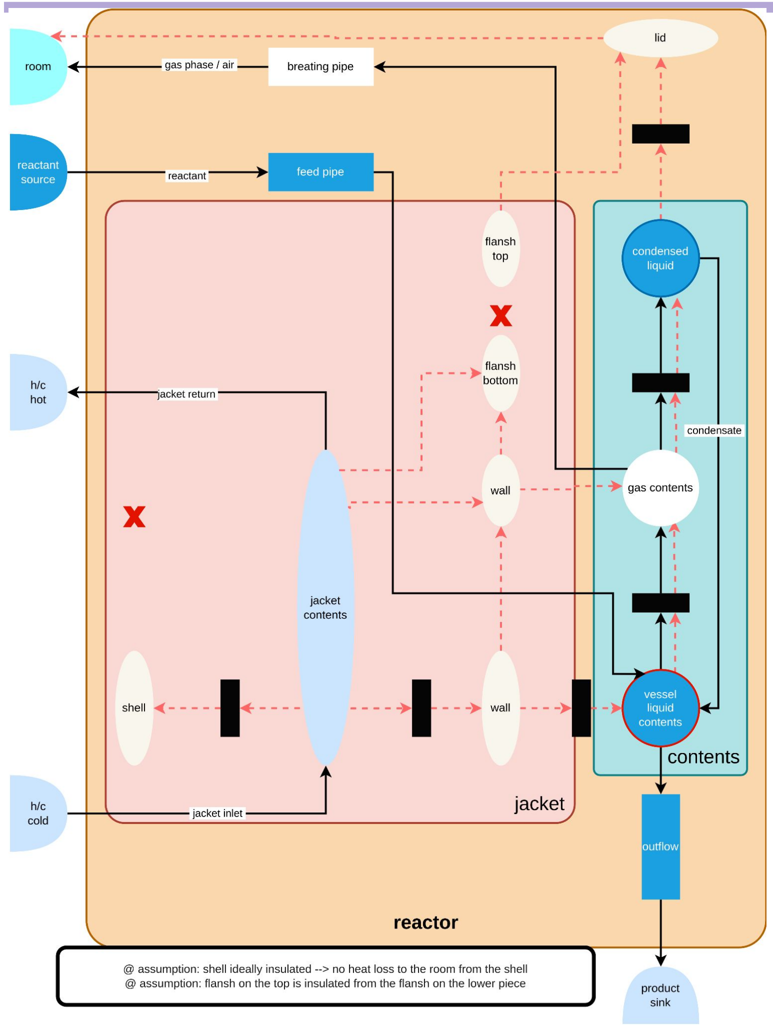
Stirred tank



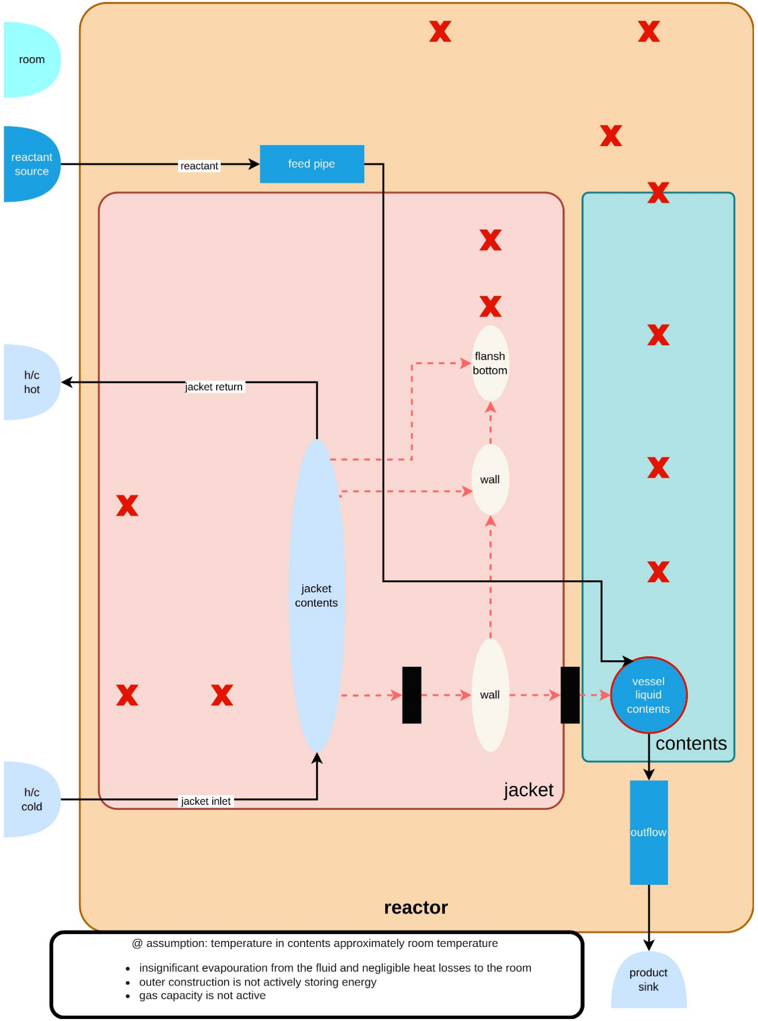
schema of a standard stirred tank



a key assumpion : contents of the tank, the liquid, is ideally stirred --> modelled as a lumped entity

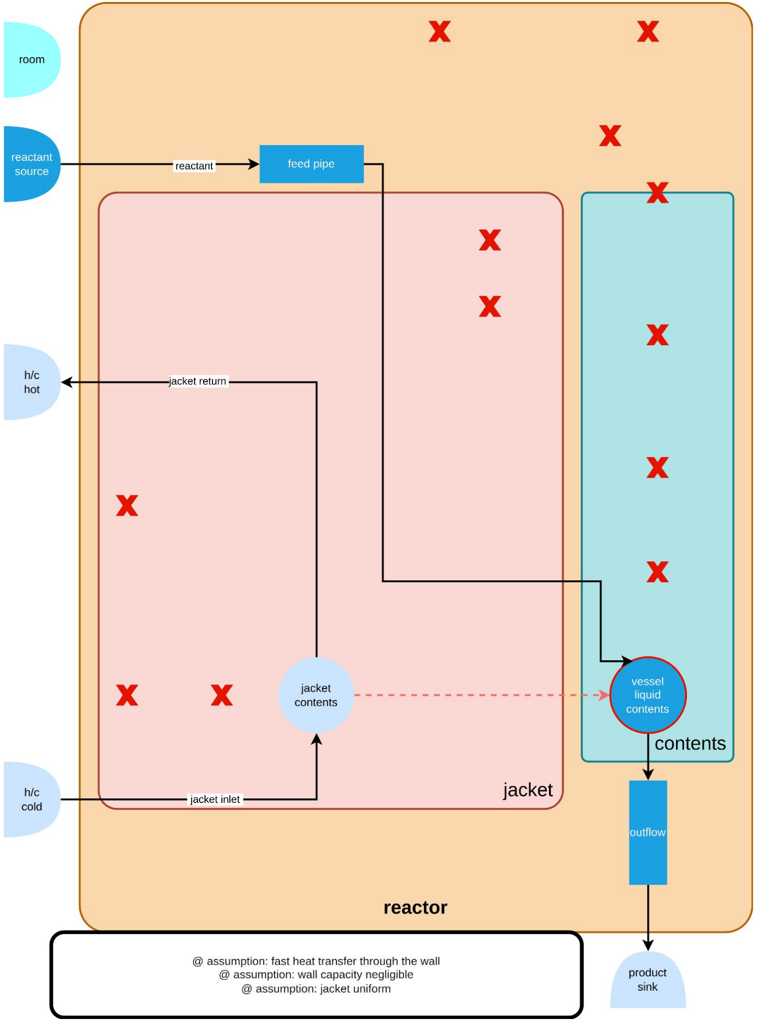


@ assumption: shell ideally insulated --> no heat loss to the room from the shell
@ assumption: flansh on the top is insulated from the flansh on the lower piece

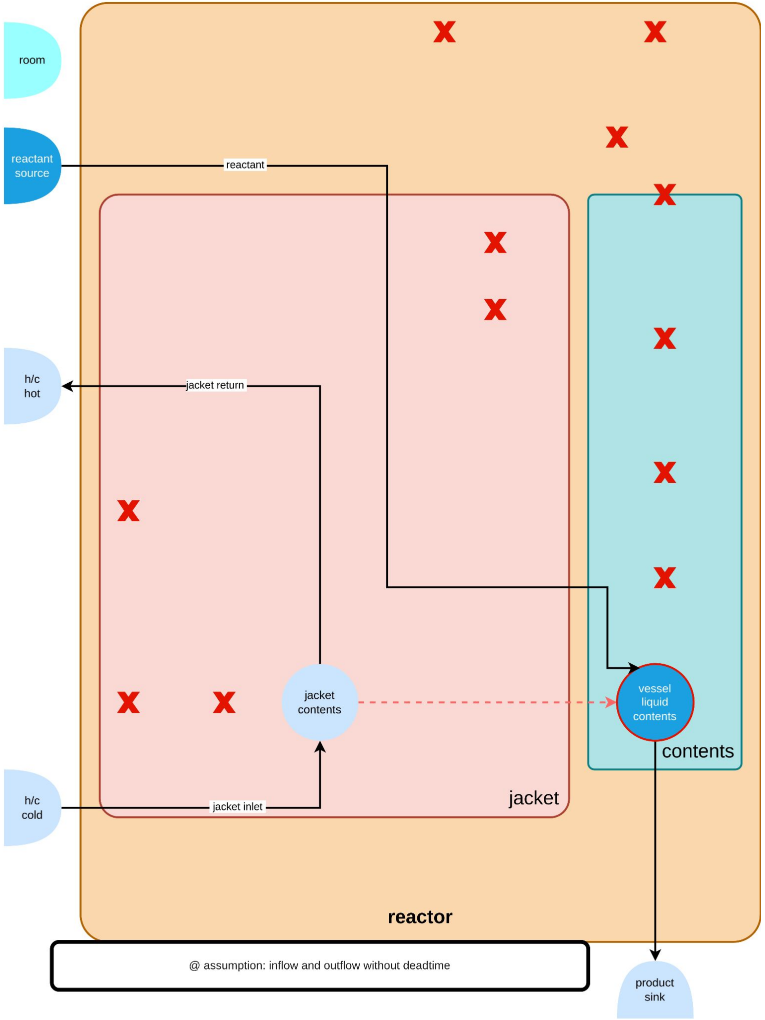


@ assumption: temperature in contents approximately room temperature

- insignificant evapouration from the fluid and negligible heat losses to the room
- outer construction is not actively storing energy
- gas capacity is not active

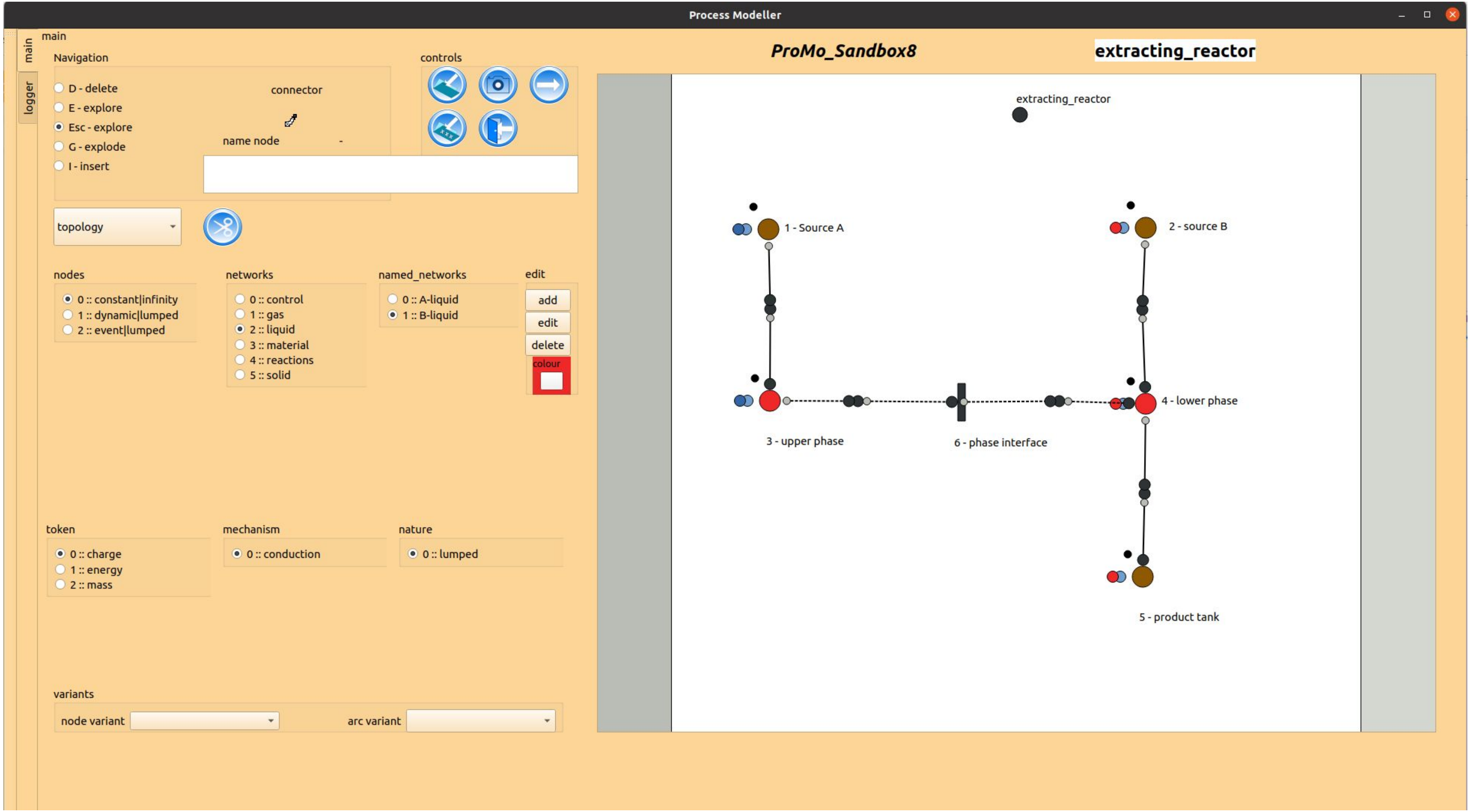


@ assumption: fast heat transfer through the wall
@ assumption: wall capacity negligible
@ assumption: jacket uniform



@ assumption: inflow and outflow without deadtime

Use in model-design software



Process Modeller

main

Navigation

- ○ D - delete
- ○ E - explore
- ● Esc - explore
- ○ G - explode
- ○ I - insert

connector

name node -

topology

ProMo_Sandbox8

extracting_reactor

nodes

- ● 0 :: constant|infinity
- ○ 1 :: dynamic|lumped
- ○ 2 :: event|lumped

networks

- ○ 0 :: control
- ○ 1 :: gas
- ● 2 :: liquid
- ○ 3 :: material
- ○ 4 :: reactions
- ○ 5 :: solid

named_networks

- ○ 0 :: A-liquid
- ● 1 :: B-liquid

token

- ● 0 :: charge
- ○ 1 :: energy
- ○ 2 :: mass

mechanism

- ● 0 :: conduction

nature

- ● 0 :: lumped

variants

node variant

arc variant

ProMo -- Process Modeller

- expert section defines primitive blocks
- translator builds models using the primitive blocks
  and generate higher-level models again as a building block
- all generated building blocks can be reused
- automatic code generation

Software suite ProMo written in Python using pyqt and deployed with ABCdesktop as browser application on https://promo-abcloudtop.io/ as beta release. First release soon. Will be announced on my webpage https://folk.ntnu.no/preisig/

Conclusions

- ModGra is a powerful model design tool
  - as a discussion tool
  - model reduction
  - multi-scale workflow generation tool
  - generation of a model library -- model dissemination