Universiteit van Amsterdam
Instituut voor Informatica

EDITH

Break out session

Integration of Resources

EDITH CSA – 1st Ecosystem Meeting on Building the VHT

Paris18th/19th January 2024

Moderator, Alfons Hoekstra, University of Amsterdam

Schedule

14.00 – 14.10	quick round of introductions
14.10 – 14.30	setting the scene – Alfons Hoekstra
14.30 – 15.30	discussion – diverge! - all
15.30 – 16.00	wrapping up, how to proceed - all

Building the European Virtual Human Twin

Call: Accelerating best use of technologies (DIGITAL-2021-DEPLOY-01)
Work program year: DIGITAL-2021-2022
Topic: ID DIGITAL-2021-DEPLOY-01-TWINS-HEALTH
Grant Agreement No: 101083771

Deliverable 3.2
First draft of the VHT roadmap

Due date of delivery: 31 July 2023
Actual submission date: 31 July 2023

Start of the project: 01 October 2022
End date: 30 September 2024

5 TECHNOLOGY FOR THE VIRTUAL HUMAN TWIN 42
5.1 INTRODUCTION 42
5.2 ORGANISATION OF RESOURCES 42
5.2.1 Multidimensional space as an organisational paradigm 42
5.2.2 Definitions 42
5.2.3 The data object 43
5.2.4 Annotation and annotation services 44
5.2.5 The model object 47
5.2.6 Execution, storage and networking Services 48
5.2.7 Workflow objects 48
5.2.8 The credibility axis in the Virtual Human Twin 49
5.3 DATA 50
5.3.1 Data generation 50
5.3.2 Data use and reuse 51
5.3.3 Data transformation services: harmonising and transcending boundaries 51
5.4 MODELS 53
5.4.1 Models as data transformation services 53
5.4.2 Models as data generation services 54
5.4.3 Models as data flow orchestrations 55
5.4.4 Model classification by context of use 56
5.5 INTEGRATION OF RESOURCES 56
5.5.1 Identification of possibilities for integration 56
5.5.2 Integration of multiscale models 56
5.5.3 Workflows 57
5.5.4 Bidirectional communication with users (including knowledge generation) 58
5.6 INFRASTRUCTURE 58
5.6.1 Catalogue & Repository 59
5.6.2 Simulation platform 60
5.6.3 Computational resources 62
5.6.4 Recommendations for VHT Infrastructure 63
5.7 CONCLUSION 63

RESOURCES

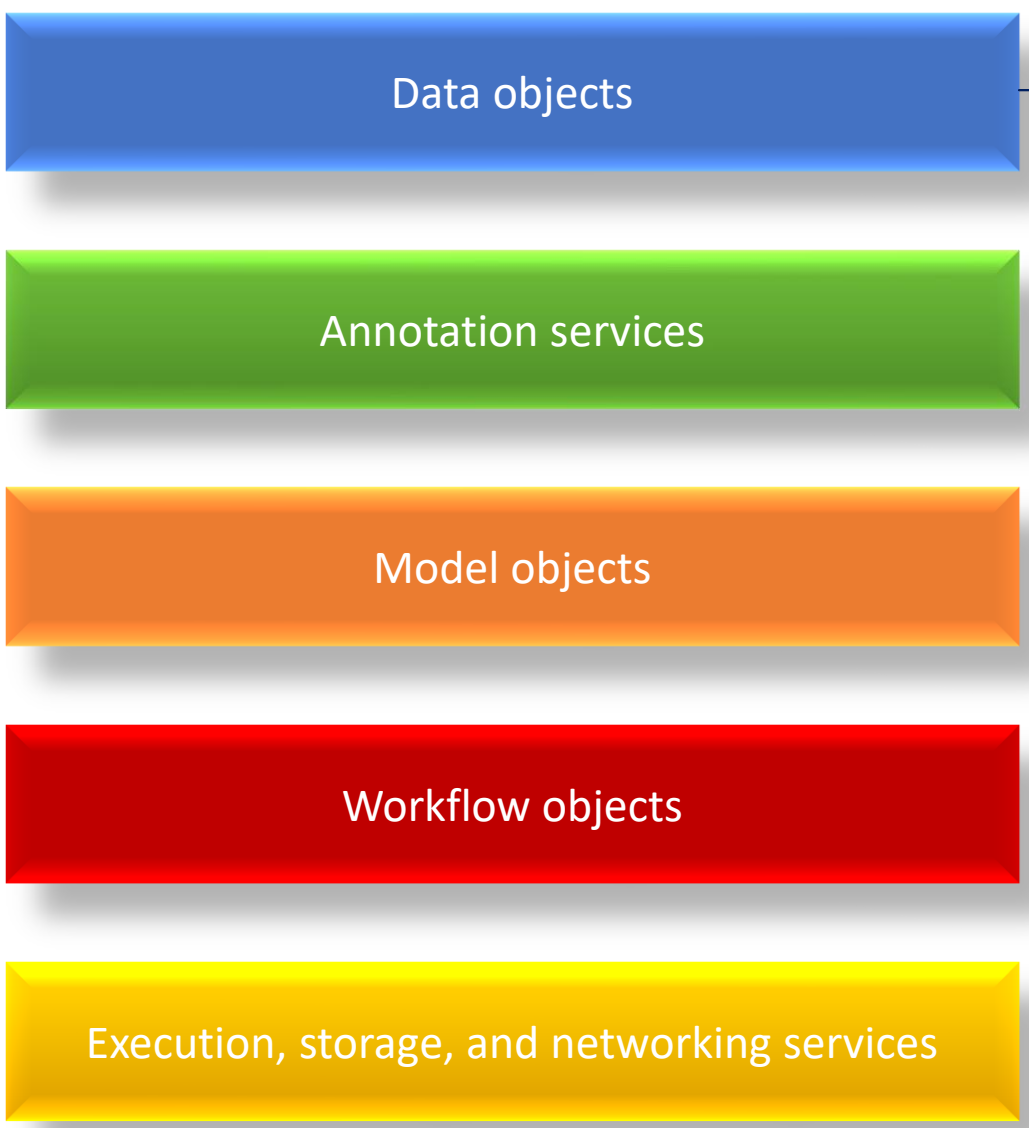
Data objects

Annotation services

Model objects

Workflow objects

Execution, storage, and networking services



The data refer to quantitative measurements

The data refer to human pathophysiology

The data are separated per each individual

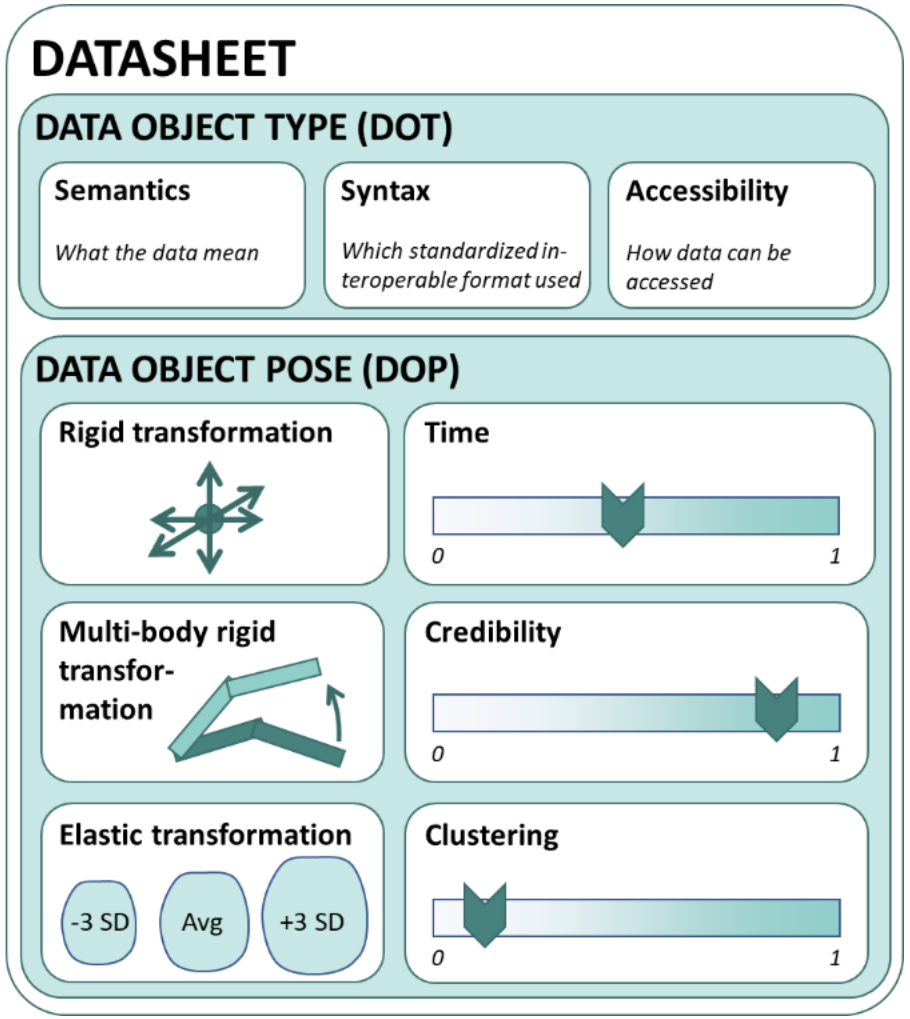
Data objects

Annotation services

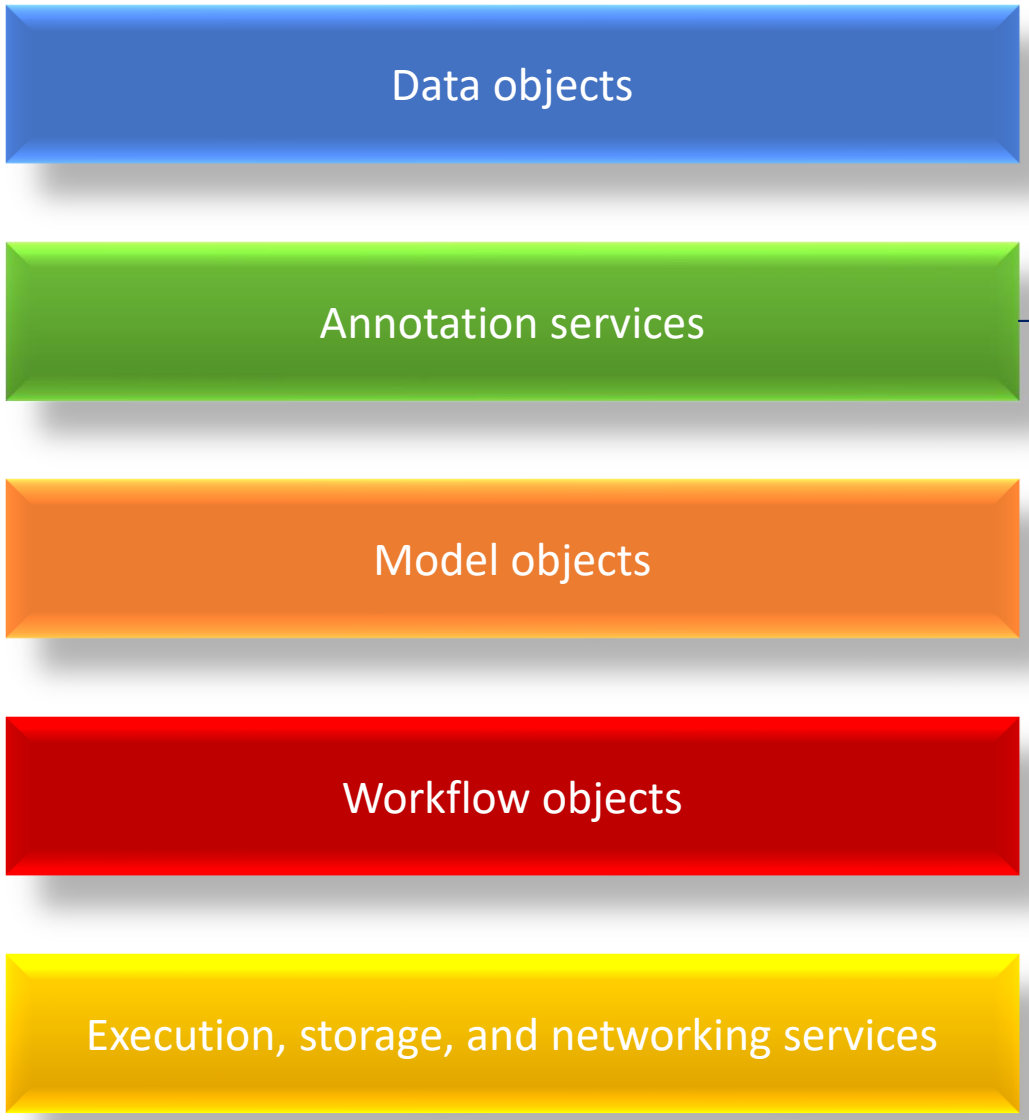
Model objects

Workflow objects

Execution, storage, and networking services

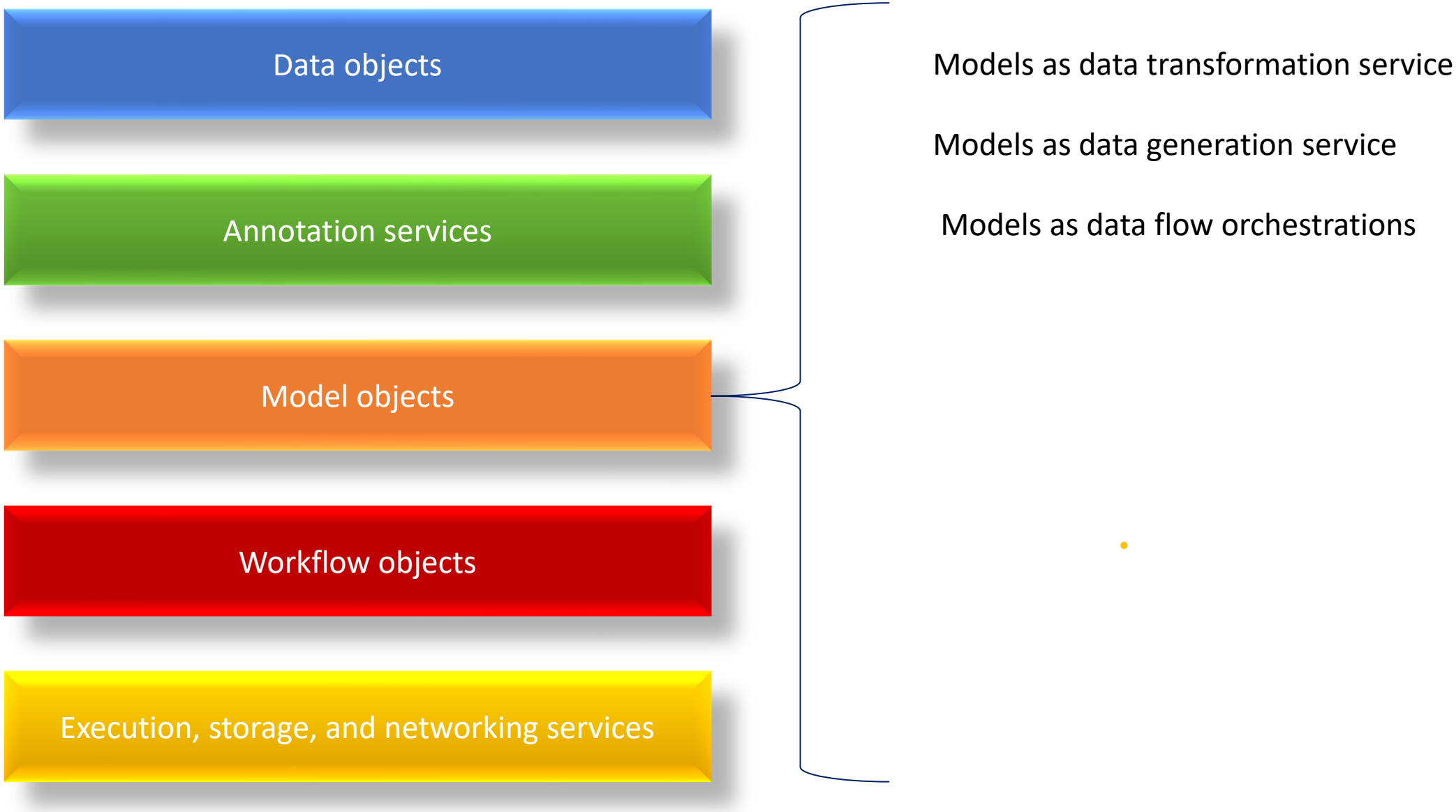


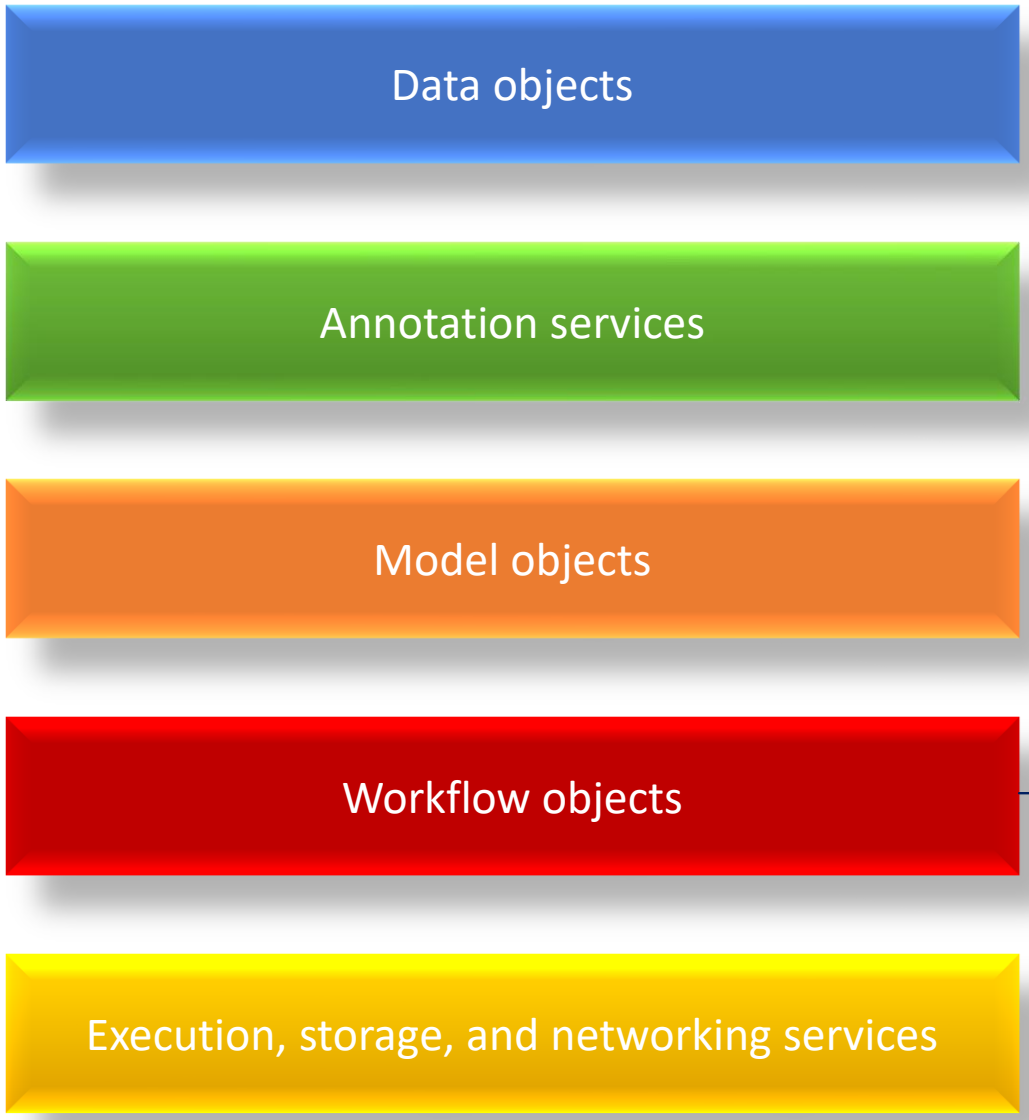
Data transformation services
- Unit conversion
- Format conversion
- Interlinking model, the IO perspective
- Personal to population level transitions
- Segmentation



Annotation of the 6D Data Object Pose

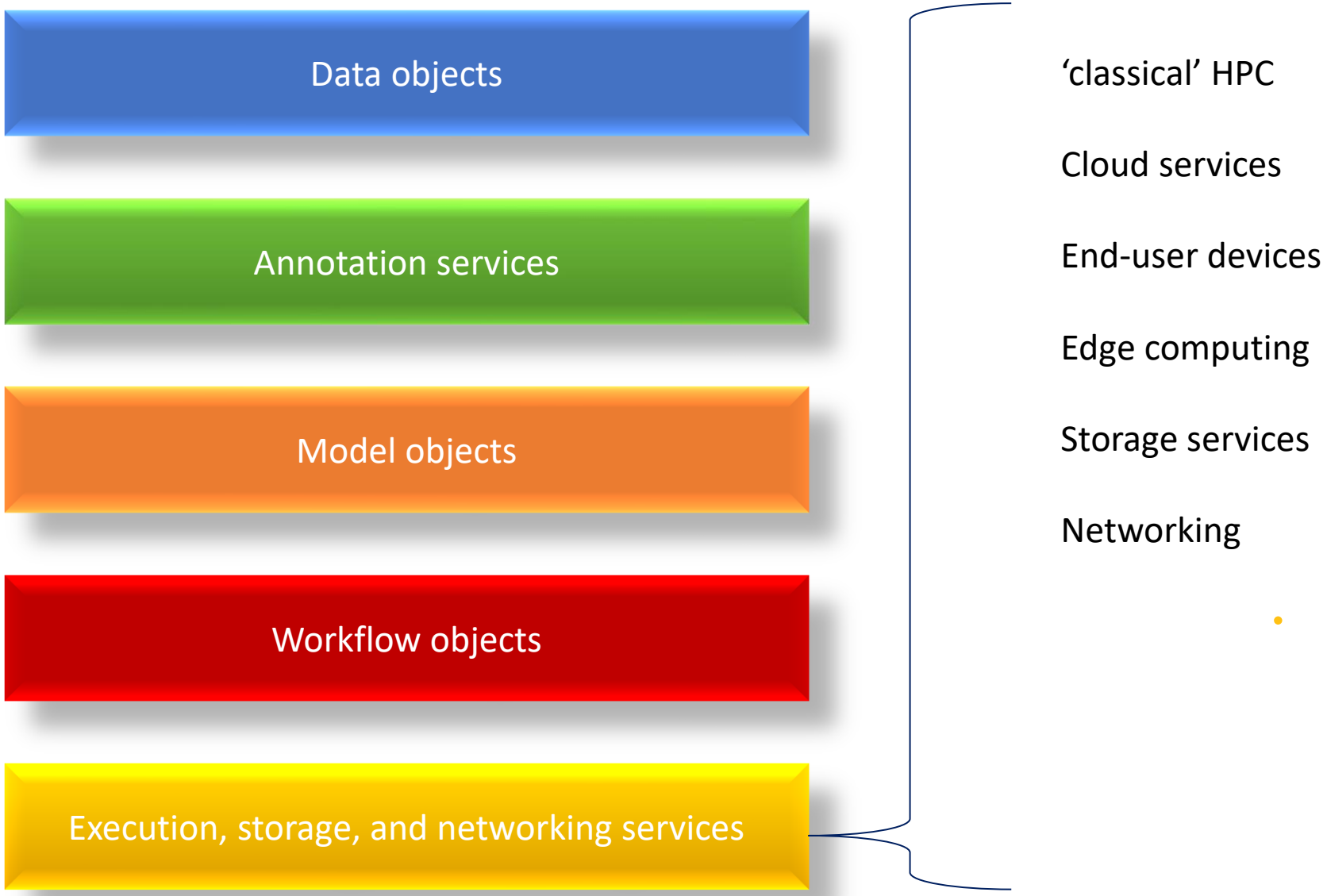
Dimension 1-3: space
Dimension 4: time
Dimension 5: clustering
Dimension 6: credibility


Data objects
Annotation services
Model objects
Workflow objects
Execution, storage, and networking services
Models as data transformation service
Models as data generation service
Models as data flow orchestrations



Orchestration of data, models, data transformation services, annotation services

Mapping on execution, storage, networking services.


Data objects
Annotation services
Model objects
Workflow objects
Execution, storage, and networking services
‘classical’ HPC
Cloud services
End-user devices
Edge computing
Storage services
Networking

Data objects

Annotation services

Model objects

Workflow objects

Execution, storage, and networking services

For details, please see section 5 of the draf roadmap

Diverge ….

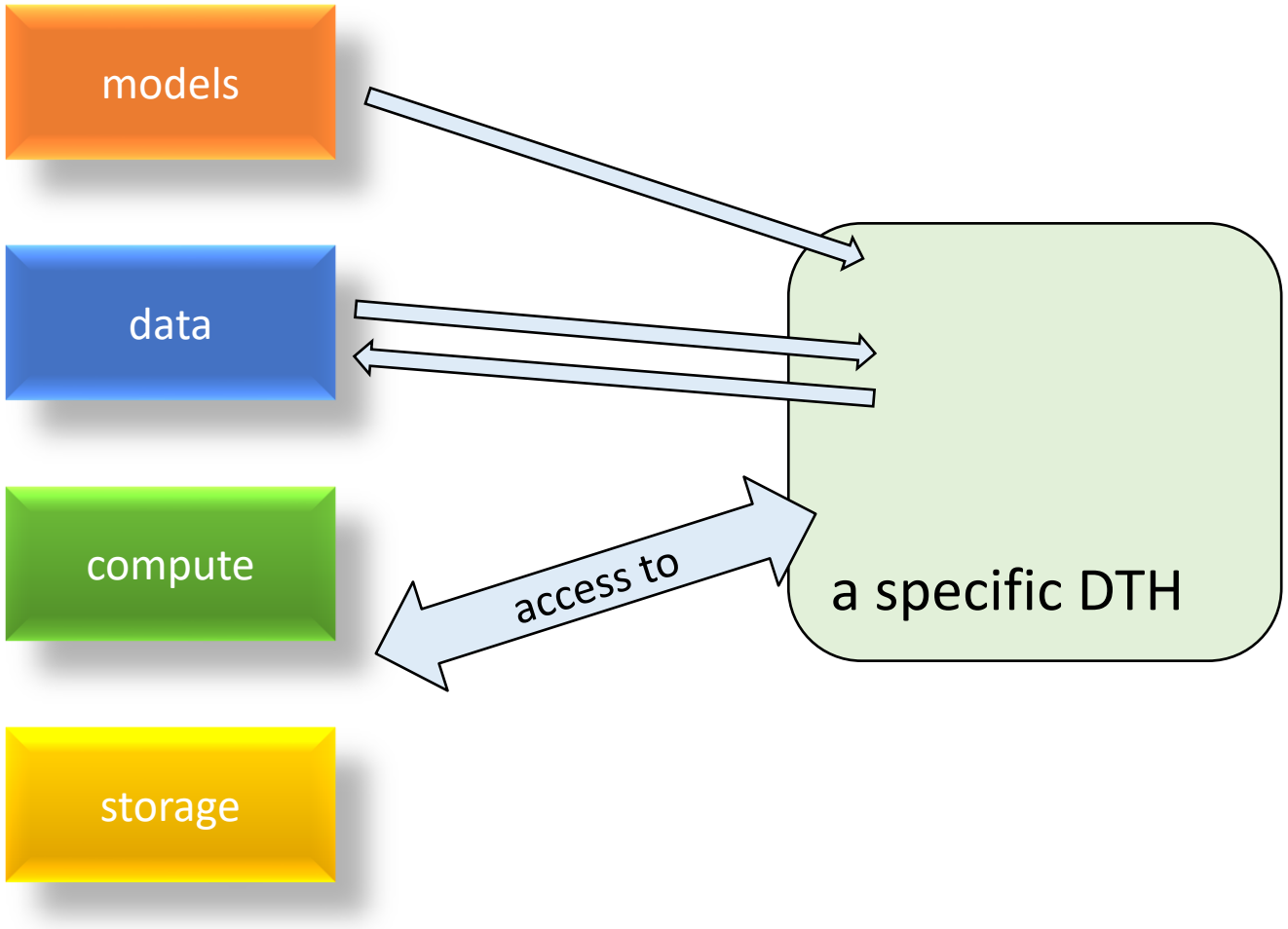
- what did we miss?
- do you disagree with items in draft?

INTEGRATION OF RESOURCES

Integration of resources on two levels

- *Inside* the models/data/compute/storage spaces
  - For *models*, e.g. integration of single scale models into multi-something models
  - For *data*, e.g. pooling of raw, synthetic, transformed, simulated data, including data transformation services, for (stratified) populations or individuals
  - For *compute* and *storage*, e.g. federating some local and remote resources
- *Between* models, data, compute, storage
  - This is actually needed to create a full blown DTH and to execute it.

Integration between models, data, compute, storage



Use workflows to achieve this.

Diverge ….

- what did we miss?
- do you disagree with items in draft?

Questions

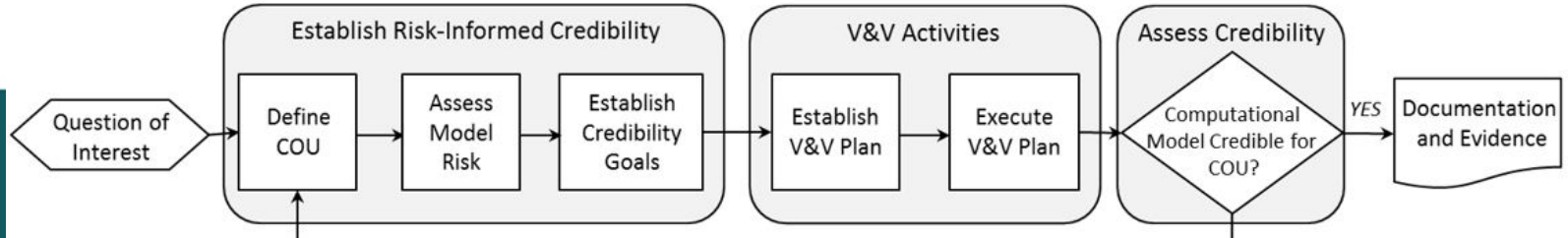
1. Do you agree that we define a workflow as the combination of models and input / output data, dynamically *requesting* access to compute / storage / networking resources?

Questions

1. Will we support a single workflow system, e.g. CWL and build on that, or support requested system?
2. Should we strive for a VHT-workflow standard, leveraging existing standards?

Some more questions

1. Are there prototypical DTH workflows, or standard components for DTH workflows?
   - Generic DTH
   - Population specific DTH
     - E.g. having standard components that automatically check the CoU/QoI for which the DTH is validated and issue warnings when DTH is used outside that context?
   - Subject specific DTH
     - As above, and maybe other functionalities that kick in when moving from population to subject specific?
   - Or for UQ campaigns
     - E.g., maybe each DTH workflow could/should be equipped with automatic non-intrusive UQ (relying e.g. on easyVVUQ developed in the EU-funded VECMA project)?
   - Validation workflows according to V&V40?
     - folllowing ASME workflow



Some more questions

1. Composing DTH workflows
   - Using advanced user interfaces, manoeuvring atlases of human anatomy?
     - Of the quality of e.g. Elsevier's complete anatomy, https://www.elsevier.com/solutions/complete-anatomy
   - Leveraging the 6D framework as backbone?
   - Exploiting advanced knowledge graphs on human (patho)physiology?
     - E.g. Elsevier's Healthcare Knowledge Graph or Biology Knowledge Graph (see https://www.elsevier.com/solutions/biology-knowledge-graph) or comparable efforts.
   - Re-using existing workflows, maybe even automating that?
   - Advanced AI to help, e.g. a ChatGPT like interface, advanced search engines in all available data/models/literature, to propose templates of workflows to be further tailored by DTH developers?

Some more questions

1. Executing DTH workflows
   - Completely automated, hiding all complexity from the DTH user?
   - Automatically sending jobs to most suitable compute resources, pulling data from the right locations, moving data around, invoking dedicated networking infra, etc?
   - For HPC jobs, advanced reservation, collocation of pooled resources, etc?
   - What are VHT specific demands in this respect, if any? Security of data? Other?

Summary of questions

1. Do you agree that we define a workflow as the combination of models and input / output data, dynamically *requesting* access to compute / storage / networking resources?
2. Will we support a single workflow system, e.g. CWL and build on that, or support requested system?
3. Should we strive for a VHT-workflow standard, leveraging existing standards?
4. Are there prototypical DTH workflows, or standard components for DTH workflows?
5. Composing DTH workflows

Diverge … ?