1st ECOSYSTEM MEETING ON BUILDING THE VHT

Contributing to the VHT resources with new use cases: how to?

Sabato Mellone

UNIBO

EDITH

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

Trinity of Software

Catalogue

- A place to share and discover research objects (publications, data, models,…)
- Unique (global) identifiers & versioning
- Metadata (manual, automated) to facilitate discovery
- Actual object may be available (stored) elsewhere
- Catalogue and harvesting services (federation, distribution)

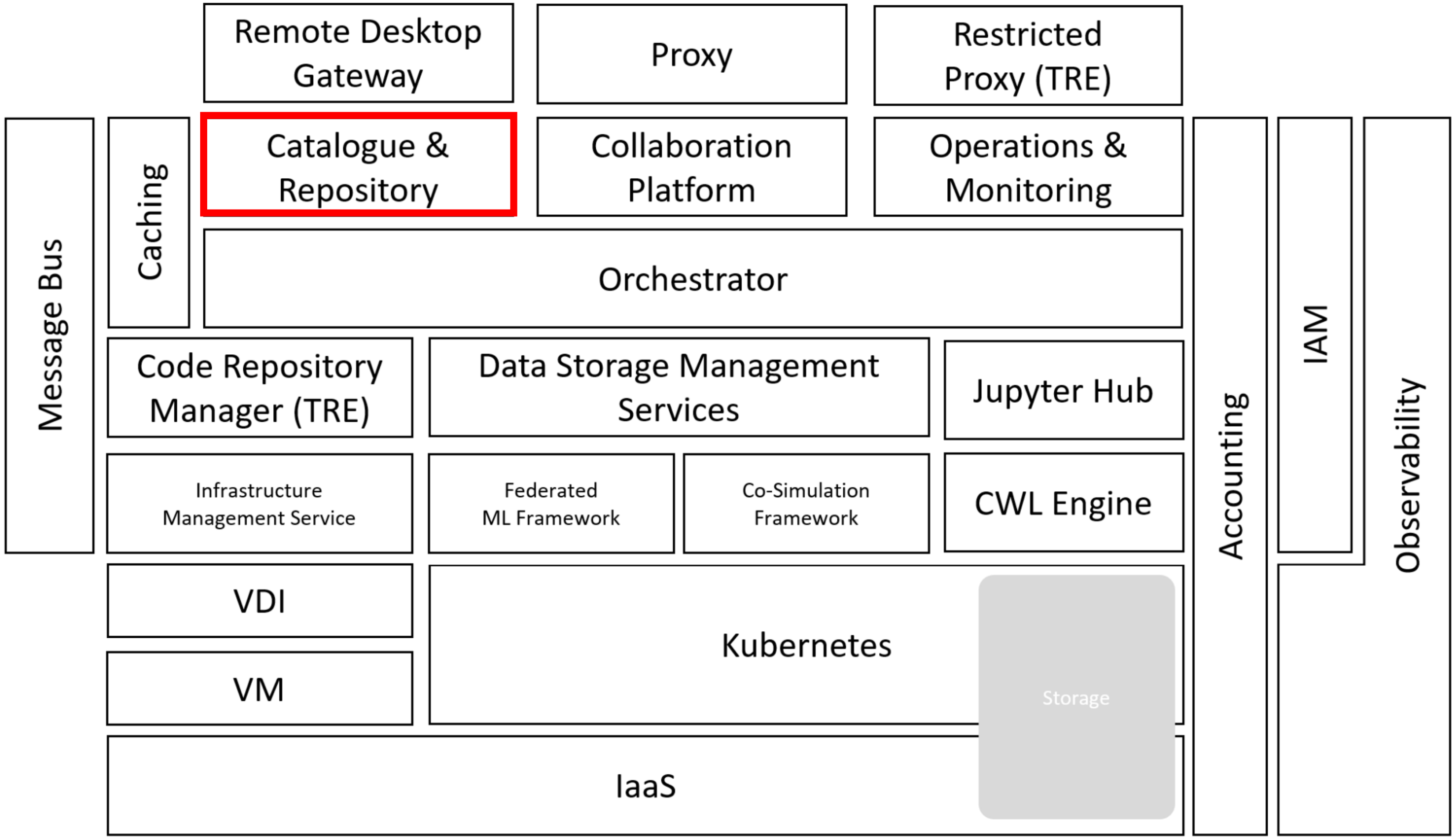
Repository

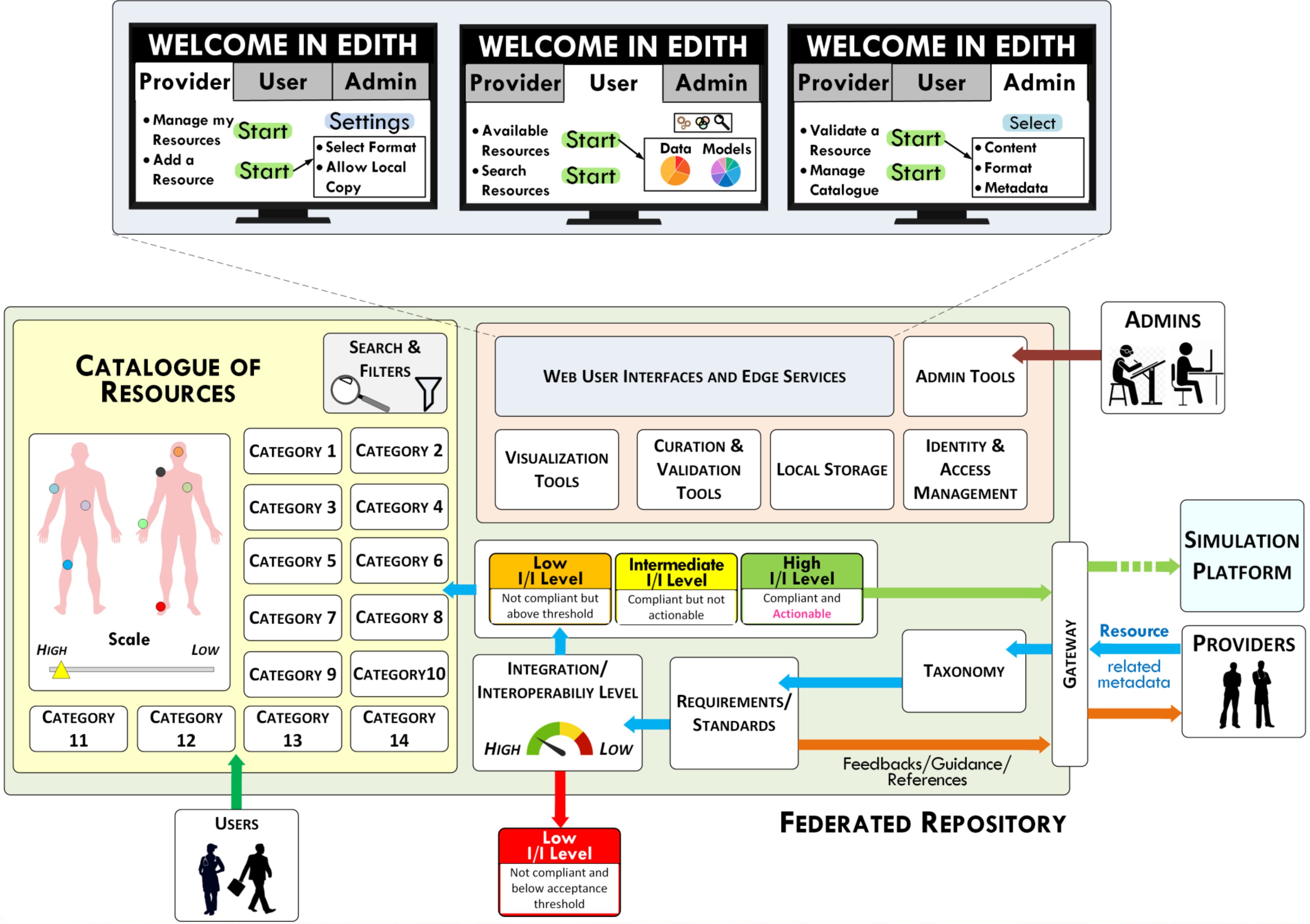
- Safely store and retrieve digital research objects (pubs, data, models,…)
- Files (versioning, metadata)
- Unique (global) identifiers & versioning
- Access policies (and sensitive data)
- Long-term preservation (storage classes)
- File & Sync (personal space)

Platform

- Analyze, simulate, visualize, process, manage, interact, …
- Software services (web apps, APIs, Jupyter notebooks, workflow engines, VDIs)
- Compute (HPC/HTC), storage and networking
- Collaboration and tiers
- Generic-purpose and domain-specific

Architecture



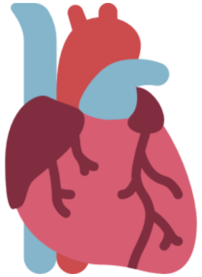


WELCOME IN EDITH
Provider
User
Admin
Manage my Resources
Start
Add a Resource
Start
Settings
Select Format
Allow Local Copy
WELCOME IN EDITH
Provider
User
Admin
Available Resources
Start
Search Resources
Start
Data
Models
WELCOME IN EDITH
Provider
User
Admin
Validate a Resource
Start
Manage Catalogue
Start
Select
Content
Format
Metadata
CATALOGUE OF RESOURCES
SEARCH & FILTERS
Scale
HIGH
LOW
CATEGORY 1
CATEGORY 2
CATEGORY 3
CATEGORY 4
CATEGORY 5
CATEGORY 6
CATEGORY 7
CATEGORY 8
CATEGORY 9
CATEGORY10
CATEGORY 11
CATEGORY 12
CATEGORY 13
CATEGORY 14
WEB USER INTERFACES AND EDGE SERVICES
ADMIN TOOLS
ADMINS
VISUALIZATION TOOLS
CURATION & VALIDATION TOOLS
LOCAL STORAGE
IDENTITY & ACCESS MANAGEMENT
Low I/I Level
Not compliant but above threshold
Intermediate I/I Level
Compliant but not actionable
High I/I Level
Compliant and Actionable
SIMULATION PLATFORM
GATEWAY
Resource
related metadata
PROVIDERS
TAXONOMY
INTEGRATION/ INTEROPERABILIY LEVEL
HIGH
LOW
REQUIREMENTS/ STANDARDS
Feedbacks/Guidance/ References
FEDERATED REPOSITORY
USERS
Low I/I Level
Not compliant and below acceptance threshold

EDITH catalogue – Use a resource

Use a resource

Select the part of the body to explore



Heart

Model #1
[Abstract]
INPUT: …
OUTPUT: …

Ontology: …
Accessibility: …
Body: …
Time (Age): …

Credibility: …
Clustering: …

Model #2
[Abstract]
INPUT: …
OUTPUT: …

Ontology: …
Accessibility: …
Body: …
Time (Age): …

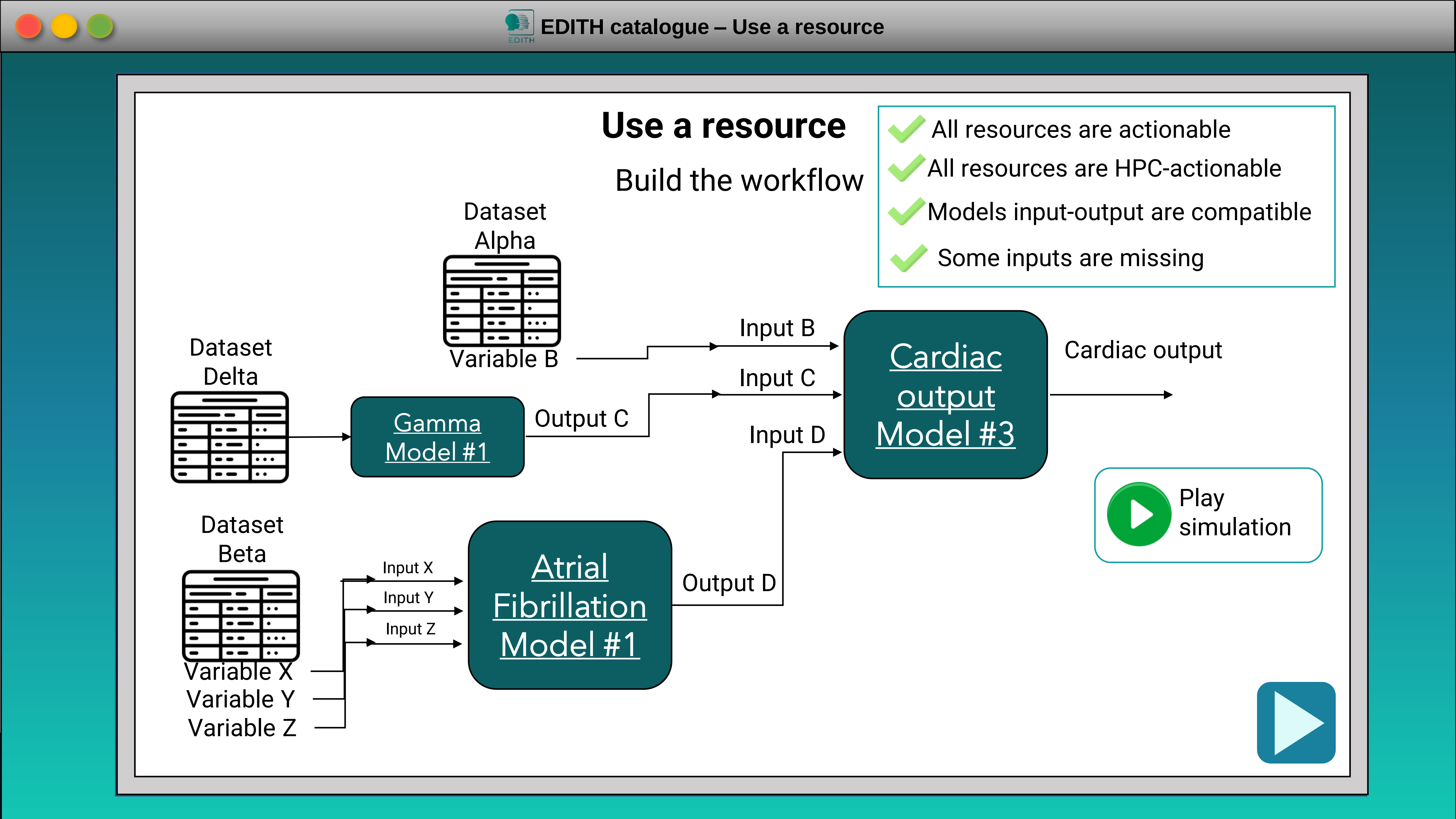
Credibility: …
Clustering: …

Model #3
[Abstract]
INPUT: …
OUTPUT: …

Ontology: …
Accessibility: …
Body: …
Time (Age): …

Credibility: …
Clustering: …

Cardiovascular system



EDITH catalogue – Use a resource
Use a resource
Build the workflow
All resources are actionable
All resources are HPC-actionable
Models input-output are compatible
Some inputs are missing
Dataset Alpha
Variable B
Input B
Dataset Delta
Gamma Model #1
Output C
Input C
Input D
Cardiac output Model #3
Cardiac output
Play simulation
Dataset Beta
Variable X
Variable Y
Variable Z
Input X
Input Y
Input Z
Atrial Fibrillation Model #1
Output D

Compliance and legal considerations

Legal considerations around the governance of the VHT infrastructure involve mainly three aspects:

- Privacy and Data Protection
- Accessibility of the resources (data/model sharing vs federation of resources)
- Licenses and terms of use

Roles and responsibilities

- **Providers**: must be authorised to make that resource available and must specify the terms of use
- **Users**: must be aware of the terms of use and comply with them
- **Administrators:** avoid making available unsuitable digital objects and/or with malicious intent. Ensure that Providers and Users can operate in agreement with their mutual obligations.

Populating the VHT

Call for external use cases

- Set of information required to upload the resource
- Data required about the resource to be uploaded are intended to make clear the terms of use (both for the provider and the user)
- Definition of
  - Terms of use
  - Intellectual property

EDITH wants you!

EDITH

We want to give due prominence to all the resources that are available today for building the Virtual Human Twin. No matter the level of maturity and complexity of what you can make available in our federated repository and catalogue, contact us and let's talk about it. We can decide together what level of integration to achieve.

What I can contribute with:		
A dataset	**A computational model**	**A platform to federate with**
What I need to provide:		
Who you are and your affiliation	Who you are and your affiliation	Who you are and your affiliation
Contact details	Contact details	Contact details
Short description of the dataset	Short description of the model including its context of use	Short description of the platform
Whether there is personal data as defined in the GDPR	Whether there is a licence and/or terms of use attached	Description of its content
If there is personal data specify the data controller and where it is geographically located.	If there is a licence, which kind of licence.	Report the owner and manager of the platform
Specify the data format (e.g. CSV)	Specify the software used (e.g. docker, script with proprietary libraries, custom script, etc.) and/or the programming language (e.g. Python, C#, etc.)	Specify the reference community
If data is in a standard format specify which one (e.g. HL7 FHIR, OMOP CDM, etc.)	Specify the computational requirements (e.g. characteristics of the HPC cluster, workstation hardware, etc.	Specify the available interfacing options (e.g. API)
Link to webpages and/or references to published papers	Link to webpages and/or references to published papers	Link to webpages and/or references to published papers

https://www.edith-csa.eu/call-for-use-cases/

Populating the VHT

Call for external use cases

13 contributions (last update: 21/12/2023)

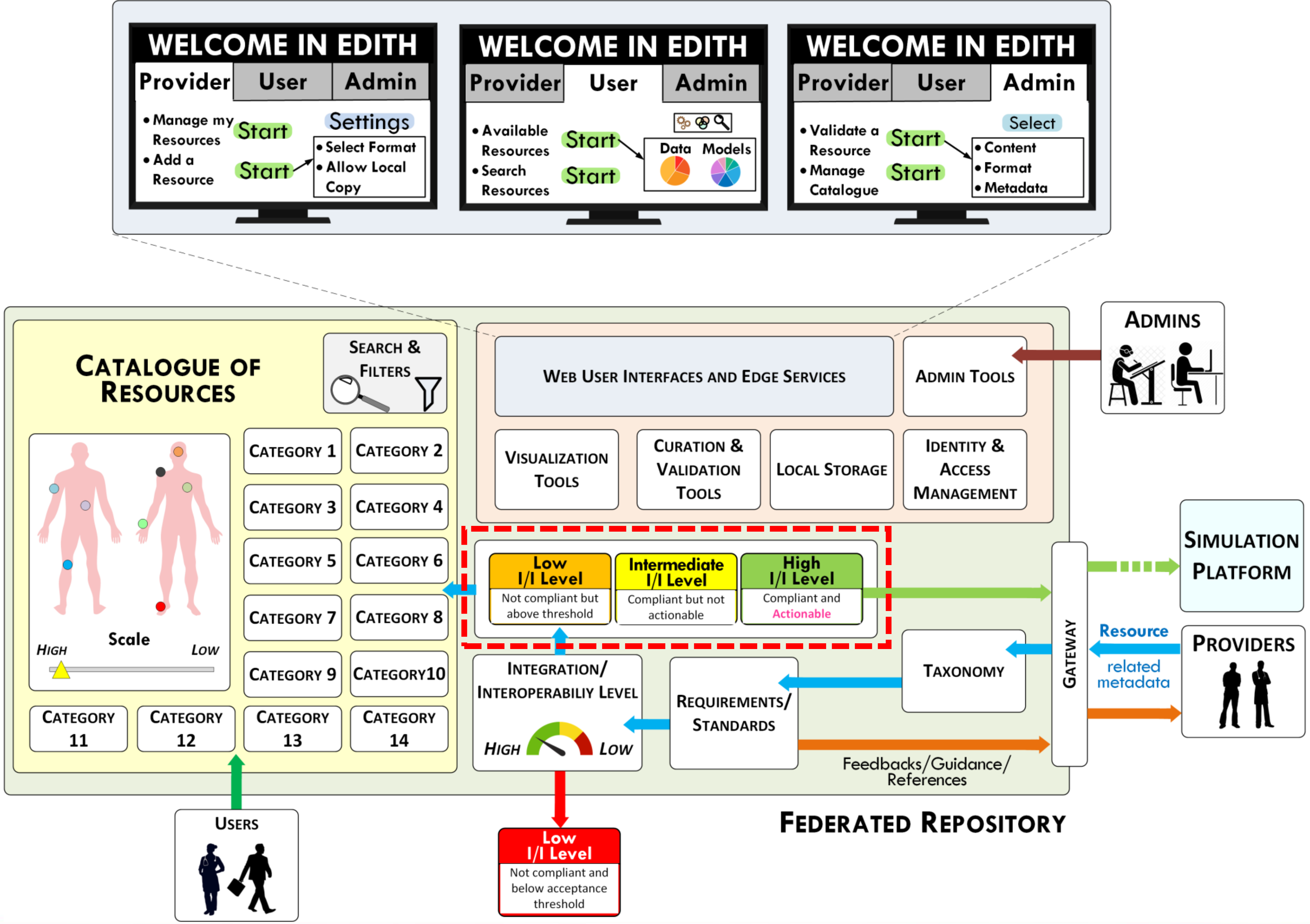
#	Contribution
7	A computational model
3	A platform to federate with
2	A dataset
1	A technology

EDITH wants you!

EDITH

We want to give due prominence to all the resources that are available today for building the Virtual Human Twin. No matter the level of maturity and complexity of what you can make available in our federated repository and catalogue, contact us and let's talk about it. We can decide together what level of integration to achieve.

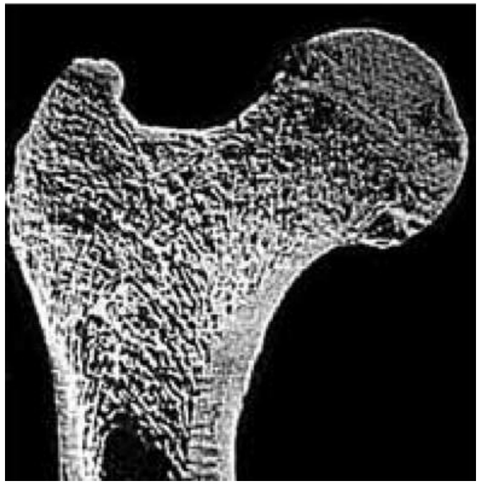
What I can contribute with:		
A dataset	**A computational model**	**A platform to federate with**
What I need to provide:		
Who you are and your affiliation	Who you are and your affiliation	Who you are and your affiliation
Contact details	Contact details	Contact details
Short description of the dataset	Short description of the model including its context of use	Short description of the platform
Whether there is personal data as defined in the GDPR	Whether there is a licence and/or terms of use attached	Description of its content
If there is personal data specify the data controller and where it is geographically located.	If there is a licence, which kind of licence.	Report the owner and manager of the platform
Specify the data format (e.g. CSV)	Specify the software used (e.g. docker, script with proprietary libraries, custom script, etc.) and/or the programming language (e.g. Python, C#, etc.)	Specify the reference community
If data is in a standard format specify which one (e.g. HL7 FHIR, OMOP CDM, etc.)	Specify the computational requirements (e.g. characteristics of the HPC cluster, workstation hardware, etc.	Specify the available interfacing options (e.g. API)
Link to webpages and/or references to published papers	Link to webpages and/or references to published papers	Link to webpages and/or references to published papers

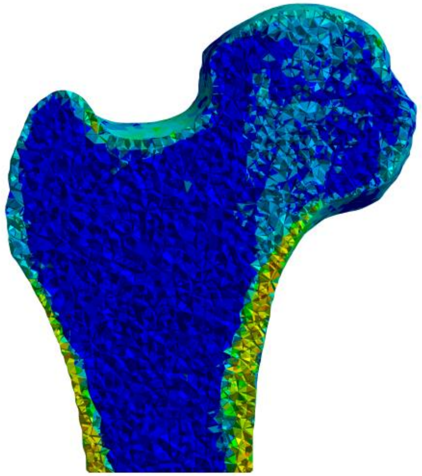

WELCOME IN EDITH
Provider User Admin
Manage my Resources Start
Add a Resource Start
Settings
Select Format
Allow Local Copy
WELCOME IN EDITH
Provider User Admin
Available Resources Start
Search Resources Start
Data Models
WELCOME IN EDITH
Provider User Admin
Validate a Resource Start
Manage Catalogue Start
Select
Content
Format
Metadata
CATALOGUE OF RESOURCES
SEARCH & FILTERS
Scale
HIGH
LOW
CATEGORY 1
CATEGORY 2
CATEGORY 3
CATEGORY 4
CATEGORY 5
CATEGORY 6
CATEGORY 7
CATEGORY 8
CATEGORY 9
CATEGORY10
CATEGORY 11
CATEGORY 12
CATEGORY 13
CATEGORY 14
WEB USER INTERFACES AND EDGE SERVICES
ADMIN TOOLS
ADMINS
VISUALIZATION TOOLS
CURATION & VALIDATION TOOLS
LOCAL STORAGE
IDENTITY & ACCESS MANAGEMENT
Low I/I Level
Not compliant but above threshold
Intermediate I/I Level
Compliant but not actionable
High I/I Level
Compliant and Actionable
SIMULATION PLATFORM
GATEWAY
Resource related metadata
PROVIDERS
TAXONOMY
INTEGRATION/ INTEROPERABILIY LEVEL
HIGH
LOW
REQUIREMENTS/ STANDARDS
Feedbacks/Guidance/ References
FEDERATED REPOSITORY
USERS
Low I/I Level
Not compliant and below acceptance threshold

EXAMPLE

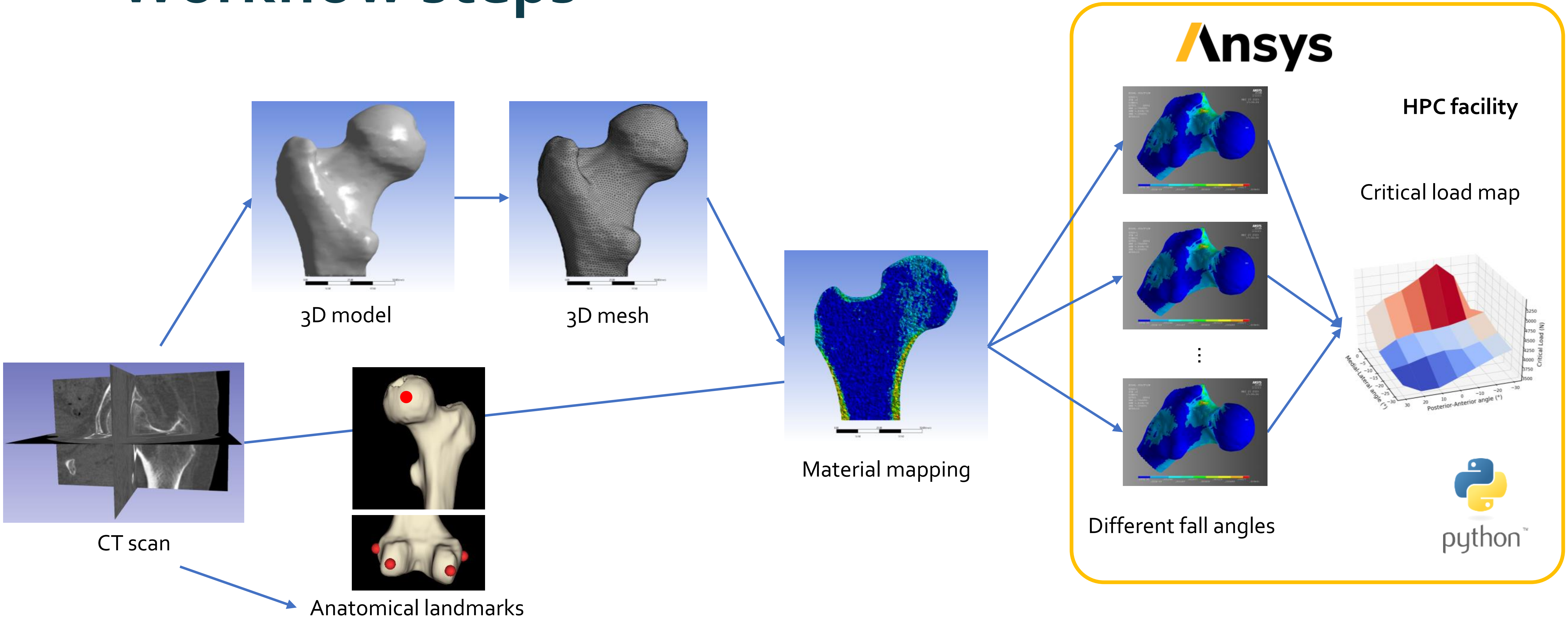
Use case BBCT

- Bologna Biomechanical Computed Tomography (BBCT) is a **Digital Twin** where a patient-specific model predicts **the risk of hip fracture** of an **osteoporotic individual** in case of fall.
- It is a better predictor of fractures than aBMD and can be used to:
  - identify the patients at high risk -> promptly start the treatments -> **reduced femur fracture incidence**
  - evaluate the efficacy of new antiresorptive drugs -> **reducing antiresorptive drug development costs and time-to-market.**

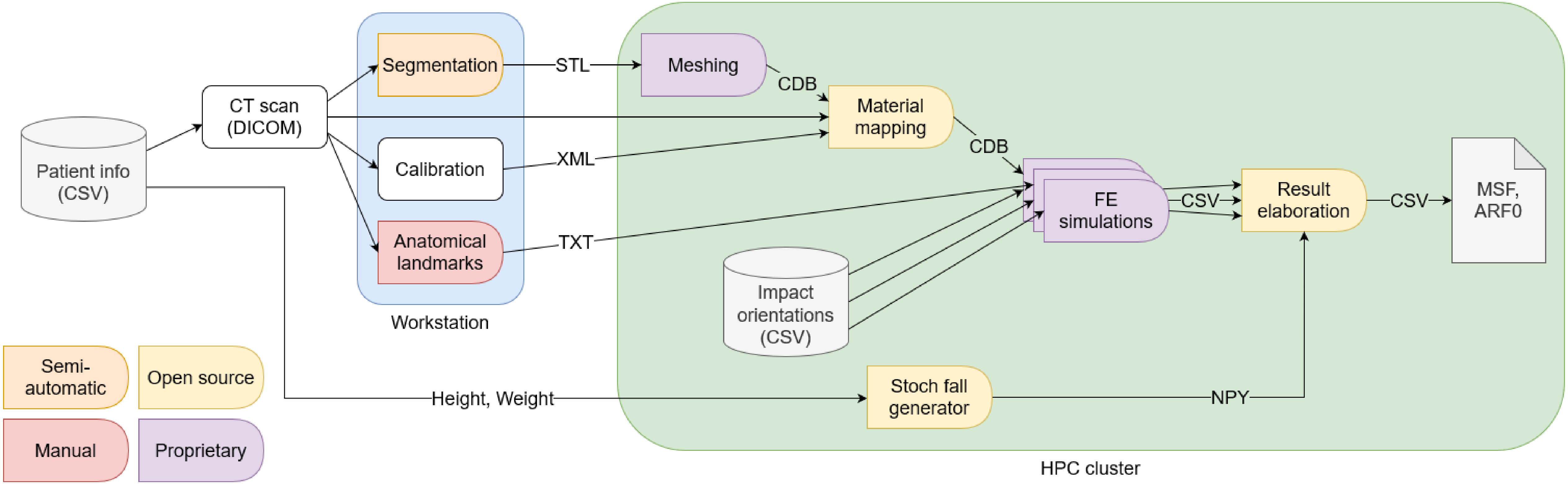




Workflow steps



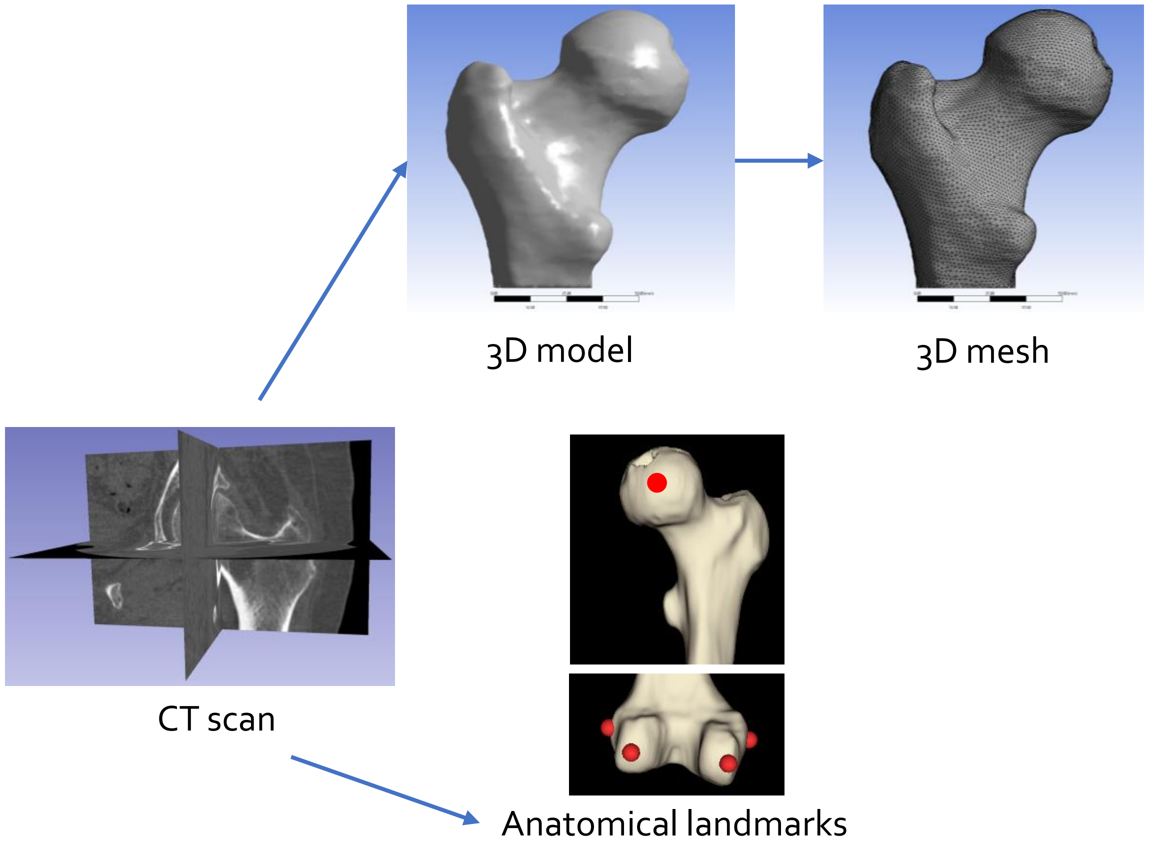
Data flow



The operator retrieves the femur CT scan and segments it and identifies the anatomical landmarks for the standard reference system on his/her workstation. The operator can then upload the CT scan, the STL file and an XML configuration file (for CT scan calibration) on the HPC cluster and submit the job indicating patient height and weight.

At the end of the simulation, the operator receives an email, and can download the folder with FE simulation results and the ARF and MSF values.

1. Image retrieval



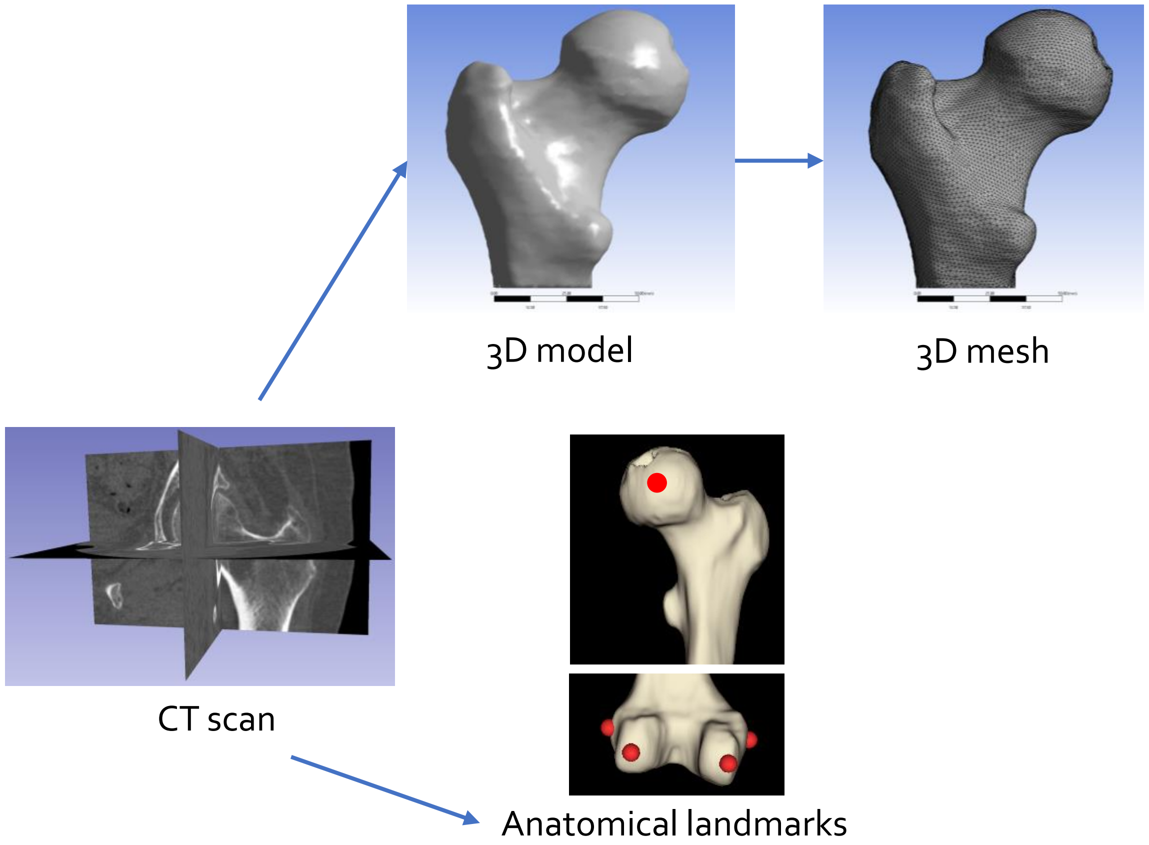
- **NOW**

The CT scans are currently (manually) uploaded to a secure shared folder on a storage server behind Rizzoli (research hospital in Bologna) firewall.
CT scans are manually copied from the server to the operator workstation.

- **OPTIMAL/EXPECTED setup**

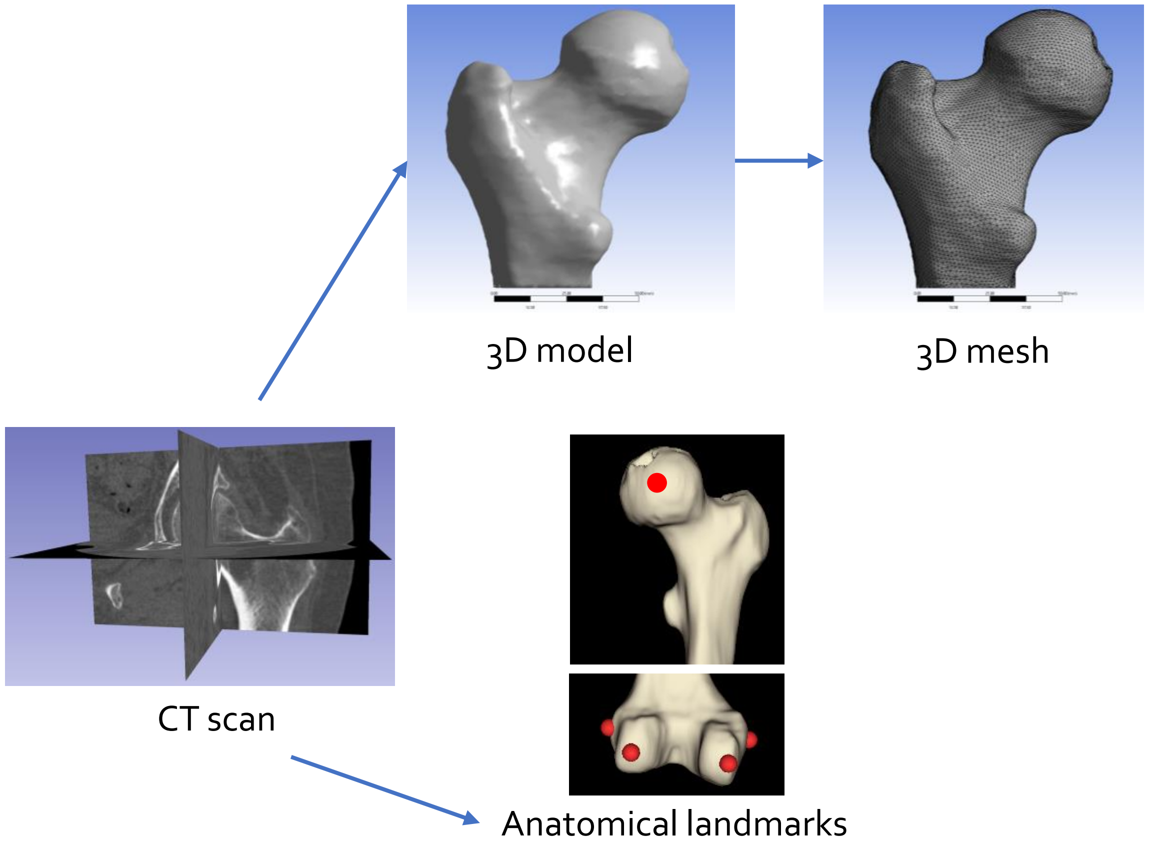
In future, automatic/semi-automatic process to request and transfer the CT scans from the hospital PACS server directly on dedicated workstations within the hospital firewall
Replace manual segmentation with a validated automatic algorithm.

2. Segmentation



- **Inputs**: **DICOM** files of the femur CT scan.
- **Outputs**: **STL** file with femur geometry.
- **Software involved**: custom Python scripts; 3DSlicer for input and final correction of the segmentation.

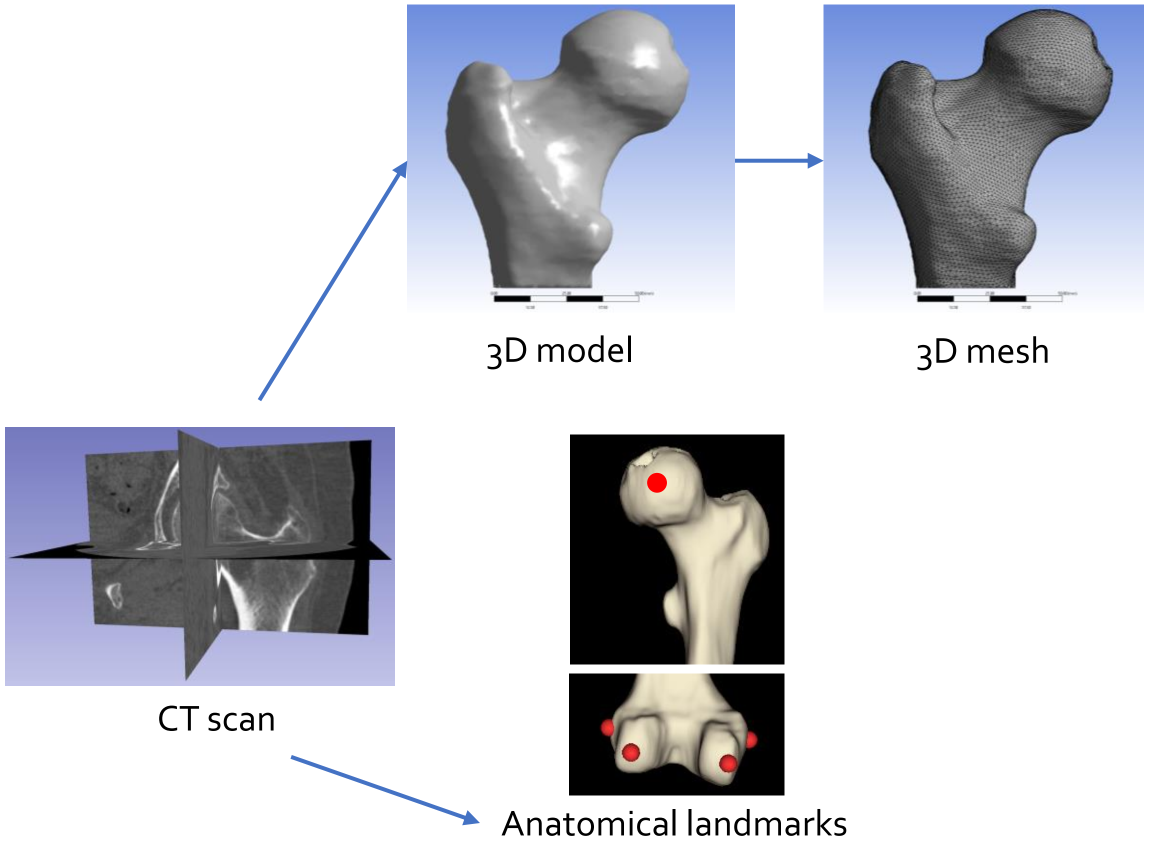
3. Anatomical landmarking



- **Inputs:** DICOM files of the demur CT scan; STL file with femur geometry.
- **Output:** INP file (simple text file in the format "variable = number") with anatomical landmarks for standard femur orientation.
- **Software involved**: Matlab on the workstation (optional, could be replaced with another software).

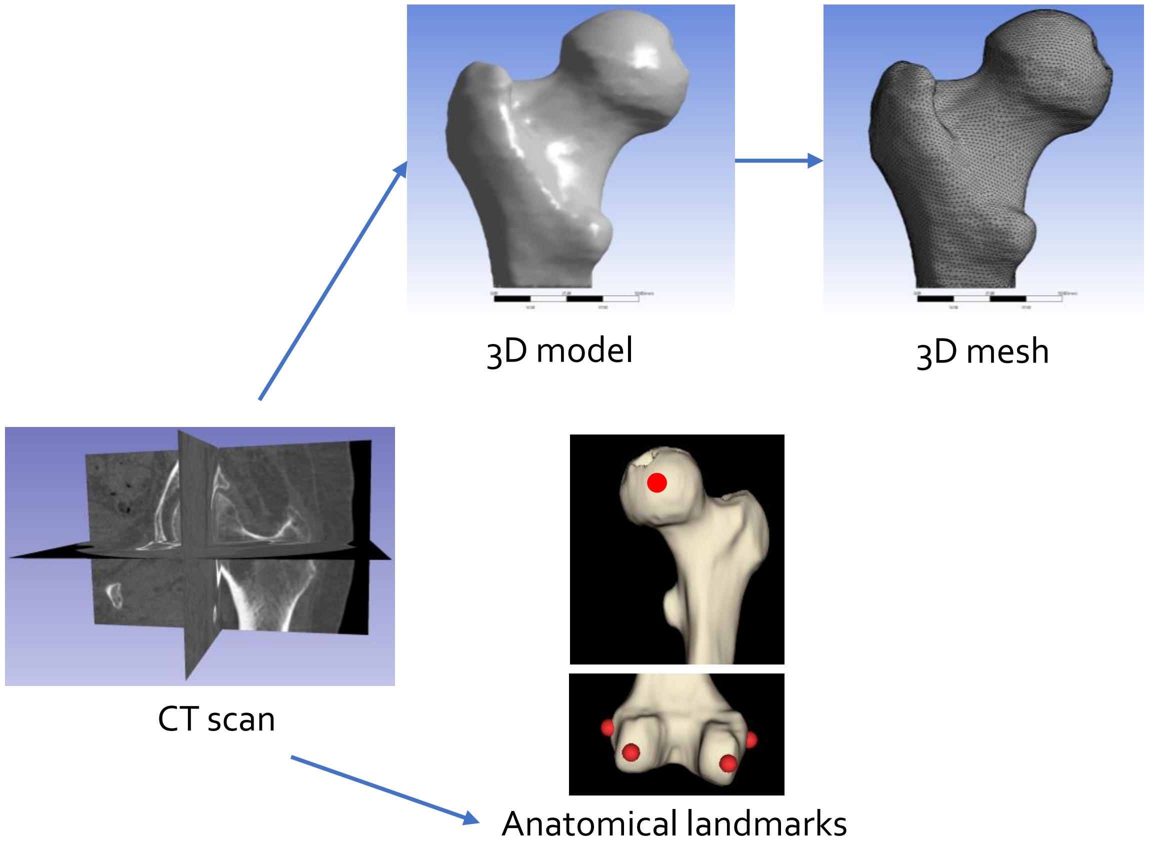
Optimal/Expected setup: Automatic landmarking with a validated algorithm.

4. Meshing



- **Input:** STL file of femur geometry.
- **Output:** CDB file (Ansys standard) of tetrahedral mesh (only geometry).
- **Software involved:** Ansys ICEM CFD (from STL to linear mesh) and Ansys Mechanical APDL (mesh refinement).

4. Meshing

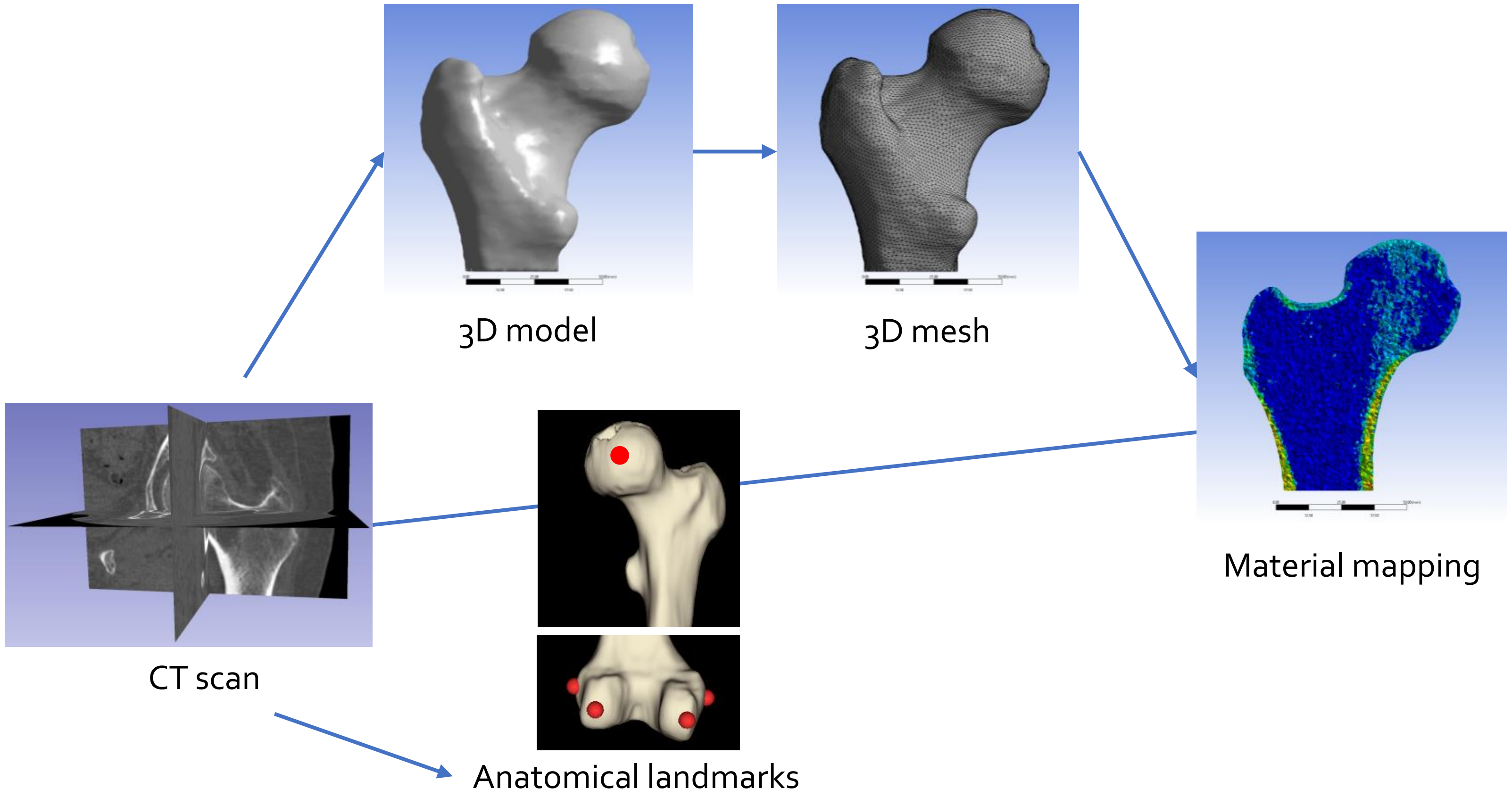


Programming languages

- TCL for Ansys ICEM CFD;
- APDL for Ansys Mechanical APDL;
- Bash for glueing

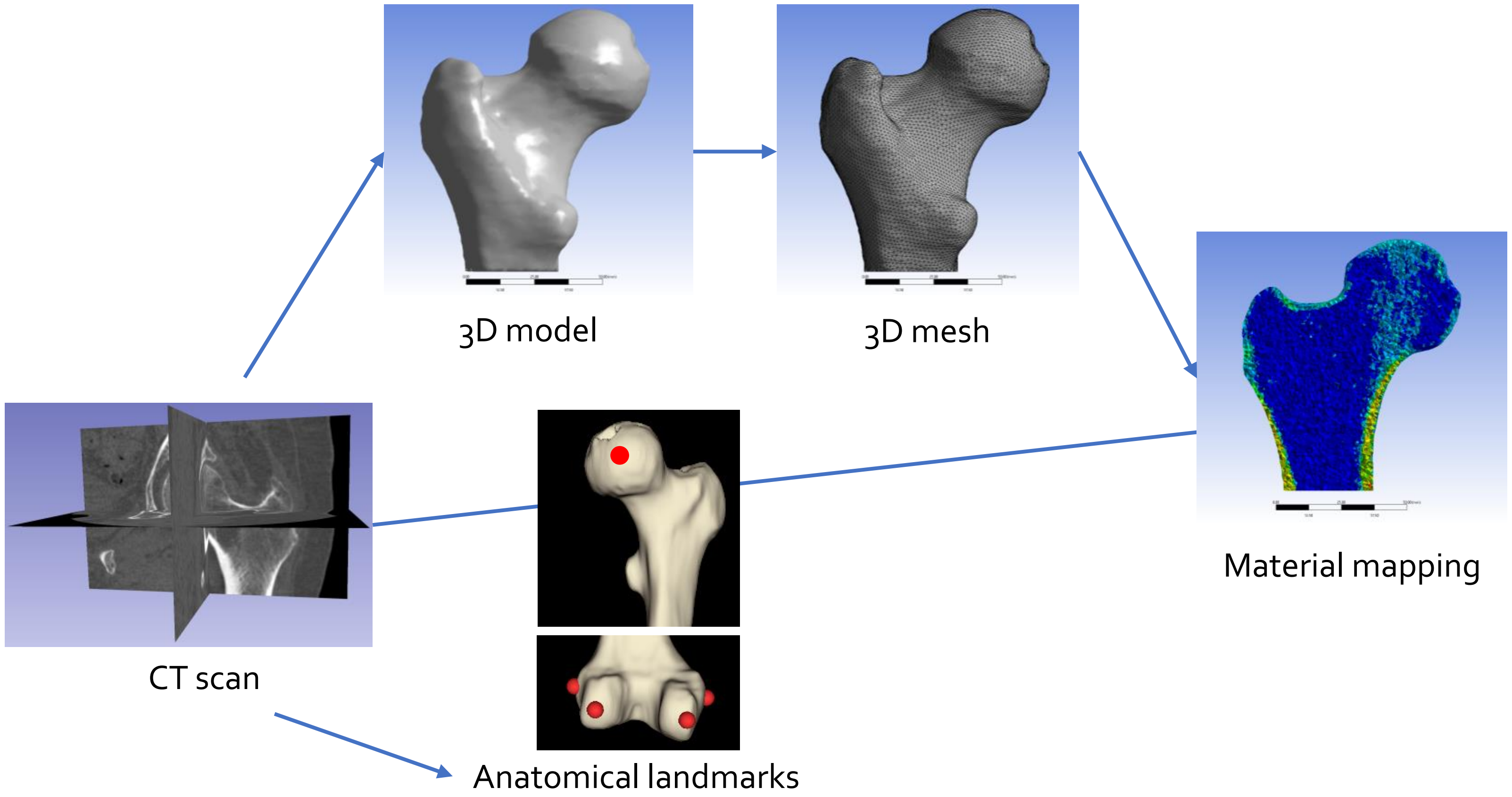
This step can be performed on the operator workstation or directly on a HPC cluster. If run in batch mode, it can require up to a couple of minutes.

5. Material Mapping



- **Inputs**: **CDB** file of refined tetrahedral mesh; **DICOM** folder containing CT scan; **XML** file (Bonemat-like, very simple) with configuration parameters.
- **Output**: **CDB** file with geometry and material properties.

5. Material Mapping

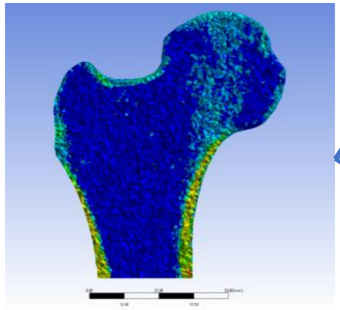


- **Software involved**: custom Python program; modules needed: NumPy, SciPy, Numba, PyDicom.

This step can be performed on the operator workstation or on a HPC cluster. It requires a couple of minutes.

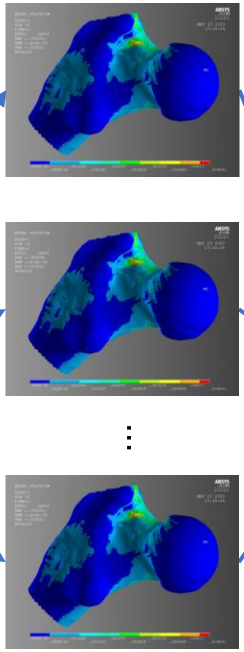
6. Stochastic falls generation

- **Inputs:** height (in cm); weight (in kg).
- **Output:** NPY (NumPy archive) file with an array of 1 million floats (possible fall loads).



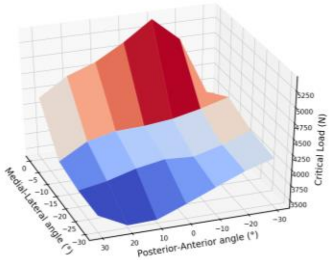
Material mapping

HPC facility



Different fall angles

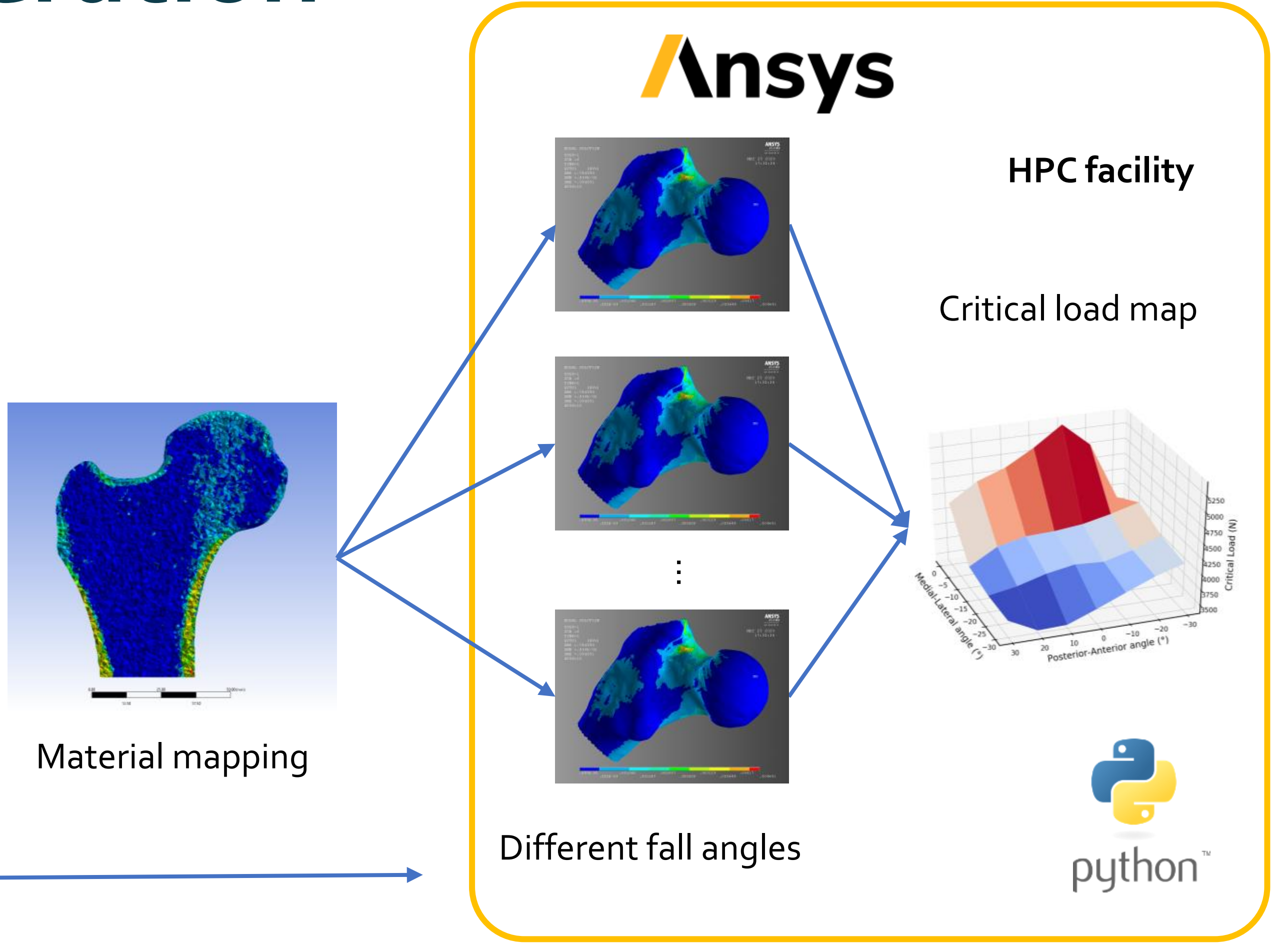
Critical load map



6. Stochastic falls generation

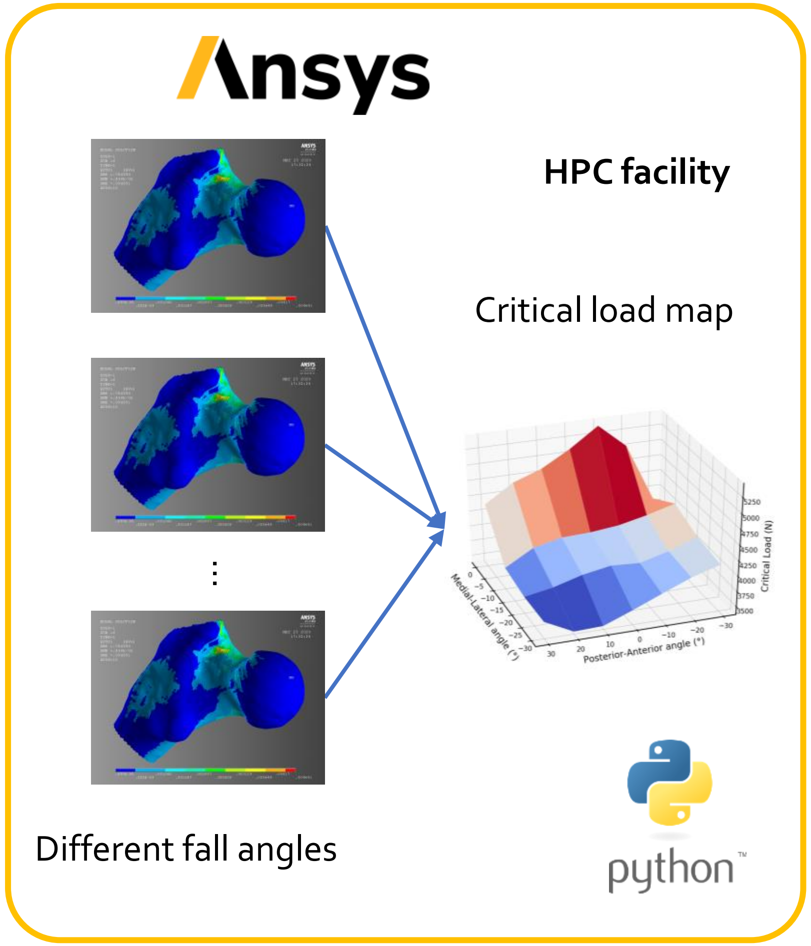
- **Software involved**: custom Python program; modules needed: NumPy, SciPy, Scikit-learn (optional).

This step can be performed on the operator workstation or on a HPC cluster. It run in seconds.



7. FE simulations

- **Inputs**: **CDB** file with geometry and material properties; **INP** file (simple text in the format "variable = number") containing the coordinates of points to standardise femur orientation; **NPY** file with possible fall loads; posAntAngle and medLatAngle (angles for impact direction in degrees); orientation name (simple string).
- **Output**: **CSV** file with tensile and compressive strains and coordinates of ROI nodes; for performance monitoring, some text logs of Ansys; **DAT** file (simple one-line CSV with critical load, fracture risk and failure location for the specific impact force orientation).
- **Software involved**: Ansys Mechanical APDL; custom Python script; modules needed: NumPy, SciPy.



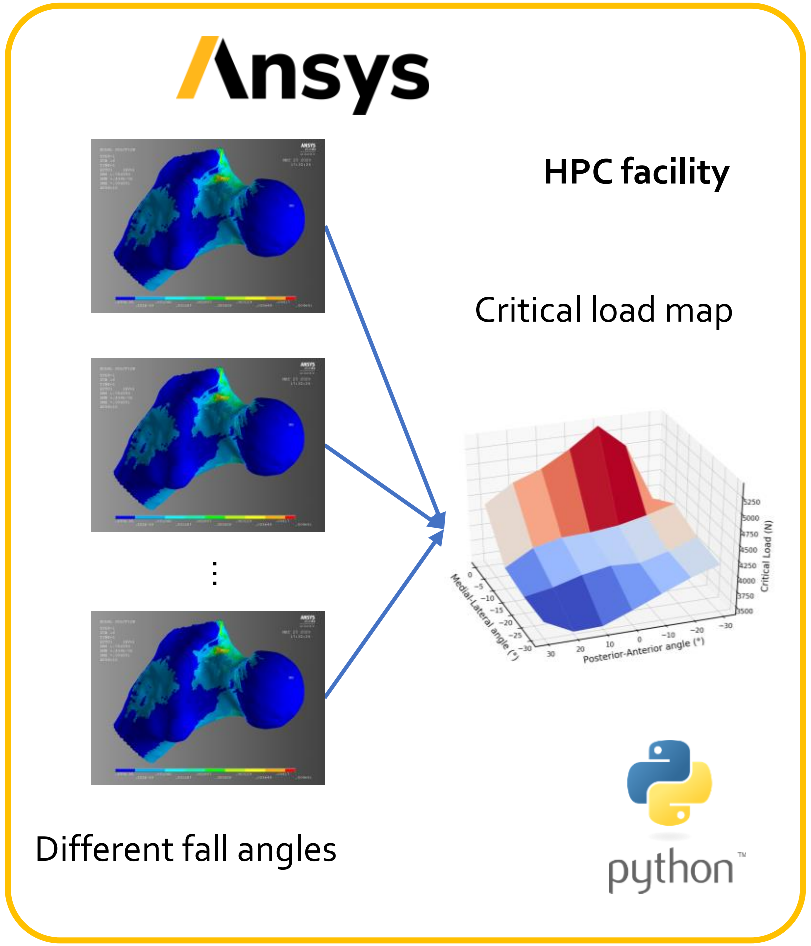
7. FE simulations

Programming Languages:

- APDL for Ansys Mechanical APDL;
- Python for post-processing;
- Bash for glueing.

Each angle simulation run on 4-6 cores within 1 hour, scalability tests to estimate the execution times with more compute cores are ongoing. The execution time also depends on the size of the model: if only proximal femur is considered in the simulation the execution time drops to 30 minutes with 3 cores. Here we need to run some basic tests.

This step is usually performed on a HPC cluster.

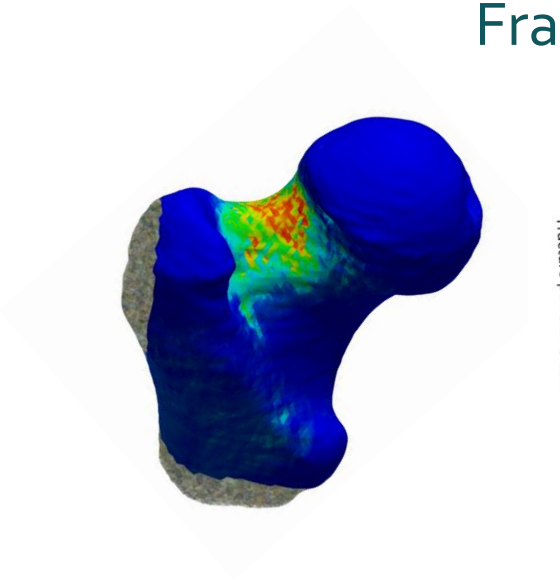


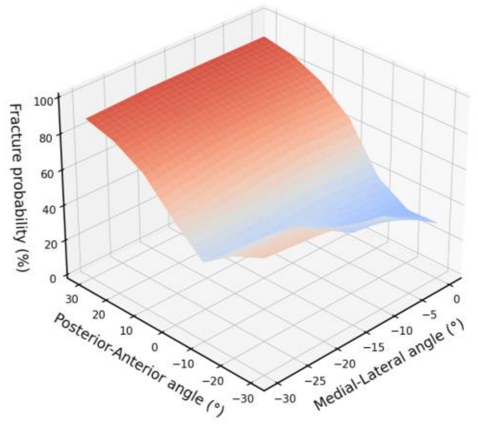
8. Collection and analysis

- **Inputs:** DAT files from FE simulations
- **Outputs:** DAT files with Absolute Risk of Fracture at time 0 (ARF0, in percentage) and Minimum Side-fall Strength (MSF, in N).
- **Software involved:** custom Python script; modules needed: NumPy, Scipy, Pandas, Matplotlib (optional).

This step can be performed on the operator workstation or on a HPC cluster. It runs in seconds.

Absolute Risk of Fracture





Minimum Side-fall Strength

HPC clusters 1/2

The most updated version of the workflow has been tested on Galileo100 cluster, hosted by CINECA. Some parts of the workflow were also deployed and tested on Cartesius (Tier-1, out of production, hosted by SURF, national HPC centre in The Netherlands), Snellius (Tier-1, SURF), and Galileo (Tier-1, out of production, CINECA). **All these clusters use SLURM as job scheduler.**

Each node of Galileo100 is equipped with 2 Intel Xeon Platinum 8260 (CascadeLake) processors with 24 cores each, and 384 GB of RAM (~7.2 GB per core).

HPC clusters 2/2

The workflow mainly consists of platform-agnostic blocks, as it is composed by custom Python scripts and proprietary multiplatform software (Ansys suite), for which the script can be easily ported from Windows to Linux.

Caveat: Ansys is not available for IBM PowerPC architectures and is officially tested only on Intel and AMD CPUs. Also, it does not officially support all the Linux distros: the only tested by Ansys company are RedHat, CentOS (formerly), and SUSE (not OpenSUSE).