# A semantically adaptable integrated visualization and natural exploration of multi-scale biomedical data

Ricardo Manuel Millán Vaquero*, Asan Agibetov†, Jan Rzepecki*, Marta Ondrésik‡§, Alexander Vais*,
Joaquim Miguel Oliveira‡§, Giuseppe Patanè†, Karl-Ingo Friese*, Rui Luís Reis‡§, Michela Spagnuolo† and Franz-Erich Wolter*

*Division of Computer Graphics, Leibniz University of Hannover, Hannover, Germany
Email: (rmillan,jrzepecki,vais,kif,few)@gdv.uni-hannover.de
†Istituto di Matematica Applicata e Tecnologie Informatiche, Consiglio Nazionale delle Ricerche, Genova, Italy
Email: (agibetov,patane,spagnuolo)@ge.imati.cnr.it
‡3B's Research Group, Biomaterials, Biodegradables and Biomimetics, Headquarters of the European Institute
of Excellence on Tissue Engineering and Regenerative Medicine, University of Minho, Caldas das Taipas, Portugal
§ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal
Email: (marta.ondresik,miguel.oliveira,rreis)@dep.uminho.pt

*Abstract*—The exploration of biomedical data which involves heterogeneous sources coming from different spatial scales and medical domains is a challenging topic in current research. In this work, we combine efforts regarding multi-scale visualization, multimodal interaction and knowledge formalization for the exploration of multi-scale biomedical data. The knowledge formalization stores and organizes the information sources, the integrated visualization captures all relevant information for the domain expertise of the user and the multimodal interaction provides a natural exploration. We present a concrete example of use of the proposed exploratory system designed for a biologist investigating multi-scale pathologies.

*Keywords*-Multi-scale biomedical exploration; knowledge formalization; ontology; natural exploration

## I. INTRODUCTION

Pathologies which simultaneously involve information relating diverse spatial scales and medical domains, such as cardiovascular, neurological or musculoskeletal diseases are often difficult to investigate. Osteoarthritis is an example, where cartilage degradation [1], [2] starts with changes evidenced at molecular and cellular scale, leads to alterations in mechanical tissue properties at the organ scale, and finally affects the range of motion of the knee at behavioral scale [3]. The expertise and data regarding this pathology is distributed among a diverse range of medical practitioners and researchers, who work on individual features involving a concrete domain and spatial scale but typically lack access to an overall picture.

Biomedical data has multiple properties that demonstrate its heterogeneity. In particular, data sets are acquired from a large variety of acquisition modalities, have numerous data formats and span several spatial scales, e.g. micro-CT scan of meniscus or cartilage, associated graphs and histological images (tissue scale), magnetic resonance imaging - MRI (organ scale) and gait motion analysis (behavioral scale).

The presentation of this plethora of information is difficult due to the massive amount of data and the diversity of its visualization properties [4], [5]. Multi-scale biomedical visualization aims at the combination of such heterogeneous biomedical data in order to provide an appropriate presentation for exploration.

Moreover, finding the relevant factors for the characterization of potential pathologies generally requires not only data but also knowledge belonging to the involved medical domains, such as radiology, biology, biomechanics and tissue engineering. Biomedical ontologies, whose role is to define the concepts and relationships between information sources that are relevant for modeling a medical domain, facilitate the organization of heterogeneous and multi-scale information sources.

This work is an extension of the exploratory system described previously in [6], augmenting it by incorporating elements of the framework proposed in [7], which introduced the combination of knowledge formalization and an integrated visualization environment in order to 1) encode multi-scale biomedical information with the help of an underlying ontology and 2) capture all data, independently of its heterogeneity, into a single view. In this work we combine efforts regarding multi-scale visualization, multimodal interaction and knowledge formalization for the exploration of multi-scale biomedical data. The proposed system tackles aspects regarding the exploration of multi-scale biomedical data, which has been addressed by several EU funded projects, such as VPH [8], MSV [9] and MSH [10]. A main focus is on the introduction of adaptive features to the environment in order to allow the specialist to explore the data according to his interest, presenting the relevant data sets and depicting the established knowledge about them. This is technically performed by customizing the underlying queries of the system to the knowledge database. The organization,

presentation and interaction facilitate the task of evaluating scientific hypotheses for the specialist.

*Outline.* First, in Sec. II, we introduce related work for biomedical exploration on multi-scale visualization, multi-modal interaction and knowledge formalization. In Sec. III we present an example scenario of the investigation of multi-scale pathologies, describing the traditional exploration methodology, and presenting the benefits of introducing our approach. In Sec. IV, we present our methodology, explaining how to organize and filter heterogeneous data based on knowledge formalization. In Sec. V we finally describe our integrated environment that captures all information relevant for the domain expertise of the user and provides a natural presentation and exploration.

## II. RELATED WORK

The interplay of domains of science is beneficial for a complete analysis of biological phenomena [8]. In the last years, researchers called for efforts to improve the exploration of biological data, emphasizing the advantage of integrated frameworks to gain insight into the biological processes [11]. In this context, multi-scale biomedical visualization, knowledge formalization and natural interaction play significant roles for the presentation, organization and exploration of multi-scale biomedical data.

*Multi-scale biomedical visualization* pursues the coupling of biomedical data coming from different spatio-temporal scales for a simultaneous presentation [12]. Authors promoted visualization techniques able to deal with a complete range of possible biological data types [13] and the formation of alliances across biomedical domains and visualization communities [14], [15], [16].

Recent works followed those premises. For instance, the MSV Project addresses interactive visualization paradigms for biomedical multi-scale data [9]. The Multimod Application Framework provides support of biomedical time-varying data [17]. Moreover, the traditional distinction between InfoVis and SciVis, and therefore the utilization of their respective techniques has begun to merge, leading to new trends of combined visualizations [18], [19].

However, the aforementioned works reveal that further improvements in terms of visualization are needed in order to substantially facilitate scientists' tasks during multi-scale biomedical exploration. For instance, standard graphical user interfaces (GUI) do not provide a simple method for transition between scales [17] and traditional approaches which conserve spatial dimensions often prevent important small scale data to have an adequate visibility in large scale views [9].

The multi-scale exploration of data sets becomes even more difficult because it does not provide per se semantic means to understand relations between data, such as the pathological relation between two data sets. Relationships generally refer to knowledge spread over several medical domains. Knowledge formalization can provide the formal tools to support the storage of knowledge and the extraction of relationships among multi-scale biomedical data for a subsequent presentation.

*Knowledge formalization*'s goal is to represent knowledge about a certain domain in such a way that a computer can perform the process of reasoning on objects of that domain in a way that a human would [20]. When dealing with data management issues, however, knowledge has to be expressed in a way that makes it applicable and useful in computational contexts. As defined in the semantic web community [21], ontologies are a means to identify relevant items in a given domain and formally define the properties or attributes that are necessary to document them in a machine-interpretable way. Using ontologies to structure and organize heterogeneous data and knowledge opens up new perspectives to medical investigation. Ontologies provide a standard way of representing terms and concepts, thereby supporting easy transmission and interpretation of data for various applications [22].

Applications of ontologies in the biomedical domain range from definitions and classifications of common medical terms (e.g. SNOMED CT [23], Gene Ontology [24], FMA [25], or GALEN [26]), to explicit specifications for a focused domain such as radiology, e.g in Radlex [27]. These ontologies have been developed in order to enable interoperability across multiple medical information systems. A detailed survey of biomedical ontologies may be found in [28].

Most of the biomedical ontologies have available implementations in the Web Ontology Language (OWL) [29]. The OWL data model is given by the Resource Description Framework (RDF) [30], consisting of so-called *RDF triples* $\langle subject, predicate, object \rangle$, which may be represented as a directed graph. Usually ontologies are queried through the standard querying mechanism SPARQL [31], which relies on RDF graph pattern matching.

However, ontologies alone are not enough to provide interactive exploration of multi-scale biomedical data and medical background knowledge [7]. Smart visualization means are needed for the interactive visualization of the ontology, supporting users to direct themselves easily to the focus of their interests, both at the conceptual level (ontology browsing) at the data level (scientific visualization of patient specific data).

*Natural interaction* aims at providing a human–machine interface that is not perceived by the user to be disruptive. Before computers became popular in the medical domain, clinicians, radiologists and researchers were accustomed to interact with patient specific data on physical, tangible media (e.g. photographic plates or film for x-ray radiography). The introduction of digital processing and storage of patient files lead to a clash between this traditional interaction and a digital exploration through keyboard and mouse control that is often perceived as artificial and limiting. Progress made

during recent years has made hand gesture interaction a more intuitive, flexible and robust alternative [32], [6]. Direct hand manipulation and exploration avoids the discrepancy between traditional and digital data exploration, freeing users from necessity to learn how to interact with data, supported by easily understandable feedback. Several previous works are focused on bringing more intuitive and more natural biomedical exploration through multimodal interaction. For a survey we refer to [33].

## III. APPLICATION SCENARIO

We present an example scenario for the investigation of multi-scale pathologies, focusing on osteoarthritis (OA). After overviewing essential features of OA in Sec. III-A, we summarize the traditional methodology of investigation applied by biologists in Sec. III-B. Based on established knowledge, biologists formulate and validate hypotheses using the evidences found in the data obtained in their experiments. However, the traditional exploration methodology does not facilitate this task. This motivates our proposal for an integrated exploratory system for automatising the organization and exploration of experimental data, which we outline in Sec. III-C and Sec. III-D, respectively.

### A. Osteoarthritis as multi-scale pathology

Osteoarthritis is an example for a multi-scale pathology. It firstly and mostly affects the articular cartilage, causing its complete degeneration [1], [2]. Changes in cellular behavior and the death of chondrocytes is a common feature of an osteoarthritic cartilage and can be evidenced through cell viability assays [34].

The altered cellular behavior consequently causes the disruption of the macromolecular tissue network at the scale of organ, which results in increased compressive stiffness and eventually in the softening of the cartilage, evidenced by variations in the mechanical properties. In more advanced phases the cartilage degradation can also be observed on MRI as thinning and progressive loss of the tissue. Finally, at the behavioral scale, the failure of cartilage will lead to joint stiffness and ultimately affect the gait pattern [3].

This model is graphically summarized in Fig. 1, where each of the large oval boxes denotes a so–called degradation process feature (DPF). Each DPF such as *loss of biomechanical function* is evidenced by several sources of evidence (SOE), e.g. *swelling*. Moreover each DPF is typically placed on a specific *spatio–temporal scale* (indicated with orange font). The arrows indicate how SOEs affect each other. Finally, some SOE are associated with corresponding measurement techniques, e.g. *nanoindentation of cartilage*.

### B. Traditional exploration methodology of biomedical data

In the following, we describe the current methodology of a biologist when dealing with multi-scale data.

1) The biologist comes up with a hypothesis $H$ of cartilage degradation, e.g. referring to Fig. 1:

   *H: The cytokine induced cell death and shift in metabolic activity will result in the disruption of extracellular matrix (ECM) integrity and inappropriate mechanical function of the cartilage.*

   This hypothesis requires investigation on several scales e.g. the cell death has to be investigated on both cellular (live/dead assay) and tissue scale (histology), while the degeneration of cartilage can be seen on tissue scale (histology) as well as on the organ scale (scan analysis of the knee). For further background on the underlying biological mechanisms, we refer to [35].

2) To support or refute this hypothesis, various measurements are performed, which produce a large amount of multi-scale and multimodal data. For example, numerous acquisition sessions are performed on individual specimens, during which several data sets (e.g. PET-CT scans) are acquired. The collected data refers to anatomical entities (e.g. femur, tibia, patella, menisci or articular cartilage).

3) During analysis, some data is processed for diagnosis or simulation purposes (e.g. understanding the structural geometry of cartilage in 3D from sections in 2D by using [36], [37] otherwise not being visible). This processing allows relevant parameters (e.g. cartilage thickness) to be calculated and evaluated.

4) The studies of the biologists are manually registered in a log book including the protocols, the steps of their optimization and also the results. To interpret the data the biologist first performs a statistical analysis and then evaluates whether the acquired data supports or refutes the assumption.

After assessment of all data, the hypothesis of cartilage degradation during osteoarthritis across biological scales may either be supported by the data that satisfies the causality assumptions of the hypothesis, or may be refuted. In either case, both types of data, supporting and refuting, are equally important.

However, the traditional process outline above suffers from several drawbacks:

- The process of data collection and analysis is typically executed manually involving several steps, which is very time-consuming. Moreover, the manual usage of log books makes it difficult to communicate the data and the steps of the experimental investigation to the collaborator colleagues who would like to repeat the experiment following the same or a modified protocol. Furthermore, as a consequence of heterogeneous data
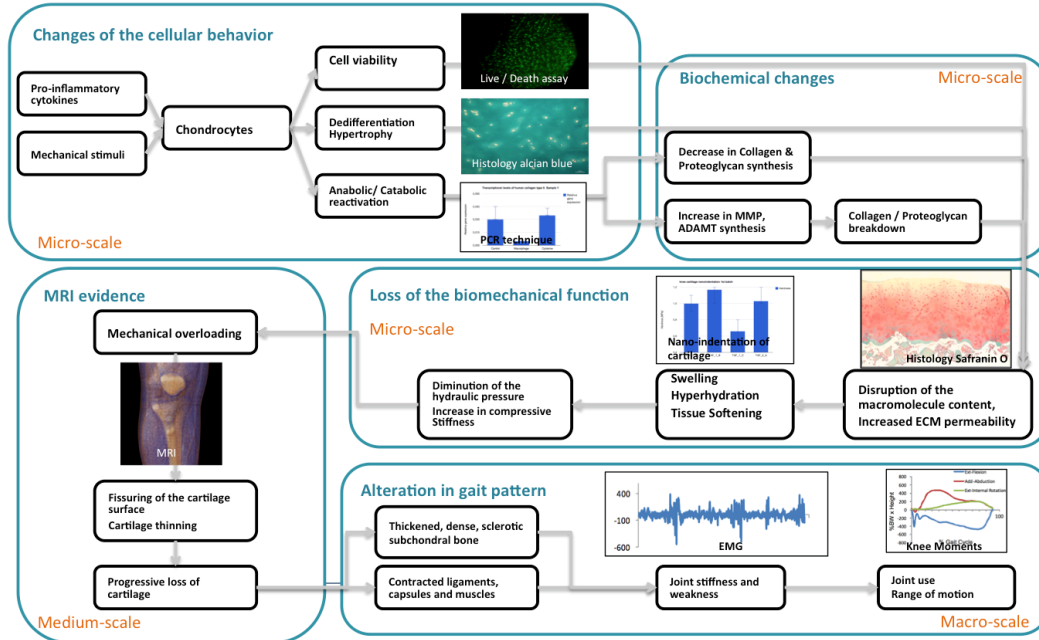
Figure 1. Model for the degradation of articular cartilage during osteoarthritis.

sets, which are often treated separately from each other, the results may be misinterpreted or treated without taking all the aspects into consideration.

- The visualization systems used for such analyses are isolated and allow only the exploration of features involving a concrete data type, but not the complete picture. Moreover, the interaction with the visualization systems is typically limited to conventional approaches, e.g. through standard mouse and keyboard interfaces, making the data hard to navigate.

### C. Organization of multi-scale data by knowledge formalization

Addressing the shortcomings of the traditional approach, our work aims to automatize the organization of relevant data sets and provide a filtering service to select the data that satisfies or refutes the hypothesis. Since the aforementioned data is heterogeneous there is a need for an integrated system, which presents relevant data in a unified view. Our approach decomposes naturally into several steps:

1) Medical background knowledge formalization encoded in an ontology, implemented in OWL, consisting of i) conceptual hierarchies: human anatomy, patient information, acquisition protocols, data representations, degradation process features and sources of evidence, and ii) relations between these concepts: patients undergoing acquisitions sessions, causal relationships between degradation process features.

2) Management of knowledge base (KB), in order to systematically store: i) available data sets (MRI, histology etc.), independently of the spatial scale or visualization technique to which they belong, ii) information on the acquisition protocol which was used to obtain data (e.g., protocols for obtaining the cartilage slices for histology analysis), and iii) relation of data sets to possible sources of evidence which they may hint to. This information can be uploaded manually, which was the case for the current work, however more interactive means of input collection and KB population supporting multi-scale and multimodal data should be envisaged as in [38].

3) Once the data sets are structurally stored in the KB, formalised relations between the concepts may be used to infer connections between data items, and thereby to generate relations, which were not explicitly stored in the KB before.

### D. Presentation and exploration of data

The visualization of the multi-scale biomedical data stored in the knowledge base is performed by means of a three-dimensional multi-layered environment. By making use of a 3D stereographic visualization, data sets and derived information are spatially distributed over the visualization scene, allowing a direct exploration of all related data of interest for the user. Data sets are rendered in independent entities, called nodes, allowing a consistent representation for heterogeneous data. Our 3D multi-layered visualization environment presents the data sets as a network of spa-
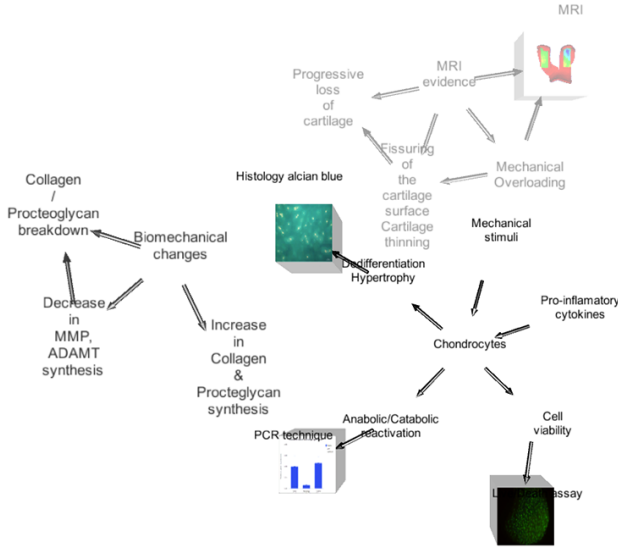
Figure 2. Example for an integrated visualization of patient-specific data set and the cartilage degradation process.

tially distributed and interconnected nodes. The nodes are equipped with several visual properties (e.g. spatial depth, geometric shape) in order to provide semantic means to identify relevant items during the exploration.

Data sets are rendered by employing suitable visualization techniques and settings, as identified in the knowledge base. They are positioned in distinguished layers, according to their spatial scale and the interest of the biologist. An example is shown Fig. 2, where the nodes belonging to the cellular scale appear to be nearer. These data sets are interconnected by arrows, highlighting the well-known relationships about them.

Our visualization scene provides semantic means to identify all related and relevant items during the exploration. This allows the simultaneous access, proper visualization and



Figure 3. Interaction with data (involving a stereoscopic display and the Leap Motion hand gesture device). Photo taken during the MultiScaleHuman [10] consortium meeting.

natural exploration of heterogeneous data across domains of medical knowledge, facilitating the task of the specialist to interpret and analyze the hypothesis.

Moreover, the user interaction (Fig. 3) is adapted to our multilayered visualization environment. Therefore, it is designed to enhance natural user instincts through hand gestures and haptic feedback. The navigation is coherent in space and time, thanks to three dimensional data organization and introduction of interaction based on simulation of physical forces. The user can manipulate data nodes through simple and intuitive gestures like grasping and dragging. Force directed organization of data graph is managed thanks to an accurate haptic [39] interface, that passes and receives forces.

## IV. KNOWLEDGE BASE

In our application, the ontology (developed within the EU FP7 MultiScaleHuman Project [10]) for multi-scale biomedical data management and visualization, referred to as *Multi-Scale Ontology* in the following, acts as a driver for data management. The queries it supports reflect the type and breadth of reasoning that medical practitioners may exploit. We augment the Multi-Scale Ontology for our use-case in such a way that:

- We model causality between DPF and SOE as OWL axioms, which allows us to interconnect concepts. In order to model hierarchical propagation of degradation process features we use the existentially quantified restriction axiom [40]. For instance, the relation between disruption of macromolecule content and tissue softening indicated in Fig. 1 is formalized as:

$$\text{Disruption\_of\_macromolecule\_content}$$
$$\sqsubseteq \exists causes \text{Tissue\_softening}$$

- Data sets are also classified with respect to their spatio–temporal dimensions, accordingly in *micro-scale* (encompassing data whose dimension ranges in micrometer size and below, e.g. nano-scale), *medium-scale* (PET, MRI, CT) and *macro-scale*. An additional *abstract scale* encompasses all the non-spatio-temporal knowledge that cannot be assigned to one of the previous scales, e.g. the anatomical entity which refers to a data set or the relations between DPF. This classification is encoded with the attribute *spatio-temporal scale*, which is going to be relevant for visualization purposes.

Fig. 4 represents the high-level representation of our augmented ontology.

*Population of the KB.* The knowledge base instantiates the corresponding classes of the Multi-Scale Ontology in order to record systematically the knowledge and data about a given patient, the acquisition sessions she or he underwent, the acquisitions protocols indicating which acquisition sessions were performed, the data sets which were acquired
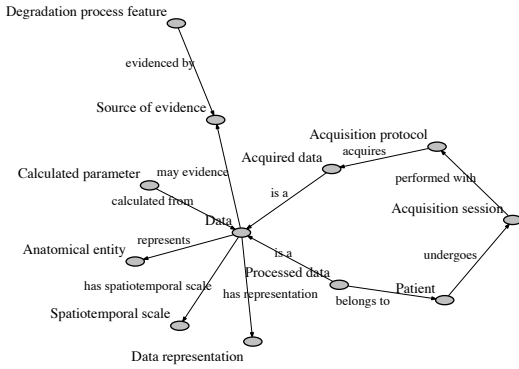
Figure 4. Ontology: big picture.

and the anatomical entities they represent. The resulting instantiation is schematically represented in Fig. 5.
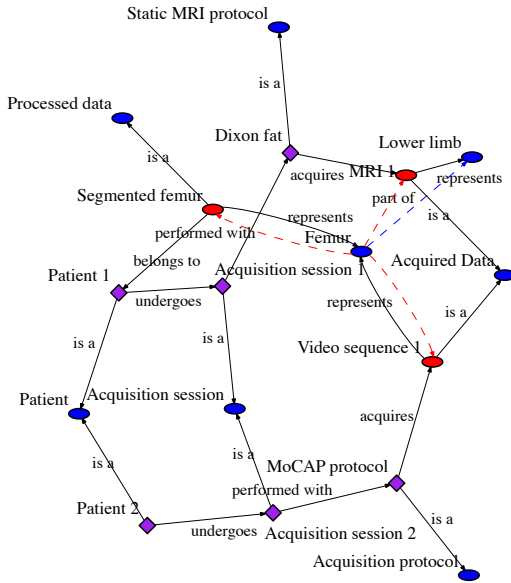


Figure 5. Instantiated KB with patient-specific data.

*Querying the KB.* Semantic queries, formulated according to the expressivity offered by the ontology, are essential but alone not sufficient to guarantee an intuitive exploration of the data and information stored in the knowledge base, especially if the user is not familiar with querying languages such as SPARQL or the internal data model of the KB. To present to the user the result of his query as well as extracted *context* information, we first transform the populated OWL ontology into a directed graph and then extract context information by using graph traversal techniques starting from the *seed node* provided by the user. The latter represents the focus entity of the user's query, e.g. all data related to *Femur*.

*OWL to graph transformation.* The ontology is converted into a directed graph by the following steps: i) The OWL ontology is loaded in one of the usual formats (RDF/XML,
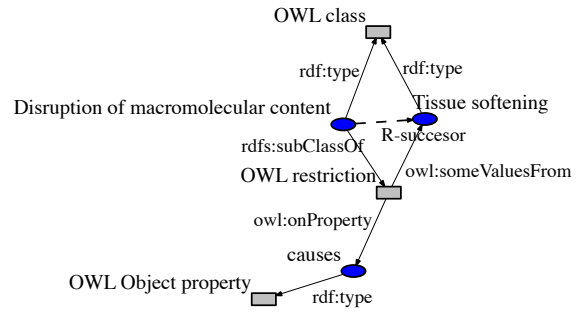
Turtle) [29], ii) The RDF graph model is used in order to identify nodes (i.e. classes, individuals) and edges of the final graph representation (i.e. `rdf : type` relations between classes and individuals, `owl : subClassOf` relations between classes, `owl : objectProperties` between instances, and `rdfs : domain`, `rdfsw : range` between classes).

During the transformation process, special treatment is done to more complex OWL axioms such as existential restriction axioms (e.g. Fig. 6), which we use to model propagation of DPFs or parthood relationships between anatomical entities (e.g. `Femur − partof − Lowerlimb`). We analyze the RDF data model of these OWL axioms in order to connect concepts in our final graph transformation. We also interconnect data instances if they refer to sources of evidence which inter-cause each other. For example Fig. 7 shows how *data1* is related to *data2* due to an existent path in the causal chain between *soe1* to *soe3* through *soe2* (transitive relation).
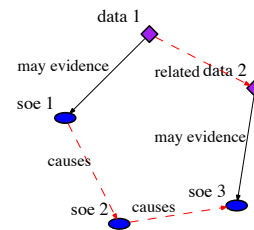


Figure 6. OWL existential restriction (R-successor).



Figure 7. Interconnection of datasets based on inference of facts in the KB.

*Inference.* Finally, the graph representation of the KB contains nodes representing concepts and instances as well as inferred relations between them. This allows us to identify semantic links between data nodes. For instance Fig 5 shows that *Video sequence 1* and *Segmented femur* are directly associated to the concept *Femur* through the *represents* relation, and would be extracted from the KB if the query contains the focus entity *Femur*. The *MRI 1* data instance is not directly connected to *Femur*, however its relation to *Femur* is inferred due to the fact that *Femur* is *part of Lower*

*limb*. Ultimately, all three data instances, *Video sequence 1*, *Segmented femur* and *MRI 1*, are going to be included in the *context information* of the *focus entity Femur*.

Answers to queries are then represented as subgraphs extracted from the graph representation of the KB, which are encoded as JSON documents [41] and served over the secured HTTPS protocol to the exploratory system (more details in Sec. V).
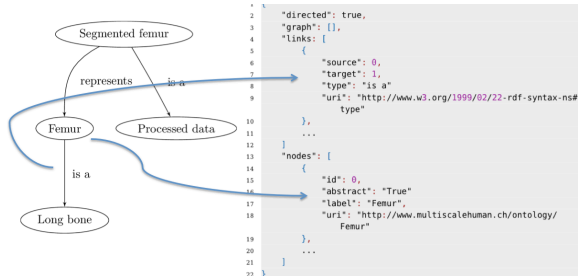


Figure 8. JSON representation of answer to the query.

## V. ARCHITECTURE OF THE EXPLORATORY SYSTEM

The exploratory system is specifically designed to be used by several users working remotely with large amounts of biomedical data. Its underlying architecture addresses portability and scalability requirements as well as network data transfer (streaming and caching) [6]. The visualization scene is realized by means of a thin-client based on the multi-platform Unity3D software framework [42], which accesses network resources, such as the the rendering system, the data repository and the knowledge base. The architecture of the exploratory system is roughly sketched in Fig. 9.

Since the *rendering system* employs a distributed scheme, which composites the visualization scene from several data sets rendered on separate rendering worker entities, it is able to deal effectively with massive data sets. The visualization system connects to a rendering worker entity, which acts as an active proxy between the client and the *data repository*. An appropriate rendering worker according to the type of data set finally renders it and sends it for the composition of the visualization scene.

Requests to the *knowledge base* yield all relevant multi-scale biomedical information of a biomedical case to present to the user, according to the *specialist*, and *patient* properties encoded in the ontology. Therefore, information regarding the multi-scale biomedical data (e.g. location of resources) is collected from the knowledge base and processed in order to create the visualization scene.

Our *exploratory system* connects with the knowledge base back-end through encrypted HTTPS protocol to ensure the security of the connection. The server accepts two path arguments for each request: modality and focus entity. Each request is identified by a session identifier. Sessions store information on the configured user profile and recently requested entities. As a result of each query a JSON document is returned consisting of the list of nodes and the list of relations between the nodes, as indicated in Fig. 8.
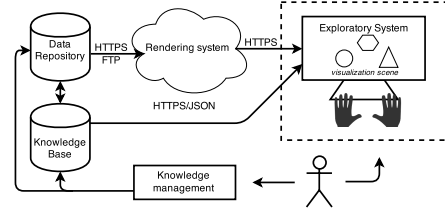


Figure 9. Architecture of the exploratory system.

In the following paragraphs we describe how the information obtained from the ontology is processed for the creation of the visualization scene, which is schematically depicted in Fig. 10.

*Appearance of multi-scale data sets*. The collected information is visually displayed either as an independent visual entity, called node, or by means of visual cues, according to the ontology attribute *spatio-temporal scale*. Data sets from *micro-scale*, *medium-scale*, *macro-scale* have a representation as nodes. These nodes encapsulate 2D, 3D or InfoVis data for a consistent simultaneous exploration of heterogeneous data. In the JSON document, each node has an associated URL leading to a related data set, which will be retrieved and cached from the data repository in order for the visualization tasks to run smoothly. The caching process will happen as soon as the exploratory system requests a rendering worker to render a data set. Each node has a 3D cube as a node representation figure, following the results obtained in the user study performed in [6] regarding visualization preferences. Moreover, information from the *abstract scale* can have a node representation (e.g. the anatomical entity which refers to a data set) or could be used for the creation of visual cues in nodes and between them in order to represent semantic relations across data sets (SOEs and DPFs).

*Positioning of data sets*. Data nodes are positioned on the multi-layered environment. The layers, distinguished by
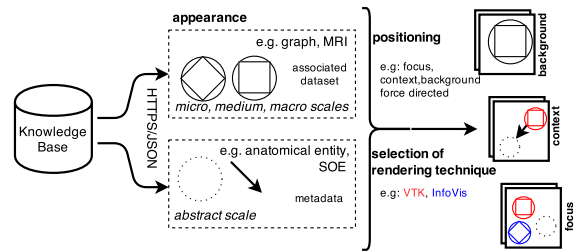


Figure 10. Process for the visualization composition.

their z-order to indicate grade of importance, are denominated *focus*, *context* and *background*. Each layer encompasses the nodes with the same spatial scale. The allocation of a node to a layer depends on the spatio-temporal scale of the data set and the interest of the user in that scale (attributes *spatio-temporal scale* and *specialist*). Focus constitutes the main scale of interest of the user. Data sets with InfoVis and SciVis suitabilities are linked and visualized on the foreground. Context is placed behind, aiming to provide context to the data on the focus layer with data coming from a contiguous spatial scale. Background is the last layer, and completes the general view across all the spatial ranges. Its nodes are presented by using the call-out technique, alleviating the differences in the order of magnitude of data.

*Selection of the rendering technique*. Data nodes are rendered by a suitable rendering entity according to their data type (3D data from VTK, STL, DICOM format files, or InfoVis, e.g. area or bar charts [43]. The visualization system identifies the visualization technique and settings to be used from the description of the data as inferred from the ontology. Concretely, the selection of visualization properties is defined by the mapping:

$$(AP, PD, CP, DR) \mapsto (VT, S),$$

where the entries in the tuple on the left denote the acquisition protocol, the processed data, the calculated parameter values and the data representation, respectively. Similarly, the tuple on the right hand side denotes the visualization technique and settings. For instance, in Fig. 11, the femoral cartilage thickness is visualized as a manifold surface mesh obtained after processing a MRI scan. Here $DR=ManifoldSurfaceMesh$ and $PD=CartilageThickness$ leads to $VT=VTKViewer$ and $S=Colormap$. Therefore, a rendering worker based on VTK is employed with visual settings for using a color map to indicate thickness in the cartilage.
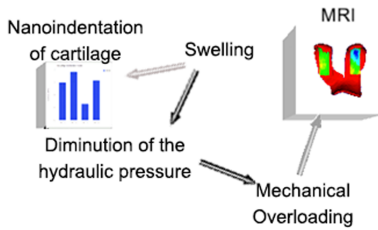


Figure 11.   Extended view of the resulting visualization.

*Determination of appearance of abstract information*. Abstract information can be used for representing nodes or for the creation of perceptual cues in the relations between such nodes in order to enrich the understanding of the complete collection of data. For example, nodes with a neutral representation and a label are used to represent anatomical entities. Examples for visual cues include:

- Colored arrows between nodes indicate relations between the nodes, e.g. SOE for the cartilage degradation process.
- Tooltips provide information on the acquisition protocol, such as the number of days the specimen preparation takes for a histological analysis.

Adaptability is a key factor for a seamless user experience, where the system not only combines data but also assists in filtering out and emphasizing core medical concepts. Different profiles of experts characterize themselves with a focus on different data sets. Adaptability allows the system to score concepts and data sets based on the frequency of user interactions. Frequency scores are stored locally along with a user profile. Every time a concept or data node is loaded, associated scores are recalculated. Similarly, every time the user interacts with a node $i$, the corresponding time of interaction $t_i$ is recorded and an associated score $S_i$ is updated according to the rule:

$$S_i^{\text{new}} = a \cdot \frac{t_i}{T} + b \cdot \frac{k_i}{N} + c \cdot S_i^{\text{old}}.$$

Here $k_i$ denotes the number of nodes that share similar properties with node $i$, $N$ is the total number of nodes and $T$ represents the total duration of the session. Finally $a, b, c$ are fixed weights. For comparison purposes, all scores are scaled to fit in the range $[0, 1]$. The next time the node $i$ is loaded, its score is compared with a threshold coefficient $d$. If $S_i < d$, the node will not appear. It is also possible to bypass completely this feature and access all elements irrespective to their score.

Regarding interaction with the exploratory system, previous work [6] was focused on two modes: i) gesture based management of nodes and ii) fine grained haptic interaction with node details. The resulting user experience was based on physical interaction between user virtual fingers and node objects. In our current approach, the line of physical immersion is pushed further. Nodes do not only interact with user's finger but their organization is driven by a simulation of physical bonds between them. Therefore, the user can reorganize the view through manual manipulation of single nodes, leading to automatic reorganization of the complete node set. Nodes can be pinned to fixed locations, which adds new boundary conditions to our force directed system [44]. This method used for the organization of the workspace also associates an inertia value to each node. Subsequently this allows to pass information about importance (or interconnectivity) of a single node through a haptic interaction device. As important nodes will have higher interconnectivity and higher intertia, they will be much more difficult to move with a haptic instrument.

## VI. Conclusion and future work

The proposed approach aims to assist specialists in the medical exploration of multi-scale biomedical data. The

knowledge formalization organizes and filters the heterogeneous data in the system. Its inference (exemplified in Sec. IV) has been proved to identify new semantic links among biomedical data which were not explicitly stored. The integrated environment (Sec. V) captures all relevant information for the domain expertise of the user and provides a natural exploration.

The exploratory system allows the use of data properties stored in the knowledge base in order to automatize the process of selection of suitable visual properties for the exploration of each subset of multi-scale biomedical information. Furthermore, the system remains scalable for the visualization of other data-types in biomedical data. For extending the system in this regard, it suffices to insert new rendering worker types and to include information about its use in the knowledge base.

The designed system is intended to be used by biomedical researchers, and experts of the medical field for research or analytic/diagnostic purposes. The scenario discussed from the point of view of a molecular biologist can be considered to be applicable for other specialists that are working on the exploration of multi-scale biomedical data. Supplementing previous user studies on visualization and interaction [6], the discussion of the use-case described in this paper indicates the utility of our approach for biomedical specialists confronted with massive amounts of multi-scale data on a daily basis.

Integrated visualization can open up new possibilities for collaboration between medical practitioners, who under normal circumstances work on their specific data sets without communicating too much with other specialists. By complementing each others' analysis with exchange of expertise and opinion, together, medical practitioners are better equipped to study and analyze multi-scale pathologies.

The realization of the exploratory system of this magnitude certainly opens up paths for further investigation regarding system adaptiveness. We would like to investigate new ideas like integration of scoring techniques on the knowledge base level or influence positioning of nodes based on user interests. This would allow different users to share experiences with each other and to collaborate. This can also lead to a whole set of collaborative scenarios, e.g. involving the optimization of protocols or the exploration of past studies, that should be investigated.

### References

[1] M. B. Goldring, "Articular cartilage degradation in osteoarthritis," *HSS Journal*, vol. 8, no. 1, pp. 7–9, 2012.

[2] X. Houard, M. B. Goldring, and F. Berenbaum, "Homeostatic mechanisms in articular cartilage and role of inflammation in osteoarthritis," *Current rheumatology reports*, vol. 15, no. 11, pp. 1–10, 2013.

[3] J. L. Astephen, K. J. Deluzio, G. E. Caldwell, M. J. Dunbar, and C. L. Hubley-Kozey, "Gait and neuromuscular pattern changes are associated with differences in knee osteoarthritis severity levels," *Journal of Biomechanics*, vol. 41, no. 4, pp. 868–876, 2008.

[4] N. J. B. McFarlane, G. J. Clapworthy, A. Agrawal, M. Viceconti, F. Taddei, E. Schileo, and F. Baruffaldi, "3d Multiscale visualisation for medical datasets," in *5th International Conference on BioMedical Visualization, MediVis'08*, 2008, pp. 47–52.

[5] R. M. Millán Vaquero, J. Rzepecki, K.-I. Friese, and F.-E. Wolter, "Visualization and user interaction methods for multiscale biomedical data," in *3D Multiscale Physiological Human*. Springer, 2014, pp. 107–133.

[6] J. Rzepecki, R. M. Millán Vaquero, A. Vais, K.-I. Friese, and F.-E. Wolter, "Multimodal approach for natural biomedical multi-scale exploration," in *Advances in Visual Computing*. Springer, 2014, pp. 620–631.

[7] A. Agibetov, R. M. Millán Vaquero, K.-I. Friese, G. Patanè, M. Spagnuolo, and F.-E. Wolter, "Integrated visualization and analysis of a multi-scale biomedical knowledge space," in *EuroVis Workshop on Visual Analytics*, Swansea, UK, 2014, pp. 25–29.

[8] P. Hunter, P. V. Coveney, B. Bono, V. Diaz, J. Fenner, A. F. Frangi, P. Harris, R. Hose, P. Kohl, P. Lawford, K. McCormack, M. Mendes, S. Omholt, A. Quarteroni, J. Skar, J. Tegner, S. R. Thomas, I. Tollis, I. Tsamardinos, J. H. G. M. v. Beek, and M. Viceconti, "A vision and strategy for the virtual physiological human (VPH) in 2010 and beyond," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 368, no. 1920, pp. 2595–2614, 2010.

[9] D. Testi, D. Giunchi, G. Clapworthy, S. Aylward, X. Planes, and R. Christie, "New interactive visualisation of multiscale biomedical data," in *ACM SIGGRAPH 2012 Posters*, New York, USA, 2012, pp. 76:1–76:1.

[10] "MultiScaleHuman Project." [Online]. Available: http://multiscalehuman.miralab.ch/

[11] S. I. O'Donoghue, A. C. Gavin, N. Gehlenborg, D. S. Goodsell, J. K. Hrich, C. B. Nielsen, C. North, A. J. Olson, J. B. Procter, and D. W. Shattuck, "Visualizing biological data: now and in the future," *Nature methods*, vol. 7, pp. S2–S4, 2010.

[12] N. J. B. Mcfarlane, X. Ma, G. J. Clapworthy, N. Bessis, and D. Testi, "A survey and classification of visualisation in multiscale biomedical applications," in *16th International Conference on Information Visualisation*, 2012, pp. 561–566.

[13] N. Gehlenborg, S. I. O'Donoghue, N. S. Baliga, A. Goesmann, M. A. Hibbs, H. Kitano, O. Kohlbacher, H. Neuweger, R. Schneider, and D. Tenenbaum, "Visualization of omics data for systems biology," *Nature methods*, vol. 7, pp. S56–S68, 2010.

[14] J. Chen, E. Dougherty, S. Demir, C. Friedman, C.-S. Li, and S. Wong, "Grand challenges for multimodal bio-medical systems," *IEEE Circuits and Systems Magazine*, vol. 5, no. 2, pp. 46–52, 2005.

[15] B. Lorensen, "On the death of visualization," in *Position Papers NIH/NSF Proc. Fall 2004 Workshop Visualization Research Challenges*, 2004.

[16] T.-M. Rhyne, "Does the difference between information and scientific visualization really matter?" *IEEE Computer Graphics and Applications*, vol. 23, no. 3, pp. 6 – 8, 2003.

[17] M. Viceconti, F. Taddei, L. Montanari, D. Testi, A. Leardini, G. Clapworthy, and S. Van Sint Jan, "Multimod data manager: A tool for data fusion," *Computer methods and programs in biomedicine*, vol. 87, no. 2, pp. 148–159, 2007.

[18] H. Hauser, "Generalizing Focus+Context visualization," in *Scientific Visualization: The Visual Extraction of Knowledge from Data*. Springer, 2006, pp. 305–327.

[19] J.-P. Balabanian, I. Viola, and E. Gröller, "Interactive illustrative visualization of hierarchical volume data," in *Proceedings of Graphics Interface 2010*, 2010, pp. 137–144.

[20] J. E. Sowa and S. C. Shapiro, *Knowledge Representation: Logical, Philosophical, and Computational Foundations Computational Foundations by John F. Sowa (Book Review)*. Springer-Verlag, 2006.

[21] F. Belleau, M.-A. Nolin, N. Tourigny, P. Rigault, and J. Morissette, "Bio2rdf: Towards a mashup to build bioinformatics knowledge systems," *Journal of Biomedical Informatics*, vol. 41, no. 5, pp. 706–716, Oct. 2008.

[22] T. R. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, vol. 5, no. 2, pp. 199–220, Jun. 1993.

[23] K. A. Spackman, P. D, K. E. Campbell, P. D, R. A. Ct, and D. S. (hon, "SNOMED RT: A reference terminology for health care," in *J. of the American Medical Informatics Association*, 1997, pp. 640–644.

[24] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig *et al.*, "Gene ontology: tool for the unification of biology," *Nature genetics*, vol. 25, no. 1, pp. 25–29, 2000.

[25] C. Rosse and J. L. V. Mejino, "A reference ontology for biomedical informatics: the foundational model of anatomy," *Journal of Biomedical Informatics*, vol. 36, p. 500, 2003.

[26] A. L. Rector, J. E. Rogers, and P. Pole, "The galen high level ontology." IOS PRESS, 1996, pp. 174–178.

[27] C. P. Langlotz, "RadLex: A new method for indexing online educational materials," *RadioGraphics*, vol. 26, no. 6, pp. 1595–1597, Nov. 2006.

[28] I. Banerjee, C. Catalano, F. Robbiano, and M. Spagnuolo, "Accessing and Representing Knowledge in the Medical Field: Visual and Lexical Modalities," in *3D Multiscale Physiological Human*, N. Magnenat-Thalmann, O. Ratib, and H. F. Choi, Eds. Springer London, 2014, pp. 297–316.

[29] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness, P. F. Patel-Schneider, and L. A. Stein, *OWL Web Ontology Language Reference*. World Wide Web Consortium, Feb. 2004. [Online]. Available: http://www.w3.org/TR/owl-ref/

[30] F. Manola and E. Miller, Eds., *RDF Primer*. World Wide Web Consortium, Feb. 2004. [Online]. Available: http://www.w3.org/TR/rdf-primer/

[31] J. Pérez, M. Arenas, and C. Gutierrez, "Semantics and complexity of SPARQL," *ACM Transactions on Database Systems*, vol. 34, no. 3, pp. 16:1–16:45, Sep. 2009.

[32] D. Gantenbein, "Kinect launches a surgical revolution," *A+ A*, vol. 9, p. 14, 2013.

[33] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1–54, 2015.

[34] M. Ondrésik, C. Correia, R. Sousa, J. Oliveira, and R. Reis, "Understanding cellular behaviour in early and late stage of MSD," *Journal of Tissue Engineering and Regenerative Medicine*, vol. 8, pp. 412–412, 2014.

[35] J. Martel-Pelletier, G. Tardif, J. Fernandes, and J.-P. Pelletier, "Metalloproteases and their modulation as treatment in osteoarthritis," in *Principles of Molecular Rheumatology*. Springer, 2000, pp. 499–513.

[36] K.-I. Friese, P. Blanke, and F.-E. Wolter, "YaDiV– an open platform for 3d visualization and 3d segmentation of medical data," *The Visual Computer*, vol. 27, no. 2, pp. 129–139, 2011.

[37] K.-I. Friese, S. B. Cichy, F.-E. Wolter, and R. E. Botcharnikov, "Analysis of tomographic mineralogical data using YaDiV– overview and practical case study," *Computers & Geosciences*, vol. 56, pp. 92–103, 2013.

[38] M. Pitikakis and I. Banerjee, "Knowledge management in medicine - A framework to organize, browse and retrieve medical data," in *Proceedings of the International Conference on Health Informatics, France, 2014*, 2014, pp. 374–380.

[39] R. Vlasov, K.-I. Friese, and F.-E. Wolter, "Haptic rendering of volume data with collision determination guarantee using ray casting and implicit surface representation," in *International Conference on Cyberworlds*, 2012, pp. 91–98.

[40] F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, Eds., *The Description Logic Handbook: Theory, Implementation, and Applications*. New York, NY, USA: Cambridge University Press, 2003.

[41] A. Agibetov, *Essence of JSON: Static Analysis for JSON*. LAP LAMBERT Academic Publishing, 2013.

[42] T. Higgins, "Unity - 3d game engine." [Online]. Available: http://unity3d.com/

[43] J.-D. Fekete, "The Infovis Toolkit," in *Proceedings of the IEEE symposium on Information Visualization*. IEEE Computer Society, 2004, pp. 167–174.

[44] T. M. Fruchterman and E. M. Reingold, "Graph drawing by force-directed placement," *Software: Practice and experience*, vol. 21, no. 11, pp. 1129–1164, 1991.