White paper on multiscale visualisation (draft)

D2.1

Work package 2: Shared vision

Debora Testi (B3C)
Daniele Giunchi (B3C)
Simone Brazzale (B3C)
Ignacio Larrabide (UPF)
Rubens Cardenes (UPF)
Gordon Clapworthy (BED)
Nigel McFarlane (BED)
Stephen R Aylward (KIT)

14/02/2011
The aim of the MSV project is to develop an open source library for the visualisation and interaction with multiscale biomedical data. The partners revised the current state of the art also in domains outside biomedicine and defined a series of needs and challenges for the MSV project which are here summarised. This shared vision document will be progressively updated in the next year and its final version will be made publicly available at the end of 2011.
<table>
<thead>
<tr>
<th>Date</th>
<th>Version</th>
<th>B3C</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>10/12/2010</td>
<td>V4</td>
<td>B3C</td>
<td>State of the art and exemplary problems section completed</td>
</tr>
<tr>
<td>15/01/2011</td>
<td>V5</td>
<td>B3C</td>
<td>Revised version including UPF contributions and comments</td>
</tr>
<tr>
<td>31/01/2011</td>
<td>V5b</td>
<td>B3C</td>
<td>Further revision of the text</td>
</tr>
<tr>
<td>14/02/2011</td>
<td>V6</td>
<td>B3C</td>
<td>Final version including contributions and comments from all partners</td>
</tr>
</tbody>
</table>
Disclaimer: This document is property of the MSV Consortium. There is no warranty for the accuracy or completeness of the information, text, graphics, links or other items contained within this material. This document represents the common view of the consortium and does not necessarily reflect the view of the individual partners.
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D</td>
<td>Three-dimensional</td>
</tr>
<tr>
<td>AUK</td>
<td>University of Auckland, NZ</td>
</tr>
<tr>
<td>B3C</td>
<td>BioComputing Competence Centre - SCS srl, Italy</td>
</tr>
<tr>
<td>BED</td>
<td>University of Bedfordshire, UK</td>
</tr>
<tr>
<td>CFD</td>
<td>Computational Fluid Dynamics</td>
</tr>
<tr>
<td>CT</td>
<td>Computed Tomography</td>
</tr>
<tr>
<td>DTI</td>
<td>Diffusion Tensor Imaging</td>
</tr>
<tr>
<td>FPS</td>
<td>Frames per second</td>
</tr>
<tr>
<td>GPU</td>
<td>Graphic Processing Units</td>
</tr>
<tr>
<td>GUI</td>
<td>Graphic User Interface</td>
</tr>
<tr>
<td>HCI</td>
<td>Human-computer interaction</td>
</tr>
<tr>
<td>KIT</td>
<td>Kitware Inc, USA</td>
</tr>
<tr>
<td>LOD</td>
<td>Level of details</td>
</tr>
<tr>
<td>MR</td>
<td>Magnetic Resonance</td>
</tr>
<tr>
<td>MSV</td>
<td>Multiscale Visualisation and interaction project</td>
</tr>
<tr>
<td>Qt</td>
<td>GUI development package from Nokia</td>
</tr>
<tr>
<td>UPF</td>
<td>Universitat Pompeu Fraba, Spain</td>
</tr>
<tr>
<td>US</td>
<td>Ultrasound</td>
</tr>
<tr>
<td>VPH</td>
<td>Virtual Physiological Human</td>
</tr>
<tr>
<td>VTK</td>
<td>Visualisation ToolKit</td>
</tr>
</tbody>
</table>
White paper on multiscale visualisation (draft)

1. Introduction .......................................................................................................................... 7
   1.1 Project domain and aim ....................................................................................................... 7

1.2 Some definitions .................................................................................................................. 8
   1.2.1 Visualisation .................................................................................................................. 8
   1.2.2 Scientific visualisation .................................................................................................. 9
   1.2.2.1 Applications ............................................................................................................. 10
   1.2.3 Interactive visualisation ................................................................................................ 13
   1.2.3.1 Human control ........................................................................................................ 13
   1.2.3.2 Rapid response to human input .............................................................................. 13
   1.2.4 Multiscale data ............................................................................................................ 15

2. State-of-the-art ..................................................................................................................... 17
   2.1 Non biomedical domains .................................................................................................. 17
   2.2 General tools ................................................................................................................... 18
   2.3 Biomedical field .............................................................................................................. 19

3. Exemplary problems .......................................................................................................... 22
   3.1 Use cases ........................................................................................................................ 22
   3.2 First categorisation ......................................................................................................... 25

4. MSV visualisation problem formalisation ......................................................................... 27
   4.1 Visualisation process ....................................................................................................... 27
   4.1.1 Definition of visualisation ........................................................................................... 28
   4.1.2 Degree of multiscaling ............................................................................................... 29
   4.1.3 Open questions and issues ....................................................................................... 30
   4.2 Formal taxonomy for multiscale visualisation ................................................................. 31

5. Multiscale Visualisation and interaction ........................................................................... 32
   5.1 Problem definition .......................................................................................................... 32
   5.2 Proposed interaction and visualisation paradigms .......................................................... 33
   5.3 Challenges ....................................................................................................................... 33
   5.3.1 Computers systems ..................................................................................................... 33
   5.3.2 Algorithms .................................................................................................................. 35
   5.3.3 Interaction and visualisation ....................................................................................... Errore. Il segnalibro non è definito.
   5.3.4 Validation .................................................................................................................... 35

6. Related efforts and potential users ..................................................................................... 36

7. Roadmap .............................................................................................................................. 37

8. References .............................................................................................................................. 38

9. ANNEX 1: Visual process glossary ..................................................................................... 40
1 Introduction

1.1 Project domain and aim

The Virtual Physiological Human (VPH)\(^1\) is a methodological and technological framework that, once established, will enable collaborative investigation of the human body as a single complex system. That framework should make it possible to interconnect predictive models that span the full range of relevant spatial and temporal scales, with diverse methods operating at varying levels of detail, into systemic networks that enable the discovery and evaluation of multi-system hypotheses [Fenner, 2008]. The project discussed in this report addresses the specific aim of integrating multiple physiological processes that operate at temporal and spatial scales that may differ by orders of magnitude (Multiscale Spatiotemporal Visualisation, MSV in brief).

As new methods for collecting and modelling multiscale data have begun to emerge from the different VPH-related projects, it has become increasingly evident that there is a shortage of appropriate tools for visualising and exploring data that are defined across a broad range of spatial and/or temporal scales; this is particularly true for data that cannot be entirely represented within the typical human perceptual range.

The number of biomedical problems that will demand multiscale visualisation in the coming years suggests that this area should start to receive urgent attention but, surprisingly, it received almost no mention in the Visualisation Research Challenges [Johnson, 2006] document produced jointly by the USA National Institute for Health and National Science Foundation.

Multiscale visualisation has been investigated in other contexts with, perhaps, the most relevant work being undertaken in the context of geographical data visualisation, as exemplified by well-known solutions such as Google Earth\(^2\). While these are extremely effective within the context of their specific target problem, not all of the available solutions can be generalised to other domains. In particular, many of the approaches do not translate well into the biomedical research area due to the volumetric, heterogeneous, and time-varying nature of the relevant data.

It is evident that this visualisation challenge is not simply related to the computational efficiency. Researchers should collaboratively work with domain experts on driving tasks to produce tools which will solve real-world needs and make systems and techniques that better leverage human characteristics.

To address this issue, an international consortium has been established involving leading groups in the development of visualisation software tools\(^3\). MSV project (FP7-IST-248032) will have the aim of designing and providing, over the course of the next two years, a first implementation of an open-source library to meet the current and future needs encountered when visualising, simulating and interacting with multiscale biomedical data. The MSV software library will provide a suitable resource for the VPH community and others to be used in this rapidly evolving area.

The aim of this document, White Paper, is to provide a review of the state-of-the-art and of the open challenges which will drive the next designing phase of the MSV library. This White Paper will be a living document growing in the next year, during which contributions and comments from external stakeholders will be collected. The final version of the White Paper will be publicly released at the end of 2011.

---


\(^2\) [http://earth.google.com/](http://earth.google.com/)

\(^3\) [http://www.msv-project.eu/consortium.html](http://www.msv-project.eu/consortium.html)
1.2 Some definitions

The literature (books, papers, conference proceedings, etc.) available on visualisation in general and on some of the specific topics we are addressing in this document is extensive and its exhaustive review was out of the scope of this report. We rather focused on those aspects relevant for the MSV project, in particular for this and for the following state-of-the-art sections.

1.2.1 Visualisation

*Visualisation* is a general term that has specific meanings in different contexts. It can refer to the research discipline, to a specific technique or to a visual result. In particular, in computer graphics, visualisation is any technique for creating images, diagrams, or animations to communicate a message. Visualisation is thus about helping people explore or explain data through software systems that provide a static or interactive visual representation [Johnson, 2006].

Visualisation has been an effective way to communicate both abstract and concrete ideas since a long time ago. However, in its early days the lack of graphics power and computational efficiency often limited its usefulness. We might consider the birth of scientific visualisation, as we currently interpret it, in 1987 with the special issue of Computer Graphics on Visualisation in Scientific Computing [McCormick, 1987]. Since then, there has been an increasing role of visualisation, and, together with computer graphics, it has today an ever-leading role in many scientific applications like science, education, engineering, multimedia, medicine, etc.

Visualisation designers exploit the high bandwidth channel of the human visual perception to allow people to comprehend information orders of magnitude more quickly than they would have by reading raw numbers or text to derive new knowledge from data. Thus, designing effective visualisation is a complex process that requires a sophisticated understanding of human information processing and, in fact, visualisation systems are explicitly designed not to replace the human but to keep him/her in the loop by extending his/her capabilities [van Wijk, 2005].

An example on the role given to the human in the visualisation process is given also by [Ware, 2004] who presents the four stages of the process giving particular emphasis to the role of the human in the loops (Figure 1).

![Visualisation Process Diagram](image)

*Figure 1 - The visualisation process adapted from [Ware, 2004]. The process is composed by four steps: i) the collection and storage of data itself, ii) the pre-processing designed to transform the data into something the user can understand, iii) the display and the graphics algorithms that produce an image on the screen, and iv) the human perceptual and cognitive system.*

Visualisation can be then categorised according to the different types of application; the most relevant for MSV are:

- *Scientific visualisation:* it is the transformation, selection or representation of data generated by simulations, observed from physical processes, or recorded during experiments. These data typically exist in an implicit or explicit geometric framework that gives context to its exploration, analysis, and understanding.
• Information visualisation: it concentrates on the use of computer-supported tools to explore, in intuitive ways, large collections of non-numerical/textual meta-data, such as lines of code in software systems, results reported in publications, and citations in bibliographic databases, networks of relations on the internet, and so forth.

As MSV project is dealing with biomedical data on different scales, we are particularly interested in scientific visualisation for which more details are given in the next section. However, information visualisation also deals with multiscale data and some approaches can be of interest, such as methods based on data cubes [Stolte, 2003].

1.2.2 Scientific visualisation

Scientific visualisation is an interdisciplinary branch of science "primarily concerned with the visualisation of 3D phenomena (architectural, meteorological, medical, biological, etc.), where the emphasis is on realistic renderings of volumes, surfaces, illumination sources, and so forth, perhaps with a dynamic (time) component" [Friendly, 2008].

Fields, which are strongly related with scientific visualisation and with MSV, are:

- **Computer simulation** is a program running on a single computer, or on a network of computers, that attempts to simulate a specific system using a numeric model. Computer simulations have become a useful part of mathematical modelling of natural systems in physics, computational biology, and chemistry; of human systems in economics, psychology, and social science; and in analogue and digital systems and manufacturing process in computer science and engineering. The goal of these simulations is to gain insight into the operation of those systems, to observe their behaviour, to diagnose their failures, and to improve their performance.

- **Computational geometry** is a branch of computer science devoted to the study of algorithms which can be stated in terms of operations on geometry. The operations include measures of surface properties, transformations that project data onto geometric forms for visualisation and analysis, and unions and intersections of geometries.

- **Signal processing** is concerned with the segmentation, registration, filtering, and characterisation of data, including images. It encompasses the fields of medical image analysis, statistical pattern recognition, computer vision, and biophysical signal recording and processing.

- **Human-Computer Interaction (HCI)** is the field of study that involves the design of graphical user interfaces (GUIs), including those involving complex visualisations and the interactive display and manipulation of data. Often neglected in early computer applications, particularly research applications; the Nintendo Wii⁴ and Apple’s iPad⁵ are just two recent examples of the power of effective HCI and the priority it is now being given in consumer systems. The surge in the adoption of Nokia’s open-source, yet highly polished, QW GUI development package⁶ similarly portrays the emphasis that research application developers are also giving to HCI.

- **Parallel and GPU computing**: Visualisation has always been computationally demanding. As soon as hardware has become capable of delivering the required output within a reasonable computational time, users have demanded greater realism, improved visual effects or additional features, and the associated computational demands have again exceeded the contemporary hardware capability. Certain graphics techniques, e.g. ray tracing or ray casting, have always been amenable to parallelisation, and there have been several examples of specific hardware being marketed to take advantage of this, though none proved commercially viable. The advent of the multicore computer and, in particular, the modern Graphics Processing Unit (GPU) has brought parallel computing to the desktop and this has greatly increased interest in parallelisation. In fact, the rapid expansion of GPGPU (general purpose GPU) programming, particularly as

---

⁴ [http://wii.com](http://wii.com)
a result of support from programming languages such as CUDA\textsuperscript{7}, has meant that many tasks in visualisation are now being addressed using parallel techniques.

Scientific visualisation is usually performed using specialised software. Some of these have been released as Open-source software (i.e. Visualisation ToolKit\textsuperscript{8}), while there are also many proprietary software packages of scientific visualisation tools.

### 1.2.2.1 Applications

Scientific visualisation is applied today in many fields of application, which are here briefly listed.

**Geography and Meteorology**

Cartography and geography have a long history in the use of scientific visualisation since the first world map in the 550 BC [Friendly, 2008]. The amount of information to be represented and their visualisation techniques have been growing steadily. At present, in this field of application, multiscale visualisation is used for rendering terrains, meteorological systems, climate changes, etc. The most famous example of multiscale representation in this field is provided by Google Earth which is described in the next state of the art section (Figure 2). Strategies for representing nearly every type of multidimensional data in this field have also been identified.

*Figure 2 – Example of visualisation of multidimensional and multiscale data in cartography and meteorology: Google earth snapshot.*

**In the applied sciences, engineering**

Engineering is also an influential field of application for scientific visualisation. It typically integrates computer simulation, computational geometry, and parallel/distributed processing to drive its visualisations. Consider, for example, the multiscale visualisation shown in Figure 3. The underlying data was generated by a computational simulation that was conducted on distributed compute system.

*Figure 3 - Streamlines and a flow speed colormap are shown in 3D and projected onto the geometry of a car, as determine by a computation fluid dynamics (CFD) simulation. Author: Renato N. Elias, Associate Researcher at the CFD Group.*

\textsuperscript{7} http://www.nvidia.com/object/cuda_home_new.html
\textsuperscript{8} http://www.vtk.org
Natural sciences

Astrophysics is also challenged by multiscale spatiotemporal data. In Figure 4 the magnetic, particle, and gravitational components of the formation of a star are depicted at one event of the complex, multi-energy process.

Figure 4 - The magnetic jet from a young stellar object rams into the magnetized molecular cloud from which the star is born by collapse. Data simulation by Rolf Walder and Doris Folini with the A-MAZE code. Adaptive refinement techniques were used to compute the visualisation to different levels of details in different regions of the volume. Author: Data courtesy of the Swiss National Supercomputing Centre.

Biomedical sciences use of visualisation techniques has also grown significantly in recent decades. Applications that involving visualisation abound, including diagnostic medical imaging, surgical guidance, cell migration analysis, genomics, proteomics, molecular modelling for drug discovery, and more. Figures 5 illustrate only a small portion of the range of scales and visualisation methods to be considered. More details on this application area will be given in the exemplary problems section.

Figure 5a – Examples of scientific visualisation in biomedicine. Visualisation of cerebral aneurysm hemodynamic simulations extracted from models based on different imaging modalities (in this case, CTA and 3DRA). Results correspond to streamlines, wall shear stresses and oscillatory shear index. Courtesy of the @neurIST project.
Figure 5b - Volume rendering of the density in an ethylene molecule shown in correspondence with a stick and ball model of the atoms and bonds that form that molecule, as a reference frame. Author: Jean M. Favre (CSCS)

Figure 5c – Visualisation generated using 3D Slicer that includes pre- and post-treatment segmentations of a tumor to illustrate its change, DTI tractography seeded the peritumor boundary, structural and DTI data colored by orientation, and segmentation of the ventricles (included for context). Author: Wendy Plesniak, Brigham and Women’s Hospital, Boston, USA
1.2.3 Interactive visualisation

Another important aspect, we will have to deal with in the MSV project, is that visualisation of biomedical data should be interactive, allowing the user to navigate within the data to gain a better understanding of the details and complexity of the information represented in them. Furthermore, in many cases interaction not only enhances the understanding of the data, but it becomes essential when the amount of data to be represented does not fit on the computer screen, or even in the computer’s available memory.

Interactive visualisation is a branch of HCI and visualisation in computer science that involves the study of how humans interact with computers to create, understand, and manipulate representation of information and how that process can be made more efficient.

For a visualisation to be considered interactive it must satisfy two criteria:

- **Human control**: control of some aspects of the visual representation of information, or of the information being represented, must be available to the user, and

- **Response time**: changes made by the user must be incorporated into the visualisation in a timely and effective manner.

1.2.3.1 Human control

Standard types of input that can be used in the human interaction phase are for example: picking some part of an existing visual representation, locating a point of interest, stroking a path, choosing an option from a list, and/or inputting numbers or text. All of these actions require a physical device. Input devices range from the common – keyboards, mice, graphics tablets, trackballs, and touchpads – to the more advanced – wired gloves, 3D mouse, haptic or other Virtual Reality tools. Recently, attention has increased on alternative interaction modalities such as human gestures devices or human tracking systems such as Microsoft’s kinect, Nintendo’s Wiimote, and multitouch panels similar to Apple’s iPad. These input methods can be used to control both what information is being represented as well as how the information is being presented. When the information being presented is altered, the visualisation is usually part of a feedback loop. More frequently, the representation of the information is changed rather than the information itself.

All these standard and new ways of interacting with data may be applicable to MSV, or new input methods, specific to MSV, may be needed. For example, MSV-specific control paradigms may be needed to interact with data that are not completely in the screen or that may be hidden by other objects at vastly different scales in time and space.

The development, evaluation, and refinement of alternative input devices, however, is outside of the scope of the MSV project discussed in this White Paper. We have instead chosen to focus our efforts on designing an interaction paradigm that is effective using a standard mouse-monitor environment. In this standard environment, beyond creating an intuitive user experience, we must also create one that responds rapidly to human cues, as discussed next.

1.2.3.2 Rapid response to human input

The term **interactive frame-rate**, measured in frames-per-second (FPS), is often used to measure how much interactive a visualisation is. Frame-rate measures the frequency with which an image (a frame) can be generated by a visualisation system. Essential for good interactivity is to maintain a high and consistent frame rate [Constantinescu, 2000]. A frame-rate between 30 and 15 frames per second (frame/s) is usually considered acceptable, while below 6 frame/s would be considered poor [MCCarthy, 2004; Constantinescu, 2000]. The interactive navigation through large datasets is usually limited by the graphics hardware capabilities and thus the maximum number of frame rate a graphical hardware can achieve depends directly on the complexity of the rendered image [Constantinescu, 2000]. However, frame-rate alone does not completely characterise interactivity as also latency is highly affecting the user perception of the interaction.

Experiments have shown that a delay greater than 10 seconds leads the user to assume that there has been an error in the processing. Other studies have confirmed the importance of the “immediacy of feedback” in learning. That is, the longer the delay between an action and its consequence, the longer it takes for a user to learn a behaviour. Yet
other studies have shown that for many tasks, the tolerable waiting time between when an input is provided and a visual representation is updated is about 2 seconds and that tolerance is reduced to 0.1 second for the user to feel as if the system is interactive [Bouch, 2000; Nah, 2004]. We have therefore chosen 6-10 FPS as the limit that an interactive MSV software application should operate when rendering data based on human input. If the system cannot provide such immediate response, continuous feedback should be provided to the user by reporting processing progress [Myers, 1985].

Several approaches have been explored to provide people with rapid visual feedback based on their input. Some include:

- **Parallel rendering and GPU acceleration**: more than one computer or video card is used simultaneously to render an image. There are different ways to technically achieve this parallel rendering like each computer can render a different region of a single frame and send the results over a network for display, or each computer can render an entire frame containing a subset of the information.

- **Progressive rendering**: where a frame-rate is guaranteed by rendering some subset of the information to be presented, like global scene illumination, and providing incremental (progressive) improvements to the rendering once the visualisation is no longer changing.

- **Level-of-detail (LOD) rendering**: where simplified or sub-sampled representations of information are rendered to achieve a desired frame-rate while the user is providing input, and then the full representation is generated once the person has ended manipulating.

- **Frameless rendering**: where the visualisation is no longer presented as a time series of images, but as a single image where different regions are updated over time.
1.2.4 Multiscale data

Besides visualisation and interaction aspects, it is important in the MSV context to define which types of data are going to be taken into consideration. We usually refer to the data we are interested in with the term “multiscale data” and its difference with respect to multidimensional and multivariate data is here described.

In engineering, physics, meteorology and computer science, multiscale modelling is the field of solving physical problems that have important features at multiple scales, particularly multiple spatial and/or temporal scales. Data associated to this type of modelling and including information at different spatial and temporal scales are called multiscale data.

As it can be seen from the above definition multiscale data are different from multivariate (data with several variables for each sampling unit) and multidimensional data (data being represented on more than one dimension like 2D, 3D, 4D etc).

Besides the differences illustrated, it should be noticed that in the case of biomedical data visualisation often we can be in presence of set of data which include multiscale, multivariate and multidimensional information at the same time. A proper combination and appropriate representation of all those aspects are important to the users in the data interpretation and understanding, and increase the complexity of the challenge to be solved by MSV.

In addition to this, biomedical data are often multiscale both in the spatial and temporal dimensions.

During data visualisation, the user focuses on a time-space window, which is the one within the human perceptual range, just spanning a part of the whole domain. Thus, even if the whole domain is available in the data (both in space and time) the user does not look at all the data at the same time but only focus on a sub-set of them (Figure 6).

By moving this time-space window from one scale to another, the user has a different perception of the information. Another important aspect of multiscale data is that most of the time there are gaps in the scales; thus during the movement of the time-space window continuity is not guaranteed by the data, but in any case a smooth transition should be provided to the user.

Usually the movement of the time-space window will not be un-constrained but limited by the type of biomedical application which is under consideration; for example, when looking and molecule interactions, the user is not interested in the meters/hours scale and the same happens, all the way around, at body level with data like for blood flow, heart studies. Two illustrations of possible ways of constrain user interaction and manage multiscale, heterogeneous data are showed in Figure 7.
Figure 7 - Left, a path through time and space contains waypoints at which additional details are available. By clicking on a waypoint, an alternative visualisation of the enclosed data is given, perhaps a PDF document is displayed, an interactive volume rendering is shown, or an internet search is issued. Right, instead of waypoints being along a path, the waypoints are embedded in volumetric data. Again, each waypoint (shown as ellipsoids in this visualisation) indicates where alternative data is available. Those data may be the pathology reports from a biopsy sample taken at that location, or biophysical recordings from probes positioned at that location, or links to publications that discuss the anatomic feature in detail.
2 State-of-the-art

2.1 Non biomedical domains

In domains other than the biomedical one, there have been attempts to visualise and interact with multiscale data. Even if the chosen interaction approach highly depends on the type of application, it is interesting to have a look at what has been already proposed to eventually identify possible methods useful when dealing also with biomedical data.

Astrophysics

Computational astrophysics has rapidly developed in the recent years, enabling, since only recently, multiscale and multi-physics simulations of entire stars, planets, galaxies, and cosmic explosions [Muller, 2004]. Within the past decade, computer facilities and development of computational methods have literally exploded, allowing huge progress in the understanding of star and planet formation, a complex field which combines radioactive- and magneto-hydrodynamics, the physics of shocks, supersonic turbulence and elaborated computational.

Moreover, the latest generations of telescopes gives an unprecedented amount of observational data. However, a proper interpretation of such data always depends on computational models that explore the origin and propagation of the observed photons. On the other hand, astrophysical objects are multiscale and multi-physics data. Flows, magnetic fields, and radiation are in perpetual dynamical interplay. Non-linear couplings among the different ingredients are crucial for the understanding of objects like stars, planets, galaxies, and cosmic explosions like supernovae. Thus, in terms of visualisation astrophysics has very similar problems to those we are trying to address with MSV; the data managed are multiscale in space ranging from $10^{25}$ to $10^{-15}$ meters [Hanson, 2000]. It is also often necessary to store and analyse huge datasets which is also one of MSV challenges.

For all these reasons, it is interesting to understand which approaches have been undertaken in astrophysics. The most interesting findings are summarised here.

The AMUSE project provides a Python based framework for the multiscale simulations of dense stellar system. The framework is meant to be executed only on Unix operative system and presents some parallel executing codes to manage large simulations. Some ideas on the data storing and multiscale data management can be taken into consideration in MSV.

Another interesting approach can be found in [Hanson, 2000]. The paper has the aim to describe a method to display data over more orders of magnitude (without significant arithmetic error) by the use of a combination of strategies:

- Definition of an homogeneous power coordinates and interpolation system;
- Systematic treatment of data at unit scale with transformation matrices;
- Systematic blending of icon and very large data group representation with object disappearance criteria and smallest visible object definition;
- Merging environment maps with octrees to make very distant data visible;
- Support to multi-resolution maps;
- Constrained multi-resolution navigation.

9 http://amusecode.org/doc/
White paper on multiscale visualisation (draft)

Cartography and geography

In geological sciences the possibility to deal with the dimensionality of the data started after 1990 and it is now challenged with an increasing volume and complexity of data to be represented. In terms of visualisation, the need in this field is to allow on-the-fly rendering across a broad range of scales.

Google earth\textsuperscript{10} is probably the most famous example. It is a virtual globe, map and geographic information program released with a free licence in 2005 by Google obtained by the superimposition of images from satellites and aerial photography. For large part of the globe only 2D images are available from vertical photography, while for others 3D images of terrains and buildings are also available (created with a digital elevation model) [Yoshimura, 2009]. The resolution is not uniform over the globe and in the most popular area goes down to 15 cm [Yoshimura, 2009]. The interaction is based on a zoom-based and Level Of Detail (LOD) approach: the more the user zooms at a higher resolution the information is rendered. The user is also allowed to add layers to the images with additional information increasing the multivariation of the information.

Even if more limited in terms of multiscale range than MSV and not dealing with the temporal scale, the user interaction has been proven to be very effective and user-friendly, and it might be taken into consideration in the navigation also of biomedical datasets.

Another software example is CoViz4D\textsuperscript{11} which is more focused on the visualisation on variation over of the data. Also in this case the navigation at different spatial scales is based on zoom-in/zoom-out interaction.

An application example which focuses more on the time scale aspect is the climate change visualisation. In this field, there are software solutions which allow the user to visualise huge amount of data over time\textsuperscript{12,13}. However, the spatial multiscale problem is usually less relevant than in other applications, while great attention is posed to the multivariate data representation.

2.2 General tools

A considerable effort has been spent in the last three decades in designing effective visualisation methods for inspecting and analysing the ever-growing data provided by experiments, medical equipments, and simulations. Generally, they have a broad scope of applicability and, as such, may cover a central role also in biomedicine.

The computer graphics community concurrently with the formidable progress in systems architectures has recently proved the feasibility of interactively and accurately rendering static volumetric datasets containing billions of multivalue grid elements on high-end distributed machines [Fogal, 2010; Guitián, 2010] as well as on commodity workstations [Crassin, 2009; Guitián, 2010] by means of sophisticated GPU (graphics processing units) algorithms and level-of-detail mechanisms. Not only visualisation efficiency is fundamental for augmenting scientific productivity, but the prerogative to provide feature-driven contextual information is essential when, for example, too many data are concomitantly displayed or some features of interest are hidden by occluding regions. This is more important for gaining insights into large and high-dimensional datasets and has been effectively addressed in [Guitián, 2010; Barakat, 2010].

Several simulations, experiments and imaging techniques provide time-varying data. The consideration of the time dimension poses additional challenges for efficiently exploring large volumes. These have been circumvented in some cases by means of advanced compression/decompression techniques [Nagayasu, 2008; Wang, 2010]. However, two common drawbacks shared by the aforementioned visualisation methods are the long pre-processing times, and the unavailability of their implementation in unspecialized solutions, e.g. existing open-source software. Furthermore, little or no attention has been dedicated to address truly heterogeneous and multiscale datasets by current visualisation approaches well-suited for handling large data.

\textsuperscript{10} http://www.google.com/intl/it/earth/index.html
\textsuperscript{11} http://www.dgi.com/coviz/cvmain.html
\textsuperscript{12} http://vis5d.sourceforge.net/doc/
\textsuperscript{13} http://www.confusion.com/
Other interesting approaches can be found in information visualisation methods like the one based on data cubes which are commonly used for abstracting relational databases [Stolte, 2003]. For example, the data cubes approach is based on both data and visual abstraction, and on the navigation trough dimensions by zooming a graph where nodes have different level of details.

2.3 Biomedical field

Computer-aided visualisation has shifted the way in which biomedicine and biology are approached. The insights gained through simulation and visualisation tools permit us to reveal fundamental aspects and sometimes lead to the formulation of new hypotheses, which can be tested by means of further experiments. The inherent complexity of biological systems, in turn, justified the development of new simulation, visualisation and imaging techniques.

Several software developments have enabled the effective exploration of large molecular assemblies [O'Donoghue, 2010]. For example, features of interest, such as spatial correlations, can be viewed on 3D structures so as to provide a mean to compactly visualise multiple properties at the same time. Additionally, the ability to superimpose structures may enhance discovery further. Global illumination effects can aid the understanding of the spatial arrangement of complex structures considerably [Gribble, 2008], albeit at greater computational and memory costs.

Visualisation of protein interactions, gene expressions and metabolic profile data has been made possible and effective through approaches based on data clustering, networks, connections between protein complexes and dimensionality reduction. The discovery process may be substantially enriched with editing tools that can synthesise new configurations or curate existing ones [Gehlenborg, 2010].

In the review recently appeared on Nature Methods [Walter, 2010], several other aspects encountered when visualising the spatial scales of the cells up to those of the organisms are pointed out. The integration of data originating from multiple simulation tools means solving the recurring problem of having them available in different file formats, and different reference systems. Moreover, the ability to efficiently and intuitively query, analyse and compare data is paramount and has recently reached a satisfactory status in some contexts; however, the visualisation of complex biological datasets is beyond the capabilities of existing software packages. Long-term patient-specific studies can reveal important time-varying disease patterns; this pushes forward the need to adopt robust registration techniques and methods that can significantly correct scanning-motion issues, nonlinear deformations and geometrical distortions. In addition, the proficiency to relate data and physical/biological phenomena may require employing algorithms developed within the fields of image processing, statistical analysis, mathematical modelling and simulation and machine learning, all possibly approached via user-defined visualisation modules, which is valuable.

The importance of on-the-fly computation to interactively refine and iterate analyses has been already emphasised [Nielsen, 2010] together with the role played by visualisation methods that enable the comparison of data of the same category, and the possibility to hide part of them when too much information is being displayed on the screen.

Another aspect of practical interest for characterising and sharing multivariate and multiscale data, common to various research domains, is represented by the skill of annotating data provenance, functionally important features alongside with their integration, and the amalgamation of simulation and clinical information sources. In particular, Gehlenborg et al. [Gehlenborg, 2010] stress that “truly integrated visualisation of systems biology data across the entire range of possible data types is still very much in its infancy”.

Moreover, the many limitations of modern visualisation tools demand more intuitive software interfaces, new ways to represent spatiotemporal datasets, and the ability to seamlessly navigate across different resolution levels whilst being able to interactively aggregate data at run-time where needed most.

If we look at software tools, we can find a couple of relevant examples here briefly described.

The BrainMaps project\(^4\), which is an interactive multi-resolution brain atlas that is based on over 20 million megapixels of sub-micron resolution, annotated, scanned images of serial sections of both primate and non-primate brains. The interface is a web-based one and it allows zooming in the different scale levels being built on top of one

\(^4\) [http://brainmaps.org/](http://brainmaps.org/)
dataset available at very high resolution. Even if very effective, this example does not deal with the time scale problem and the data are only represented with 2D slices which can be limiting in some biomedical applications.

Another approach with a web-based interface has been recently presented by Google: the Body Browser tool\(^{15}\). The tool allows to explore the body organs using a navigation widget and different level of transparencies. The tool is new and so far mainly a demonstrator for the WebGL technology but for sure it will increase its capabilities in the future. At present its main limitation in the MSV perspective is that the data are not multiscale (just organs are available to be navigated) and are not patient-specific (what is visualised is a pre-processed model a “generic” human). In any case, the ideas and widgets to manage transparency and navigate objects, which are one hidden from another, might be an interesting approach also for some MSV use cases.

In terms of software tools, another recent development in the biomedical domain is one of the results of the EU-funded LHDL project [McFarlane, 2008; Viceconti, 2010]. Two prototypes were implemented based on the Multimod Application Framework\(^{16}\) providing an interactive visualisation environment for biomedical data defined over different spatial or temporal scales.

In summary, the temporal data visualisation and interaction approach, tested with Electromyography (EMG) scalar data at high frequency, was based on the perceptual limitation of our visual system: there is a physiological limitation to the frequency humans can visually perceive. However, by providing the user the possibility to slow down or to speed up the time in a certain interval of the data, the user was allowed to explore the entire range of frequencies represented in a single signal, or in multiple fused signals. For what concerns the spatial scales, even if more complex, as dealing with 3D dataset (CT and microCT scans), the basic approach was the same. Given a current view, the user could interactively define the level of zoom: when an object is too small or too big to be perceived, it disappeared from the visualisation using a placeholder only for objects that are smaller than the current view. With this approach all data objects in the scene are always visible, independently from their size allowing the user to have an understanding on where the different data are. These two prototypes are very interesting and already provide some understanding on a possible type of interaction which can be of interest for MSV objectives. The spatial scale example examined many aspects of the click-and-zoom interaction in 3D, such as a minimal set of GUI controls, the speed of zoom, the behaviour of the sub-scale placeholders, coupling between the scale and the GUI, and the importance of knowing which objects were the subject of the camera’s current “attention”. The main limitations to be overcome with respect to these prototypes are: the missing coupling between spatial and time multiscale navigation; the generalisation of the approach to different types of data (so far only 3D volumes were considered); the visualisation of placeholders embedded inside larger objects; the problems of units and ill-conditioning when combining scales many orders of magnitude apart; and the management of the glyphs in case of a high number of data in each scale should be improved.

Figures 8 show a sequence of example images from the click-and-zoom view. The data consists of a femur volume CT image, visualised in a slice view, with embedded microCT and nanoCT images of the bone structure.

---

\(^{15}\) [http://bodybrowser.googlelabs.com/body.html#](http://bodybrowser.googlelabs.com/body.html#)

\(^{16}\) [http://www.openmaf.org/](http://www.openmaf.org/)
Figure 8b - Slice view of femur head (50mm) showing microCT data of trabecular structure, with further placeholder for embedded nanoCT.

Figure 8c - Slice view of trabecular structure (20mm) with placeholder for nanoCT; the femur is out-of-scale (too large) and is no longer visualised.

Figure 8d - Result of clicking on nanoCT placeholder: the dataset is expanded to show the single trabecula (700 micron). The larger out-of-scale datasets are not visualised.
3 Exemplary problems

In this section some use cases are briefly described, in order to motivate better this problem, convincing the reader about the importance of multiscale visualisation in very different biomedical domains. Additionally, from the examples shown here and others that are going to be collected in the future, MSV partners will be able to extract patterns that are common to more than one application, and therefore will be good candidates to perform the design of general solutions to the visualisation needs in the medical imaging field, that eventually, can become important for other areas of research as well.

Contact has been established with VPH, US and New Zealand biomedical projects and information on their multiscale data visualisation needs have been collected. As a result of this assessment, we summarise the output obtained from the most relevant VPH projects and applications. In particular, for each project, a summary of the data produced in the different scales have been collected together with the availability of software tools for their visualisation and the possibility to share example data with the MSV consortium (Table 1).

<table>
<thead>
<tr>
<th>Project</th>
<th>Modelling approach</th>
<th>Multiscale</th>
<th>Able to provide data</th>
</tr>
</thead>
<tbody>
<tr>
<td>LHD L</td>
<td>Medical imaging and simulation</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>VPHOP</td>
<td>Medical imaging and simulation</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>@neurIST</td>
<td>Medical imaging, simulation and genetics</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>COAST</td>
<td>Simulation</td>
<td>Yes</td>
<td>Pending consortium approval</td>
</tr>
<tr>
<td>ARCH</td>
<td>Medical imaging and simulation</td>
<td>No</td>
<td>Pending consortium approval</td>
</tr>
<tr>
<td>SCATh</td>
<td>Medical imaging and simulation</td>
<td>No</td>
<td>Pending consortium approval</td>
</tr>
<tr>
<td>euHeart</td>
<td>Medical Imaging and simulation</td>
<td>Yes</td>
<td>Pending consortium approval</td>
</tr>
</tbody>
</table>

Table 1 – Summary of the VPH projects involved in the assessment exercise

The projects were then grouped in terms of similar biomedical field of application.

- Orthopaedics and trauma modelling
- Vascular modelling and treatment planning
- Heart modelling
- Cancer modelling

Few use case examples in the categories above are given in the next section.

3.1 Use cases

Cerebral aneurysm modelling and treatment

The example described here concerns the analysis of intracranial aneurysms, see [Villa-Uriol, 2010] for details. In Figure 9 is also shown a set of related data in multiple scales.

- Geometrical studies. The basic image information to start with this problem is 3DRA datasets, but MR-TOF and CTA data can also be used. Although in clinical practice only one of these 3 modalities is typically acquired for a patient, it is not rare to have two of them available for diagnosis or treatment planning. Then, to analyse geometrically the case under study with some of these modalities, a 3D computational model of the vessels is obtained. The visualisation of these image modalities together with the derived models, is a resolved issue, but
joining these visualisation capabilities to support other datasets needed to completely model this problem, is still open as discussed below.

- **Vessel motion analysis.** In addition to the geometrical analysis, it is possible to perform a dynamic study of the vessel wall. The data required here (namely, 3D+t) as well as the information obtained from this study, has increasing visualisation requirements, and is limited to a cardiac cycle period.

- **Hemodynamic studies.** In order to assess the influence of blood flow in the aneurysm, hemodynamic simulations are commonly performed using the 3D surface models previously obtained. These studies would give us a high amount of information over one cardiac cycle, including the blood flow velocity, wall shear stress, pressure, etc., whose visualisation is not an easy task.

- **Integration of biomechanical models.** In order to fully understand the mechanisms of aneurysm formation, progression and rupture, models simulating these processes can be incorporated, as well as approaches that simulate the formation of thrombus that may occur within an aneurysm after the implantation of endovascular devices. The information derived from these models has different spatial and temporal scales that are not trivial to integrate with the previous data for visualisation purposes.

- **Genetic data integration.** It is also known that genetic patterns can influence the probability of appearance/rupture of aneurysm. Therefore, the incorporation of this information is also quite valuable, and its integration in the visualisation framework will also help and speed up the understanding of the clinicians about this disease.

- **Follow up studies.** Additionally, follow-up studies can be undertaken after several months or years of the initial study, increasing the temporal scale. Visualisation of the image data at these distant time instances is also desirable for a better understanding of the evolution of the disease.

- **Clinical data.** Finally, integration of clinical data in the decision-making process can be crucial. Visualisation of clinical information, either as labels overlays, colours information, or symbolic, can be a good solution to merge in the same space many useful information required for the treatment decision.

![Figure 9 - Two different use cases from VPH: vascular modelling and cardiac modelling.](image-url)
Heart analysis modelling

As a second example, some of the visualisation issues that are required for the study of this complex and important organ are reported.

- **Anatomical studies.** The basic information to explore the anatomy of the heart are 3D image modalities such as CT, MR, or US, from where it is possible to analyse the shape and the morphology of different structures (ventricles, valves, coronaries etc). Again, it is common to have different modalities of the same patient, and their concurrent visualisation within the same reference system becomes important.

- **Dynamic studies.** Of course, capturing and studying heart motion is extremely important, and therefore 3D+t image modalities such as CT, MR, US are usually employed. Dynamic visualisation of these medical images, and fusion between them is also of key importance in these studies.

- **A priori information (atlases, models).** Models and atlas resembling common population features, are usually required 1) for inclusion of a priori information in some studies, or 2) for comparison between individual subject against a population. The visualisation of these kind of data together with common image datasets, have particular visualisation and interaction requirements due to the difference in reference systems, sizes, resolution, etc.

- **Fibre distribution.** Other modalities such as Diffusion Tensor Imaging (DTI) are important in the study of the orientation of the heart fibres, that are correlated to the mechanical behaviour of the heart. Hence, the visualisation of tensor field is an additional issue to tackle.

- **Microscopic studies.** The correlation between the macroscopic and microscopic behaviour is also important, and visualisation of the microscopic level, (cell level) will allow, for instance, to visually confront the macroscopic and microscopic fibre orientation in given locations of the heart.

- **Hemodynamic simulations.** The dynamic studies performed with standard image modalities can be also complemented with hemodynamic studies, where blood flow can be computed and visualised together with the relevant anatomical information.

- **Electro-physiological studies.** Together with the image modalities, the physiological signals acquired through electrodes are crucial to study electrical behaviour of the heart, and are used, for instance, to model and plan the success of resynchronization therapies. Correct visualisation of these signals in space and time is therefore and important issue in these cases.

- **Follow up.** Finally, as in the previous example, follow up studies, give rise to the problem of visualisation of individual data in distant time instances, and is another topic to be solved.

Musculoskeletal modelling

- **Anatomical study.** In many different musculoskeletal clinical applications there is the need to extract from the medical images information on both the hard and the soft tissues; these information are acquired with different imaging modalities (CT and MRI) which need to be visualised together to have the complete information on the subject’s musculoskeletal system. At the same time, the evolution over time of anatomical characteristics as visible in the medical images is often used in the clinical context, thus the user has to deal with medical imaging data over time.

- **Different data sources.** In clinical applications, like for example the prediction of osteoporotic fractures, the medical images should be visualised together with geometries of different structure and/or with simulation results coming from Finite Element Models and/or motion analysis data. The merging into the visualisation of these heterogeneous data types does not have yet a complete and fully accepted solution.

- **Multiscale modelling.** For example, in the prediction of osteoporotic fracture, the fracture risk is a multifactorial risks in which multiple-scales processes contribute. Thus it can be evaluated only if data from different scales are available, like: whole body postural and loading data, medical imaging at the organ level,
information at the tissue level on the bone structure (i.e. microCT), cellular level information on the remodelling capabilities of the tissues, etc.

- **Clinical data.** Finally, integration of clinical data in the decision-making process can be crucial. Visualisation of clinical information, either as labels overlays, colours information, or symbolic, can be a good solution to merge in the same space many useful information required for the treatment decision.

### 3.2 First categorisation

When a multiscale visualisation problem is faced, there are a series of problems to be addressed. These issues have been ordered into categories classifying them depending on what they entail and require. It seems natural to think on a first high level categorization to separate the kind of problem at hand, as visualisation, interaction and data management. Still, all these categories are not exclusive. From this sub-categories in each one of them can be derived. Three major categories of problems (and sub-problem therein) are here reported:

1. **Visualisation of multiscale data:**
   - **Visualisation across scales** in time and/or space (different orders of magnitude): going from the molecule level (nano-meters and nano-seconds) to the systemic level or even entire populations (meters and years). This is one of the most challenging problems to address, since VPH problems deals with modelling of different levels, and the interaction between them. In this sense here we can divide this problem into two:
     - **Spatial:** when the data covers multiple spatial levels: molecules, genes, cells, tissues, organs, systems, populations.
     - **Temporal:** when the data covers multiple temporal levels: from nano-seconds to years.
   Of course both types of problems are coupled, being molecules and cell processes orders of magnitude faster than organs or systems processes and being both types of multi-scaling present in many biomedical applications.
   - **Gaps between different scales** (not all levels are available): the different spatial and temporal resolution of the acquisition devices, makes the representation of biomedical data a multiple data problem, where only sparse scales are modelled. With the current technology, gaps between different levels are always present. With these gaps, the amount of information described by one level in a given space scale, will cover a small amount of space in the level below it, thus, only a very small proportion of the data at a given level can be actually available at the lower scales. It is evident that this will be one of the major issues when dealing with multiscale problems. For this reason, visualising in a continuous way across different scales either is not possible or it has to be simulated.
   - **Integration of data in the same reference system:** different datasets have to be placed at a correct spatial localization using a common reference system, even if they pertain to very different scales. The proper alignment in a common reference system or the possibility for the user to identify correspondences among the data is mandatory for an adequate interpretation of several datasets either if they are being visualised simultaneously or not. This problem arises from the visualisation of data from different acquisition systems, different acquisition procedures, data obtained from simulations, etc.
   - **Multivariate data:** even within the same dimensionality, the data can be of different types and nature, multiple medical image modalities (CT, MR, US), different sizes, spatial resolutions, origins, and also computational models; this entails concurrently approaching several visualisation capabilities, which need to be merged into the same virtual environment [Van Sint Jan, 2006].

2. **Interaction with multiscale data:**
   - **Access at interactive frame rates:** the ability to explore such complex datasets at interactive rates is essential for effectively capturing and understanding the phenomena of interest [Lum, 2006], but its
realisation represents a major challenge underpinning the need to improve or extend existing rendering methods.

- **Scene exploration**: the interaction with multiscale data allows the user to actually browse them, being able to change the point of view of the observer, change between scales, zooming in and out within the same scale, etc.

- **Selection of data**: finally the user needs to be able to select a portion of the data, in a given scale for many possible reasons: visualise it better (zooming in or out), centre the view on it, hide or show it, save it, remove it, crop it, perform measures, perform post processing, etc. This can be implemented in different ways, but must be compatible with the multiscale nature of the data, for example using a command panel, where the user can select data from a list, or directly from the viewers, picking a zone, selecting a ROI, using a region based criteria, etc.

3. Management of multiscale data:

- **High dimensionality datasets**: in addition to the integration of different data dimensionality, it is more and more common to have high dimensionality datasets, such as time varying vector fields (6D+t), or tensor fields [Cardenes, 2010] (9D, 18D, 29D, etc.). These entail adopting special visualisation paradigms due to the large amount of memory and processing requirements in the CPUs and GPUs, and the difficulty in pursuing correct and meaningful visualisation. The visualisation of these kind of datasets, have been explored in different fields, especially in Diffusion Tensor Imaging (DTI), and Diffusion Spectrum Imaging (DSI) where tensor fields are rendered to explore the fibre structures of the brain. The integration of these solutions together with the rest of multiscale visualisation paradigms will certainly provide the most advanced visualisation solutions for biomedical data.

- **Heterogeneous dimensionality of data**: there is also a need for the integration of data with different dimensions: 0D, such as tags or scalar values, 1D, such as physiological signals, 2D, such as images, 3D such as volumetric data, 3D+t such as time-varying ones [Wang, 2010]. The fusion of information with different dimensionalities represents another challenging issue when dealing with different biomedical data from different sources. There exists several solutions to these kind of problems, such as for instance, the use of volume and surface rendering techniques, or the use of colour maps to represent scalar values, etc. However, the integration of these techniques along different scales requires a common strategy for a seamless visualisation of multiple data.

- **Very large (Gb) amounts of data in each scale**: at the same scale, medical images, biomedical signals, computational simulations of biological processes, etc. require a high amount of space either for storage, and for loading and visualisation. For this reason, smart solutions for efficient rendering of large datasets are required [Crassin, 2009; Agrawal, 2010]. Due to the increasing performance of acquisition machines, simulation servers, computer software, etc., the amount of pre and post processed data available for clinical studies are also increasing. Therefore, there is a need for fast filtering, storage and visualisation of the data, to show what is relevant to the clinicians to take diagnosis and treatment decisions, and also to researchers to implement and validate their models.

- **Management of data classification and annotation**: in many biomedical applications the information on what is data visualised representing (i.e. which organ, which tissue etc.) and its anatomical relationship with the other data might be important for a correct interpretation and navigation of the data themselves. There are available metadata ontologies for the anatomical representation17,18 and also software tools for the annotation; however, this becomes another challenge when the textual information of the metadata is associated to multiscale binary objects.

4 MSV visualisation problem formalisation

In this section, we try to provide a formal representation of the visualisation process and in particular of its multiscale aspects.

4.1 Visualisation process

In the literature it is possible to find different attempts to formalise visualisation [van Wijji, 2005; Onodera, 1990; Hutchins, 1999]. However, each of them focused on a specific aspect of visualisation, like the process, its cost model etc. An example of visualisation process representation is given in Figure 10.

![Figure 10 - From [Johnson, 2006]. The visualisation discovery process. Data encompasses the range from a single bit, to time-varying 3D tensor fields, to multi-modal data sources requiring alignment, registration, and fusion, and to non-spatial information sources integrating broad areas of human knowledge. The visualisation specification includes the hardware, the algorithms, and the specific parameters. Users adjust the specification of the visualisation, requiring the introduction of interactive controls. The resulting image will often be an image in the usual sense, but it can also be an animation, or auditory or haptic feedback.](image)

However to the authors’ knowledge, nothing is available with respect to the multiscale aspects of visualisation. In this section, we tried to provide a formalisation of the visualisation aspects which relate with multiscale data representation. A glossary of different terms used in the visual process can be found in Annex 1.
4.1.1 Definition of visualisation

First of all we have to define in formal terms the objects which are visualised. These can be generally represented by a dataset which can be multidimensional (i.e. a 3D volume), a series of attributes or variables (i.e. attenuation of radiation through a tissue), and a mapping of those variables on the elements of the dataset (i.e. the attenuation value in each cell of the 3D volume); which results in:

- a multidimensional dataset \( D = \{e_1, \ldots, e_n\} \) contains \( n \) samples points or data elements \( e_i \);
- \( D \) represents the data attributes \( A = \{A_1, \ldots, A_m\}, m \geq 1 \);
- The data elements encode values for each attribute, that is, \( e_i = \{a_{i,1}, \ldots, a_{i,m}\}, a_{i,j} \in A_j \).

As described in section 1.2.1, visualisation converts the raw data into images that are presented to the viewer, and it can be represented by a mapping \( M(V,\phi) \) between the dataset attributes above and a visual feature (i.e. colour mapping) thus

- \( V = \{V_1, \ldots, V_m\} \) identifies a visual feature \( V_j \) used to display data attribute \( A_j \);
- \( \phi_j : A_j \rightarrow V_j \) maps the domain of \( A_j \) to the range of displayable values in \( V_j \).

Based on these definitions, visualisation is the selection of appropriate mapping \( M(V,\phi) \) together the viewer’s interpretation of the images produced by \( M \). An effective visualisation chooses among all possible \( M \) functions the best one to support the exploration and analysis tasks the viewer wants to perform in the specific field of application. The whole chain of the visualisation formalisation can be represented as in Figure 11.

![Visualisation formalisation](image)

**Figure 11 - Visualisation formalisation**

Visualisation is thus dependent on a number of aspects: the physical characteristics of the display device (e.g. resolution in terms of the total number of pixels, and the physical size of the display), on the viewer’s visual acuity, on the visualisation technique (the methods used to map a data element values to a visual representation), and on the properties of the data (e.g. its dimensionality, number of elements), and the specific task to be performed by the viewer.

If we consider more in details, the case of a multiscale and/or multidimensional dataset:

- The elements of the dataset \( D \) are sampled over space and time, where the attributes \( A_j \) they encode are space-time related.
- In order to be representative the visual feature \( V_j \) used to represent the attribute \( A_j \) is also space-time related.
- The mapping \( M(V,\phi) \) usually preserves the space-time ratios (i.e. zoom, or isotropic scaling)

In choosing the appropriate visualisation approach (the definition of the mapping \( M \)) an important aspect is objects distinguishability or resolution acuity which represents the condition in which two visual features are seen as distinct by the user. As also from the definitions in Annex 1, this characteristics depends on the resolution of the display device and on the visual angle, and in terms of the above \( V_j \) can be schematised as:
Figure 12 - Distinguishability graph

In reality the distinguishability graph has also upper boundaries for the visual angle (maximum view angle) and for the display resolution (maximum display size). Usually this is not relevant, because we can choose the mapping function $M(V, \phi)$ to scale the size of visual feature $V_j$ (zoom) used to represent the attribute $A_j$ so that it fits the upper and lower boundaries (i.e., between the pixel and the whole screen).

But what happens if we have to display simultaneously multiple $V_j$ that map very different space-time regions? No single scaling can fit them all in the screen!

### 4.1.2 Degree of multiscaling

In the above conditions, it would be important to have a formal indicator of the multiscaling degree. Thus, we propose to define a degree to indicate severity of the multiscaling ($S_d$ and $T_d$):

- In space: $S_d = L/a_s$ where:
  - $S_d$ spatial dynamics
  - $L$ largest diagonal of bounding box
  - $a_s \in A_j$ is the smallest size of the smallest attribute of the dataset we need to visualise

- Similarly for time $T_d = T/a_t$
Based on the above indicators we will have for example that the following degrees of multiscaling:

- **Whole VPH:** from body $10^1 \text{m}$ to atoms $10^{-10} \text{m} \implies S_d = 10^{11}$
- **Bone example:** CT $10^{-1} \text{m}$; nanoCT $10^{-6} \text{m} \implies S_d = 10^5$

But which is the limit an MSV application can tackle? Defining a threshold value to define MSV is very difficult as representational and perceptual limits tend to overlap. Roughly speaking we can say that it is hard to perceive visual features for:

- $S_d > 100$
- $T_d > 100$.

Based on this threshold we will have that a single multiscale dataset contains data attributes to be visualised defined over space-time with a degree of multiscaling $> 100$, while a multiscale scene contains multiple datasets, each with a degree of multiscaling $< 100$, but that when displayed together exceed this threshold.

### 4.1.3 Open questions and issues

More work is needed in the future to address some still open issues in the above attempt of formalisation which are here listed:

- **Mathematical formalism is incomplete and need improvements:**
  - Definition of the limits for each element of the chain (multiple limits for colour, contrast, spatial resolution, recognition, etc.) is missing;
  - Definition of a degree of multiscaling that factors in all elements of the chain shall be investigated.
4.2 Formal taxonomy for multiscale visualisation

This section will provide a taxonomy which should cover all the cases of multiscale visualisation and interaction independently from the nature of data and type of chosen visual representation. It will be added in the future versions of the White Paper document and completed by end 2011.
5 Multiscale Visualisation and interaction

5.1 Problem definition

MSV essential problem is the visualisation of data and information that can be on a wide range of scales both in space and time (from nano scale to the body level scale). The typical data will be a 3D+t dataset of which multiple instances at different scales will have to be displayed together. However, we should also account for the presence of data of other dimensions (0D, 1D or 2D) as described in Section 4. As for the visualisation device, the MSV target will be usually a standard monitor with a resolution of about 1000x1000 and a mouse as input device. The use of advanced hardware (like VR technologies or devices with high degrees of freedom) is not considered in this phase as mandatory for the designing and development of the MSV library. However, the integration of “advanced” devices will not be prevented by the final implementation for testing or use in special application fields.

Thus in summary, different constraints will have to be taken into account in the design and development of the MSV library:

- information will be
  - on very different spatial and temporal scales, going from the molecule (nano-metres and nano-seconds) up to the body level (metres and years) (Figure 13 as example);
  - in different forms (medical images, computer models, signals, etc);
  - of heterogeneous dimensionality (1D, 2D, 3D, 3D+t);
- visualisation should be interactive even if very large (GB) volumes of data are available at each scale;
- depending on the problem at hand, there may be gaps in the scales (not all levels are available) but visualisation should, nevertheless, be continuous across the scales;
- data in different scales usually have different systems of reference: proper definition of the relative position or correspondence in a common reference system is essential for an adequate interpretation of this information.

![Figure 13 - Example of multiscale data integration and visualisation needs from the VPHOP project for the prediction of fracture in osteoporotic patients.](image-url)
5.2 Proposed interaction and visualisation paradigms

This section will provide a description possible strategies and paradigms to be adopted by MSV. It will be added in the following versions of the white paper and completed by end 2011.

It will also include information on:

- if the MSV approach to multiscale data management will be dimensional or dimensionless;
- which will be the interaction model;
- how MSV will manage the visualisation of data objects bigger than the field of view;
- if the multiscale interaction can generalised independently from the type of data/renderer.

Even if it is too early to fill this section, the MSV partners are already thinking and working on a possible approaches for the multiscale visualisation and interaction.

First of all, the zoom-based approach used in the previous LHDL project and proven to be very effective and user friendly in many other applications seems to be very promising in the data navigation given that the user is provided with inputs on where the lower scale data are placed with respect to the whole data. Moreover, a way to understand the relative position of the data is necessary: a data tree which allows also the data selection, or a small map on the corner showing what part of the big map the user is actually looking at can be useful at this purpose.

5.3 Challenges

From the point of view of the data to be managed, an extensive description of the challenges has been given in Section 4; in the following section we address a bit more in details other challenges MSV will have to deal with.

5.3.1 Computers systems

As previously mentioned in different sections of this document, many aspects of multiscale visualisation are affected by the hardware and computer system capabilities. A particularly important aspect that will arise when dealing with large data objects will be the data management and I/O on the disk and the possibility to have an efficient memory management during the different data processing.

Particularly interesting on this aspect is the VTK multi-resolution streaming\(^1^9\). VTK's streaming has an interesting feature in that it is possible to ask for data in arbitrarily small chunks.

Data streaming enables the processing of large datasets by partitioning the data into smaller pieces that can be processed sequentially in a single machine or in parallel on a multi-processor system. In Figure 14 streaming has been applied for producing a full-body visualisation of the CT Visible Female dataset using VTK. The colour bar in the figure illustrates which sections of the data were processed by select nodes of a parallel cluster.

\(^1^9\) [http://www.kitware.com/products/html/MultiResolutionStreamingInVTKAndParaView.html](http://www.kitware.com/products/html/MultiResolutionStreamingInVTKAndParaView.html)
Level-of-detail enables users to interact with large datasets by rendering data at different resolutions depending on the frame rate that the user specifies. Low resolution, interactive rendering is useful for navigation and pose selection. Once a point-of-view is selected, a higher resolution but slower rendering of the data is automatically generated. This mechanism makes it possible to offer an interactive experience to the user and to still obtain high quality visualisations. Figure 15 illustrates the use of level-of-detail on Kitware’s volume rendering application VolView. The image on the left shows the coarse level of detail used during user interaction in order to increase the frame rate. The image on the right shows the high quality rendering that is produced once the user selects a particular scene point-of-view.
Another approach that might be evaluated for the efficient I/O of large data from and to disk is a hierarchical, bricked, partition-based, out-of-core strategy to balance the usage of main and external memories [Agrawal, 2010]. Such an algorithm for exploring large volume datasets requires excellent cooperation between the rendering and data-retrieval algorithms. Often, however, rendering and data retrieval are two independent processes, which introduces some limitations.

When a user wants to import a volume that is larger than some given threshold, the volume is not loaded into the memory but is reprocessed to produce subsampled versions (resolution levels), which are stored on disk in an optimised form. The algorithm successively samples the input dataset until the size of the lowest resolution level drops below some given threshold.

The processing of the input dataset has three main steps: sampling, bricking and fast compression. As the input dataset is too large to be fully loaded into memory, it must be processed in smaller parts, with the steps being repeated for each part. In the first step, the loaded part is subsampled; next, the voxels are reorganised into blocks (bricks); then, they are stored in the output file. An indexing strategy is used that reflects both the likely manner in which the data will later be retrieved during visualisation and the characteristics of the type of data for which the technique was designed; for these data types, it was shown to provide lossless compression to about 30% of the original size in the examples studied.

Such approaches have been commonly applied in rendering geographical terrain but less frequently in the biomedical area. Further work needs to be done to make them applicable to GPUs, which will normally have significantly less addressable memory than the CPU.

5.3.2 Algorithms

One big issue when dealing with data coming from different and heterogeneous sources is that they are usually in different position in time and space and need to be registered. The appropriate technique/algorithm to be used is highly dependent on the type of data and application and the development of new general registration techniques is out of the scope of MSV.

However at least the possibility to create a correspondence among the different data in terms of position of the smaller objects within the bigger ones should be provided to the user to allow to create a structure between the data and an easier navigation.

5.3.3 Validation

An important issue for the MSV project is the validation of the proposed approached. Due to the available time-frame, an extensive validation is out of the scope of the project.

A quantitative measurement of the improvements obtained by introducing a new visualisation paradigm is difficult when these are related to the human perception and interaction. Moreover, the efficiency of a new paradigm can be measured only when a specific task has been identified which is highly related to the field of application. At the same time, being visualisation efficacy not independent from the human perception, aspects of HCI might be taken into consideration. Even if a HCI evaluation if out of the scope and expertise of the consortium, it will be important to understand if the proposed approach is useful and effective as expected when applied to real-case scenarios.

First of all, it is important to identify the end-users of the MSV results. Being the output of the project mainly an open-source software library, our primarily end-users will be programmers who have to implement a software solution with multiscale visualisation aspects/issues.

However, in order to understand if the proposed approach is meaningful also to the those who will navigate and interpret the biomedical data with the MSV proposed approach, a series of prototypes (benchmarks) will be implemented relying on the exemplary problems definition. For those fields of application in which we will be able to access data, demos will be provided to the interested research projects through the participation at dissemination events (like conferences and workshops) or through direct communication with the interest external stakeholders so to collect feedbacks on the chosen approach.
6 Related efforts and potential users

Potential users for the open-source library, the MSV project is going to develop, are obviously (apart the project partner which will exploit the library in their internal software development) the different software initiative which are ongoing at international level. The most important ones are here summarised:

- **CTK**: Common ToolKit international initiative\(^{20}\), whose goals are to provide a unified set of basic programming constructs that are useful for medical imaging applications development, to facilitate the exchange and combination of code and data, to document, integrate, and adapt successful solutions, and to avoid the duplication of code and data; some of the MSV partners are also participating to the CTK initiative and this will facilitate the communication between the two groups, the identification of synergies with other visualisation expert groups, and cross-fertilisation of ideas;

- **SOFA**: SOFA project\(^ {21} \) an Open Source framework primarily targeted at real-time simulation, with an emphasis on medical simulation promoted by INRIA (France). This framework is more focused on the simulation aspects and thus quite complementary to MSV; in this context it may take advantage of the MSV library to deal with the visualisation aspects;

- **XIP**: the eXtensible Imaging Platform\(^ {22}\) is another open source initiative mainly promoted in the US; being an environment for the rapid development of medical imaging applications from an extensible set of modular elements, it might integrate the MSV library as well.

From the biomedical research perspective potential users of the library will be:

- **VPH toolkit initiative**: The VPH ToolKit\(^ {23} \) is a technical and methodological framework to support and enable VPH research - the collaborative investigation of the human body as a single complex system. It is supported by the VPH Network of Excellence\(^ {24}\) and it might be a perfect mean of diffusion of the MSV library within the VPH community.

- **VPH related research efforts**: as stated before in the document different VPH-related projects have multiscale data visualisation issues in their researches. As already done, direct contacts will be established in the different VPH dissemination events in order to let them aware of the project results. The projects already contacted will be invited to evaluate the MSV prototypes for preliminary feedbacks.

- **NAMIC**: the National Alliance for Medical Image Computing\(^ {25} \) is a multi-institutional, interdisciplinary community of researchers, who share the recognition that modern healthcare demands improved technologies to ease suffering and prolong productive life. Members of NAMIC already showed interest in the MSV objectives in the early stage of the project. Further contacts will be established in the future.

\(^{20}\) [http://www.commonkt.org](http://www.commonkt.org)

\(^{21}\) [http://www.sofa-framework.org/home](http://www.sofa-framework.org/home)

\(^{22}\) [http://www.openxip.org/](http://www.openxip.org/)


\(^{24}\) [http://www.vph-noe.eu/](http://www.vph-noe.eu/)

7 Roadmap

The first aim of the MSV project is to define a shared vision for the most fundamental multiscale interaction and visualisation challenges, along with an identification of suitable categories of exemplary problems. After this, it will follow the definition of best practice, which will set guidelines for future developments, and the implementation of an open-source library as an extension of the widely used Visualisation Toolkit (VTK)\(^{26}\) from which the whole biomedical community will benefit.

The building of the shared vision started with a consensus meeting which took place in February 2010 in San Diego (USA) as part of the SPIE Medical Imaging conference. This first MSV meeting involved consortium members and invited external experts, and focused on the analysis of multiscale visualisation needs and of the current state of the art. The discussion confirmed the need for multiscale visualisation in many different fields of biomedical research (from both industrial and research points of view) and highlighted the current lack of appropriate software tools to address this issue.

Soon after the meeting, a discussion started within the consortium of which the results are summarised in this first draft of the White Paper. This document is not conclusive and its aim is to share the ideas developed within the consortium with external experts and the biomedical communities so to define a global vision of the multiscale visualisation interaction problem and its challenges.

At the beginning of 2011, also the best practices definition has started and the discussion will be ongoing with the final version of the White Paper to be delivered by end of 2011.

All 2012 will be devoted to the implementation of the library, even if already in 2011 prototypes will be implemented so to dissemination the MSV ideas and also test the approach on real-case data coming from the exemplary problems.

The public release of the MSV library as open-source extension of VTK will be release at the end 2012.

\(^{26}\) [http://www.vtk.org/](http://www.vtk.org/)
8 References


9 ANNEX 1: Visual process glossary

- **Visual Angle**: Visual angle is the angle subtended by an object on the eye of an observer. Visual angles are generally defined in degrees, minutes, and seconds of arc (a minute is 1/60 degree and a second is 1/60 minute).

  ![Visual Angle Diagram](image)

  If $O$ is the size of an object and $d$ the distance of the object from the eye, the visual angle $\alpha$ can be calculated as:

  $\tan(\alpha/2) = (O/2)/d$ and therefore, $\alpha = 2 \arctan (O/2)/d$

- **Visual Acuity**: Acuteness or clearness of vision, which is dependent on the sharpness of the retinal focus within the eye and the sensitivity of the interpretative faculty of the brain. Visual acuities are measurements of our ability to see details and are important because they define absolute limits on the information densities that can be perceived. Normally visual acuity refers to the ability to resolve two separated points or lines, but there are other measures of the ability of the visual system to discern spatial differences. There are different types of acuities and relative measurements, here are just some examples:

  - **Resolution Acuity**: This is sometimes called "minimum separable" acuity. It refers to the ability to detect a separation, or gap, between objects. A gap is made progressively smaller until the two bars cannot be distinguished from a single bar. This is different from grating acuity where the bars of the grating are progressively narrowed until the pattern cannot be distinguished from a uniform gray field. Resolution acuity thresholds are typically around 0.5 minutes of arc.

  - **Stereoscopic Acuity**: Is the ability to detect tiny differences in depth with the two eyes. For more complex targets, stereo-acuity is similar to normal monocular visual acuity, or around 0.6-1.0 arc minutes, but for much simpler targets, such as vertical rods, may be as low as only 2 arc seconds.

  - **Recognition Acuity**: Probably the most familiar kind of acuity. The task requires the viewer to name the target stimuli. Many eyecharts use either the Snellen or Sloan lettering system to measure this form of acuity. The rows of letters are made progressively smaller until the subject cannot reliably recognize them. Recognition acuity is typically measured in optotype units, e.g., 20/100, 20/40, 20/20. The optotype line which health, young emmetropic eyes can read at a distance of 20 feet sets the standard.

  - **Detection Acuity**: This is sometimes called "minimum visible" acuity. The subject's task is to detect a either a light or dark target against a background of opposite luminance polarity. As target size gets very small we can no longer resolve differences in size, so the minimum visible acuity reduces to a test of sensitivity (since small targets reflect fewer photons to our eye).

  - **Localization Acuity**: Task is to detect a misalignment of two (or more) objects. One type of localization acuity is called Vernier acuity. This type of acuity is often referred to a hyperacuity, since humans are extraordinarily sensitive to misalignment. The displacement is made progressively smaller until the subject can no longer reliably distinguish the direction. Vernier acuity thresholds are typically around 5-10 seconds of arc.
Dynamic Acuity: All of the above types of acuity are static. Each can be put in motion to measure dynamic visual acuity.

- **Display Resolution:** The resolution of an element on a particular display device (i.e. the number of distinct pixels in each dimension that can be displayed). It can be an ambiguous term especially as the displayed resolution is controlled by all different factors in cathode ray tube (CRT) and flat panel or projection displays using fixed picture-element (pixel) arrays.