# Display and Management of Geomatics Research Data

Timothy Trewartha Michiel Johan Baird
Supervisor: Hussein Suleman
Department of Computer Science
University of Cape Town

June 11, 2012

## 1 Project Description

The Zamani project, started by the UCT Department of Geomatics, aims to preserve African cultural heritage by documenting heritage sites and producing laser scanned models. Currently they have documented about 40 sites in 12 African countries with close to 100 models. Some of the models are very detailed, containing billions of points. With a fast-growing volume of data, the Department of Geomatics is facing several challenges, ranging from basic storage of the data to viewing and interacting with the large models in real-time.

There are two main components to this project. The first is to develop a system to enable both viewing and streaming of the Zamani models in their full detail, at interactive frame rates. The second is to investigate various ways of automating GIS (Geographic Information Systems) workflow. The aim is to develop a software tool to enable more efficient manipulation of GIS data. The real-time streaming infrastructure will additionally be integrated into the software developed for workflow automation, as shown in the following diagram.

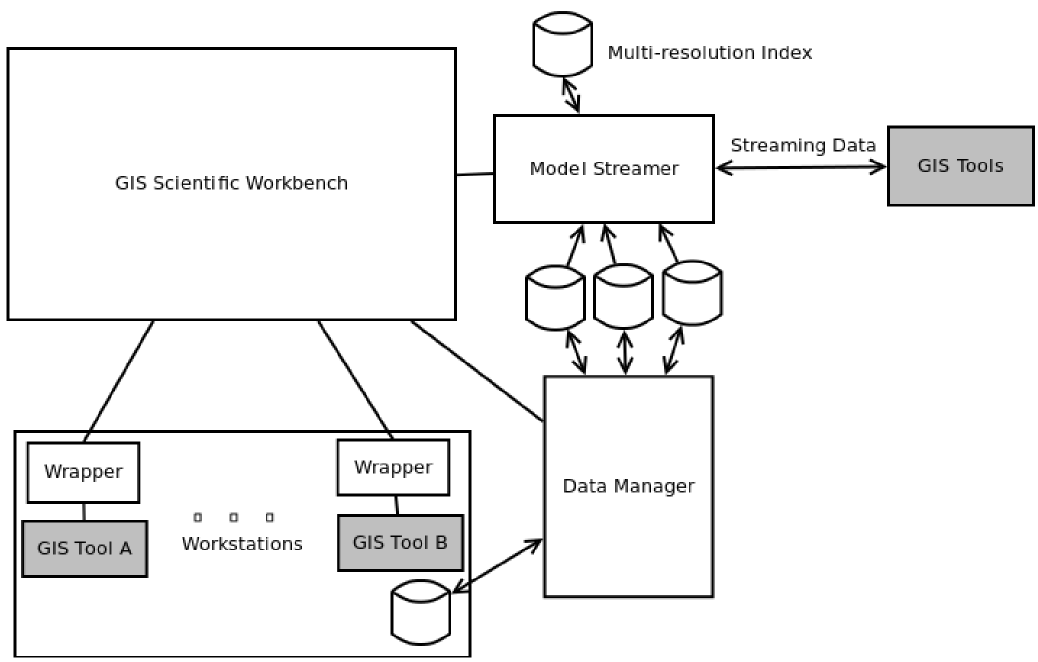


Figure 1: Interface between various project components

# 2 Problem Statement and Research Questions

This project aims to tackle two key issues faced by the Geomatics Department: the inability to interact with large models in real-time as well as the lack of tools enabling workflow to be automated. As such, the following key research questions have been proposed:

## 2.1 Is it feasible to support real time viewing of models containing billions of points?

The UCT Department of Geomatics has indicated that they have difficulties handling the sizes of some of their models. These laser scanned models of cultural heritage sites are often very large, some of them containing over 8 billion points. Given this vast scale of data, traditional viewing methods and the current hardware and software systems are not able to cope. Consequently, before viewing or manipulating the data, one must go through a process of decimating the original data by a factor of 10, 100 or more. This compromise is often unacceptable, as one may often require the full original detail. This level of detail is often necessary for cultural heritage sites in order to view details such as cracks and flaws, with a view to preserving the site and preventing damage.

This project will investigate the feasibility of real-time interaction with the Zamani models in their full detail. Answering this research question in the affirmative would enable exploration of these models interactively without decreasing the resolution beforehand.

## 2.2 How effective is an automated workflow system in the GIS context?

Developing a geographic information system involves the capture, storage, manipulation, analysis and management of geographic data. This data is very diverse and, as such, has to be handled in quite diverse ways. The data gets abstracted into various forms. This presents a rather unique challenge in managing the data as it could be used by anyone of the research staff at any point in the process. Currently, data is being manually moved or copied to where it is required. This takes up a lot of time, the movement is laborious, and it could benefit from automation.

Workflow Management Systems aim to decompose complicated projects and processes into small atomic chucks [Taylor et al., 2006]. This decomposition can then be optimised to improve the efficiency. GIS research projects generally have multi-person teams where the work is done in a parallel fashion. Under these conditions workflow management systems are particularly effective.

The aim is to provide a workflow management that is applicable for GIS projects. This system should be able to: interface with the current systems, track and manage the workflow, provide local data availability and content delivery, and increase overall efficiency within the discipline.

# 3 Procedures and Methods

Given the above problems, the following procedures and methods are being proposed:

## 3.1 Implement a Hierarchical Data Structure

From researching the literature it seems that the most common way of dealing with large point based models containing billions of points is to build a multiresolution data structure to divide the model into manageable chunks. Initially, only a small subset of the number of available points is required. As one zooms into the model, additional points are fetched from the data structure until the full original detail is available. Using such a level-of-detail structure should enable the Department of Geomatics to view even very large models at interactive frame rates, without having to decimate the original data.

The data structure being proposed is an octree [Wand et al., 2007]. All data is stored in the leaf nodes and inner nodes provide simplified multi-resolution representations. Additionally the data structure imposes the constraint that no leaf node should contain more than a specified number of points. This number of points is a parameter in the system which will need to be determined experimentally, but for this structure a value of around 30,000 was found to give good performance [Wand et al., 2007]. The following figure shows the first three levels in such an octree structure, for a given point based model of a sphere.

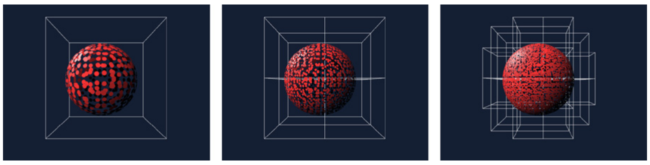

Figure 2: Each level of the octree structure provides greater detail

## 3.2 Use an existing workflow management system as a base

Various platforms already exist that are designed to manage workflow. These systems, however, need to be adapted for GIS research. As these systems already have a large number of features, such as, flow building toolkits, flow optimisation, integration capabilities and monitoring, writing such a system from the ground up would be a pointless task.

The decision on which system to use is partly dependent on the requirements of the Geomatics department. This will be assessed and discussed. After this decision is made the core functionalities would be established.

These components would then be implemented in a modular fashion, on top of the existing system. The goal of these components is to address the core functionalities that were determined.

The aim is to develop a system that integrates with the current GIS tools. A high level of collaboration with the Geomatics department will be required. The feedback will be used to prioritise components and functionalities.

## 3.3 Testing and Evaluation

A key evaluation criteria will be to demonstrate that the new system is able to render the large point based models in real-time. Since this functionality was not available previously, it will be a significant success. Additionally, it will be important to test whether real-time streaming from the server to client machines is feasible.

Further testing and refinement is required for the workbench to determine the following: the speed of the content delivery system, the effectiveness of imposed workflow, the effectiveness of the user interface and overall system stability.

It will be important to evaluate the success of the system in various usage scenarios. These usage scenarios will provide a useful mesure for evaluating the success of the resulting system. For example, consider the following usage scenario for the workbench:

1. GIS researcher defines the process for a particular research outcome

2. The system automatically transfers the appropriate data to the computers where it needs to be used
3. As soon as a step is done it is logged and the appropriate data is moved down the pipeline

# 4 Ethical, Professional and Legal Issues

## 4.1 User Testing

One of the components of the project involves the design and evaluation of a user interface. Ideally, the design process for this would require input from the user as well as testing. This testing requires that ethical clearance be obtained.

## 4.2 Data Privacy

The Department of Geomatics has indicated that some of the data collected by the Zamani Project is sensitive and is not to be made freely available. It is important to ensure that, during the course of this project, this wish is respected and that nothing is done to compromise the privacy of sensitive data. As such, the data, once received, will only be stored on the server used as part of the project. Special permission will be required if the data is requested for testing at an external location.

# 5 Related Work

## 5.1 Hierarchical Data Structure

Common methods for structuring 3D data include octrees [Wand et al., 2007], R-trees [Zhu et al., 2007], bounding sphere hierarchies [Rusinkiewicz and Levoy, 2000], and Hilbert Space Filling Curves [Wang and Shan, 2005], each with their own advantages and disadvantages. Based on the experimental results of each method, a dynamic octree structure seems to have the best performance [Wand et al., 2007]. Wand et al. demonstrated its ability to handle large models efficiently by using this data structure to achieve interactive walkthroughs of a data set containing over 2.2 billion points.

## 5.2 Meshes and Point Clouds

Applying rendering techniques to point clouds can present many challenges and acquiring consistent renderings is non-trivial. This is because there is no explicit connectivity between points in a point cloud, in contrast to mesh-based models. Traditional approaches involve the generation of a triangular mesh from the point cloud generated by the laser scanner[Cazals and Giesen, 2006]. However, the level-of-detail datastructure is applicable to point clouds and not meshes. Consequently the renderer chosen for this project should employ local surface reconstruction methods. This process is linear and should scale well, even in the case of very large models[Dobrev et al., 2010].

## 5.3 Automated Workflow Management

Various fields of science have benefited from automated workflow. It has seen good increases in productivity [Brahe and Schmidt, 2007] and the ability to share workflow between colleges has aided quite significantly in th the reproducibility of the science[De Roure and Goble, 2009]. GIS has been evaluated to be highly applicable to workflow systems[Migliorini et al., 2011]. However, some limitations were noted. Various components, such as the modeling and processing of spatial data, would need to be added to make it feasible.

# 6 Anticipated Outcomes

There are two key anticipated outcomes from this project. First is the implementation of a multi-resolution data structure to enable real-time interaction with large point based models. Second is the development of a software system to enable automated GIS workflow. It is expected that if both of these outcomes are achieved, the results could have a significant impact in the Department of Geomatics. It will enable the viewing of models in full detail without decimation and could greatly increase the efficiency in Geomatics research. Key evaluation criteria are:

- Can the system render the largest of the Zamani models at interactive frame rates?
- Does the system enable streaming of large models from a central server?
- Has dataflow been automated to provide transparent local access?
- Has Geomatics research been successfully mapped to a workbench environment?

# 7 Project Plan

## 7.1 Risks

### Request for Hardware Denied

*Severity:* High
*Likelihood:* Low
It is possible that the request for a server will be denied. If this happens, a considerable amount of restructuring will be required and it would have a significant impact on the course of the project.

### Network Constraints

*Severity:* Medium
*Likelihood:* Low
One of the core functionalities of the system would be to provide content delivery of data that is required for a specific task. Providing this local data allows the task to get completed without unnecessary fetching delays. There is a risk that this content delivery system could saturate the network. This would cause the system to be slow and unusable.

### Middleware

*Severity:* High
*Likelihood:* Low
For this project to be successful, the workflow management system will have to interface heavily with existing software used to perform GIS operations. This will require large amounts of middleware to be developed that understand the input and output formats of this software. Since many of these formats are proprietary, a significant amount of effort will have to be made for the system to function. If these formats can not be integrated, it presents a huge risk to the project.

#### Hardware Limitations

*Severity:* Medium
*Likelihood:* Medium
There is a risk that the hardware available will not be able to cope with the load that will be required. Since a distributed system is not being proposed, there is a risk that the system will become a bottleneck. In such an eventuality either the scale of the project would have to be decreased, or a solution would have to be proposed. Such a solution would most likely involve distributing the data on multiple servers.

#### Large Indices

*Severity:* Medium
*Likelihood:* Low
When indexing the models, a significant amount of data will be generated. Given that many of the models are already very large, these indices might become infeasibly large. Dealing with such large indices will be an important part of the project and this risk will have to be handled carefully.

#### Integration with Existing GIS Software

*Severity:* Low
*Likelihood:* Medium
The hierarchical data structure required as part of this project aims to facilitate level of detail streaming and real-time interaction. However, ideally, one should not have to re-implement tools which are already available, such as ArcGIS. The aim is to integrate the data structure into a pre-existing software package to prevent unnecessary work. However, this may be difficult and there are several associated risks including unavailability of source code and lack of documentation for the software.

#### Indexing takes too long

*Severity:* High
*Likelihood:* Low
In the literature it had been noted that indexing the 3D data requires a significant amount of time [Wand et al., 2007]. While it is hoped that this will not be a problem, steps will have to be taken to ensure that the duration of the indexing process does not pose a risk to the completion of the project. It will also be important to allow time for the possible event of a system failure, in which case the indexing process would have to be restarted.

## 7.2 Timeline, including Gantt chart

Figure 3 shows the anticipated timeline in the form of a Gantt chart.

## 7.3 Resources required

#### Hardware

A server will be required for this project to enable central storage of the 3D models. A large amount of storage will be required, as the Zamani data is over 20TB. The aim is to initially store a subset of the data, and expand as the project progresses. Multiple hard drives will

| Name | Work |
|---|---|
| **Literature Review** | **20d** |
| First Draft | 15d |
| Second Draft | 3d |
| Final Handin | 2d |
| **Project Proposal** | **11d** |
| First Draft | 7d |
| Second Draft | 2d |
| Final Handin | 2d |
| **Project Proposal Presentation** | **6d** |
| Proposal Dry run | 4d |
| Final Presentation | 2d |
| Proposal Revision | 7d |
| Project Web Presence | 18d |
| **Initial Feasibility Demonstratio** | **48d** |
| Design | 10d |
| Implementation | 27d |
| Evaluation | 11d |
| **Background Chapter** | **28d** |
| First Draft | 21d |
| Second Draft | 4d |
| Final Handin | 3d |
| **Design Chapter** | **31d** |
| First Draft | 17d |
| Second Draft | 12d |
| Final Handin | 2d |
| **First Implementation + Writeu** | **87d** |
| Design | 13d |
| Implementation | 31d |
| Evaluation | 14d |
| Writeup | 29d |
| **Final Implementation + Writeu** | **60d** |
| Design | 9d |
| Implementation | 14d |
| Evaluation | 7d |
| Writeup | 30d |
| Report Outline Complete | 12d |
| **Final complete draft of Report** | **14d** |
| First Draft | 4d |
| Second Draft | 4d |
| Third Draft | 4d |
| Final Handin | 2d |
| Project Report Final Handin | 7d |
| Poster Due | 4d |
| Webpage | 5d |
| Reflection Paper | 3d |
| Final project presentation | 7d |

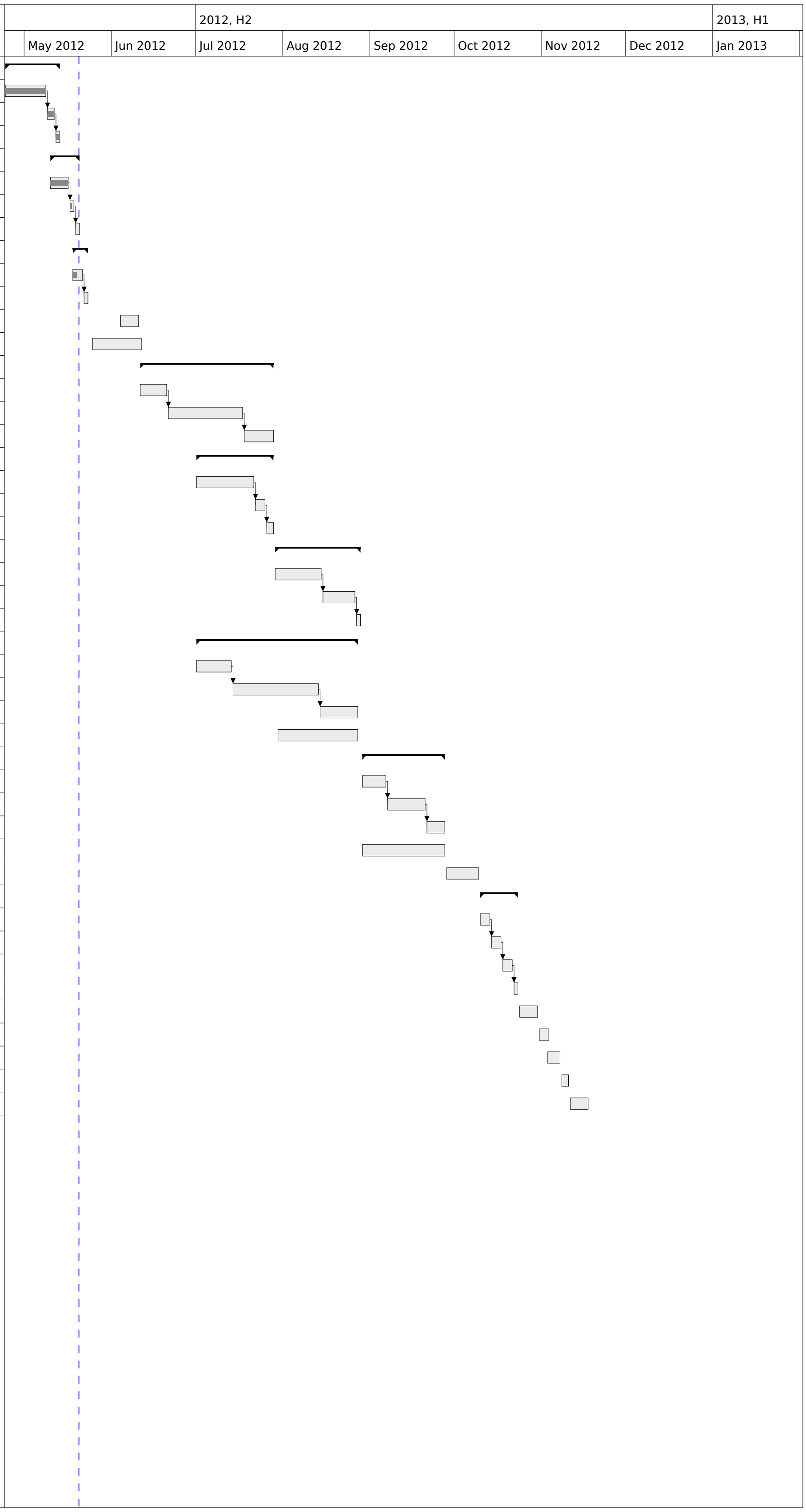


Figure 3: Project Schedule

be required and this will allow for a certain amount of parallelism in data access. The exact specifications of the server are still being finalised.

### Geographic Data

The Department of Geomatics has indicated that they are willing to make their data available for the purposes of this project. It will be important to obtain the models at an early stage as any delays in obtaining the models will delay the entire project.

## 7.4 Milestones

### 7.4.1 Project Milestones

| Description | Duration | Start | Finish |
|---|---|---|---|
| Literature Review | 18 days | 24 April 2012 | 14 May 2012 |
| Fist Draft Due | 15 days | 24 April 2012 | 9 May 2012 |
| Second Draft | 3 days | 9 May 2012 | 12 May 2012 |
| Project Proposal | 12 days | 10 May 2012 | 21 May 2012 |
| First Draft Due | 7 days | 10 May 2012 | 17 May 2012 |
| Second Draft Due | 2 days | 17 May 2012 | 19 May 2012 |
| Project Proposal Presentation | 6 days | 18 May 2012 | 24 May 2012 |
| Proposal Dry run | 4 days | 18 May 2012 | 22 May 2012 |
| Proposal Revision Completed | 7 days | 4 June 2012 | 11 June 2012 |
| Project Web Presence | 18 days | 25 May 2012 | 12 June 2012 |
| Initial Feasibility Demonstration | 48 days | 11 June 2012 | 29 July 2012 |
| Design | 10 days | 11 June 2012 | 21 June 2012 |
| Implementation | 27 days | 21 June 2012 | 18 July 2012 |
| Evaluation | 11 days | 18 July 2012 | 29 July 2012 |
| Background Chapter | 28 days | 1 July 2012 | 29 July 2012 |
| First Draft Due | 21 days | 1 July 2012 | 22 July 2012 |
| Second Draft Due | 4 days | 22 July 2012 | 26 July 2012 |
| Design Chapter | 31 Days | 29 July 2012 | 29 August 2012 |
| First Draft Due | 17 days | 29 July 2012 | 15 August 2012 |
| Second Draft Due | 14 days | 15 August 2012 | 26 August 2012 |
| First implementation and Writeup | 31 days | 1 July 2012 | 29 August 2012 |
| Design | 13 days | 1 July 2012 | 14 July 2012 |
| Implementation | 31 days | 14 July 2012 | 15 August 2012 |
| Evaluation | 14 days | 15 August 2012 | 26 August 2012 |
| Writeup | 29 days | 30 July 2012 | 29 August 2012 |
| Final implementation and Writeup | 30 days | 29 August 2012 | 28 September 2012 |
| Design | 9 days | 29 August 2012 | 7 September 2012 |
| Implementation | 14 days | 7 Semptember 2012 | 21 September 2012 |
| Evaluation | 7 days | 21 September 2012 | 28 September 2012 |
| Writeup | 30 days | 29 August 2012 | 28 September 2012 |
| Report Outline Complete | 12 days | 28 September 2012 | 10 October 2012 |
| Final Complete Draft of Report | 14 days | 10 October 2012 | 24 October 2012 |
| First Draft Due | 4 days | 10 October 2012 | 14 October 2012 |
| Second Draft Due | 4 days | 14 October 2012 | 18 October 2012 |
| Third Draft Due | 4 days | 18 October 2012 | 22 October 2012 |
| Project Report Final Handin | 7 days | 24 October 2012 | 31 October 2012 |
| Poster Due | 4 days | 31 October 2012 | 3 November 2012 |
| Website | 4 days | 31 October 2012 | 3 November 2012 |
| Project Demonstrations | 5 days | 3 November 2012 | 8 November 2012 |
| Write Reflection paper | 3 days | 8 November 2012 | 11 November 2012 |
| Final project presentation | 7 days | 11 November 2012 | 18 November 2012 |

## 7.5 Deliverables

### GIS Workbench

A key component of the project is to produce a GIS workbench. This will be the framework that ties all the components together. This will involve using an existing Workflow System and setting it up to represent the flow of a GIS project.

### Middleware for Core Functionalities

Once the GIS workflow is properly understood and modeled, it is important to create middleware that interfaces with the systems that are currently being used.

### Data Flow Facilitator

To be able produce the content delivery that is required, a dataflow facilitator will need to be developed that is integrated with the workflow management system.

### Hierarchical Data Structure

In order to facilitate level of detail streaming, it will be essential to implement a hierarchical data structure that can support interactive viewing of models containing billions of points.

### Streaming Infrastructure

A real-time streaming infrastructure from the server to client machines will be an important deliverable. This will also need to be implemented as early as possible.

## 7.6 Work Allocation

Timothy Trewartha will be implementing the hierarchical data structure to support real-time interaction with the Zamani models. Michiel Johan Baird will be developing the GIS workbench.

# References

[Brahe and Schmidt, 2007] Brahe, S. and Schmidt, K. (2007). The story of a working workflow management system. In *Proceedings of the 2007 international ACM conference on Supporting group work*, GROUP '07, pages 249–258, New York, NY, USA. ACM.

[Cazals and Giesen, 2006] Cazals, F. and Giesen, J. (2006). Delaunay triangulation based surface reconstruction: Ideas and algorithms. In *EFFECTIVE COMPUTATIONAL GEOMETRY FOR CURVES AND SURFACES*, pages 231–273. Springer.

[De Roure and Goble, 2009] De Roure, D. and Goble, C. (2009). Software design for empowering scientists. *Software, IEEE*, 26(1):88 –95.

[Dobrev et al., 2010] Dobrev, P., Rosenthal, P., and Linsen, L. (2010). Interactive image-space point cloud rendering with transparency and shadows. In Skala, V., editor, *Communication Papers Proceedings of WSCG, The 18th International Conference on Computer Graphics, Visualization and Computer Vision*, pages 101–108, Plzen, Czech Republic. UNION Agency – Science Press. dobrevrosenthallinsenvcglsmoothvispointsinsight.

[Migliorini et al., 2011] Migliorini, S., Gambini, M., Belussi, A., Negri, M., and Pelagatti, G. (2011). Workflow technology for geo-processing: the missing link. In *Proceedings of the 2nd International Conference on Computing for Geospatial Research & Applications*, page 36. ACM.

[Rusinkiewicz and Levoy, 2000] Rusinkiewicz, S. and Levoy, M. (2000). QSplat: A multiresolution point rendering system for large meshes. In *Proceedings of ACM SIGGRAPH 2000*, pages 343–352.

[Taylor et al., 2006] Taylor, I. J., Deelman, E., Gannon, D. B., and Shields, M. (2006). *Workflows for e-Science: Scientific Workflows for Grids*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

[Wand et al., 2007] Wand, M., Berner, A., Bokeloh, M., Fleck, A., Hoffmann, M., Jenke, P., Maier, B., Staneker, D., and Schilling, A. (2007). Interactive editing of large point clouds. In Chen, B., Zwicker, M., Botsch, M., and Pajarola, R., editors, *Symposium on Point-Based Graphics 2007 : Eurographics / IEEE VGTC Symposium Proceedings*, pages 37–46, Prague, Czech Republik. Eurographics Association.

[Wang and Shan, 2005] Wang, J. and Shan, J. (2005). Space-filling curve based point clouds index. *Geocomputation*.

[Zhu et al., 2007] Zhu, Q., Gong, J., and Zhang, Y. (2007). An efficient 3d r-tree spatial index method for virtual geographic environments. *ISPRS Journal of Photogrammetry and Remote Sensing*, 62(3):217 – 224.