The Significant Properties of Vector Images

David A. Duce[1]
daduce@brookes.ac.uk

[1]School of Technology
Oxford Brookes University
Oxford OX33 1HX

JISC/BL/DPC workshop 7 April 2008

Outline

1. The Study
   - Activities
   - Context
   - Methodology
2. Approach
   - Approach to SPs and Metrics
   - Tool Support
   - Examples
3. Summary
   - Conclusions
   - Recommendations

Acknowledgement

- Observations based study JISC-funded study *Signficant Properties of Vector Images*; JISC Digital Preservation and Records Management Programme (2007)
- Mike Coyne, George Mallen and Mike Stapleton (System Simulation)
- David Duce and Bob Hopgood (Oxford Brookes University)
- Reviewed by Arnold (University of Brighton)
- Thanks to Kevin Ashley, Jon Blower, Paul Brown, David Cruickshank, Douglas Dodds, David Duke, Gareth Knight, Nick Lambert, Brian Matthews, Kieran Niven, Alan Shipman, Alan Smith and Will Wilcox

Outline

1. The Study
   - Activities
   - Context
   - Methodology
2. Approach
   - Approach to SPs and Metrics
   - Tool Support
   - Examples
3. Summary
   - Conclusions
   - Recommendations

Activities

- Defining Scope of objects to be considered
- Surveying potential range of application areas
- Surveying range of digital formats in use and potential archival formats
- Enumerating hierarchy of Significant Properties of relevance to the community
- Establishing easily understood metric
- Exploring whether capture of properties could be (semi-) automated
- Trying approach on real examples

Outline

1. The Study
   - Activities
   - Context
   - Methodology
2. Approach
   - Approach to SPs and Metrics
   - Tool Support
   - Examples
3. Summary
   - Conclusions
   - Recommendations

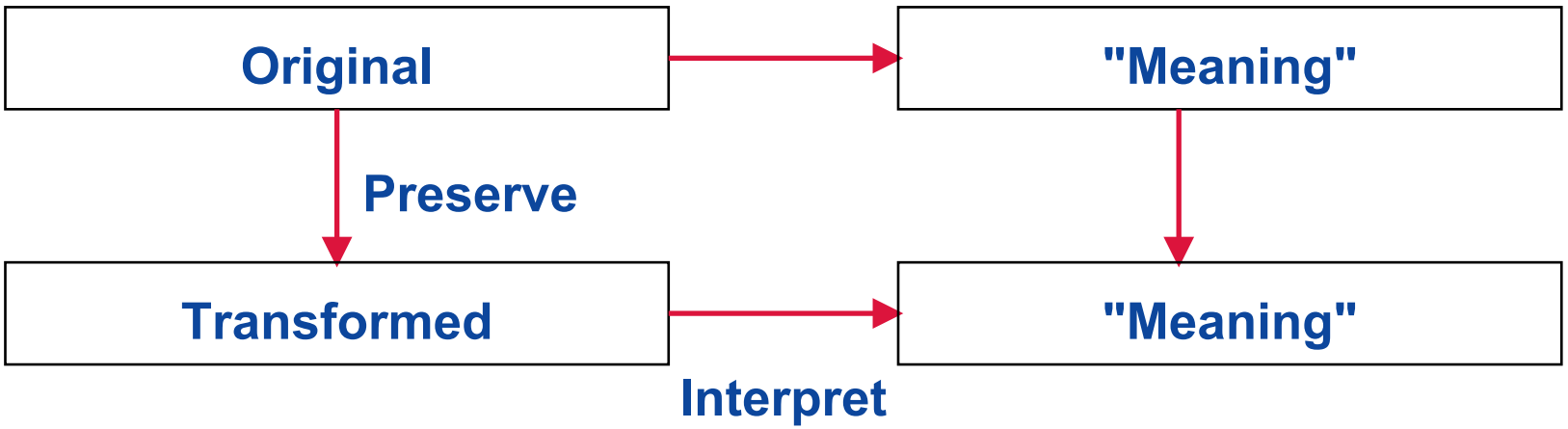
InSPECT Project

- Carried out in context of a framework provided by the ‘Investigating the Significant Content of Electronic Content over Time (InSPECT)’ project
- InSPECT is committed to:
  - The *conceptual model* developed by the National Archives of Australia (NAA), known as the *performance model* and the associated concept of *essence*, which InSPECT believes is equivalent to the term *significant properties*.
  - The *data-centric* approach that focuses on maintaining objects in current data formats rather than the *process-centric* approach that keeps objects in their original form and attempts to emulate the original environment

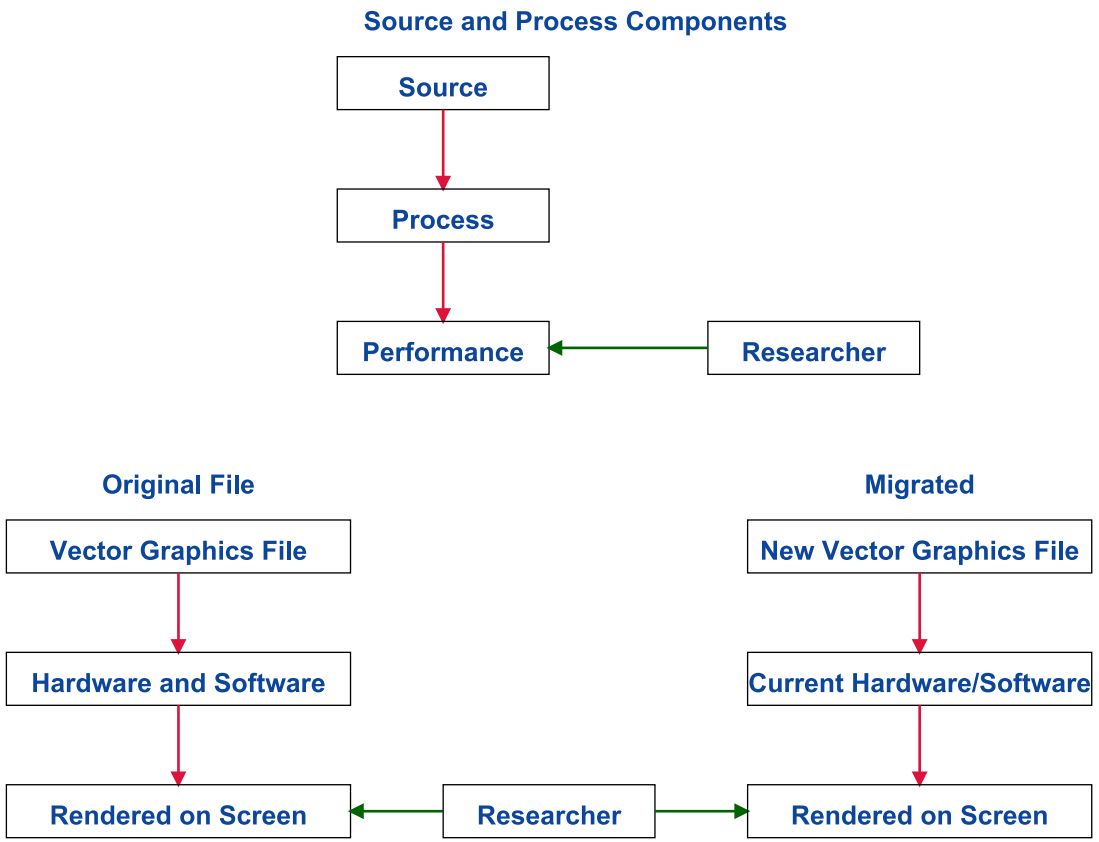
NAA Conceptual Model

- *Source* of an object combined with *process* creates a *performance*
- Archival strategy is to transform the original object with related information to produce a transformed source that retains the *essence* of the original
- i.e. retains its *significant properties*
- Familiar questions from visualization and accessibility
  - Do you see what I mean?
  - Do I mean what you see?
- Challenge to identify *significant properties* of the original object and retain under transformation

Transformation Model



Conceptual Model



Vector Graphics

```
<svg width="511" height="320">
  <g transform="translate(10,−200)">
    <g transform="scale(1.0)">
      <path class="duck" id="duck"
        d="M  0  312
           C  40 360 120 280 ... Z"/>
      <text class="subcode">
        <textPath xlink:href="#duck">
         We go up, then we go down, ...
        </textPath>
      </text>
    </g>
  </g>
</svg>
```

Outline

1. The Study
   - Activities
   - Context
   - Methodology
2. Approach
   - Approach to SPs and Metrics
   - Tool Support
   - Examples
3. Summary
   - Conclusions
   - Recommendations

Methodology

- Refinement from initial idea
- Project workshops focused on three candidate formats
  - Computer Graphics Metafile (ISO/IEC)
  - Scalable Vector Graphics (W3C)
  - PDF/A (ISO)
- Consultations across representative application areas
- Three ways to generate vector images
  - from data
  - directly
  - (extract from raster image)

Outline

1 The Study
- Activities
- Context
- Methodology

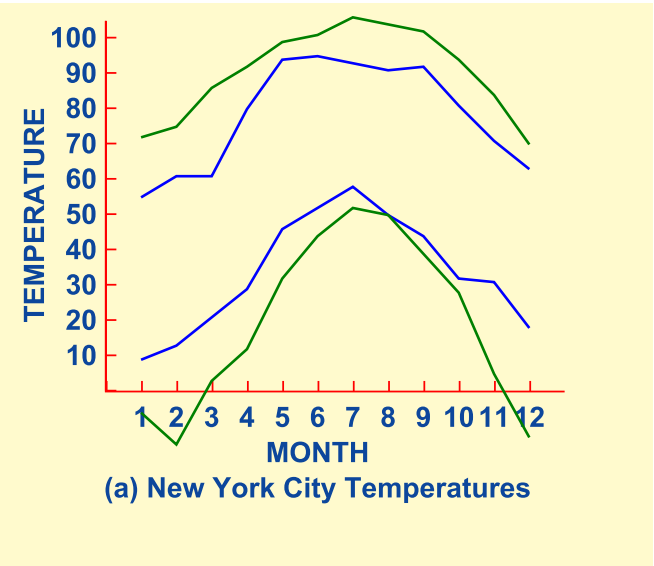
2 Approach
- Approach to SPs and Metrics
- Tool Support
- Examples

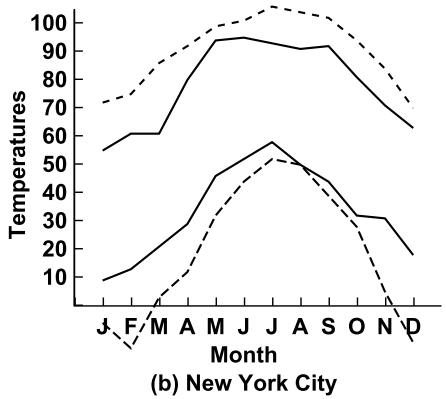
3 Summary
- Conclusions
- Recommendations

Approach

- Can think of process of creating a graphical representation as a binding process
- Computer Graphics Reference Model (ISO/IEC 11072:1992) supports this view
- Link significance to trade-off, e.g. precise colour vs. colour for differentiation

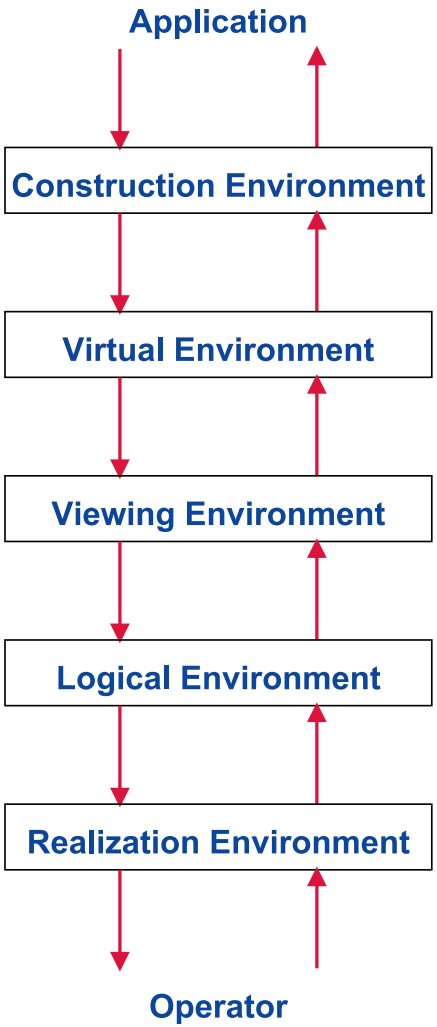
Example





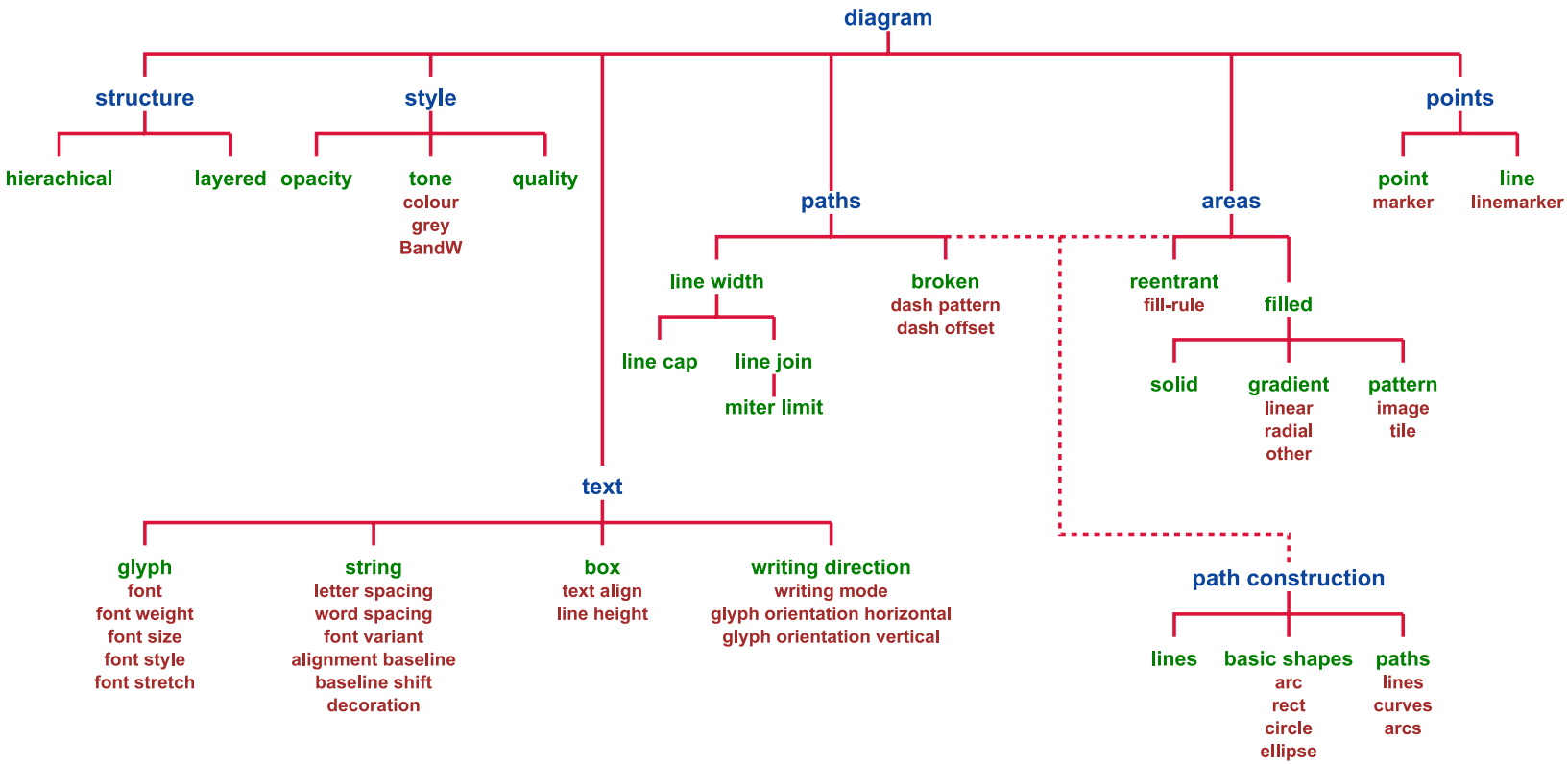
- Visualizations of New York City dataset
- Same dataset for both

Computer Graphics Reference Model



- Virtual *Scene* binds intrinsically important properties
- Logical *Graphical image* binds properties regarded as styling

Property Hierarchy



Levels of Significance

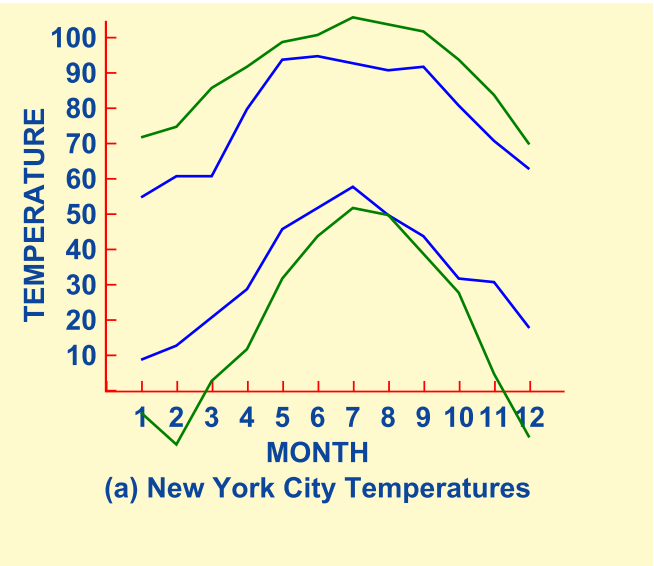
- Property is significant in the *scene* of the *virtual* environment.
- Property is bound to the *graphical image* of the *logical environment*.
- Property is used, but in a minor way and, if missing, little information would be lost.
- Property is not used at all or has no significance.

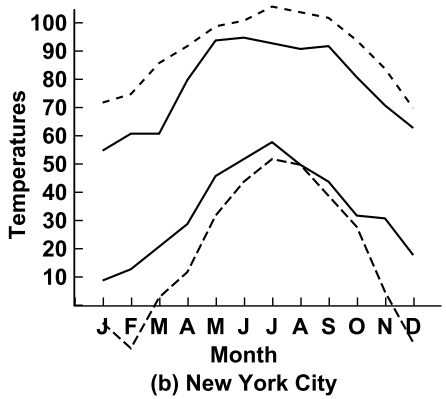
Significance Values

Give a significance value to a property, e.g.

- 0: property has no significance, it is not used.
- 1 to 3: property is used but does not have any major effect on the diagram.
- 4 to 6: property is used and different values of this property must be differentiable in the diagram. However, substitution by another property would not be significant.
- 7 to 9: property is used and is significant. Not rendering it or substituting another property for it will cause a serious loss of information. The aim would be to use the values 0, 2, 5, and 8 as the main differentiators and then use the values above and below to shade the significance

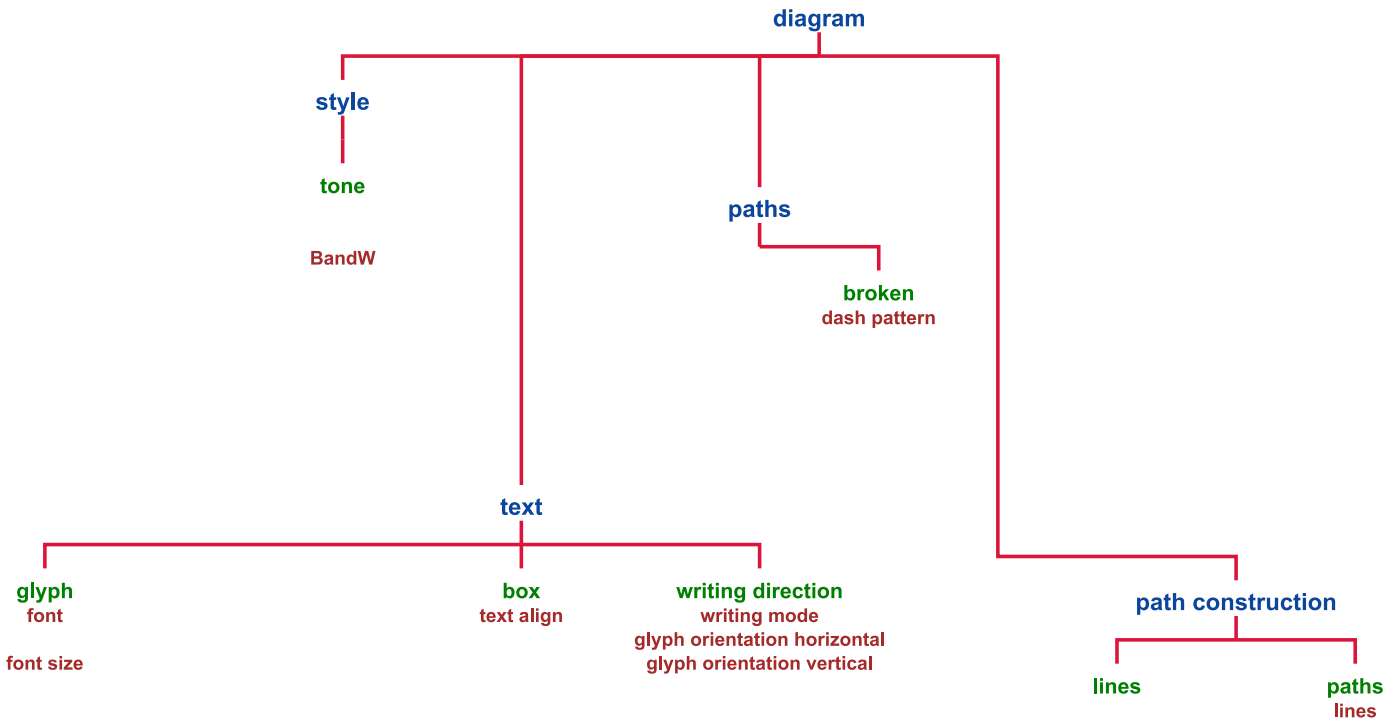
Example





- Visualizations of New York City dataset
- Same dataset for both

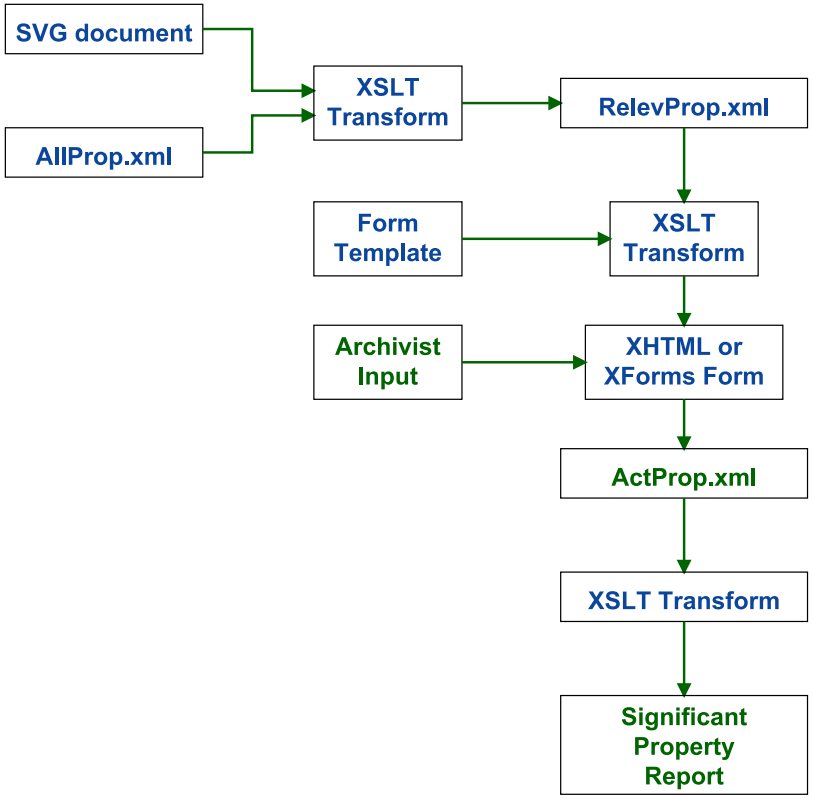
Significant Properties



Outline

1. The Study
   - Activities
   - Context
   - Methodology
2. Approach
   - Approach to SPs and Metrics
   - Tool Support
   - Examples
3. Summary
   - Conclusions
   - Recommendations

Eliciting Significant Properties



Outline

1. The Study
   - Activities
   - Context
   - Methodology
2. Approach
   - Approach to SPs and Metrics
   - Tool Support
   - Examples
3. Summary
   - Conclusions
   - Recommendations

Archaeology Data Service - Channel Tunnel Excavation (SVG)

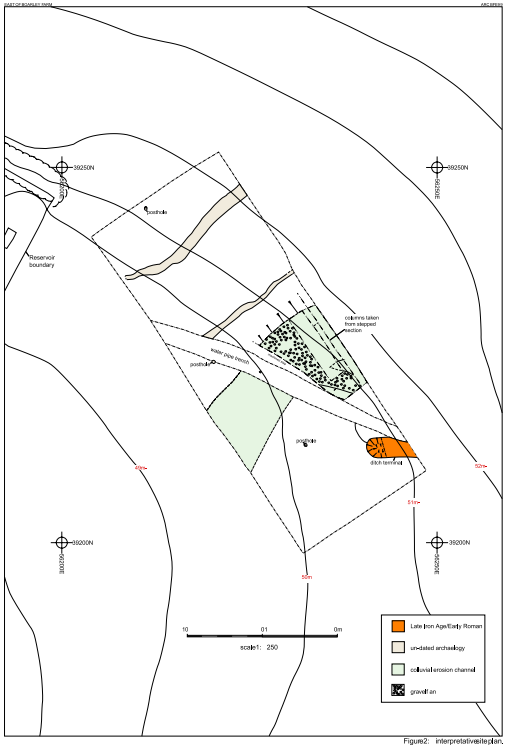
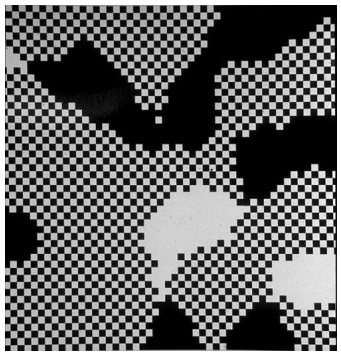
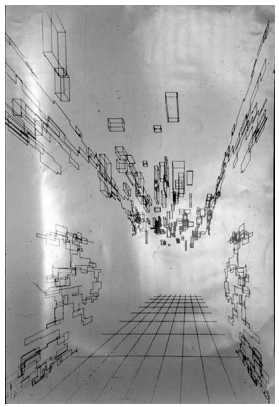


Figure2: interpretativesiteplan.

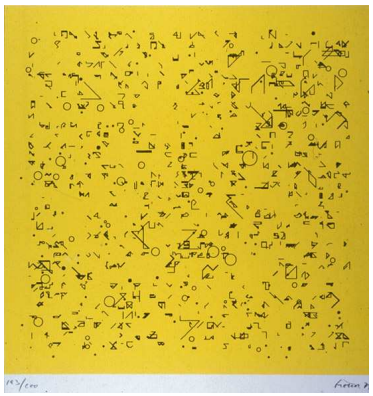
Computer Art Images



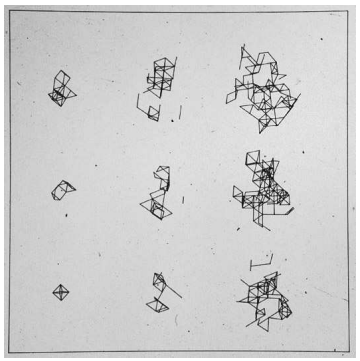
Lambert Meertens and Leo Geurtz (circa 1973)



Georg Nees



Manfred Mohr 1971



Paul Brown 1979

Significant Properties

- Significant properties of early computer art works vary
- In only one case was text significant at a metric value of 9
- Several were significant at a metric value around 5 and a few had values around 2
- Most of the line drawing needed a reasonable fine line but otherwise not significant. Only one line drawing made significant use of thick lines
- Many consisted of a regular pattern of characters used because of their overall grey scale intensity
- Only a few of the area filled works required specification of fill-rule (defining inside/outside)
- Colour was primarily used for differentiation. One or two had sufficiently precise differences between colours that they could be called significant.

From Paul Brown ...

I've reconstructed many of my early plotter works using contemporary technology. To me these are essentially 'identical' or maybe even better than the originals but the art world is unlikely to agree. This is the distinction between an artist working in the "conceptual domain" and an art world addicted to the unique artefact. Somewhere in here is the concept of the 'original'.

Thus ...

- Document *why* object is being preserved and precisely why it is being preserved in a particular way
- Broader perspective
  - different players/artist, critics, conversationists etc may have different perspectives on the ‘original’
  - many ways to preserve, modulo the perspective
  - importance of metadata

Outline

1 The Study
- Activities
- Context
- Methodology

2 Approach
- Approach to SPs and Metrics
- Tool Support
- Examples

3 Summary
- Conclusions
- Recommendations

Conclusions

- Broadly endorse InSPECT approach
- Preservation of vector graphics not widely practised
- Process of creating and modifying vector image distinguishes from other types of object
- Preserve at application data level, if generated from application data and no intrinsic value in visualization
- Candidate formats
  - WebCGM mainly of interest in engineering (significant use)
  - SVG is an XML application; includes font definition and animation capability
  - PDF/A specific archival format
- Archive binary; all have well-defined compression schemes

Conclusions

- All can capture associaterd metadata
- Little ability to capture constraints, e.g. box A is joined to box B
- CGRM framework invaluable
- Scope for semi-automated process to determine SPs

Outline

1 The Study
- Activities
- Context
- Methodology

2 Approach
- Approach to SPs and Metrics
- Tool Support
- Examples

3 Summary
- Conclusions
- Recommendations

Recommendations

- Endorse recommendations of earlier Digital Image Archiving Study to maintain registrty of file formats
- Recommend use of WebCGM, SVG and PDF/A for 2d vector graphics
- Review of conversion tools
- Significant properties report could be used to drive conversion process
- Further investigation of tools for extracting Significant Properties
- Importance of test suites (and their preservation)

Recommendations

- Investigate W3C RDF/A and related work for adding metadata to XML applications, including ontology for vector graphics
- Extend study to 3d graphics and time-dependent vector images
- (Arnold) Note work in cultural heritage sector on recording series of events object has been engaged in; link to problem of archiving image and archiving intent of image

Thank You

References I

Mike Coyne, David Duce, Bob Hopgood, George Mallen, Mike Stapleton,

Significant Properties of Vector Images.
System Simulation and Oxford Brookes University, 2007.
Available at
http://www.jisc.ac.uk/media/documents/programmes/preservation/vector_images.pdf