SVG and the Preservation of Vector Images

David Duce\textsuperscript{1}  Bob Hopgood\textsuperscript{1}  Mike Coyne\textsuperscript{2}  Mike Stapleton\textsuperscript{2}  George Mallen\textsuperscript{2}

\textsuperscript{1}\textit{School of Technology, Oxford Brookes University, UK}
\textsuperscript{2}\textit{System Simulation,}

SVG Open 26-28 August 2008
Acknowledgements

- Based on JISC-funded study *Significant Properties of Vector Images*; JISC Digital Preservation and Records Management Programme (2007), led by the authors
- Thanks to participants in the study 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
Background

- Digital objects are just bits, requiring conversion to become accessible
- Hardware, operating systems, application programs, all are subject to continuous change
- Poses challenges for continuing access
Approaches to the problem

- **Bitstream preservation**
- **Technology preservation** - hardware and software
- **Migration** - to new media and/or file formats
- **Emulation** - replicate original look and feel on modern hardware
- **Universal Virtual Computer**
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
- Links to test suites, e.g. for SVG
1 The Study
   • Background
   • Context
   • Methodology

2 Approach
   • The Computer Graphics Reference Model
   • Tool Support

3 Case Studies

4 SVG as an Archival Format

5 Conclusions and Recommendations
Methodology

- Initial idea: use ISO/IEC Computer Graphics Reference Model to provide a framework
- Focused on three candidate formats
  - Computer Graphics Metafile (ISO/IEC 7942) and WebCGM profile
  - Scalable Vector Graphics (W3C)
  - PDF/A (ISO 19005:2005)
- Three ways to generate vector images
  - from data
  - directly
  - (extract from raster image)
- Consultations across representative application areas, representing first two approaches
Outline

1. The Study
   - Background
   - Context
   - Methodology

2. Approach
   - The Computer Graphics Reference Model
   - Tool Support

3. Case Studies

4. SVG as an Archival Format

5. Conclusions and Recommendations
Can think of process of creating graphical output as a binding process


Link significance to application conventions, e.g. hatch patterns denoting specific materials, linestyle denoting specific diameter of co-axial cable

Link significance to trade-off, e.g. precise colour vs. colour for differentiation, colour for differentiation vs. linestyle for differentiation
Example

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

Duce, Hopgood, Coyne, Stapleton, Mallen
Virtual *Scene* binds intrinsically important properties

Logical *Graphical image* binds properties regarded as styling
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

Use the values 0, 2, 5, and 8 as the main differentiators and the values above and below to shade the significance.
Example

- Properties not in the figure have no significance
- Use this to derive restricted property hierarchy
- Potentially significant properties
- Values provided by human judgement
Significant Properties

diagram

style

tone

BandW

text

glyph

font

font size

box

text align

writing direction

writing mode

glyph orientation horizontal

glyph orientation vertical

paths

broken

dash pattern

path construction

lines

paths

lines
Eliciting Significant Properties

SVG document -> XSLT Transform -> RelevProp.xml

AllProp.xml -> Form Template -> XSLT Transform

Archivist Input -> XHTML or XForms Form -> ActProp.xml

ActProp.xml -> XSLT Transform

Significant Property Report
From Archaeology Data Service

Interpretative site plan
- Line properties are significant
- Fill colour used for differentiation (archaeological periods, erosion channel)
- Text is significant, but font isn’t
Computer Art Images

Lambert Meertens and Leo Geurtz (circa 1973)

Manfred Mohr 1971

Georg Nees

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
- Fill-rule only specified in a few of the area filled works
- Colour primarily used for differentiation. One or two had sufficiently precise differences between colours for colour to be significant

Duce, Hopgood, Coyne, Stapleton, Mallen
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
Pros

- Only PDF/A specifically defined as an archival format
- SVG 1.1 and WebCGM rely on external information not automatically contained in archival file
- Fonts can be defined in SVG 1.1, but not in WebCGM
- Potential of SVG declarative animation
- SVG XML representation enables transformation
- Availability of conversion tools and import/export capability in applications
Cons

- Lack of explicit support for layering (\textit{g} element is overloaded)
- Reasonable set of primitives; addition of e.g. NURBS would widen applicability
- Attribute model is deficient w.r.t. CGRM
Conclusions

- Broadly endorse applicability of InSPECT approach to vector images
- 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 image is 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
Conclusions

- All can capture associated 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
Recommendations

- 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
- Address limitations of SVG as a presentation format
- Define SVG archival profile
Thank You!