

Connecting and communicating our digital heritage with the future



An Expedition to European Digital Cultural Heritage

Collecting, Connecting — and Conserving?

21-22 June 2006, Residenz zu Salzburg

Prof Seamus Ross

Director, HATII (University of Glasgow)

<http://www.hatii.arts.gla.ac.uk>

Construct

so we can SPEAK to the Future

The Growing Wealth of Europe

- Centrality of cultural applications research
 - We are moving ahead
 - There is more to do
- Need more evidence as to the role of European Culture in Wealth Creation
- Engaging Society – participatory, interactive, contributory
- Outside the walls of libraries, archives, and museums
- The post-modern concept of textuality is alive and vibrant
- Intellectual Capital—community capital
- Some Loss is unavoidable



Library at Hadrian's Villa at Tivoli

© HATII UofGlasgow, 2005

Brace the doors....

- The widespread use of digital technologies is creating a tidal wave of digital information waiting to be ingested into memory institutions—new technologies are needed to make this happen
- Users expectations for access to analogue content is being re-defined by how they can use content in digital form
- User expectations for representations are changing



Records arriving at the Hocken Library in Dunedin (NZ), September 2003

Expectations

- Protection and conservation of cultural memory
- Global access to knowledge
- Preservation accountability
- Reduction of costs (e.g. info reuse)
- Scalable approaches
- Foundation for the knowledge and creative economy
- Delivery and preservation strategies applicable across diverse media and dynamic objects
- Responsive to multilingual and multicultural collections
- Knowledge representation



© HATII UofGlasgow, 2005

Preservation

- Content
- Context
- Provenance
- Interpretability
- Reliability
- Authenticity
- Integrity
- Scalability

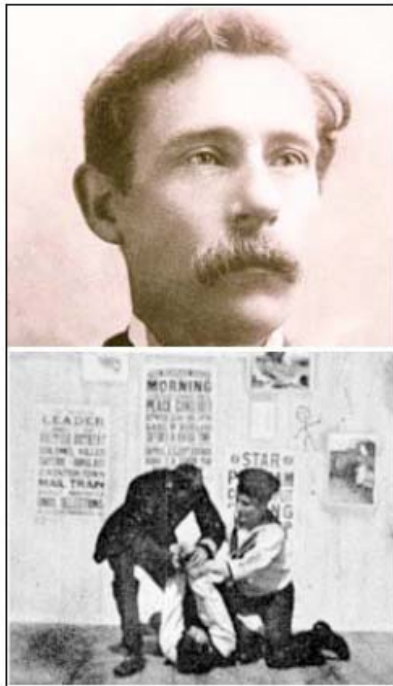


Proactive Collecting Behaviour: There will still be accidental discoveries

UK's earliest film found in attic after 111 years

By NIKKI MURFITT and MATT BORN, Daily Mail 09:43am 19th June 2006

[Reader comments \(5\)](#)



Director Birt Acres (top) made the 50-second film about a pickpocket.

Look here too



Personal, Portable, & Persistent Storage

- Quality consumer storage devices
- Interoperable
- Durable
- Recognisable what they are
- Device independent



Hard disk undergoing one stage of forensic analysis as part of data recover planning at the *Tunstall and Tunstall* (Nepean, Ontario, Canada) in 2003.

ERPANET Case Studies

Variation in awareness of risk

Value of information depends upon recognition of business/organisational dependency, potential of re-use, or risk associated with information (not necessarily cost)

Responsibility rarely taken at corporate level

Few organisations have adequate strategies

Activity is fragmentary: practices tend to be *incomplete, ad hoc, and unitary*

waiting for external solutions

- Preservation and storage poorly understood
- Lack of policies and procedures

- **www.erpanet.org**

Rights Issues

- **Copyright**
- **Privacy**
- **Freedom of Information**
- **Public Sector Information Directive**
- **Approaches to citation**

Collection Building

- Recognition that we need to act proactively
- Need more thought around-- what is a collection?
- Diversity of collections
- Debate surrounds whether we should select



Digitisation & Representation



© HATII UofGlasgow, 2005



© HATII UofGlasgow, 2005

- Integrated multi-modal digitisation technologies
- Automated conversion of manuscript materials
- Technologies that can be used by small and medium sized institutions
- Automated workflow
- New kinds of representation necessary

Representation Technologies

- At Content Level

- Support for automated and deep encoding/markup of documents (post modern sense)
- Multilingual representations
- Emphasis on creation of ontologies and tools for using them

- At Digital Object Level

- Representation Information
 - e.g. file format registeries

- At the Process Level

- Process (e.g. performance)



© HATII UofGlasgow, 2002



© HATII UofGlasgow, 2002

Automation: Collection Building & Management

- Human effort does not scale to the size and complexity of material
- Workflow characterisation
- Preservation process selection (e.g. utility analysis)
- Acquisition of content
 - Web archiving
 - digitisation
- Organisation of content
- Description
- Management
- Collection personalisation
- Scalability



© HATII UofGlasgow, 2005

Semantic and Syntactic Metadata Extraction

- Automated extraction of metadata
- Understandability of objects
- Scalability and cross-media applicability
- Responsiveness to semantic layering
- Applicability to unstructured or semi-structured materials

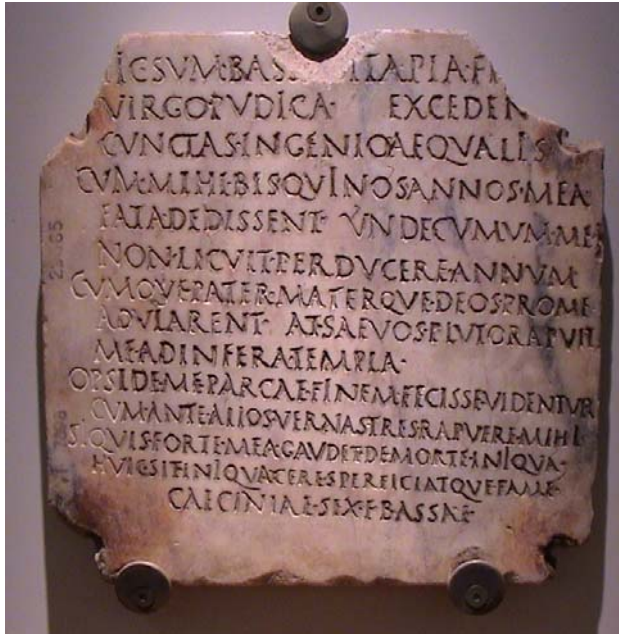


© HATII UofGlasgow, 2002



Example: Bridging the semantic gap – contrast in imaging searches on colour, shape, or texture with those for objects (e.g slippers) or concepts (success) or in the case of audio emotion, decision points, interaction patterns.

Automated Description & Appraisal



- Annotation and Provenance
 - What about structural differences—say between radiological and linguistic data sets and their annotations
- Summarisation technologies (point of view)
 - What about composite documents
 - What about databases, images,
- Knowledge representation developments & tools
- What if collection development testing

Digital Libraries

- Dynamic
- Ubiquitous
- Knowledge driven
- Participatory
- Multi-lingual
- Service layered

European Digital Library, Bricks, The WWW

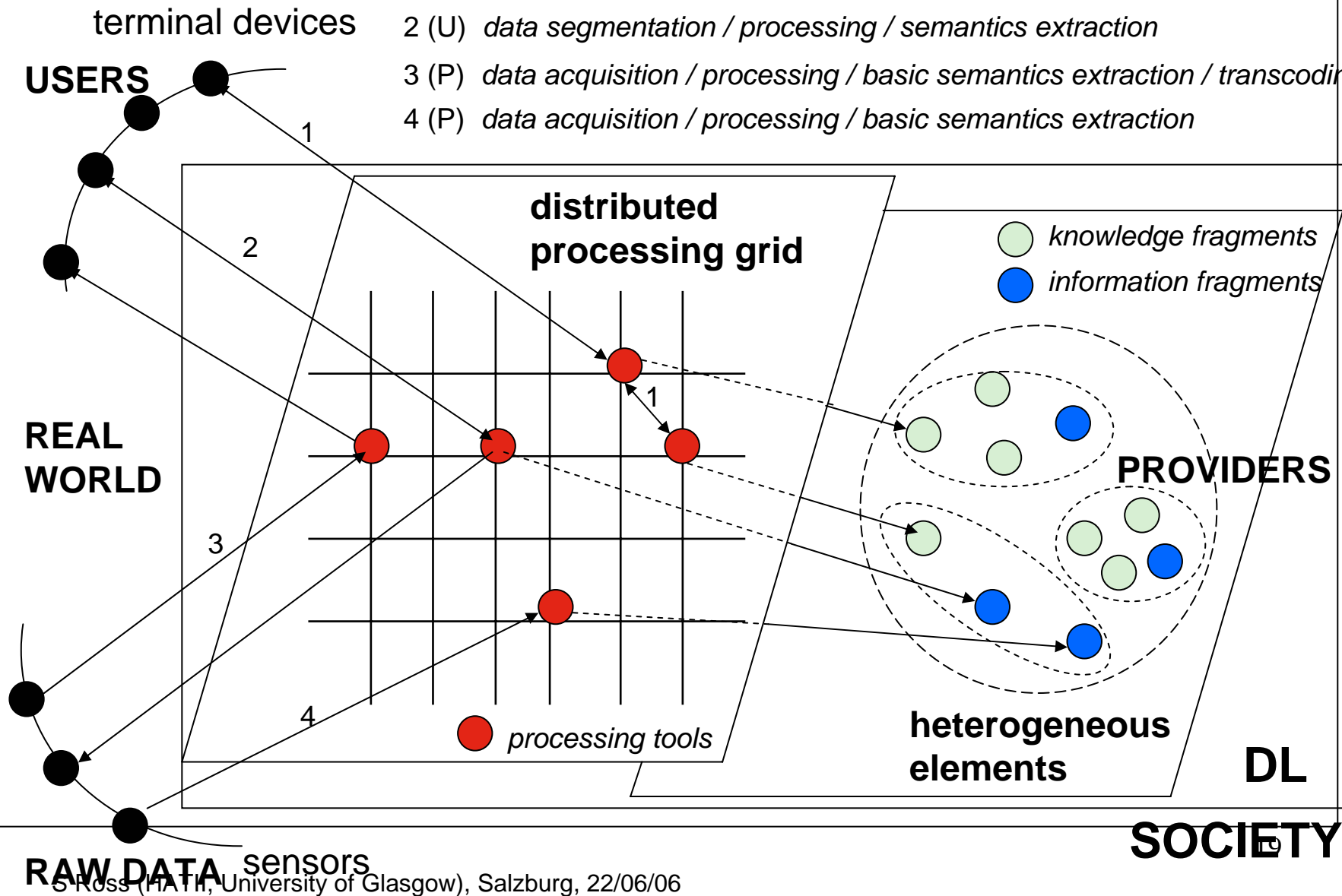


© ERPANET, 2005

Stuart Weibel of OCLC at the grave of George Boole following the ERPANET seminar on *Persistent Identifiers*, Cork, June 2004

DELOS Corvara (del Bimbo, et al 2004)

- 1 (U) *processing / semantics extraction / cooperation / transcoding*
- 2 (U) *data segmentation / processing / semantics extraction*
- 3 (P) *data acquisition / processing / basic semantics extraction / transcoding*
- 4 (P) *data acquisition / processing / basic semantics extraction*



Access Pathways

- Cross media discovery and retrieval (e.g. across audio, images, text, and three dimensional models)
- Interoperability of heterogeneous resources
- Abstract query layers applicable across multilingual contexts
- Personalisation responsive to e.g. cognitive abilities or work tasks or environments what data can and should be collected, how should it be collected, how should it be packaged



© HATII, UofGlasgow, 2004

Annotation

- **Semantic Annotation**
- **Annotation services and applications**
- **User Contributions (e.g. folksonomies)**
- **There are technologies and these are improving all the time, but we need methodologies**

Participatory

- **Storytelling**
- **Communities of practice**
- **Communities of life**
- **Collaboration and contributions**



© HATII UofGlasgow, 2005

Historic Media on Display at the Launch of the UK Digital Curation Centre (DCC), November 04
<http://www.dcc.ac.uk>

Connections and Connectivity

- Interconnections of data and information resulting in knowledge
- Connecting participants and building knowledge communities
- Creating Quality



From Frameworks to Repositories

© ERPANET, 2004



Museums (Vienna)

© ERPANET, 2002



Archives (Wellington NZ)

Digital Repositories

- **Change will be a feature of repositories**
 - **Storage technologies**
 - **Services, close down of some and initiation of others**
 - **Workflows**
 - **Verification mechanisms**
 - **Migration, refreshing, emulation—and**



Digital Repository Infrastructure, Swiss Federal Archives, Berne, October 2004

© ERPANET, 2005

Audit and Certification

- Consider such traditional archives as those held at the Tate or The National Archives (Kew)
- How do we know that our data are secure
- Trust
 - How is it established?
 - How is it maintained?
 - How is it secured?
 - What happens when it is lost?
 - How can it be verified?
- BUT will audit and certification increase our trust in services provided by digital repositories?



© ERPANET, 2005

Speakers at the Lisbon Seminar on The Selection, Appraisal and Retention of Digital Scientific Data,

Managing Complex and Dynamic Digital Entities

- There has been little research to address how interrelationships between the components of compound documents might be maintained.
 - How can complex and dynamic entities be authenticated and their integrity verified?
 - How can dynamic entities be accessioned and managed in an archive?
 - To what aspects of a dynamic document should metadata be attached and what metadata would be required?
 - How do we ensure dynamic qualities across time?
 - At what level is loss of dynamic qualities acceptable? What measurement metrics?

Cost, Risk, and Value

- **Cost**
 - Increasing understanding of collecting, managing, curating, and using
 - Need to know and control
- **Risk**
 - Digital Preservation ‘is a risk management activity at all points along the digital longevity pathway’
- **Value & Benefits**
 - Needs to be the focus of activity

Collaboration and Co-ordination

Competition

Orteig Prize



ID: C-1967-03418, **Credit:**

http://grcimagenet.grc.nasa.gov/GRCDigitalImages/1967/1967_03418L.jpg₃₁

Conclusions

- **Acceptable Loss**
- **Excite**
- **Educate –students, professionals, and popular audiences**
- **Good Practice Guidance (Competence Centres)**
- **Collaboration**
- **Co-ordination**
- **Timing**
- **Competition**

- **Conference provides an intellectual framework**