Digital Forensics as a Retrospective Science: Towards Concepts of Parametric Curation and Adaptive Introspection

Jeremy Leighton John; Department of Digital Scholarship, British Library; London, United Kingdom

Abstract

Archives are created in order to provide information about the past, even if it is the recent past. Not all information is captured in the first place and there are limits to how much information can be retained due to resource constraints. Future analysis of the information will depend on the nature of the original selection process and the information that is subsequently available for analysis. Retrospective science and scholarship concerns the reconstruction of the past using extant information. The effectiveness of future retrospection can be boosted by optimizing the quantity and quality of the information retained: avoiding unnecessary redundancy (through parametric curation) and by adaptively appraising and retaining the information that is most potent for retrospection (through adaptive introspection). The diversity and scale of the digital universe necessitates efficiency across the archival life cycle. The paper aims to advance digital curation theory, albeit using a practical prospect.

Summary of Long Paper

The field of digital forensics has been adopted by curatorial, archival and scholarly researchers and practitioners. It is an example of a retrospective science, aiming to reconstruct past events and objects based on authentic evidence. Other fields of scholarship and science are likewise directed at the reconstruction of the past using traces of evidence existing in the present: astronomy, palaeontology and geology, phylogenetics, comparative linguistics, archaeology and history, to mention the most obvious ones. The theme also emerges in a literary context in Voltaire's story of Zadig, which is itself inspired by ancient folklore. The unifying process is the identification of related information shared by descendants and uniquely individualistic information that is novel, arising independently. It is akin to the evolutionary concept of 'descent with modification'. The existence of conserved information, sometimes highly conserved information, offers possible insights in understanding longterm preservation as well as potential efficiencies in handling critical information.

The use of raw image formats in digital photography is familiar to image scientists. Underlying the approach is the concept of 'parametric image editing', where very large photographic files may be edited and handled through the use of metadata that store information about changes made to a single master file thereby avoiding the burdensome retention of many large edited variant files. It is founded on the recording of metadata about the changes that take place, allowing the process to be reversed or repeated; and thus the ultimate expression of the digital image is the outcome of manipulation by metadata.

In this context again there is conservative information (the master file) and novel information (the metadata that controls the way the image is rendered). This phenomenon may be recognised in other contexts, notably in the manner with which genetic information in living organisms may be expressed in diverse ways depending on the influence of other genetic information (represented as 'regulatory genes'). Thus the terminology of digital imaging science motivates the phrase 'parametric curation' in this paper. In the archival context it may be seen as an approach that aims to use – as far as feasible with available technologies at any time – metadata about changes made rather than unduly replicating identical or similar information. Parametric curation seeks judicious efficiencies where there would otherwise be duplication without benefit, while acknowledging the importance of necessary redundancy, replication as a means of securing backup copies of information held in reserve. Specific examples from the literature will be proffered. Clearly such an approach will depend on the reliability of any reversible procedure, and on convincing indicators of authenticity. In this way the theme returns full circle to the application of digital forensics.

The optimal balance between redundancy and efficiency is a critical one for digital curation and this is true for living organisms too, and so much of the paper will explore parallels with natural systems and evolutionary biology drawing in techniques and perspectives from phylogenetics and biologically inspired adaptive computing. The attempt to balance necessary redundancy and desirable informational diversity with judicious avoidance of duplication without benefit can be usefully informed by insights gleaned from cross-disciplinary research. The interplay between conservative and novel information is considered as is the role of selection and inheritance with ephemeral as well as continual materiality, the archive being a material entity.

Selected References

[1] Jeremy Leighton John, Digital Forensics and Preservation (Digital Preservation Coalition Technology Watch Series, UK, 2012).

[2] Christopher A. Lee, Kam Woods, Matthew Kirschenbaum, and Alexandra Chassanoff. From bitstreams to heritage: putting digital forensics into practice in collecting institutions, BitCurator white paper, 30 September 2013.

[3] Kari Kraus 2009. Conjectural criticism: computing past and future texts. Digital Humanities Quarterly, 3.

[4] Hany Farid, June 2008. Digital image forensics. Scientific American: p. 66-71.

Author Biography

Jeremy Leighton John has a BSc (Hons) in zoology from University College London, a MSc in wildlife ecology from University of New Hampshire, and a DPhil in zoology from Merton College and University of Oxford, having researched aspects of evolution and phylogenetics. Curator of eMANUSCRIPTS at the British Library. Fellow of the Linnean Society; Committee Member of the Section for Archives and Technology (ARA: UK & Ireland); formerly the Library Committee of the Royal Society.