Authenticity and Functionality at Digital Archaeological Data

David BIBBY[1] / Reiner GÖLDNER[2]

[1] State Heritage Management Baden-Württemberg / [2] Archaeological Heritage Office Saxony

Abstract: There is no way around it. Archaeological information has to be archived in digital format. Excavation reports, context descriptions, photo documentation, excavation drawings, maps of find spots, monuments and protected sites, 3D scans of excavation areas and finds - data is created digitally with increased regularity. It is stored increasingly less often as analogue copy on paper or film in filing cabinets, boxes and sleeves - indeed, for some data it is impossible, difficult or even makes no sense to create analogue copies. Digital data provide many more possibilities and more functionality then analogue data. But how far can we trust digital data? Are they *true*? Are they still authentic after years and years in the archive?

Keywords: authenticity, digital archive, archaeological data.

Classic Authenticity

Over the last two decades a number of papers on digital archiving have appeared and some of them also touch on *authenticity*. But the meanings are diverse and it seems that no generally accepted concept exists. Even the „Kriterienkatalog vertrauenswürdige digitale Langzeitarchive“ (→ Nestor 2008b), an important document of the German Nestor Project (→ Nestor 2008a), includes a detailed discussion of a conventional interpretation of *authenticity*, but almost entirely ignores the conflicting poles of authenticity and functionality. *Authenticity* in general means „*verbürgt, echt* (vouched, genuine) (→ Brockhaus 1997, p. 369), Wikipedia describes the meaning thus: *„Authenticity refers to the truthfulness of origins, attributions, commitments, sincerity, devotion, and intentions.”* (→ Wikipedia 2013a). In classic archiving „authenticity” means, that documents and objects remain original, unchanged and undamaged during the archiving process. They don't come to any harm and after many years of archiving their state is identical to the state in which they were in at the time they were ingested into the archive. In other words: content, form and appearance remain untouched and integrous. It must be guaranteed, that documents are not manipulated or even falsified during archiving. The task of an archivist is *„to hand on the documents as nearly as possible in the state in which he received them, without adding or taking away, physically or morally, anything: to preserve unviolated, without the possibility of suspicion, every element in them, every quality they possessed when they came to him”* (→ Jenkinson, 1965).

Modern Authenticity

It is a natural assumption, that digital data should remain unchanged and unviolated to be considered authentic. But the goal of 1965 doesn't match the whole diverse world of digital data. The classical concept of authenticity in the digital domain is problematic:

- A digital object can be cloned (reproduced identically, copied) - so which one is the authentic original?
- A digital document needs a system environment to be used (viewed, read, …). Consider changed system environments and changed presentation software. Do they cause loss of authenticity?

- Life cycles of digital systems are very short. What about data migration to newer versions of data formats?
- Digital documents can be very complex, not only supporting viewing and reading. What about the preservation of functional features?

Clearly, an unspecified absolute authenticity cannot be achieved in archiving digital data. But what can be achieved?

Another glance at some ideas regarding authenticity in the field of information technologies may provide some clues.

- The German Wikipedia presents: In information security authenticity means: *The property of being genuine and able to be verified and be trusted. The verification of a given property is called authentication*. (→ Wikipedia 2013b, translated[1]) and refers to the Internet Security Glossary (→ RFC 4949, p. 29): *„The property of being genuine and able to be verified and be trusted.“* This provides us with a first approach – authenticity is a matter of being verified, genuine, with truthfulness of given properties.
- The National Archives (UK) states: *„an authentic record is what it purports to be and is free from alteration or corruption”* (→ TNA 2002, section 4.1.1, p. 14). Elsewhere in the same document, authenticity is described by identity (attributes that document the singularity, such as the name of the author, the date of origin, the subject) and integrity. Regarding integrity it is stated that: *„a record has integrity if it remains complete and uncorrupted in all its essential respects throughout the course of its existence. This does not mean that a record must be precisely the same as it was when first created, for its integrity to exist and be demonstrated. A record can be considered to be essentially complete and uncorrupted if the message that it is meant to communicate in order to achieve its purpose is unaltered.”* (→ TNA 2002, section 3.1.7, p. 8).
- In the German standard DIN „Information und Dokumentation – Leitfaden zur Informationsübernahme in digitale Langzeitarchive“ we find: *Authenticity implies, that only intended, documented changes in terms of preservation actions were performed on the object. Criteria of authenticity are the significant properties of the information object*. (→ DIN 31645, p. 6, translated[2])

For the sake of completeness it should also be mentioned, that in information technologies authenticity is often reduced to (personal) identity: T*he term authenticity means the property which guarantees that a communication partner is really who he purports to be*. (→ BSI 2013, translated[3]). In this context authentication (the process of verifying authenticity) is limited to the identification (of a person) to, say, get access to a computer system. This specific concept little concerns archaeology and will therefore not be discussed here. In an archaeological context the concept of authenticity is more general.

[1] *„In der Informationssicherheit bezeichnet Authentizität die Eigenschaften der Echtheit, Überprüfbarkeit und Vertrauenswürdigkeit. Die Überprüfung einer behaupteten Eigenschaft wird als Authentifikation bezeichnet.“*

[2] *„Authentizität beinhaltet, dass nur beabsichtigte, dokumentierte Veränderungen im Sinne der Erhaltungsmaßnahmen am Objekt durchgeführt wurden. Maßstab für die Authentizität sind die signifikanten Eigenschaften des Informationsobjekts.“*

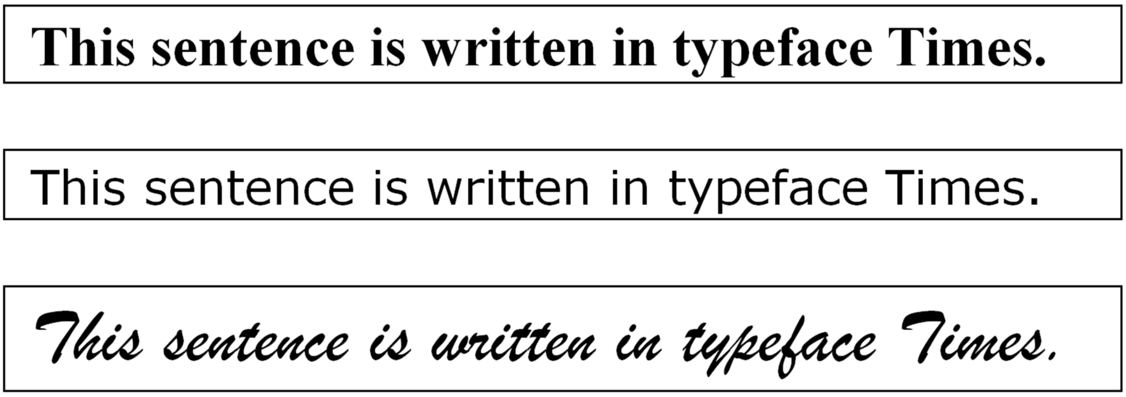
[3] 3 *„Mit dem Begriff Authentizität wird die Eigenschaft bezeichnet, die gewährleistet, dass ein Kommunikationspartner tatsächlich derjenige ist, der er vorgibt zu sein“*

Authenticity is a matter of visible properties, of dedicated significant features. It is not all encompassing. There is a wide choice of properties which can be used to verify authenticity: file size, check sum, digital certificate, font type/size/color, …, content, form, meaning, function and so on. But, which of those are appropriate in our case? Which authenticity features must remain unchanged in the long term? Which can we allow to change? Are there differences in the criteria and weighting according to data types? And how is it possible to verify which properties are changed and which are unchanged after possible migration processes? These issues are of minor importance in classical archiving, but they are highly significant in digital archiving!

Examples of Functionality

It is usual to use hash codes (check sum) or digital watermarks to protect digital data, but most common are digital signatures (e.g. the RSA process). If needed, the binary integrity of a dataset can be tested with these methods. At first glance these methods seem sufficient for simple text and digital photographs. But is it that simple? Some examples:

Example 1:
A text document was created with the font Times. Would it be authentic to display or print this document – with exact the same wording – but with the font Verdana or Script?



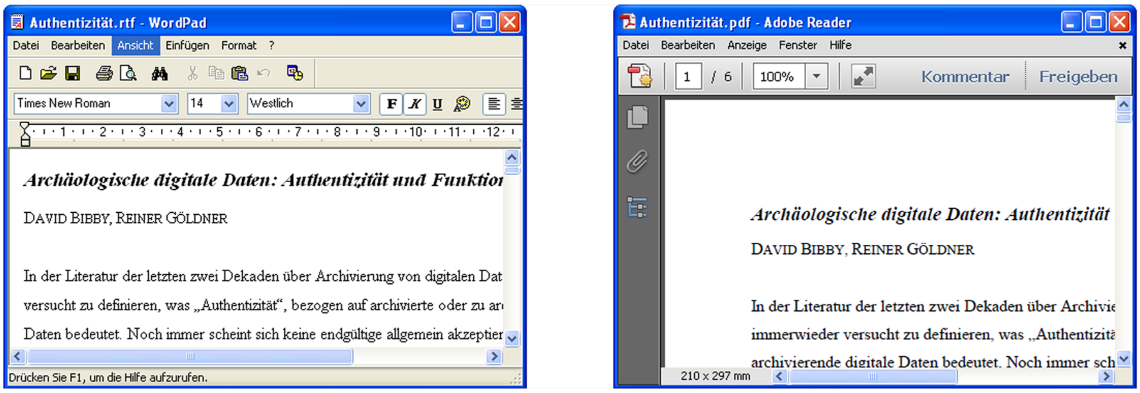
This sentence is written in typeface Times.

This sentence is written in typeface Times.

This sentence is written in typeface Times.

Fig. 1 – Text passage written with various fonts. Which one is authentic?

Example 2:
A text document was created as Rich Text Format (RTF). Would it be authentic to output it as Portable Document Format (PDF) – with the exact same wording and exact the same font – in which case size and inner file structure have changed and therefore the hash code and digital signature are no longer valid?


Authentizität.rtf - WordPad
Datei Bearbeiten Ansicht Einfügen Format ?
Times New Roman
14
Westlich
F K U

Archäologische digitale Daten: Authentizität und Funktion

David Bibby, Reiner Göldner

In der Literatur der letzten zwei Dekaden über Archivierung von digitalen Dat
versucht zu definieren, was „Authentizität", bezogen auf archivierte oder zu ar
Daten bedeutet. Noch immer scheint sich keine endgültige allgemein akzeptier

Drücken Sie F1, um die Hilfe aufzurufen.

Authentizität.pdf - Adobe Reader
Datei Bearbeiten Anzeige Fenster Hilfe
1 / 6
100%
Kommentar
Freigeben
210 x 297 mm

Archäologische digitale Daten: Authentizität

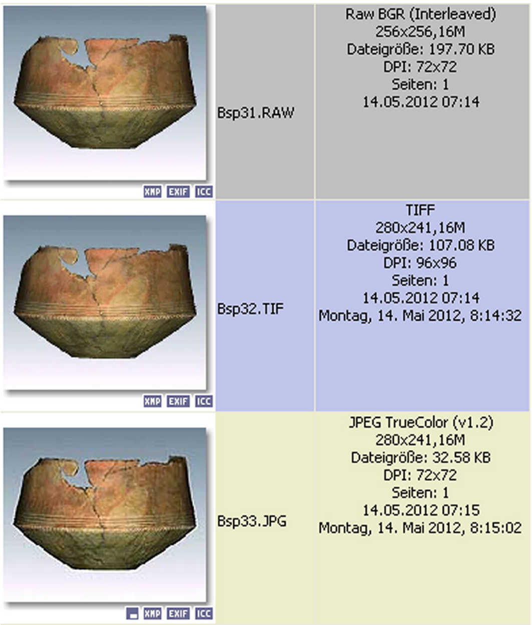
David Bibby, Reiner Göldner

In der Literatur der letzten zwei Dekaden über Archivie
immerwieder versucht zu definieren, was „Authentizitä
archivierende digitale Daten bedeutet. Noch immer sch

Fig. 2 – Presentation of a text: RTF in WordPad (left), PDF in Acrobat Reader (right).

Example 3:

A digital image was created in RAW format, then developed to TIFF and subsequently saved as JPEG. Dynamic range and color space were changed, the appearance was „improved" and the data was compressed with loss – without visible difference. Which of these steps are authentic?



Bsp31.RAW	Raw BGR (Interleaved) 256x256,16M Dateigröße: 197.70 KB DPI: 72x72 Seiten: 1 14.05.2012 07:14
Bsp32.TIF	TIFF 280x241,16M Dateigröße: 107.08 KB DPI: 96x96 Seiten: 1 14.05.2012 07:14 Montag, 14. Mai 2012, 8:14:32
Bsp33.JPG	JPEG TrueColor (v1.2) 280x241,16M Dateigröße: 32.58 KB DPI: 72x72 Seiten: 1 14.05.2012 07:15 Montag, 14. Mai 2012, 8:15:02

Fig. 3 – Short information regarding an image in several data formats (RAW, TIFF, JPEG).

Superficially seen, these questions appear banal, but on closer examination they become less trivial. It might be possible to formulate an adequate definition of authenticity criteria for the above mentioned data types without too much difficulty. But archaeological information in digital form often contains not only textual or pictorial content but also complex information structures and „functionality". Therefore we argue that in our case the conventional definition and understanding of authenticity is lacking.

Example 4:

Excavation databases contain not only data sets with simple textual or visual information about the excavation, but also forms to fill in or edit and reports to output the information. There may also be, functionalities including queries, filters, SQL functions and diagram tools etc. to ease the use of information within the database and promote the evaluation of the excavation data. Here we need an extended concept of authenticity. Not only has the textual content to be preserved, but also (first and foremost) the functionality. Maybe, the appearance can be preserved, maybe not. For example, there is reason for hope, that, say, a PDF/A document created in the year 2010 will appear visually unchanged in the year 2030, even after migration steps. But there is little ground for optimism, that it will be possible to access and use an excavation database in, say, „Paradox" format from 1993 in all its functionality even today. During a migration step an almost complete re-engineering of the old database would probably be necessary to preserve all its functionality.

Fig. 4 – Conglomeration of old and new databases: 1982 dBASE II, 1996 Paradox 7 on floppy disks, 1997 excavation database on floppy disks, 2003 Access on CD, 2008 PostgreSQL in the web.

It is quite clear, that the original appearance of the database doesn't have to be preserved. In this scenario it is not important, that all forms and reports look exactly like the original ones, but content and functionality have to be preserved, associated with descriptions of the database schema and the (archaeological) concepts used in data modelling.

Example 5:

In considering CAD excavation plans we find another complex situation: functionality composed of content (recorded objects, finds etc.) and structure (relations of layers and graphical elements). These plans represent multiplex spatial and content-based associations of the recorded finds, contexts, strata, plana etc., leading to many possible variants of presentation with varying messages to the viewer, all within the same file. A CAD file is not only a static drawing (like a printed plan), but a complex data source for dynamic use of functions such as filtering, querying, layout etc. Archiving of CAD data does not mean preserving a fixed visual appearance. Archiving of CAD data means preserving the ability to adequately reproduce the aforementioned multiplex spatial and content-based associations

Example 6:

Registers of find spots or heritage sites are based on geographic references, hence they are often handled with geographic information systems (GIS). As geographic information can, in simple terms, be considered as a combination of CAD data and database and therefore the criteria from the last example can be reemployed in dealing with pat of the GIS-data. But they alone are not adequate. Dynamic functionality has to be considered from both sides: from the point of view of the database and from the graphical data side. Especially in the case of GIS one important point becomes quite clear: the level of functionality is not only determined by the data itself, but also by the GIS software used and its tools (and ultimately by the whole system environment). For an adequate future use of such data it will be necessary to work with GIS tools, whose functionally is comparable with (or can reproduce) that which was used while creating the data.

Example 7:

The extensive use of the third dimension is still new, but already popular in the field of archaeology. Whilst 3D data represent complex geometric realities they, unlike GIS and CAD, (still) do not contain much functionality. In that way they can be compared to text and image data. Point clouds in ASCII format with columns of X, Y, Z (or X, Y, Z, R, G, B with color information) are often recommended as data type for data exchange and archiving. Such point clouds may become very large, but their structure remains simple. Mesh data is more complicated because of many proprietary file formats. Some data types for meshes seem to be developing into quasi standards (e.g. STL, PLY) but their internal structures are not yet standardized and their archiving qualities are therefore still questionable. This considered the authenticity of 3D data can for the moment be preserved using ASCII data types. At the time of writing it is questionable as to whether the authenticity of more complex 3D formats (mesh data, 3D data with photo-realistic texture etc.) can be verified at all during migration.[4]

3D models are clearly of great interest and use for archaeologists, especially in visualizing their results in an easily accessible form to a wider audience. The unchallenged scientific value of these data sets would benefit from defining criteria for their long term preservation.

Résumé

The exact preservation of the original data is not really a measure of authenticity when archiving digital archaeological data. The aim must be to ensure that content, appearance, structure and functionality of the deposited data remain accessible to the user - to varying extents, according to the type of information, the original data contain. Differing digital archaeological documents have different characteristics. The challenge is to preserve almost all of these characteristics. We cannot avoid the fact that migration steps will certainly influence, even change specific document properties. So it is necessary to decide upon migration methods which don't harm the defined „authenticity properties" and can be used without loss of significant features. Hence: first and foremost, *authenticity properties* and significant features must be defined for each document or for each type of document, keeping in mind that: *„a record can be considered to be essentially complete and uncorrupted if the message that it is meant to communicate in order to achieve its purpose is unaltered."* (→ TNA 2002, section 3.1.7, p. 8).

In DIN 31645 (p. 12) we find: *The digital long-term archive and/or producer have to define the significant features of the information objects to be ingested. The involved parties decide which representation deviations of an information object are still acceptable in future technical environments and which are not.*

[4] Comment R. Göldner: I see some open questions. Is there a method available to compare 3D models of different data types at all? And which properties should be verified? What do we create all these 3D models for? ;-)

Essential criteria are the needs of the users of the digital long-term archive (target group). (translated[5]) also: *The digital long-term archive has to record the significant features, that are chosen and uniquely named for each information object, as meta-data. This meta-data is the indicator for the evaluation of pending preservation measures (converters, emulators) and future platforms*. (translated[6]). Here the InSPECT project is of considerable interest. It analyzed and examined some significant features based on non-specialist audio files, emails, digital images and structured texts. This work should continue and be extended to the field of archaeology!

Guaranteeing authenticity as defined here calls for fairly extensive meta-data with technical, content-based and functional context:

- meta-data on the state of the data at point of ingest into the archive and
- ongoing meta-data entries on all processes, influencing the object within the archive.

These meta-data are official records and must be protected carefully to ensure a trustworthy presentation of the information contained therein (It is self-evident, that original data is not replaced by newer versions during migration, but remain layered within the archive).

The whole range of the topics *authenticity of digital data* and *significant features* is too complex for a short and comprehensive approach. The aim of this paper is simply to attract attention to this as yet neglected theme. Even though we are continually producing digital data, the problems associated with not just their adequate preservation but also with guaranteeing the *authenticity* of that preserved data have so far received little attention. But we must *work it out* if we want to ensure the integrity, accessibility, usability and last but not least the survival of our special and unique archaeological information, with no expiry date.

Resources, References, Web Links

Brockhaus (1997). Brockhaus in 15 Bänden, Bd 1. Leipzig-Mannheim.

BSI (2013). Authentizität. Bundesamt für Sicherheit in der Informationstechnik. (https://www.bsi.bund.de/DE/Themen/Cyber-Sicherheit/Glossar/cs_Glossar_node.html → glossary “Authentizität”)

DigiCULT. Technology Challenges for Digital Culture. (http://www.digicult.info)

DIN 31645. Information und Dokumentation – Leitfaden zur Informationsübernahme in digitale Langzeitarchive.

Duranti, Luciana (1995) „Reliability and Authenticity: The Concepts and Their Implications.” Archivaria 39.

Hirtle, Peter B. Archival Authenticity in a Digital Age. Council on Library and Information Resources, Publication 92. (http://www.clir.org/pubs/reports/pub92/hirtle.html)

InSPECT-Project. Final Report. (http://www.significantproperties.org.uk/inspect-finalreport.pdf)

InterPARES. International Research on Permanent Authentic Records in Electronic Systems. (http://www.interpares.org)

Jenkinson, Hilary (1995) A Manual of Archive Administration. London, 1965.

[5] *„Digitales Langzeitarchiv und/oder Produzent müssen die signifikanten Eigenschaften der zu übernehmenden Informationsobjekte definieren. Die Beteiligten entscheiden, welche Darstellungsabweichungen eines Informationsobjekts in zukünftigen technischen Umgebungen noch akzeptabel sind und welche nicht. Ein maßgebliches Kriterium sind die Bedürfnisse der Nutzer des digitalen Langzeitarchivs (Zielgruppe)."*

[6] *„Das digitale Langzeitarchiv muss die für jedes Informationsobjekt ausgewählten und eindeutig benannten signifikanten Eigenschaften als Metadaten verzeichnen. Diese Metadaten sind der Gradmesser für die Beurteilung anstehender Erhaltungsmaßnahmen (Konverterprogramme, Emulatoren) und zukünftiger Nutzungsplattformen."*

Nestor (2008a). German network of excellence on digital long term preservation. (http://www.langzeitarchivierung.de)

Nestor (2008b). Kriterienkatalog vertrauenswürdige digitale Langzeitarchive, Version 2, Frankfurt 2008. (http://edoc.hu-berlin.de/series/nestor-materialien/8/PDF/8.pdf)

Park , Eun G. (2003). The Role of Authenticity in the Life Cycle of Digital Documents. Graduate School of Library & Information Studies, McGill University. Montreal.

RFC 4949, Internet Security Glossary, Version 2. IETF. (https://tools.ietf.org/html/rfc4949)

TNA (2002). - The National Archives: Generic requirements for sustaining electronic information over time: 1 Defining the characteristics for authentic records.
(http://www.nationalarchives.gov.uk/rmcas/CapacityResourceDetail.asp?Id=213 ,
http://webarchive.nationalarchives.gov.uk/20100604215648/https:/www.nationalarchives.gov.uk/documents/generic_reqs1.pdf)

Wikipedia (2013a): Authenticity. (http://en.wikipedia.org/wiki/Authenticity)

Wikipedia (2013b): Authentizität. (http://de.wikipedia.org/wiki/Authentizität)

(web links last verified at 2013/10/07)