

Significant properties of email

Introduction

Significant properties are those aspects of a digital record that must be preserved over time in order for the Information Object to remain accessible and meaningful. The InSPECT Project is funded by JISC to investigate methods to maintain the authenticity of digital resources across transformation processes and over time. It is developing a framework that will allow institutions to identify, measure, and declare the significant properties of a specified group of digital object types.

This paper contains a brief summary of significant properties, the data model and taxonomy adopted by the InSPECT project. It also provides several examples of the properties that may be considered significant for e-mail messages.

Methodology

The identification of properties that should be considered significant for different classes of record content is not a simple task that can be automated based on a set of universal rules. A set of rules defined for one category of resource may prove to be too restrictive when applied to unusual variations, or inappropriate for other object types. Instead, the InSPECT Project team has sought to take an epistemological approach, by considering the intended purpose of the Information Content during evaluation.

The methodology developed for the project may be separated into four key stages:

1. Identification of the Components of the digital object that an evaluator wishes to maintain.
2. Analysis of the properties of each Component that are required or beneficial for its recreation.
3. Classify the function performed by each property and assess its value.
4. Obtain measurements for each property

To distinguish the properties that are essential from those that are superfluous, an assessor must have a defined understanding of the function performed by each property and its contribution to the whole. The InSPECT project has recognized that the significant properties of information objects may be categorized into two broad categories:

1. *Properties that directly describe characteristics of the intellectual content.*
Properties that refer to the intellectual content itself are technology-independent, remaining the same in analogue and digital forms. Examples include the length of the information content (e.g. duration of an audio recording, number of characters contained in a paragraph) and spatial position (e.g. playback of audio through the left or right speaker, the position and size of a shape in a vector image, the sequential order of several paragraphs on a page).
2. *Properties that indicate the environment in which the intellectual content may be reproduced*
The properties classified into the second category are likely to be derived from the technology (e.g. the encoding format in use) that is used to store the information content. Examples include the quality level (e.g. a higher bit depth and sampling rate to provide an accurate reproduction of a sound recording) or access status of content (e.g. the ability to recreate text, audio, or moving image for viewing or editing).

The ability to classify a property into one of the two categories is influenced by its intended purpose.

Research by the Digital Preservation Testbed and several other projects has identified that the properties of digital artefacts may be classified in five broad categories:

- 1) *Content*: Content is an abstract term to describe the expression of intellectual Work. In a digital environment, Content may describe aspects of the text, still and moving images, audio, and other intellectual productions. Examples: duration, character count.
- 2) *Context*: Context may be applied to any information contained in the digital record that describes the environment in which the Content was created or that affect its intended

meaning. *Examples:* Creator name, date of creation, description of the intellectual work, computer environment in which the Source was created (possibly).

- 3) *Rendering:* The rendering category refers to any information that contributes to the re-creation of the performance. For example, it may be applied to a visual or audible Component. *Examples:* font type, colour and size, bit depth.
- 4) *Structure:* Structure refers to any information that describes the relationship between two or more types of Content, as required to reconstruct the performance. It may be applied to the intrinsic or extrinsic relationships contained in the performance. *Examples:* E-mail attachment
- 5) *Behaviour:* Behaviour is applicable to any information that describes the method in which the Content interacts with other stimuli. Stimuli may include the interaction of the user with the software, or the interaction with other sources of information, such as an external resource that affects the context, content, structure, or appearance of the resource. Behaviour is considered to be the most difficult characteristic to preserve – it is often tied to the capabilities of particular software applications and may be difficult to translate. It is also difficult to define all behavioural characteristics in a quantitative manner. *Examples:* Hyperlinks

The application of the five categories varies for different object types. It is common for text-based objects to possess a large number of artefacts and properties that fit into the content or context categories, whereas images and audio objects are likely to possess properties that are classified in the rendering category.

Data model

The InSPECT Project developed a simple data model to organise the key entities of an object. The data model is influenced by work undertaken by the PREMIS Working Group¹, National Archives Seamless Flow programme², PLANETS Project³, CASPAR Project⁴, OAI-ORE⁵ and other JISC-funded Significant Properties projects⁶. The significant properties data model defines four entities that are considered important for describing the underlying properties that comprise a digital object:

- 1) *Object:* An Object represents “a discrete unit of information in digital form” (PREMIS 2, p6). In a practical digital environment, it may be represented a single file (PREMIS File or Filestream) or two or more files (PREMIS Representation) that are required to re-create the Performance. The Object provides a logical distinction between the macro environment of a digital storage system and the micro environment that is the topic of analysis for this report.
- 2) *Component:* A Component is a unit of information that form a logical group. A Component may represent an intellectual entity (e.g. an image, a text paragraph, a shape in a vector diagram, and other examples) in an Object. A Component may be encoded in one or more files that are interpreted to create a Performance.
- 3) *Property:* A property represents the technical or semantic characteristics that must be recorded to recreate the performance of the Component. It may directly contribute to the recreation of the Component, or indirectly through being required by a second Property. An evaluation of the contribution that it makes to the re-creation of a Component or an Object as a whole is subjective and is likely to require evaluation prior to a judgement being made that it should be removed or quality degraded.
- 4) *Agent:* A person, organisation, or software program/system that is associated with the definition, categorisation and/or evaluation of significant properties contained in an Object. Examples of an agent may include: a Depositor that has submitted a digital object to an

¹ <http://www.loc.gov/standards/premis/>

² http://www.nationalarchives.gov.uk/electronicrecords/seamless_flow/default.htm

³ <http://www.planets-project.eu/>

⁴ <http://www.casparpreserves.eu/>

⁵ <http://www.openarchives.org/ore/>

⁶ <http://www.dpconline.org/graphics/events/080407workshop.html>

digital repository and has defined the intellectual components that must be maintained; an employee of a digital repository that is responsible for performing curatorial work.

Figure 1, based on the PREMIS Data Model⁷ indicates the relationship between the four entities.

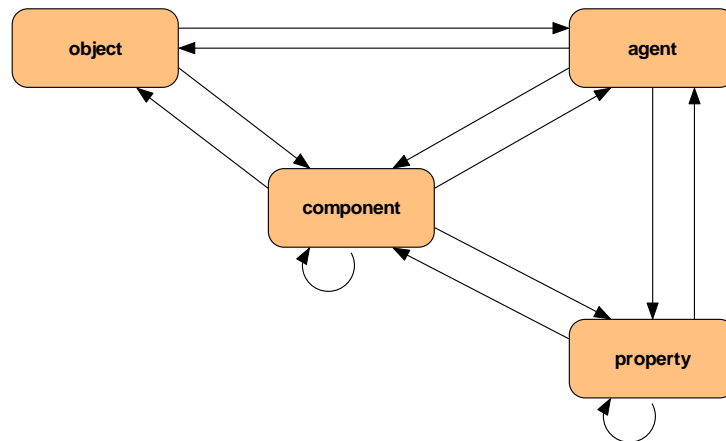


Figure 1. InSPECT data model

The four entities may be linked to other entities using a defined set of rules:

1. An Object may be associated with one or more Components
2. A Component may be associated with one or more Properties or Components;
3. A Property may be linked to one or more Properties if they have some form of mutual dependency.
4. An Agent can be assigned for each entity. However, an Agent cannot be related to other Agents.

⁷ Data Dictionary for Preservation Metadata: PREMIS version 2.0, p5, figure 1

Worked example: Email

Electronic mail is a method of transmitting messages over an electronic communication system. The specification for email messages is defined in several documents, collectively called the Multipurpose Internet Mail Extensions (MIME). The specifications indicate that an email is a compound of data that performs several different purposes that collectively establish the context and content of a message.

1. Component Analysis

By applying the InSPECT data model, the project team has identified that an email may possess three or more logical components.

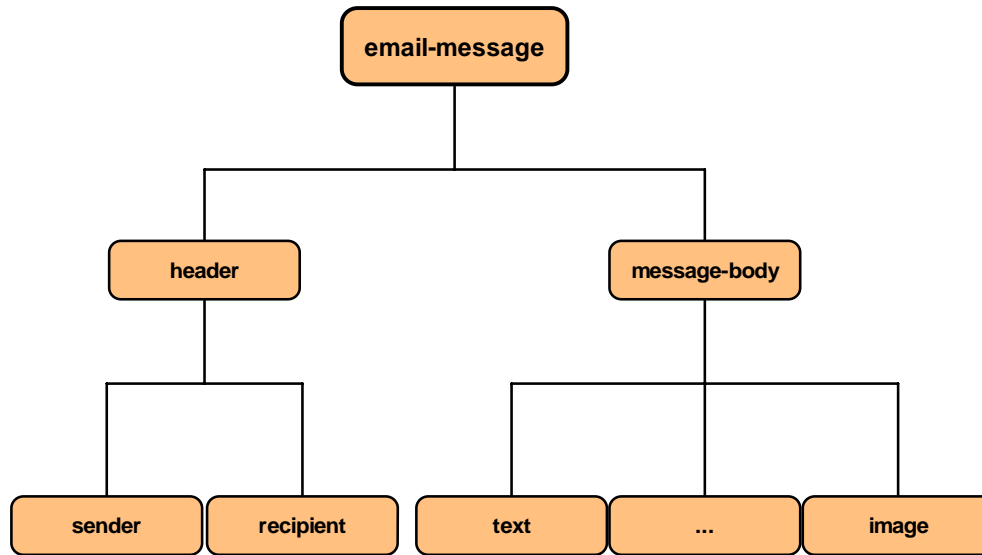


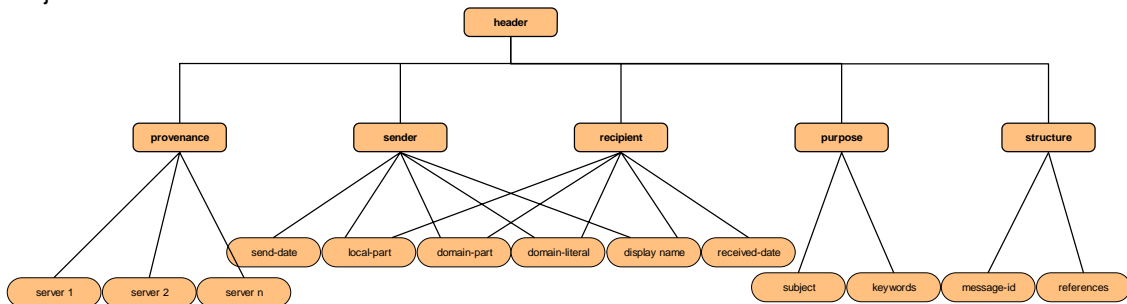
Figure 2: Components of an email

Email message

To understand the message that is being communicated, a digital archive must maintain two high-level components:

1. Header

Structured data that provides information about the sender (name, e-mail address), the path that was taken to deliver the message, the intended recipient (name, e-mail address), an indication of the subject and other relevant information.

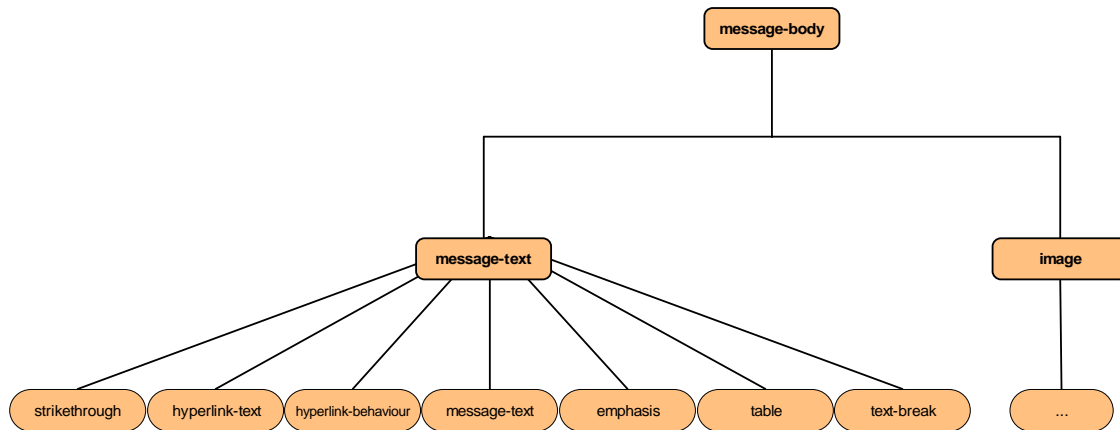


- 1) *Sender*: structured information that indicates the person responsible for transmitting the email message and the date it was despatched
- 2) *Recipient*: structured information that indicates one or more persons that are the intended recipient of the email message. The recipient may be indicated as the primary, secondary, or tertiary recipient through the use of the To, Carbon Copy (CC) and Blind Carbon Copy fields.
- 3) *Provenance*: The route that the email has taken to reach the recipient's mailbox, indicating the servers through which it passed.

- 4) *Purpose*: The purpose of the message (subject, keywords) and its role within the email system (message id).

2. Message-Body

The message body often contains a combination of unstructured or semi-structured information that may be displayed as text paragraphs, tables, lists and other categorization methods. Information may be encoded as plain text, HTML, or Microsoft Rich Text format. These three allow the creator to specify attributes of the appearance, such as layout, colour, size, etc.



The type of information contained in the message body is less consistent in comparison to the organisational structure of the header. It may (or may not contain) one or more sub-components, such as message text and images that possess specific properties.

2. Property Analysis

Sender component

<i>Title:</i>	sender
<i>Description:</i>	The agent responsible for sending the message.
<i>Function: Class</i>	Content / Context
<i>Preservation Level: Valuation</i>	0 1 2 3 4 5 6 7 8 9 [10]

Property

<i>Title:</i>	local-part	
<i>Description:</i>	The username or other identifier in use by the creator, prior to the @ symbol	
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level: Valuation</i>	0 1 2 3 4 5 6 7 8 9 [10]	
<i>Rationale</i>		
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	Equality	Equality
<i>Measurement Unit:</i>		
<i>Measurement Value:</i>	gordon.brown	gordon.brown
<i>Measurement data type:</i>	alphanumeric	alphanumeric

Property

<i>Title:</i>	domain-part	
<i>Description:</i>	A host name or domain name that is used by a DNS to indicate the origin of the message	
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level: Valuation</i>	0 1 2 3 4 5 6 7 8 9 [10]	
<i>Rationale</i>		
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	Equality	Equality
<i>Measurement Unit:</i>		
<i>Measurement Value:</i>	number10.gov.uk	number10.gov.uk
<i>Measurement data type:</i>	alphanumeric	alphanumeric

Property

<i>Title:</i>	domain-literal	
<i>Description:</i>	An indicator of the source domain of the message specified by its IP (numeric) address.	
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level: Valuation</i>	0 1 2 3 4 [5] 6 7 8 9 10	
<i>Rationale</i>		
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	Equality	Equality
<i>Measurement Unit:</i>		
<i>Measurement Value:</i>	222.222.222.222	222.222.222.222
<i>Measurement data type:</i>	. 0-9	. 0-9

Property

<i>Title:</i>	display-name	
<i>Description:</i>	An indicator of the source domain of the message specified by its IP (numeric) address.	
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	

<i>Preservation Level: Valuation</i>	0 1 2 3 4 [5] 6 7 8 9 10
<i>Rationale</i>	
<i>Function:</i>	Preservation Distribution
<i>Measurement Type:</i>	Equality Equality
<i>Measurement Unit:</i>	
<i>Measurement Value:</i>	Gordon Brown Gordon Brown
<i>Measurement data type:</i>	alphanumeric alphanumeric

Property

<i>Title:</i>	send-date
<i>Description:</i>	The date and time that an email was submitted to a mail server by a Sender
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure
<i>Preservation Level: Valuation</i>	0 1 2 3 4 5 6 7 8 9 [10]
<i>Rationale</i>	
<i>Function:</i>	Preservation Distribution
<i>Measurement Type:</i>	Equality Equality
<i>Measurement Unit:</i>	
<i>Measurement Value:</i>	2008-09-23T16:34Z 2008-09-23T16:34Z
<i>Measurement data type:</i>	ISO 8601 ISO 8601

Recipient Component

The Recipient component contains properties similar to those outlined for the Sender. An exception is send-date, which is substituted for received-date

<i>Title:</i>	recipients-primary1
<i>Description:</i>	The agent that receives the message.
<i>Function: Class</i>	Content / Context
<i>Preservation Level: Valuation</i>	0 1 2 3 4 5 6 7 8 9 [10]

Property

<i>Title:</i>	received-date
<i>Description:</i>	The date and time that an email was received by a recipient.
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure
<i>Preservation Level: Valuation</i>	0 1 2 3 4 5 6 7 8 9 [10]
<i>Rationale</i>	
<i>Function:</i>	Preservation Distribution
<i>Measurement Type:</i>	Equality Equality
<i>Measurement Unit:</i>	
<i>Measurement Value:</i>	2008-09-23T16:34Z 2008-09-23T16:34Z
<i>Measurement data type:</i>	ISO 8601 ISO 8601

Purpose Component

<i>Title:</i>	purpose
<i>Description:</i>	A summary of the intended purpose of the message
<i>Function: Class</i>	Content / Context
<i>Preservation Level: Valuation</i>	0 1 2 3 4 [5] 6 7 8 9 10
<i>Rationale:</i>	

Property

<i>Title:</i>	subject
<i>Description:</i>	A short string that identifies the topic of the message

<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level:</i>	0 1 2 3 4 5 6 7 8 9 [10]	
<i>Valuation</i>		
<i>Rationale</i>		
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	Equality	Equality
<i>Measurement Unit:</i>	alphanumeric	alphanumeric
<i>Measurement Value:</i>	Re: Your message	Re: Your message
<i>Measurement data type:</i>	UTF-8	UTF-8

Property

<i>Title:</i>	keywords	
<i>Description:</i>	A list of important words and phrases that might be useful for searching	
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level:</i>	0 1 2 3 4 [5] 6 7 8 9 10	
<i>Valuation</i>		
<i>Rationale</i>		
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	Equality	Equality
<i>Measurement Unit:</i>	alphanumeric	alphanumeric
<i>Measurement Value:</i>	Dublin core, metadata, rdf	Dublin core, metadata, rdf
<i>Measurement data type:</i>	alphanumeric	alphanumeric

Provenance Component

<i>Title:</i>	provenance	
<i>Description:</i>	The route that the email has taken to reach the recipient's mailbox, indicating the servers through which it passed	
<i>Function: Class</i>	Content / Context	
<i>Preservation Level:</i>	0 1 2 3 4 (5) 6 7 8 9 10	
<i>Valuation</i>		
<i>Rationale:</i>	Text is, in many circumstances, used as the primary form of communication in an email message.	

Property

<i>Title:</i>	server	
<i>Description:</i>	The server through which the message has passed to reach its destination.	
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level:</i>	0 1 2 3 4 (5) 6 7 8 9 10	
<i>Valuation</i>		
<i>Rationale</i>		
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	Equality	
<i>Measurement Unit:</i>	alphanumeric	
<i>Measurement Value:</i>		
<i>Measurement data type:</i>		

Structure Component

<i>Title:</i>	message-id	
<i>Description:</i>		
<i>Function: Class</i>	Content / Context	
<i>Preservation Level:</i>	0 1 2 3 4 (5) 6 7 8 9 10	
<i>Valuation</i>		
<i>Rationale:</i>	Structural information assigned to an email message that indicates its placement in a threaded discussion.	

Property

<i>Title:</i>	message-id	
<i>Description:</i>	An identifier assigned to an email message that indicates its placement in an threaded discussion.	
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level: Valuation</i>	0 1 2 3 4 (5) 6 7 8 9 10	
<i>Rationale</i>		
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	Equality	
<i>Measurement Unit:</i>	alphanumeric	
<i>Measurement Value:</i>	19980506192030.26456.gmail@cr.yip.to	
<i>Measurement data type:</i>		

Property

<i>Title:</i>	references	
<i>Description:</i>	The emails that precede the current message and may contribute to an understanding of the message in a larger context.	
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level: Valuation</i>	0 1 2 3 4 (5) 6 7 8 9 10	
<i>Rationale</i>		
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	Equality	
<i>Measurement Unit:</i>	alphanumeric	
<i>Measurement Value:</i>	19951223192543.3034.gmail@silverton.berkeley.edu	
<i>Measurement data type:</i>		

Message-text Component

<i>Title:</i>	message-text	
<i>Description:</i>	The text of an email message that contains intellectual information for interpretation.	
<i>Function: Class</i>	Content / Context	
<i>Preservation Level: Valuation</i>	0 1 2 3 4 5 6 7 8 9 [10]	
<i>Rationale:</i>	Text is, in many circumstances, used as the primary form of communication in an email message.	

Property

<i>Title:</i>	text	
<i>Description:</i>	The textual information that is to be communicated	
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level: Valuation</i>	0 1 2 3 4 5 6 7 8 9 [10]	
<i>Rationale</i>		
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	Equality	Equality
<i>Measurement Unit:</i>	alphanumeric	
<i>Measurement Value:</i>	For attention of Tony Blair. Sincerely, G. Brown	
<i>Measurement data type:</i>	UTF-8	

Property

<i>Title:</i>	hyperlink-text	
<i>Description:</i>	An indication of the hyperlink embedded within plain text	

<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level: Valuation</i>	0 1 2 3 4 5 6 7 8 9 [10]	
<i>Rationale</i>		
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	equality	equality
<i>Measurement Unit:</i>	URI	URI
<i>Measurement Value:</i>	http://www.significantproperties.org.uk	http://www.significantproperties.org.uk
<i>Measurement data type:</i>	alphanumeric	alphanumeric

Property

<i>Title:</i>	hyperlink-behaviour	
<i>Description:</i>	An indication of the method in which the hyperlink should be treated by rendering software	
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level: Valuation</i>	0 1 [2] 3 4 5 6 7 8 9 10	
<i>Rationale</i>	The distribution version is intended for use by a researcher, who may find it beneficial for hyperlinks to be active (i.e. the user can click on them to progress to a destination URL).	
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	Equality	Equality
<i>Measurement Unit:</i>		
<i>Measurement Value:</i>	Inactive	Active
<i>Measurement data type:</i>	Boolean / controlled vocabulary	Boolean / controlled vocabulary

Property

<i>Title:</i>	text-break	
<i>Description:</i>	A distinct portion of written text that may indicate a logical distinction of text. A text break may indicate an paragraph that represents an intentional division of information or an unintentional separation imposed by technology limitations.	
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level: Valuation</i>	0 1 [2] 3 4 5 6 7 8 9 10	
<i>Rationale</i>		
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	Range	Range
<i>Measurement Unit:</i>		
<i>Measurement Value:</i>	<p></p> 0D 0A	<p></p> 0D 0A
<i>Measurement data type:</i>	controlled vocabulary	controlled vocabulary

Property

<i>Title:</i>	table	
<i>Description:</i>	A combination of words, numbers, or other information that is arranged in parallel columns. The distinction between rows and columns provide semantic meaning.	
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level: Valuation</i>	0 1 2 3 4 5 6 7 [8] 9 10	
<i>Rationale</i>		
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	Range	Range

<i>Measurement Unit:</i>		
<i>Measurement Value:</i>	<table><tr></td></tr></table> able “ ”	<table><tr></td></tr></table> “ ”
<i>Measurement data type:</i>	controlled vocabulary	controlled vocabulary

Property

<i>Title:</i>	strikethrough	
<i>Description:</i>	A typographical presentation of words with a line through the centre. A strikethrough may be used to indicate legitimate information that has been recently deleted, or words that are not intended for inclusion	
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level: Valuation</i>	0 1 2 3 4 5 6 7 [8] 9 10	
<i>Rationale</i>		
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	Range	Range
<i>Measurement Unit:</i>		
<i>Measurement Value:</i>	text text <strike></strike> <s></s>	text text <strike></strike> <s></s>
<i>Measurement data type:</i>	controlled vocabulary	controlled vocabulary

Property

<i>Title:</i>	emphasis	
<i>Description:</i>	The accentuation of one or more words for emphasis.	
<i>Function:</i>	Content / Context / Rendering / Behaviour / Structure	
<i>Preservation Level: Valuation</i>	0 1 2 3 4 5 6 7 [8] 9 10	
<i>Rationale</i>		
<i>Function:</i>	Preservation	Distribution
<i>Measurement Type:</i>	Range	Range
<i>Measurement Unit:</i>		
<i>Measurement Value:</i>	<emphasis></emphasis> 	<emphasis></emphasis>
<i>Measurement data type:</i>	controlled vocabulary	controlled vocabulary