



Document Name			
Document Title	Significant Properties Data Dictionary		
Work Package			
Author(s) & project role	Gareth Knight, Digital Curation Specialist		
Date	12 February 2010	Filename	sigprop-dictionary
Access	<input checked="" type="checkbox"/> Project and JISC internal	<input checked="" type="checkbox"/> General dissemination	

Document History		
Version	Date	Comments
1.1	12 February 10	Copied text into revised template
1.0	21 August 08	Removed duplicate relationshipSubType entry and changed endnotes to footnotes.
0.5	05 August 08	Corrections to layout. Replaced AHDS with CERCH logo. Addition of section on other description languages.
0.4	31 July 08	Description of Agent entity and semantic units provided in documents, expansion of relationship units.
0.3	29 June 08	Addition of genre and measurement weighting semantic unit
0.2	10 June 08	Revised conceptual model and data elements on the basis of feedback from SG, AB & LM
0.1	05 February 08	First draft of data dictionary outlined in 'Framework for the definition of significant properties'

Contributors

The following people have made direct or indirect contribution to this report: Adrian Brown, Richard Davis, David Duce, Stephen Grace, Ian Hodges, Brian Matthews, Lynne Montague, Maureen Pennock, Mike Stapleton and Paul Wheatley.

Intended Audience

This document is written for use by the InSPECT project team, the JISC community and those interested in digital preservation.



Table of Contents

Table of Contents	1
Table of Contents	2
Introduction	3
Purpose of the significant properties data dictionary	3
Distinction from other object description languages	3
Properties that may be recorded in the data dictionary	3
Implementation	3
Data Model	3
Overview	3
Hierarchy of semantic units in the Significant Properties Data Dictionary	3
1.1 Identifier	3
1.2. title	3
1.3. definition	3
1.4. function	3
1.5. preservationLevel	3
1.6. specificationRegistry	3
1.7. measurement	3
1.8. genre	3
2 Agent	3
2.1. agentIdentifier	3
2.2. agentName	3
2.3. agentRole	3
3. Relations	3
3.1.relationship	3

Introduction

Significant properties are those aspects of a digital record that must be preserved over time in order for it 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 report provides an overview of a data dictionary for the description of significant properties for one or more digital objects, indicating the type of information that may be recorded and the method in which it may be structured.

Purpose of the significant properties data dictionary

In recent years it has been recognised that curation and preservation activities are required to maintain information in a form that is accessible and authentic. The difficulties presented by the creation of digital information arise as a result of the changing state of the technical environment and subject domain in which information may be accessed over a period of many years. New techniques must be developed to maintain access to information, including the recreation of the environment in which information was accessible in its original form and its conversion to new formats for use by contemporary hardware and software. To ascertain that the information presented to the user has maintained its authenticity and integrity, some form of validation is required to compare the information in its original form with its current manifestation to ensure that it is completed and unchanged in the aspects that are considered to be significant.

The significant properties data dictionary, as outlined in this report provides a framework in which institutions may specify and record the aspects of digital information that they consider to be useful to maintain. It may be used to describe a diverse range of properties associated with one or more types of digital object. The data model on which it is based is deliberately abstract and application-independent to allow it to be applied to a range of content types (still raster and vector images, moving images, audio, datasets, and text based data) and expression methods (RDF, database).

The significant properties data dictionary will support the management of digital objects by enabling institutions to:

1. Identify, analyse and record the components and properties of a digital object
2. Evaluate the subjective value of each component and property that represent the information content.
3. Assign quantitative and qualitative quality thresholds for the recreation of information content
4. Evaluate the recreation of information content by comparing properties stored in different manifestations and evaluate if information is intact or incomplete.
5. Obtain information regarding the ability to maintain each property when converting to a different encoding format, by querying a third-party service.

Although many institutions possess procedural lists or guidelines that indicate the properties associated with specific types of digital object, they are often stored in different locations and forms¹. The use of the data dictionary and its various implementations enable institutions to record information that is relative to the object of analysis and store it in the same storage environment, for subsequent retrieval. The information may also be supplemented by information provided by a third-party, such as a format registry.

¹ Many institutions continue to store details of digital objects in printed records.

Distinction from other object description languages

Many data dictionaries exist that are able to store information on the internal structure of a digital object. These have been created for use by many academic fields and cater for different genres and formats of digital object. Examples include The Data Format Description Language (DFDL)², Binary Format Description (BFD) Language, Bitstream Syntax Description Language (BSDL), Binary XML Description Language (BinX)³, and Earth Science Markup Language (ESML)⁴, as well as related work such as the Message Transmission Optimization Mechanism (MTOM) and Extensible Scientific Interchange Language (XSIL)⁵.

The significant properties data dictionary is distinct from existing description languages in several key aspects: it may be applied to any type information irrespective of resource type, file format and subject discipline; the data model on which the significant properties data dictionary is built is abstract and non-prescriptive in its implementation, allowing implementers to produce a framework that is suitable for diverse types of resource; and it has been built and tailored to the requirements of the digital preservation, enabling institutions to record the relationship between properties of different encoding formats, among other factors⁶.

Properties that may be recorded in the data dictionary

The choice of properties that must be recorded for each digital object is a management activity that requires the consideration of issues associated with the perception of value by the assessor and institution. Although work being performed by PLANETS and other projects demonstrate it is possible to provide a common framework in which to make decisions, it has demonstrated that many choices are inherently subjective⁷. A discussion of the factors that contribute to an evaluation of significance is outside the scope of this report. However, it is useful to provide the reader with an overview of the type of information that may be recorded, based on experience gained in the InSPECT Project and the four JISC-funded significant properties projects. The significant properties of a digital object may be organised into two broad categories:

1. *Properties that describe characteristics of the intellectual content.* Properties that may be classified in this category are, in many examples, likely to be technology-independent. The same type of information would be recorded for intellectual content if it was stored on analogue or digital media. Examples include the:
 - a. Length of the information content, e.g. duration of an audio recording, number of characters contained in a paragraph.
 - b. Placement of the information content, 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 and indirectly refer to the intellectual content.* 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:

² <http://forge.gridforum.org/projects/dfdl-wg/>

³ A language used to describe the content, structure and physical layout (endian-ness, blocksize) of binary files.

⁴ ESML may be used to describe the structure of various earth science data formats, which can be used to parse and decode the data format.

⁵ XSIL is a hierarchical, extensible, transport language for the description of scientific data objects.

⁶ In addition, a distinction may be made between the conceptual basis of description languages that are intended to replicate the information contained in a digital object (e.g. XCDL) and the Significant Properties data dictionary which is intended to supplement the digital object.

⁷ See Knight (2008) 'Deciding Factors', available at <http://www.significantproperties.org.uk> for criteria that may influence an institution's choice, including the existing investment and availability of future funding.

- a. Quality level that is used to store the information content, e.g. a higher bit depth and sampling rate to provide an accurate reproduction of a sound recording.
- b. Access status of the information 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. For a small number of properties it is possible to classify them into both categories.

Implementation

The data dictionary may be implemented using several forms, including Resource Description Framework (RDF), an XML-based metadata schema, a database, or other expression. The InSPECT Project team are creating a XML schema using a 'significantProperties' and 'PREMIS' namespace.

Data Model

Overview

A data or content model provides a structural definition for a type of entity. They are commonly defined to indicate an intended structure, accompanied by a set of rules that specify the methods in which the entity may be utilised. Several conceptual models exist that have been developed or have been applied to the modelling of digital objects. It is common for each model to introduce concepts and terminology developed in a particular discipline for application to digital objects and, as a result, the conceptual basis and terminology in use often vary. Common data modeling techniques for digital objects that may be familiar to the reader include the FRBR⁸, ABC⁹ and PREMIS¹⁰ data models. The TNA conceptual data model was produced by The National Archives for the Seamless Flow programme and has subsequently been adapted by the PLANETS Project¹. However, they are broadly compatible. Table 1 indicates the key terms in each data model and the relationship between the various entities.

TNA	PREMIS	FRBR
-	-	Work
Deliverable Unit	Intellectual Entity	Expression
	Representation	
Manifestation	-	Manifestation
-	Object	Item
File	File	-
Bitstream	Bitstream	-
-	Filestream	-

Table 1: Key terms in the TNA, PREMIS and FRBR data models

InSPECT 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 Group¹¹, National Archives Seamless Flow programme¹², PLANETS Project¹³, CASPAR Project¹⁴, OAI-ORE¹⁵ and other JISC-funded Significant Properties projects¹⁶. The project team have adopted terminology and definitions from these studies when appropriate to avoid unnecessary reinvention.

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). The InSPECT Object is a compound of many types of information (e.g. text, images, sounds, etc.) consisting of intellectual or technical components. It is broadly compatible with the FRBR ‘Item’¹⁷, PREMIS ‘Object’ (which may be a File, Filestream, or Bitstream,

⁸ <http://www.ifla.org/VII/s13/frbr/frbr.pdf>

⁹ <http://www.metadata.net/harmony/ABCv2.htm>

¹⁰ <http://www.oclc.org/research/projects/pmwg/>

¹¹ <http://www.loc.gov/standards/premis/>

¹² http://www.nationalarchives.gov.uk/electronicrecords/seamless_flow/default.htm

¹³ <http://www.planets-project.eu/>

¹⁴ <http://www.casparpreserves.eu/>

¹⁵ <http://www.openarchives.org/ore/>

¹⁶ <http://www.dpconline.org/graphics/events/080407workshop.html>

¹⁷ <http://www.ifla.org/VII/s13/frbr/frbr.htm>

OAI-ORE 'Resource' and Software 'Download'¹⁸. 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 represents 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. The InSPECT definition of Component is broadly compatible with that defined in The Significant Properties of Software study¹⁹.
- 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 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 Model²⁰ indicates the relationship between the four entities.

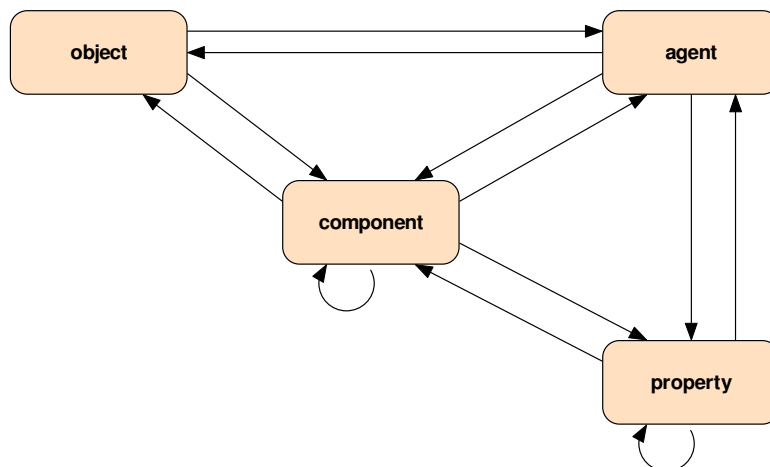


Figure 1. InSPECT data model

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;

¹⁸ http://www.jisc.ac.uk/media/documents/programmes/preservation/spsoftware_report_redacted.pdf

¹⁹ http://www.jisc.ac.uk/media/documents/programmes/preservation/spsoftware_report_redacted.pdf

²⁰ Data Dictionary for Preservation Metadata: PREMIS version 2.0, p5, figure 1

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.

A further description of the interaction between entities is outlined in the Relationship entity (p3).

Hierarchy of semantic units in the Significant Properties Data Dictionary

The Significant Properties data dictionary provides a hierarchy of semantic units that may be used to describe the internal composition of a digital object at the Component (first layer, second layer, and so on) and property, as well as the relationship between these entities. Figure 2 and 3 indicate the optional and mandatory units that should be recorded at the component and property-level.

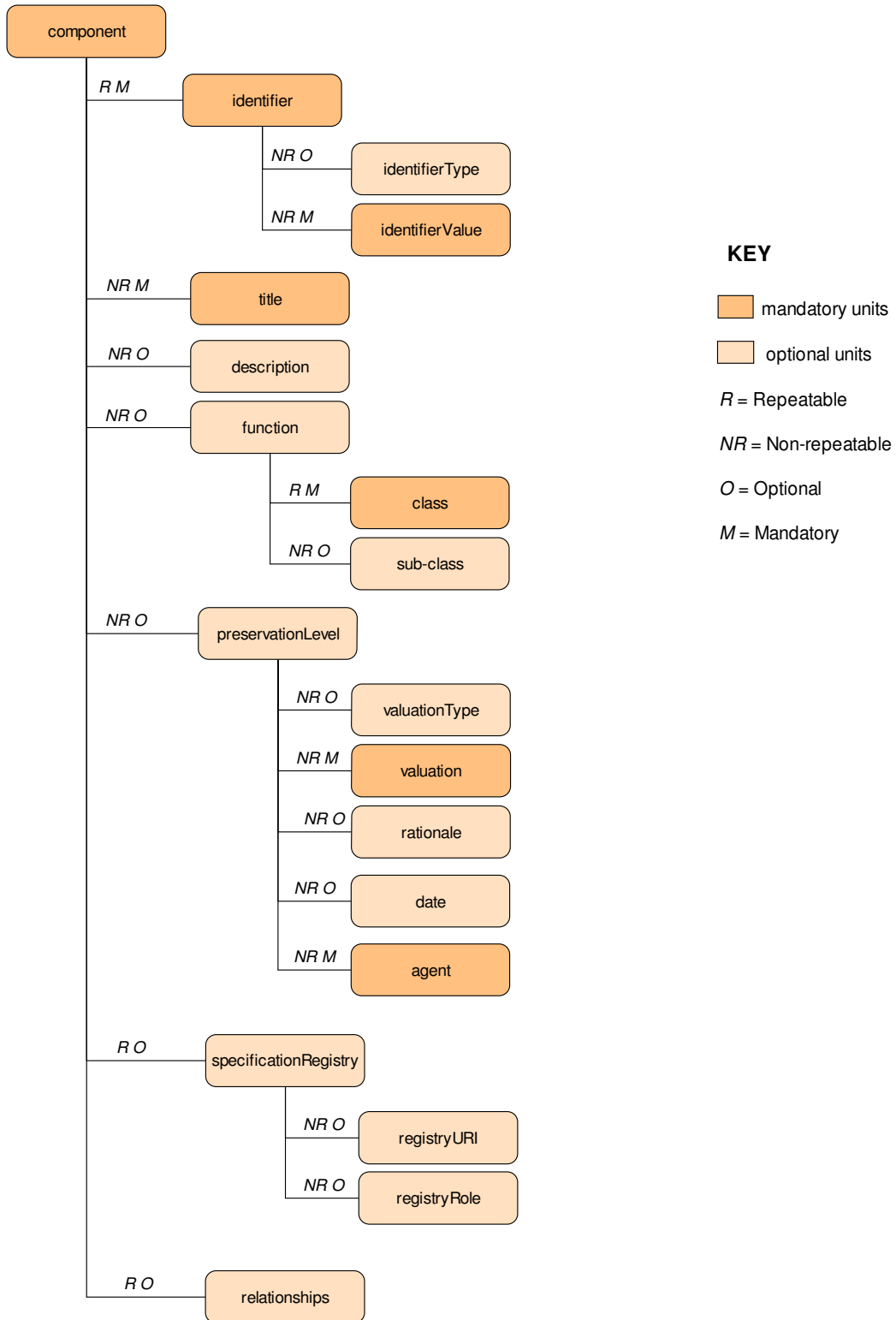


Figure 2: semantic units that may be recorded at the component-level

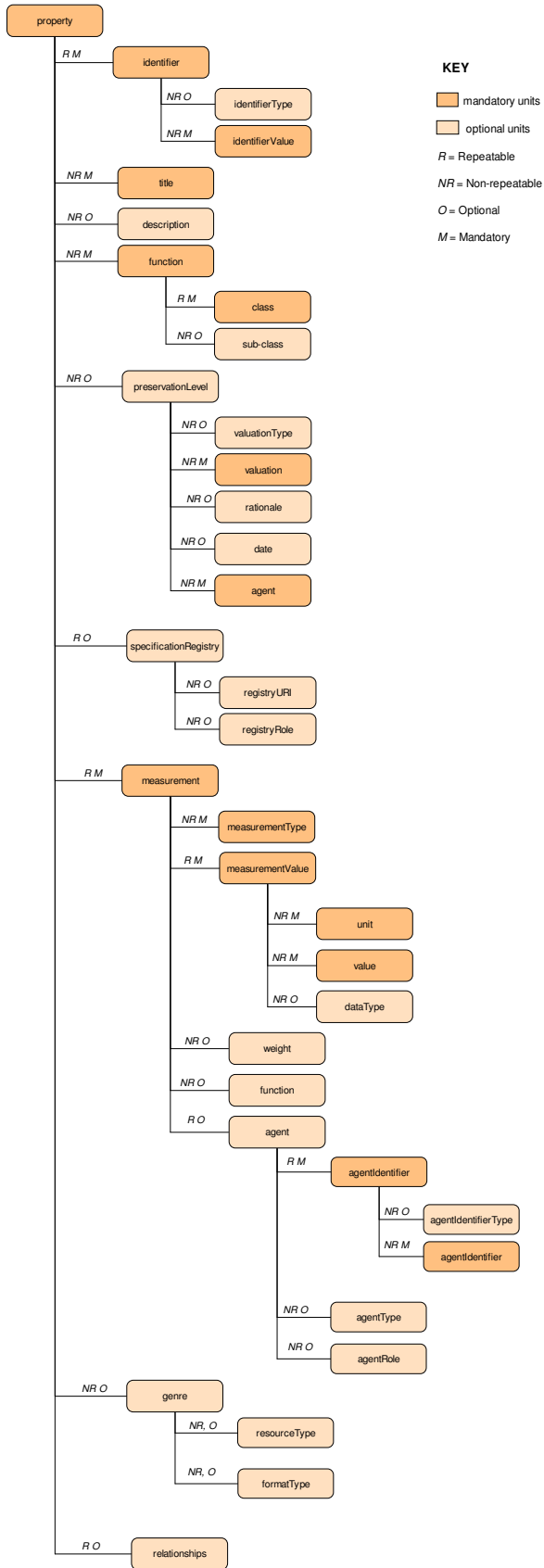


Figure 3: semantic units that may be recorded at the property-level

1. The Significant Properties Data Dictionary

1.1 Identifier

Semantic Unit	identifier		
Semantic Component	1.1.1. identifierType 1.1.2. identifierValue		
Definition	A designation used to uniquely identify the Component or Property within the format registry or preservation repository system in which it is stored.		
Rationale	Each Component and Property must have a unique identifier to allow them to be associated with one or more Objects.		
Data Constraint	container		
Category	object	component	property
Cardinality	1... n	1... n	1... n
Obligation	mandatory	mandatory	mandatory
Examples	n/a	n/a	n/a
Notes	The object category may be equivalent to a PREMIS File or a Representation. The use of identifiers to specify		

1.1.1. IdentifierType

Semantic Unit	identifierType		
Semantic Component	none		
Definition	A designation of the domain within which the identifier is unique.		
Rationale	The Identifier values for components and properties may be the same in two or more domains. The combination of identifierType and identifierValue should ensure uniqueness.		
Data Constraint	A local controlled vocabulary should be defined.		
Category	object	component	property
Cardinality	0...1	0...1	0...1
Obligation	optional	optional	optional
Examples	CERCH AHDS	CERCH AHDS	CERCH AHDS
Notes	The definition and rationale of identifierType is influenced by the PREMIS objectIdentifierType. The identifierType is intended to indicate the domain in which the identifier is unique – this may be the institution, department, or repository software, or other environment identifier.		

1.1.2. IdentifierValue

Semantic Unit	identifierValue		
Semantic Component	none		
Definition	A machine-processable identifier that is unique within the domain.		
Rationale	The identifierValue of a Component or Property must be readily identified.		
Data Constraint	A local controlled vocabulary may be defined.		
Category	object	component	property
Cardinality	1	1	1
Obligation	mandatory	mandatory	mandatory
Examples	00010202	00010202:02	00010202:02:02
Notes	The definition and rationale of identifierValue is influenced by the PREMIS		

	objectIdentifierType
--	----------------------

1.2. title

Semantic Unit	title		
Semantic Component	none		
Definition	A short title that indicates the function of the Component or Property.		
Rationale	Provides a short (often single word) description of the purpose of the component or property.		
Data Constraint	A controlled vocabulary may be defined that indicates the title of component and properties for each content type.		
Category	object	component	property
Cardinality	0	1	1
Obligation	n/a	Mandatory	Mandatory
Examples		still image moving image Dublin core audio-stream1 audio-stream2 text-paragraph10	Sample rate Local-part
Notes	Consistency should be maintained across content types for properties that perform the same function through the use of common terminology. The same title should not be used for two properties that fulfil different purposes.		

1.3. definition

Semantic Unit	definition		
Semantic Component	none		
Definition	A formal statement that human-readable description of the purpose of the component or property		
Rationale			
Data Constraint	Free text		
Category	object	component	property
Cardinality	0	0...1	0...1
Obligation	n/a	optional	optional
Examples		1) A host name or domain name that is used by a DNS to indicate the origin of the message 2) A short string that identifies the topic of the message	See examples in work packages
Notes	The definition should be written and stored by an appropriate service provider, such as The National Archives' PRONOM and does not necessarily need to be stored in the object metadata itself.		

1.4. function

Semantic Unit	function		
Semantic Component	1.4.1. class 1.4.2. sub-class		
Definition	A container that indicates function that the Component or Property performs in the Object.		
Rationale	The function provides a controlled set of terms to describe the functionality of the Component and Property enabling comparison and analysis.		
Data Constraint	container		
Category	object	component	property
Cardinality	0	0...1	1
Obligation	n/a	optional	mandatory
Examples			
Notes			

1.4.1. class

Semantic Unit	class		
Semantic Component			
Definition	The categories that indicate the function that the Component performs for the Object or the Property performs for the Component.		
Rationale	The categories serve as a simple method to understand the purpose of the Component or Property, without the requirement to analyse the specific function of the object type.		
Data Constraint	Controlled vocabulary (content, context, structure, rendering, behaviour)		
Category	object	component	property
Cardinality	0	0...n	1
Obligation	n/a	mandatory	mandatory
Examples		content, context structure	content, context structure rendering behaviour
Notes	The categories were created by Rothenberg (1999) and subsequently utilised by the InterPARES ²¹ , Digital Preservation Testbed ²² and CRIB ²³ projects. The InSPECT Project has replaced the original use of 'appearance' with 'rendering' to enable its use to describe the re-creation of non-visual aspects. However, the reader may choose to use 'appearance' in its original definition if necessary. See InSPECT Significant Properties workflow report for an explanation of the five categories.		

1.4.2. sub-class

Semantic Unit	sub-class		
Semantic Component			

²¹ <http://www.interpares.org/>

²² <http://www.digitaleduurzaamheid.nl/index.cfm?paginakeuze=286&lang=en>

²³ <http://crib.dsi.uminho.pt/>

Definition	An extensible descriptor set that indicates the function that the Component performs for the Object or the Property performs for the Component.		
Rationale	The extensible list enables a service provider to describe the function at greater level of granularity, suitable for specific purposes or object types.		
Data Constraint	The development of a controlled vocabulary is recommended.		
Category	object	component	property
Cardinality	0	0...1	0...1
Obligation	n/a	Optional	Optional
Examples		creator sender raster image still image audio-waveform moving image text	creator sender
Notes			

1.5. preservationLevel

Semantic Unit	preservationLevel		
Semantic Component	1.5.1. valuationType 1.5.2. valuation 1.5.3. rationale 1.5.4. date 1.5.5. agent		
Definition	A container that stores information of the significance of a Component or Property to the recreation of a Performance.		
Rationale			
Data Constraint	container		
Category	object	component	property
Cardinality	0	0...1	0...1
Obligation	n/a	Optional	Optional
Examples			
Notes			

1.5.1. valuationType

Semantic Unit	valuationType		
Semantic Component			
Definition	The basis on which a valuation of the Component or Property is made.		
Rationale	It is useful to define the criteria on which a decision is made, in order to understand the value in its intended context		
Data Constraint	alphanumeric		
Category	object	component	property

Cardinality	0	0...1	0...1
Obligation	n/a	Optional	Optional
Examples		InSPECT-evaluation 1.0 SigPropVectorGraphics [Institution] preservation handbook PLANETS	InSPECT-evaluation 1.0 SigPropVectorGraphics [Institution] preservation handbook PLANETS
Notes			

1.5.2. valuation

Semantic Unit	valuation		
Semantic Component			
Definition	A performance indicator that indicates the intellectual value or contribution that the entity makes to the recreation of the performance.		
Rationale	Valuation may be used to subsequently evaluate the need to recreate the entity in a different environment.		
Data Constraint	Controlled vocabulary		
Category	object	component	property
Cardinality	0	1	1
Obligation	n/a	mandatory	Mandatory
Examples		7.1 10 essential	7.1 10 essential
Notes	The data constraint has been defined as numeric. However, the valuation may use alternative values if required.		

1.5.3. rationale

Semantic Unit	rationale		
Semantic Component			
Definition	A statement of one or more reasons that justifies the valuation, in addition to criteria referenced in valuationType.		
Rationale	A decision on the valuation of an entity may require criteria specific to the digital object or institution to be considered.		
Data Constraint	Free text		
Category	object	component	property
Cardinality	0	0...1	0...1
Obligation	n/a	Optional	Optional

Examples		Institutional mandate	
Notes			

1.5.4. date

Semantic Unit	date		
Semantic Component			
Definition	The date on which the valuation was performed.		
Rationale	Some institutions may consider it to be useful to establish the period in which a valuation occurred.		
Data Constraint	ISO8601 recommended, though other date formats may be used		
Category	object	component	property
Cardinality	0	0...1	0...1
Obligation	n/a	Optional	Optional
Examples		YYYYMMDD YYYY-MM- DD:hh:mm:ss	YYYYMMDD YYYY-MM-DD:hh:mm:ss
Notes	The time period and responsibility for significance evaluation may be record as a PREMIS event or stored in the significant properties schema.		

1.5.5. agent

Semantic Unit	agent		
Semantic Component	agentIdentifier agentName agentRole		
Definition	The agent that is responsible for specifying the valuation		
Rationale	It is useful to identify the agent responsible for defining the valuation of the property.		
Data Constraint	container		
Category	object	component	property
Cardinality			0...1
Obligation			optional
Examples			
Notes			

1.6. specificationRegistry

Semantic Unit	specificationRegistry		
Semantic Component	1.6.1 registryURI 1.6.2. registryRole		
Definition	A third-party resource that can provide further information about the Object, Component or Property. This may indicate a format registry, location of a		

	specification on a trusted site, or other location.		
Rationale	Format registries and published specifications provide useful information for understanding the compositional structure of digital objects.		
Data Constraint	container		
Category	object	component	property
Cardinality	0... n	0... n	0... n
Obligation	Optional	optional	optional
Examples			
Notes	The object category is equivalent to a format specification and should be referenced accordingly.		

1.6.1. registryURI

Semantic Unit	registryURI		
Semantic Component	None		
Definition	A designation that indicates the name or URL of the specification source.		
Rationale			
Data Constraint	None		
Category	object	component	property
Cardinality	0... 1	0... 1	0... 1
Obligation	Optional	Optional	Optional
Examples	www.nationalarchives.gov.uk/pronom	http://www.biblioscape.com/rtf15_spec.htm	http://www.biblioscape.com/rtf15_spec.htm
Notes	The object category is equivalent to a format specification and should be referenced accordingly.		

1.6.2. registryRole

Semantic Unit	registryRole		
Semantic Component	None		
Definition	The purpose or expected use of the registry		
Rationale			
Data Constraint	None		
Category	object	component	property
Cardinality	0... 1	0... 1	0... 1
Obligation	Optional	Optional	Optional
Examples	<ul style="list-style-type: none"> formal specification format registry 	<ul style="list-style-type: none"> specification-formal specification-informal 	<ul style="list-style-type: none"> Format registry
Notes	The object category may be equivalent to a PREMIS File or a Representation.		

1.7. measurement

Semantic Unit	measurement
Semantic Component	measurementType measurementValue weight function

	agent		
Definition	The method in which a property of the Object or Component is measured.		
Rationale	This serves as a method to identify that the Component as been transferred in its entirety and that no errors have occurred that would result in a reduction in quality.		
Data Constraint	container		
Category	object	component	property
Cardinality	0	0	1...n
Obligation	n/a	n/a	mandatory
Examples			
Notes			

1.7.1. measurementType

Semantic Unit	measurementType		
Semantic Component			
Definition	An indicator of the interpretation of the value of the property.		
Rationale	The measurement value may require additional information to be understood in context. This is particularly important for properties that require two or more values to be recorded.		
Data Constraint	A controlled vocabulary should be defined.		
Category	object	component	property
Cardinality	0	0	1
Obligation	n/a	n/a	mandatory
Examples			equality minimum maximum range
Notes	<p>Three constraints are currently recognized:</p> <ul style="list-style-type: none"> • equality: the property stored in the Record must be equal to one or more values stored in the metadata. • minimum: if a numeric measurement is used, minimum indicates the lowest numeric value that is allowed. The minimum and maximum measurement types must be used in combination. • maximum: if a numeric measurement is used, maximum indicates the highest number value that is allowed. For example, the highest sampling rate of an audio recording. • range: the value is one of several that are recorded. 		

1.7.2. measurementValue

Semantic Unit	measurementValue		
Semantic Component	unit value dataType		
Definition	A container that stores the measured value of a property		
Rationale	The value may be measured by one of several measurement scales, e.g. hertz, megahertz.		
Data Constraint	container		
Category	object	component	property
Cardinality	0	0	1... n

Obligation	n/a	n/a	mandatory
Examples			
Notes			

1.7.2.1. unit

Semantic Unit	unit		
Semantic Component	n/a		
Definition	The unit in which the value is measured		
Rationale	The measurement unit indicates the measurement scale in which the property is measured.		
Data Constraint	A controlled vocabulary may be defined for each object type.		
Category	object	component	property
Cardinality	0	0	1
Obligation	n/a	n/a	mandatory
Examples			hertz megahertz no. of characters
Notes			

1.7.2.2. value

Semantic Unit	value		
Semantic Component	n/a		
Definition	A measured value of the property.		
Rationale	This serves as a method to identify that the Component as been transferred in its entirety and that no errors have occurred that would result in a reduction in quality.		
Data Constraint			
Category	object	component	property
Cardinality	0	0	1
Obligation	n/a	n/a	mandatory
Examples			42 96000 blue
Notes	A controlled vocabulary may be defined for each object type, e.g. positive integer for sample rate. See constraint.		

1.7.2.3. dataType

Semantic Unit	dataType		
Semantic Component	n/a		
Definition	The constraint placed on the measured value		
Rationale	This serves as a method to constrain the type of value that may be stored		
Data Constraint	controlled vocabulary		
Category	object	component	property
Cardinality	0	0	0...1
Obligation	n/a	n/a	optional
Examples			positive integer boolean controlled vocabulary

Notes	A controlled vocabulary may be defined for each object type, e.g. positive integer for sample rate
-------	--

1.7.2.4. agent

Semantic Unit	agent		
Semantic Component	agentIdentifier agentName agentRole		
Definition	The agent that is responsible for defining the unit and value.		
Rationale	It is useful to identify the agent responsible for defining the unit and value of the property. This may be a software tool that indicates the property measurement in a digital object (e.g. a sample rate of 96,000), or the person responsible for making a decision on the acceptable variance in a property value.		
Data Constraint	container		
Category	object	component	property
Cardinality	n/a	n/a	0...n
Obligation			optional
Examples			
Notes			

1.7.3. weight

Semantic Unit	weight		
Semantic Component	n/a		
Definition	A declaration of any weighting that may need to be applied to the measurement value		
Rationale	Additional weight may need to be applied in specific circumstances, e.g. the quality of an audio recording is often better if stored in WMA in comparison to MP3 at the same sample rate.		
Data Constraint			
Category	object	component	property
Cardinality	0	0	1
Obligation	n/a	n/a	optional
Examples			wma
Notes			

1.7.4. function

Semantic Unit	function		
Semantic Component	n/a		
Definition	The function or circumstances in which the property measurement value should be used.		
Rationale	The function semantic unit enables the digital repository to declare the situation in which the measurement value should be used.		
Data Constraint	Controlled vocabulary (extensible to support designated community profile of the digital repository)		
Category	object	component	property

Cardinality	0	0	0...n
Obligation	n/a	n/a	optional
Examples			preservation distribution - designated community profile1 distribution - designated community profile2
Notes	The function replaces the designated community profile/intended community semantic unit in earlier versions of the data dictionary.		

1.8. genre

Semantic Unit	genre		
Semantic Component	1.7.3.1. resourceType 1.7.3.2. formatType		
Definition	The category in which a property may be classified.		
Rationale	It is useful to distinguish between properties of different genre, such as content (raster images, audio, text) and format (tiff, gif, html). A relationship may be defined between format-specific properties and general properties that fulfil the same or similar function for other formats that share the same content type. This may be useful when converting technical properties between formats, e.g. TIFF to PNG.		
Data Constraint	container		
Category	object	property	property
Cardinality	0	0	0...1
Obligation	n/a	n/a	optional
Examples			
Notes	function:subclass performs the same task at the component layer		

1.8.1. resourceType

Semantic Unit	contentType		
Semantic Component			
Definition	The category in which a property may be classified.		
Rationale	It is useful to identify and distinguish between properties that are applicable in one or more genre.		
Data Constraint	controlled vocabulary		
Category	object	component	property
Cardinality	0	0	1
Obligation	n/a	n/a	optional
Examples			raster image still image audio-waveform moving image text
Notes	A relationship may be defined between format-specific properties and general properties that fulfil the same or similar function for other formats that share the same content type. This may be useful when converting technical properties between formats, e.g. TIFF to PNG. For the PRONOM format registry, the obligation should be mandatory.		

1.8.2. formatType

Semantic Unit	formatType		
Semantic Component			
Definition	The encoding format in which a property is defined.		
Rationale	It is useful to distinguish between properties of format (tiff, gif, html) and those that share the same genre.		
Data Constraint	controlled vocabulary (e.g. mime type)		
Category	object	component	property
Cardinality	0	0	0..n
Obligation	n/a	n/a	optional
Examples			image/tiff [pronom ID]
Notes	A relationship may be defined between format-specific properties and general properties that fulfil the same or similar function for other formats that share the same content type. This may be useful when converting technical properties between formats, e.g. TIFF to PNG.		

2 Agent

The Agent entity represents agents, who may be human agents, organisations, or software tools, involved in the lifecycle of digital objects. The Agent entity performs three functions in the significant properties data dictionary:

1. Indicates the properties that should be maintained in the long-term.
2. Specifies the preservation level that should be assigned to the property or component
3. Indicates the acceptable quality threshold for each property

Implementers are allowed to describe the Agent using the PREMIS, FOAF (Friend of a Friend), or other namespace that are considered suitable. For the most part, the sub-elements of the Agent entity duplicate those specified in the PREMIS Data Dictionary 2.0 (<http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>), with the addition of a 'role' element.

2.1. agentIdentifier

Semantic Unit	agentIdentifier
Semantic Component	agentIdentifierType agentIdentifierValue
Definition	The designation used to uniquely identify the agent within a preservation repository system.
Rationale	Each agent associated with the definition and description of significant properties in the format registry and preservation repository must be assigned a unique identifier to associate it with various decision making process.
Data Constraint	container
Category	n/a
Cardinality	1... n
Obligation	mandatory
Examples	n/a
Notes	Identifiers must be unique within the repository. The <i>agentIdentifier</i> is repeatable in order to allow both repository assigned and externally-assigned identifiers to be recorded.

2.1.1. agentIdentifierType

Semantic Unit	agentIdentifierType
Semantic Component	None
Definition	A designation of the domain within which the identifier is unique.
Rationale	The Identifier values for agents may be similar between format registries and digital repositories. The combination of identifierType and identifierValue should ensure some degree of uniqueness.
Data Constraint	A controlled vocabulary should be defined.
Category	n/a
Cardinality	1
Obligation	optional
Examples	Cerch Pronom
Notes	

2.1.2. agentIdentifierValue

Semantic Unit	agentIdentifierValue
Semantic Component	None
Definition	The identifier of an agent that is unique in the domain.
Rationale	The identifierValue of a Component or Property must be readily identified.
Data Constraint	A controlled vocabulary or guidance on appropriate layout of the element should be defined.
Category	n/a
Cardinality	1
Obligation	mandatory
Examples	43434020272
Notes	As indicated in the PREMIS Data Dictionary 2.0, agentIdentifiervalue may be a unique key or a controlled textual form of name.

2.2. agentName

Semantic Unit	agentName
Semantic Component	None
Definition	A text string which could be used in addition to <i>agentIdentifier</i> to identify an agent.
Rationale	A readable version of the machine-processable identifier specified in <i>agentIdentifier</i> .
Data Constraint	A controlled vocabulary or guidance on appropriate layout of the element should be defined.
Category	n/a
Cardinality	0...1
Obligation	optional
Examples	Surname, Forename [company title] [Product title] [version number]
Notes	As indicated in the PREMIS Data Dictionary 2.0, agentIdentifiervalue may be a unique key or a controlled textual form of name.

2.3. agentRole

Semantic Unit	agentRole
Semantic Component	None
Definition	The capacity in which the agent is operating.
Rationale	The reader should be able to understand the role that the agent is performing.
Data Constraint	A controlled vocabulary or guidance on appropriate layout of the element should be defined.
Category	n/a
Cardinality	0... 1
Obligation	optional
Examples	Format registry Creator Archiving institution
Notes	

3. Relations

An Object, Component, or Property may be related to one or more entities of each type. The data model in use will influence the type of relationships specified in a working system: an ‘atomic’ data model that specifies that each object should be self contained will possess Components and Properties that relate to that Object in isolation. Alternatively, for a ‘multiple manifestation’ model, such as that advocated for OAI-ORE, the Components and Properties may be associated with a Resource that has two or more ‘Views’. The InSPECT Project team has taken a similar approach to the PREMIS Working Group, recognizing three key relationship types that may be applied in various scenarios:

1. Structural: the relationships between the three entity types located within the Object. Structural relationships may be applied to five scenarios: relationship between an Object and one or more Components, relationship between a Component and other Components, and the relationship between a Component and one or more Properties.
2. Dependency: relationships between Objects, Components, or Properties that indicate a mutual or one-sided reliance on another entity. Dependency relationships are likely to occur between multiple properties that must be considered in combination to recreate an aspect of the Performance.
3. Derivation: the relationship between an Object, Component, or Property with another of the same type as a result of being a type of derivative. Derivation relationships are likely to be recorded between format-specific and genre-specific properties.

The method in which these relationships are to be expressed in the PRONOM Format Registry and the preservation metadata schema has yet to be finalised. For the purpose of this document, the authors have adopted the PREMIS Relationship entity as a basis for specifying the relationships that may be recorded between Objects, Components and Properties. These may be replaced by an appropriate RDF format when implemented.

3.1.relationship

Semantic Unit	relationship		
Semantic Component	relationshipType relationshipSubType relatedIdentification		
Definition	A container for relationship metadata		
Rationale	The information is useful for understanding the relationship between entities.		
Data Constraint	container		
Category	object	component	property
Cardinality	0...n	0...n	0...n
Obligation	optional	optional	optional
Examples			
Notes			

3.1.1. relationshipType

Semantic Unit	relationshipType
Semantic Component	n/a
Definition	A high-level categorisation of the nature of the relationships

Rationale	The information is useful for understanding the relationship between entities.		
Data Constraint	Controlled vocabulary		
Category	object	component	property
Cardinality	1	1	1
Obligation	Mandatory	Mandatory	Mandatory
Examples	structural dependency derivation	structural	structural dependency derivation
Notes	The definition is taken from the PREMIS Data Dictionary 2.0, p111.		

3.1.2. relationshipSubType

Semantic Unit	relationshipSubType		
Semantic Component	n/a		
Definition	A specific characterization of the nature of the relationship documented in relationshipType.		
Rationale	The information is useful for understanding the relationship between entities.		
Data Constraint	Controlled vocabulary		
Category	object	component	property
Cardinality	1	1	1
Obligation	Mandatory	Mandatory	Mandatory
Examples	isPartOf hasPart	isPartOf hasPart hasSiblings	isPartOf hasPart
Notes	The definition is taken from the PREMIS Data Dictionary 2.0, p112.		

3.1.3. relatedIdentification

Semantic Unit	relatedIdentification		
Semantic Component	relatedIdentifierType relatedIdentifierValue		
Definition	The identifier and sequential context of the related resource.		
Rationale			
Data Constraint	container		
Category	object	component	property
Cardinality	1...n	1...n	1...n
Obligation	mandatory	mandatory	mandatory
Examples			
Notes	The related object may or may not be held within the preservation repository. Recommended practice is that objects reside within the repository unless there is a good reason to reference an object outside. Internal and external references should be clear. The definition is taken from the PREMIS Data Dictionary 2.0, p113.		

3.1.3.1 relatedIdentifierType

Semantic Unit	relatedObjectIdentifierType		
Semantic Component	n/a		
Definition	A designation of the domain within which the identifier is unique.		
Rationale	Identifiers may be used in multiple domains, which may cause confusion or		

	errors to occur.		
Data Constraint	controlled vocabulary		
Category	object	component	property
Cardinality	1...n	1...n	1...n
Obligation	Mandatory	Mandatory	Mandatory
Examples	See identifierType	See identifierType	See identifierType
Notes	If the related object is held within the preservation repository, this should be the value of the identifierType The definition is taken from the PREMIS Data Dictionary 2.0, p114.		

3.1.3.2 relatedIdentifierValue

Semantic Unit	relatedIdentifierValue		
Semantic Component	n/a		
Definition	The value of the related identifier		
Rationale	Indicates Objects, Components or Properties that have some relationship to the entity.		
Data Constraint	controlled vocabulary		
Category	object	component	property
Cardinality	1...n	1...n	1...n
Obligation	Mandatory	Mandatory	Mandatory
Examples	See PREMIS:ObjectIdentifierValue	See identifierValue	See identifierValue
Notes	If the related object is held within the preservation repository, this should be the value of the identifierValue or objectIdentifierValue. A Component may refer to the objectIdentifierValue of a file if it is embedded in the Component. E.g. an image. The definition is taken from the PREMIS Data Dictionary 2.0, p115.		

ⁱ Anon (2007-09-10). PLANETS Project. Retrieved on February 4, 2008 from: <http://www.planets-project.eu/>