TECHNICAL REPORT

ISO/IEC TR 21000-1

Second edition
2004-11-01

Information technology — Multimedia framework (MPEG-21) —

Part 1: Vision, Technologies and Strategy

Technologies de l'information — Cadre multimédia (MPEG-21) —

Partie 1: Vision, technologies et stratégie

Reference number
ISO/IEC TR 21000-1:2004(E)

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

Page

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of the joint technical committee is to prepare International Standards. Draft International Standards adopted by the joint technical committee are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75 % of the national bodies casting a vote.

In exceptional circumstances, the joint technical committee may propose the publication of a Technical Report of one of the following types:

— type 1, when the required support cannot be obtained for the publication of an International Standard, despite repeated efforts;

— type 2, when the subject is still under technical development or where for any other reason there is the future but not immediate possibility of an agreement on an International Standard;

— type 3, when the joint technical committee has collected data of a different kind from that which is normally published as an International Standard ("state of the art", for example).

Technical Reports of types 1 and 2 are subject to review within three years of publication, to decide whether they can be transformed into International Standards. Technical Reports of type 3 do not necessarily have to be reviewed until the data they provide are considered to be no longer valid or useful.

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights.

ISO/IEC TR 21000-1, which is a Technical Report of type 3, was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 29, *Coding of audio, picture, multimedia and hypermedia information*.

This second edition cancels and replaces the first edition (ISO/IEC TR 21000-1:2001), which has been technically revised.

ISO/IEC 21000 consists of the following parts, under the general title *Information technology — Multimedia framework (MPEG-21)*:

— *Part 1: Vision, Technologies and Strategy [TR]*

— *Part 2: Digital Item Declaration*

— *Part 3: Digital Item Identification*

— *Part 5: Rights Expression Language*

— *Part 6: Rights Data Dictionary*

— *Part 7: Digital Item Adaptation*

— *Part 8: Reference Software*

— *Part 9: File Format*

— *Part 11: Evaluation Tools for Persistent Association Technologies [TR]*

The following part is under preparation:

— *Part 10: Digital Item Processing*

Executive Summary

Currently, multimedia technology provides content creators and consumers with a myriad of coding, access and distribution possibilities. At the same time, communication infrastructure is being put into place to enable access to information and multimedia services from almost anywhere at anytime. Still, no global end-to-end solutions exist allowing all different user communities to interact in an interoperable way. This lack of interoperable (and thus standardized) solutions is stalling the deployment of advanced multimedia packaging and distribution applications although most of the individual technologies are indeed already present.

This motivated MPEG (ISO/IEC JTC1 SC29 WG11) in June 2000 to start to work on the definition of enabling normative technology for the multimedia applications of the 21st century: MPEG-21 "Multimedia Framework". MPEG-21's approach is to define a framework to support transactions that are interoperable and highly automated, specifically taking into account Intellectual Property Management and Protection requirements and targeting multimedia access and delivery using heterogeneous networks and terminals. The significant progress that was made in developing MPEG-21 since the publication of first edition of this Technical Report in 2001 (ISO/IEC 21000-1:2001) has led to this updated version.

Based on the above observations, MPEG-21 aims at defining a normative albeit open framework for multimedia creation and sharing for use by all the players in the delivery and consumption chain. This open framework will provide content creators and service providers with equal opportunities in the MPEG-21 enabled open market. This will also be to the benefit of the content consumer providing them access to a large variety of content in an interoperable manner.

The MPEG-21 vision can thus be summarized as follows: *to define a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices used by different communities*.

MPEG-21 identifies and defines the normative technologies needed to support the multimedia delivery chain as described above as well as the relationships between and the operations supported by them. Within the parts of MPEG-21, these elements are elaborated by defining the syntax and semantics of their characteristics, such as interfaces to the elements.

Part 1 of MPEG-21 (ISO/IEC 21000-1) provides:

a) A *vision* for a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices to meet the needs of all Users[1)].

b) A method to facilitate the integration of components and standards in order to harmonise *technologies* for the creation, management, manipulation, transport, distribution and consumption of content.

c) A *strategy* for achieving a multimedia framework by the development of specifications and standards based on well-defined functional requirement through collaboration with other bodies.

1) A User is any entity that interacts in the MPEG-21 environment or makes use of a Digital Item (all capitalised terms are used as defined in Clause 2).

Introduction

End users' appetite for content and the accessibility of information is increasing at an incredible pace. Access devices, with a myriad set of different terminal and network capabilities, are making their way into end users' lives. Additionally, these access devices are used in different locations and environments. Their users, however, are currently not given tools to deal efficiently with all the intricacies of this new multimedia usage environment.

Enabling "ease of use" is becoming increasingly important as individuals are producing more and more digital media for their personal use and for sharing among family and friends (as is evidenced by the large number of amateur music, photo and media sharing web sites). These amateur "content providers" have many of the same concerns as commercial content providers (management of content, re-purposing of content based on consumer/device capabilities, protection of rights, protection from unauthorised access/modification, protection of privacy of providers and consumers, etc.).

Such developments provide new models for distributing and trading digital content electronically in addition to existing business models for trading physical goods. Such new business models mean that the boundaries between the delivery of audio sound (music and spoken word), accompanying artwork (graphics), text (lyrics), video (visual) and synthetic spaces will become increasingly blurred. Indeed, it is becoming increasingly difficult to identify the different intellectual property rights that are associated with multimedia content. New solutions are required to manage the access and delivery process of these different content types in an integrated and harmonized way, entirely transparent to the user of multimedia services.

This motivates the MPEG-21 Multimedia Framework initiative that aims to enable transparent and augmented use of multimedia resources across a wide range of networks and devices, specifically taking into account Intellectual Property Management and Protection and the heterogeneity of the access and delivery infrastructure.

Information technology — Multimedia framework (MPEG-21) —

Part 1:
Vision, Technologies and Strategy

1 Scope

This Technical Report has been prepared within ISO/IEC JTC 1/SC 29/WG 11 to reflect the progress made on the definition of the MPEG-21 Multimedia Framework. This progress is a result of the combination of WG 11's efforts to standardise the parts of the multimedia framework where it has the appropriate expertise, and the integration with standards initiatives which are being developed by other bodies. The intention of this collaborative approach is to maximise harmonisation of efforts and enable effective standards solutions to be implemented in the shortest possible time.

The Technical Report is introduced by a problem statement and a solution statement. The problem statement describes a multimedia usage environment founded upon ubiquitous networks that is encouraging new business models for trading digital content. In this environment, the distinction between content types is less clear as their integration as multimedia resources in new products and services makes the traditional boundaries less distinct. In addition, individuals are becoming increasingly aware of the value, both commercial and intrinsic, of their own digital asset resources and new possibilities presented by the tools which enable them to create and collect, package and distribute content. The solution statement introduces the vision of the Multimedia Framework to support transactions that are interoperable and highly automated, which is required to support these new types of commerce.

This MPEG-21 Multimedia Framework is based on two essential concepts: the definition of a fundamental unit of distribution and transaction (the Digital Item) and the concept of Users interacting with Digital Items. The Digital Items can be considered the "what" of the Multimedia Framework (e.g., a video collection, a music album) and the Users can be considered the "who" of the Multimedia Framework.

The goal of MPEG-21 can be phrased as: defining the technology needed to support Users to exchange, access, consume, trade and otherwise manipulate Digital Items in an efficient, transparent and interoperable way. This Technical Report gives an overview of the technologies that have been identified to enable this goal (and that are consequently being reflected into the different parts of the MPEG-21 standard).

In creating its definition of a multimedia framework and in making its proposals and recommendations for further standardisation, MPEG-21 is taking into account other related multimedia activities. The Technical Report identifies other multimedia initiatives that are currently in progress that are considered as candidates for future interaction and collaboration with the standards work plan agreed by MPEG-21 (see also Annex A).

2 Terms and definitions

For the purposes of this Technical Report, the following terms and definitions apply:

2.1
Abstraction
Distinct intellectual or artistic creation or concept.

2.2
Asset
Manifestation, i.e. a physical or digital embodiment of an Expression.

2.3
Digital Item
Structured digital object with a standard representation, identification and metadata within the MPEG-21 framework. [ISO/IEC 21000-2]

NOTE: This entity is also the fundamental unit of distribution and transaction within this framework.

2.4
End User
User taking the role of consumer, i.e. being at the end of a value or delivery chain

EXAMPLE: a human consumer, an agent operating on behalf of a human consumer, etc.

NOTE: "User" refers to all participants in the value or delivery chain.

2.5
Expression
Intellectual or artistic realisation of an Abstraction.

2.6
Manifestation
The physical or digital embodiment of an Expression.

2.7
Peer
Device or application that compliantly processes a Digital Item[2)].

2.8
Resource
Individually identifiable Asset such as a video or audio clip, an image, or a textual Asset. [ISO/IEC 21000-2]

NOTE: A Resource may also potentially be a physical object.

2.9
User
Any entity that interacts in the MPEG-21 environment or makes use of Digital Items.

EXAMPLE: Creator, rights holders, distributors and consumers of Digital Items.

3 Symbols and abbreviated terms

List of symbols and abbreviated terms

3.1
API
Application Program Interface

2) The term "Terminal" is usually avoided within the MPEG-21 documents because of its connotation as being the end point in a chain of communication. Besides such applications, the term Peer explicitly also includes devices or applications that create or alter Digital Items, and that handle Digital Items "in transit". How compliance will be achieved is currently being discussed.

3.2
CATV
Community Aerial Television

3.3
CD
Compact Disc

3.4
DI
Digital Item

3.5
DIA
Digital Item Adaptation

3.6
DIBO
Digital Item Base Operations

3.7
DID
Digital Item Declaration

3.8
DII
Digital Item Identification

3.9
DIP
Digital Item Processing

3.10
DIME
Digital Item Method Engine

3.11
DMIF
Multimedia Integration Framework

3.12
DVB
Digital Video Broadcasting

3.13
ECMA
European Computer Manufacturer Association

3.14
HTML
Hypertext Mark-up Language

3.15
ID
IDentifier

3.16
IEEE
Institute of Electrical and Electronic Engineers

3.17
IPMP
Intellectual Property Management and Protection

3.18
ITU
International Telecommunication Union

3.19
JPEG
Joint Photographic Experts Group

3.20
MHP
Multimedia Home Platform

3.21
LAN
Local Area Network

3.22
MIDI
Musical Industry Digital Interface

3.23
MPEG
Motion Picture Experts Group

3.24
MSF
Multiservice Switching Forum

3.25
QoS
Quality of Service

3.26
RDD
Rights Data Dictionary

3.27
REL
Rights Expression Language

3.28
SNR
Signal to Noise Ratio

3.29
TR
Technical Report

3.30
TV
TeleVision

3.31
XML
eXtensible Mark-up Language

4 Structure of the Technical Report

The Technical Report is introduced by a problem statement and a solution statement. The problem statement describes a multimedia usage environment founded upon ubiquitous networks that is encouraging new business models for trading digital content. The solution statement introduces the vision of the Multimedia Framework to support transactions that are interoperable and highly automated, which is required to support these new types of commerce.

Then, the two essential concepts of the Multimedia Framework are introduced: the definition of a fundamental unit of distribution and transaction (the Digital Item) and the concept of Users interacting with Digital Items.

Clause 6 provides some brief background on the first edition of this document (ISO/IEC 21000-1:2001). In that document, seven architectural key elements needed to support MPEG-21's goals have been identified. These architectural key elements eventually lead to the formulation of requirements for the different MPEG-21 parts.

Following the publication of first edition of this Technical Report in 2001 (ISO/IEC 21000-1:2001), various Calls for Proposals based upon requirements have been and are being issued by MPEG. Eventually the responses to the calls result in different parts of the MPEG-21 standard (i.e. ISO/IEC 21000-N) after intensive discussion and harmonization efforts. Clause 7 provides a high-level overview of all MPEG-21 parts defined so far.

Clause 8, "MPEG-21 Achievements" lists those MPEG-21 parts that have already reached Final Draft International Standard (FDIS) or Technical Report (TR) status, obviously excluding this document. For each part, the goal and the rationale are listed. Consequently, the resulting key concepts and basic approach are presented. Finally, Clause 9 highlights those MPEG-21 parts that are currently well advanced in MPEG-21.

5 Overview

5.1 Problem Statement

End users' appetite for content and the accessibility of information is increasing at an incredible pace. Access devices, with a myriad set of differing terminal and network capabilities, are making their way into end users' lives. Additionally, these access devices are used in different locations and environments. Their users, however, are currently not given tools to deal efficiently with all the intricacies of this new multimedia usage environment.

Enabling "ease of use" is becoming increasingly important as individuals are producing more and more digital media for their personal use and for sharing among family and friends (as is evidenced by the large number of amateur music, photo and media sharing web sites). These "content providers" have many of the same concerns as commercial content providers (management of content, re-purposing content based on consumer/device capabilities, protection of rights, protection from unauthorised access/modification, protection of privacy of providers and consumers, etc.).

Such developments rewrite existing business models for trading physical goods with new models for distributing and trading digital content electronically. Indeed, it is becoming increasingly difficult to identify the different intellectual property rights that are associated with multimedia content. The boundaries between the delivery of audio sound (music and spoken word), accompanying artwork (graphics), text (lyrics), video (visual) and synthetic spaces will become increasingly blurred. New solutions are consequently required to manage the access and delivery process of these different content types in an integrated and harmonized way, entirely transparent to the user of multimedia services.

5.2 Solution Statement

Normative technology is required to enable interoperable new multimedia usage cases. MPEG's approach is to define a Multimedia Framework to ensure that the systems that deliver multimedia content are *interoperable* and that transactions are simplified and, if possible, *automated*. This should apply to the infrastructure

requirements for content delivery, content security, rights management, secure payment, and to the technologies enabling them – and this list is not exhaustive.

The scope of MPEG-21 could therefore be described as the integration of the critical technologies enabling transparent and augmented use of multimedia resources across a wide range of networks and devices to support functions such as: content creation, content production, content distribution, content consumption and usage, content packaging, intellectual property management and protection, content identification and description, financial management, user privacy, terminals and network resource abstraction, content representation and event reporting.

From its background in key technology and information management standards related to the management, delivery and representation of multimedia content, MPEG was well positioned to initiate such an activity. However, MPEG has recognised from the very start of MPEG-21 that the integration of disparate multimedia technologies can only be achieved by working in collaboration with other bodies.

5.3 Normative Implications

The Multimedia Framework is being developed through a combination of MPEG's efforts to standardise the parts of the Multimedia Framework where it has the appropriate expertise, and the integration with other multimedia initiatives which have been or are being developed by other bodies. MPEG hence contributes to the definition of the Multimedia Framework by developing new standards or by developing interfaces for other existing or future standards and services to provide the required interoperability or architectural elements.

MPEG-21's normative recommendations will be determined by interoperability requirements, and their level of detail may vary for each framework element. The actual instantiation and implementation of the framework elements below the abstraction level required to achieve interoperability, will not be specified.

MPEG-21 also provides several non-normative Technical Reports providing information on how to best use MPEG-21 technology.

5.4 Conformance and Reference Software

Conformance is an essential element of each MPEG standard. However, within the scope of this Technical Report, no conformance criteria are given. Instead, Part 14 of MPEG-21 (ISO/IEC 21000-14) will contain such criteria; they are currently under development.

A similarly important issue, reference software, is also not addressed in this Technical Report. It will be covered in ISO/IEC 21000-8, though.

5.5 Description of a Multimedia Framework Architecture

5.5.1 Introduction

To define where standards are required in a Multimedia Framework which is capable of supporting the delivery of digital content, it is necessary first to reach a shared understanding about common concepts. This presents a difficulty, as there are many examples of different architectures that evolve in response to a variety of models for the use of content. In order to avoid giving undue preference to one model above another, it is proposed to describe the multimedia framework as a generic architecture of conceptual design. Such a broad and high-level approach will allow for more specific use cases to be elaborated, which can be mapped back against the generic architecture as the work continues.

The intent is to maintain an MPEG-21 Use Case Scenario document in conjunction with the Technical Report to provide examples of potential MPEG-21 applications.

MPEG-21 is based on two essential concepts: the definition of a fundamental unit of distribution and transaction (the Digital Item) and the concept of Users interacting with Digital Items. The Digital Items can be considered the "what" of the Multimedia Framework (e.g., a video collection, a music album) and the Users can be considered the "who" of the Multimedia Framework.

The goal of MPEG-21 can thus be rephrased to: defining the technology needed to support Users to exchange, access, consume, trade and otherwise manipulate Digital Items in an efficient, transparent and interoperable way.

During the MPEG-21 standardization process, Calls for Proposals based upon requirements have been and are being issued by MPEG. The responses to the calls lead to the development of different parts of the MPEG-21 standard (i.e. ISO/IEC 21000-*n*).

5.5.2 Digital Items

A Digital Item is a structured digital object with a standard representation, identification and metadata within the MPEG-21 framework. This entity is the fundamental unit of distribution and transaction within this framework.

MPEG-21 describes a set of abstract terms and concepts to form a useful model for defining Digital Items (Digital Item Declaration). Within this model, a Digital Item is the digital representation of an Asset, and as such, it is the unit that is acted upon (managed, described, exchanged, collected, etc.) within the model. The goal of this model is to be as flexible and general as possible, while providing for the "hooks" that enable higher-level functionality. This, in turn, allows the model to serve as a key foundation in the building of higher-level models in other MPEG-21 elements.

In practice, a Digital Item is a combination of resources, metadata, and structure. The resources are the individual Assets or (distributed) content. The metadata comprises informational data about or pertaining to the Digital Item as a whole or to the individual Resources included in the Digital Item. Finally, the structure relates to the relationships among the parts of the Digital Item, both resources and metadata.

An example of a Digital Item may be a music compilation including the music but also photos, videos, animation graphics, lyrics, scores, MIDI files, interviews with the singers, news related to the songs, statements by an opinion maker, ratings of an agency, position in the hit list, navigational information driven by user preferences, bargains, etc.

5.5.3 Users

In MPEG-21 a User (with capitalized "U") is any entity that interacts in the MPEG-21 environment or makes use of Digital Items. Such Users e.g., include individuals, consumers, communities, organizations, corporations, consortia, and governments. Users are identified specifically by their relationship to another User for a certain interaction. From a purely technical perspective, MPEG-21 makes no distinction between a "content provider" and a "consumer"—both are Users. A single entity may use content in many ways (publish, deliver, consume, etc.), and so all parties interacting within MPEG-21 are categorized as Users equally. However, a User may assume specific rights and responsibilities according to their interaction with other Users within MPEG-21.

At its most basic level, MPEG-21 can be seen as providing a framework in which one User interacts with another User and the object of that interaction is a Digital Item. Some such interactions are creating content, providing content, modifying content, archiving content, rating content, enhancing and delivering content, aggregating content, delivering content, syndicating content, retail selling of content, consuming content, subscribing to content, regulating content, facilitating transactions that occur from any of the above, and regulating transactions that occur from any of the above. Any of these are "uses" of MPEG-21, and the parties involved are Users.

5.6 Example MPEG-21 Use Case

5.6.1 Introduction

To illustrate the innovative uses MPEG-21 enables, an example use case is described here. For the sake of clarity, it is not intended to cover all MPEG-21 parts and concepts.

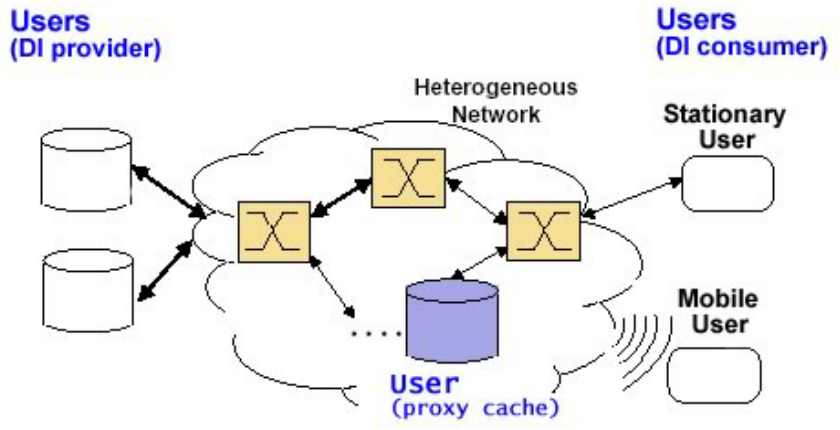

Figure 1 — Example Distributed Multimedia System

Consider a distributed multimedia system comprising a certain number of Users exchanging Digital Items across a wide range of networks such as the Internet, mobile phone connections, etc. to a variety of terminals (Figure 1).

Rights holders and their authorised representatives will want to set the usage conditions for the different Digital Items. These can be simple – the Resources of this Digital Item can be used – or it may contain complex conditions – the resources of this Digital Item can be used provided that certain conditions are met, which may, for instance, include temporal, spatial or group membership provisions. Complex Rights Expressions of this nature will enable rights holders to develop a multiplicity of business models.

Additionally, in order to guarantee a smooth delivery of such Digital Items over heterogeneous networks "adaptations" of Resources may be required, for example to overcome network congestion or to allow Digital Items to be routed at the same time to Users connected via mobile devices and to Users on fixed lines. A large number of adaptation possibilities exist, e.g., videos can be adapted by simple frame dropping or by modifying the quantization coefficients. Such adaptations can happen anywhere in the delivery chain from the Digital Item provider to the Digital Item consumer and can be governed by Rights Expressions.

Assume that a mobile User requests a Digital Item containing the newest movie trailers that are available without payment from another User providing a repository of Digital Items. In addition, the mobile User only accepts videos with a bitrate lower than 500 kbps. On its way to the Digital Item repository, the request for the trailer passes another User that has the functionality of a proxy cache and, fortunately, the requested resource is cached there. Assume, however, that the bandwidth requirement of the cached version exceeds the resource constraints.

Based on watching the movie trailers, the User decides to watch a movie at high resolution in a stationary environment. This high-resolution version is not free and cannot be redistributed to other Users.

5.6.2 Enabling MPEG-21 Technologies

This subclause highlights MPEG-21 technologies that enable the above use case. For the sake of clarity, only a subset of these technologies are listed.

5.6.2.1 Digital Item Declaration

Part 2 of MPEG-21 (ISO/IEC 21000-2) defines the structure of the Digital Item. These Digital Items comprise:

— The Digital Item Declaration (DID); and

— The Resources.

The DID is an XML file describing the Digital Item whereas the Resources are the individually identifiable multimedia Assets of the Digital Item (DI). The DID file may include information such as unique identifiers for the complete DI as well as for Resources, expressions on rights and permissions pertaining to the DI (or parts thereof) and generic metadata describing the Digital Item and its Resources. Finally, the DID contains references to the Resources[3]. Typical examples of Resources include AAC audio files, MPEG-2 video clips, JPEG images, MPEG-4 presentations, HTML pages – but also e.g., video clips or text in proprietary formats.

5.6.2.2 Rights Expression Language/Rights Data Dictionary

MPEG-21 Part 5 provides a machine-readable Rights Expression Language (REL) that can declare rights and permissions using the terms as defined in the Rights Data Dictionary (RDD, MPEG-21 Part 6). The provider of the movie will use the REL to express under what conditions which User shall be able to access the item.

After paying for the high-resolution movie, the User will obtain an MPEG-21 REL License (e.g., in a second DI) containing the conditions under which the User can view the movie. This License contains a grant with four parts:

— keyHolder: identifies the User that is being granted the rights;

— Play: identifies the permission that the User is being granted; The RDD provides the semantic definition of what this permission includes (in the example, the User is only allowed to render the video once);

— diReference: identifies the Digital Item, e.g., urn:grid:a1-abcde-1234567890-f, that the User is being granted rights to (in accordance with MPEG-21 Part 3, Digital Item Identification[4]); and

— validityInterval: identifies the time interval during which the User is allowed to play the Digital Item.

The task of enforcing these rights is the job of the IPMP system (MPEG-21 Part 4).

5.6.2.3 Resource Adaptation

MPEG-21 Part 7 Digital Item Adaptation (DIA) provides the means to steer Resource adaptation, e.g., in the proxy cache, and to carry it out in a more efficient way. The first tool to this end is the DIA AdaptationQoS Descriptor, which specifies the relationship between resource constraints, feasible operations on the resource and associated resulting qualities. The second descriptor is the DIA Bitstream Syntax Description (BSD) which describes the high-level structure of the media bit stream and which is appropriate for the adaptation of media resources on devices with constrained computational or memory resources. The advantage of using a BSD for adaptation is that it has a considerably smaller size than the media described and that adaptation engines need not be aware of the specific coding format of the movie. Rather, an adaptation module may rely on the BSD to modify the description/structure of the bit stream (by “editing” the BSD) and to further generate the adapted media data from the adapted BSD. This results in a more lightweight, coding format-independent way of adapting media streams, as compared e.g., to a sophisticated transcoding technique. Additional MPEG-21 tools provide so-called Usage Environment Descriptors which specify terminal and network characteristics and User preferences, among others.

Let us assume that the Digital Item cached in the proxy cache has, beside the resource, MPEG-7 descriptors available on the media format and media profile, also the MPEG-21 DIA components AdaptationQoS and BSD available. The proxy cache has to determine the bandwidth requirements of the version cached. This may be obtained from MPEG-7 descriptors in the Digital Item. The User constraint is then matched against the bandwidth requirement. If an adaptation is required, the AdaptationQoS associated to this video is consulted and the appropriate adaptation operation is extracted, together with a link to the adaptation process, which acts on the BSD description available for the video.

3) It is however also possible to have Resources contained in the DID itself.

4) The Digital Item Identification specification allows established industry identifiers (e.g. International Standard Book Numbers (ISBN), International Standard Recording Codes (ISRC), etc) to be used in the context of MPEG-21.

5.7 Collaboration with other multimedia standardisation initiatives

In creating its definition of a multimedia framework and making its proposals and recommendations for further standardisation, MPEG continuously takes other related multimedia activities into account. MPEG has hence sought collaboration (e.g., through liaisons) with relevant initiatives to expedite the work.

This Technical Report identifies some other relevant multimedia initiatives in the context of MPEG-21: a non-exhaustive list is given in Annex A. During its previous standards developments, MPEG has always recognised the importance of establishing liaisons with other bodies and organisations with which it shares complementary or common objectives. These liaisons have provided a useful channel for communicating between the parties to ensure that any overlap between concurrent standards activities is minimised and that, where necessary, common technology can be shared.

The broad scope of the task of defining a multimedia framework presents new challenges and opportunities for collaboration between those initiating standards activities in this area. The value of an integrated framework for the management and delivery of multimedia content is considerable and is attracting the interest and enthusiasm of major standards bodies.

Within this Technical Report, MPEG is describing a vision of a multimedia framework, pinpointing the components of the framework which require standardisation. However, it makes no assumption that MPEG itself will undertake the task of actually standardising all of the identified components. Rather, MPEG tries to co-ordinate its work with other standards bodies to ensure that it can concentrate on those areas which are best suited to and compatible with its mandate. As a result, a high level of practical collaboration with other standards bodies has taken place in order to complete some standardisation tasks successfully.

6 About ISO/IEC 21000-1:2001

From the very start, MPEG-21 has opted for a User centric approach. The first edition of this document (ISO/IEC 21000-1:2001) identifies and defines the seven architectural key elements needed to support MPEG-21's goals. Although it was already foreseen that eventually these elements would be elaborated into one or more MPEG-21 parts, this partitioning into seven architectural key elements proved to be an efficient working method to focus efforts when defining and refining the MPEG-21 requirements.

The seven key elements that have been defined in ISO/IEC 21000-1:2001 are:

a) Digital Item Declaration (a uniform and flexible abstraction and interoperable schema for declaring Digital Items);

b) Digital Item Identification and Description (a framework for identification and description of any entity regardless of its nature, type or granularity);

c) Content Handling and Usage (provide interfaces and protocols that enable creation, manipulation, search, access, storage, delivery, and (re)use of content across the content distribution and consumption value chain);

d) Intellectual Property Management and Protection (the means to enable Digital Items and their rights to be persistently and reliably managed and protected across a wide range of networks and devices);

e) Terminals and Networks (the ability to provide interoperable and transparent access to content across networks and terminals);

f) Content Representation (how the media resources are represented);

g) Event Reporting (the metrics and interfaces that enable Users to understand precisely the performance of all reportable events within the framework).

As an example, Digital Item Adaptation (MPEG-21 part 7) addresses requirements initially formulated in the Terminals and Networks architectural key element.

7 Overview of the MPEG-21 Parts

7.1 Introduction

Following the publication of first edition of this Technical Report in 2001 (ISO/IEC 21000-1:2001), various Calls for Proposals based upon requirements have been and are being issued by MPEG. Eventually the responses to the calls result in different parts of the MPEG-21 standard (i.e. ISO/IEC 21000-N) after intensive discussion and harmonization efforts. This new edition aims to reflect the significant progress that was made in developing MPEG-21 since the publication of first edition of this Technical Report (ISO/IEC 21000-1:2001).

An advantage to develop the MPEG-21 standard in several rather independent parts is to allow the various pieces of technology to be useful as stand-alones and thus used as much as possible, even if in conjunction with proprietary technologies. This has been the case, for example, for MPEG-2 Video, which today is being used together with MPEG-2 Systems but not with MPEG-2 Audio in the context of the U.S. digital TV system. However, although the various parts may be used independently, they were developed to give optimal results when they are used together; this principle is particularly important for MPEG-21 since an integrated multimedia framework is the purpose of this standard.

This clause gives a high-level overview of the different MPEG-21 parts that have already been developed or are currently under development. Those most advanced in the standardisation process are further elaborated in Clause 8 and 9.

7.2 Part 1 – Vision, Technologies and Strategy

Part 1 is this document. The title “Vision, Technologies and Strategy” has been chosen to reflect the fundamental purpose of the Technical Report. This is to:

— Define a 'vision' for a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices to meet the needs of all users

— Achieve the integration of components and standards to facilitate harmonisation of 'technologies' for the creation, modification, management, transport, manipulation, distribution, and consumption of Digital Items.

— Define a 'strategy' for achieving a multimedia framework by the development of specifications and standards based on well-defined functional requirements through collaboration with other bodies.

The first edition of this TR is ISO/IEC 21000-1:2001. This document is the second edition.

7.3 Part 2 – Digital Item Declaration (DID)

The purpose of the Digital Item Declaration (DID) specification is to describe a set of abstract terms and concepts to form a useful model for defining Digital Items. Within this model, a Digital Item is the digital representation of an Asset, and as such, it is the thing that is acted upon (managed, described, exchanged, collected, etc.) within the model. The goal of this model is to be as flexible and general as possible, while providing for the “hooks” that enable higher level functionality. This, in turn, will allow the model to serve as a key foundation in the building of higher level models in other MPEG-21 elements (such as Identification & Description or IPMP). This model specifically does not define a language in and of itself. Instead, the model helps to provide a common set of abstract concepts and terms that can be used to define such a scheme, or to perform mappings between existing schemes capable of Digital Item Declaration, for comparison purposes.

The DID technology is described in three normative sections:

— Model: The Digital Item Declaration Model describes a set of abstract terms and concepts to form a useful model for defining Digital Items. Within this model, a Digital Item is the digital representation of an Asset, and as such, it is the thing that is acted upon (managed, described, exchanged, collected, etc.) within the model.

— Representation: Normative description of the syntax and semantics of each of the Digital Item Declaration elements, as represented in XML. This section also contains some non-normative examples for illustrative purposes.

— Schema:Normative XML schema comprising the entire grammar of the Digital Item Declaration representation in XML.

7.4 Part 3 – Digital Item Identification (DII)

ISO/IEC 21000-3 comprises three parts:

— The possibility to uniquely identify Digital Items and parts thereof (such as resources) as well as other entities and relationships, such as Abstractions associated with Digital Items. DII provides a schema that can be used to include identifiers into a Digital Item Declaration.

— A method how to uniquely identify Description Schemes (using XML namespaces). This element enables the association of well-defined descriptions (so-called metadata) with Digital Items and their parts.

— A method to identify the type of a Digital Item.

It should be noted, however, that the DII specification does not specify new identification systems itself. For example, ISO/IEC 21000-3 does not attempt to replace the ISRC (as defined in ISO 3901) for sound recordings but allows ISRCs to be used within MPEG-21. In fact, the DII specification allows for existing and new identification systems to be ingenerated into the MPEG-21 domain through the mechanism of a Registration Authority.

7.5 Part 4 – Intellectual Property Management and Protection (IPMP)

This part of MPEG-21 will define an interoperable framework for Intellectual Property Management and Protection (IPMP). Fairly soon after MPEG-4, with its IPMP hooks, became an International Standard, concerns were voiced within MPEG that many similar devices and players might be built by different manufacturers, all MPEG-4, but many of them not working together. This is why MPEG decided to start a new project on more interoperable IPMP systems and tools. The project includes standardized ways of retrieving IPMP tools from remote locations, exchanging messages between IPMP tools and between these tools and the terminal. It also addresses authentication of IPMP tools, and has provisions for integrating Rights Expressions according to the Rights Data Dictionary and the Rights Expression Language.

Efforts are currently ongoing to define the requirements for the management and protection of intellectual property in the various parts of the MPEG-21 standard currently under development.

7.6 Part 5 – Rights Expression Language (REL)

A Rights Expression Language is seen as a machine-readable language that can declare rights and permissions using the terms as defined in the Rights Data Dictionary.

The REL is intended to provide flexible, interoperable mechanisms to support transparent and augmented use of digital resources in publishing, distributing, and consuming of digital movies, digital music, electronic books, broadcasting, interactive games, computer software and other creations in digital form, in a way that protects digital content and honours the rights, conditions, and fees specified for digital contents. It is also intended to support specification of access and use controls for digital content in cases where financial exchange is not part of the terms of use, and to support exchange of sensitive or private digital content.

Each REL Grant contains four elements;

a) Permission - this articulates what usage of the Digital Item one user is providing to another.

b) Condition - this articulate the constraint that the first user places on the second as a result of the use of their Digital Item. Example: you can play my movie, in return for a payment of $1. This statement includes the Permission "You can play my movie" and the condition "in return for $1".

c) Principle - this identifies exactly the User to who the Permission is being granted. i.e. "User XXX (You) can play my movie".

d) Resource - this identifies exactly the Digital Item (or part thereof) for which a Permission is being granted. E.g., "User XXX can play Item YYY (my movie)".

From this simple model of Grant = Permission + Condition + Principle + Resource very flexible Rights Expressions can be generated.

The Rights Expression Language is intended to provide a flexible interoperable mechanism to ensure personal data is processed in accordance with individual rights and to meet the requirement for Users to be able to express their rights and interests in a way that addresses issues of privacy and use of personal data.

The standard Rights Expression Language is able to support guaranteed end-to-end interoperability, consistency and reliability between different systems and services. To do so, it must offer richness and extensibility in declaring rights, conditions and obligations, ease and persistence in identifying and associating these with digital contents, and flexibility in supporting multiple usage/business models.

MPEG has developed its REL to meet the requirements it defined. MPEG recognizes that several industries and User communities will need to modify the language to better meet their specific needs. To facilitate easy mapping of the REL to these specific applications MPEG has developed a process of Extension and Profiling of the Language.

The Extension process allows individuals to define elements of the language specific to their needs. This includes development of new verbs and schematic elements to improve efficiency in their specific domain. MPEG has also developed the Rights Data Dictionary to ensure the semantic interpretation of new verbs is unambiguously understood to promote interoperability.

The Profile process allows individuals to select only the parts of the language applicable to their application. This optimises payload of Digital Items and computation requirements of MPEG terminals. MPEG recognizes that different applications require different levels of complexity and flexibility in the REL. The Profiling process allows one community to select only the elements they feel they need to meet a specific application need.

It is important to note that an Extension and Profile can be used concurrently to optimise the applicability of the REL to one specific application.

7.7 Part 6 – Rights Data Dictionary (RDD)

The Rights Data Dictionary (RDD) comprises a set of clear, consistent, structured, integrated and uniquely identified Terms to support the MPEG-21 Rights Expression Language.

The structure of the dictionary is specified, along with a methodology for creating the dictionary. The means by which further Terms may be defined is also explained.

The Dictionary is a prescriptive Dictionary, in the sense that it defines a single meaning for a Term represented by a particular RDD name (or Headword), but it is also inclusive in that it recognizes the prescription of other Headwords and definitions by other Authorities and incorporates them through mappings. The RDD also supports the circumstance that the same name may have different meanings under different Authorities. The RDD specification has audit provisions so that additions, amendments and deletions to Terms and their attributes can be tracked.

RDD recognises legal definitions as and only as Terms from other Authorities that can be mapped into the RDD. Therefore Terms that are directly authorized by RDD neither define nor prescribe intellectual property rights or other legal entities.

As well as providing definitions of Terms for use in the REL, the RDD specification is designed to support the mapping and transformation of metadata from the terminology of one namespace (or Authority) into that of another namespace (or Authority) in an automated or partially-automated way, with the minimum ambiguity or loss of semantic integrity.

The dictionary is based on a logical model (the Context Model) which is the basis of the dictionary ontology. The model is described in detail in a normative annex to the specification. It is based on the use of verbs which are contextualised so that a dictionary created using the model with it can be as extensible and granular as required. An annex explaining how the model can be applied to generate new Terms is also provided.

7.8 Part 7 – Digital Item Adaptation (DIA)

One of the goals of MPEG-21 is to achieve interoperable transparent access to (distributed) advanced multimedia content by shielding users from network and terminal installation, management and implementation issues. This will primarily enable the provision of network and terminal resources on demand so that multimedia content can be created and ubiquitously shared, always with the agreed/contracted quality, reliability and flexibility.

Towards this goal, the adaptation of Digital Items is required. As shown in Figure 2, Digital Items are subject to a resource adaptation engine, as well as a descriptor adaptation engine, which together produce the adapted Digital Items. The target for this part of the standard is to specify tools that provide input to the adaptation engine, so that any constraints on the delivery and consumption of resources can be satisfied, and the quality of the user experience can be guaranteed.

It is important to emphasise that the adaptation engines themselves are non-normative tools of Digital Item Adaptation. However, descriptions and format-independent mechanisms that provide support for Digital Item Adaptation in terms of resource adaptation, descriptor adaptation, and/or Quality of Service management are within the scope of this part of the standard.

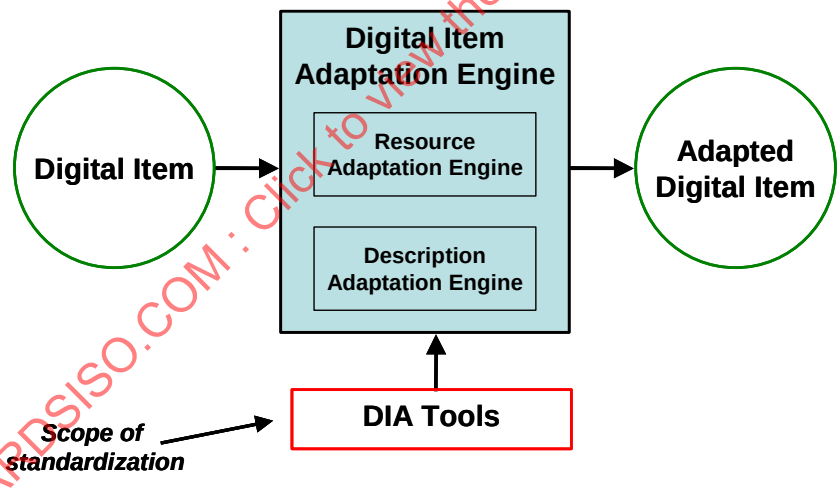

Figure 2 — Illustration of Digital Item Adaptation

7.9 Part 8 – Reference Software

Reference software is the first systems-related specifications in MPEG-21. The development of the Reference Software is based on the requirements that have been defined for an architecture for processing Digital Items.

7.10 Part 9 – File Format

An MPEG-21 Digital Item can be a complex collection of information. Both still and dynamic media (e.g., images and movies) can be included, as well as Digital Item information, metadata, layout information, and so on. It can include both textual data (e.g., XML) and binary data (e.g., an MPEG-4 presentation or a still picture). For this reason, the MPEG-21 file format inherits several concepts from MP4, in order to make 'multi-

purpose' files possible. A dual-purpose MP4 and MP21 file, for example, would play just the MPEG-4 data on an MP4 player, and would play the MPEG-21 data on an MP21 player. A 'resource map' allows the inclusion of multiple referenced resources in the same or other files, and for systems-level management of those resources (e.g., consolidating them into one file, or file compacting). The DID and resources can also be protected.

7.11 Part 10 – Digital Item Processing (DIP)

The Digital Item Declaration Language, as it is currently standardized, is an XML language for defining a Digital Item. This implies that a Digital Item Declaration is, as it should be, a static declaration. Digital Item Methods allow the User to add functionality to a Digital Item Declaration.

The standardization of Digital Item Processing allows interoperability at the processing level. The main idea behind the Digital Item Processing Architecture is that, on receipt of a DID, a list of DI Methods that can be applied to the Digital Item is presented to the User. The User chooses one Method that is then executed by the DIP Engine.

7.12 Part 11 – Evaluation Methods for Persistent Association Technologies

This Technical Report documents best practice in the evaluation of persistent association technologies, i.e., technologies that link information to identify and describe content using the content itself. This part does not contain any normative behaviour.

The purpose of this part is to allow evaluations of such technologies to be conducted using a common methodology rather than to standardise the technologies themselves. This is intended to give confidence to those relying on the technologies that the results are:

— Appropriate tests of the technology that will predict performance under real-world conditions

— Comparable with results obtained from other tests conducted using the same methodology.

This Technical Report focuses on the evaluation of two classes of technology: watermarks and fingerprints when applied to audio. It is expected that the scope of this Technical Report may be enhanced in future to cover other media types including still pictures.

7.13 Part 12 – Test Bed for MPEG-21 Resource Delivery

This part provides a (software-based) test bed for (i) the delivery of scalable media delivery, and (ii) testing/evaluating this scalable media delivery in streaming environments (e.g., by taking into account varying network environments).

7.14 Part 13 – Scalable Video Coding

One important application requirement in the MPEG-21 multimedia framework is efficient and seamless scalable coding. This means that compression efficiency should not be inferior when compared to non-scalable coders. Further, scalability should include different dimensions: spatial, temporal, SNR and complexity. Even though this had been a goal in the development of previous MPEG standards, it has not been achieved in a generic sense, because of sacrifice in performance. Hence, technologies such as wavelet coding which inherently possess scalability features can be potential candidates to achieve this objective, if their performance matches the state of the art.

A video coding technology providing flexible scalability of a single bitstream allows seamless integration of servers, heterogeneous networks, terminals, acquisition and storage devices which have different characteristics in the MPEG-21 multimedia framework. Examples of potential applications that can benefit from such improved scalable coding technologies are: video over the internet, video over wireless LANs, video over mobile systems for live broadcasting, specific video object decoding and conversational video services such as video telephony and video conferencing, multi-channel content production and distribution, surveillance-and-storage applications, and layered protection of contents.

Moreover, the new highly efficient scalable technologies follow an open loop approach (without recursion of frame-to-frame prediction). This overcomes the problem of error propagation effects which often occur when previous video coding standards were used in network environments with data losses. In addition, the effect of transmission losses can be minimized when unequal error protection is applied to parts of the scalable stream according to their respective relevance. When transmitted over networks with QoS support, a similar effect can be achieved by assigning higher priority to base layer parts of scalable streams which represent the video signal with a guaranteed minimum quality.

Other requirements which may be interesting are:

— A unique scalable solution serving broad range of rates, qualities and formats, potentially extending the definition of conformance points

— Support for complexity scalability by limiting the impact of incomplete decoding, allow to break decoding at any point to achieve a high degree of freedom in the design of complexity-scalable decoders.

— Temporal random access in both streaming and storage related applications. For storage, as further requirements, reverse play, fast forward/reverse, efficient editing and coding of manipulated frames.

The most important application requirement is scalability which allows arbitrary combinations of spatial, temporal and SNR dimensions. Such functionality would improve transmission over variable-bandwidth networks, storage and different-capability display devices. Multicast/broadcast over variable-bandwidth networks might be seen as potential major application, also in combination with storage on servers or in home networks or for surveillance applications. Environments with "encoding once, decoding multiple times at various rates and resolutions" paradigm would be very cost-effective. However, it is important that scalable coding does not impact rate-distortion performance and complexity when compared to state-of-the art single-layer coders. Efficient scalable representations can easily provide QoS capabilities, regardless of the provisions offered by the network. This will complement other MPEG-21 technology such as DIA.

8 MPEG-21 Achievements

8.1 Introduction

This clause lists those MPEG-21 parts, introduced in Clause 7, that have already reached Final Draft International Standard (FDIS) or Technical Report (TR) status, obviously excluding this document (ISO/IEC 21000-1). For each part, the goal and the rationale are listed. Consequently, the resulting key concepts and basic approach are presented.

8.2 ISO/IEC 21000-2: Digital Item Declaration (DID)

8.2.1 Goal

To establish a uniform and flexible model and interoperable representation schema for defining Digital Items.

8.2.2 Rationale

Within any system (such as MPEG-21) that proposes to facilitate a wide range of actions involving *Digital Items*, there is a strong need for a concrete representation of any individual *Digital Item*. Clearly there are many kinds of content, and probably just as many possible ways of representing it. This presents a strong challenge to design a powerful and flexible model for *Digital Items* that accommodates the many types of content, as well as any and all new forms it may assume in the future. Such a model is only truly useful if it yields a schema that can be used to unambiguously represent and interoperably communicate about any *Digital Items* defined within the model.

8.2.3 Key concepts and basic approach

The purpose of the Digital Item Declaration (DID) specification is to describe a set of abstract terms and concepts to form a useful model for defining Digital Items. Within this model, a Digital Item is the digital representation of an Asset, and as such, it is the thing that is acted upon (managed, described, exchanged, collected, etc.) within the model. The goal of this model is to be as flexible and general as possible, while providing for the "hooks" that enable higher-level functionality. This, in turn, will allow the model to serve as a key foundation in the building of higher-level models in other MPEG-21 elements (such as Identification & Description or IPMP). This model specifically does not define a language in and of itself. Instead, the model helps to provide a common set of abstract concepts and terms that can be used to define such a scheme, or to perform mappings between existing schemes capable of Digital Item Declaration, for comparison purposes.

The DID technology is described in three normative sections:

— Model: The Digital Item Declaration Model describes a set of abstract terms and concepts to form a useful model for defining Digital Items. Within this model, a Digital Item is the digital representation of an Asset, and as such, it is the thing that is acted upon (managed, described, exchanged, collected, etc.) within the model.

— Representation: Normative description of the syntax and semantics of each of the Digital Item Declaration elements, as represented in XML. This section also contains some non-normative examples for illustrative purposes.

— Schema:Normative XML schema comprising the entire grammar of the Digital Item Declaration representation in XML.

The following subclauses describe the semantic "meaning" of the principle elements of the Digital Item Declaration Model. Please note that in the descriptions below, the defined elements in italics are intended to be unambiguous terms within this model.

8.2.3.1 Container

A container is a structure that allows items and/or containers to be grouped. These groupings of items and/or containers can be used to form logical packages (for transport or exchange) or logical shelves (for organization). Descriptors allow for the "labelling" of containers with information that is appropriate for the purpose of the grouping (e.g. delivery instructions for a package, or category information for a shelf).

It should be noted that a container itself is not an item; containers are groupings of items and/or containers.

8.2.3.2 Item

An item is a grouping of sub-items and/or components that are bound to relevant descriptors. Descriptors contain information about the item, as a representation of an Asset. Items may contain choices, which allow them to be customized or configured. Items may be conditional (on predicates asserted by selections defined in the choices). An item that contains no sub-items can be considered an entity -- a logically indivisible Asset. An item that does contain sub-items can be considered a compilation -- a work composed of potentially independent sub-parts. Items may also contain annotations to their sub-parts.

The relationship between items (as defined in ISO/IEC 21000-2) and Digital Items could be stated as follows: items are declarative representations of Digital Items.

8.2.3.3 Component

A component is the binding of a resource to all of its relevant descriptors. These descriptors are information related to all or part of the specific resource instance. Such descriptors will typically contain control or structural information about the resource (such as bit rate, character set, start points or encryption information) but not information describing the "content" within.

It should be noted that a component itself is not an item; components are building blocks of items.

8.2.3.4 Anchor

An anchor binds descriptors to a fragment, which corresponds to a specific location or range within a resource.

8.2.3.5 Descriptor

A descriptor associates information with the enclosing element. This information may be a component (such as a thumbnail of an image, or a text component), or a textual statement.

8.2.3.6 Condition

A condition describes the enclosing element as being optional, and links it to the selection(s) that affect its inclusion. Multiple predicates within a condition are combined as a conjunction (an AND relationship). Any predicate can be negated within a condition. Multiple conditions associated with a given element are combined as a disjunction (an OR relationship) when determining whether to include the element.

8.2.3.7 Choice

A choice describes a set of related selections that can affect the configuration of an item. The selections within a choice are either exclusive (choose exactly one) or inclusive (choose any number, including all or none).

8.2.3.8 Selection

A selection describes a specific decision that will affect one or more conditions somewhere within an item. If the selection is chosen, its predicate becomes true; if it is not chosen, its predicate becomes false; if it is left unresolved, its predicate is undecided.

8.2.3.9 Annotation

An annotation describes a set of information about another identified element of the model without altering or adding to that element. The information can take the form of assertions, descriptors, and anchors.

8.2.3.10 Assertion

An assertion defines a full or partially configured state of a choice by asserting true, false or undecided values for some number of predicates associated with the selections for that choice.

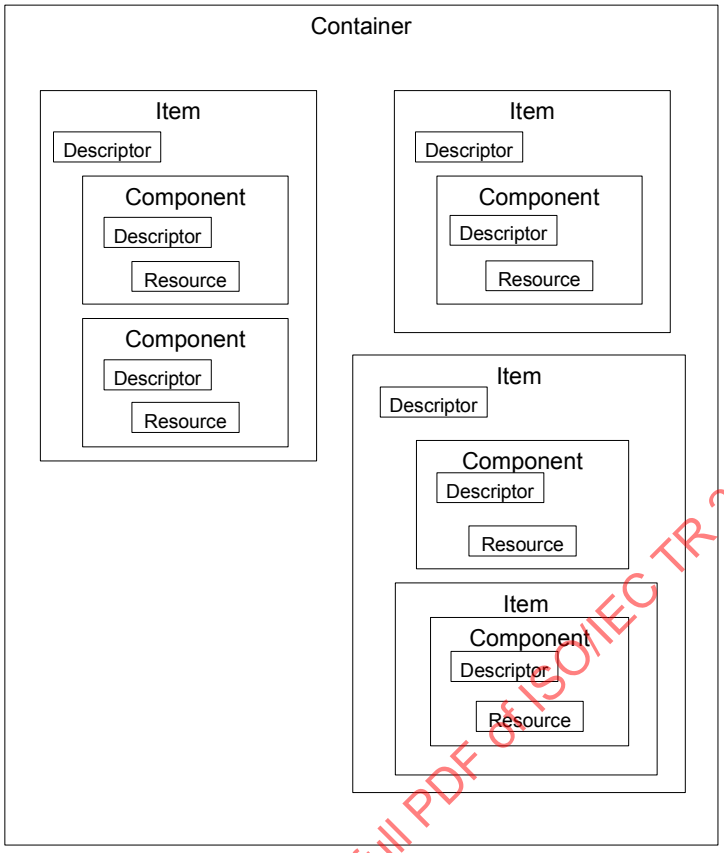

Figure 3 — Relationship of the principle elements within the Digital Identification Declaration Model

8.2.3.11 Resource

A resource is an individually identifiable Asset such as a video or audio clip, an image, or a textual Asset. A resource may also potentially be a physical object. All resources must be locatable via an unambiguous address.

8.2.3.12 Fragment

A fragment unambiguously designates a specific point or range within a resource. Fragment may be resource type specific.

8.2.3.13 Statement

A statement is a literal textual value that contains information, but not an Asset. Examples of likely statements include descriptive, control, revision tracking or identifying information.

8.2.3.14 Predicate

A predicate is an unambiguously identifiable Declaration that can be true, false or undecided.

Figure 3 shows the most important elements within this model, how they are related, and as such, the hierarchical structure of the Digital Item Declaration Model.

8.3 ISO/IEC 21000-3: Digital Item Identification (DII)

8.3.1 Goal

To enable the unique identification of Digital Items, different Digital Item Types and metadata schemes for describing Digital Items.

8.3.2 Rationale

The unique identification of Digital Items and the ascription of metadata to Digital Items, are crucial elements in many applications within MPEG-21's domain and have been successfully used in many areas. For example, ISBNs[5)] are used in the book industry in conjunction with a set of metadata to uniquely identify and describe books. This has enabled publishers and booksellers to manage their stock as well as customers to easily search and purchase books, especially in eCommerce transactions. Similar applications in the digital world are similarly dependent on unique and persistent (see Subclause 7.12) identification of Digital Items and their parts and the ability to associate well-formed metadata with such items.

Furthermore, to allow the identification of Digital Items themselves, DII needs to enable the association of a "related identifier" with a Digital Item (or part thereof). This latter concept is necessary to, for example, associate identifiers to intellectual property rights related to Digital Items (e.g., an ISWC[6)])

As common types of Digital Items emerge, DII provides a means to identify which type of Digital Item a Digital Item is.

8.3.3 Key concepts and basic approach

DII works in close conjunction with the Digital Item Declaration Language as provided in Part 2 and provides a set of XML elements that allow the inclusion of content identifiers and DI type identifiers into Digital Item Declarations.

In addition, DII uses the XML namespace mechanism to allow the identification of metadata schemes that can be used to describe Digital Items.

8.4 ISO/IEC 21000-5: Rights Expression Language (REL)

8.4.1 Goal

To establish a well-structured, flexible language for the unambiguous and machine-interpretable expression of permissions

8.4.2 Rationale

If a User is to honour another User's intellectual property rights and other interests in a Digital Item, the first User needs a way to understand the permissions that the second User has given him with respect to that Digital Item. To accurately communicate this understanding between Users, an unambiguous REL for expressing the permissions is needed. To enable Users to utilize machines to help them honour intellectual property rights and other interests, the REL must be machine-interpretable. The REL must also be well-structured and flexible, flexible so that it can be used to express a wide variety of permissions and well-structured so that, when it *is* used to express a wide variety of permissions, the semantics of the structure of each expression is still clear.

5) International Standard Book Number

6) International Standard Work Code

8.4.3 Key concepts and basic approach

The REL adopts a simple and extensible data model for many of its key concepts and elements. The REL defines conceptually abstract elements to represent the following four key concepts:

— Principal: represents the identity of a participant (such as a person or device) involved in activities relating to granting or consuming rights.

— Right: represents an act (such as playing or printing) usually in the sense of permitting the performance of that act.

— Resource: represents an object that may be used (such as a piece of multimedia content, a digital item, a web service, a name, or an email address).

— Condition: represents terms, conditions, or obligations and can be used to constrain the exercise of permissions.

These four conceptually abstract elements represent the *concepts* that are used to convey rights. In practical use, concrete elements are derived from these conceptually abstract elements to define specific instances of these key concepts. It is these concrete elements that actually appear in Rights Expressions.

Four of these concrete elements (one derived from each conceptually abstract element) can come together to form a *primitive grant*, which, in general terms, conveys to the identified Principal permission to perform the identified Right over the identified Resource if the identified Conditions are satisfied. Figure 4 illustrates the basic structure of a primitive grant (the conceptually abstract elements would be replaced with concrete elements of the appropriate type):

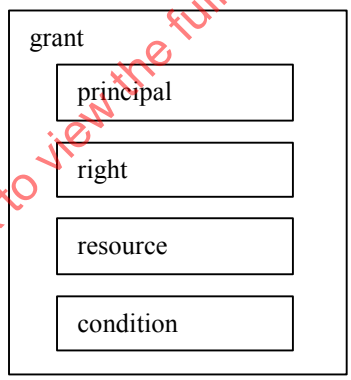

Figure 4 — Structure of a Grant

The REL also provides mechanisms for defining and referencing variables to construct non-primitive grants, grants that usually convey many related permissions at once. For instance, by using a variable in place of the conceptually abstract principal element, a non-primitive grant can convey permission to each of several Principals (for instance, all members in a particular club).

While a grant is more complete than, for instance, a Principal, it is not a complete Rights Expression that can be transferred unambiguously from one User to another. A full Rights Expression is called a License. A typical License consists of one or more grants and an issuer element. The issuer element can contain a principal or a Signature to identify the User who issued the License; the latter option (Signature) is useful to address trust and authentication issues and to ensure License integrity. The issuer element can also contain details relating to the issuance of the License (for instance, the time it was issued or the mechanism by which it might be revoked). The following figure illustrates the structure of a simple License that contains a single grant and is issued by a single User:

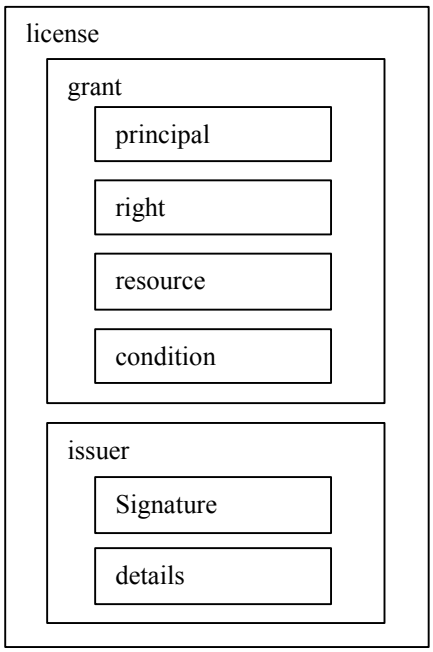

Figure 5 — Structure of a License

In addition to the defining the structure of Licenses down to the conceptually abstract level principal, rights, resource, and condition, the REL provides several concrete elements that can be used in place of these conceptually abstract ones in real Licenses. The REL also allows others to extend the language by defining their own concrete elements and to profile the language by specifying additional rules on the structure of Licenses and the types of concrete elements they may contain. An extension and a profile can be used concurrently to optimise the applicability of the REL to one specific situation while preserving the option for interoperability with other tools and systems.

8.5 ISO/IEC 21000-6: Rights Data Dictionary (RDD)

8.5.1 Goal

To provide a set of clear, consistent, structured, integrated and uniquely identified Terms to support the MPEG-21 Rights Expression Language (REL), ISO/IEC 21000-5 and to support the mapping of Terms from different namespaces.

8.5.2 Rationale

It is widely recognised that there is a need for semantic standards in the area of digital rights management. Not only are unambiguous semantics essential for the Rights Expression Language, but their lack has been a basic inhibitor of progress towards interoperability in that there can be no satisfactory way of exchanging data between different proprietary digital rights management applications. With such semantic standards every communication will depend on imprecise and unmaintainable many-to-many semantic mappings. This will be costly and de-incentivising for rights holders, who will have to bear the cost. In turn, this will damage the interests of DRM companies that will depend on a plentiful supply of interoperable content (interoperable because it is identified and described in a consistent manner). In short, for everyone involved, the cost of maintaining communications between proprietary applications without a standard semantic layer will be unsupportable.

8.5.3 Key concepts

8.5.3.1 RDD Database

The tool containing the RDD Dictionary and supporting its maintenance.

8.5.3.2 RDD Dictionary

The Terms and their TermAttributes defined according to this Standard.

8.5.3.3 RDD System

A system comprising the RDD Dictionary, the RDD Database and the specifications contained in Annex A.

8.5.3.4 RDD Registration Authority

The Registration Authority appointed to administer this Standard.

8.5.3.5 Term

Term is defined in the RDD Dictionary as “A semantic element with a defined Meaning and an RddIdentifier”. A Term is the basic unit of the RDD Dictionary structure. Terms have up to twelve standardised attributes.

8.5.3.6 RddIdentifier

RddIdentifier is defined in the RDD Dictionary as “the unique Identifier of a Term in the RDD Dictionary”.

8.5.3.7 Headword

Headword is defined in the RDD Dictionary as “The primary, human-readable Name of a Term according to its Authority”.

8.5.3.8 Term Description

TermDescription is defined in the RDD Dictionary as “A natural language Description of the Meaning of a Term.”

8.5.3.9 The Context Model

The fundamental building block of the RDD Ontology. The ContextModel defines a group of five Terms (the “BasicTermSet”) with associated Classes and RelatingTerms whose application to a specific ActType or ContextType results in the definition of a Family group of new Terms with DerivedMeanings and PartlyDerivedMeanings.

8.5.3.10 Family

Family is defined in the RDD Dictionary as “A group of Relationships that determine attribute inheritance from one Term to others according to the ContextModel”. There are two Types of Families of Terms: ActionFamily and ContextFamily.

8.6 ISO/IEC 21000-7: Digital Item Adaptation (DIA)

8.6.1 Goal

To define description tools to enable and enhance the adaptation of Digital Items.

8.6.2 Rationale

As the number of networks, types of terminals and content representation formats increase, interoperability between different systems and different networks is becoming a greater challenge. Common devices such as personal computers, televisions, and mobile devices support a variety of different multimedia content representation formats, and vary in their display capabilities, processing power and memory capacity. Furthermore, the networks that these devices are connected to are often characterized by different capabilities and conditions. Additionally, user preferences and other factors of the usage environment must also be accounted for to enable the delivery of personalized services. Given this complex and dynamic usage environment, it becomes necessary to have a standardized description of the usage environment, as well as a set of standardized tools that enable the scaling of resources and, in general, the adaptation of Digital Items.

8.6.3 Key concepts and basic approach

The Digital Item Adaptation tools are clustered into eight major categories as illustrated in Figure 6. The categories are clustered according to their functionality and use for Digital Item Adaptation. Common to all tools and depicted in the centre of the diagram are the *Schema Tools* and *Low-Level Data Types*. The schema tools provide root elements and referencing mechanisms for all DIA descriptions, while the low-level data types specify some basic data types that can be used generically by several DIA tools.

— The first major category is the *Usage Environment Description Tools*, which include User characteristics, terminal capabilities, network characteristics and natural environment characteristics. These tools provide descriptive information about the various properties of the usage environment, which originate from Users, to accommodate, for example, the adaptation of Digital Items for transmission, storage and consumption.

— The second category is referred to as *BSDLink*, which provides the facilities to create a rich variety of adaptation architectures based on tools specified within ISO/IEC 21000 and ISO/IEC 15398, among others.

— *Bitstream Syntax Description* tools comprise the third major category of Digital Item Adaptation tools. A BSD describes the syntax – in most cases, the high level structure – of a binary media resource. Using such a description, a Digital Item resource adaptation engine can transform the bitstream and the corresponding description using editing-style operations such as data truncation and simple modifications.

— The fourth category of tools is referred to as *Terminal and Network Quality of Service*. The tools specified in this category describe the relationship between QoS constraints (e.g., on network bandwidth or a terminal's computational capabilities), feasible adaptation operations satisfying these constraints and associated media resource qualities that result from adaptation. This set of tools therefore provides the means to trade-off these parameters with respect to quality so that an adaptation strategy can be formulated and optimal adaptation decisions can be made in constrained environments.

— The *Universal Constraints Description Tools* form the fifth category of tools, which enables the possibility to describe limit and optimization constraints on adaptations.

— The sixth category is referred to as *Metadata Adaptability*. This tool specifies hint information that can be used to reduce the complexity of adapting the metadata contained in a Digital Item. On the one hand they are used for filtering and scaling and on the other hand for integrating XML instances.

— For *Session Mobility*, the seventh category of tools, the configuration state information that pertains to the consumption of a Digital Item on one device is transferred to a second device. This enables the Digital Item to be consumed on the second device in an adapted way.

— Finally, the eighth category of tools is referred to as *DIA Configuration Tools*, which provides information required for the configuration of a Digital Item Adaptation Engine.

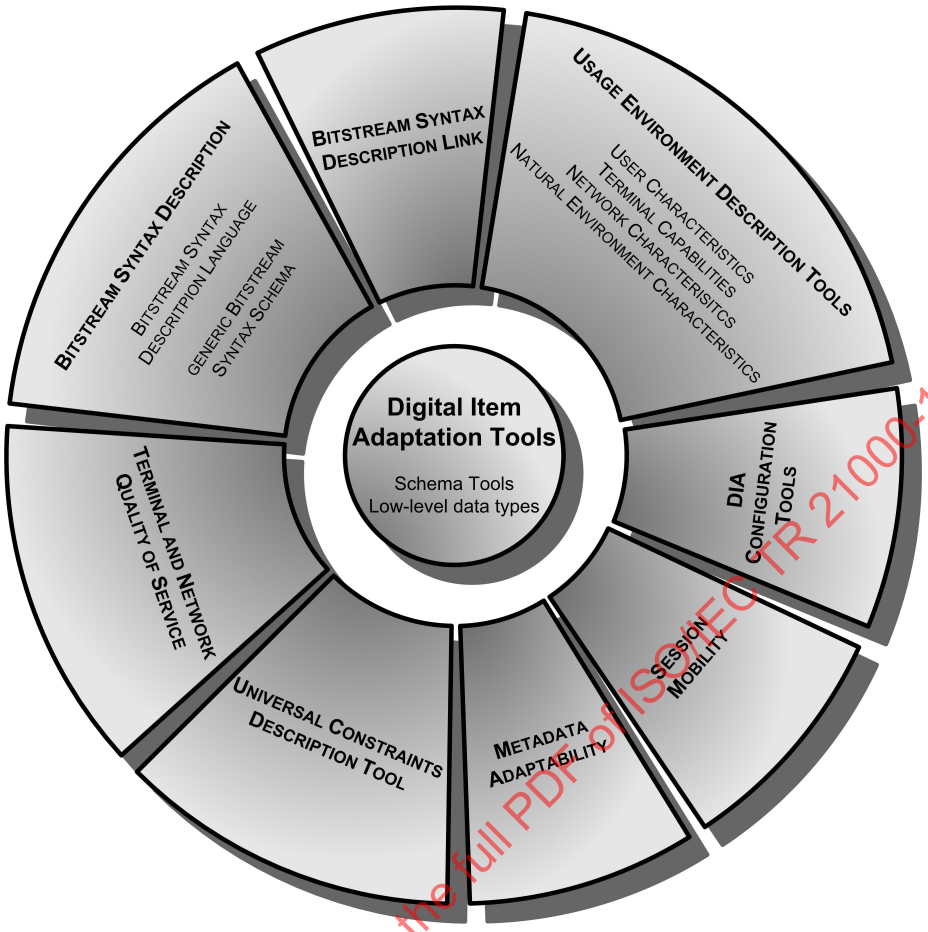

Figure 6 — Overview and Organisation of Digital Item Adaptation Tools

According to the DIA model shown in Figure 2 in Subclause 7.8, a Digital Item is input to a Digital Item Adaptation Engine. The Adaptation Engine can modify the input Digital Item by adapting the resources or metadata within the Digital Item or the declaration of the Digital Item to the usage environment. Additionally, the identifiers and rights expressions pertaining to the adapted Digital Item need not be the same as those pertaining to the input Digital Item. The DIA specification deals with adaptation, but specifically does not address the relationship of rights and permissions to adaptations. The relationship of rights and permissions is to be addressed in an amendment to ISO/IEC 21000-7. It is expected that users ISO/IEC 21000-7 will register terms describing their specific adaptations with the Registration Authority described in ISO/IEC 21000-6 in order to provide interoperability.

9 Highlights of the MPEG-21 Parts under Development

9.1 Introduction

This clause gives a more detailed insight in the MPEG-21 parts that are currently well advanced in MPEG-21. For each part, the goal and the rationale are listed. Consequently, the resulting key concepts and basic approach are presented.

9.2 ISO/IEC 21000-10: Digital Item Processing

9.2.1 Goal

To provide tools that allow Users creating Digital Items to provide other Users suggested mechanisms by which they may interact with that Digital Item. This extends to methods for creating, altering, interacting with and consuming elements of the Digital Item.

9.2.2 Rationale

One of the important aspects related to concept of the Digital Item is that the Digital Item Declaration (see Clause 7.3) is a static declaration of only the information (structure, resources and metadata) in the Item; in the declaration itself there is no implied processing of that information. This is important as it contrasts with e.g. HTML based Web pages where the presentation of the information is intermingled with the information itself. While the separation of the information from the usage of the information in Digital Items has significant advantages (different Users can receive very different presentations of information), it also means that on receipt of the static declaration, a User has nothing that indicates how the information should be processed.

9.2.3 Key Concepts

Thus, Digital Item Processing is the area of MPEG-21 that encompasses all the aspects of processing a static Digital Item from a User (and application) perspective. It includes such processing as downloading the Digital Item; requesting IPMP and Rights processing of the DI; downloading and presenting individual media resources; and performing tasks ("methods") that an author requires for a particular DI. While much of Digital Item Processing is still in the early stages of consideration, the last aspect mentioned (Digital Item Methods, DIMs) has been identified of particular importance to Digital Item usage and defined in more detail.

The concept of Digital Item Methods is that, on receipt of a Digital Item, a User will have available a "list" of processes ("methods") that can be applied to the Digital Item. These methods provide a mechanism for a User (author, publisher, distributor, etc) to specify a preferred set of procedures by which the DI should be handled. A simple illustration of this functionality would be a music album DI with an "Add Track" method. The method would allow a User to add a new track to the DI in the preferred format of the DI (i.e., in the correct place in the hierarchy and with the correct descriptors). MPEG intends the methods to be run on a method engine (called the Digital Item Method Engine, DIME), which supports standard base operations (Digital Item Base Operations, or DIBOs; analogous to the standard library of functions of a programming language). Users will be able to create methods for their Digital Items from these base operations and using a standardised syntax (Digital Item Method Language, DIML).

The DIML is based on the third edition of ECMAScript [1]. This provides lexical conventions and syntactical structure, flow control structures (i.e. if/then, while, etc.), primitive data types (String, Boolean and Integer), composite data types (objects and arrays), standard arithmetic, logical and bitwise operators, exception handling, error definitions and support for Regular Expressions.

In the DIP context, the ECMAScript host environment is the MPEG-21 Peer (or more precisely the DIME). DIML extends ECMAScript by defining additional object types, global object properties, and functions that are standardised by MPEG-21 DIP. Additional object types defined in DIML represent such things as a DID document, DIP exceptions, resource status, and the object map. The object map is required to map from DID elements, to objects with semantic significance to a particular DIM (or DIMs). For example, an Item element in a DID can be mapped to a Music Track object, which is understood by an Add Track method.

Additional properties of the global object defined in DIML include a reference to the current DID document object.

Additional functions defined in DIML include the standardized set of DIBOs. MPEG-21 DIP specifies the normative syntax and normative semantics of each DIBO (the “what”), but does not mandate how an MPEG-21 Peer might implement those semantics (the “how”). Any given Peer will be expected to provide some base level of functionality, for example to “play” a resource. The DIBOs provide a standard interface for a DIM author to access this functionality of the Peer. The standardized semantics of the DIBO indicate the generic functionality that must be provided by an implementation of a DIBO. However, the manufacturer of the Peer is free to implement the DIBO to provide that functionality as they wish. The standardized DIML and DIBOs provides for interoperability of DIMs, yet the freedom to implement DIBOs still allows for such things as competition among manufacturers of MPEG-21 Peers, and MPEG-21 Peers suited to specific purposes.

The current set of DIBOs specified can be placed in to one of the following categories.

— Operations that manipulate the DID at the XML level, such as inserting or removing elements, setting the value of an element attribute.

— Operations that manipulate the state of a DI, such as setting the state of selections within a choice.

— Operations that are representations of or associated with an RDD verb, such as play.

Digital Item extension Operations, DIxOs extend DIP to allow optimal execution of more complex tasks. The means for invoking a DIxO from DIML is normatively specified, however the actual language used to implement a DIxO is not normative. Currently DIxOs are restricted to being able to call the DIBOs only, and hence a binding for the DIBOs is required for the language in which a DIxO is implemented. In addition an execution environment for the DIxOs implemented in a given language needs to be specified. Currently the DIP working draft version 2 includes a Java binding for the DIBOs and a specification of a Java execution environment for Java DIxOs, J-DIxOs.

9.3 ISO/IEC 21000-11: Evaluation Methods for Persistent Association Technologies

9.3.1 Goal

To provide an objective and rigorous framework for the assessment of Persistent Association Technologies by setting out a framework for guiding the assessment process. Within that framework, to provide *recommendations* detailing best practice in undertaking the various aspects of an assessment process, utilising *standardised testing methodologies* wherever possible.

9.3.2 Rationale

A general task that has been widely addressed with numerous technologies concerns the linking of content items with metadata. By far the most widely deployed approach to date involves juxtaposing metadata and content elements – either within file formats or databases. However, these approaches are “fragile” in so far as the link between content and metadata can easily be deleted or otherwise lost. This has led to the development of families of “Persistent Association Technologies” (PAT) that can provide a degree of resilience in establishing and maintaining the association between content and metadata. Technologies presently fall into two main categories of watermarking and fingerprinting. In recent years there have been some significant cases where PAT has been assessed and/or deployed in commercial applications. In tandem with this, there has been significant debate – especially on engineering issues – and some considerable progress in the techniques used to assess PAT in a variety of application scenarios. In some cases PAT has provided excellent solutions and in other cases results have been less positive.

The rationale of this present document has been to gather the experience and best practice of practitioners who have worked with PAT development, assessment or deployment and to set out this experience for the benefit of others performing similar tasks. Presently, the focus of contributors to this effort has been on PAT for audio applications. However, this document is has been deliberately structured in such a way that future contributions on video or still image technologies could readily be incorporated.

9.3.3 Key concepts

9.3.3.1 Application Framework

Whilst the requirements of PAT and of its assessment in different applications may differ, there are nonetheless many generic factors that can and should be assessed or considered in a structured way. These factors have been set out to form a framework for approaching the assessment task.

9.3.3.2 PAT Parameters

Key parameters of PAT have been identified for assessment, as:

— Fingerprint Size;

— Watermark Payload;

— Granularity;

— Perceptibility;

— Robustness;

— Reliability; and

— Computational Performance.

There is a need to consider the performance trade off that occurs as these parameters are varied. The document sets out a description of these factors along with detailed recommendations for assessing them.

9.3.3.3 Issues in PAT

There are issues in PAT, in particular the resistance to malicious attacks, scalability and interactions between different PAT systems. The document provides some information on these issues.

9.3.3.4 Practical Testing

Constructing a practical test environment complete with test items, impairment tools, and automation tools is a considerable engineering exercise. The document sets out advice and guidelines on these topics.

9.4 ISO/IEC 21000-12: Test Bed for MPEG-21 Resource Delivery

9.4.1 Goal

To provide a platform for evaluating standards within the MPEG-21 framework in the context of various resource delivery technologies and scenarios.

9.4.2 Rationale

In order to enable maximum use of its standards WG11 encourages the exploitation of synergies among the different MPEG specifications. ISO/IEC 21000-12 will deliver this synergy beyond MPEG-21 by providing a platform for the integration and evaluation of MPEG technologies in the context of media resource delivery.

9.4.3 Key Concepts

The Test Bed is designed for assessment of the performance of scalable video codecs in a streaming application and can be extended to a generic environment for the evaluation of resource delivery technologies over unreliable packet-switched networks. Currently, the test bed is designed for testing and verification of the following technologies in a dynamic environment:

— MPEG-21 Digital Item Declaration;

— MPEG-21 Digital Item Adaptation;

— Scalable video and audio coding technologies;

— MPEG-4 delivery using the Internet Protocol; and

— MPEG-2/4 IPMP compliant content protection technologies.

It is important to note that the Test Bed is an informative implementation of how various MPEG technologies which can be integrated into a working system for testing and verification purposes. The specific implementation of MPEG technologies used in the Test Bed represents only one way of achieving this

integration and is not part of the normative standards of respective technologies, unless they are part of the reference software of the corresponding MPEG standards.

The Test Bed is mainly composed of three pieces of software:

— Server,

— Client, and

— Network emulator.

A schematic overview of the Test Bed is illustrated in the figure below.

In particular, the network emulator allows the user to specify network parameters such as channel bandwidth, packet loss rate, network buffer size, delay, and jitter, in a network profile and controls the network condition accordingly in real time. The Test Bed employs different audio and video codecs for representing the media content and contains an MPEG-21 Digital Item Adaptation Engine that demonstrates MPEG-21 Network Adaptation QoS mechanisms.

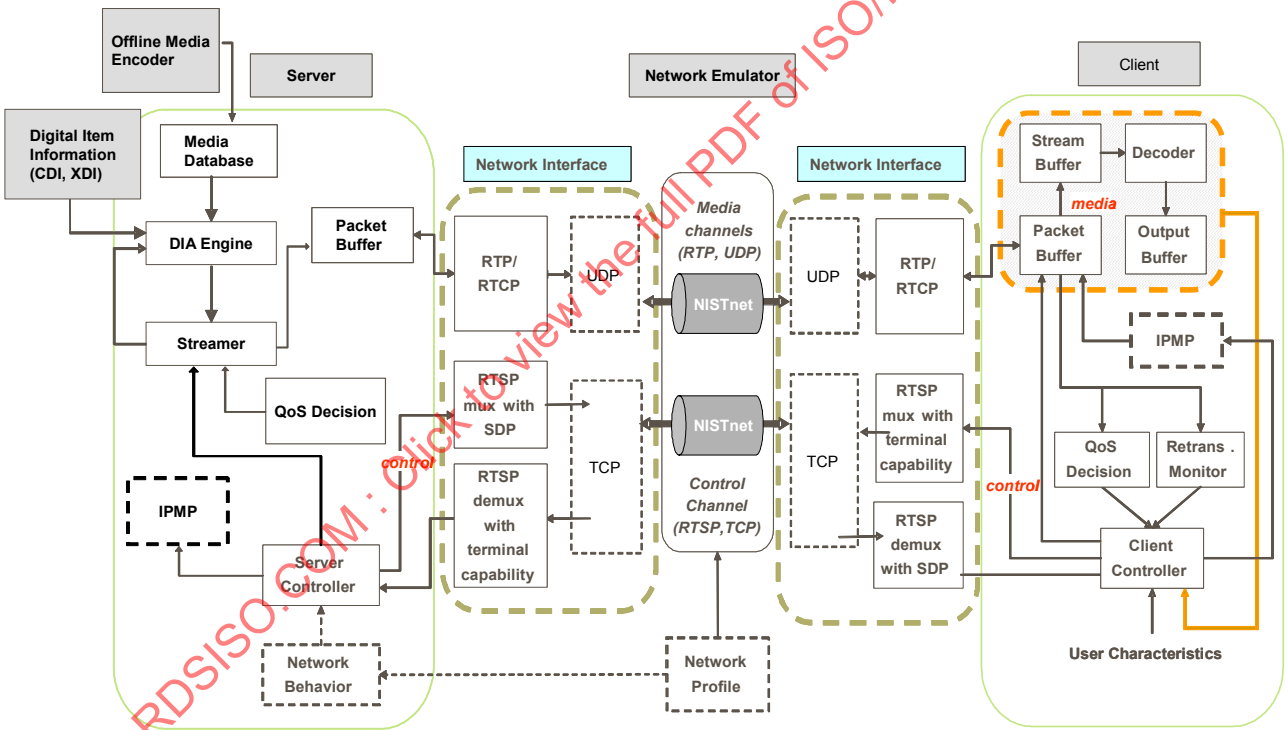

Figure 7 — Schematic Overview of the Multimedia Test Bed

Annex A
(informative)

List of Activities Related to the Multimedia Framework

A.1 Introduction

This annex gives a non-exhaustive list of activities related to the Multimedia Framework. The activities are grouped into four different categories:

— Standardised models and frameworks

— Current standardisation activities

— Forums and consortia

— Specifications produced by forums and consortia

A.2 Standardised Models and Frameworks

A.2.1 CWA (CEN Workshop Agreement) 13699 “Model for Metadata for Multimedia Information”

It defines the terms information resource, metadata and multimedia, comments on the nature of metadata and explores the relationship between metadata and the information to which it refers. A conceptual model for metadata for multimedia information resources was elaborated based on three concepts: metadata classes, roles and actions. Three major roles were identified in the metadata model, creators, service providers and users, who perform actions that apply equally to both information resources and metadata. A life cycle model was described which can be applied either to an information resource or to its associated metadata and illustrates the phases through which an information resource and its associated metadata may pass from creation or acquisition to retention or disposal. The Agreement recommended the development of standardised mechanisms for performing actions on multimedia information resources and metadata.

A.2.2 CWA 13700 “Requirements for metadata for multimedia information”

It presents a requirements taxonomy that identifies a set of basic general requirements. They are classified into a) metadata, and b) facilities needed in association with metadata. The classifications are aligned, respectively, with related concepts in the metadata model (metadata classes) and in the metadata framework (framework decomposition into delivery, access, protocols, discovery and asset management and interoperability). Most of the identified requirements come from several application domains such as Retail, Services, Public Administration, Entertainment, Scientific and Cultural heritage, Publishing, Education and Healthcare. Requirements from two application examples, audio-visual publishing and brokerage, are described in more detail and the taxonomy is used to list requirements in real applications such as the Danish Library use of Dublin Core metadata.

A.2.3 The Metadata for Multimedia Information Framework

It provided a structured classification of information and activities involving metadata for multimedia information in early 1999; providing brief information on a topic, the specifications and standards being developed, what consortia or research projects were actively working in a field and what reference material (books, magazine articles and web sites) were relevant. The framework was never progressed to a CEN Workshop Agreement since it was intended to be a “live” resource with constant updating. Another CEN/ISSS