ECP-2007-DILI-517006

EFG – The European Film Gateway

Common interoperability schema for archival resources and filmographic descriptions

Report on the common interoperability schema

Deliverable number	*D 2.2*
Dissemination level	*Public*
Delivery date	*24 June 2009*
Status	*Final*
Authors	*Franca Debole (CNR - ISTI), Pasquale Savino (CNR - ISTI), Detlev Balzer (for DIF) with contributions by Marco Rendina (for IL), Go Sugimoto, (EDL), Georg Eckes (DIF), Francesca Schulze (DIF)*

1 Introduction

This deliverable describes the common interoperability schema developed for EFG. The schema allows to consolidate the data held in the various heterogeneous source databases of the EFG content providers in a common format. A first draft outline of the schema had been established in the beginning of March 2009. Since then it has been refined in a series of meetings and telephone conferences. Mainly involved in the establishment of the finalised version were Consiglio Nazionale delle Ricerche – Istituto di Scienza e Tecnologie dell'Informazione (CNR-ISTI), Deutsches Filminstitut – DIF (DIF), Istituto Luce (IL) and Europeana (EDL). An advanced draft version has been presented to the project partners during the WG 3 workshop carried out in Copenhagen from 11 to 13 May 2009. On the basis of the discussion held in Copenhagen the data model has then been finalised. It will be circulated to all content providers before the end of June 2009. As defined in the DOW (M2.9 "Evaluation and approval of EFG interoperability by EDL"), evaluation and approval of the existing EFG metadata schema by the Europeana interoperability staff is expected until beginning of July. Part of the metadata schema described in this deliverable is an XML expression of the same, which is currently being developed. Following the work on the EFG metadata schema, the mapping from the EFG to the Europeana format ESE[1] will be carried out shortly.

This deliverable is divided into five chapters. Chapter 2 gives a general overview of the EFG metadata model, while chapter 3 introduces the major entities. Chapter 4 focuses on common elements that are applied within more than one entity of the EFG metadata model. Finally, chapter 5 describes the relationships that can exist between the various entities.

---

[1] Europena Semantic Elements

2 Preliminary remarks to the EFG metadata model

The EFG data model described in this deliverable is composed of eight major entities. Entities are defined as the topmost level of description that can have relationships to other entities. In order to describe objects, the EFG metadata model basically distinguishes between three levels: Creation, Manifestation and Item. "Creation" is the topmost level of description and basically refers to the concept of Cinematographic Work as defined in the EN 15907 reference model. Other reference models have been used, such as the FRBR (concepts: Work / Expression) as well as the concept of WPE (Work / Primary Expression) as defined by the OLAC[2]. A creation in EFG can be either an audiovisual work (represented in the "AVCreation" entity) or a non-audiovisual work (represented in the "NonAVCreation" entity). This applies also to the "Manifestation" level which is represented in the entities "AVManifestation" and "NonAVManifestation". The "Item" entity functions as a logical wrapper for the digital file presented in the EFG portal. Moreover, "Agents", "Event" and "Collection" also have entity status.

Each major entity of the schema comes with elements, attributes and relationships. A thorough introduction of these is done in chapter 3. Simple elements are defined redundantly while common elements are defined in a separate chapter. Common elements, in contrast to regular elements, can occur as part of more than one major entity. Moreover, the description of the metadata model includes proposals for value spaces that can be controlled either through vocabularies or syntax definitions.

The last chapter provides details about the relationships that can occur between entities. Relationships can have a type attribute for values that define the semantics of the relationship. These values will be taken from controlled vocabularies which specify a forward and inverse form of the relationship, and possible restrictions depending on the classes of entities for which the relationship is valid. The vocabularies for the relationship type attributes will be defined within Work Package 3.

The EFG metadata model applies existing metadata standards and schemas as well as reference models from the audiovisual and digital library sector. These are:

- EN 15907, Film identification – Enhancing interoperability of metadata – Elements sets and structures.[3]
- Specification for the Europeana Semantic Elements. V3.1 Date: 25.02.2009. – URL: http://dev.europeana.eu/public_documents/Specification_for_metadata_elements_in_the_Europeana_prototype.pdf

---

[2] Online Audiovisual Cataloguers
[3] http://www.filmstandards.org

- IFLA: Functional Requirements for Bibliographic Records. Corrected version of 2009.
  – URL: http://www.ifla.org/en/publications/functional-requirements-for-bibliographic-records
- Moving Image Work-Level Records Task Force (Online Audiovisual Cataloguers): Report and Recommendations for Moving Image Works. – URL: http://www.olacinc.org/drupal/?q=node/27. – Cited: 16.06.2009.
- Encoded Archival Context (EAC). – URLs: http://jefferson.village.virginia.edu/eac/; http://jefferson.village.virginia.edu/eac/. – Cited: 16.06.2009

Moreover, this documentation provides proposals for standards that can be used to define controlled vocabularies or syntactic rules for value spaces. These are:

- ISO 639-1:2002. Codes for the representation of names of languages - Part 1: Alpha 2 code
- ISO 3166-1: Codes for the representation of names of countries and their subdivisions. - Part 1: Country codes
- ISO-3166-2: Codes for the representation of names of countries and their subdivisions – Part 2: Country subdivision code
- ISO 8601:2004, Data elements and interchange formats - Information interchange - Representation of dates and times.
- Library of Congress: MARC Code List for Countries. Date: April 2008 – URL: http://www.loc.gov/marc/countries/cou_home.html
- ISO 15511:2003 Information and documentation – International Standard Identifier for Libraries and Related Organisations (ISIL)
- Internet Media Types (MIME). - http://www.iana.org/assignments/media-types

At the core, the EFG data model is an entity relationship model, although it applies some parts from an RDF[4] logic namely semantically defined types for relationships (see chapter 5). This entity-relationship view is illustrated in figure 1 below which provides an overview of all major entities and the relationships between them. For relationships, the cardinalities are given in a notation that is explained in figure 2. The model implies that all many-to-many relationships are bi-directional, i.e. that they can have distinct meanings in each direction.

---

[4] Resource Description Framework
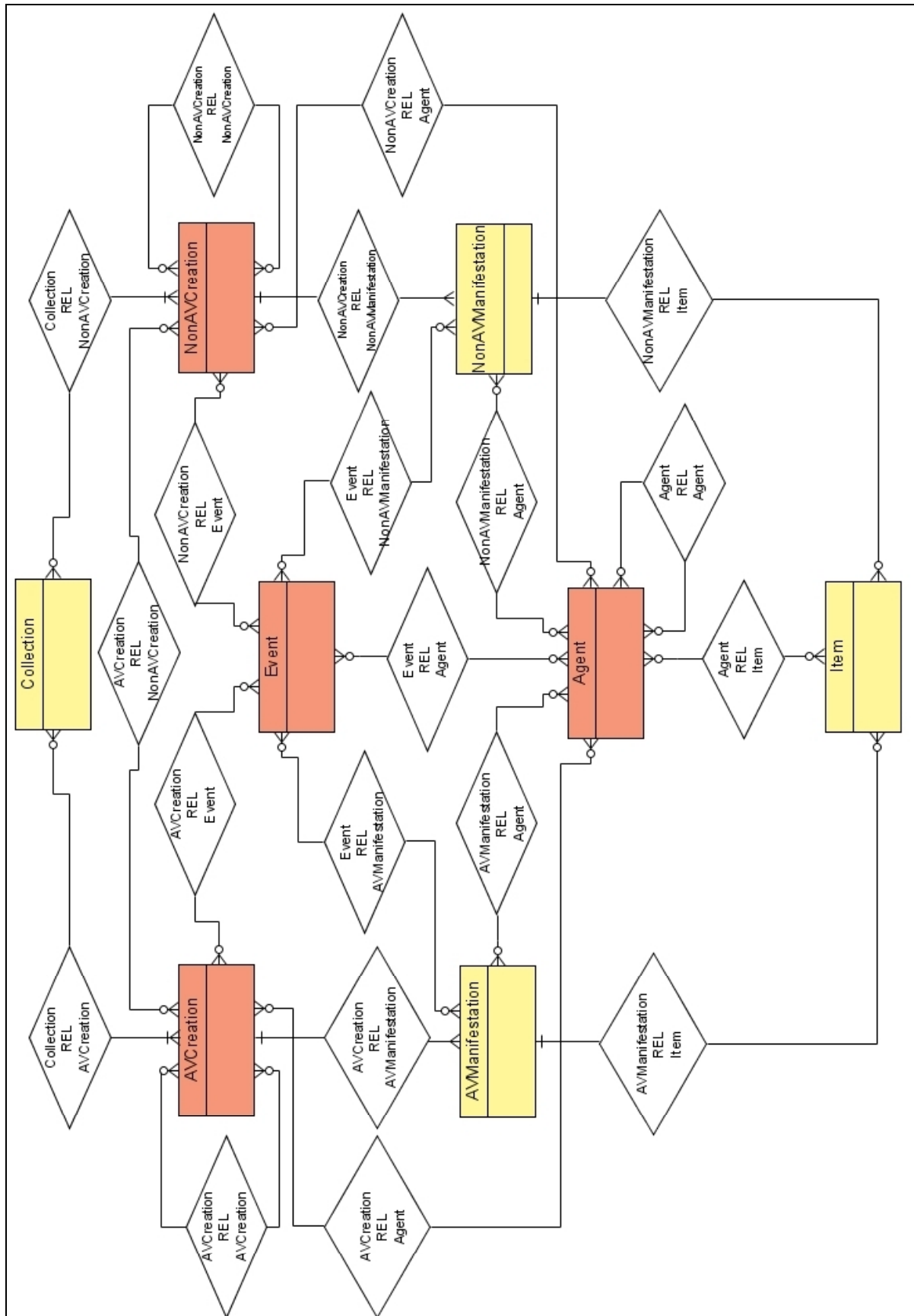
Figure 1: Entities and relationships of the EFG metadata model
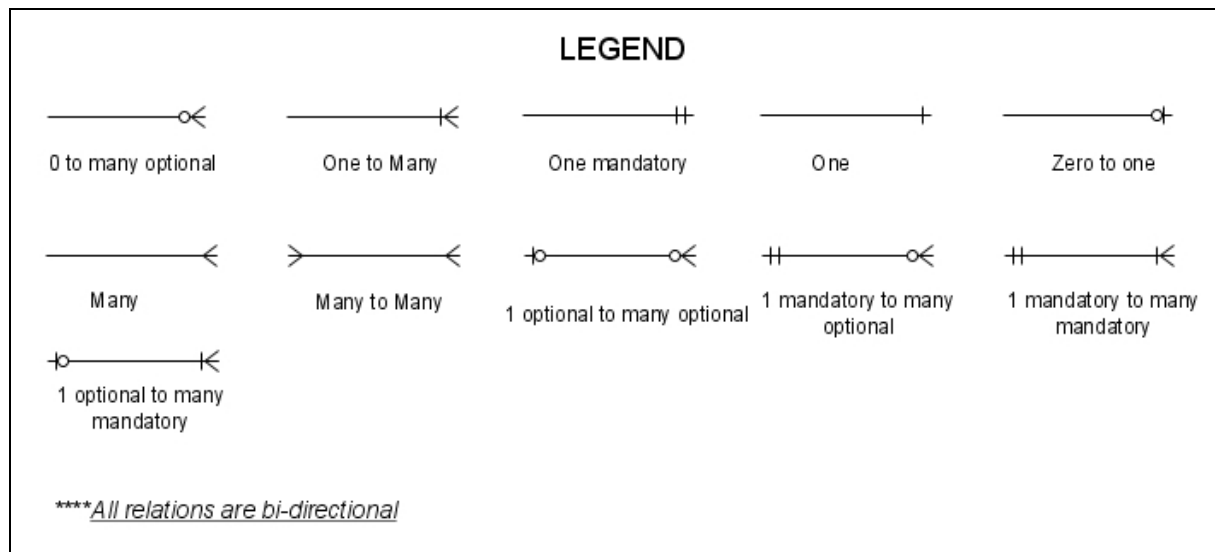
Figure 2: Cardinalities of relationships

For a better understanding of the metadata schema documentation provided hereafter, a short definition of the used terms is given:

- **Entity**: A conceptual unit defining a focus of description. Entities can enter relationships with other entities. An entity can have attributes and can be composed of elements.

- **Element**: A semantic unit that is a major property of one or more entities. An element can have attributes.

- **Attribute**: A named property, of elements and/or relationships, whose value domain can be restricted by a data type. An attribute usually consists of a name and a value.

- **Relation**: A relation is a property that associates an entity to another entity. A relation can have attributes.

- **Cardinality**: The number of times an instance of a relationship is permitted. An element with the cardinality status zero is optional. A relationship with the status one or more is mandatory.

- **Occurrences**: The number of times an instance of an element can exist (zero, one or more). An element with the occurrence status zero is optional. An element with the status one or more is mandatory.

3 Entities

For each entity listed here, a list of elements and attributes is supplied. For some of these, proposals for value spaces are given, either through controlled vocabulary or through syntax definitions. The number of occurrences, the entity's status as well as their allowed relationships are indicated in all cases. It should be noted, that some occurrences can be subject to revision.

3.1 AVCreation

This entity refers to the concept of cinematographic work as defined in EN 15907 (clause 4.1). An audiovisual creation should only have properties that remain constant throughout all of its manifestations.

3.1.1 Elements

- **Identifier** (described in 4.1)
  In this context it is the identifier of the audiovisual creation.
  Status: Mandatory; Occurrences: 1

- **Title** (described in 4.2)
  Status: Mandatory; Occurrences: 1..n

- **Record Source** (described in 4.3)
  Status: Mandatory: Occurrences: 1..n

- **Identifying Title**
  A short phrase for identifying the audiovisual creation, to be used e.g. in human-readable result lists from database queries. The identifying title should not be constructed by a cataloguer according to cataloguing rules but algorithmically during the metadata conversion process for EFG.
  Status: Mandatory; Occurrences: 1

  Attribute:
  - Origin
    Acronym or other identifier indicating the origin of the element content. For identifying titles created automatically, this should be the name and version of the software agent.
    Status: Optional

- **Country of Reference**

  The geographic origin of a audiovisual creation. This should be the country or countries where the production facilities are located. If production information is missing, this element can refer to countries where the audiovisual creation was filmed or distributed, or where copies are known to exist in archives.
  Value space: controlled vocabulary (e.g. "ISO 3166-1")
  Status: Mandatory; Occurrences: 1..n

  Attribute:

  - Reference

    The relationship between a geographic area and the audiovisual creation. Defaults to "production", use of this attribute is only required if the point of reference is different from the country of production.
    Value space: controlled vocabulary (e.g. "production", "shooting location" etc.)

- **Production Year**

  The year associated with the production of the audiovisual creation. A typical use of this element is the chronological ordering of AVCreations.
  Status: Mandatory; Occurrences: 1..n

  Attribute:

  - derived from

    If this attribute is empty or omitted, then it is assumed that the "Production Year" has been validated as a real production year. Otherwise, it may contain a textual string which specifies how it has been derived (e.g. inferred from censorship).
    Value space: free text

- **Keywords** (described in 4.5)
  Status: Optional; Occurrences: 0..n

- **Description** (described in 4.6)
  Status: Optional; Occurrences: 0..n

- **Note**

  This element includes all data derived from content providers, that cannot fit into other elements.
  Value space: text
  Status: Optional; Occurrences: 0..n

Attribute:

- <u>Source</u>
  This attribute specifies the source that produced the note element.
  Value space: text

- **UserTag**
  This is a tag created by the user through the EFG interface.
  Value space: text
  Status: Optional; Occurrences: 0..1

3.1.2 Allowed Relationships

- AVCreationRELAgent (described in 5.1)
- AVCreationRELEvent (described in 5.7)
- AVCreationRELAVManifestation (described in 5.8)
- AVCreationRELAVCreation (described in 5.2)
- AVCreationRELNonAVCreation(described in 5.6)
- CollectionRELAVCreation (described in 5.9)

3.2 AVManifestation

This entity groups all properties that can change during the lifecycle of an audiovisual creation without affecting the identity of a film work as such. A manifestation is the physical embodiment of an audiovisual creation. Examples are archival copies (analogue or digital) or distribution files. This entity was modelled by using the EN 15907 (clause 4.3).

3.2.1 Elements

- **Identifier** (described in 4.1)
  In this context it is the identifier of the audiovisual manifestation.
  Status: Mandatory; Occurrences: 1

- **Title** (described in 4.2)
  Status: Mandatory; Occurrences: 1..n

- **Record Source** (described in 4.3)
  In this context it is the identifier of the audiovisual manifestation.
  Status: Mandatory; Occurrences: 1..n

- **Language**
  The language of the spoken, sung or written content.
  Value space: controlled vocabulary (ISO 639-1 based on RFC 4646)
  Status: Optional; Occurrences: 0..n

  Attribute:

  - Language usage
    This indicates the relationship between the language and the manifestation.
    Defaults to "unspecified".
    Value space: controlled vocabulary (e.g. "unspecified", "original spoken dialogue", "dubbing", "subtitles", "voice-over commentary", etc.)

- **Dimension**
  The total physical dimension of the manifestation represented as numeric value, with decimal places if required. For example, if the carrier type is a DVD, it represents the DVD size. If the carrier type is a file, it represents the size in bytes, etc.
  Status: Optional; Occurrences: 0..n

  Attributes:

  - unit
    Value space: controlled vocabulary (e.g. "bytes", "meters", "inches")
  - reference
    If the manifestation is composed of more than one physical carrier for which individual lengths are known, then this attribute should be used to refer the length information to a particular carrier (e.g. "reel 3").
    Value space: text

- **Duration**
  The running time of the audiovisual manifestation measured in minutes and seconds.
  Status: Optional; Occurrences: 0..1

  Attribute:

  - frame rate
    Optional value for the projection speed, given in frames per second, to which the given duration refers.
    Value space: text

- **Format** (described in 4.4)
  Status: Optional; Occurrences: 0..1

- **Note**
  This element includes all data derived from content providers, that cannot fit into other elements.
  Value space: text
  Status: Optional; Occurrences: 0..n

  Attribute:

- Source
  This attribute specifies the source that produced the note element.
  Value space: text

3.2.2 Allowed relationships

- EventRELAVManifestation (described in 5.12)
- AVManifestationRELItem (described in 5.10)
- AVManifestationRELAgent (described in 5.11)
- AVCreationRELAVManifestation (described in 5.8)

3.3 NonAVCreation

Non-audiovisual creations in EFG can be pictures, photos, correspondence, books or periodicals. The descriptions of NonAVCreations and their associated manifestations follows an early version of the Europeana Semantic Elements specification. For the data exchange with the Europeana system mappings to the current ESE version will be constructed. Digital representations of non-film works available for viewing are accommodated in an associated instance of the item entity.

3.3.1 Elements

- **Identifier** (described in 4.1)
  In this context it is the identifier of the non-audiovisual creation.
  Status: Mandatory; Occurrences: 1

- **Title** (described in 4.2)
  Status: Mandatory; Occurrences: 1..n

- **Record Source** (described in 4.3)
  Status: Mandatory; Occurrences: 1..n

- **Type**
  The general type of the non-audiovisual creation.
  Value space: controlled vocabulary (e.g. "text", "image", "sound")
  Status: Mandatory; Occurrences: 1

- **Keywords** (described in 4.5)
  Status: Optional; Occurrences: 0..n

- **Description** (described in 4.6)
  Status: Optional; Occurrences: 0..n

- **Date**
  A point or period of time associated with an event in the lifecycle of the non-audiovisual creation.
  Value space: syntax follows ISO 8601 ("YYYY-MM-DD" or "YYYY")
  Status: Optional; Occurrences: 0..n

  Attributes:
  - Created
    Date of the creation of the non-audiovisual creation.
  - Issued
    Date of formal issuance (e.g. publication) of the non-audiovisual creation.

- **Language**
  The language of the major verbal content of the document.
  Value space: controlled vocabulary (ISO 639-1 based on RFC 4646)
  Status: Optional; Occurrences: 0..1

- **Coverage**
  The spatial or temporal topic of the non-audiovisual manifestation.
  Status: Optional; Occurrences: 0..n

  Attributes:
  - spatial
    This may be a named place, a location, a spatial coordinate or a named administrative entity.
    Value space: controlled vocabulary (e.g. "ISO-3166-1"")

- temporal
  This may be a period, date or range date.
  Value space: text or controlled syntax (ISO 8601 in form of YYYY-MM-DD)

- **Note**
  This element includes all data derived from content providers, that cannot fit into other elements.
  Value space: text
  Status: Optional; Occurrences: 0..n

  Attribute:

- Source
  This attribute specifies the source that produced the note element.
  Value space: text

- **UserTag**
  This is a tag created by the user through the EFG interface.
  Value space: text
  Status: Optional; Occurrences: 0..1

3.3.2 Allowed relationships

- NonAVCreationRELAgent (described in 5.1)
- NonAVCreationRELNonAVManifestation (described in 5.8)
- NonAVCreationRELNonAVCreation (described in 5.3)
- NonAVCreationRELEvent (described in 5.7)
- CollectionRELNonAVCreation (described in 5.9)
- AVCreationRELNonAVCreation (described in 5.6)

3.4 NonAVManifestation

This entity has the function to keep track of copies made of non-audiovisual objects (e.g. a letter gets scanned and then copied, etc.).

3.4.1 Elements

- **Identifier** (described in 4.1)
  In this context it is the identifier of the non-audiovisual manifestation.
  Status: Mandatory; Occurrences: 1

D 2.2 Common interoperability schema for archival resources and filmographic descriptions, report on the common interoperability schema

EFG
european film gateway

- **Title** (described in 4.2)
  Status: Mandatory; Occurrences: 1..n

- **Record Source** (described in 4.3)
  Status: Mandatory; Occurrences: 1..n

- **Type**
  The type of the NonAVManifestation.
  Value space: controlled vocabulary (e.g. "image", "text", "audio")
  Status: Mandatory; Occurrences: 1

- **Language**
  The language of the spoken, sung or written content.
  Value space: controlled vocabulary (ISO 639-1 based on RFC 4646)
  Status: Optional; Occurrences: 0..n

  Attribute:

  - <u>Language usage</u>
    This indicates the relationship between the language and the manifestation.
    Defaults to "unspecified".
    Value space: controlled vocabulary (e.g. "unspecified", "original spoken dialogue", "dubbing", "subtitles", "voice-over commentary", etc.)

- **Size**
  The total physical dimension of the manifestation represented as numeric value. Usually expressed in bytes.
  Status: Optional; Occurrences: 0..1

- **Digital format**
  The digital format of the manifestation.
  Value Space: Controlled vocabulary (Internet Media Types (MIME), e.g. "bmp", "jpg", "tiff", "pdf", "web site", "rss")
  Status: Optional; Occurrences: 0..1

- **Note**
  This element includes all data derived from content providers, that cannot fit into other elements.
  Value space: text
  Status: Optional; Occurrences: 0..n

*D 2.2 Common interoperability schema for archival resources
and filmographic descriptions, report on the common
interoperability schema*

Attribute:

- <u>Source</u>
  This attribute specifies the source that produced the note element.
  Value space: text

3.4.2 Allowed relationships

- NonAVManifestationRELAgent (described in 5.11)
- NonAVManifestationRELItem (described in 5.10)
- NonAVCreationRELNonAVManifestation (described in 5.8)
- EventRELNonAVManifestation (described in 5.12)

3.5 Item

The item entity points to the digital file held in the source repository. Since the item is not a part of the EFG database the Item entity will always contain a reference to an outside resource, either for immediate viewing or for ordering from an archive or distributor. The item functions as a logical wrapper for the digital object displayed in EFG. This entity is modelled according to the current version of the ESE.

3.5.1 Elements

- **Identifier** (described in 4.1)
  In this context the item identifier.
  Status: Mandatory; Occurrences: 1

- **isShownBy**
  An unambiguous URL reference to the digital object on the content provider´s web site in its full best available resolution/quality.
  Status: Mandatory; Occurrences: 1

- **isShownAt**
  An unambiguous URL reference to the digital object on the content provider´s web site in its full information context.
  Status: Mandatory; Occurrences: 1

- **Format**
  The file format of the viewing item.
  Value space: controlled vocabulary (Internet Media Types (MIME). In particular cases, format parameters relative to the MIME type should be included: e.g. resolution, unusual codecs required for viewing, etc.)
  Status: Mandatory; Occurrences: 1

- **Provider**
  Name of the organisation that sends the data to EFG.
  Value space: controlled vocabulary (e.g.)
  Status: Mandatory; Occurrences: 1

- **URI**
  An unambiguous URI to the resource within the EFG context. This is a record identifier for the object in the EFG system. It is created based on unique identifiers provided in the source metadata.
  Status: Optional; Occurrences: 0..1

- **Country**
  The name of the country in which the content provider is based (or "Europe" in case of Europe-wide projects).
  Value space: Controlled vocabulary (e.g. "ISO 3166-1")
  Status: Optional; Occurrences: 0..1

- **Note**
  This element includes all data derived from content providers, that cannot fit into other elements.
  Value space: text
  Status: Optional; Occurrences: 0..n

  Attribute:
  - Source
    This attribute specifies the source that produced the note element.
    Value space: text

3.5.2 Allowed relationships

- NonAVManifestationRELItem (described in 5.10)
- AVManifestationRELItem (described in 5.10)
- AgentRELItem (described in 5.5)

3.6 Agent

Agent is defined as an entity that can perform any kind of action. The EFG metadata model defines two agent types: Person and Corporate Body. As many content providers do not manage their agent data in authority files it is envisaged to store normalized and non-normalized agent metadata "side by side". For persons a common authority registry will be established. This entity was modelled to be compliant with the EAC standard that will be applied as an exchange format for authority data between EFG and the partners' local databases.

3.6.1 Elements (Agent type: Person)

- **Identifier** (described in 4.1)
  In this context it is the person identifier in the EFG database.
  Status: Mandatory; Occurrences: 1

- **Record Source** (described in 4.3)
  Status: Mandatory; Occurrences: 1..n

- **Person name**
  A name by which the person is (or was) known.
  Value space: e.g. preferred name (VN), bureaucratic name (BN), pseudonym (PN), spelling variant (SV), other name (WN)
  Status: Mandatory; Occurrences: 1..n

  Elements:
  - Part of person name
    A distinct part of a person's name.
    Value space: name prefixes, forename, family name, name suffixes
    Status: Mandatory; Occurrences: 1..n

  - Geographical scope of person name
    Indicates the region in which a particular name was used by the person.
    Value space: free text or controlled vocabulary (e.g. ISO 3166-1 country codes, except for historical geographic entities)
    Status: Optional; Occurrences: 0..1

  - Temporal scope of person name
    Time when the name was used by the person.
    Value space: Syntax follows ISO 8601 date range
    Status: Optional; Occurrences: 0..1

- **Date**
  Temporal properties of the person relating to its existence and activity.
  Value space: syntax follows ISO 8601 ("YYYY-MM-DD" or "YYYY")
  Status: Optional; Occurrences: 0..n

  Attribute:
  - Type
    This attribute specifies the type of temporal properties.
    Value space: defined by EAC (e.g. "date of birth", "date of death", "year of activity")

- **Region of Activity**
  A country from the geographic area in which the person is or was active.
  Value space: controlled vocabulary (e.g. "ISO 3166-1")
  Status: Optional; Occurrences: 0..1

- **Place**
  The places related to the person.
  Value space: text
  Status: Optional; Occurrences: 0..n

  Attribute:
  - Type
    This attribute specifies the type of place properties.
    Value space: defined by EAC (e.g. "place of birth", "place of death")

- **Sex**
  The person´s sex.
  Value space: controlled vocabulary
  Status: Mandatory; Occurrences: 1

- **Type of activity**
  A film-related activity of the person, taken from relationship records or from secondary sources.
  Value space: controlled vocabulary (e.g. "director", "actor", "screenplay", etc).
  Status: Optional; Occurrences: 0..n

- **Note**
  This element includes all data derived from content providers, that cannot fit into other elements.
  Status: Optional; Occurrences: 0..n

Attribute:

- Source
  This attribute specifies the source that produced the note element.

3.6.2 Elements (Agent type: Corporate Body)

- **Identifier** (described in 4.1)
  In this context the corporate body identifier in the EFG database.
  Status: Mandatory; Occurrences: 1

- **Record Source** (described in 4.3)
  A reference to the archive or other organisation supplying the record.
  Status: Mandatory; Occurrences: 1..n

- **Date**
  Temporal properties of the corporate body relating to its existence and activity.
  Value space: syntax follows ISO 8601 ("YYYY-MM-DD" or "YYYY")
  Status: Optional; Occurrences: 0..n

  Attribute:

  - Type
    This attribute specifies the type of temporal properties.
    Value space: by EAC (e.g. "founded", "dissolved")

- **Region of Activity**
  A country from the geographic area in which the corporate body is or was active.
  Value space: controlled vocabulary (e.g. "ISO 3166-1")
  Status: Optional; Occurrences: 0..1

- **Corporate body name**
  A name by which the corporate body is or was known.
  Controlled vocabulary: e.g. preferred name (VN), bureaucratic name (BN), pseudonym (PN), spelling variant (SV), other name (WN)
  Status: Mandatory; Occurrences: 1..n

  Elements:

  - Part of the corporate body name
    A distinct part of a corporate body´s name.
    Status: Optional; Occurrences: 0..n

- Geographical scope
  Indicates the region in which a particular name was used by the corporate body.
  Value space: text or controlled vocabulary (e.g. "ISO 3166-1" country codes, except for historical geographic entities)
  Status: Optional; Occurrences: 0..1

- Temporal scope
  Time when the name was used by the corporate body.
  Value space: Syntax follows ISO 8601 date range
  Status: Optional; Occurrences: 0..1

- **Type of activity**
  A film-related activity of the corporate body, taken from relationship records or from secondary sources.
  Value space: controlled vocabulary (e.g. "distribution company", "production company")
  Status: Optional; Occurrences: 0..n

- **Note**
  This element includes all data derived from content providers, that cannot fit into other elements.
  Status: Optional; Occurrences: 0..n

  Attribute:
  - Source
    This attribute specifies the source that produced the note element.

3.6.3 Allowed relationships

- AgentRELAgent (described in 5.4)
- AgentRELItem (described in 5.5)
- NonAVManifestationRELAgent (described in 5.11)
- AVManifestationRELAgent (described in 5.11)
- NonAVCreationRELAgent (described in 5.1)
- AVCreationRELAgent (described in 5.1)
- EventRELAgent (described in 5.13)

3.7 Event

Events are defined as primary entities that can occur within the lifecycle of an audiovisual or non-audiovisual creation. This entity was modelled by using EN 15907 (clause 5.2). The EFG metadata model distinguishes between the following event types:

- **Publication event:** This can be a public screening, broadcast or a public release on a physical distribution medium or online.
- **Decision event:** An event in which a manifestation of a creation was evaluated by a censorship body or an accredited rating agency.
- **IPR registration:** A registration of intellectual property rights.
- **Award:** A bestowal of an award relating to the audiovisual creation or to an achievement by an Agent in the context of an audiovisual creation.
- **Production event:** A specific event in the course of production of a creation. Examples are dates and locations where castings took place or dates and locations of shootings or other recordings.

3.7.2 Elements (Event type: Publication Event)

- **Identifier** (described in 4.1)
  Status: Mandatory; Occurrences: 1

- **Source ID**
  An identification for the event record, if this exists in the database from which the filmographic record was published, else the value is "undefined".
  Status: Mandatory; Occurrences: 1..n

- **Type**
  The type of an exhibition event (e.g. "Premiere", "Cinema release") or type of publication ("e.g. Paperback book with DVD") if not evident with the type of associated audiovisual manifestation.
  Value space: free text or controlled vocabulary
  Status: Mandatory; Occurrences: 1

- **Date**
  The date on which the exhibition took place or on which copies of the manifestation became available for hire.
  Value space: Syntax follows ISO 8601 (YYYY-MM-DD)
  Status: Optional; Occurrences: 0..1

- **Regional Scope**
  The country or large-scale geographic entity where the event took place (e.g. exhibition) or where published items have been made available (e.g. distribution area)
  Value space: controlled vocabulary (e.g. "ISO-3166-1")
  Status: Optional; Occurrences: 0..1

- **Place**
  The name of the city of smaller geographic entity (e.g. building) where the event took place.
  Value space: text
  Status: Optional; Occurrences: 0..1

- **Event Name**
  The name of an event that the exhibition was a part of (e.g. name of a film festival, distribution channel of a broadcaster, etc.).
  Value space: text
  Status: Optional; Occurrences: 0..1

- **Exhibition organiser**
  The name of an Agent that was responsible for the exhibition (e.g. name of a cinema company, film club, broadcaster, etc.).
  Value space: text
  Status: Optional; Occurrences: 0..n

- **Publisher**
  If not evident from an association of Agent with Manifestation, this element can be used to state the name of the publisher.
  Status: Optional; Occurrences: 0..n

- **Access conditions**
  Specific restrictions for accessing the content of the manifestation (e.g. "for rental only").
  Status: Optional; Occurrences: 0..n

3.7.3 Elements (Event type: Decision Event)

- **Identifier** (described in 4.1)
  Status: Mandatory; Occurrences: 1

- **Source ID**
  An identification for the event record, if this exists in the database from which the filmographic record was published, else the value is "undefined".
  Status: Mandatory; Occurrences: 1..n

- **Type**
  Types include e.g. censorship decisions, rating decisions etc.
  Value space: free text or controlled vocabulary
  Status: Mandatory; Occurrences: 1

- **Date**
  The date on which the verdict was announced or from which on the verdict was declared valid.
  Value space: syntax follows ISO 8601 (YYYY-MM-DD)
  Status: Optional; Occurrences: 0..1

- **Regional scope**
  The country or large-scale geographic entity for which the decision is (was) valid.
  Value space: controlled vocabulary (e.g. "ISO-3166-1")
  Status: Optional; Occurrences: 0..n

- **Agency**
  The name of the agency performing the rating or censorship.
  Value space: text
  Status: Optional; Occurrences: 0..n

- **Filing number**
  An identifier issued by the agency uniquely identifying the act of rating or censorship and associated documents such as censorship visa or rating certificates.
  Status: Optional; Occurrences: 0..1

- **Verdict**
  The outcome of the act of rating or censorship.
  Value space: text
  Status: Optional; Occurrences: 0..1

3.7.4 Elements (Event type: IPR Registration)

- **Identifier** (described in 4.1)
  Status: Mandatory; Occurrences: 1

- **Source ID**
  An identification for the event record, if this exists in the database from which the filmographic record was published, else the value is "undefined".
  Status: Mandatory; Occurrences: 1..n

- **Date**
  The date on which the registration was filed or the date on which registration became effective.
  Value space: Syntax follows ISO 8601 (YYYY-MM-DD) (if no value is available in the source archive, the value must be set to a default value)
  Status: Mandatory; Occurrences: 1

- **Registration agency**
  Name of agency issuing the registration certificate (e.g. "UK Intellectual Property Office").
  Value space: text (if no value is available in the source archive, the value must be set to ""not available in the source archive")
  Status: Mandatory; Occurrences: 1

- **Regional scope**
  The geographic region for which copyright is claimed.
  Value space: controlled vocabulary (e.g. "ISO 3166-2", "AFNOR XP Z44" for historical countries, "MARC" for extraterrestrial regions)
  Status: Optional; Occurrences: 0..n

- **Name of applicant**
  Name of the Agent claiming copyright in the audiovisual or non-audiovisual creation.
  Value space: free text or retrieved from an Agent relationship
  Status: Optional; Occurrences: 0..n

3.7.5 Elements (Event type: Award)

- **Identifier** (described in 4.1)
  Status: Mandatory; Occurrences: 1

- **Source ID**
  An identification for the event record, if this exists in the database from which the filmographic record was published, else the value is "undefined".
  Status: Mandatory; Occurrences: 1..n

- **Date**
  The date the award was bestowed to an Agent.
  Value space: Syntax follows ISO 8601 in form of YYYY-MM-DD (if no value is available in the source archive, the value must be set to a default value)
  Status: Mandatory; Occurrences: 1

- **Nomination only**
  If the audiovisual creation was nominated but not among the winners, then this element should be set to the logical value "true".
  Value space: "true" or "false"
  Status: Optional; Occurrences: 0..1

- **Award Name**
  The name of the award or trophy, possibly including a numeric designation (e.g. 2nd Price).
  Value space: text (if no value is available in the source archive, the value must be set to "not available in the source archive")
  Status: Mandatory; Occurrences: 1

- **Achievement**
  A phrase describing a specific achievement for which the award was given, if not for the audiovisual creation in total.
  Value space: text
  Status: Optional; Occurrences: 0..1

- **Sponsor**
  Name of the Agent(s) that have sponsored the award.
  Value space: free text or retrieved from an Agent relationship
  Status: Optional; Occurrences: 0..n

3.7.6 Elements (Event type: Production event)

- **Identifier** (described in 4.1)
  Status: Mandatory; Occurrences: 1

- **Type**
  The type of event this element instance refers to. Examples are "outdoor shooting", "indoor shooting", "post-production", etc.
  Value space: free text or controlled vocabulary
  Status: Optional; Occurrences: 0..1

- **Location**
  Any geographic name or address of the location where the event took place.
  Value space: text
  Status: Optional; Occurrences: 0..n

- **Regional scope**
  The country or other large-scale geographic entity where the event took place.
  Value space: free text or controlled vocabulary
  Status: Optional; Occurrences: 0..n

- **Date**
  The date or time-span during which the event took place.
  Value space: Syntax follows ISO 8601 (YYYY-MM-DD or date range)
  Status: Optional; Occurrences: 0..n

3.7.7 Allowed relationships

- EventRELAgent (described in 5.13)
- EventRELAVManifestation(described in 5.12)
- EventRELNonAVManifestation(described in 5.12)
- NonAVCreationRELEvent (described in 5.7)
- AVCreationRELEvent (described 5.7)

3.8 Collection

In the EFG context, a collection is defined by a compilation of creation (audiovisual or non-audiovisual). This entity functions like a "basket" of instances from EFG metadata elements that can be assigned by users. It does not refer to the notion of collection professionally curated in an archive.

3.8.1 Elements

- **Identifier** (described in 4.1)
  Status: Mandatory; Occurrences: 1

- **Collection Title**
  A textual title of the collection
  Value space: text
  Status: Mandatory; Occurrences: 1

- **Description** (described in 4.6)
  Status: Optional; Occurrences: 0..n

3.8.2 Allowed relationships

- CollectionRELAVCreation (described in 5.9)
- CollectionRELNonAVCreation (described in 5.9)

4 Common Elements

These are elements that occur as part of more than one major entity. This chapter defines all common elements of the EFG metadata model. Moreover, attributes and proposals for value spaces are given as in the previous chapter.

4.1 Identifier

An unambiguous reference to the resource within a given context.
Value space: GUID or ID chosen from an external naming schema

Attributes:

- **scheme**
  An unambiguous reference of scope within which the identifier is unique. References can be taken from an internationally known identification scheme, an XML namespace identifier, or a URN namespace (e.g. URN: UUID).
  Status: Mandatory; Occurrences: 1

4.2 Title

Titles can occur several levels of description and can change during the lifecycle of a creation. Titles can be a word, phrase or character, naming the work, a group of works, a particular manifestation or an individual item. Film titles in EFG will be managed through a common authority registry. The title element is modelled after the EN 15907 (clause 6.3).

Attribute:

- **Language**
  The language of the title.
  Value space: controlled vocabulary (ISO 639-1 based on RFC 4646; if no value is available in the source archive, the value must be set to a default value)
  Status: Mandatory; Occurrences: 1

Elements:

- **Title Text**
  The textual expression of the title.
  Status: Mandatory; Occurrences: 1

- **Part designation**
  A combination of the name of a structuring unit and the count value that identifies the current creation as an individual part of a complex work. The element "Part Designation" is used only for the audiovisual and non-audiovisual creation and not for manifestations.
  Status: Optional; Occurrences: 0..1

  Attributes:

  - unit
    The name of the unit that represents within a series, serial, or multi-part work.
    Value space: controlled vocabulary (e.g. "part", "episode", "issue")

  - value
    An ordinal number, numeration symbol, or date value that identifies the individual creation within a complex work.
    Value space: text

- **Title relationship**
  The type of relationship between the title and the entity to which it is assigned.
  Value space: controlled vocabulary (e.g. "original title", "working title", "distribution title" etc.)
  Status: Optional; Occurrences: 0..1

- **Temporal scope**
  Date when a change of a title occurred.
  Value space: Syntax follows ISO 8601 (YYYY-MM-DD)
  Status: Optional; Occurrences: 0..1

- **Geographic scope**
  Indication to which region(s) the title applies if a work was distributed under different titles in distinct geographic regions.
  Status: Optional; Occurrences: 0..n

  Elements:
  - Code
    Region code.
    Value space: controlled vocabulary (e.g. "ISO-3166-2", "AFNOR XP Z44" for historical countries, "MARC" for extraterrestrial regions)
    Status: Optional; Occurrences: 0..n

- region name
  Name of a region other than a country.
  Value space: controlled vocabulary
  Status: Optional; Occurrences: 0..n

4.3 Record Source

A reference to the EFG content provider.

Elements:
- **Source**
  The name of the archive supplying the record.
  Value space: controlled vocabulary
  Status: Mandatory; Occurrences: 1

- **Source identifier**
  An unambiguous reference to the archive supplying the record.
  Value Space: URI or a registered namespace identifier (e.g. ISIL - ISO 15511)
  Status: Mandatory; Occurrences: 1

4.4 Format

The description of the physical artefact or digital file on which an audiovisual manifestation is fixed.

Elements:
- **Digital format**
  The format of the digital file.
  Status: Optional; Occurrences: 0..1

  Elements:
  - Container
    A container or wrapper format is a file format, or often a stream format whose specifications regard only the way data are stored but *not* coded within the file.
    Value space: controlled vocabulary (e.g. "AVI", "MP4", "3GP", "RealMedia")
    Status: Optional; Occurrences: 0..1

- Coding
  Specifies the codification of the data.
  Value space: controlled vocabulary (Internet Media Types (MIME), e.g. "WMA","WMV", "MPEG-4", "RealVideo")
  Status: Optional; Occurrences: 0..1

- **Carrier type**
  The type of physical carrier used for the manifestation.
  Value space: controlled vocabulary (e.g. "DVD", "online streaming media", "podcasts")
  Status: Optional; Occurrences: 0..1

- **Gauge**
  The width of the film stock or other carrier used for the manifestation.
  Value space: controlled vocabulary (e.g. "35 mm", "1/2 in", etc.).
  Status: Optional; Occurrences: 0..1

- **Aspect ratio**
  The ratio between width and height of the image.
  Value space: text (numeric values or names, e.g. "full frame", "Cinemascope")
  Status: Optional; Occurrences: 0..1

- **Sound**
  The name of the system by which sound is recorded either on the carrier, on a separate medium, or as part of the digital encoding (e.g. Western Electric (Westrex) Movietone, RCA Photophone, Tobis, Dolby Digital, etc.).
  Status: Optional ; Occurrences: 0..1

  Attributes*:*
  - hasSound
    A boolean value indicating if the manifestation includes recorded sound
    Value space: "true", "false"
  - Recording system
    A boolean value indicating if the system and/or method was used for the primary recording.
    Value space: "true", "false"

- **Colour**
  The name of the colour system or process. For digital formats the name of the colour space. For analogue video the name of video colour system.
  Value space: text
  Status: Optional; Occurrences: 0..1

4.5 Keywords

A term or set of vocabulary terms describing the content of an audiovisual or non-audiovisual creation. Controlled and uncontrolled terms can be used together, but not within a single set of terms. Likewise, if more than one controlled vocabulary is used, then terms from each of these must be contained in a separate instance of this element. A separate instance is also required for each language if terms in more than one language are taken from a multilingual vocabulary.

Attributes:

- **Type**
  Type of information described by the keywords.
  Value space: controlled vocabulary (e.g. "subject", "genre", "form")

- **Language**
  The language of the content of each subject. Can be omitted if the language is evident from the scheme.
  Value space: controlled vocabulary (ISO 639-1 based on RFC 4646)

- **Scheme**
  A unique identifier denoting the controlled vocabulary (preferably URI). If the subject terms are not from a controlled vocabulary, the value of this element should be set to "uncontrolled".
  Value space: URI, ID or "uncontrolled"

Elements:

- **Term**
  An element containing a single term. This can be the textual value of the term. For non-textual terms the classification codes, preferably a combination of the code and the verbal class description should be indicated.
  Value space: text or controlled vocabulary (If no value is available in the source archive, the value must be set to ""not available in the source archive".)*
  Status: Mandatory; Occurrences: 1..n

  Attributes*:*

- Term ID
  A non-text identifier that can be combined with scheme ID to from a unique resource identifier for the term within a controlled vocabulary.

4.6 Description

Textual descriptions include synopses, plot summaries, reviews, transcripts or shot lists. They can occur in more than one language and they can have statements of authorship or references to external resources.

Attributes:

- **Language**
  The language of the content of description text.
  Value space: controlled vocabulary (ISO 639-1 based on RFC 4646)

- **Type**
  A keyword denoting the type of description.
  Value space: controlled vocabulary (e.g. "Synopsis", "Shotlist", "Review" etc.)

- **Source**
  Either the name of the description author, or a URI identifying the source directly or via a reference system such as an on-line catalogue.
  Value space: text or URI

4.7 Segmentation

A segment is defined as a sequential part of a time-based medium, thus of an audiovisual manifestation. In contrast to semantically defined sequences (e.g. news item in a newsreel) this entity allows for structural divisions of audio-visual content (e.g. time points in a media stream from which frames have been taken). This element is subject to further specification. Segments should at least have an individual identifier (ID), a start and an end frame but also parts of the MPEG-7[5] standard could be applied.

---

[5] ISO 15938. URLs to the documentation of this metadata standard:
http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm;
http://www.m4if.org/resources.php#Section40

5 Relation Definition

This chapter specifies the relationships that can occur between entities. For each relation the cardinality restriction is indicated. Moreover, this chapter lists the relationships with a type attribute for values that define the semantics of the relationships. These values will be taken from controlled vocabularies which specify a forward and inverse form of the relationship, and possible restrictions depending on the classes of entities for which the relationship is valid.

5.1 AVCreation REL Agent ; NonAVCreation REL Agent; AVManifestation REL Agent ; NonAVManifestation REL Agent

An agent could be an actor and at the same time a director, but it could happen for an AVCreation he is just one of this. This relation, using the type attribute, specifies which relationship exists between a particular Agent and an AVCreation/NonAVCreation or an AVManifestation/NonAVManifestation.
Cardinality: n..m

Attribute:

- **Type**
  A keyword denoting the type of relation between the
  AVCreation/AVManifestation/NonAVCreation/NonAVManifestation and the Agent.
  Value Space: Type vocabulary
  Status: Mandatory; Occurrences: 1

5.2 AVCreation REL AVCreation

An AVCreation could be a part of another AVCreation or the two could share the same content. This relation, using the type attribute, specifies which relationship exists between a particular AVCreation and another one.
Cardinality: n..m

Attribute:

- **Type**
  A keyword denoting the type of relation between an AVCreation and another.
  Value Space: Type vocabulary
  Status: Mandatory; Occurrences: 1

5.3 NonAVCreation REL NonAVCreation

A NonAVCreation could be a part of another NonAVCreation or the two could share the same content. This relation, using the type attribute, specifies which relationship exists between a particular NonAVCreation and another one.
Cardinality: n..m

Attribute:

- **Type**
  A keyword denoting the type of relation between a NonAVCreation and another NonAVCreation.
  Value Space: Type vocabulary
  Status: Mandatory; Occurrences: 1

5.4 Agent REL Agent

An Agent could be a parent of another Agent or the two could be the owner of a company. This RelatedToAgent, using the type attribute, specifies which relationship exists between a particular Agent and another one.
Cardinality: n..m

Attribute:

- **Type**
  A keyword denoting the type of relation between an Agent and another one.
  Value Space: Type vocabulary
  Status: Mandatory; Occurrences: 1

5.5 Agent REL Item

An agent could be an actor and at the same time a director, but it could happen for an Item he is just one of this. This relation, using the type attribute, specifies which relationship exists between a particular Agent and an Item.
Cardinality: n..m

Attribute:

- **Type**
  A keyword denoting the type of relation between the
  Item and the Agent.
  Value Space: Type vocabulary
  Status: Mandatory; Occurrences: 1

5.6 AVCreation REL NonAVCreation

An AVCreation could have documentation related, as article on newspaper or poster. This relation, using the type attribute, specifies which relationship exists between a particular AVCreation and a NonAVCreation.
Cardinality: n..m

Attribute:

- **Type**
  A Keywords denoting the type of relation between the AVCreation and a NonAVCreation.
  Value Space: Type vocabulary
  Status: Mandatory; Occurrences: 1

5.7 AVCreation REL Event; NonAVCreation REL Event

An AVCreation or a NonAVCreation can be related to a specific Event. This relationship has no attribute.
Cardinality: n..m

5.8   AVCreation   REL   AVManifestation;   NonAVCreation   REL NonAVManifestation

This relationship relates an AVCreation/NonAVCreation to its AVManifestation/ NonAVManifestation. This relationship has no attribute.
Cardinality: 1..n

5.9 Collection REL AVCreation; Collection REL NonAVCreation

This relationship relates an a Collection to the AVCreation/NonAVCreation that compose it. This relationship has no attribute.
Cardinality: n..m

5.10 AVManifestation REL Item; NonAVManifestation REL Item

This relationship relates an AVManifestation/NonAVManifestation to the Item related. This relationship has no attribute.
Cardinality: 1..n

5.11 AVManifestation REL Agent; NonAVManifestation REL Agent

This relationship relates an AVManifestation/NonAVManifestation to the Agent related. This relationship has no attribute.
Cardinality: n..m

5.12 Event REL AVManifestation; Event REL NonAVManifestation

This relationship relates an Event with an AVManifestation/NonAVManifestation related. This relationship has no attribute.
Cardinality: n..m

5.13 Event REL Agent

This relationship relates an Event with an Agent related. This relationship has no attribute.
Cardinality: n..m