

# Age label data model – Discussion paper

Sep 2<sup>nd</sup> 2013

## Background

Starting from the insight that currently harmonisation of classification schemes in European Member States is neither desirable nor feasible due to different socio-cultural contexts of classification, technical interoperability between the existing schemes and their electronic labels is seen as a possible way ahead to optimise the efficiency of machine-readable age labels and their comprised rating knowledge beyond national borders, to extend availability of online classification labels as well as to wider and more innovative ways of use of user-side information tools and filter software in general.

As these outcomes are important parts of the [CEO Coalition](#)'s tasks, the consortium of companies decided to establish a task force that focuses on proposing an interoperable data model for existing classification schemes and, by doing so, aiming at achieving better interoperability of classification data and electronic age labels in practice.

However, the relevance of interoperable labels for companies and users differ, depending on both the context of system environments (closed gardens: low relevance, open internet: high relevance) and the business case including the respective data transfers (B2B: probably lower relevance, B2C: high relevance). This led to a legitimate variance in company interest among the Coalition members, while other presumably more relevant stakeholders staid out of the debate, especially rating bodies, filter software providers, additional content providers, consumer associations, family associations and so on. The Task Force plans to incorporate those bodies and/or their opinions in the discussion in 2013.

## Objective of the task force

The main objective of the task force is to propose a technology-neutral data model for electronic content labels including agreed categories and fields that may contain content-specific information. The proposal is planned to include documentation, code snippet examples and probable queries to support implementing the data model in existing classification contexts. In a second step it will discuss by what means (interoperable standards, program languages, etc.) machine-readability of interoperable electronic labels can be implemented and plans to publish a proposal for implementation options in different technical contexts. Both proposals are planned to serve as guidelines for either existing players to implement the data model in their existing schemes or for new players that plan to label online content and thus reduce the risk of sunk costs.

# A data model for online labels

## Basic principles of a data model

The data model has to build on currently existing practices, as it otherwise would undermine the efforts already taken by both companies and rating bodies as well as the classification knowledge that goes with such labels. For companies and bodies that already label online content electronically, no disadvantages should result from the proposals made.

The three basic requirements the data model will therefore take into account are:

- (1) The data model has to be **technology-neutral** to reach maximum openness and compatibility between different systems and languages.
- (2) It has to **consider existing electronic labelling systems** to ensure that these are not undermined by the interoperable data model.
- (3) It has to thoroughly **take into account existing national and supranational classification schemes**. By doing so, existing visual labels can easily be extended by respective electronic labels while at the same time ensuring compatibility with the data model.

One fundamental principle of the data model is that neither existing approaches and schemes nor future ones have to provide information in *all* categories – as long as the data that is provided by the label does fit into any of the categories, the system is technically interoperable. However, the more information a system or label provides, the better other systems will be able to use and process the data.

## Categories and fields of a data model

Main categories of data fields within the data model that have been agreed on are age labels, content descriptors and information about the body issuing the age classification. While most of the single fields are within these categories will become the focus of the task force's attention, as well as their openness or possible variables.

### Category: age label

Age labels are a common approach in content rating systems worldwide. However, there are different schemes of how to provide the age information (specific age, age group or additional age information like parental guidance).

Field	Possible values	Comments
rating_age	Numeric	Minimum age that is suitable for the content.  Most age classification systems already use numeric values. However, systems that do not will have to provide translation tables to provide numeric values here (e.g. US-ESRB: E→0; E10+→10; T→13; M→17; A→18).  <i>This field is mandatory.</i>
rating_additional	shortText	If a system uses additional values to specify an age classification, these additional age information

		<p>has to go here, e.g. PG, R18, 12A, XXX etc.</p> <p>The Task Force will have to discuss whether systems that rely on non-numeric classification (e.g. ESRB) can provide their original rating in this field, too.</p>
rating_icon	URL	<p>To provide trustful classifications, this field provides the URL to the original age rating icon that can be used in cases where a visual age rating is deemed necessary.</p> <p>The Task Force members check whether there are systems that use two or more icons for displaying the age rating. In this case, the group will discuss possibilities to extend this data field.</p>

As numeric values are superior for machine-based processing than open text fields, it seems more feasible to translate textual age classification like „teens“ or „mature“ into numbers on side of the classification scheme (and its API) rather than to translate textual values in the data model into number on processor-/client-side. An additional field might be necessary to cope with age-label-specific additional information, e.g. „parental guidance“. To refer to rating bodies' icons technically within a label might offer to transfer trust in a more visual way – here, however, the task force will have to identify potential IP-related issues.

### **Content descriptors**

Many existing classification schemes use content descriptors to give additional information about the content and the reasons for the respective age rating. The categories of these descriptors are quite comparable worldwide. However, there are and always will be peculiarities of single systems or schemes, requesting a flexible approach of the data model. The Task Force decided to opt for a mixed data field model here, where common and agreed categories are predefined, while the category allows for providing additional content descriptor fields for the sake of flexibility. However, it already is foreseeable that the uptake of such additional field will be comparably slow, since fragmented forms of content categories will result in the loss of synergy effects of such an interoperable model. To minimise these effects it is being planned to monitor the additional or new content categories and regularly decide on potential new fields that will become pre-defined..

In this category, too, icons are commonly used as content descriptors. The data model has to provide URL references to these icons, too, as long as there are no severe IP or brand issues.

If a pre-defined content descriptor is applicable, the allowed values are yes/no (or 1/0) only. The Task Force has opted against scalable values (low, medium, strong) as these comparatives are heavily depending on the region and classification context and are deemed more confusing for consumers than helpful.

*Proposal for pre-defined data fields (standardised content descriptors)*

<b>Field name</b>	<b>Possible values</b>	<b>Comments</b>
descriptor_content_sexuality	yes, no / 1, 0	
descriptor_content_sexuality_icon	URL	Address of original sex/erotic/nudeness icon
descriptor_content_violence	yes, no / 1, 0	
descriptor_content_violence_icon	URL	Address of original violence/weapons/blood icon
descriptor_content_discrimination	yes, no / 1, 0	
descriptor_content_discrimination_icon	URL	Address of original discrimination/ racism/ hate speech icon
descriptor_content_cursing	yes, no / 1, 0	
descriptor_content_cursing_icon	URL	Address of original obscene/bad language/cursing icon
descriptor_content_drugs	yes, no / 1, 0	
descriptor_content_drugs_icon	URL	Address of original drugs/tobacco/alcohol icon
descriptor_content_fear	yes, no / 1, 0	
descriptor_content_fear_icon	URL	Address of original fear/shock icon
descriptor_content_gambling	yes, no / 1, 0	
descriptor_content_gambling_icon	URL	Address of original gambling icon

If new relevant content categories emerge or a system is making use of other content descriptors than the ones pre-defined by the data model, it is possible to provide additional content descriptors in a specified way.

*Example for additional data fields (additional content descriptors)*

<b>Field name</b>	<b>Possible values</b>
descriptor_content_younameit (e.g. descriptor_content_selfharm; descriptor_content_antisocial)	yes, no

descriptor_content_younameit_desc	shortText (short description of additional content descriptor)
descriptor_younameit_icon	Address of original additional icon

Some systems do not provide systematic content descriptors, but offer additional information regarding the reasoning for a specific age rating in text form. Such information is harder to structure and to process technically but it still provides relevant information. Later on, systems will be able to process this data automatically, too. As some systems use such open textual descriptions (e.g. BBFC), the data model will have take into account such fields, too.

Field name	Possible values
descriptor_content_opentext	shortText

A third type of content descriptors relate to information about features or functionalities that the content (or better: the content-related service or platform) provides to the user. It is being agreed that such features might pose risks for minors, too. E.g. user generated content might contain relevant depictions that would change existing age classifications, chat functionalities result in unknown people approaching the underage user in a harmful way or location-based services log and display the movement and/or other person-related information to third parties. PEGI already started to implement such descriptors, hence a first step could be to take those as predefined ones, while leaving the feature descriptor field open to new ones, too.

*Proposal for pre-defined data fields regarding features (standardised feature descriptors)*

Field name	Possible values	Comments
descriptor_feature_inapppurchase	yes, no / 1, 0	The service contains elements enabling the consumer to purchase additional content or functionality, regardless of whether the app itself was acquired for free or not.
descriptor_feature_inapppurchase_icon	URL	Address of original icon for in-app purchase features.
descriptor_feature_personaldatasharing	yes, no / 1, 0	The service gives its provider (or a third party) access to personal data such as home address, contact

		details or bank account numbers.
descriptor_feature_personaldatasharing_icon	URL	Address of original icon for personal data sharing features.
descriptor_feature_locationdatasharing	yes, no / 1, 0	The service contains the option to share exact location on a map when using the service. The location information may be shared publicly or with a specific network inside the service or elsewhere online.
descriptor_feature_locationdatasharing_icon	URL	Address of original icon for location data sharing features.
descriptor_feature_chat	yes, no / 1, 0	The service includes an option for a user to chat with other users of the app. These users may operate under a pseudonym or anonymously.
descriptor_feature_chat_icon	URL	Address of original icon for chat features.

Similarly to the content descriptors, additional feature descriptors will emerge during time. Hence, the data model should be open to new descriptors.

*Example for additional feature fields (additional feature descriptors)*

<b>Field name</b>	<b>Possible values</b>
descriptor_feature_younameit (e.g. descriptor_feature_upload)	yes, no
descriptor_additional_younameit_desc	shortText (short description of additional feature descriptor)
descriptor_additional_younameit_icon	Address of original additional icon

### **Issuing body and type of classification system**

For all content ratings, it is important to refer to the body that issued the specific label. First, this information links the age label and content descriptors to brand or

institutional trust. Moreover, this category allows the assessment of the geographical origin of the label and its potential legal relevance. Also, additional information on the type of classification behind the rating procedure can be an important asset when it comes to trust. By providing the date of the classification, the label can also show the actuality of a rating.

<b>Field name</b>	<b>Possible values</b>	<b>Comments</b>
origin_body	shortText (FSM, PEGI, NICAM, USK, FSK, „own“, „Company Name“ etc.)	Since age classification can be based on self-classification, this field has to be open for all kinds of issuing bodies.  <i>This field is mandatory. (to be discussed)</i>
origin_body_url	URL	Reference to address of issuing body with additional information on institutional background and rating procedures.
origin_type	formal, owner, user-generated (to be discussed)	To give a pre-defined information on the institutional setting and procedural form of classification decisions, a categorisation could significantly increase machine-readable trust-based configurations.  The Task Force will discuss if this form of categorised information on the age classification procedure is (a) necessary, (b) feasible and (c) what categories are deemed suitable (additional suggestions: rating-body; producer; provider; user). Ideas are welcomed.
origin_date	Date (e.g. 2013-04-24)	The date of the most recent decision on an age classification.

### Label metadata (details to be discussed)

During the Task Force's discussions the need for the category „label metadata“ has become obvious: Depending on the form of age classification data (data dump that for instance can be applied to external data later-on vs. a label attached to a specific online content), each piece of interoperable classification information has to provide information that refers to the specific content the information applies to („scope“). This scope might either aim at specific, isolated media content that – sometimes in a nation- or region-specific version - has already been classified or it aims at a list of categories and sub-categories of a website or service.

The problem relating to the scope of application of a label is that its form depends on the context of implementation: For labels that are provided with the content, the scope has to clarify to what parts of the content the label applies. As the content comes together with the label, there is no need to provide any unique identifiers to link the label to the content. The issue here is to clarify the scope of the label, any exemptions or overriding special cases. Since most content and services online rely on URI structures, the metadata for these kinds of labels will be URI-based, too. The basic approach will be to take the age label for the whole URI authority, usually the second level domain. However, possibilities for URI-based exemptions (specific paths or folders) have to be possible.

Contrary to content-wise attached labels another form of providing classification data is to offer centralised databases with thousands of classifications. Usually, existing rating bodies will opt for such forms of data provisions. The issue here is that for instance an online shop or a VOD service wants to query the database for valid age classification information. To get the correct information out of the database both the data provider as well as the demander will have to use unique identifiers. As the label is detached from the content, the database query has to aim at getting the one correct information back from the database. In practice, such UIDs are not being used throughout all existing rating bodies. And even if a rating body uses UIDs, they aren't the same among different rating bodies. It is thus foreseeable that an UID field is necessary, but hard to implement in practice. An alternative can be to base queries on the title of the media content (e.g. movie title, game title). This might, however, lead to several results as in many cases different versions of a game or film have been classified by a rating body (cinematic version, DVD version, TV version etc.). Both approaches regarding this form of detached classification information have drawbacks and will have to be discussed in the Task Force.

#### *Proposal for metadata fields*

<b>Field name</b>	<b>Possible values</b>	<b>Comments</b>
metadata_scope	URL	FQDN-based scope of application of a label (not suitable for central databases of classification information that is detached from the content)



metadata_UID	Numeric	UID of classified content (system-specific, probably not suitable for content-wise attached labels)
metadata_title	shortText	Title of classified content (system-specific)

### **Additional field proposals (to be discussed)**

- Positive labels  
The Task Force will have to discuss if and where to put a data field that positively shows child-oriented content (positive labels). Two places that seem suitable are either within the age label category, where a specific “child-oriented content” shows that content is programmed in a way to suit children’s needs, or within the content descriptor category, where such a specific label could be provided.
- Regarding a proposed “alternate tag”, e.g. a redirect URL in case of blocking by a user-side parental software, the Task Force had the impression that this will depend on each specific implementation and thus should not be described by the data model. However, it might be an option to provide such a field as an optional data field, too (to be discussed).
- The same goes for a data field that holds information on the variableness of a specific age classification: Even if a field “revisit-after-days” seems suitable, its application will heavily depend on the context. The Task Force will discuss whether the data model proposal will include such a field.