A Unified Model for Socially Interconnected Multimedia-Enriched Objects

Theodora Tsikrika, Katerina Andreadou, Anastasia Moumtzidou, Emmanouil Schinas, Symeon Papadopoulos, Stefanos Vrochidis, and Ioannis Kompatsiaris

Information Technologies Institute, CERTH, Thessaloniki, Greece

Abstract. Enabling effective multimedia information processing, analysis, and access applications in online social multimedia settings requires data representation models that capture a broad range of the characteristics of such environments and ensure interoperability. We propose a flexible model for describing Socially Interconnected MultiMedia-enriched Objects (SIMMO) that integrates in a unified manner the representation of multimedia and social features in online environments. Its specification is based on a set of identified requirements and its expressive power is illustrated using several diverse examples. Finally, a comparison of SIMMO with existing approaches demonstrates its unique features.

Keywords: multimedia description, social interactions, interoperability

1 Introduction

The massification of web publishing, together with the proliferation of social media hosting and sharing platforms, and the widespread use of mobile devices have led to an unceasing generation of large volumes of online content objects, such as Web pages and social media posts. Such objects are enriched with multimedia and are interconnected in various ways, including through social interactions. To satisfy the information needs of the diverse users in such online social multimedia environments (ranging from laypeople simply searching for information to media monitoring professionals) requires effective multimedia information processing, analysis, and access applications that support tasks such as clustering, classification, summarisation, search, recommendation, and retrieval. To enable such applications, there is a need to employ data representation models that capture a broad range of the characteristics of such settings and ensure interoperability across diverse multimedia objects, hosting services, and tasks.

Existing models for the description of multimedia content and structure, such as MPEG-7 [5], RUCoD [6], and WebLab [7], do not take into account the social characteristics and interconnections in current web settings, since most were developed prior to the advent of social Web or/and for different purposes. On the other hand, more recent approaches that capture such social aspects, such as the SIOC [1] and FOAF [2] ontologies, do not consider the potential multimodality of online content and the variety of its extracted annotations.

This work aims to bridge this gap by proposing a flexible and expressive framework for the represention of Socially Interconnected MultiMedia-enriched Objects (SIMMO), that avoids the complexity of previous models (e.g., [5]). Our main contribution is a model that describes in a unified manner interconnected multimedia content on the Web, generated by a variety of sources and hosted in diverse platforms, by also considering social features, such as users' interactions with such multimedia objects and with each other. The data representation requirements in such online social multimedia environments are identified (Section 2) in order to support the specification of the model and its implementation (Section 3). The expressive power of SIMMO is illustrated using a number of diverse examples of online multimedia objects with social interconnections, such as multimodal Web pages and YouTube videos (Section 4), while a comparison with existing approaches demonstrates its unique features (Section 5).

2 Data Representation Requirements

The proposed model needs to capture a number of characteristics stemming from the nature of online social multimedia with particular focus on those that are typically taken into account in the context of multimedia information processing, analysis, and access applications. The most salient such characteristics [8, 9] are:

- **C1 Host heterogeneity and fragmentation**: There is a plethora of diverse types of online services hosting and sharing media content, ranging from Web sites hosted on dedicated servers to social media sharing platforms, such as Flickr, Instagram, YouTube, etc. This heterogeneity is further exacerbated by the fact that each of the latter is typically associated with a number of attributes that are particular to the platform. It is possible though to identify cross-platform mappings for several of these attributes, such as among those conveying endorsement, e.g., likes in Facebook and favourites in Twitter.
- **C2 Media objects diversity**: Online media content is expressed in a variety of modalities (such as text, images, video, and audio) and contained within diverse media objects, ranging from simple media items (e.g., an online image or video file) to complex multimedia documents (e.g., Web pages and social media posts) consisting of heterogenous media items.
- **C3 Online links and relations**: Online media content does not live in isolation; there are in fact various relations that can be established among media objects. As mentioned above, multimedia documents can contain media items (e.g., YouTube videos can be embedded in Web pages or be shared through tweets), while they can also be interconnected with other media objects (e.g., Web pages or social media posts can contain links to other Web pages).
- **C4 Social links and interactions**: The users of social media sharing and networking platforms are connected with each other through explicit links (e.g., followship, friendship) and interact with the posted content and with each other (often using content), e.g., they like Facebook posts, comment on YouTube videos, add Vines to replies in Twitter, etc. Such social user behaviour is also supported outside the context of such platforms by several

Web sites that allow social interactions with their Web pages through, e.g., posting comments on them or sharing them on Twitter or Facebook.

C5 Dynamic content: Multimedia documents can also be classified based on their relationship with time. *Static* multimedia documents do not have a temporal dimension, whereas *dynamic* Web pages change over time, e.g., through comments being continuously posted on them.

C6 Automatically generated metadata: The digital devices currently used for generating media items (e.g., images) have the capability of automatically creating a wealth of metadata to annotate them, such as the geographical identification metadata tagging the photographs taken by smartphones and digital cameras. Such automatically generated metadata typically accompany the raw content, but social media sharing platforms may replace them with explicit metadata fields or even completely remove them in some cases.

The proposed model also needs to support a number of tasks commonly performed in multimedia information processing, analysis, and access applications [8, 9], such as search, clustering, and summarisation. Typical such tasks in online social multimedia settings include those listed below; their characteristics are influenced, to a large extent, by the properties of such settings outlined above:

T1 Cross-host search: In a variety of settings, end users are interested in retrieving media content in response to their information needs irrespective of the environment hosting the relevant media objects (see also C1), e.g., both Web pages and tweets relevant to a submitted query. Establishing the relevance and importance of media objects hosted in multiple and widely different environments is particularly challenging given their heterogeneity.

T2 Multimodal search: End users are interested in retrieving relevant information irrespective of the media in which it is encoded, while also having the freedom to express their queries in whichever media they find intuitive, e.g., using similar images when searching for an image or keywords to find their favourite song, and combinations thereof. Enabling unified retrieval that is transparent to users given queries expressed in any number of modalities is a difficult task given also the heterogeneity of available media objects (see also C2) and annotations (as discussed next in T3).

T3 Layered annotation: Multimedia content can be currently described in a multitude of ways and at different levels of abstraction, including descriptive metadata (e.g., creation date) (see also C6), textual annotations (e.g., keywords), low-level features (e.g., visual features such as SIFT), high-level features (e.g., concepts), and events. Many such annotations are interdependent, e.g., high-level features are generated based on the extracted low-level features, while events may be determined using the identified concepts. Establishing relations among annotations (e.g., determining which visual features were used for the concept annotation process) is important in many settings, particularly when end users are search professionals or researchers.

T4 Varied granularity access: In many cases, end users are interested in accessing media content at a granularity level different to that of a multimedia object. When searching for information, for instance, retrieval of only the specific media segments that contain relevant information, instead of entire multimedia objects, reduces users' cognitive load and increases their satisfaction. Such focussed retrieval applications include finding only the shots in a news video relevant to a story or only the image segments where e.g., a specific logo appears. Furthermore, representation at higher levels of granularity, e.g., multimedia collections, is also useful in many contexts. For instance, an aggegated view created by summarising a set of social media posts on the same subject or story provides a snapshot of public opinion on that topic.

T5 **Content provenance**: In several applications, it is important to track the original source of a content item posted online, e.g., to establish whether an image has been previously published in a different context. The ease with which media content is embedded within multimedia documents and shared across diverse platforms (see also C3 and C4) indicates the significance, but also the difficulty of this task. This is further the case when online content undergoes manipulations and alterations, and is subsequently reposted for entertainment (e.g., memes) or malicious (e.g., propaganda) purposes.

T6 **Influentials identification**: When researching a particular story or topic, several types of users (e.g., journalists, analysts, etc.) are interested in identifying influential and relevant content and also, particularly in the case of social media, the content contributors who publish such content. As this is typically achieved by analysing the Web and social link structures, it is paramount to model such relations between multimedia objects, between users and multimedia objects, and also between users (see also C3 and C4).

Based on the above characteristics and tasks, we extract a set of requirements for an effective data representation model that would enable multimedia information processing, analysis, and access applications in online socially interconnected environments. Without claiming that the above lists are exhaustive, they do cover both the principal aspects of online multimedia settings and their social features (see also Section 5). Therefore, our model should represent (in brackets the relevant items from the above lists giving rise to each requirement):

R1 media content of various modalities (C2, T2),
R2 diverse media objects, ranging from mono-modal media items to composite multimedia documents (C2, C3, T2),
R3 media objects across heterogeneous hosting environments in a unified manner (C1, T1),
R4 online relations between media objects (C2, C3, T5, T6),
R5 social interactions between users and media objects (C3, C4, C5, T5, T6),
R6 content contributors, their relations and interactions, as expressed through their accounts in social Web platforms (C4, T6),
R7 granularity at various levels, ranging from media segments to multimedia collections (T4),
R8 rich heterogeneous annotations describing media objects of any granularity, and the relationships between such annotations (T2, T4, T3), and
R9 descriptive metadata as attributes of media objects (C6, T3).

Next, we propose a data representation model based on the above requirements.

3 Data Representation Model

This section presents the proposed framework for the unified representation of Socially Interconnected MultiMedia-enriched Objects (SIMMO) available in web environments. SIMMO consists of a number of core entities and their sub-classes, attributes, and relations that have been determined based on the requirements (R1-R9) identified in Section 2. While similar entities are also encountered, at least in part, in other models (e.g., [5, 6, 1]) that have also formed part of our inspiration (see Section 5), it is the interconnections among SIMMO elements and the novel approach of bridging the gap between multimedia and social features that make SIMMO unique in its ability to support a wide range of applications.

Figure 1 presents a conceptual model of SIMMO with the following core entities and their sub-classes:

- ***Object*** is a generic entity representing media content ranging from mono-modal ***Item***s to multimedia ***Document***s. Each Item represents the actual media content consisting of a single modality, such as ***Text***, ***Image***, ***Video***, or ***Audio***, whereas Documents may be viewed as container objects consisting of potentially multiple such Items, and thus modalities. The most common instantiations of Web Documents are ***Webpage***s (e.g., pages in news sites, in entertainment portals, etc.) or ***Post***s in media sharing platforms with social characteristics (e.g., Facebook posts, tweets, etc.). There are also cases of Webpages consisting of Posts; a forum page, for instance, can be viewed as a container object consisting of posts on the same topic. The ***Media*** entity is introduced as an abstraction of Image, Video, and Audio so as to represent their common characteristics, such as the fact that they all may be associated with a Text item modelling the text associated with them (e.g., a caption) or extracted from them through e.g., ASR (Automatic Speech Recognition) for Video and Audio, and OCR (Optical Character Recognition) for Image and Video. Finally, further media (e.g., 3D objects) may be added as Item instantiations depending on the requirements of the particular application.
- ***Source*** is a generic entity representing media content contributors. This includes ***UserAccount***s representing users generating content, mainly posts in social media sharing platforms where they hold accounts, and ***WebDomains*** representing the Web sites hosting media content generated by their contributors. WebDomains are viewed as content contributors, even though they do not correspond to the actual person who contributed the content, given that in many cases the information regarding such people may not be available, or may be of much lesser importance in this specific context.
- ***Segment*** locates the media content of Items at a finer level of granularity (e.g., a passage in text, a region in an image, or a portion of a video) by including positional information as attributes. Instantiations of Segments (not depicted in Figure 1) include ***LinearSegment***s (e.g., with start/end positions as attributes for referring to text parts), ***SpatialSegment***s (e.g., with (x, y) pairs as attributes for referring to image regions), ***TemporalSegment***s (e.g., with start/end times as attributes for referring to video

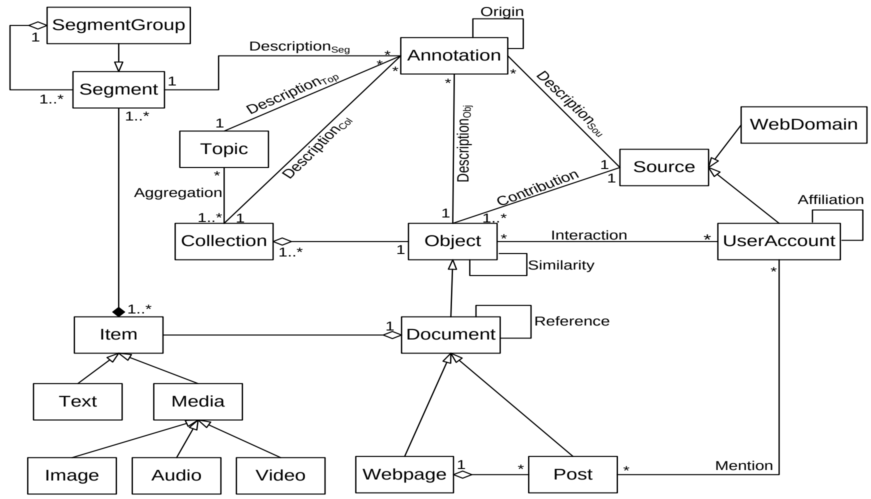

Fig. 1. SIMMO conceptual model presenting its elements and their relations. For simplicity, association relations that have attributes are depicted as simple associations.

frames/shots/scenes), and ***SpatioTemporalSegment***s. A ***SegmentGroup*** represents a collection of Segments; it is also modelled as a sub-class of Segment, thus allowing it to contain both Segments and other SegmentGroups.

- ***Collection*** models aggregates of Objects (i.e., higher levels of granularity), such as corpora of Web documents, sets of tweets, and image collections.
- ***Annotation*** is a generic entity representing together with its sub-classes (not depicted in Figure 1) a wide range of descriptions extracted from media content. These include annotations typically extracted from text (e.g., keywords, named entities, summaries, categories, etc.), media content features (e.g., low level descriptors, concepts and events), affective descriptions (e.g., sentiment and polarity), veracity scores reflecting the reliability of information and thus the trust that should be placed on it, and many others.
- ***Topic*** refers to any subject of interest in the context of an information processing, analysis, or access applications that users would like to keep track of. Its explicit representation allows to support a broad range of tasks, such as information filtering, topic tracking, and classification.

The main relations between these SIMMO elements, excluding the generalisation and aggregation/composition relations already discussed, are:

- The generation of media objects is modelled through a ***Contribution*** association between Source and Object.
- Explicit relations between Documents are modelled as ***Reference*** associations, with attributes such as the type of the relation. By considering that a Document may Reference another Document, we also consider (through inheritance) that a Webpage may Reference another Webpage (e.g., link to

it) and a Post may Reference another Post (e.g., reply to it or comment on it). We consider that this association is also able to model the References to Webpages from Posts (e.g., the Web links embedded in tweets) and to Posts from Webpages (e.g., to the comments dynamically posted on a Webpage).
- Objects may also be implicitly related to other Objects, e.g., through a computation of their similarity. Such ***Similarity*** relations are modelled as recursive associations between Objects, with attributes such as the type of the relation and the similarity score. This is useful in several applications and tasks, including clustering and verification of content provenance.
- A UserAccount may be involved in several relations, e.g., (i) be mentioned in a Post, (ii) be affiliated with (be friends with, follow etc.) another UserAccount, or (iii) interact with an Object (through likes, shares, views, etc.); the latter is more common for Posts, but users also interact with (e.g., like) whole Webpages. These three relations are modelled through the ***Mention***, ***Affiliation***, and ***Interaction*** assosiations, respectively, with attributes, such as the type of relation and the date it was established. As mentioned above, commenting is not modelled as a relation between Documents and UserAccounts, but rather as a Reference between two Documents (e.g., two Posts).
- All types of entities (i.e., Objects, Segments, Collections, Sources, and Topics) and their sub-classes may be associated with Annotations that are used for describing them. Such ***Description*** relations represent, for instance, the annotation of an Image with the SIFT features extracted from it, a TemporalSegment of a Video (such as a shot) with Concepts, or a UserAccount with Keywords reflecting the users' profile. Furthermore, links between different annotations (e.g., low-level descriptors and the concepts obtained from them) are modelled through the reflexive relation ***Origin*** between Annotations to denote the provenance of one with respect to the other.
- Each Topic is associated with a Collection of Objects on the particular subject of interest and may also be annotated itself. For instance, the Topic "Tour de France 2014" bicycle race would be associated with a Collection of Documents, such as Webpages and tweets on the subject, and could be annotated with the concepts "cycling" and "yellow yersey", the entity "Union Cycliste Internationale", and extracted locations, such as "Grenoble, France".

SIMMO elements and their relations also have several attributes representing their properties. For example, each Object is associated with a *URI*, *creation date*, and *crawl date*. Text is described by its *format* (e.g., HTML), an Image by its *size*, *EXIF* data, and associated *thumbnail*, a Video by its *duration*, *number of frames*, and associated *thumbnail*, and an Audio by its *duration*. Documents also have attributes related to the statistics regarding their social interactions, e.g., numbers of likes, comments, views, etc. The properties of a UserAccount include a *stream ID* denoting the platform hosting the account, the user's *name*, and the *number of followers/following/friends*. A complete list of the available attributes can be found in our implementation of SIMMO, discussed next.

Implementation: We have implemented the SIMMO framework in Java 1.7. We used Maven for controlling the project's build process, unit testing, and documentation creation, and the Google GSON library for converting Java objects into their JSON representation. This does by no means constrain the user from choosing another JSON library or another serialisation method. The SIMMO framework is open-source, released under the Apache License v2, and available at: `https://github.com/MKLab-ITI/simmo`. Based on this implementation, examples of commonly encountered online multimedia objects are presented next.

4 SIMMO Examples

To illustrate the flexibility and expressive power of the proposed framework, this section presents (using JSON) the SIMMO specifications of three diverse examples of online multimedia objects with social interconnections: (i) a Web page with content in different modalities and various annotations, (ii) a YouTube video with comments by several users, and (iii) a tweet with social interactions.

Consider, for instance, a Web page from an online newspaper discussing the recent World Cup 2014 final (Figure 2 (left)). SIMMO models this as a Webpage corresponding to the following JSON (some URLs have been shortened):

```
{  /* Webpage.json */
   "id":"180444840287",
   "url":"http://goo.gl/5JRsHi",
   "title":"World Cup was biggest event yet for Twitter with 672m tweets",
   "description":"Germany's demolition of Brazil ... peak tweets-per-minute",
   "tags":["Twitter", "World Cup", "Social networking", ...,  "Digital media", "Internet" ],
   "creationDate":"Jul 15, 2014 8:01:20 AM",
   "items":[
      {
         "type": "TEXT",
         "textFormat":"HTML",
         "content":"<p>Germany may have beaten Argentina to win the World Cup, ... "
      },
      {
         "type":"IMAGE",
         "url":"http://goo.gl/Uh4okO",
         "width":620,
         "height":372
         "describedBy":[ { "type":"LOWLEVELDESCRIPTOR", "annotationId":"A8423" }, ... ],
      }
   ],
   "references":[ { "type":"LINK", "referencedDocumentId":"180444840289" }, ...  ],
   "describedBy":[ { "type":"SUMMARY", "annotationId":"A9765" }, ....  ]
}
```

The Webpage has particular attributes, such as title and description, and contains HTML Text and an Image, each with its own properties. Both the Webpage and its constituent Items may be annotated (e.g., the Webpage with a summary and the Image with visual features, listed below as separate JSON entries). The Webpage also connects to other Webpages through References of type "LINK".

```
{  /* Summary.json */
   "id":"A9765"
   "summaryMethod":"Manual",
   "content":"Germany may have beaten Argentina to win the World Cup, ..."
}
{  /* LowLevelDescriptor.json */
   "id":"A8423"
   "descriptorType":"SURF",
   "numberOfFeatures":128,
   "descriptorValue":"128 1035 <CIRCLE 470 276 1 0 0>; 0.000537204 0.000681045 ... 0.00020111"
}
```

Fig. 2. Illustrative examples of multimedia documents (l-r): a Web page in the news domain, a YouTube video with comments, and a tweet with replies.

The next example corresponds to a YouTube video (Figure 2 (middle)) contributed by a UserAccount and modelled as a Post consisting of the actual video content and References to its comments, each also modelled as a Post. Several social interaction features are also modelled as attributes, such as the number of subscriptions to the UserAccount and the number of views of the video.

```
{  /* Post.json */
   "id":"wtt2aSV8wdw",
   "url":"https://www.youtube.com/watch?v=wtt2aSV8wdw",
   "title":"Internet Citizens: Defend Net Neutrality",
   "description":"Tell the FCC to reclassify broadband internet ...",
   "creationDate":"May 5, 2014 4:07:17 PM",
   "createdBy":"acc98754",
   "items":[
      {
         "type":"VIDEO",
         "url":"https://www.youtube.com/v/wtt2aSV8wdw",
         "width":1280,
         "height":720,
         "duration":213,
      }
   ],
   "numComments":4538,
   "numViews":919353,
   "positiveVotes":43615,
   "negativeVotes":394,
   "references":[ { "type":"COMMENT", "referencedDocumentId":"409sfh" }, ...  ]
}
{  /* UserAccount.json */
   "id":"acc98754",
   "name":"CGP Grey",
   "numSubscriptions":1361024,
   "avatarSmall":"http://goo.gl/YJS4PG"
}
{  /* Post.json */
   "id":"409sfh",
   "createdBy":"acc74528"
   "items":[
      {
         "type":"TEXT"
         "textFormat":"HTML",
         "content":"<div class="Ct">Learn about this and pass it on! ... </div>",
      }
   ],
   "numComments":72,
   "positiveVotes":739,
   "negativeVotes":0
}
```

The final example corresponds to a tweet (Figure 2 (right)) modelled as a Post that contains both Text and an Image, together with Mentions to specific UserAccounts, while statistics of social interactions are represented by attributes. Replies to the tweet are also modelled as Posts (not listed here).

```
{  /* Post.json */
  "id": "491252639225901056",
  "createdBy": "digitalocean"
  "creationDate": "Jul 21, 2014 4:05:30 PM",
  "items": [
    {
      "type": "TEXT"
      "textFormat": "HTML",
      "content": "We sent @jedgar out to meet DigitalOcean customer @KrakenIO and all ..."
    }
    {
      "type": "IMAGE"
      "url": "http://pbs.twimg.com/media/BtFHq9ZCUAAhyho.jpg:large",
      "height": 768,
      "width": 1024,
    }
  ],
  "mentions":[ { "mentioned":"jedgar" }, ...  ],
  "numShares": 4,
  "positiveVotes": 19,
  "negativeVotes": 0,
  "references":[ { "type":"REPLY", "referencedDocumentId":"491255375912370176" }, ...  ]
}
```

5 Comparison to Existing Approaches

To assess the expressive power of SIMMO, we compare it to other multimedia data representation models. First, existing approaches are presented and then a comparison is performed on the basis of the requirements identified in Section 2.

Early attempts to describe the content and structure of multimedia data (e.g., [4]) were soon superseded by the MPEG-7 standard [5], a generic, but complex, framework that enables highly structural, detailed descriptions of multimedia content at different granularity levels. MPEG-7 relies on: (i) *Descriptors* (*D*) defining the syntax and semantics of diverse features, (ii) *Description Schemes* (*DS*) describing the structure and semantics of relations among D or DS, (iii) a *Description Definition Language* allowing the creation and modification of DS and D, and (iv) *Systems Tools*, supporting various tasks, e.g., synchronisation of descriptions with content.

MPEG-21 [3] followed soon as an open framework for multimedia delivery and consumption, focussing on how the elements of a multimedia application infrastructure should relate, integrate, and interact. To this end, it centres around the concept of *Digital Items*, i.e., structured objects with multimedia content and metadata, and *Users* interacting with them; it also puts particular emphasis on Intellectual Property issues and mechanisms for the management of rights.

More recently, the Rich Unified Content Description (RUCoD) [6] framework was introduced for representing intrinsic properties of multimedia *Content Objects*, enhanced with real-world information (e.g., geo-location) and affective descriptions (e.g., in the valence/arousal 2D space). Each RUCoD consists of: (i) a header containing descriptive metadata (e.g., id and creation date) together with information about the media it contains and their descriptors, (ii) low-level descriptors, (iii) real-world descriptors, and (iv) user-related descriptors.

In addition, the infrastructure developed by WebLab [7] for integrating multimedia information processing components also defined a common exchange format to support the communication between such components. This exchange format, in essence a multimedia data representation model, centres on the notion of *Resource* that models several types of entities, including content in various modalities, multimedia documents and their segments, and diverse annotations.

The models discussed above focus on the description of multimedia content and thus satisfy requirements R1, R2, R7, R8 and R9, listed in Section 2; see also Table 1 for an overview of the requirements satisfied by each model. Given though that most were developed prior to the explosion of social media, they do not take into account the social characteristics and interconnections in current web environments (requirement R5). Such aspects have been addressed by ontologies, such as SIOC [1] and FOAF [2]. SIOC (Socially-Interlinked Online Communities) captures the nature, structure, and content of online communities (such as forums) through the representation of *User*s creating *Post*s organised in *Forum*s that are hosted on *Site*s; modelled as sub-classes of the generic concepts *Item*, *Container*, and *Space*, respectively. SIOC is commonly used in conjunction with the FOAF (Friend Of A Friend) vocabulary to express users' personal information and social networking interactions. These approaches are not concerned though with the potential multimodality of Posts/Items and the annotations extracted from such multimedia content (requirements R2 and R8), that are of paramount importance in information processing, analysis, and access tasks.

SIMMO bridges the gap between these perspectives by modelling both multimedia content (and its descriptions) and also users' social interactions with such content and with each other; see Table 1. To this end, SIMMO has borrowed several elements from the aforementioned approaches, while it has also introduced new aspects to support the emerging needs and requirements. For instance, the SIMMO multimedia content description draws many ideas from MPEG standards, but eschews their complexity, while SIMMO Annotations instantiated as LowLevelDescriptors could be mapped to standardised MPEG-7 Descriptors. Modelling granularity at the Segment level has been inspired by WebLab, while RUCoD has motivated the incorporation of affective and real-world features. The

Table 1. Comparison of different models w.r.t. the requirements identified in Section 2 (✓= requirement is satisfied; ∼ = requirement is partly satisfied)

Requirement: brief description	MPEG-7	RUCoD	WebLab	SIOC+FOAF	SIMMO
R1: multiple modalities	✓	✓	✓		✓
R2: diverse media objects	✓	✓	✓		✓
R3: heterogenous hosts	✓	✓	✓	✓	✓
R4: online links		✓		✓	✓
R5: social interactions				✓	✓
R6: contributors(description+relations)	∼			✓	✓
R7: granularity at different levels	✓	✓	✓		✓
R8: various annotations	✓	✓	✓		✓
R9: descriptive metadata	✓	✓	✓	✓	✓

concept of UserAccount has been borrowed by FOAF, while the Post and Forum SIOC elements could be mapped to the Post and Webpage SIMMO components. Finally, many of the attributes of media Objects, Documents, and Items are equivalent to those proposed by Dublin Core (`http://www.dublincore.org`).

6 Conclusions

We have proposed a model that integrates in a unified manner the representation of multimedia and social features in online environments. Its flexibility and expressive power allow it to embrace the heterogeneity of multimedia content and its interconnections, thus making it unique in its ability to support a wide range of multimedia information processing, analysis, and access applications. Our aim is for SIMMO to be a reusable data model across such applications; to facilitate its adoption, we plan to extend its documentation, add utility methods (such as the implementation of standard indexing and retrieval operations), and identify and implement mappings to established data models (such as SIOC).

Acknowledgements. This work was supported by MULTISENSOR (contract no. FP7-610411), SocialSensor (contract no. FP7-287975), and REVEAL (contract no. FP7-610928) projects, partially funded by the European Commission.

References

1. U. Bojars, J. G. Breslin, V. Peristeras, G. Tummarello, and S. Decker. Interlinking the social web with semantics. *IEEE Intelligent Systems*, 23(3):29–40, 2008.
2. D. Brickley and L. Miller. FOAF vocabulary specification 0.98. *Namespace Document*, 9, 2012.
3. I. S. Burnett, R. V. de Walle, K. Hill, J. Bormans, and F. Pereira. MPEG-21: Goals and achievements. *IEEE MultiMedia*, 10(4):60–70, 2003.
4. A. Caetano and N. Guimaraes. A model for content representation of multimedia information. In *Proceedings of the 1st workshop on the Challenge of Image Retrieval (CIR 1998), organised by the British Computer Society*, 1998.
5. S.-F. Chang, T. Sikora, and A. Purl. Overview of the MPEG-7 standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):688–695, 2001.
6. P. Daras, A. Axenopoulos, V. Darlagiannis, D. Tzovaras, X. L. Bourdon, L. Joyeux, A. Verroust-Blondet, V. Croce, T. Steiner, A. Massari, A. Camurri, S. Morin, A.-D. Mezaour, L. F. Sutton, and S. Spiller. Introducing a unified framework for content object description. *International Journal of Multimedia Intelligence and Security*, 2(3):351–375, 2011.
7. P. Giroux, S. Brunessaux, S. Brunessaux, J. Doucy, G. Dupont, B. Grilheres, Y. Mombrun, A. Saval, and P. d. des Portes. Weblab: An integration infrastructure to ease the development of multimedia processing applications. In *Proceedings of the International Conference on Software and System Engineering and their Applications (ICSSEA)*, 2008.
8. S. Papadopoulos and Y. Kompatsiaris. Social multimedia crawling for mining and search. *IEEE Computer*, 47(5):84–87, 2014.
9. N. Ramzan, R. van Zwol, J.-S. Lee, K. Clüver, and X.-S. Hua, editors. *Social Media Retrieval.* Computer Communications and Networks. Springer, 2013.