CLUE Working Group
Internet-Draft
Intended status: Informational
Expires: December 21, 2014

R. Presta
S P. Romano
University of Napoli

June 19, 2014

An XML Schema for the CLUE data model
draft-ietf-clue-data-model-schema-05

Abstract

This document formally describes the data model adopted in CLUE. It provides an XML schema file for the definition of CLUE data model types.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on December 21, 2014.

Copyright Notice

Copyright (c) 2014 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction 4
2. Terminology 4
3. XML Schema 7
4. <mediaCaptures> 17
5. <encodingGroups> 17
6. <captureScenes> 17
7. <simultaneousSets> 17
8. <globalSceneEntries> 17
9. <captureEncodings> 17
10. <mediaCapture> 17
10.1. <capturedMedia> 20
10.2. <captureSceneIDREF> 20
10.3. <encGroupIDREF> 20
10.4. <spatialInformation> 20
10.4.1. <capturePoint> 21
10.4.2. <captureArea> 22
10.5. <nonSpatiallyDefinable> 23
10.6. <content> 23
10.7. <synchronizationID> 24
10.8. <composed> 24
10.9. <switched> 24
10.10. <policy> 24
10.11. <maxCaptures> 25
10.12. <individual> 25
10.13. <description> 25
10.14. <priority> 26
10.15. <lang> 26
10.16. <mobility> 26
10.17. <maxCaptureEncodings> 26
10.18. <relatedTo> 26
10.19. <view> 26
10.20. <presentation> 27
10.21. <capturedPeople> 27
10.21.1. <personIDREF> 27
10.22. captureID attribute 27
11. Audio captures 27
11.1. <audioChannelFormat> 28
12. Video captures 28
12.1. <embeddedText> 29
13. Text captures 29
14. <captureScene> 29
14.1. <sceneInformation> 30
14.2. <sceneEntries> 30
14.3. sceneID attribute 31
14.4. scale attribute 31
15. <sceneEntry> 32
15.1. <mediaCaptureIDs> 32
15.2. sceneEntryID attribute 33

15.3. mediaType attribute 33
16. <encodingGroup> 33
16.1. <maxGroupBandwidth> 33
16.2. <encodingIDList> 33
16.3. encodingGroupID attribute 34
17. <simultaneousSet> 34
17.1. <captureIDREF> 34
17.2. <sceneEntryIDREF> 35
18. <globalSceneEntry> 35
19. <people> 35
19.1. <person> 36
19.1.1. personID attribute 37
19.1.2. <personInfo> 37
19.1.3. <personType> 37
20. <captureEncoding> 37
20.1. <captureID> 38
20.2. <encodingID> 38
20.3. <configuredContent> 38
21. <clueInfo> 38
22. Sample XML file 39
23. MCC example 45
24. Diff with draft-ietf-clue-data-model-schema-02 version . . . 52
25. Diff with draft-ietf-clue-data-model-schema-03 version 52
26. Diff with draft-ietf-clue-data-model-schema-04 version 53
27. Informative References 53

1. Introduction

This document formally describes the data model associated with the description of a CLUE telepresence session. It defines the involved data model types and provides a formal description of each of them in terms of an XML schema file.

The schema is based on information contained in [I-D.ietf-clue-framework]. It encodes information and constraints defined in the aforementioned document in order to provide a formal representation of the concepts therein presented. The schema definition is intended to be modified according to changes applied to the above mentioned CLUE document.

The document actually represents a proposal aiming at the definition of a coherent structure for all the information associated with the description of a telepresence scenario.

2. Terminology

This document refers to the same terminology used in [I-D.ietf-clue-framework]. We briefly recall herein some of the main terms adopted in that document.

Audio Capture: Media Capture for audio. Denoted as ACn in the example cases in this document.

Camera-Left and Right: For Media Captures, Camera-Left and Camera-Right are from the point of view of a person observing the rendered media. They are the opposite of Stage-Left and Stage-Right.

Capture: Same as Media Capture.

Capture Device: A device that converts audio and video input into an electrical signal, in most cases to be fed into a media encoder.

Capture Encoding: A specific encoding of a Media Capture, to be sent by a Media Provider to a Media Consumer via RTP.

Capture Scene: An abstraction grouping semantically-coupled Media Captures available at the Media Provider's side, representing a precise portion of the local scene that can be transmitted remotely. Capture Scene MAY correspond to a part of the telepresence room or MAY focus only on the presentation media. A Capture Scene is characterized by a set of attributes, as well as by a set of Capture Scene Entries.

Capture Scene Entry: A list of Media Captures of the same media type that constitute a possible representation of a Capture Scene. Media Captures belonging to the same Capture Scene Entry can be sent simultaneously by the Media Provider.

CLUE Participant: An entity able to use the CLUE protocol within a telepresence session. It can be an Endpoint or a MCU able to use the CLUE protocol.

Consumer: Same as Media Consumer.

Encoding or Individual Encoding: The representation of an encoding technology. In the CLUE datamodel, for each encoding, a set of parameters is provided. Such parameters represent encoding constraints, like for example the maximum bandwidth of the Media Provider the encoding can consume.

Encoding Group: The representation of a group of encodings. For each group, a set of parameters is provided, representing the constraints to be applied to the group as a whole. An example is the maximum bandwidth that can be consumed when using the contained encodings simultaneously.

Endpoint The logical point of final termination through receiving, decoding and rendering, and/or initiation through capturing, encoding, and sending of media streams. An endpoint consists of one or more physical devices which source and sink media streams, and exactly one SIP Conferencing Framework Participant (which, in turn, includes exactly one SIP User Agent). Endpoints can be anything from multiscreen/multicamera room controllers to handheld devices.

MCU: Multipoint Control Unit (MCU) - a device that connects two or more endpoints together into one single multimedia conference. An MCU may include a Mixer.

Media: Any data that, after suitable encoding, can be conveyed over RTP, including audio, video or timed text.

Media Capture: A "Media Capture", or simply "Capture", is a source of Media of a single type (i.e., audio or video or text).

Media Stream: The term "Media Stream", or simply "Stream", is used as a synonymous of Capture Encoding.

Media Provider: A CLUE participant (i.e., an Endpoint or an MCU) able to send Media Streams.

Media Consumer: A CLUE participant (i.e., an Endpoint or an MCU) able to receive Media Streams.

Scene: Same as Capture Scene.

Scene Entry: Same as Capture Scene Entry.

Stream: Same as Media Stream.

Multiple Content Capture: A Capture that can contain different Media Captures of the same media type. It is denoted as MCC in this document. In the Stream resulting from the MCC, the Stream coming from the encoding of the composing Media Captures can appear simultaneously, if the MCC is the result of a mixing operation, or can appear alternatively over time, according to a certain switching policy.

Plane of Interest: The spatial plane containing the most relevant subject matter.

Provider: Same as Media Provider.

Render: The process of reproducing the received Streams like, for instance, displaying the remote video on the Media Consumer's screens, or playing out the remote audio through loudspeakers.

Simultaneous Transmission Set: A set of Media Captures that can be transmitted simultaneously from a Media Provider.

Single Media Capture: A Capture representing the Media coming from a single-source Capture Device.

Spatial Information: Data about the spatial position of a Capture Device that generates a Single Media Capture within the context of a Capture Scene representing a physical portion of a Telepresence Room.

Stream Characteristics: The union of the features used to describe a Stream in the CLUE environment and in the SIP-SDP environment

Video Capture: A Media Capture for video.

3. XML Schema

This section contains the proposed CLUE data model schema definition.

The element and attribute definitions are formal representations of the concepts needed to describe the capabilities of a Media Provider, as well as the streams that are requested by a Media Consumer given the Provider's offer.

The main groups of information are:

<mediaCaptures>: the list of media captures available (Section 4)

<encodingGroups>: the list of encoding groups (Section 5)

<captureScenes>: the list of capture scenes (Section 6)

<simultaneousSets>: the list of simultaneous transmission sets (Section 7)

<globalSceneEntries>: the list of global capture entry sets (Section 8)

<people>: meta data about the participants represented in the telepresence session (Section 19). [to be discussed]

<captureEncodings>: the list of instantiated capture encodings (Section 9)

All of the above refers to concepts that have been introduced in [I-D.ietf-clue-framework] and further detailed in the following of this document.

```
<?xml version="1.0" encoding="UTF-8" ?>
<xs:schema
   targetNamespace="urn:ietf:params:xml:ns:clue-info"
   xmlns:tns="urn:ietf:params:xml:ns:clue-info"
   xmlns:xs="http://www.w3.org/2001/XMLSchema"
   xmlns="urn:ietf:params:xml:ns:clue-info"
   xmlns:xcard="urn:ietf:params:xml:ns:vcard-4.0"
   elementFormDefault="qualified"
   attributeFormDefault="unqualified">

<!-- Import xcard XML schema -->
<xs:import namespace="urn:ietf:params:xml:ns:vcard-4.0"
schemaLocation="xcard.xsd"/>
```

```
<!-- ELEMENT DEFINITIONS -->
<xs:element name="mediaCaptures" type="mediaCapturesType"/>
<xs:element name="encodingGroups" type="encodingGroupsType"/>
<xs:element name="captureScenes" type="captureScenesType"/>
<xs:element name="simultaneousSets" type="simultaneousSetsType"/>
<xs:element name="globalSceneEntries" type="globalSceneEntriesType"/>
<xs:element name="people" type="peopleType"/>

<xs:element name="captureEncodings" type="captureEncodingsType"/>


<!-- MEDIA CAPTURES TYPE -->
<!-- envelope of media captures -->
<xs:complexType name="mediaCapturesType">
 <xs:sequence>
   <xs:element name="mediaCapture" type="mediaCaptureType"
   maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>


<!-- DESCRIPTION element -->
<xs:element name="description">
 <xs:complexType>
  <xs:simpleContent>
   <xs:extension base="xs:string">
     <xs:attribute name="lang" type="xs:language"/>
   </xs:extension>
  </xs:simpleContent>
 </xs:complexType>
</xs:element>

<!-- MEDIA CAPTURE TYPE -->
<xs:complexType name="mediaCaptureType" abstract="true">
  <xs:sequence>
    <!-- mandatory fields -->
    <xs:element name="capturedMedia" type="xs:string"/>
    <xs:element name="captureSceneIDREF" type="xs:IDREF"/>
    <xs:element name="encGroupIDREF" type="xs:IDREF" minOccurs="0"/>
    <xs:choice>
      <xs:sequence>
        <xs:element name="spatialInformation"
                    type="tns:spatialInformationType"/>
      </xs:sequence>
      <xs:element name="nonSpatiallyDefinable" type="xs:boolean" fixed="true"/>
    </xs:choice>
    <!-- for handling multi-content captures: -->
    <xs:choice>
```

```
      <xs:sequence>
        <xs:element name="synchronizationID" type="xs:ID" minOccurs="0"/>
        <xs:element name="content" type="contentType" minOccurs="0"/>
        <xs:element name="composed" type="xs:boolean" minOccurs="0"/>
        <xs:element name="switched" type="xs:boolean" minOccurs="0"/>
        <xs:element name="policy" type="xs:string" minOccurs="0"/>
        <xs:element name="maxCaptures" type="xs:unsignedInt" minOccurs="0"/>
      </xs:sequence>
      <xs:element name="individual" type="xs:boolean" fixed="true"/>
    </xs:choice>
    <!-- optional fields -->
    <xs:element ref="description" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="priority" type="xs:unsignedInt" minOccurs="0"/>
    <xs:element name="lang" type="xs:language" minOccurs="0"/>
    <xs:element name="mobility" type="mobilityType" minOccurs="0"/>
    <xs:element name="presentation" type="presentationType" minOccurs="0"/>
    <xs:element name="view" type="viewType" minOccurs="0"/>
    <xs:element name="capturedPeople" type="capturedPeopleType" minOccurs="0"/>
    <xs:element name="maxCaptureEncodings" type="xs:unsignedInt"
    minOccurs="0"/>
    <xs:element name="relatedTo" type="xs:IDREF" minOccurs="0"/>
    <xs:any namespace="##other" processContents="lax" minOccurs="0"
    maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="captureID" type="xs:ID" use="required"/>
  <xs:anyAttribute namespace="##other" processContents="lax"/>
</xs:complexType>


<!-- CONTENT TYPE -->
<xs:complexType name="contentType">
 <xs:sequence>
   <xs:element name="captureIDREF" type="xs:IDREF"
   minOccurs="0" maxOccurs="unbounded"/>
   <xs:element name="sceneEntryIDREF" type="xs:IDREF"
   minOccurs="0" maxOccurs="unbounded"/>
   <xs:any namespace="##other" processContents="lax" minOccurs="0"
   maxOccurs="unbounded"/>
 </xs:sequence>
 <xs:anyAttribute namespace="##other" processContents="lax"/>
</xs:complexType>

<!-- CAPTURED PEOPLE TYPE -->
<xs:complexType name="capturedPeopleType">
<xs:sequence>
<xs:element name="personIDREF" type="xs:IDREF" maxOccurs="unbounded"/>
</xs:sequence>
</xs:complexType>
```

```
<!-- PEOPLE TYPE -->
<xs:complexType name="peopleType">
<xs:sequence>
<xs:element name="person" type="personType"
                    maxOccurs="unbounded"/>
</xs:sequence>
</xs:complexType>

<!-- PERSON TYPE -->
<xs:complexType name="personType">
  <xs:sequence>
     <xs:element name="personInfo" type="xcard:vcardType" maxOccurs="1"
                         minOccurs="0"/>
     <xs:element name="personType" type="personTypeType"
                         minOccurs="0"
                 maxOccurs="unbounded"/>
     <xs:any namespace="##other" processContents="lax" minOccurs="0"
                 maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="personID" type="xs:ID"/>
  <xs:anyAttribute namespace="##other" processContents="lax"/>
</xs:complexType>

<!-- PERSON TYPE TYPE -->
<xs:simpleType name="personTypeType">
        <xs:restriction base="xs:string">
        <xs:enumeration value="chairman"/>
        <xs:enumeration value="vice-chairman"/>
        <xs:enumeration value="minute taker"/>
        <xs:enumeration value="presenter"/>
        <xs:enumeration value="translator"/>
        <xs:enumeration value="timekeeper"/>
        <xs:enumeration value="attendee"/>
        </xs:restriction>
</xs:simpleType>



<!-- VIEW TYPE -->
<xs:simpleType name="viewType">
 <xs:restriction base="xs:string">
  <xs:enumeration value="room"/>
  <xs:enumeration value="table"/>
  <xs:enumeration value="lectern"/>
  <xs:enumeration value="individual"/>
  <xs:enumeration value="audience"/>
 </xs:restriction>
</xs:simpleType>
```

```
<!-- PRESENTATION TYPE -->
<xs:simpleType name="presentationType">
 <xs:restriction base="xs:string">
  <xs:enumeration value="slides"/>
  <xs:enumeration value="image"/>
  <xs:enumeration value=""/>
 </xs:restriction>
</xs:simpleType>

<!-- SPATIAL INFORMATION TYPE -->
<xs:complexType name="spatialInformationType">
 <xs:sequence>
  <xs:element name="capturePoint" type="capturePointType"/>
  <xs:element name="captureArea" type="captureAreaType"
  minOccurs="0"/>
  <xs:any namespace="##other" processContents="lax" minOccurs="0"
  maxOccurs="unbounded"/>
 </xs:sequence>
 <xs:anyAttribute namespace="##other" processContents="lax"/>
</xs:complexType>

<!-- MOBILITY TYPE -->
<xs:simpleType name="mobilityType">
 <xs:restriction base="xs:string">
  <xs:enumeration value="static"/>
  <xs:enumeration value="dynamic"/>
  <xs:enumeration value="highly-dynamic"/>
 </xs:restriction>
</xs:simpleType>

<!-- TEXT CAPTURE TYPE -->
<xs:complexType name="textCaptureType">
 <xs:complexContent>
  <xs:extension base="tns:mediaCaptureType">
  </xs:extension>
 </xs:complexContent>
</xs:complexType>


<!-- AUDIO CAPTURE TYPE -->
<xs:complexType name="audioCaptureType">
 <xs:complexContent>
  <xs:extension base="tns:mediaCaptureType">
   <xs:sequence>
    <xs:element name="audioChannelFormat" type="audioChannelFormatType"
    minOccurs="0"/>
   </xs:sequence>
  </xs:extension>
```

```
 </xs:complexContent>
</xs:complexType>


<!-- AUDIO CHANNEL FORMAT TYPE -->
<xs:simpleType name="audioChannelFormatType">
 <xs:restriction base="xs:string">
  <xs:enumeration value="mono"/>
  <xs:enumeration value="stereo"/>
 </xs:restriction>
</xs:simpleType>

<!-- VIDEO CAPTURE TYPE -->
<xs:complexType name="videoCaptureType">
 <xs:complexContent>
  <xs:extension base="tns:mediaCaptureType">
   <xs:sequence>
    <xs:element ref="embeddedText" minOccurs="0"/>
   </xs:sequence>
  </xs:extension>
 </xs:complexContent>
</xs:complexType>

<!-- EMBEDDED TEXT ELEMENT -->
<xs:element name="embeddedText">
 <xs:complexType>
  <xs:simpleContent>
   <xs:extension base="xs:boolean">
    <xs:attribute name="lang" type="xs:language"/>
   </xs:extension>
  </xs:simpleContent>
 </xs:complexType>
</xs:element>

<!-- CAPTURE SCENES TYPE -->
<!-- envelope of capture scenes -->
<xs:complexType name="captureScenesType">
 <xs:sequence>
  <xs:element name="captureScene" type="captureSceneType"
  maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>

<!-- CAPTURE SCENE TYPE -->
<xs:complexType name="captureSceneType">
 <xs:sequence>
  <xs:element ref="description" minOccurs="0" maxOccurs="unbounded"/>
  <xs:element name="sceneInformation" type="xcard:vcardType" minOccurs="0"/>
```

```
  <xs:element name="sceneEntries" type="sceneEntriesType"/>
  <xs:any namespace="##other" processContents="lax" minOccurs="0"
  maxOccurs="unbounded"/>
 </xs:sequence>
 <xs:attribute name="sceneID" type="xs:ID" use="required"/>
 <xs:attribute name="scale" type="scaleType" use="required"/>
 <xs:anyAttribute namespace="##other" processContents="lax"/>
</xs:complexType>

<!-- SCALE TYPE -->
<xs:simpleType name="scaleType">
 <xs:restriction base="xs:string">
  <xs:enumeration value="millimeters"/>
  <xs:enumeration value="unknown"/>
  <xs:enumeration value="noscale"/>
 </xs:restriction>
</xs:simpleType>

<!-- CAPTURE AREA TYPE -->
<xs:complexType name="captureAreaType">
 <xs:sequence>
  <xs:element name="bottomLeft" type="pointType"/>
  <xs:element name="bottomRight" type="pointType"/>
  <xs:element name="topLeft" type="pointType"/>
  <xs:element name="topRight" type="pointType"/>
 </xs:sequence>
</xs:complexType>

<!-- POINT TYPE -->
<xs:complexType name="pointType">
 <xs:sequence>
  <xs:element name="x" type="xs:decimal"/>
  <xs:element name="y" type="xs:decimal"/>
  <xs:element name="z" type="xs:decimal"/>
 </xs:sequence>
</xs:complexType>

<!-- CAPTURE POINT TYPE -->
<xs:complexType name="capturePointType">
 <xs:complexContent>
  <xs:extension base="pointType">
   <xs:sequence>
    <xs:element name="lineOfCapturePoint" type="tns:pointType"
    minOccurs="0"/>
   </xs:sequence>
   <xs:attribute name="pointID" type="xs:ID"/>
  </xs:extension>
 </xs:complexContent>
```

```
</xs:complexType>


<!-- SCENE ENTRIES TYPE -->
<!-- envelope of scene entries of a capture scene -->
<xs:complexType name="sceneEntriesType">
 <xs:sequence>
  <xs:element name="sceneEntry" type="sceneEntryType"
  maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>

<!-- SCENE ENTRY TYPE -->
<xs:complexType name="sceneEntryType">
 <xs:sequence>
  <xs:element ref="description" minOccurs="0" maxOccurs="unbounded"/>
  <xs:element name="mediaCaptureIDs" type="captureIDListType"/>
 </xs:sequence>
 <xs:attribute name="sceneEntryID" type="xs:ID" use="required"/>
 <xs:attribute name="mediaType" type="xs:string" use="required"/>
</xs:complexType>


<!-- CAPTURE ID LIST TYPE -->
<xs:complexType name="captureIDListType">
 <xs:sequence>
  <xs:element name="captureIDREF" type="xs:IDREF"
  maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>

<!-- ENCODING GROUPS TYPE -->
<xs:complexType name="encodingGroupsType">
 <xs:sequence>
  <xs:element name="encodingGroup" type="tns:encodingGroupType"
  maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>

<!-- ENCODING GROUP TYPE -->
<xs:complexType name="encodingGroupType">
 <xs:sequence>
  <xs:element name="maxGroupBandwidth" type="xs:unsignedInt"/>
  <xs:element name="encodingIDList" type="encodingIDListType"/>
  <xs:any namespace="##other" processContents="lax" minOccurs="0"
  maxOccurs="unbounded"/>
 </xs:sequence>
 <xs:attribute name="encodingGroupID" type="xs:ID" use="required"/>
```

```
 <xs:anyAttribute namespace="##any" processContents="lax"/>
</xs:complexType>

<!-- ENCODING ID LIST TYPE -->
<xs:complexType name="encodingIDListType">
 <xs:sequence>
  <xs:element name="encID" type="xs:string" maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>

<!-- SIMULTANEOUS SETS TYPE -->
<xs:complexType name="simultaneousSetsType">
 <xs:sequence>
  <xs:element name="simultaneousSet" type="simultaneousSetType"
  maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>

<!-- SIMULTANEOUS SET TYPE -->
<xs:complexType name="simultaneousSetType">
 <xs:sequence>
   <xs:element name="captureIDREF" type="xs:IDREF"
   minOccurs="0" maxOccurs="unbounded"/>
   <xs:element name="sceneEntryIDREF" type="xs:IDREF"
   minOccurs="0" maxOccurs="unbounded"/>
     <xs:any namespace="##other" processContents="lax" minOccurs="0"
   maxOccurs="unbounded"/>
 </xs:sequence>
 <xs:attribute name="setID" type="xs:ID" use="required"/>
 <xs:attribute name="mediaType" type="xs:string"/>
 <xs:anyAttribute namespace="##any" processContents="lax"/>
</xs:complexType>

<!-- GLOBAL SCENE ENTRIES TYPE -->
<xs:complexType name="globalSceneEntriesType">
 <xs:sequence>
  <xs:element name="globalSceneEntry" type="globalSceneEntryType"
  maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>

<!-- GLOBAL SCENE ENTRY TYPE -->
<xs:complexType name="globalSceneEntryType">
 <xs:sequence>
   <xs:element name="sceneEntryIDREF" type="xs:IDREF"
    maxOccurs="unbounded"/>
   <xs:any namespace="##other" processContents="lax" minOccurs="0"
   maxOccurs="unbounded"/>
```

```
 </xs:sequence>
 <xs:attribute name="globalSceneEntryID" type="xs:ID"/>
 <xs:attribute name="mediaType" type="xs:string" use="required"/>
 <xs:anyAttribute namespace="##any" processContents="lax"/>
</xs:complexType>

<!-- CAPTURE ENCODINGS TYPE -->
<xs:complexType name="captureEncodingsType">
 <xs:sequence>
  <xs:element name="captureEncoding" type="captureEncodingType"
  maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>

<!-- CAPTURE ENCODING TYPE -->
<xs:complexType name="captureEncodingType">
 <xs:sequence>
  <xs:element name="captureID" type="xs:string"/>
  <xs:element name="encodingID" type="xs:string"/>
  <xs:element name="configuredContent" type="contentType" minOccurs="0"/>
  <xs:any namespace="##other" processContents="lax" minOccurs="0"
   maxOccurs="unbounded"/>
 </xs:sequence>
 <xs:attribute name="ID" type="xs:ID"/>
 <xs:anyAttribute namespace="##any" processContents="lax"/>
</xs:complexType>

<!-- CLUE INFO ELEMENT -->
<!-- the <clueInfo> envelope can be seen
     as the ancestor of an <advertisement> envelope -->
<xs:element name="clueInfo" type="clueInfoType"/>

<!-- CLUE INFO TYPE -->
<xs:complexType name="clueInfoType">
  <xs:sequence>
   <xs:element ref="mediaCaptures"/>
   <xs:element ref="encodingGroups"/>
   <xs:element ref="captureScenes"/>
   <xs:element ref="simultaneousSets" minOccurs="0"/>
   <xs:element ref="globalSceneEntries" minOccurs="0"/>
   <xs:element ref="people" minOccurs="0"/>
   <xs:any namespace="##other" processContents="lax" minOccurs="0"
   maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="clueInfoID" type="xs:ID" use="required"/>
  <xs:anyAttribute namespace="##other" processContents="lax"/>
</xs:complexType>
</xs:schema>
```

The following sections describe the XML schema in more detail.

4. <mediaCaptures>

<mediaCaptures> represents the list of one ore more media captures available on the media provider's side. Each media capture is represented by a <mediaCapture> element (Section 10).

5. <encodingGroups>

<encodingGroups> represents the list of the encoding groups available at the media provider's side. Each encoding group is represented by an <encodingGroup> element (Section 16).

6. <captureScenes>

<captureScenes> represents the list of capture scenes available at the media provider's side. Each capture scene is represented by a <captureScene> element. (Section 14).

7. <simultaneousSets>

<simultaneousSets> contains the simultaneous sets indicated by the media provider. Each simultaneous set is represented by a <simultaneousSet> element. (Section 17).

8. <globalSceneEntries>

<globalSceneEntries> contains a set of alternative representations of all the scenes that are offered by a Media Provider to a Media Consumer. Each alternative is named "global scene entry" and is represented by a <globalSceneEntry> element. (Section 18).

9. <captureEncodings>

<captureEncodings> is a list of capture encodings. It can represent either the list of the desired capture encodings indicated by the media consumer or the list of instantiated captures on the provider's side. Each capture encoding is represented by a <captureEncoding> element. (Section 20).

10. <mediaCapture>

According to the CLUE framework, a media capture is the fundamental representation of a media flow that is available at the provider's side. Media captures are characterized (i) by a set of features that are independent from the specific type of medium, and (ii) by a set of features that are media-specific. The features that are common to

all media types appear within the media capture type, that has been designed as an abstract complex type. Media-specific captures, such as video captures, audio captures and others, are specializations of that abstract media capture type, as in a typical generalization-specialization hierarchy.

The following is the XML Schema definition of the media capture type:

```
<!-- MEDIA CAPTURE TYPE -->
<xs:complexType name="mediaCaptureType" abstract="true">
  <xs:sequence>
    <!-- mandatory fields -->
    <xs:element name="capturedMedia" type="xs:string"/>
    <xs:element name="captureSceneIDREF" type="xs:IDREF"/>
    <xs:element name="encGroupIDREF" type="xs:IDREF" minOccurs="0"/>
    <xs:choice>
      <xs:sequence>
        <xs:element name="spatialInformation"
                    type="tns:spatialInformationType"/>
      </xs:sequence>
      <xs:element name="nonSpatiallyDefinable" type="xs:boolean" fixed="true"/>
    </xs:choice>
    <!-- for handling multi-content captures: -->
    <xs:choice>
      <xs:sequence>
        <xs:element name="synchronizationID" type="xs:ID" minOccurs="0"/>
        <xs:element name="content" type="contentType" minOccurs="0"/>
        <xs:element name="composed" type="xs:boolean" minOccurs="0"/>
        <xs:element name="switched" type="xs:boolean" minOccurs="0"/>
        <xs:element name="policy" type="xs:string" minOccurs="0"/>
        <xs:element name="maxCaptures" type="xs:unsignedInt" minOccurs="0"/>
      </xs:sequence>
      <xs:element name="individual" type="xs:boolean" fixed="true"/>
    </xs:choice>
    <!-- optional fields -->
    <xs:element ref="description" minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="priority" type="xs:unsignedInt" minOccurs="0"/>
    <xs:element name="lang" type="xs:language" minOccurs="0"/>
    <xs:element name="mobility" type="mobilityType" minOccurs="0"/>
    <xs:element name="presentation" type="presentationType" minOccurs="0"/>
    <xs:element name="view" type="viewType" minOccurs="0"/>
    <xs:element name="participantIDs" type="participantIDsType" minOccurs="0"/>
    <xs:element name="maxCaptureEncodings" type="xs:unsignedInt"
    minOccurs="0"/>
    <xs:element name="relatedTo" type="xs:IDREF" minOccurs="0"/>
    <xs:any namespace="##other" processContents="lax" minOccurs="0"
    maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="captureID" type="xs:ID" use="required"/>
  <xs:anyAttribute namespace="##other" processContents="lax"/>
</xs:complexType>
```

10.1. <capturedMedia>

<capturedMedia> is a mandatory field specifying the media type of the capture ("audio", "video", "text",...).

10.2. <captureSceneIDREF>

<captureSceneIDREF> is a mandatory field containing the identifier of the capture scene the media capture belongs to. Indeed, each media capture must be associated with one and only one capture scene. When a media capture is spatially definable, some spatial information is provided along with it in the form of point coordinates (see Section 10.4). Such coordinates refer to the space of coordinates defined for the capture scene containing the capture.

10.3. <encGroupIDREF>

<encGroupIDREF> is a mandatory field containing the identifier of the encoding group the media capture is associated with.

10.4. <spatialInformation>

Media captures are divided into two categories: (i) non spatially definable captures and (ii) spatially definable captures.

Captures are spatially definible when it is possible to provide at least the coordinates of the device position within the telepresence room of origin. Such coordinates are expressed according to the coordinate space of the capture scene the media captures belongs to.

Non spatially definable captures cannot be characterized within the physical space of the telepresence room of origin. Captures of this kind are for example those related to recordings, text captures, DVDs, recorded presentations, or external streams, that are played in the telepresence room and transmitted to remote sites. Another example is represented by switched captures: their content, in fact, comes from different devices over time.

Spatially definable captures represent a part of the telepresence room. The captured part of the telepresence room is described by means of the <spatialInformation> element. Non spatially definable captures do not show in their XML description such element: they are instead characterized by having the <nonSpatiallyDefinable> tag set to "true" (see Section 10.5).

The definition of the spatial information type is the following:

```
<!-- SPATIAL INFORMATION TYPE -->
<xs:complexType name="spatialInformationType">
 <xs:sequence>
  <xs:element name="capturePoint" type="capturePointType"/>
  <xs:element name="captureArea" type="captureAreaType"
  minOccurs="0"/>
  <xs:any namespace="##other" processContents="lax" minOccurs="0"
  maxOccurs="unbounded"/>
 </xs:sequence>
 <xs:anyAttribute namespace="##other" processContents="lax"/>
</xs:complexType>
```

The <capturePoint> contains the coordinates of the capture device that is taking the capture, as well as, optionally, the pointing direction (see Section 10.4.1). It is a mandatory field when the media capture is spatially definable, independently from the media type.

The <captureArea> is an optional field containing four points defining the captured area covered by the capture (see Section 10.4.2).

10.4.1. <capturePoint>

The <capturePoint> element is used to represent the position and the line of capture of a capture device. The XML Schema definition of the <capturePoint> element type is the following:

```
<!-- POINT TYPE -->
<xs:complexType name="pointType">
 <xs:sequence>
  <xs:element name="x" type="xs:decimal"/>
  <xs:element name="y" type="xs:decimal"/>
  <xs:element name="z" type="xs:decimal"/>
 </xs:sequence>
</xs:complexType>

<!-- CAPTURE POINT TYPE -->
<xs:complexType name="capturePointType">
 <xs:complexContent>
  <xs:extension base="pointType">
   <xs:sequence>
    <xs:element name="lineOfCapturePoint" type="tns:pointType"
    minOccurs="0"/>
   </xs:sequence>
   <xs:attribute name="pointID" type="xs:ID"/>
  </xs:extension>
 </xs:complexContent>
</xs:complexType>
```

The point type contains three spatial coordinates (x,y,z) representing a point in the space associated with a certain capture scene.

The capture point type extends the point type, i.e., it is represented by three coordinates identifying the position of the capture device, but can carry further information. Such further information is conveyed by the <lineOfCapturePoint>, which is one more point-type element representing the "point on line of capture", that gives the pointing direction of the capture device.

The coordinates of the point on line of capture MUST NOT be identical to the capture point coordinates. If the point on line of capture is not specified, no assumptions are made about the pointing direction of the capturing device.

10.4.2. <captureArea>

<captureArea> is an optional element that can be contained within the spatial information associated with a media capture. It represents the spatial area captured by the media capture.

The XML representation of that area is provided through a set of four point-type elements, <bottomLeft>, <bottomRight>, <topLeft>, and <topRight>, as it can be seen from the following definition:

```
<!-- CAPTURE AREA TYPE -->
<xs:complexType name="captureAreaType">
 <xs:sequence>
  <xs:element name="bottomLeft" type="pointType"/>
  <xs:element name="bottomRight" type="pointType"/>
  <xs:element name="topLeft" type="pointType"/>
  <xs:element name="topRight" type="pointType"/>
 </xs:sequence>
</xs:complexType>
```

<bottomLeft>, <bottomRight>, <topLeft>, and <topRight> should be co-planar.

By comparing the capture area of different media captures within the same capture scene, a consumer can better determine the spatial relationships among them and render them correctly.

10.5. <nonSpatiallyDefinable>

When media captures are non spatially definable, they are marked with the boolean <nonSpatiallyDefinable> element set to "true" and no <spatialInformation> is provided. Indeed, <nonSpatiallyDefinable> and <spatialInformation> are mutually exclusive tags, according to the <choice> section within the XML Schema definition of the media capture type.

10.6. <content>

A media capture can be (i) an individual media capture or (ii) a multiple content capture (MCC). A multiple content capture is made of different captures that can be arranged spatially (by a composition operation), or temporally (by a switching operation), or that can result from the proper orchestration of both techniques. If a media capture is a MCC, then it MUST show in its XML data model representation the <content> element. It is a mandatory element composed of a list of media capture identifiers ("captureIDREF") and capture scene entry identifiers ("sceneEntryIDREF"), where the last ones are used as shortcuts to refer to multiple capture identifiers.

```
<!-- CONTENT TYPE -->
<xs:complexType name="contentType">
 <xs:sequence>
   <xs:element name="captureIDREF" type="xs:IDREF"
   minOccurs="0" maxOccurs="unbounded"/>
   <xs:element name="sceneEntryIDREF" type="xs:IDREF"
   minOccurs="0" maxOccurs="unbounded"/>
   <xs:any namespace="##other" processContents="lax" minOccurs="0"
   maxOccurs="unbounded"/>
 </xs:sequence>
 <xs:anyAttribute namespace="##other" processContents="lax"/>
</xs:complexType>
```

10.7. <synchronizationID>

<synchronizationID> is an optional element for multiple content captures that contains a numeric identifier. Multiple content captures marked with the same identifier in the <synchronizationID> contain, at each point in time, captures coming from the same room endpoint.

10.8. <composed>

<composed> is an optional boolean element that can be used only for multiple content captures. It indicates whether or not a multiple content capture is a mix (audio) or a composition (video) of streams. If set to "true", it means that the capture can result at a certain point in time from more than one capture. This attribute is useful for a media consumer for example to avoid nesting a composed video capture into a different composed capture.

10.9. <switched>

<switched> is an optional boolean element that can be used only for multiple content captures. It indicates whether or not a multiple content capture switches over time. If set to "true", it means that the content of the MCC (in terms of the actual composing captures) can change over time.

10.10. <policy>

<policy> is an optional element that can be used only for multiple content captures. It indicates the criteria applied to build the multiple content capture using the media captures referenced in <contentCaptureIDs>. Such element can be set to a list of pre-defined values ([todo]).

10.11. <maxCaptures>

<maxCaptures> is an optional element that can be used only for multiple content captures. It indicates the maximum number of media captures that can be represented in the multiple content capture at a time.

10.12. <individual>

<individual> is a boolean element that MUST be used for single-content captures. Its value is fixed and set to "true". Such element indicates that the capture being described is not a multiple content capture. Indeed, <individual> and the aforementioned tags related to MCC attributes (from Section 10.6 to Section 10.11) are mutually exclusive, according to the <choice> section within the XML Schema definition of the media capture type.

10.13. <description>

<description> is used to optionally provide human-readable textual information about a media capture. The same element is used to describe, besides media captures, capture scenes and capture scene entries, as it is included in their XML representation. A media capture can be described by using multiple <description> elements, each providing information in a different language. The <description> element definition is the following:

```
<!-- DESCRIPTION element -->
<xs:element name="description">
 <xs:complexType>
  <xs:simpleContent>
   <xs:extension base="xs:string">
     <xs:attribute name="lang" type="xs:language"/>
   </xs:extension>
  </xs:simpleContent>
 </xs:complexType>
</xs:element>
```

As it can be seen from the above snippet, <description> is a string element with an attribute ("lang") indicating the language used in the textual description.

10.14. <priority>

<priority> is an optional unsigned integer field indicating the importance of a media capture according to the media provider's perspective. It can be used on the receiver's side to automatically identify the most relevant contribution from the media provider. The higher the importance, the lower the contained value. When media captures are marked with a "0" priority value, it means that they are "not subject to priority".

10.15. <lang>

<lang> is an optional element containing the language used in the capture, if any.

10.16. <mobility>

<mobility> is an optional element indicating whether or not the capture device originating the capture may move during the telepresence session. Such an optional element can assume one of the three following values: (i) static, (ii) dynamic or (iii) highly dynamic, defined as in [I-D.ietf-clue-framework].

10.17. <maxCaptureEncodings>

The optional <maxCaptureEncodings> contains an unsigned integer indicating the maximum number of capture encodings that can be simultaneously active for the media capture. If absent, this parameter defaults to 1. The minimum value for this attribute is 1. The number of simultaneous capture encodings is also limited by the restrictions of the encoding group the media capture refers to by means of the <encGroupIDREF> element.

10.18. <relatedTo>

The optional <relatedTo> element contains the value of the ID attribute of the media capture it refers to. The media capture marked with a <relatedTo> element can be for example the translation of a main media capture in a different language.

10.19. <view>

The <view> element is an optional tag describing what is represented in the spatial area covered by a media capture. The current possible values are: "table", "lectern", "individual", and "audience", as listed in the enumerative view type in the following.

10.20. <presentation>

The <presentation> element is an optional tag used for media captures conveying information about presentations within the telepresence session. The current possible values are "slides" and "images", as listed in the enumerative presentation type in the following.

10.21. <capturedPeople>

This optional element is used to indicate which telepresence session participants are represented within the media captures. For each participant, a <personIDREF> element is provided.

[to be discussed]

10.21.1. <personIDREF>

<personIDREF> contains the identifier of the represented person. Metadata about the represented participant can be retrieved by accessing the <people> list (Section 19).

10.22. captureID attribute

The "captureID" attribute is a mandatory field containing the identifier of the media capture.

11. Audio captures

Audio captures inherit all the features of a generic media capture and present further audio-specific characteristics. The XML Schema definition of the audio capture type is reported below:

```
<!-- AUDIO CAPTURE TYPE -->
<xs:complexType name="audioCaptureType">
 <xs:complexContent>
  <xs:extension base="tns:mediaCaptureType">
   <xs:sequence>
    <xs:element name="audioChannelFormat" type="audioChannelFormatType"
    minOccurs="0"/>
   </xs:sequence>
  </xs:extension>
 </xs:complexContent>
</xs:complexType>
```

Audio-specific information about the audio capture is contained in

<audioChannelFormat> (Section 11.1).

11.1. <audioChannelFormat>

The optional <audioChannelFormat> element is a field with enumerated values ("mono" and "stereo") which describes the method of encoding used for audio. A value of "mono" means the audio capture has one channel. A value of "stereo" means the audio capture has two audio channels, left and right. A single stereo capture is different from two mono captures that have a left-right spatial relationship. A stereo capture maps to a single RTP stream, while each mono audio capture maps to a separate RTP stream.

The XML Schema definition of the <audioChannelFormat> element type is provided below:

```
<!-- AUDIO CHANNEL FORMAT TYPE -->
<xs:simpleType name="audioChannelFormatType">
 <xs:restriction base="xs:string">
  <xs:enumeration value="mono"/>
  <xs:enumeration value="stereo"/>
 </xs:restriction>
</xs:simpleType>
```

12. Video captures

Video captures, similarly to audio captures, extend the information of a generic media capture with video-specific features, such as <embeddedText> (Section 12.1).

The XML Schema representation of the video capture type is provided in the following:

```
<!-- VIDEO CAPTURE TYPE -->
<xs:complexType name="videoCaptureType">
 <xs:complexContent>
  <xs:extension base="tns:mediaCaptureType">
   <xs:sequence>
    <xs:element ref="embeddedText" minOccurs="0"/>
   </xs:sequence>
  </xs:extension>
 </xs:complexContent>
</xs:complexType>
```

12.1. <embeddedText>

The <embeddedText> element is a boolean element indicating that there is text embedded in the video capture. The language used in such embedded textual description is reported in the "lang" attribute.

The XML Schema definition of the <embeddedText> element is:

```
<!-- EMBEDDED TEXT ELEMENT -->
<xs:element name="embeddedText">
 <xs:complexType>
  <xs:simpleContent>
   <xs:extension base="xs:boolean">
    <xs:attribute name="lang" type="xs:language"/>
   </xs:extension>
  </xs:simpleContent>
 </xs:complexType>

</xs:element>
```

13. Text captures

Also text captures can be described by extending the generic media capture information, similarly to audio captures and video captures.

The XML Schema representation of the text capture type is currently lacking text-specific information, as it can be seen by looking at the definition below:

```
<!-- TEXT CAPTURE TYPE -->
<xs:complexType name="textCaptureType">
 <xs:complexContent>
  <xs:extension base="tns:mediaCaptureType">
  </xs:extension>
 </xs:complexContent>
</xs:complexType>
```

14. <captureScene>

A media provider organizes the available captures in capture scenes in order to help the receiver both in the rendering and in the selection of the group of captures. Capture scenes are made of

capture scene entries, that are a set of media captures of the same media type. Each capture scene entry represents an alternative to thoroughly represent a capture scene for a fixed media type.

The XML Schema representation of a <captureScene> element is the following:

```
<!-- CAPTURE SCENE TYPE -->
<xs:complexType name="captureSceneType">
 <xs:sequence>
  <xs:element ref="description" minOccurs="0" maxOccurs="unbounded"/>
  <xs:element name="sceneInformation" type="xcard:vcardType" minOccurs="0"/>
  <xs:element name="sceneEntries" type="sceneEntriesType"/>
  <xs:any namespace="##other" processContents="lax" minOccurs="0"
  maxOccurs="unbounded"/>
 </xs:sequence>
 <xs:attribute name="sceneID" type="xs:ID" use="required"/>
 <xs:attribute name="scale" type="scaleType" use="required"/>
 <xs:anyAttribute namespace="##other" processContents="lax"/>
</xs:complexType>
```

The <captureScene> element can contain zero or more textual <description> elements, defined as in Section 10.13. Besides <description>, there is the <sceneInformation> element (Section 14.1), which contains structured information about the scene in the vcard format, as well as the <sceneEntries> element (Section 14.2), which is the list of available capture scene entries.

14.1. <sceneInformation>

The <sceneInformation> element contains optional information about the capture scene according to the vcard format.

14.2. <sceneEntries>

The <sceneEntries> element is a mandatory field of a capture scene containing the list of scene entries. Each scene entry is represented by a <sceneEntry> element (Section 15).

```
<!-- SCENE ENTRIES TYPE -->
<!-- envelope of scene entries of a capture scene -->
<xs:complexType name="sceneEntriesType">
 <xs:sequence>
  <xs:element name="sceneEntry" type="sceneEntryType"
  maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>
```

14.3. sceneID attribute

The sceneID attribute is a mandatory attribute containing the identifier of the capture scene.

14.4. scale attribute

The scale attribute is a mandatory attribute that specifies the scale of the coordinates provided in the spatial information of the media capture belonging to the considered capture scene. The scale attribute can assume three different values:

"millimeters" - the scale is in millimeters. Systems which know their physical dimensions (for example professionally installed telepresence room systems) should always provide those real-world measurements.

"unknown" - the scale is not necessarily millimeters, though it is the same for every media capture in the capture scene. Systems which don't know specific physical dimensions but still know relative distances should select "unknown" in the scale attribute of the capture scene to be described.

"noscale" - there is no common physical scale among the media captures of the capture scene. That means the scale might be different for each media capture.

```
<!-- SCALE TYPE -->
<xs:simpleType name="scaleType">
 <xs:restriction base="xs:string">
  <xs:enumeration value="millimeters"/>
  <xs:enumeration value="unknown"/>
  <xs:enumeration value="noscale"/>
 </xs:restriction>
</xs:simpleType>
```

15. <sceneEntry>

A <sceneEntry> element represents a capture scene entry, which contains a set of media captures of the same media type describing a capture scene.

A <sceneEntry> element is characterized as follows.

```
<!-- SCENE ENTRY TYPE -->
<xs:complexType name="sceneEntryType">
 <xs:sequence>
  <xs:element ref="description" minOccurs="0" maxOccurs="unbounded"/>
  <xs:element name="mediaCaptureIDs" type="captureIDListType"/>
 </xs:sequence>
 <xs:attribute name="sceneEntryID" type="xs:ID" use="required"/>
 <xs:attribute name="mediaType" type="xs:string" use="required"/>
</xs:complexType>
```

One or more optional <description> elements provide human-readable information about the contents of a scene entry. <description> is defined as already seen in Section 10.13.

The remaining child elements are described in the following subsections.

15.1. <mediaCaptureIDs>

The <mediaCaptureIDs> is the list of the identifiers of the media captures included in the scene entry. It is an element of the captureIDListType type, which is defined as a sequence of <captureIDREF> elements, each containing the identifier of a media capture listed within the <mediaCaptures> element:

```
<!-- CAPTURE ID LIST TYPE -->
<xs:complexType name="captureIDListType">
 <xs:sequence>
  <xs:element name="captureIDREF" type="xs:IDREF"
  maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>
```

15.2. sceneEntryID attribute

The sceneEntryID attribute is a mandatory attribute containing the identifier of the capture scene entry represented by the <sceneEntry> element.

15.3. mediaType attribute

The mediaType attribute contains the media type of the media captures included in the scene entry.

16. <encodingGroup>

The <encodingGroup> element represents an encoding group, which is made of a set of one or more individual encodings and some parameters that apply to the group as a whole. Encoding groups contain references to individual encodings that can be applied to captures of the same media type. In other words, they can group audio encodings or, alternatively, video encodings. The definition of the <encodingGroup> element is the following:

```
<!-- ENCODING GROUP TYPE -->
<xs:complexType name="encodingGroupType">
 <xs:sequence>
  <xs:element name="maxGroupBandwidth" type="xs:integer"/>
  <xs:element name="encodingIDList" type="encodingIDListType"/>
  <xs:any namespace="##other" processContents="lax" minOccurs="0"
  maxOccurs="unbounded"/>
 </xs:sequence>
 <xs:attribute name="encodingGroupID" type="xs:ID" use="required"/>
 <xs:anyAttribute namespace="##any" processContents="lax"/>
</xs:complexType>
```

In the following, the inner elements are further described.

16.1. <maxGroupBandwidth>

<maxGroupBandwidth> is an optional field containing the maximum bitrate that can be shared by the individual encodings included in the encoding group.

16.2. <encodingIDList>

<encodingIDList> is the list of the individual encodings grouped together in the encoding group. Each individual encoding is

represented through its identifier contained within an <encID> element.

```
<!-- ENCODING ID LIST TYPE -->
<xs:complexType name="encodingIDListType">
 <xs:sequence>
  <xs:element name="encID" type="xs:IDREF" maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>
```

16.3. encodingGroupID attribute

The encodingGroupID attribute contains the identifier of the encoding group.

17. <simultaneousSet>

<simultaneousSet> represents a simultaneous transmission set, i.e., a list of captures of the same media type that can be transmitted at the same time by a media provider. There are different simultaneous transmission sets for each media type.

```
<!-- SIMULTANEOUS SET TYPE -->
<xs:complexType name="simultaneousSetType">
 <xs:sequence>
   <xs:element name="captureIDREF" type="xs:IDREF"
   minOccurs="0" maxOccurs="unbounded"/>
   <xs:element name="sceneEntryIDREF" type="xs:IDREF"
   minOccurs="0" maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>
```

Besides the identifiers of the captures (<captureIDREF> elements), also the identifiers of capture scene entries can be exploited, as shortcuts (<sceneEntryIDREF> elements).

17.1. <captureIDREF>

<captureIDREF> contains the identifier of the media capture that belongs to the simultaneous set.

17.2. <sceneEntryIDREF>

<captureIDREF> contains the identifier of the scene entry containing a group of captures that can be sent simultaneously with the companion captures in the simultaneous set.

18. <globalSceneEntry>

<globalSceneEntry> represents a set of captures of the same media type representing a summary of the complete Media Provider's offer. The media type of a global scene entry is reported in the "mediaType" attribute. The content of a global scene entry is expressed by leveraging scene entry identifiers, put within <sceneEntryIDREF> elements. Each global scene entry is identified by a unique identifier within the "globalSceneEntryID" attribute.

```
<!-- GLOBAL SCENE ENTRY TYPE -->
<xs:complexType name="globalSceneEntryType">
 <xs:sequence>
   <xs:element name="sceneEntryIDREF" type="xs:IDREF"
   minOccurs="0" maxOccurs="unbounded"/>
   <xs:any namespace="##other" processContents="lax" minOccurs="0"
   maxOccurs="unbounded"/>
 </xs:sequence>
 <xs:attribute name="globalSceneEntryID" type="xs:ID"/>
 <xs:attribute name="mediaType" type="xs:string" use="required"/>
 <xs:anyAttribute namespace="##any" processContents="lax"/>
</xs:complexType>
```

19. <people>

Information about the participants that are represented in the media captures is conveyed via the <people> element. As it can be seen from the XML Schema excerpt reported below, for each participant, a <person> element is provided.

[to be discussed]

```
<!-- PEOPLE TYPE -->
<xs:complexType name="peopleType">
<xs:sequence>
<xs:element name="person" type="personType"
                    maxOccurs="unbounded"/>
</xs:sequence>
</xs:complexType>

<!-- PERSON TYPE -->
<xs:complexType name="personType">
  <xs:sequence>
     <xs:element name="personInfo" type="xcard:vcardType" maxOccurs="1"
                         minOccurs="0"/>
     <xs:element name="personType" type="personTypeType"
                         minOccurs="0"
                  maxOccurs="unbounded"/>
     <xs:any namespace="##other" processContents="lax" minOccurs="0"
                  maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="personID" type="xs:ID"/>
  <xs:anyAttribute namespace="##other" processContents="lax"/>
</xs:complexType>

<!-- PERSON TYPE TYPE -->
<xs:simpleType name="personTypeType">
        <xs:restriction base="xs:string">
        <xs:enumeration value="chairman"/>
        <xs:enumeration value="vice-chairman"/>
        <xs:enumeration value="minute taker"/>
        <xs:enumeration value="presenter"/>
        <xs:enumeration value="translator"/>
        <xs:enumeration value="timekeeper"/>
        <xs:enumeration value="attendee"/>
        </xs:restriction>
</xs:simpleType>
```

19.1. <person>

<person> includes all the metadata related to a person represented within one or more media captures. Such an element currently provides at least the vcard of the person (via the <personInfo> element, see Section 19.1.2) and optionally his conference role(s) (via one or more <personType> elements, see Section 19.1.3). Furthermore, it has a mandatory "personID" attribute (Section 19.1.1).

19.1.1. personID attribute

The "personID" attribute carries the identifier of a represented person. Such identifier can be used to refer to the participant, as in the <capturedPeople> element inside the media captures representation (Section 10.21).

19.1.2. <personInfo>

The <personInfo> element is the XML representation of all the fields composing a vcard as specified in the Xcard RFC [RFC6351]. The vcardType is imported by the Xcard XML Schema provided by [I-D.ietf-ecrit-additional-data]. As such schema specifies, the <fn> element within <vcard> is mandatory.

19.1.3. <personType>

The value of the <personType> element determines the role of the represented participant within the telepresence session. It can be one of the following terms, that are defined in the framework document: "presenter", "timekeeper","attendee", "minute taker", "translator", "chairman", "vice-chairman".

A participant can play more than one conference role. In that case, more than one <personType> element will appear in his description.

20. <captureEncoding>

A <captureEncoding> is composed of the association between a media capture and an individual encoding, forming a capture stream as defined in [I-D.ietf-clue-framework]. A possible solution to model such an entity is provided in the following.

```
<!-- CAPTURE ENCODING TYPE -->
<xs:complexType name="captureEncodingType">
 <xs:sequence>
  <xs:element name="captureID" type="xs:string"/>
  <xs:element name="encodingID" type="xs:string"/>
  <xs:element name="configuredContent" type="contentType" minOccurs="0"/>
  <xs:any namespace="##other" processContents="lax" minOccurs="0"
   maxOccurs="unbounded"/>
 </xs:sequence>
 <xs:attribute name="ID" type="xs:ID"/>
 <xs:anyAttribute namespace="##any" processContents="lax"/>
</xs:complexType>
```

20.1. <captureID>

<captureID> is the mandatory element containing the identifier of the media capture that has been encoded to form the capture encoding.

20.2. <encodingID>

<encodingID> is the mandatory element containing the identifier of the applied individual encoding.

20.3. <configuredContent>

<configuredContent> is an optional element to be used in case of configuration of MCCs. It contains the list of capture identifiers and capture scene entry identifiers the Media Consumer wants within the MCC. That element is structured as the <content> element used to describe the content of a MCC. The total number of media captures listed in the <configuredContent> must be lower than or equal to the value carried within the <maxCaptures> attribute of the MCC.

21. <clueInfo>

The <clueInfo> element has been left within the XML Schema for representing a drafty version of the body of an ADVERTISEMENT message (see the example section).

```
<!-- CLUE INFO ELEMENT -->
<!-- the <clueInfo> envelope can be seen
     as the ancestor of an <advertisement> envelope -->
<xs:element name="clueInfo" type="clueInfoType"/>

<!-- CLUE INFO TYPE -->
<xs:complexType name="clueInfoType">
  <xs:sequence>
   <xs:element ref="mediaCaptures"/>
   <xs:element ref="encodingGroups"/>
   <xs:element ref="captureScenes"/>
   <xs:element ref="simultaneousSets" minOccurs="0"/>
   <xs:element ref="globalSceneEntries" minOccurs="0"/>
   <xs:element ref="people" minOccurs="0"/>
   <xs:any namespace="##other" processContents="lax" minOccurs="0"
   maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="clueInfoID" type="xs:ID" use="required"/>
  <xs:anyAttribute namespace="##other" processContents="lax"/>
</xs:complexType>
```

22. Sample XML file

The following XML document represents a schema compliant example of a CLUE telepresence scenario. Taking inspiration from the examples described in the framework draft ([I-D.ietf-clue-framework]), the XML representation of an endpoint-style Media Provider's offer is herein provided.

There are three cameras, where the central one is also able to capture a zoomed-out view of the overall telepresence room. Besides the three video captures coming from such cameras, the MP makes available a further multi-content capture about the loudest segment of the room, obtained by switching the video source across the three cameras. For the sake of simplicity, only one audio capture is advertised for the audio of the whole room.

The three cameras are placed in front of three participants (Alice, Bob and Ciccio), whose vcard and conference role details are also provided.

Media captures are arranged into four capture scene entries:

1. (VC0, VC1, VC2) - left, center and right camera video captures

2. (VC3) - video capture associated with loudest room segment

3. (VC4) - video capture zoomed out view of all people in the room

4. (AC0) - main audio

There are two encoding groups: (i) EG0, for video encodings, and (ii) EG1, for audio encodings.

As to the simultaneous sets, only VC1 and VC4 cannot be transmitted simultaneously since they are captured by the same device, i.e., the central camera (VC4 is a zoomed-out view while VC1 is a focused view of the front participant). The simultaneous sets would then be the following:

SS1 composed of VC3 and all the captures in the first capture scene entry (VC0,VC1,VC2);

SS2 composed of VC3, VC0, VC2, VC4

```xml
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<clueInfo xmlns="urn:ietf:params:xml:ns:clue-info"
xmlns:ns2="urn:ietf:params:xml:ns:vcard-4.0" clueInfoID="NapoliRoom">
    <mediaCaptures>
        <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="AC0">
            <capturedMedia>audio</capturedMedia>
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <encGroupIDREF>EG1</encGroupIDREF>
            <spatialInformation>
                <capturePoint>
                    <x>0.5</x>
                    <y>1.0</y>
                    <z>0.5</z>
                    <lineOfCapturePoint>
                        <x>0.5</x>
                        <y>0.0</y>
                        <z>0.5</z>
                    </lineOfCapturePoint>
                </capturePoint>
            </spatialInformation>
            <individual>true</individual>
            <description lang="en">main audio from the room</description>
            <priority>1</priority>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>room</view>
            <capturedPeople>
                <personIDREF>alice</personIDREF>
                <personIDREF>bob</personIDREF>
                <personIDREF>ciccio</personIDREF>
            </capturedPeople>
            <maxCaptureEncodings>1</maxCaptureEncodings>
        </mediaCapture>
        <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="VC0">
            <capturedMedia>video</capturedMedia>
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <encGroupIDREF>EG0</encGroupIDREF>
            <spatialInformation>
                <capturePoint>
                    <x>0.5</x>
                    <y>1.0</y>
                    <z>0.5</z>
                    <lineOfCapturePoint>
                        <x>0.5</x>
                        <y>0.0</y>
                        <z>0.5</z>
```

```
                </lineOfCapturePoint>
            </capturePoint>
        </spatialInformation>
        <individual>true</individual>
        <description lang="en">left camera video capture</description>
        <priority>1</priority>
        <lang>it</lang>
        <mobility>static</mobility>
        <view>individual</view>
        <capturedPeople>
            <personIDREF>ciccio</personIDREF>
        </capturedPeople>
        <maxCaptureEncodings>2</maxCaptureEncodings>
    </mediaCapture>
    <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:type="videoCaptureType" captureID="VC1">
        <capturedMedia>video</capturedMedia>
        <captureSceneIDREF>CS1</captureSceneIDREF>
        <encGroupIDREF>EG0</encGroupIDREF>
        <spatialInformation>
            <capturePoint>
                <x>0.5</x>
                <y>1.0</y>
                <z>0.5</z>
                <lineOfCapturePoint>
                    <x>0.5</x>
                    <y>0.0</y>
                    <z>0.5</z>
                </lineOfCapturePoint>
            </capturePoint>
        </spatialInformation>
        <individual>true</individual>
        <description lang="en">central camera video capture</description>
        <priority>1</priority>
        <lang>it</lang>
        <mobility>static</mobility>
        <view>individual</view>
        <capturedPeople>
            <personIDREF>alice</personIDREF>
        </capturedPeople>
        <maxCaptureEncodings>2</maxCaptureEncodings>
    </mediaCapture>
    <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:type="videoCaptureType" captureID="VC2">
        <capturedMedia>video</capturedMedia>
        <captureSceneIDREF>CS1</captureSceneIDREF>
        <encGroupIDREF>EG0</encGroupIDREF>
        <spatialInformation>
```

```
                <capturePoint>
                    <x>0.5</x>
                    <y>1.0</y>
                    <z>0.5</z>
                    <lineOfCapturePoint>
                        <x>0.5</x>
                        <y>0.0</y>
                        <z>0.5</z>
                    </lineOfCapturePoint>
                </capturePoint>
            </spatialInformation>
            <individual>true</individual>
            <description lang="en">right camera video capture</description>
            <priority>1</priority>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>individual</view>
            <capturedPeople>
                <personIDREF>bob</personIDREF>
            </capturedPeople>
            <maxCaptureEncodings>2</maxCaptureEncodings>
        </mediaCapture>
        <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="VC3">
            <capturedMedia>video</capturedMedia>
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <encGroupIDREF>EG0</encGroupIDREF>
            <nonSpatiallyDefinable>true</nonSpatiallyDefinable>
            <composed>false</composed>
            <switched>true</switched>
            <policy>Soundlevel:0</policy>
            <maxCaptures>1</maxCaptures>
            <description lang="en">loudest room segment</description>
            <priority>1</priority>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>individual</view>
            <maxCaptureEncodings>1</maxCaptureEncodings>
        </mediaCapture>
        <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="VC4">
            <capturedMedia>video</capturedMedia>
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <encGroupIDREF>EG0</encGroupIDREF>
            <spatialInformation>
                <capturePoint>
                    <x>0.5</x>
                    <y>1.0</y>
```

```
                    <z>0.5</z>
                    <lineOfCapturePoint>
                        <x>0.5</x>
                        <y>0.0</y>
                        <z>0.5</z>
                    </lineOfCapturePoint>
                </capturePoint>
            </spatialInformation>
            <individual>true</individual>
            <description lang="en">zoomed out view of all people in the room
            </description>
            <priority>1</priority>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>room</view>
            <capturedPeople>
                <personIDREF>alice</personIDREF>
                <personIDREF>bob</personIDREF>
                <personIDREF>ciccio</personIDREF>
            </capturedPeople>
            <maxCaptureEncodings>1</maxCaptureEncodings>
        </mediaCapture>
    </mediaCaptures>
    <encodingGroups>
        <encodingGroup encodingGroupID="EG0">
            <maxGroupBandwidth>600000</maxGroupBandwidth>
            <encodingIDList>
                <encID>ENC1</encID>
                <encID>ENC2</encID>
                <encID>ENC3</encID>
            </encodingIDList>
        </encodingGroup>
        <encodingGroup encodingGroupID="EG1">
            <maxGroupBandwidth>300000</maxGroupBandwidth>
            <encodingIDList>
                <encID>ENC4</encID>
                <encID>ENC5</encID>
            </encodingIDList>
        </encodingGroup>
    </encodingGroups>
    <captureScenes>
        <captureScene scale="unknown" sceneID="CS1">
            <sceneEntries>
                <sceneEntry mediaType="video" sceneEntryID="SE1">
                    <mediaCaptureIDs>
                        <captureIDREF>VC0</captureIDREF>
                        <captureIDREF>VC1</captureIDREF>
                        <captureIDREF>VC2</captureIDREF>
```

```
                    </mediaCaptureIDs>
                </sceneEntry>
                <sceneEntry mediaType="video" sceneEntryID="SE2">
                    <mediaCaptureIDs>
                        <captureIDREF>VC3</captureIDREF>
                    </mediaCaptureIDs>
                </sceneEntry>
                <sceneEntry mediaType="video" sceneEntryID="SE3">
                    <mediaCaptureIDs>
                        <captureIDREF>VC4</captureIDREF>
                    </mediaCaptureIDs>
                </sceneEntry>
                <sceneEntry mediaType="audio" sceneEntryID="SE4">
                    <mediaCaptureIDs>
                        <captureIDREF>AC0</captureIDREF>
                    </mediaCaptureIDs>
                </sceneEntry>
            </sceneEntries>
        </captureScene>
    </captureScenes>
    <simultaneousSets>
        <simultaneousSet setID="SS1">
            <captureIDREF>VC3</captureIDREF>
            <sceneEntryIDREF>SE1</sceneEntryIDREF>
        </simultaneousSet>
        <simultaneousSet setID="SS2">
            <captureIDREF>VC0</captureIDREF>
            <captureIDREF>VC2</captureIDREF>
            <captureIDREF>VC4</captureIDREF>
            <captureIDREF>VC3</captureIDREF>
        </simultaneousSet>
    </simultaneousSets>
    <people>
        <person personID="bob">
            <personInfo>
                <ns2:fn>
                    <ns2:text>Bob</ns2:text>
                </ns2:fn>
            </personInfo>
            <personType>minute taker</personType>
        </person>
        <person personID="alice">
            <personInfo>
                <ns2:fn>
                    <ns2:text>Alice</ns2:text>
                </ns2:fn>
            </personInfo>
            <personType>presenter</personType>
```

```
        </person>
        <person personID="ciccio">
            <personInfo>
                <ns2:fn>
                    <ns2:text>Ciccio</ns2:text>
                </ns2:fn>
            </personInfo>
            <personType>chairman</personType>
            <personType>timekeeper</personType>
        </person>
    </people>
</clueInfo>
```

23. MCC example

Enhancing the scenario presented in the previous example, the media provider is able to advertise a composed capture VC7 composed of a big picture representing the current speaker (VC3) and two picture-in-picture boxes representing the previous speakers (the previous one -VC5- and the oldest one -VC6). The provider does not want to instantiate and send VC5 and VC6, so it does not associate any encoding group with them. Their XML representations are provided for enabling the description of VC7.

A possible description for that scenario could be the following:

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<clueInfo xmlns="urn:ietf:params:xml:ns:clue-info"
xmlns:ns2="urn:ietf:params:xml:ns:vcard-4.0"
clueInfoID="NapoliRoom">
    <mediaCaptures>
        <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="AC0">
            <capturedMedia>audio</capturedMedia>
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <encGroupIDREF>EG1</encGroupIDREF>
            <spatialInformation>
                <capturePoint>
                    <x>0.5</x>
                    <y>1.0</y>
                    <z>0.5</z>
                    <lineOfCapturePoint>
                        <x>0.5</x>
                        <y>0.0</y>
```

```
                    <z>0.5</z>
                </lineOfCapturePoint>
            </capturePoint>
        </spatialInformation>
        <individual>true</individual>
        <description lang="en">main audio from the room</description>
        <priority>1</priority>
        <lang>it</lang>
        <mobility>static</mobility>
        <view>room</view>
        <capturedPeople>
            <personIDREF>alice</personIDREF>
            <personIDREF>bob</personIDREF>
            <personIDREF>ciccio</personIDREF>
        </capturedPeople>
        <maxCaptureEncodings>1</maxCaptureEncodings>
    </mediaCapture>
    <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:type="videoCaptureType" captureID="VC0">
        <capturedMedia>video</capturedMedia>
        <captureSceneIDREF>CS1</captureSceneIDREF>
        <encGroupIDREF>EG0</encGroupIDREF>
        <spatialInformation>
            <capturePoint>
                <x>0.5</x>
                <y>1.0</y>
                <z>0.5</z>
                <lineOfCapturePoint>
                    <x>0.5</x>
                    <y>0.0</y>
                    <z>0.5</z>
                </lineOfCapturePoint>
            </capturePoint>
        </spatialInformation>
        <individual>true</individual>
        <description lang="en">left camera video capture</description>
        <priority>1</priority>
        <lang>it</lang>
        <mobility>static</mobility>
        <view>individual</view>
        <capturedPeople>
            <personIDREF>ciccio</personIDREF>
        </capturedPeople>
        <maxCaptureEncodings>2</maxCaptureEncodings>
    </mediaCapture>
    <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:type="videoCaptureType" captureID="VC1">
        <capturedMedia>video</capturedMedia>
```

```
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <encGroupIDREF>EG0</encGroupIDREF>
            <spatialInformation>
                <capturePoint>
                    <x>0.5</x>
                    <y>1.0</y>
                    <z>0.5</z>
                    <lineOfCapturePoint>
                        <x>0.5</x>
                        <y>0.0</y>
                        <z>0.5</z>
                    </lineOfCapturePoint>
                </capturePoint>
            </spatialInformation>
            <individual>true</individual>
            <description lang="en">central camera video capture</description>
            <priority>1</priority>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>individual</view>
            <capturedPeople>
                <personIDREF>alice</personIDREF>
            </capturedPeople>
            <maxCaptureEncodings>2</maxCaptureEncodings>
        </mediaCapture>
        <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="VC2">
            <capturedMedia>video</capturedMedia>
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <encGroupIDREF>EG0</encGroupIDREF>
            <spatialInformation>
                <capturePoint>
                    <x>0.5</x>
                    <y>1.0</y>
                    <z>0.5</z>
                    <lineOfCapturePoint>
                        <x>0.5</x>
                        <y>0.0</y>
                        <z>0.5</z>
                    </lineOfCapturePoint>
                </capturePoint>
            </spatialInformation>
            <individual>true</individual>
            <description lang="en">right camera video capture</description>
            <priority>1</priority>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>individual</view>
```

```
            <capturedPeople>
                <personIDREF>bob</personIDREF>
            </capturedPeople>
            <maxCaptureEncodings>2</maxCaptureEncodings>
        </mediaCapture>
        <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="VC3">
            <capturedMedia>video</capturedMedia>
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <encGroupIDREF>EG0</encGroupIDREF>
            <nonSpatiallyDefinable>true</nonSpatiallyDefinable>
            <content>
                <sceneEntryIDREF>SE1</sceneEntryIDREF>
            </content>
            <composed>false</composed>
            <switched>true</switched>
            <policy>Soundlevel:0</policy>
            <maxCaptures>1</maxCaptures>
            <description lang="en">loudest room segment</description>
            <priority>1</priority>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>individual</view>
            <maxCaptureEncodings>1</maxCaptureEncodings>
        </mediaCapture>
        <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="VC4">
            <capturedMedia>video</capturedMedia>
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <encGroupIDREF>EG0</encGroupIDREF>
            <spatialInformation>
                <capturePoint>
                    <x>0.5</x>
                    <y>1.0</y>
                    <z>0.5</z>
                    <lineOfCapturePoint>
                        <x>0.5</x>
                        <y>0.0</y>
                        <z>0.5</z>
                    </lineOfCapturePoint>
                </capturePoint>
            </spatialInformation>
            <individual>true</individual>
            <description lang="en">zoomed out view of all people in the room
            </description>
            <priority>1</priority>
            <lang>it</lang>
            <mobility>static</mobility>
```

```
            <view>room</view>
            <capturedPeople>
                <personIDREF>alice</personIDREF>
                <personIDREF>bob</personIDREF>
                <personIDREF>ciccio</personIDREF>
            </capturedPeople>
            <maxCaptureEncodings>1</maxCaptureEncodings>
        </mediaCapture>
        <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="VC5">
            <capturedMedia>video</capturedMedia>
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <nonSpatiallyDefinable>true</nonSpatiallyDefinable>
            <content>
                <sceneEntryIDREF>SE1</sceneEntryIDREF>
            </content>
            <composed>false</composed>
            <switched>true</switched>
            <policy>Soundlevel:1</policy>
            <maxCaptures>1</maxCaptures>
            <description lang="en">penultimate loudest room segment
            </description>
            <priority>1</priority>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>individual</view>
            <maxCaptureEncodings>1</maxCaptureEncodings>
        </mediaCapture>
        <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="VC6">
            <capturedMedia>video</capturedMedia>
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <nonSpatiallyDefinable>true</nonSpatiallyDefinable>
            <content>
                <sceneEntryIDREF>SE1</sceneEntryIDREF>
            </content>
            <composed>false</composed>
            <switched>true</switched>
            <policy>Soundlevel:2</policy>
            <maxCaptures>1</maxCaptures>
            <description lang="en">last but two loudest room segment
            </description>
            <priority>1</priority>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>individual</view>
            <maxCaptureEncodings>1</maxCaptureEncodings>
        </mediaCapture>
```

```
        <mediaCapture xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="VC7">
            <capturedMedia>video</capturedMedia>
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <nonSpatiallyDefinable>true</nonSpatiallyDefinable>
            <content>
                <captureIDREF>VC3</captureIDREF>
                <captureIDREF>VC5</captureIDREF>
                <captureIDREF>VC6</captureIDREF>
            </content>
            <composed>true</composed>
            <switched>true</switched>
            <maxCaptures>1</maxCaptures>
            <description lang="en">big picture of the current speaker + pips
            about previous speakers</description>
            <priority>1</priority>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>individual</view>
            <maxCaptureEncodings>1</maxCaptureEncodings>
        </mediaCapture>
    </mediaCaptures>
    <encodingGroups>
        <encodingGroup encodingGroupID="EG0">
            <maxGroupBandwidth>600000</maxGroupBandwidth>
            <encodingIDList>
                <encID>ENC1</encID>
                <encID>ENC2</encID>
                <encID>ENC3</encID>
            </encodingIDList>
        </encodingGroup>
        <encodingGroup encodingGroupID="EG1">
            <maxGroupBandwidth>300000</maxGroupBandwidth>
            <encodingIDList>
                <encID>ENC4</encID>
                <encID>ENC5</encID>
            </encodingIDList>
        </encodingGroup>
    </encodingGroups>
    <captureScenes>
        <captureScene scale="unknown" sceneID="CS1">
            <sceneEntries>
                <sceneEntry mediaType="video" sceneEntryID="SE1">
                    <description lang="en">participants' individual videos
                    </description>
                    <mediaCaptureIDs>
                        <captureIDREF>VC0</captureIDREF>
                        <captureIDREF>VC1</captureIDREF>
```

```
                        <captureIDREF>VC2</captureIDREF>
                    </mediaCaptureIDs>
                </sceneEntry>
                <sceneEntry mediaType="video" sceneEntryID="SE2">
                    <description lang="en">loudest segment of the room
                    </description>
                    <mediaCaptureIDs>
                        <captureIDREF>VC3</captureIDREF>
                    </mediaCaptureIDs>
                </sceneEntry>
                <sceneEntry mediaType="video" sceneEntryID="SE5">
                    <description lang="en">loudest segment of the room + pips
                    </description>
                    <mediaCaptureIDs>
                        <captureIDREF>VC7</captureIDREF>
                    </mediaCaptureIDs>
                </sceneEntry>
                <sceneEntry mediaType="audio" sceneEntryID="SE4">
                    <description lang="en">room audio</description>
                    <mediaCaptureIDs>
                        <captureIDREF>AC0</captureIDREF>
                    </mediaCaptureIDs>
                </sceneEntry>
                <sceneEntry mediaType="video" sceneEntryID="SE3">
                    <description lang="en">room video</description>
                    <mediaCaptureIDs>
                        <captureIDREF>VC4</captureIDREF>
                    </mediaCaptureIDs>
                </sceneEntry>
            </sceneEntries>
        </captureScene>
    </captureScenes>
    <simultaneousSets>
        <simultaneousSet setID="SS1">
            <captureIDREF>VC7</captureIDREF>
            <sceneEntryIDREF>SE1</sceneEntryIDREF>
        </simultaneousSet>
        <simultaneousSet setID="SS2">
            <captureIDREF>VC0</captureIDREF>
            <captureIDREF>VC2</captureIDREF>
            <captureIDREF>VC4</captureIDREF>
            <captureIDREF>VC7</captureIDREF>
        </simultaneousSet>
    </simultaneousSets>
    <people>
        <person personID="bob">
            <personInfo>
                <ns2:fn>
```

```
                    <ns2:text>Bob</ns2:text>
                </ns2:fn>
            </personInfo>
            <personType>minute taker</personType>
        </person>
        <person personID="alice">
            <personInfo>
                <ns2:fn>
                    <ns2:text>Alice</ns2:text>
                </ns2:fn>
            </personInfo>
            <personType>presenter</personType>
        </person>
        <person personID="ciccio">
            <personInfo>
                <ns2:fn>
                    <ns2:text>Ciccio</ns2:text>
                </ns2:fn>
            </personInfo>
            <personType>chairman</personType>
            <personType>timekeeper</personType>
        </person>
    </people>
</clueInfo>
```

24. Diff with draft-ietf-clue-data-model-schema-02 version

capturePrameters and encodingParameters have been removed from the captureEncodingType

data model example has been updated and validated according to the new schema. Further description of the represented scenario has been provided.

A multiple content capture example has been added.

Obsolete comments and references have been removed.

25. Diff with draft-ietf-clue-data-model-schema-03 version

encodings section has been removed

global capture entries have been introduced

capture scene entry identifiers are used as shortcuts in listing the content of MCC (similarly to simultaneous set and global capture entries)

Examples have been updated. A new example with global capture entries has been added.

<encGroupIDREF> has been made optional.

<single> has been renamed into <individual>

Obsolete comments have been removed.

participants information has been added.

26. Diff with draft-ietf-clue-data-model-schema-04 version

globalCaptureEntries/Entry renamed as globalSceneEntries/Entry;

sceneInformation added;

Only capture scene entry identifiers listed within global scene entries (media capture identifiers removed);

<participants> renamed as <people> in the >clueInfo< template

<vcard> renamed as <personInfo> to synch with the framework terminology

<participantType> renamed as <personType> to synch with the framework terminology

<participantIDs> renamed as <capturedPeople> in the media capture type definition to remove ambiguity

Examples have been updated with the new definitions of <globalSceneEntries> and of <people>.

27. Informative References

[I-D.ietf-clue-framework] Duckworth, M., Pepperell, A., and S. Wenger, "Framework for Telepresence Multi-Streams", draft-ietf-clue-framework-15 (work in progress), May 2014.

[I-D.ietf-ecrit-additional-data] Rosen, B., Tschofenig, H., Marshall, R., Randy, R., and J. Winterbottom, "Additional Data related to an Emergency Call", draft-ietf-ecrit-additional-data-22 (work in progress), April 2014.

[RFC4796] Hautakorpi, J. and G. Camarillo, "The Session Description Protocol (SDP) Content Attribute", RFC 4796, February 2007.

[RFC6351] Perreault, S., "xCard: vCard XML Representation", RFC 6351, August 2011.

Authors' Addresses

Roberta Presta
University of Napoli
Via Claudio 21
Napoli 80125
Italy

EMail: roberta.presta@unina.it

Simon Pietro Romano
University of Napoli
Via Claudio 21
Napoli 80125
Italy

EMail: spromano@unina.it