CLUE Working Group
Internet-Draft
Intended status: Standards Track
Expires: July 31, 2016

R. Presta
S P. Romano
University of Napoli

January 28, 2016

An XML Schema for the CLUE data model
draft-ietf-clue-data-model-schema-12

Abstract

This document 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 July 31, 2016.

Copyright Notice

Copyright (c) 2016 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. Definitions 4
4. XML Schema 6
5. <mediaCaptures> 18
6. <encodingGroups> 18
7. <captureScenes> 18
8. <simultaneousSets> 18
9. <globalViews> 18
10. <captureEncodings> 18
11. <mediaCapture> 18
11.1. captureID attribute 20
11.2. mediaType attribute 20
11.3. <captureSceneIDREF> 20
11.4. <encGroupIDREF> 20
11.5. <spatialInformation> 20
11.5.1. <captureOrigin> 22
11.5.2. <captureArea> 23
11.6. <nonSpatiallyDefinable> 23
11.7. <content> 23
11.8. <synchronizationID> 24
11.9. <allowSubsetChoice> 24
11.10. <policy> 25
11.11. <maxCaptures> 25
11.12. <individual> 26
11.13. <description> 26
11.14. <priority> 27
11.15. <lang> 27
11.16. <mobility> 27
11.17. <relatedTo> 28
11.18. <view> 28
11.19. <presentation> 28
11.19.1. <embeddedText> 29
11.19.2. <capturedPeople> 29
11.20. Audio captures 30
11.20.1. <sensitivityPattern> 30
11.21. Video captures 31
11.22. Text captures 31
11.23. Other capture types 32
11.24. <captureScene> 32
11.24.1. <sceneInformation> 33
11.24.2. <sceneViews> 33
11.24.3. sceneID attribute 34
11.24.4. scale attribute 34
11.25. <sceneView> 35
11.25.1. <mediaCaptureIDs> 35
11.25.2. sceneViewID attribute 36
11.26. <encodingGroup> 36

11.26.1. <maxGroupBandwidth> 36
11.26.2. <encodingIDList> 36
11.26.3. encodingGroupID attribute 37
11.27. <simultaneousSet> 37
11.27.1. setID attribute 38
11.27.2. mediaType attribute 38
11.27.3. <mediaCaptureIDREF> 38
11.27.4. <sceneViewIDREF> 38
11.27.5. <captureSceneIDREF> 38
11.28. <globalView> 38
11.29. <people> 39
11.29.1. <person> 39
12. <captureEncoding> 41
12.1. <captureID> 42
12.2. <encodingID> 42
12.3. <configuredContent> 42
13. <clueInfo> 42
14. XML Schema extensibility 43
14.1. Example of extension 43
15. Security considerations 46
16. IANA considerations 46
16.1. XML namespace registration 46
16.2. XML Schema registration 47
16.3. MIME Media Type Registration for
'application/clue_info+xml' 47
17. Sample XML file 48
18. MCC example 57
19. Diff with draft-ietf-clue-data-model-schema-10 version 67
20. Diff with draft-ietf-clue-data-model-schema-10 version 67
21. Diff with draft-ietf-clue-data-model-schema-09 version 67
22. Diff with draft-ietf-clue-data-model-schema-08 version 68
23. Diff with draft-ietf-clue-data-model-schema-07 version 68
24. Diff with draft-ietf-clue-data-model-schema-06 version 68
25. Diff with draft-ietf-clue-data-model-schema-04 version 69
26. Diff with draft-ietf-clue-data-model-schema-03 version 69
27. Diff with draft-ietf-clue-data-model-schema-02 version 70
28. Acknowledgments 70
29. Informative References 70

1. Introduction

This document provides an XML schema file for the definition of CLUE data model types.

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 document aims at the definition of a coherent structure for information associated with the description of a telepresence scenario. Such information is used within the CLUE protocol messages ([I-D.ietf-clue-protocol]) enabling the dialogue between a Media Provider and a Media Consumer. CLUE protocol messages, indeed, are XML messages allowing (i) a Media Provider to advertise its telepresence capabilities in terms of media captures, capture scenes, and other features envisioned in the CLUE framework, according to the format herein defined and (ii) a Media Consumer to request the desired telepresence options in the form of capture encodings, represented as described in this document.

2. Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

3. Definitions

This document refers to the same definitions used in [I-D.ietf-clue-framework], except for the "CLUE Participant" definition. We briefly recall herein some of the main terms used in the document.

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

Capture: Same as Media Capture.

Capture Device: A device that converts physical input, such as audio, video or text, 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: A structure representing a spatial region captured by one or more Capture Devices, each capturing media representing a portion of the region. The spatial region represented by a Capture Scene MAY correspond to a real region in physical space, such as a room. A Capture Scene includes attributes and one or more Capture Scene Views, with each view including one or more Media Captures.

Capture Scene View: A list of Media Captures of the same media type that together form one way to represent the entire Capture Scene..

CLUE Participant: This term is not imported from the framework terminology. A CLUE Participant identifies a generic entity (either an Endpoint or a MCU) making use of the CLUE protocol.

Consumer: Short for Media Consumer.

Encoding or Individual Encoding: A set of parameters representing a way to encode a Media Capture to become a Capture Encoding.

Encoding Group: A set of encoding parameters representing a total media encoding capability to be sub-divided across potentially multiple Individual Encodings.

Endpoint A CLUE-capable device which is 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 [RFC4353] Participant (which, in turn, includes exactly one SIP User Agent). Endpoints can be anything from multiscreen/multicamera rooms to handheld devices.

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

Media Capture: A source of Media, such as from one or more Capture Devices or constructed from other Media streams.

Media Consumer: A CLUE-capable device that intends to receive Capture Encodings.

Media Provider: A CLUE-capable device that intends to send Capture Encodings.

Multiple Content Capture: A Capture that mixes and/or switches other Captures of a single type. (E.g. all audio or all video.) Particular Media Captures may or may not be present in the resultant Capture Encoding depending on time or space. Denoted as MCCn in the example cases in this document.

Multipoint Control Unit (MCU): A CLUE-capable device that connects two or more endpoints together into one single multimedia conference [RFC7667]. An MCU includes an [RFC4353] like Mixer, without the [RFC4353] requirement to send media to each participant.

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 of the remote video on the Media Consumer's screens, or playing of the remote audio through loudspeakers.

Scene: Same as Capture Scene.

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

Single Media Capture: A capture which contains media from a single source capture device, e.g. an audio capture from a single microphone, a video capture from a single camera.

Spatial Relation: The arrangement in space of two objects, in contrast to relation in time or other relationships.

Stream: A Capture Encoding sent from a Media Provider to a Media Consumer via RTP [RFC3550].

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.

4. XML Schema

This section contains the 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 and the streams that are requested by a Media Consumer given the

Media Provider's ADVERTISEMENT ([I-D.ietf-clue-protocol]).

The main groups of information are:

<mediaCaptures>: the list of media captures available (Section 5)

<encodingGroups>: the list of encoding groups (Section 6)

<captureScenes>: the list of capture scenes (Section 7)

<simultaneousSets>: the list of simultaneous transmission sets (Section 8)

<globalViews>: the list of global views sets (Section 9)

<people>: meta data about the participants represented in the telepresence session (Section 11.29).

<captureEncodings>: the list of instantiated capture encodings (Section 10)

All of the above refers to concepts that have been introduced in [I-D.ietf-clue-framework] and further detailed in 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"
   version="1.0">

<!-- 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="globalViews" type="globalViewsType"/>
<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="captureSceneIDREF" type="xs:IDREF"/>
    <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="policy" type="policyType" minOccurs="0"/>
        <xs:element name="maxCaptures" type="maxCapturesType"
         minOccurs="0"/>
        <xs:element name="allowSubsetChoice" type="xs:boolean"
         minOccurs="0"/>
```

```
      </xs:sequence>
      <xs:element name="individual" type="xs:boolean" fixed="true"/>
    </xs:choice>
    <!-- optional fields -->
    <xs:element name="encGroupIDREF" type="xs:IDREF" minOccurs="0"/>
    <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"
     maxOccurs="unbounded"/>
    <xs:element name="mobility" type="mobilityType" minOccurs="0"/>
    <xs:element name="presentation" type="presentationType"
     minOccurs="0"/>
    <xs:element ref="embeddedText" minOccurs="0"/>
    <xs:element name="view" type="viewType" minOccurs="0"/>
    <xs:element name="capturedPeople" type="capturedPeopleType"
     minOccurs="0"/>
    <xs:element name="relatedTo" type="xs:IDREF" minOccurs="0"/>
  </xs:sequence>
  <xs:attribute name="captureID" type="xs:ID" use="required"/>
  <xs:attribute name="mediaType" type="xs:string" use="required"/>

</xs:complexType>

<!-- POLICY TYPE -->
<xs:simpleType name="policyType">
 <xs:restriction base="xs:string">
      <xs:pattern value="([a-zA-Z0-9])+[:]([0-9])+"/>
    </xs:restriction>
</xs:simpleType>

<!-- CONTENT TYPE -->
<xs:complexType name="contentType">
 <xs:sequence>
   <xs:element name="captureIDREF" type="xs:IDREF"
   minOccurs="0" maxOccurs="unbounded"/>
   <xs:element name="sceneViewIDREF" 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>

<!-- MAX CAPTURES TYPE -->
<xs:complexType name="maxCapturesType">
  <xs:simpleContent>
    <xs:extension base="xs:unsignedShort">
      <xs:minInclusive value="1"/>
```

```
      <xs:attribute name="exactNumber" type="xs:boolean"/>
    </xs:extension>
  </xs:simpleContent>
</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" use="required"/>
  <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:enumeration value="observer"/>
        </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:restriction>
</xs:simpleType>

<!-- SPATIAL INFORMATION TYPE -->
<xs:complexType name="spatialInformationType">
 <xs:sequence>
  <xs:element name="captureOrigin" type="captureOriginType"
   minOccurs="0"/>
  <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>


<!-- 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 ORIGIN TYPE -->
<xs:complexType name="captureOriginType">
  <xs:sequence>
        <xs:element name="capturePoint" type="pointType"></xs:element>
        <xs:element name="lineOfCapturePoint" type="pointType"
         minOccurs="0">
        </xs:element>
```

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


<!-- 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>


<!-- 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:sequence>
  <xs:any namespace="##other" processContents="lax" minOccurs="0"
    maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:anyAttribute namespace="##other" processContents="lax"/>
  </xs:extension>
 </xs:complexContent>
</xs:complexType>


<!-- OTHER CAPTURE TYPE -->
<xs:complexType name="otherCaptureType">
 <xs:complexContent>
  <xs:extension base="tns:mediaCaptureType">
  <xs:sequence>
  <xs:any namespace="##other" processContents="lax" minOccurs="0"
    maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:anyAttribute namespace="##other" processContents="lax"/>
```

```
   </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="sensitivyPattern" type="sensitivyPatternType"
     minOccurs="0"/>
     <xs:any namespace="##other" processContents="lax" minOccurs="0"
     maxOccurs="unbounded"/>
    </xs:sequence>
    <xs:anyAttribute namespace="##other" processContents="lax"/>
   </xs:extension>
  </xs:complexContent>
 </xs:complexType>


 <!-- SENSITIVITY PATTERN TYPE -->
 <xs:simpleType name="sensitivyPatternType">
  <xs:restriction base="xs:string">
   <xs:enumeration value="uni"/>
   <xs:enumeration value="shotgun"/>
   <xs:enumeration value="omni"/>
   <xs:enumeration value="figure8"/>
   <xs:enumeration value="cardioid"/>
   <xs:enumeration value="hyper-cardioid"/>
  </xs:restriction>
 </xs:simpleType>


 <!-- VIDEO CAPTURE TYPE -->
 <xs:complexType name="videoCaptureType">
  <xs:complexContent>
   <xs:extension base="tns:mediaCaptureType">
    <xs:sequence>
     <xs:any namespace="##other" processContents="lax" minOccurs="0"
     maxOccurs="unbounded"/>
    </xs:sequence>
    <xs:anyAttribute namespace="##other" processContents="lax"/>
   </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="sceneViews" type="sceneViewsType" minOccurs="0"/>
   <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="mm"/>
   <xs:enumeration value="unknown"/>
   <xs:enumeration value="noscale"/>
  </xs:restriction>
 </xs:simpleType>



 <!-- SCENE VIEWS TYPE -->
 <!-- envelope of scene views of a capture scene -->
 <xs:complexType name="sceneViewsType">
  <xs:sequence>
```

```
   <xs:element name="sceneView" type="sceneViewType"
   maxOccurs="unbounded"/>
  </xs:sequence>
 </xs:complexType>

 <!-- SCENE VIEW TYPE -->
 <xs:complexType name="sceneViewType">
  <xs:sequence>
   <xs:element ref="description" minOccurs="0" maxOccurs="unbounded"/>
   <xs:element name="mediaCaptureIDs" type="captureIDListType"/>
  </xs:sequence>
  <xs:attribute name="sceneViewID" type="xs:ID" 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:unsignedLong"/>
   <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="mediaCaptureIDREF" type="xs:IDREF"
    minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="sceneViewIDREF" type="xs:IDREF"
    minOccurs="0" maxOccurs="unbounded"/>
    <xs:element name="captureSceneIDREF" 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 VIEWS TYPE -->
 <xs:complexType name="globalViewsType">
  <xs:sequence>
   <xs:element name="globalView" type="globalViewType"
   maxOccurs="unbounded"/>
  </xs:sequence>
 </xs:complexType>

 <!-- GLOBAL VIEW TYPE -->
 <xs:complexType name="globalViewType">
  <xs:sequence>
    <xs:element name="sceneViewIDREF" type="xs:IDREF"
     maxOccurs="unbounded"/>
    <xs:any namespace="##other" processContents="lax" minOccurs="0"
    maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:attribute name="globalViewID" type="xs:ID"/>
  <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" use="required"/>
 <xs:anyAttribute namespace="##any" processContents="lax"/>
</xs:complexType>

<!-- CLUE INFO ELEMENT -->
<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="globalViews" 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>
```

Following sections describe the XML schema in more detail. As a general remark, please notice that optional elements that don't define what their absence means are intended to be associated with undefined properties.

5. <mediaCaptures>

<mediaCaptures> represents the list of one or more media captures available at the Media Provider's side. Each media capture is represented by a <mediaCapture> element (Section 11).

6. <encodingGroups>

<encodingGroups> represents the list of the encoding groups organized on the Media Provider's side. Each encoding group is represented by an <encodingGroup> element (Section 11.26).

7. <captureScenes>

<captureScenes> represents the list of the capture scenes organized on the Media Provider's side. Each capture scene is represented by a <captureScene> element. (Section 11.24).

8. <simultaneousSets>

<simultaneousSets> contains the simultaneous sets indicated by the Media Provider. Each simultaneous set is represented by a <simultaneousSet> element. (Section 11.27).

9. <globalViews>

<globalViews> 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 view" and it is represented by a <globalView> element. (Section 11.28).

10. <captureEncodings>

<captureEncodings> is a list of capture encodings. It can represent 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 12).

11. <mediaCapture>

A Media Capture is the fundamental representation of a media flow that is available on 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="captureSceneIDREF" type="xs:IDREF"/>
    <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="policy" type="policyType" minOccurs="0"/>
        <xs:element name="maxCaptures" type="maxCapturesType" minOccurs="0"/>
        <xs:element name="allowSubsetChoice" type="xs:boolean" minOccurs="0"/>
      </xs:sequence>
      <xs:element name="individual" type="xs:boolean" fixed="true"/>
    </xs:choice>
    <!-- optional fields -->
    <xs:element name="encGroupIDREF" type="xs:IDREF" minOccurs="0"/>
    <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"
     maxOccurs="unbounded"/>
    <xs:element name="mobility" type="mobilityType" minOccurs="0"/>
    <xs:element name="presentation" type="presentationType" minOccurs="0"/>
    <xs:element ref="embeddedText" minOccurs="0"/>
    <xs:element name="view" type="viewType" minOccurs="0"/>
    <xs:element name="capturedPeople" type="capturedPeopleType" minOccurs="0"/>
    <xs:element name="relatedTo" type="xs:IDREF" minOccurs="0"/>
  </xs:sequence>
  <xs:attribute name="captureID" type="xs:ID" use="required"/>
  <xs:attribute name="mediaType" type="xs:string" use="required"/>
</xs:complexType>
```

11.1. captureID attribute

The "captureID" attribute is a mandatory field containing the identifier of the media capture. Such an identifier serves as the way the capture is referenced from other data model elements (e.g., simultaneous sets, capture encodings, and others).

11.2. mediaType attribute

The "mediaType" attribute is a mandatory attribute specifying the media type of the capture. Common values are "audio", "video", "text". Other values can be provided. It is assumed that implementations agree on the interpretation of those other values.

11.3. <captureSceneIDREF>

<captureSceneIDREF> is a mandatory field containing the value of the identifier of the capture scene the media capture is defined in, i.e., the value of the sceneID (Section 11.24.3) attribute of that capture scene. Indeed, each media capture must be defined within 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 11.5). Such coordinates refer to the space of coordinates defined for the capture scene containing the capture.

11.4. <encGroupIDREF>

<encGroupIDREF> is an optional field containing the identifier of the encoding group the media capture is associated with, i.e., the value of the encodingGroupID (Section 11.26.3) attribute of that encoding group. Media captures that are not associated with any encoding group can not be instantiated as media streams.

11.5. <spatialInformation>

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

Captures are spatially definable when at least (i) it is possible to provide the coordinates of the device position within the telepresence room of origin (capture point) together with its capturing direction specified by a second point (point on line of capture), or (ii) it is possible to provide the represented area within the telepresence room, by listing the coordinates of the four co-planar points identifying the plane of interest (area of capture). The coordinates of the abovementioned points must be 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, registered presentations, or external streams that are played in the telepresence room and transmitted to remote sites.

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. By comparing the <spatialInformation> element of different media captures within the same capture scene, a consumer can better determine the spatial relationships between them and render them correctly. Non spatially definable captures do not embed such element in their XML description: they are instead characterized by having the <nonSpatiallyDefinable> tag set to "true" (see Section 11.6).

The definition of the spatial information type is the following:

```
<!-- SPATIAL INFORMATION TYPE -->
<xs:complexType name="spatialInformationType">
 <xs:sequence>
  <xs:element name="captureOrigin" type="captureOriginType"
   minOccurs="0"/>
  <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 <captureOrigin> contains the coordinates of the capture device that is taking the capture (i.e., the capture point), as well as, optionally, the pointing direction (i.e., the point on line of capture) (see Section 11.5.1).

The <captureArea> is an optional field containing four points defining the captured area covered by the capture (see Section 11.5.2).

The scale of the points coordinates is specified in the scale (Section 11.24.4) attribute of the capture scene the media capture belongs to. Indeed, all the spatially definable media captures referring to the same capture scene share the same coordinate system

and express their spatial information according to the same scale.

11.5.1. <captureOrigin>

The <captureOrigin> element is used to represent the position and optionally the line of capture of a capture device. <captureOrigin> MUST be included in spatially definable audio captures, while it is optional for spatially definable video captures.

The XML Schema definition of the <captureOrigin> element type is the following:

```
<!-- CAPTURE ORIGIN TYPE -->
<xs:complexType name="captureOriginType">
  <xs:sequence>
        <xs:element name="capturePoint" type="pointType"/>
        <xs:element name="lineOfCapturePoint" type="pointType"
         minOccurs="0"/>
  </xs:sequence>
  <xs:anyAttribute namespace="##any" processContents="lax"/>
</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>
```

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

The <captureOrigin> element includes a mandatory <capturePoint> element and an optional <lineOfCapturePoint> element, both of the type "pointType". <capturePoint> specifies the three coordinates identifying the position of the capture device. <lineOfCapturePoint> is another pointType 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. For a spatially definable video

capture, if the point on line of capture is provided, it MUST belong to the region between the point of capture and the capture area. For a spatially definable audio capture, if the point on line of capture is not provided, the sensitivity pattern should be considered omnidirectional.

11.5.2. <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. <captureArea> MUST be included in the spatial information of spatially definable video captures, while it MUST NOT be associated with audio captures.

The XML representation of that area is provided through a set of four point-type elements, <bottomLeft>, <bottomRight>, <topLeft>, and <topRight> that MUST be co-planar. The four coplanar points are identified from the perspective of the capture device. The XML schema definition is the following:

```
<!-- 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>
```

11.6. <nonSpatiallyDefinable>

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

11.7. <content>

A media capture can be (i) an individual media capture or (ii) a multiple content capture (MCC). A multiple content capture is made by different captures that can be arranged spatially (by a composition operation), or temporally (by a switching operation), or

that can result from the orchestration of both the techniques. If a media capture is an MCC, then it MAY show in its XML data model representation the <content> element. It is composed by a list of media capture identifiers ("captureIDREF") and capture scene view identifiers ("sceneViewIDREF"), where the last ones are used as shortcuts to refer to multiple capture identifiers. The referenced captures are used to create the MCC according to a certain strategy. If the <content> element does not appear in a MCC, or it has no child elements, then the MCC is assumed to be made of multiple sources but no information regarding those sources is provided.

```
<!-- CONTENT TYPE -->
<xs:complexType name="contentType">
 <xs:sequence>
   <xs:element name="captureIDREF" type="xs:IDREF"
   minOccurs="0" maxOccurs="unbounded"/>
   <xs:element name="sceneViewIDREF" 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>
```

11.8. <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 all times captures coming from the same sources. It is the Media Provider that determines what the source for the captures is. In this way, the Media Provider can choose how to group together single captures for the purpose of keeping them synchronized according to the <synchronizationID> element.

11.9. <allowSubsetChoice>

<allowSubsetChoice> is an optional boolean element for multiple content captures. It indicates whether or not the Provider allows the Consumer to choose a specific subset of the captures referenced by the MCC. If this attribute is true, and the MCC references other captures, then the Consumer MAY specify in a CONFIGURE message a specific subset of those captures to be included in the MCC, and the Provider MUST then include only that subset. If this attribute is

false, or the MCC does not reference other captures, then the Consumer MUST NOT select a subset. If <allowSubsetChoice> is not shown in the XML description of the MCC, its value is to be considered "false".

11.10. <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>. The <policy> value is in the form of a token that indicates the policy and an index representing an instance of the policy, separated by a ":" (e.g., SoundLevel:2, RoundRobin:0, etc.). The XML schema defining the type of the <policy> element is the following:

```
<!-- POLICY TYPE -->
<xs:simpleType name="policyType">
 <xs:restriction base="xs:string">
      <xs:pattern value="([a-zA-Z0-9])+[:]([0-9])+"/>
    </xs:restriction>
</xs:simpleType>
```

At the time of writing, only two switching policies are defined in [I-D.ietf-clue-framework]:

SoundLevel: the content of the MCC is determined by a sound level detection algorithm. The loudest (active) speaker (or a previous speaker, depending on the index value) is contained in the MCC. Index 0 represents the most current instance of the policy, i.e., the currently active speaker, 1 represents the previous instance, i.e., the previous active speaker, and so on.

RoundRobin: the content of the MCC is determined by a time based algorithm.

Other values for the <policy> element can be used. In this case, it is assumed that implementations agree on the meaning of those other values and/or those new switching policies are defined in later documents.

11.11. <maxCaptures>

<maxCaptures> is an optional element that can be used only for multiple content captures (MCC). It provides information about the number of media captures that can be represented in the multiple

content capture at a time. If <maxCaptures> is not provided, all the media captures listed in the <content> element can appear at a time in the capture encoding. The type definition is provided below.

```
<!-- MAX CAPTURES TYPE -->
<xs:complexType name="maxCapturesType">
  <xs:simpleContent>
    <xs:extension base="xs:unsignedShort">
      <xs:minInclusive value="1"/>
      <xs:attribute name="exactNumber" type="xs:boolean"/>
    </xs:extension>
  </xs:simpleContent>
</xs:complexType>
```

When the "exactNumber" attribute is set to "true", it means the <maxCaptures> element carries the exact number of the media captures appearing at a time. Otherwise, the number of the represented media captures MUST be considered "<=" the <maxCaptures> value.

For instance, an audio MCC having the <maxCaptures> value set to 1 means that a media stream from the MCC will only contain audio from a single one of its constituent captures at a time. On the other hand, if the <maxCaptures> value is set to 4 and the exactNumber attribute is set to "true", it would mean that the media stream received from the MCC will always contain a mix of audio from exactly four of its constituent captures.

11.12. <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 the capture that is being described is not a multiple content capture. Indeed, <individual> and the aforementioned tags related to MCC attributes (from Section 11.7 to Section 11.11) are mutually exclusive, according to the <choice> section within the XML Schema definition of the media capture type.

11.13. <description>

<description> is used to provide human-readable textual information. This element is included in the XML definition of media captures, capture scenes and capture scene views to the aim of providing human-readable description of, respectively, media captures, capture scenes and capture scene views. According to the data model definition of a media capture (Section 11)), zero or more <description> elements can

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

11.14. <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. If no priority is assigned, no assumptions regarding relative importance of the media capture can be assumed.

11.15. <lang>

<lang> is an optional element containing the language used in the capture. Zero or more <lang> elements can appear in the XML description of a media capture.

11.16. <mobility>

<mobility> is an optional element indicating whether or not the capture device originating the capture may move during the telepresence session. That optional element can assume one of the three following values:

static SHOULD NOT change for the duration of the CLUE session, across multiple ADVERTISEMENT messages.

dynamic MAY change in each new ADVERTISEMENT message. Can be assumed to remain unchanged until there is a new ADVERTISEMENT message.

highly-dinamic MAY change dynamically, even between consecutive ADVERTISEMENT messages. The spatial information provided in an ADVERTISEMENT message is simply a snapshot of the current values at the time when the message is sent.

11.17. <relatedTo>

The optional <relatedTo> element contains the value of the captureID attribute (Section 11.1) of the media capture to which the considered media capture refers. The media capture marked with a <relatedTo> element can be for example the translation of the referred media capture in a different language.

11.18. <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.

```
<!-- 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>
```

11.19. <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.

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

11.19.1. <embeddedText>

The <embeddedText> element is a boolean element indicating that there is text embedded in the media capture (e.g., in a video capture). The language used in such embedded textual description is reported in <embeddedText> "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>
```

11.19.2. <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.

11.19.2.1. <personIDREF>

<personIDREF> contains the identifier of the represented person, i.e., the value of the related personID attribute (Section 11.29.1.1). Metadata about the represented participant can be retrieved by accessing the <people> list (Section 11.29).

11.20. 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="sensitivyPattern" type="sensitivyPatternType"
    minOccurs="0"/>
    <xs:any namespace="##other" processContents="lax" minOccurs="0"
    maxOccurs="unbounded"/>
   </xs:sequence>
   <xs:anyAttribute namespace="##other" processContents="lax"/>
  </xs:extension>
 </xs:complexContent>
</xs:complexType>
```

An example of audio-specific information that can be included is represented by the <sensitivyPattern> element. (Section 11.20.1).

11.20.1. <sensitivityPattern>

The <sensitivityPattern> element is an optional field describing the characteristics of the nominal sensitivity pattern of the microphone capturing the audio signal.

The XML Schema definition is provided below:

```
<xs:simpleType name="sensitivyPatternType">
 <xs:restriction base="xs:string">
  <xs:enumeration value="uni"/>
  <xs:enumeration value="shotgun"/>
  <xs:enumeration value="omni"/>
  <xs:enumeration value="figure8"/>
  <xs:enumeration value="cardioid"/>
  <xs:enumeration value="hyper-cardioid"/>
 </xs:restriction>
</xs:simpleType>
```

11.21. Video captures

Video captures, similarly to audio captures, extend the information of a generic media capture with video-specific features.

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:any namespace="##other" processContents="lax" minOccurs="0"
    maxOccurs="unbounded"/>
   </xs:sequence>
   <xs:anyAttribute namespace="##other" processContents="lax"/>
  </xs:extension>
 </xs:complexContent>
</xs:complexType>
```

11.22. 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:sequence>
  <xs:any namespace="##other" processContents="lax" minOccurs="0"
    maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:anyAttribute namespace="##other" processContents="lax"/>
  </xs:extension>
 </xs:complexContent>
</xs:complexType>
```

Text captures SHOULD be marked as non spatially definable (i.e., they should present in their XML description the <nonSpatiallyDefinable> (Section 11.6) element set to "true").

11.23. Other capture types

Other media capture types can be described by using the CLUE data model. They can be represented by exploiting the "otherCaptureType" type. This media capture type is conceived to be filled in with elements defined within extensions of the current schema, i.e., with elements defined in other XML schemas (see Section 14 for an example). The otherCaptureType inherits all the features envisioned for the abstract mediaCaptureType.

The XML Schema representation of the otherCaptureType is the following:

```
<!-- OTHER CAPTURE TYPE -->
<xs:complexType name="otherCaptureType">
 <xs:complexContent>
  <xs:extension base="tns:mediaCaptureType">
  <xs:sequence>
  <xs:any namespace="##other" processContents="lax" minOccurs="0"
    maxOccurs="unbounded"/>
  </xs:sequence>
  <xs:anyAttribute namespace="##other" processContents="lax"/>
  </xs:extension>
 </xs:complexContent>
</xs:complexType>
```

When defining new media capture types that are going to be described by means of the <otherMediaCapture> element, spatial properties of such new media capture types SHOULD be defined (e.g., whether or not they are spatially definable, whether or not they should be associated with an area of capture, etc.).

11.24. <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 media captures and capture scene views, that are sets of media captures of the same media type. Each capture scene view is an alternative to represent completely 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="sceneViews" type="sceneViewsType" minOccurs="0"/>
  <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>
```

Each capture scene is identified by a "sceneID" attribute. The <captureScene> element can contain zero or more textual <description> elements, defined as in Section 11.13. Besides <description>, there is the optional <sceneInformation> element (Section 11.24.1), which contains structured information about the scene in the vcard format, and the optional <sceneViews> element (Section 11.24.2), which is the list of the capture scene views. When no <sceneViews> is provided, the capture scene is assumed to be made of all the media captures which contain the value of its sceneID attribute in their mandatory captureSceneIDREF attribute.

11.24.1. <sceneInformation>

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

11.24.2. <sceneViews>

The <sceneViews> element is a mandatory field of a capture scene containing the list of scene views. Each scene view is represented by a <sceneView> element (Section 11.25).

```
<!-- SCENE VIEWS TYPE -->
<!-- envelope of scene views of a capture scene -->
<xs:complexType name="sceneViewsType">
 <xs:sequence>
  <xs:element name="sceneView" type="sceneViewType"
  maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>
```

11.24.3. sceneID attribute

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

11.24.4. 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:

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

"unknown" - the scale is the same for every media capture in the capture scene but the unity of measure is undefined. Systems which are not aware of specific physical dimensions yet 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 could be different for each media capture.

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

11.25. <sceneView>

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

A <sceneView> element is characterized as follows.

```
<!-- SCENE VIEW TYPE -->
<xs:complexType name="sceneViewType">
 <xs:sequence>
  <xs:element ref="description" minOccurs="0" maxOccurs="unbounded"/>
  <xs:element name="mediaCaptureIDs" type="captureIDListType"/>
 </xs:sequence>
 <xs:attribute name="sceneViewID" type="xs:ID" use="required"/>
</xs:complexType>
```

One or more optional <description> elements provide human-readable information about what the scene view contains. <description> is defined as already seen in Section 11.13.

The remaining child elements are described in the following subsections.

11.25.1. <mediaCaptureIDs>

The <mediaCaptureIDs> is the list of the identifiers of the media captures included in the scene view. It is an element of the captureIDListType type, which is defined as a sequence of <captureIDREF>, 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>
```

11.25.2. sceneViewID attribute

The sceneViewID attribute is a mandatory attribute containing the identifier of the capture scene view represented by the <sceneView> element.

11.26. <encodingGroup>

The <encodingGroup> element represents an encoding group, which is made by 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 media captures. The definition of the <encodingGroup> element is the following:

```
<!-- ENCODING GROUP TYPE -->
<xs:complexType name="encodingGroupType">
 <xs:sequence>
  <xs:element name="maxGroupBandwidth" type="xs:unsignedLong"/>
  <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 contained elements are further described.

11.26.1. <maxGroupBandwidth>

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

11.26.2. <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:string" maxOccurs="unbounded"/>
 </xs:sequence>
</xs:complexType>
```

11.26.3. encodingGroupID attribute

The encodingGroupID attribute contains the identifier of the encoding group.

11.27. <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="mediaCaptureIDREF" type="xs:IDREF"
   minOccurs="0" maxOccurs="unbounded"/>
   <xs:element name="sceneViewIDREF" type="xs:IDREF"
   minOccurs="0" maxOccurs="unbounded"/>
   <xs:element name="captureSceneIDREF" 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>
```

Besides the identifiers of the captures (<mediaCaptureIDREF> elements), also the identifiers of capture scene views and of capture scene can be exploited as shortcuts (<sceneViewIDREF> and <captureSceneIDREF> elements). As an example, let's consider the situation where there are two capture scene views (S1 and S7). S1 contains captures AC11, AC12, AC13. S7 contains captures AC71, AC72. Provided that AC11, AC12, AC13, AC71, AC72 can be simultaneously sent to the media consumer, instead of having 5 <mediaCaptureIDREF>

elements listed in the simultaneous set (i.e., one <mediaCaptureIDREF> for AC11, one for AC12, and so on), there can be just two <sceneViewIDREF> elements (one for S1 and one for S7).

11.27.1. setID attribute

The "setID" attribute is a mandatory field containing the identifier of the simultaneous set.

11.27.2. mediaType attribute

The "mediaType" attribute is an optional attribute containing the media type of the captures referenced by the simultaneous set.

When only capture scene identifiers are listed within a simultaneous set, the media type attribute MUST appear in the XML description in order to determine which media captures can be simultaneously sent together.

11.27.3. <mediaCaptureIDREF>

<mediaCaptureIDREF> contains the identifier of the media capture that belongs to the simultanous set.

11.27.4. <sceneViewIDREF>

<sceneViewIDREF> contains the identifier of the scene view containing a group of captures that are able to be sent simultaneously with the other captures of the simultaneous set.

11.27.5. <captureSceneIDREF>

<captureSceneIDREF> contains the identifier of the capture scene where all the included captures of a certain media type are able to be sent together with the other captures of the simultaneous set.

11.28. <globalView>

<globalView> is a set of captures of the same media type representing a summary of the complete Media Provider's offer. The content of a global view is expressed by leveraging only scene view identifiers, put within <sceneViewIDREF> elements. Each global view is identified by a unique identifier within the "globalViewID" attribute.

```
<!-- GLOBAL VIEW TYPE -->
<xs:complexType name="globalViewType">
 <xs:sequence>
   <xs:element name="sceneViewIDREF" type="xs:IDREF"
    maxOccurs="unbounded"/>
   <xs:any namespace="##other" processContents="lax" minOccurs="0"
   maxOccurs="unbounded"/>
 </xs:sequence>
 <xs:attribute name="globalViewID" type="xs:ID"/>
 <xs:anyAttribute namespace="##any" processContents="lax"/>
</xs:complexType>
```

11.29. <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 depicted below, for each participant, a <person> element is provided.

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

11.29.1. <person>

<person> includes all the metadata related to a person represented within one or more media captures. Such element provides the vcard of the subject (via the <personInfo> element, see Section 11.29.1.2) and his conference role(s) (via one or more <personType> elements, see Section 11.29.1.3). Furthermore, it has a mandatory "personID" attribute (Section 11.29.1.1).

```
<!-- 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" use="required"/>
  <xs:anyAttribute namespace="##other" processContents="lax"/>
</xs:complexType>
```

11.29.1.1. personID attribute

The "personID" attribute carries the identifier of a represented person. Such an identifier can be used to refer to the participant, as in the <capturedPeople> element in the media captures representation (Section 11.19.2).

11.29.1.2. <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.

11.29.1.3. <personType>

The value of the <personType> element determines the role of the represented participant within the telepresence session organization. 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.

```
<!-- 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:enumeration value="observer"/>
        </xs:restriction>
</xs:simpleType>
```

12. <captureEncoding>

A capture encoding is given from the association of a media capture with an individual encoding, to form a capture stream as defined in [I-D.ietf-clue-framework]. Capture encodings are used within CONFIGURE messages from a Media Consumer to a Media Provider for representing the streams desired by the Media Consumer. For each desired stream, the Media Consumer needs to be allowed to specify: (i) the capture identifier of the desired capture that has been advertised by the Media Provider; (ii) the encoding identifier of the encoding to use, among those advertised by the Media Provider; (iii) optionally, in case of multi-content captures, the list of the capture identifiers of the desired captures. All the mentioned identifiers are intended to be included in the ADVERTISEMENT message that the CONFIGURE message refers to. The XML model of <captureEncoding> 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" use="required"/>
 <xs:anyAttribute namespace="##any" processContents="lax"/>
```

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

12.1. <captureID>

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

12.2. <encodingID>

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

12.3. <configuredContent>

<configuredContent> is an optional element to be used in case of configuration of MCC. It contains the list of capture identifiers and capture scene view identifiers the Media Consumer wants within the MCC. That element is structured as the <content> element used to describe the content of an 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.

13. <clueInfo>

The <clueInfo> element includes all the information needed to represent the Media Provider's description of its telepresence capabilities according to the CLUE framework. Indeed, it is made by:

- the list of the available media captures (<mediaCaptures> (Section 5))

- the list of encoding groups (<encodingGroups> (Section 6))

- the list of capture scenes (<captureScenes> (Section 7))

- the list of simultaneous transmission sets (<simultaneousSets> (Section 8))

- the list of global views sets (<globalViews> (Section 9))

- meta data about the participants represented in the telepresence session (<people> (Section 11.29)).

It has been conceived only for data model testing purposes and, though it resembles the body of an ADVERTISEMENT message, it is not actually used in the CLUE protocol message definitions. The

telepresence capabilities descriptions compliant to this data model specification that can be found in Section 17 and Section 18 are provided by using the <clueInfo> element.

```
<!-- 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="globalViews" 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>
```

14. XML Schema extensibility

The telepresence data model defined in this document is meant to be extensible. Extensions are accomplished by defining elements or attributes qualified by namespaces other than "urn:ietf:params:xml:ns:clue-info" and "urn:ietf:params:xml:ns:vcard-4.0" for use wherever the schema allows such extensions (i.e., where the XML Schema definition specifies "anyAttribute" or "anyElement"). Elements or attributes from unknown namespaces MUST be ignored.

14.1. Example of extension

When extending the CLUE data model, a new schema with a new namespace associated with it needs to be specified.

In the following, an example of extension is provided. The extension defines a new audio capture attribute ("newAudioFeature") and an attribute for characterizing the captures belonging to an "otherCaptureType" defined by the user. An XML document compliant with the extension is also included. The XML file results validated against the current CLUE data model schema.

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

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

<!-- Import CLUE XML schema -->
<xs:import namespace="urn:ietf:params:xml:ns:clue-info"
schemaLocation="clue-data-model-schema.xsd"/>

<!-- ELEMENT DEFINITIONS -->
<xs:element name="newAudioFeature" type="xs:string"/>
<xs:element name="otherMediaCaptureTypeFeature" type="xs:string"/>

</xs:schema>
```

```
<?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"
xmlns:ns3="urn:ietf:params:xml:ns:clue-info-ext"
clueInfoID="NapoliRoom">
    <mediaCaptures>
        <mediaCapture
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="audioCaptureType"
        captureID="AC0"
        mediaType="audio">
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <nonSpatiallyDefinable>true</nonSpatiallyDefinable>
            <individual>true</individual>
            <encGroupIDREF>EG1</encGroupIDREF>
            <ns3:newAudioFeature>newAudioFeatureValue
            </ns3:newAudioFeature>
        </mediaCapture>
        <mediaCapture
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="otherCaptureType"
        captureID="OMC0"
        mediaType="other media type">
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <nonSpatiallyDefinable>true</nonSpatiallyDefinable>
            <encGroupIDREF>EG1</encGroupIDREF>
            <ns3:otherMediaCaptureTypeFeature>OtherValue
            </ns3:otherMediaCaptureTypeFeature>
        </mediaCapture>
    </mediaCaptures>
    <encodingGroups>
        <encodingGroup encodingGroupID="EG1">
            <maxGroupBandwidth>300000</maxGroupBandwidth>
            <encodingIDList>
                <encID>ENC4</encID>
                <encID>ENC5</encID>
            </encodingIDList>
        </encodingGroup>
    </encodingGroups>
    <captureScenes>
        <captureScene scale="unknown" sceneID="CS1"/>
    </captureScenes>
</clueInfo>
```

15. Security considerations

This document defines an XML Schema data model for telepresence scenarios. The modeled information is identified in the CLUE framework as necessary in order to enable a full-optional media stream negotiation and rendering. Indeed, the XML elements herein defined are used within CLUE protocol messages to describe both the media streams representing the Media Provider's telepresence offer and the desired selection requested by the Media Consumer. Security concerns described in [I-D.ietf-clue-framework], Section 15, apply to this document.

Data model information carried within CLUE messages SHOULD be accessed only by authenticated endpoints. Indeed, some information published by the Media Provider might reveal sensitive data about who and what is represented in the transmitted streams. The vCard included in the <personInfo> elements (Section 11.29.1) mandatorily contains the identity of the represented person. Optionally vCards can also carry the person's contact addresses, together with his/her photo and other personal data. Similar privacy-critical information can be conveyed by means of <sceneInformation> elements (Section 11.24.1) describing the capture scenes. The <description> elements (Section 11.13) also can specify details about the content of media captures , capture scenes and scene views that should be protected.

Integrity attacks to the data model information encapsulated in CLUE messages can invalidate the success of the telepresence session's setup by misleading the Media Consumer's and Media Provider's interpretation of the offered and desired media streams.

The assurance of the authenticated access and of the integrity of the data model information is up to the involved transport mechanisms, namely the CLUE protocol [I-D.ietf-clue-protocol] and the CLUE data channel [I-D.ietf-clue-datachannel].

16. IANA considerations

This document registers a new XML namespace, a new XML schema and the MIME type for the schema.

16.1. XML namespace registration

URI: urn:ietf:params:xml:ns:clue-info

Registrant Contact: IETF CLUE Working Group <clue@ietf.org>, Roberta Presta <roberta.presta@unina.it>

XML:

BEGIN

```
<?xml version="1.0"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML Basic 1.0//EN"
 "http://www.w3.org/TR/xhtml-basic/xhtml-basic10.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
 <head>
  <meta http-equiv="content-type"
        content="text/html;charset=iso-8859-1"/>
  <title> CLUE Data Model Namespace</title>
 </head>
 <body>
  <h1>Namespace for CLUE Data Model</h1>
  <h2>urn:ietf:params:xml:ns:clue-info</h2>
  <p>See
   <a href="http://www.rfc-editor.org/rfc/rfcXXXX.txt"> RFC XXXX</a>.
   <!--[[NOTE TO IANA/RFC-EDITOR: Please update RFC URL
    and replace XXXX with the RFC number for this specification.]]-->
  </p>
 </body>
</html>
```

END

16.2. XML Schema registration

This section registers an XML schema per the guidelines in [RFC3688].

URI: urn:ietf:params:xml:schema:clue-info

Registrant Contact: CLUE working group (clue@ietf.org), Roberta Presta (roberta.presta@unina.it).

Schema: The XML for this schema can be found as the entirety of Section 4 of this document.

16.3. MIME Media Type Registration for 'application/clue_info+xml'

This section registers the "application/clue_info+xml" MIME type.

To: ietf-types@iana.org

Subject: Registration of MIME media type application/clue+xml

MIME media type name: application

MIME subtype name: clue_info+xml

Required parameters: (none)

Optional parameters: charset
Same as the charset parameter of "application/xml" as specified in [RFC7303], Section 3.2.

Encoding considerations: Same as the encoding considerations of "application/xml" as specified in [RFC7303], Section 3.2.

Security considerations: This content type is designed to carry data related to telepresence information. Some of the data could be considered private. This media type does not provide any protection and thus other mechanisms such as those described in Section 15 are required to protect the data. This media type does not contain executable content.

Interoperability considerations: None.

Published specification: RFC XXXX [[NOTE TO IANA/RFC-EDITOR: Please replace XXXX with the RFC number for this specification.]]

Applications that use this media type: None.

Additional Information: Magic Number(s): (none),
File extension(s): .clue,
Macintosh File Type Code(s): TEXT.

Person & email address to contact for further information: Roberta Presta (roberta.presta@unina.it).

Intended usage: LIMITED USE

Author/Change controller: The IETF

Other information: This media type is a specialization of application/xml [RFC7303], and many of the considerations described there also apply to application/clue_info+xml.

17. 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]), it is provided the XML representation of an endpoint-style Media Provider's

ADVERTISEMENT.

There are three cameras, where the central one is also capable of capturing a zoomed-out view of the overall telepresence room. Besides the three video captures coming from the cameras, the Media Provider makes available a further multi-content capture of 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 views:

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, 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). On the other hand, VC3 and VC4 cannot be simultaneous either, since VC3, the loudest segment of the room, might be at a certain point in time focusing on the central part of the room, i.e., the same as VC1. The simultaneous sets would then be the following:

SS1 made by VC3 and all the captures in the first capture scene view (VC0,VC1,VC2);

SS2 made by VC0, VC2, VC4

```
<?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="audioCaptureType" captureID="AC0" mediaType="audio">
        <captureSceneIDREF>CS1</captureSceneIDREF>
        <spatialInformation>
            <captureOrigin>
                    <capturePoint>
                    <x>0.0</x>
                    <y>0.0</y>
                    <z>10.0</z>
                </capturePoint>
                <lineOfCapturePoint>
                    <x>0.0</x>
                    <y>1.0</y>
                    <z>10.0</z>
                </lineOfCapturePoint>
            </captureOrigin>
        </spatialInformation>
        <individual>true</individual>
        <encGroupIDREF>EG1</encGroupIDREF>
        <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>
    </mediaCapture>
    <mediaCapture
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
     xsi:type="videoCaptureType" captureID="VC0" mediaType="video">
        <captureSceneIDREF>CS1</captureSceneIDREF>
        <spatialInformation>
            <captureOrigin>
                    <capturePoint>
                    <x>-2.0</x>
                    <y>0.0</y>
                    <z>10.0</z>
                </capturePoint>
            </captureOrigin>
            <captureArea>
                    <bottomLeft>
                            <x>-3.0</x>
```

```
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomLeft>
                        <bottomRight>
                                <x>-1.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomRight>
                        <topLeft>
                                <x>-3.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topLeft>
                        <topRight>
                                <x>-1.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topRight>
                </captureArea>
            </spatialInformation>
            <individual>true</individual>
            <encGroupIDREF>EG0</encGroupIDREF>
            <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>
        </mediaCapture>
        <mediaCapture
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="VC1" mediaType="video">
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <spatialInformation>
                <captureOrigin>
                        <capturePoint>
                        <x>0.0</x>
                        <y>0.0</y>
                        <z>10.0</z>
                    </capturePoint>
                </captureOrigin>
                <captureArea>
                        <bottomLeft>
                                <x>-1.0</x>
                        <y>20.0</y>
```

```
                        <z>9.0</z>
                        </bottomLeft>
                        <bottomRight>
                                <x>1.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomRight>
                        <topLeft>
                                <x>-1.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topLeft>
                        <topRight>
                                <x>1.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topRight>
                </captureArea>
            </spatialInformation>
            <individual>true</individual>
            <encGroupIDREF>EG0</encGroupIDREF>
            <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>
        </mediaCapture>
        <mediaCapture
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="VC2" mediaType="video">
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <spatialInformation>
                <captureOrigin>
                        <capturePoint>
                        <x>2.0</x>
                        <y>0.0</y>
                        <z>10.0</z>
                    </capturePoint>
                </captureOrigin>
                <captureArea>
                        <bottomLeft>
                                <x>1.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
```

```
                        </bottomLeft>
                        <bottomRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomRight>
                        <topLeft>
                                <x>1.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topLeft>
                        <topRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topRight>
                </captureArea>
            </spatialInformation>
            <individual>true</individual>
            <encGroupIDREF>EG0</encGroupIDREF>
            <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>
        </mediaCapture>
        <mediaCapture
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="VC3" mediaType="video">
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <spatialInformation>
                <captureArea>
                        <bottomLeft>
                                <x>-3.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomLeft>
                        <bottomRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomRight>
                        <topLeft>
                                <x>-3.0</x>
```

```
                        <y>20.0</y>
                        <z>11.0</z>
                        </topLeft>
                        <topRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topRight>
                    </captureArea>
                </spatialInformation>
                <content>
                    <sceneViewIDREF>SE1</sceneViewIDREF>
                </content>
                <policy>Soundlevel:0</policy>
                <encGroupIDREF>EG0</encGroupIDREF>
                <description lang="en">loudest room segment</description>
                <priority>2</priority>
                <lang>it</lang>
                <mobility>static</mobility>
                <view>individual</view>
            </mediaCapture>
            <mediaCapture
             xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
            xsi:type="videoCaptureType" captureID="VC4" mediaType="video">
                <captureSceneIDREF>CS1</captureSceneIDREF>
                <spatialInformation>
                    <captureOrigin>
                        <capturePoint>
                        <x>0.0</x>
                        <y>0.0</y>
                        <z>10.0</z>
                       </capturePoint>
                    </captureOrigin>
                    <captureArea>
                        <bottomLeft>
                                <x>-3.0</x>
                        <y>20.0</y>
                        <z>7.0</z>
                        </bottomLeft>
                        <bottomRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>7.0</z>
                        </bottomRight>
                        <topLeft>
                                <x>-3.0</x>
                        <y>20.0</y>
                        <z>13.0</z>
```

```
                        </topLeft>
                        <topRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>13.0</z>
                        </topRight>
                </captureArea>
            </spatialInformation>
            <individual>true</individual>
            <encGroupIDREF>EG0</encGroupIDREF>
            <description lang="en">zoomed out view of all people in the
            room</description>
            <priority>2</priority>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>room</view>
            <capturedPeople>
                <personIDREF>alice</personIDREF>
                <personIDREF>bob</personIDREF>
                <personIDREF>ciccio</personIDREF>
            </capturedPeople>
        </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">
            <sceneViews>
                <sceneView sceneViewID="SE1">
                    <mediaCaptureIDs>
                        <captureIDREF>VC0</captureIDREF>
                        <captureIDREF>VC1</captureIDREF>
                        <captureIDREF>VC2</captureIDREF>
```

```
                    </mediaCaptureIDs>
                </sceneView>
                <sceneView sceneViewID="SE2">
                    <mediaCaptureIDs>
                        <captureIDREF>VC3</captureIDREF>
                    </mediaCaptureIDs>
                </sceneView>
                <sceneView sceneViewID="SE3">
                    <mediaCaptureIDs>
                        <captureIDREF>VC4</captureIDREF>
                    </mediaCaptureIDs>
                </sceneView>
                <sceneView sceneViewID="SE4">
                    <mediaCaptureIDs>
                        <captureIDREF>AC0</captureIDREF>
                    </mediaCaptureIDs>
                </sceneView>
            </sceneViews>
        </captureScene>
    </captureScenes>
    <simultaneousSets>
        <simultaneousSet setID="SS1">
            <mediaCaptureIDREF>VC3</mediaCaptureIDREF>
            <sceneViewIDREF>SE1</sceneViewIDREF>
        </simultaneousSet>
        <simultaneousSet setID="SS2">
            <mediaCaptureIDREF>VC0</mediaCaptureIDREF>
            <mediaCaptureIDREF>VC2</mediaCaptureIDREF>
            <mediaCaptureIDREF>VC4</mediaCaptureIDREF>
        </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>
```

18. MCC example

Enhancing the scenario presented in the previous example, the Media Provider is able to advertise a composed capture VC7 made by 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="audioCaptureType" captureID="AC0" mediaType="audio">
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <spatialInformation>
                <captureOrigin>
                        <capturePoint>
                        <x>0.0</x>
                        <y>0.0</y>
                        <z>10.0</z>
                    </capturePoint>
                    <lineOfCapturePoint>
                        <x>0.0</x>
                        <y>1.0</y>
                        <z>10.0</z>
```

```
                    </lineOfCapturePoint>
                </captureOrigin>
            </spatialInformation>
            <individual>true</individual>
            <encGroupIDREF>EG1</encGroupIDREF>
            <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>
        </mediaCapture>
        <mediaCapture
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="VC0" mediaType="video">
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <spatialInformation>
                <captureOrigin>
                        <capturePoint>
                        <x>0.5</x>
                        <y>1.0</y>
                        <z>0.5</z>
                    </capturePoint>
                    <lineOfCapturePoint>
                        <x>0.5</x>
                        <y>0.0</y>
                        <z>0.5</z>
                    </lineOfCapturePoint>
                </captureOrigin>
            </spatialInformation>
            <individual>true</individual>
            <encGroupIDREF>EG0</encGroupIDREF>
            <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>
        </mediaCapture>
        <mediaCapture
```

```
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:type="videoCaptureType" captureID="VC1" mediaType="video">
    <captureSceneIDREF>CS1</captureSceneIDREF>
    <spatialInformation>
        <captureOrigin>
                <capturePoint>
                <x>0.0</x>
                <y>0.0</y>
                <z>10.0</z>
            </capturePoint>
        </captureOrigin>
        <captureArea>
                <bottomLeft>
                        <x>-1.0</x>
                <y>20.0</y>
                <z>9.0</z>
                </bottomLeft>
                <bottomRight>
                        <x>1.0</x>
                <y>20.0</y>
                <z>9.0</z>
                </bottomRight>
                <topLeft>
                        <x>-1.0</x>
                <y>20.0</y>
                <z>11.0</z>
                </topLeft>
                <topRight>
                        <x>1.0</x>
                <y>20.0</y>
                <z>11.0</z>
                </topRight>
        </captureArea>
    </spatialInformation>
    <individual>true</individual>
    <encGroupIDREF>EG0</encGroupIDREF>
    <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>
</mediaCapture>
<mediaCapture
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
```

```
        xsi:type="videoCaptureType" captureID="VC2" mediaType="video">
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <spatialInformation>
                <captureOrigin>
                        <capturePoint>
                        <x>2.0</x>
                        <y>0.0</y>
                        <z>10.0</z>
                    </capturePoint>
                </captureOrigin>
                <captureArea>
                        <bottomLeft>
                                <x>1.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomLeft>
                        <bottomRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomRight>
                        <topLeft>
                                <x>1.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topLeft>
                        <topRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topRight>
                </captureArea>
            </spatialInformation>
            <individual>true</individual>
            <encGroupIDREF>EG0</encGroupIDREF>
            <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>
        </mediaCapture>
        <mediaCapture
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
        xsi:type="videoCaptureType" captureID="VC3" mediaType="video">
```

```
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <spatialInformation>
                <captureArea>
                        <bottomLeft>
                                <x>-3.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomLeft>
                        <bottomRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomRight>
                        <topLeft>
                                <x>-3.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topLeft>
                        <topRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topRight>
                </captureArea>
            </spatialInformation>
            <content>
                <sceneViewIDREF>SE1</sceneViewIDREF>
            </content>
            <policy>Soundlevel:0</policy>
            <encGroupIDREF>EG0</encGroupIDREF>
            <description lang="en">loudest room segment</description>
            <priority>2</priority>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>individual</view>
        </mediaCapture>
        <mediaCapture
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
         xsi:type="videoCaptureType" captureID="VC4" mediaType="video">
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <encGroupIDREF>EG0</encGroupIDREF>
            <spatialInformation>
                <captureOrigin>
                        <capturePoint>
                        <x>0.0</x>
                        <y>0.0</y>
                        <z>10.0</z>
                    </capturePoint>
```

```
                </captureOrigin>
                <captureArea>
                        <bottomLeft>
                                <x>-3.0</x>
                        <y>20.0</y>
                        <z>7.0</z>
                        </bottomLeft>
                        <bottomRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>7.0</z>
                        </bottomRight>
                        <topLeft>
                                <x>-3.0</x>
                        <y>20.0</y>
                        <z>13.0</z>
                        </topLeft>
                        <topRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>13.0</z>
                        </topRight>
                </captureArea>
            </spatialInformation>
            <individual>true</individual>
            <encGroupIDREF>EG0</encGroupIDREF>
            <description lang="en">
              zoomed out view of all people in the room
            </description>
            <priority>2</priority>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>room</view>
            <capturedPeople>
                <personIDREF>alice</personIDREF>
                <personIDREF>bob</personIDREF>
                <personIDREF>ciccio</personIDREF>
            </capturedPeople>
        </mediaCapture>
        <mediaCapture
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
         xsi:type="videoCaptureType" captureID="VC5" mediaType="video">
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <spatialInformation>
                <captureArea>
                        <bottomLeft>
                                <x>-3.0</x>
                        <y>20.0</y>
```

```
                        <z>9.0</z>
                        </bottomLeft>
                        <bottomRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomRight>
                        <topLeft>
                                <x>-3.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topLeft>
                        <topRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topRight>
                </captureArea>
            </spatialInformation>
            <content>
                <sceneViewIDREF>SE1</sceneViewIDREF>
            </content>
            <policy>Soundlevel:1</policy>
            <description lang="en">penultimate loudest room segment
            </description>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>individual</view>
        </mediaCapture>
        <mediaCapture
         xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
         xsi:type="videoCaptureType" captureID="VC6" mediaType="video">
            <captureSceneIDREF>CS1</captureSceneIDREF>
            <spatialInformation>
                <captureArea>
                        <bottomLeft>
                                <x>-3.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomLeft>
                        <bottomRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomRight>
                        <topLeft>
                                <x>-3.0</x>
                        <y>20.0</y>
```

```
                        <z>11.0</z>
                        </topLeft>
                        <topRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topRight>
                    </captureArea>
                </spatialInformation>
                <content>
                    <sceneViewIDREF>SE1</sceneViewIDREF>
                </content>
                <policy>Soundlevel:2</policy>
                <description lang="en">last but two loudest room segment
                </description>
                <lang>it</lang>
                <mobility>static</mobility>
                <view>individual</view>
            </mediaCapture>
            <mediaCapture
             xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
             xsi:type="videoCaptureType" captureID="VC7" mediaType="video">
                <captureSceneIDREF>CS1</captureSceneIDREF>
                <spatialInformation>
                    <captureArea>
                        <bottomLeft>
                                <x>-3.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomLeft>
                        <bottomRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>9.0</z>
                        </bottomRight>
                        <topLeft>
                                <x>-3.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topLeft>
                        <topRight>
                                <x>3.0</x>
                        <y>20.0</y>
                        <z>11.0</z>
                        </topRight>
                    </captureArea>
                </spatialInformation>
                <content>
```

```
                <captureIDREF>VC3</captureIDREF>
                <captureIDREF>VC5</captureIDREF>
                <captureIDREF>VC6</captureIDREF>
            </content>
            <maxCaptures exactNumber="true">3</maxCaptures>
            <encGroupIDREF>EG0</encGroupIDREF>
            <description lang="en">big picture of the current speaker +
            pips about previous speakers</description>
            <priority>3</priority>
            <lang>it</lang>
            <mobility>static</mobility>
            <view>individual</view>
        </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">
            <sceneViews>
                <sceneView sceneViewID="SE1">
                    <description lang="en">participants' individual
                    videos</description>
                    <mediaCaptureIDs>
                        <captureIDREF>VC0</captureIDREF>
                        <captureIDREF>VC1</captureIDREF>
                        <captureIDREF>VC2</captureIDREF>
                    </mediaCaptureIDs>
                </sceneView>
                <sceneView sceneViewID="SE2">
                    <description lang="en">loudest segment of the
                    room</description>
                    <mediaCaptureIDs>
                        <captureIDREF>VC3</captureIDREF>
```

```
                    </mediaCaptureIDs>
                </sceneView>
                <sceneView sceneViewID="SE5">
                    <description lang="en">loudest segment of the
                    room + pips</description>
                    <mediaCaptureIDs>
                        <captureIDREF>VC7</captureIDREF>
                    </mediaCaptureIDs>
                </sceneView>
                <sceneView sceneViewID="SE4">
                    <description lang="en">room audio</description>
                    <mediaCaptureIDs>
                        <captureIDREF>AC0</captureIDREF>
                    </mediaCaptureIDs>
                </sceneView>
                <sceneView sceneViewID="SE3">
                    <description lang="en">room video</description>
                    <mediaCaptureIDs>
                        <captureIDREF>VC4</captureIDREF>
                    </mediaCaptureIDs>
                </sceneView>
            </sceneViews>
        </captureScene>
    </captureScenes>
    <simultaneousSets>
        <simultaneousSet setID="SS1">
            <mediaCaptureIDREF>VC3</mediaCaptureIDREF>
            <mediaCaptureIDREF>VC7</mediaCaptureIDREF>
            <sceneViewIDREF>SE1</sceneViewIDREF>
        </simultaneousSet>
        <simultaneousSet setID="SS2">
            <mediaCaptureIDREF>VC0</mediaCaptureIDREF>
            <mediaCaptureIDREF>VC2</mediaCaptureIDREF>
            <mediaCaptureIDREF>VC4</mediaCaptureIDREF>
        </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>
```

19. Diff with draft-ietf-clue-data-model-schema-10 version

Applied modifications as per Mark Duckworth's review (example corrections and maxCapturesType modification)

maxCapturesType has been changed from positiveInteger to unsignedShort excluding value 0.

20. Diff with draft-ietf-clue-data-model-schema-10 version

Minor modifications have been applied to address nits at page https://www.ietf.org/tools/idnits?url=https://www.ietf.org/archive/id/draft-ietf-clue-data-model-schema-10.txt.

21. Diff with draft-ietf-clue-data-model-schema-09 version

- We have introduced a <captureOrigin> element containing a mandatory <capturePoint> and an optional <lineOfCapturePoint> in the definition of <spatialInformation> as per Paul's review

- A new type definition for switching policies (resembled by <policy> element) has been provided in order to have acceptable values in the form of "token:index".

- Minor modifications suggested in WGLC reviews have been applied.

22. Diff with draft-ietf-clue-data-model-schema-08 version

- o Typos correction

23. Diff with draft-ietf-clue-data-model-schema-07 version

- o IANA Considerations: text added
- o maxCaptureEncodings removed
- o personTypeType values aligned with CLUE framework
- o allowSubsetChoice added for multiple content captures
- o embeddedText moved from videoCaptureType definition to mediaCaptureType definition
- o typos removed from section Terminology

24. Diff with draft-ietf-clue-data-model-schema-06 version

- o Capture Scene Entry/Entries renamed as Capture Scene View/Views in the text, <sceneEntry>/<sceneEntries> renamed as <sceneView>/ <sceneViews> in the XML schema.
- o Global Scene Entry/Entries renamed as Global View/Views in the text, <globalSceneEntry>/<globalSceneEntries> renamed as <globalView>/<globalViews>
- o Security section added.
- o Extensibility: a new type is introduced to describe other types of media capture (otherCaptureType), text and example added.
- o Spatial information section updated: capture point optional, text now is coherent with the framework one.
- o Audio capture description: <sensitivityPattern> added, <audioChannelFormat> removed, <captureArea> disallowed.
- o Simultaneous set definition: added <captureSceneIDREF> to refer to capture scene identifiers as shortcuts and an optional mediaType attribute which is mandatory to use when only capture scene identifiers are listed.
- o Encoding groups: removed the constraint of the same media type.

- o Updated text about media captures without <encodingGroupIDREF> (optional in the XML schema).

- o "mediaType" attribute removed from homogeneous groups of capture (scene views and globlal views)

- o "mediaType" attribute removed from the global view textual description.

- o "millimeters" scale value changed in "mm"

25. 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>.

26. 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.

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

captureParameters 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.

28. Acknowledgments

The authors thank all the CLUErs for their precious feedbacks and support.

29. Informative References

[I-D.ietf-clue-datachannel] Holmberg, C., "CLUE Protocol data channel", draft-ietf-clue-datachannel-11 (work in progress), November 2015.

[I-D.ietf-clue-framework] Duckworth, M., Pepperell, A., and S. Wenger, "Framework for Telepresence Multi-Streams", draft-ietf-clue-framework-25 (work in progress), January 2016.

[I-D.ietf-clue-protocol] Presta, R. and S. Romano, "CLUE protocol", draft-ietf-clue-protocol-06 (work in progress), October 2015.

[I-D.ietf-ecrit-additional-data] Gellens, R., Rosen, B., Tschofenig, H., Marshall, R., and J. Winterbottom, "Additional Data Related to an Emergency Call",

draft-ietf-ecrit-additional-data-37 (work in progress), October 2015.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, DOI 10.17487/RFC3550, July 2003, <http://www.rfc-editor.org/info/rfc3550>.

[RFC3688] Mealling, M., "The IETF XML Registry", BCP 81, RFC 3688, DOI 10.17487/RFC3688, January 2004, <http://www.rfc-editor.org/info/rfc3688>.

[RFC4353] Rosenberg, J., "A Framework for Conferencing with the Session Initiation Protocol (SIP)", RFC 4353, DOI 10.17487/RFC4353, February 2006, <http://www.rfc-editor.org/info/rfc4353>.

[RFC6351] Perreault, S., "xCard: vCard XML Representation", RFC 6351, DOI 10.17487/RFC6351, August 2011, <http://www.rfc-editor.org/info/rfc6351>.

[RFC7303] Thompson, H. and C. Lilley, "XML Media Types", RFC 7303, DOI 10.17487/RFC7303, July 2014, <http://www.rfc-editor.org/info/rfc7303>.

[RFC7667] Westerlund, M. and S. Wenger, "RTP Topologies", RFC 7667, DOI 10.17487/RFC7667, November 2015, <http://www.rfc-editor.org/info/rfc7667>.

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