Digital Imaging and Communications in Medicine (DICOM)

Supplement 202: ~~Real~~ Real-Time Video

Prepared by:

DICOM Standards Committee, Working Group 13

1300 N. 17th Street

Rosslyn, Virginia 22209 USA

VERSION: Draft First Read, April 24, 2017

Developed in accordance with: DICOM Workitem 2016-12-D

Table of Contents

TO BE COMPLETED…

**** *Editorial content – to be removed before Final Text* ****

TODO:

Editor's Notes

External sources of information

Editorial Issues and Decisions

#	Issue	Status

Closed Issues

#	Issues

Open Issues

#	Issues	Status
1	Name of the supplement ("~~Real~~ Real-Time Video" proposed)?	Open
2	Do we specify use case(s) and which level of detail?	Open
3	Do we embrace also multi-frame medical imaging (e.g.; live US, live RF) or only (visible light) video?	Open
4	How shall we deal with proper understanding and proper referencing of SMPTE/VSF documents	Open
5	How we proceed with the medical metadata, either using a VSF/SMPTE defined mechanism or a pure RTP one, respecting the classical DICOM encoding?	Open
6	Provide a table where we list of kind of information to convey in the metadata along with the video. Look at part 18 (how to define recoding e.g. media type/DICOM) and enhanced CT/MR objects (list of information which are constant vs. variable).	Open
7	Selection of metadata to be conveyed and why (justified based on the use cases). Be very selective. Which frequency for sending the metadata (every frame?).	Open
8	Is there a mechanism to register (in SMPTE or others) for a domain specific options?	Open

**** *End of Editorial content – to be removed before Final Text* ****

Scope and Field of Application

This Supplement describes a new standard for the transport of real-time video and associated medical data, titled Real-Time Video Transport Protocol.

DICOM has developed several standards for the storage of medical video in endoscopy, microscopy or echography, typically. But medical theaters such as the operating room (OR) are for the moment still using proprietary solutions to handle communication of real-time video and associated information like patient demographics, study description or 3D localization of imaging sources.

The new Real-Time Video standard will enable to deploy interoperable devices inside the OR and beyond, enabling a better management of imaging information, impacting directly the quality of care.

Professional video (e.g. TV studios) equipment providers and users have defined a new standardized approach for conveying video and associated information (audio, ancillary data, metadata…) enabling the deployment of equipment in a distributed way (vs. peer-to-peer).

The supplement defines an IP-based new DICOM [SOP Class++, Services TO BE COMPLETED…] for the transport of real-time video including quality compatible with the communication inside the operating room (OR). SMPTE ST 2110 suite, elaborated on the basis of Technical Recommendation TR03 originated by the VSF (Video Services Forum) is used as a platform. The specific level of requirements (size and complexity of metadata, quality of image, ultra low latency, variety of image resolution, restriction to pixel ratio of 1… TO BE CHECKED AND COMPLETED) introduce some necessary restrictions of the SMPTE ST 2110 suite recommendations. In addition to these recommendations, DICOM is defining a mechanism enabling to convey specific medical metadata along with the video while respecting the architecture defined in TR03.

This proposed Supplement includes a number of Addenda to existing Parts of DICOM:

- PS 3.1 Introduction and Overview
(will add introduction of the new protocol)

- PS 3.2 Conformance
(will add conformance for real-time communication)

- PS 3.3 Information Object Definitions
(may add new Information Object Definitions if existing IODs are not sufficient)

- PS 3.4 Service Class Specifications
(will add new Service Class Specifications for real-time communication)

- PS 3.5 Data Structures and Encoding
(may add new Data Structure and Semantics for data related to real-time communication)

- PS 3.6 Data Dictionary
(may add new Data definition related to real-time communication and video description)

- PS 3.7 Message Exchange
(will add new Message Exchange definition for real-time communication)

- PS 3.8 Network Communication Support for Message Exchange
(will add new Network Communication Support For Message Exchange (e.g. synch.))

- PS 3.17: Explanatory Information
(may add new explanatory information (e.g. video transports standards))

- PS 3.18 Web Services

(may add new Web Services for supporting the real-time communication (e.g. config.))

Potentially a new Part may be created for specifying the real-time communication Services.

PS3.17: Add a new Annex Real-Time Video Use Cases as indicated.

XX Real-Time Video Use Cases (Informative)

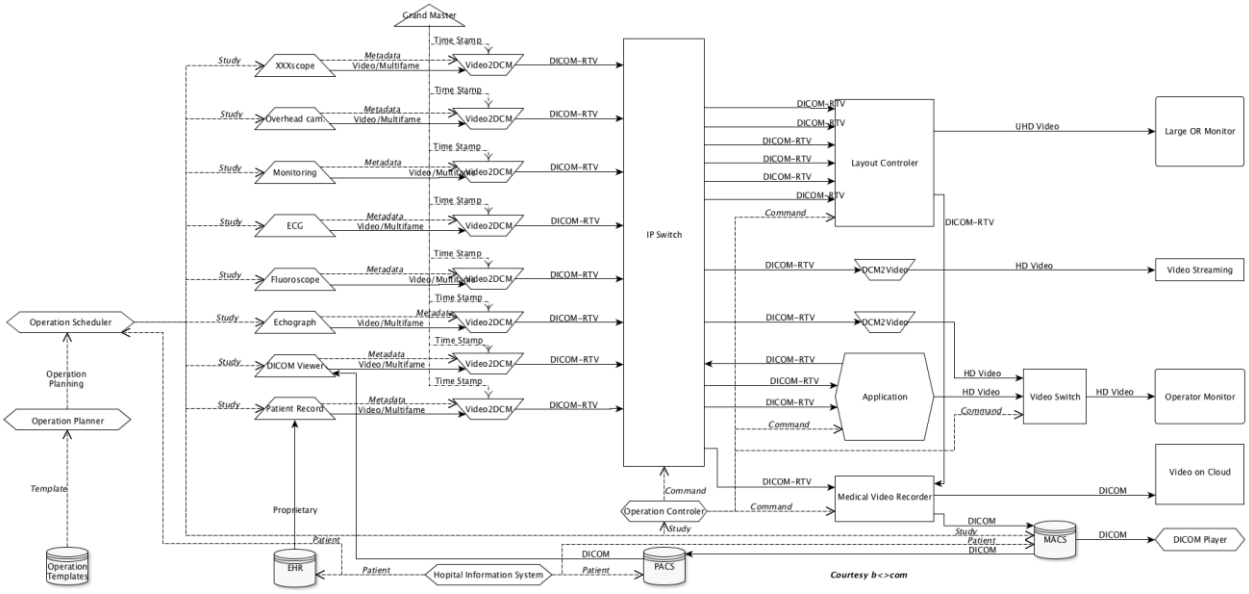

Figure XX-1: Overview diagram of operating room

As shown on Figure XX-1, the DICOM Real-Time Video (DICOM-RTV) communication is used to connect various video or multiframe sources to various destinations, through a standard IP switch.

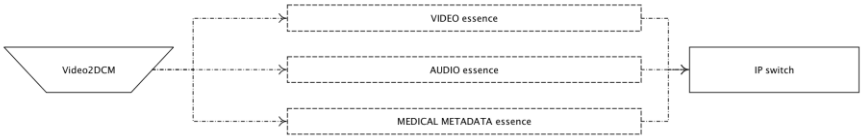

Figure XX-2: Real-Time Video flow content overview

As shown on figure Figure XX-2, the DICOM Real-Time Video flow is comprised of typically three different sub-flows (“essences”) for respectively video, audio and medical-metadata information. Using the intrinsic capability of IP to convey different flow on the same support, the information conveyed on the Ethernet cable will include three kinds of blocks for video (thousands for each video frame), audio (hundreds for each frame) and medical-metadata (units for each frame), respectively represented as “V” (video), “A” (audio) and “M” (metadata) on the Figure XX-3. The information related to one frame will ~~be~~ comprise alternate blocks of the three types, the video related ones being largely more frequent.

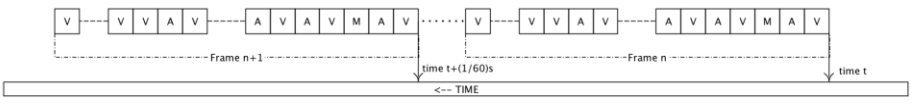

Figure XX-3: Real-Time Video flow details

XX.1 Generic Use Case 1: Duplicating video on additional monitors

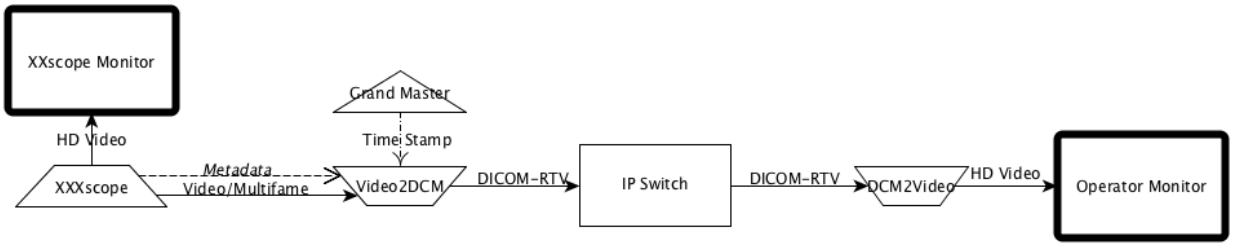

Figure XX-4: Duplicating on additional monitor

In the context of image guided surgery, two operators are directly contributing to the procedure:

- a surgeon performing the operation itself, using relevant instruments;
- an assistant controlling the imaging system (e.g. coelioscope).

In some situations, both operators cannot stand on the same side of the patient. Because the control image has to be in front of each operator, two monitors are required, a primary one, directly connected to the imaging system, and the second one being on the other side.

Additional operators (e.g. surgery nurse) also have to see what is happening in order to anticipate actions (e.g. providing instrument).

The live video image has to be transferred on additional monitors with a minimal latency, without modifying the image itself (resolution…). The latency between the two monitors (see Figure XX-4) should be compliant with collaborative activity on surgery where the surgeon is operating based on the second monitor and the assistant is controlling the endoscope based on the primary monitor. This supplement addresses only the communication aspects, not the presentation. Some XXscopes are now producing UHD video, with the perspective to support also HDR (High Dynamic Range) for larger color gamut management (upo to 10 bits per channel) as well as HFR (High Frame Rate), i.e.; up to 120 Hz.

XX.2 Generic Use Case 2: Post Review by Senior

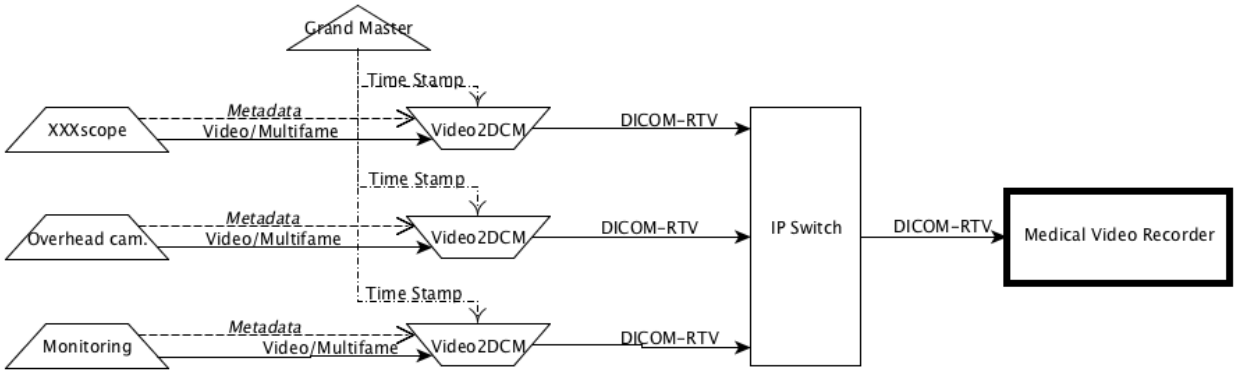

Figure XX-5: Recording multiple video sources

A junior surgeon performs a procedure which apparently goes well. The next day, the patient state is not ok, requiring the surgeon to refer the patient to a senior surgeon.

In order to decide what to do, the senior surgeon:

- has to review and understand what happened;

- takes the decision to re-operate the patient or not;
- if a new operation is performed, needs to have access to the sequence of the first operation which is suspected.

Moreover, the junior surgeon has to review her/his own work in order to prevent against a new mistake.

A good quality recording of video needs to be kept, at least for ~~some time~~a certain duration, including all the video information (endoscopy, overhead, monitoring, …) and associated metadata (see Figure XX-5). Storing the video has to be doable in real-time. The recording has to maintain time consistency between the different video channels. The format of recording is out of the scope of the supplement, as well as the way of replaying the recorded videos. Only the method for feeding the recorder with the synchronized videos and associated metadata is specified by the present supplement.

XX.3 Generic Use Case 3: Automatic display in Operating Room (OR)

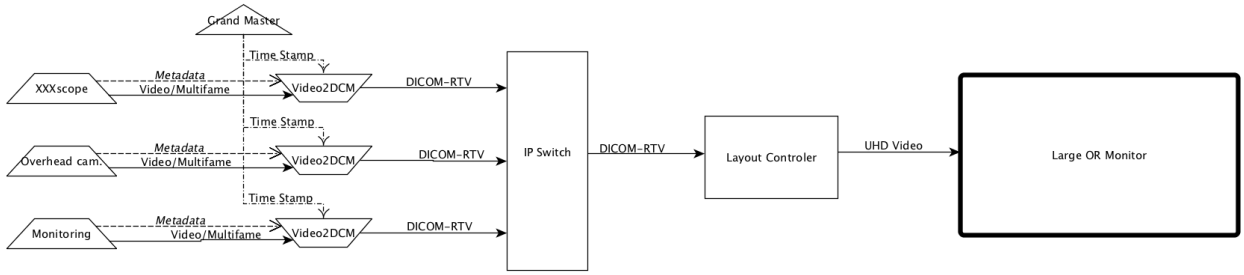

Figure XX-6: Displaying multiple source on one unique monitor

OR are more and more equipped with large monitors displaying all the necessary information. Depending on the stage of the procedure, the information to display is changing. In order to improve the quality of the real-time information shared inside the OR, it is relevant to automate the set-up of such a display, based on the metadata conveyed along with the video (e.g. displaying the XXXscope image only when relevant).

All the video streams have to be transferred with the relevant information (patient, study, equipment…), as shown on the Figure XX-6. The mechanisms relative to the selection and execution of layout of images on the large monitor are out of the scope of this supplement. Only the method for conveying the multiple synchronized video along with the metadata, used as parameters for controlling the layout, ~~are~~ is specified in the present supplement.

XX.4 Generic Use Case 4: Augmented Reality

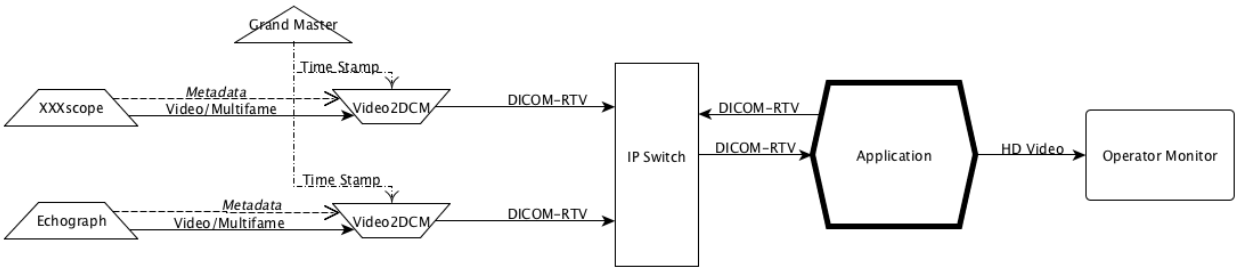

Figure XX-7: Application combining multiple real-time video sources

Image guided surgery is gradually becoming mainstream, mainly because minimally invasive. In order to guide the surgeon gesture, several procedures are based on the 3D display of patient anatomy reconstructed from MR or CT scans. But real-time medical imaging (3D ultrasound typically) can also be used as reference. Display devices (glasses, tablets…) will be used to show real-time “composite” image

merged from the main video imaging (endoscopy, overhead, microscopy…) and the multi-frame medical imaging. The real-time "composite" image could be also exported as a new video source, through the DICOM Real-Time Video protocol.

All video streams have to be transferred with ultra-low latency and very strict synchronization between frames (see Figure XX-7). Metadata associated with the video has to be updated at the frame rate (e.g. 3D position of the US probe). The mechanisms used for combining multiple video sources or to detect and follow 3D position of devices are out of scope of this supplement. Only the method for conveying the multiple synchronized video/multiframe sources along with the parameters, that may change at evering frame, is specified in the present supplement.

XX.5 Generic Use Case 5: Robotic aided surgery

Robotic assisted surgery is emerging. Image guided robots or cobots are gradually used for diffent kinds of procedures. In the near future, different devices will have to share the information provided by the robot synchronized with the video produced by imaging sources. I order to be able to process properly the information provided by the robot, it should be possible to convey such information at a frequency bigger that the video frequency, i.e.; 400 Hz vs. 60 Hz for present HD.

PS3.17: Add a new Annex Transport of Elementary Stream over IP as indicated.

YY Transport of Elementary Stream over IP (Informative)

Carriage of audiovisual signals in their digital form across television plants has historically been achieved using coaxial cables that interconnect equipments through Serial Digital Interface (SDI) ports. The SDI technology provides a reliable transport method to carry a multiplex of video, audio and metadata with strict timing relationships.

The features and throughput of IP networking equipment having improved steadily, it has become practical to use IP switching and routing technology to convey and switch video, audio, and metadata essence within television facilities.

Existing standards such as SMPTE ST 2022-6:2012 have seen a significant adoption in this type of application where they have brought distinct advantages over SDI albeit only performing Circuit Emulation of SDI (ie. Perfect bit-accurate transport of the SDI signal contents).

However, the essence multiplex proposed by the SDI technology may be considered as somewhat inefficient in many situations where a significant part of the signal is left unused if little or no audio &/or ancillary data has to be carried along with the video raster, as depicted in figure YY-1 below:

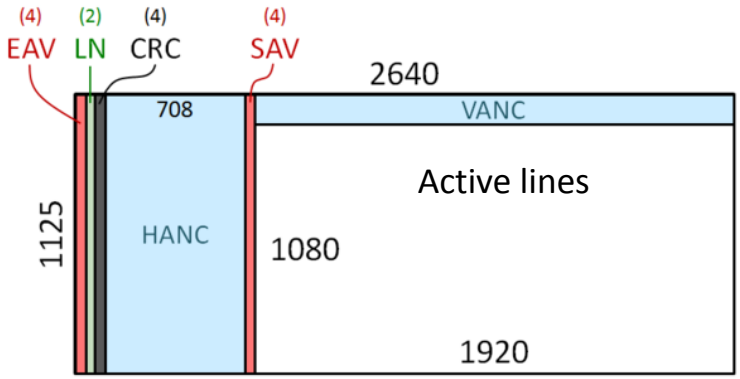

Figure YY-1
structure of a High Definition SDI signal

As new image formats such as UHD get introduced, the corresponding SDI bit-rates increase, way beyond 10Gb/s and the cost of equipments that need to be used at different points in a TV plant to embed, de-embed, process, condition, distribute, etc the SDI signals becomes a major concern.

Consequently there has been a desire in the industry to switch and process different essence elements separately, leveraging on the flexibility and cost-effectiveness of commodity networking gear and servers.

The Video Services Forum (VSF) has authored its Technical Recommendation #3 (a.k.a. VSF-TR03) describing the principles of a system where streams of different essences (namely video, audio, metadata to begin with) can be carried over an IP-based infrastructure whilst preserving their timing characteristics. VSF TR03 leverages heavily on existing technologies such as RTP, AES67, PTP, mostly defining how they can be used together to build the foundations of a working ecosystem .

The TR03 work prepared by VSF has been handed off to the Society of Motion Picture & Television Engineers (SMPTE) for due standardization process. The 32nf60 Drafting Group has broken down the TR03 work into different documents addressing distinct aspects of the system. This family of standards *(once approved, which is not yet the case at the time of this writing)* bears the ST 2110 prefix.

The initial documents identified in the family are:

- ST 2110-10: System Timing and definitions;
- ST 2110-20: Uncompressed active video;
- ST 2110-30: Uncompressed PCM audio;
- ST 2110-40: Ancillary data;
- ST 2110-50: ST 2022-6 as an essence.

The system is intended to be extensible to a variety of essence types, its pivotal point being the use of the RTP protocol. In this sytem, essence streams are encapsulated separately into RTP before being individually forwarded through the IP network.

A system is built from devices that have senders and/or receivers. Streams of RTP packets flow from senders to receivers. RTP streams can be either unicast or multicast, in which case multiple receivers can receive the stream over the network.

Devices may be adapters that convert from/to existing standard interfaces like HDMI or SDI, or they may be processors that receive one or more streams from the IP network, transform them in some way and transmit the resulting stream(s) to the IP network. Cameras and monitors may transmit and receive elementary RTP streams directly through an IP-connected interface, eliminating the need for legacy video interfaces.

Proper operation of the ST 2110 environment relies on a solid timing infrastructure that has been largely inspired by the one used in AES67 for Audio over IP.

Inter-stream synchronization relies on timestamps in the RTP packets that are sourced by the senders from a common Reference Clock. The Reference Clock is distributed over the IP network to all participating senders and receivers via PTP (Precision Time Protocol version 2, IEEE 1588-2008).

Synchronization at the receiving device is achieved by the comparison of RTP timestamps with the common Reference Clock. The timing relationship between different streams is determined by their relationship to the Reference Clock.

Each device maintains a Media Clock which is is frequency locked to its internal timebase and advances at an exact rate specified for the specific media type. The media clock is used by senders to sample media and by receivers when recovering digital media streams. For video and ancillary data, the rate of the media clock is 90 kHz, whereas for audio it can be 44.1 kHz, 48 kHz, or 96KHz.

For each specific media type RTP stream, the RTP Clock operates at the same rate as the Media Clock.

ST 2110-20 proposes a very generic mechanism for RTP encapsulation of a video raster. It supports arbitrary resolutions, frame rates, and proposes a clever pixel packing accommodating an extremely wide variety of bit depths and sampling modes. It is very heavily inspired from IETF RFC4175.

ST 2110-30 provides a method to encapsulate PCM digital audio using AES67 to which it applies a number of constraints.

ST 2110-40 provides a simple method to tunnel packets of SDI ancillary data present in a signal over the IP network and enables a receiver to reconstruct an SDI signal that will embed the ancillary data at the exact same places it occupied in the original stream.

Devices that contain one or more sender have to construct one SDP (Session Description Protocol) object per RTP Stream.These SDP objects are made available through the management interface of the device, thereby publishing the characteristics of the stream they encapsulate. This provides the basic information a system needs to gather in order to identify the available signal sources on the network.

It is worth noting that although ST 2110 currently describes the method for transporting video and audio as uncompressed essence, the same principles may be applied to other types of media by selecting the appropriate RTP payload encapsulation scheme, and complying to the general principles defined by ST 2110-10.