# New Work Item Proposal on Data Migration

Submitted by WG-33

**Introduction.**

A use case of steadily increasing significance is porting large DICOM repositories from one image management system (PACS or VNA) to another. Users typically replace their PACS after ~12-15 years, often with change of vendor. Replacement requires migrating historical data to the new system. Thus, every year, 5-10% of user organizations may be doing a PACS data migration.

Old images are now routinely retained “forever”, and data set sizes are increasing with 3D/4D and multimodality studies. Archives in many institutions now store over a billion instances with data volumes over one petabyte. Migration approaches need to operate at large scales, and handle both on-premises and remote (e.g., cloud-based) storage.

Migration often occurs when either the source system or the destination, or both, are in clinical operation, but systems designed and configured to handle the throughput of regular operations may not have capacity for the additional massive input/output requirements of migration. With a data transfer rate of 1 terabyte / day (quite high for even the most advanced PACS), the time to transfer a petabyte archive is 3 years. Performance constraints exist on both the source and destination systems.

Similar needs arise when healthcare institutions merge previously disparate PACS into an enterprise PACS. The archives from the old PACS need to be migrated. This is an increasing need with the accelerated pace of healthcare organization consolidations. Conversely, large sets of archived data may sometimes need to be migrated out of a PACS to support business divestment or realignment in a healthcare organization.

There are also research use cases, including artificial intelligence and machine learning, where bulk access to the archive is desirable, and such uses might leverage some of the same mechanisms developed for migration. PACS audit and quality control may also utilize some of the standardized functionality developed for migration, such as an archive inventory and metadata to identify the data produced by a particular unit or by a particular modality.

Recognizing the opportunity for DICOM to establish standards to facilitate the process of migration, DICOM WG-33 (Data Archive and Management) was chartered in December 2019, and has been meeting bi-weekly since February 2020. Each meeting has typically seen about 30 participants, reflecting the level of interest in this topic. This New Work Item Proposal represents the current consensus among the several classes of stakeholders.

**Limitations of Current Standard.**

The current DICOM Standard does not address the use case and technical interoperability requirements for migration of a full enterprise archive data set, and it is currently ill-suited for the major performance issues of migration.

The current Standard is designed for routine daily department operational workflows – acquisition, storage, analysis, and reading of imaging studies associated with individual patients. The Standard is optimized for identifying and transferring the objects for, at most, just a few studies or patients at a time. Network Query/Retrieve operations are synchronous, and the network connection must remain open until the operation is complete. The number of items in a response therefore is typically restricted; the Standard explicitly defines A70x – Out of Resources error statuses for these restrictions. Even media-based data exchange is specified only for the use cases of limited file sets, basically what can fit on a DVD.

A key requirement for migration (and other use cases) is the ability to have an inventory of all studies, series, and instances in an archive. While the current Query Service (DIMSE or equivalent DICOMweb) could be used, limitations on number of responses and the synchronous protocol require the use of a possibly very large number of partial Query requests. This makes producing such an inventory difficult.

While the current Standard Retrieve and Storage Services make moving data possible, they require significant overhead for the transfer of each object. Although there is current work to define the transfer of a bundle of objects under the DICOMweb services, which would reduce that overhead somewhat, that would still be insufficient for migration of an entire enterprise scale archive. A standardized method of direct filesystem access to stored object files is needed.

**Description of Proposal.**

The Work Item will specify an Information Object Definition capable of encoding an inventory of all studies, series, and instances in an archive. This is functionally equivalent to a Query response that returns an inventory of the entire PACS database (or a described subset thereof). Inventory of non-patient objects are out of scope for this Work Item.

The Work Item will define a mechanism to initiate the production of the inventory and allow it to proceed asynchronously. The inventory SOP Instance would be available when the operation is complete. The proposal will address any additional issues arising from inventory of an archive in active use. Access to the inventory SOP Instances would be similar to mechanisms specified for other DICOM Non-Patient Object classes (such as Color Palettes).

Since archive systems may optionally support direct filesystem access to DICOM Part 10 compliant files, for all or some of their stored instances, the inventory IOD will allow a link to such accessible files. To support a common PACS implementation design, wherein the archive may retain metadata updates (e.g., changed patient IDs) in its database and not propagate them to the stored instances, the inventory will provide the current metadata and flags indicating the stored instances have not been updated.

Interoperability via file exchange is specified in the DICOM Media Storage Service. The Work Item will specify the means by which an implementation, either the File Set Creator (the archive) or the File Set Reader (the migration destination), may claim conformance to this file exchange capability, and specifically to required functionality regarding accessing current metadata.

Although DICOM Interchange Media allow the inclusion in the file set of HL7 Clinical Document Architecture objects (e.g., radiology reports) as native files not encapsulated under a DICOM header, the scope of the Work Item is limited to migration of DICOM instances. While the Work Item may consider appropriate alignment with HL7 standards (e.g., FHIR), it is not expected to propose a solution for migrating non-DICOM archived data.

**Parts of Standard Affected.**

Part 1 – update overview of DICOM capabilities for shared file access

Part 2 - update the Network and Media Service conformance claims.

Part 3 - add the new inventory IOD.

Part 4 - update the Media Storage Service Class for requirements on file-based access, update the Non-Patient Object Storage Service Class for the inventory SOP Class, add a Q/R Service Class for the inventory objects, and add a service for initiating production of inventory instances.

Part 6 - add new SOP Class UIDs and new attributes.

Part 10 - update to address archive file exchange.

Part 11 - add a new Media Application Profile for archive file exchange.

Part 12 - add one or more new shared file system Media Formats.

Part 15 – address security profiles for file access

Part 17 - add an informative Annex.

Part 18 - update the DICOMweb Non-Patient Instance Service and Resources to include the inventory objects.

**Workload and Timeline.**

Drafting of the Supplement will be assigned to WG-33 (Data Archive and Management), which will coordinate as needed with WG-14 (Security), WG-20 (Integration of Imaging and Information Systems), WG-23 (Artificial Intelligence and Application Hosting), and WG-27 (Web Technology). Harry Solomon (Laitek) has volunteered to serve as editor. Production of the first draft is expected to take 4 months.

WG-06 may require up to 14 hours to review the drafts prior to release for Public Comment, 8 hours for comment resolution and Letter Ballot preparation, and 6 hours for Final Text. This will involve five or six WG-06 meetings, thus extending over 12-15 months. The WG-06 reviews will be facilitated by having an editor with long-time participation in WG-06 and extensive experience in writing and reviewing Supplements (including participating in developing the original media exchange standard – Supplements 1, 2, and 3).