



# Automatic processing of CERN video, audio and photo archives

Michał Kwiatek, CERN IT Department, email: Michal.Kwiatek@cern.ch

CHEP 2007

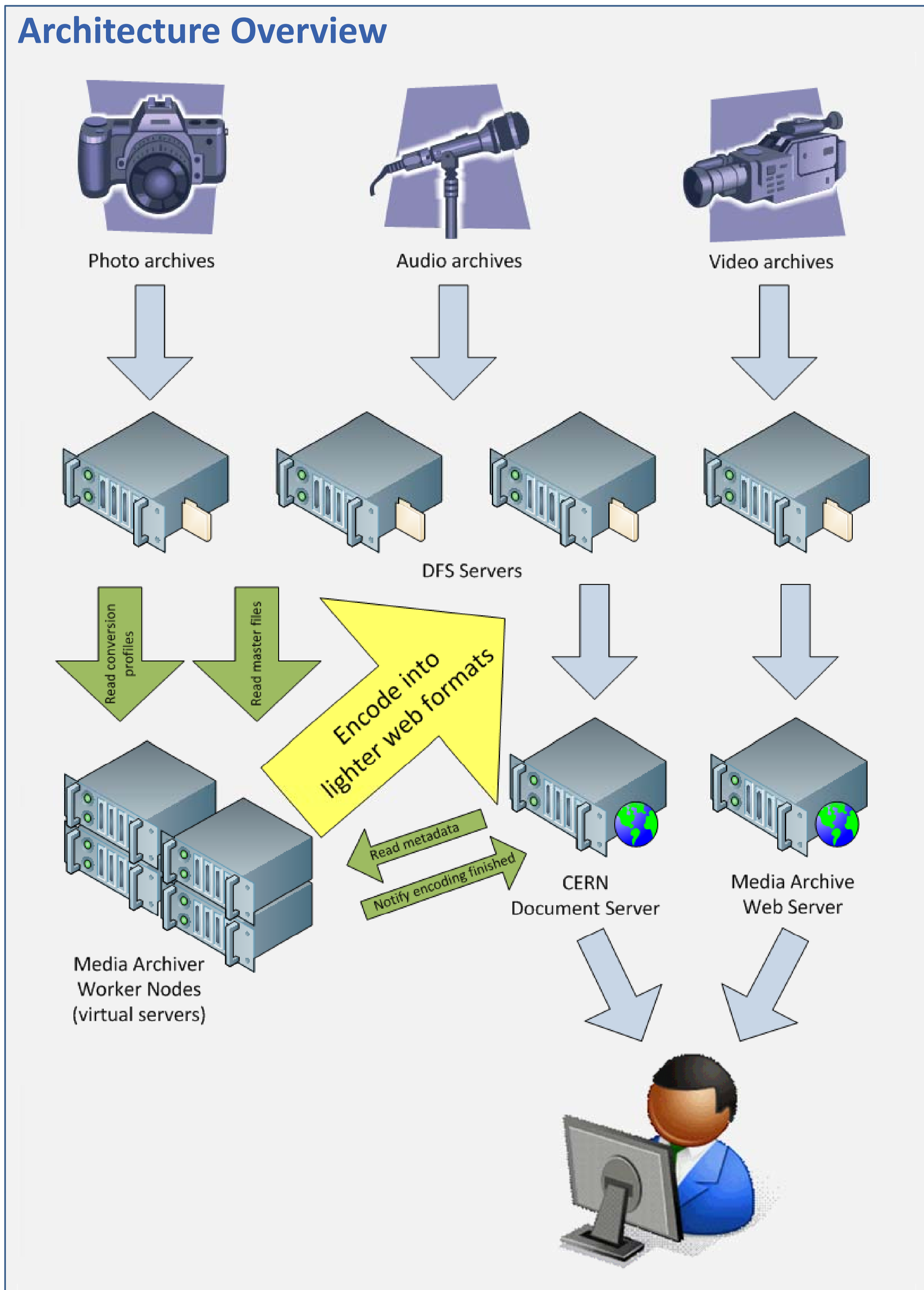
International Conference on Computing in High Energy and Nuclear Physics  
2-7 September 2007, Victoria BC Canada

## Abstract

The digitalization of CERN audio-visual archives, a major task currently in progress, will generate over 40 TB of video, audio and photo files. Storing these files is one issue, but a far more important challenge is to provide long-time coherence of the archive and to make these files available on-line with minimum manpower investment.

An infrastructure, based on standard CERN services, has been implemented, whereby master files, stored in the CERN Distributed File System (DFS), are discovered and scheduled for encoding into lightweight web formats based on predefined profiles. Changes in master files, conversion profiles or in the metadata database (read from CDS, the CERN Document Server) are automatically detected and the media re-encoded whenever necessary. The encoding processes are run on virtual servers provided on-demand by the CERN Server Self Service Centre, so that new servers can be easily configured to adapt to higher load. Finally, the generated files are made available from the CERN standard web servers with streaming implemented using Windows Media Services.

## Architecture Overview



## Archiving workflow

CDS, the CERN Document Server, is the central database of CERN audio, video and photo archives. Information about media files is entered into CDS by members of Photolab and Audio-visual teams. Media files are put on DFS, CERN's standard storage system. The Media Archiver software discovers new files immediately and initiates the off-line encoding process.

Media Archiver first queries CDS for metadata: format, title, description and date of recording. Encoding profiles are attributed based on the format. The metadata is then included in the generated files. When the files are ready, CDS is notified and links to the generated files are added to the CDS record.

The end-user discovers audio-visual records by browsing through CDS, available directly from CERN's public welcome page. The files are linked from CDS, but served from the Media Archive web server running Internet Information Server and Windows Media Services.



## Media Archiver software

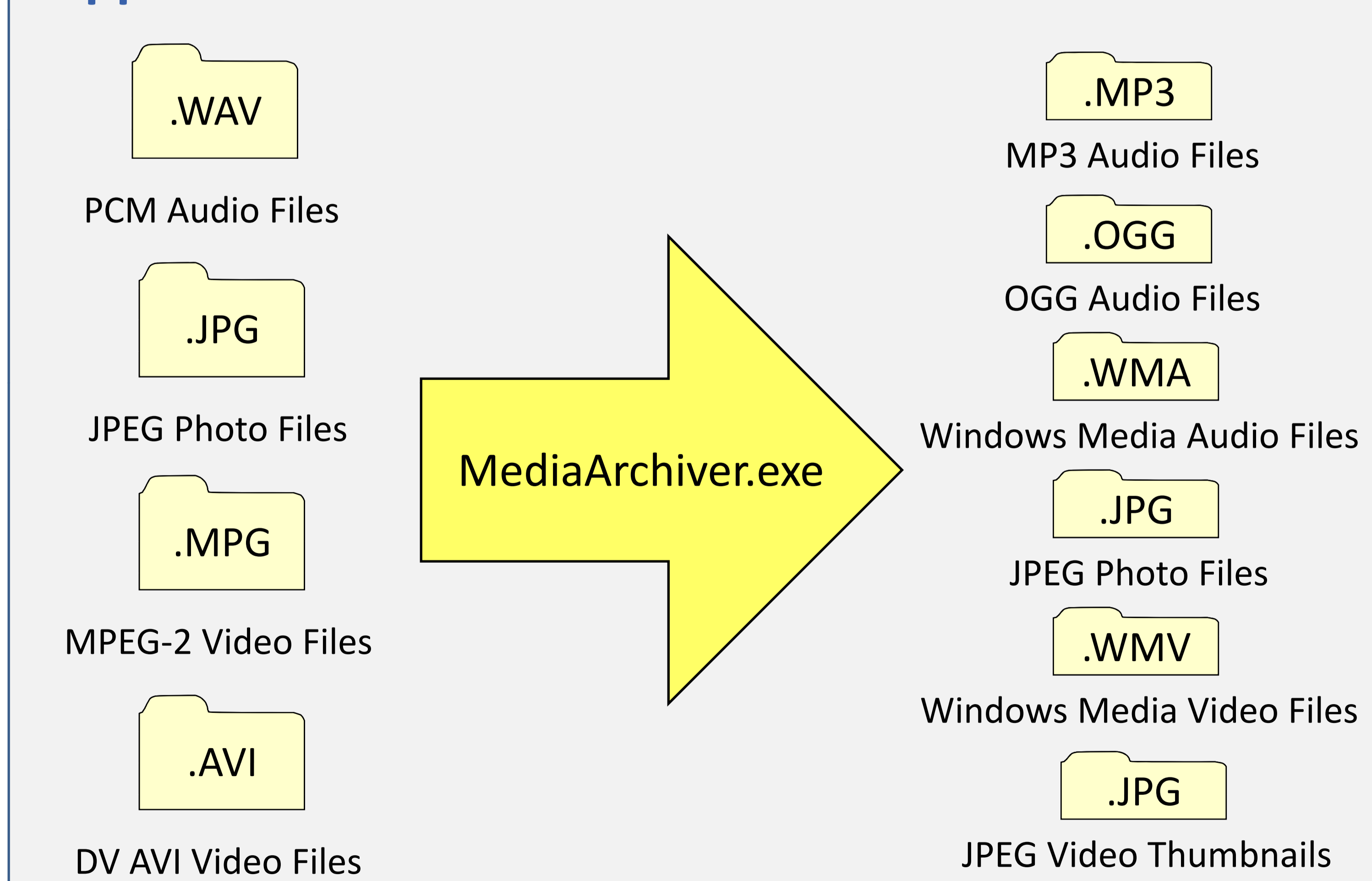
Encoding of multimedia files into light-weight formats is done using the Media Archiver software developed within CERN IT-IS group. Media Archiver automatically discovers new or modified files and starts off-line encoding. Changes to encoding profiles or metadata information are also automatically discovered. What is more, Media Archiver also enforces that files stored on DFS are described in CDS, which is important to make sure that the archive will be coherent in the long run.

The Media Archiver software is based on Windows Media Encoder 9 series, which provides APIs to encode audio and video files into Windows Media Audio and Video formats. Such files can then be streamed from standard CERN web servers.

However, the architecture of Media Archiver is open and enables third-party encoders to be used. In fact, two open source products are already incorporated in the process: FFmpeg, for creating thumbnails of MPEG-2 videos and Ogg Vorbis to encode audio files in the OGG format.

The Media Archiver software creates added-value by effectively integrating standard CERN services: CERN Document Server supported by IT-UDS group and DFS and web servers supported by IT-IS group. Consequently, the synergies created allow to run the services effectively and the end-user gets a consistent experience.

## Supported formats



## References

1. CERN Document Server: <http://cdsweb.cern.ch/>
2. CERN Windows On Demand service: <https://www.cern.ch/winservices/Services/WoD/Default.aspx>
3. Windows Media Encoder 9 Series <http://microsoft.com/windows/windowsmedia/forpros/encoder/default.mspix>
4. Ogg Vorbis encoder: <http://www.rarewares.org/ogg-oggenc.php>
5. FFmpeg encoder: <http://ffmpeg.mplayerhq.hu/>