

# Digital Dunhuang: A Case Study for Digital Preservation and Digital Asset Management



Peter Zhou  
UC Berkeley

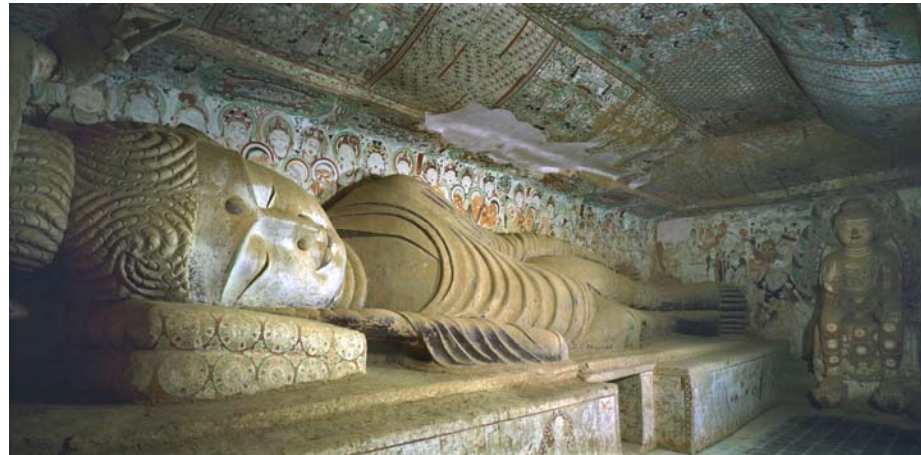
PRDLA Annual Meeting, Los Angeles, November 9, 2011





























# Teams that have worked on Dunhuang Project

- Dunhuang Academy
- The Getty Conservation Institute
- Northwestern University
- Zhejiang University
- Other universities and research institutions in China and elsewhere in the world





# Functional Requirements

- A platform for storing current and future content
- Permanent preservation of large files of varying formats, texts or images
- Delivery of content globally





# Components

## DAM

Facilitate asset creation, cataloging, image, video, and text file management and delivery, version control, and track digital preservation actions. Push metadata and content to the Digital Dunhuang platform. Manages master high resolution files and original documents





# Digital Preservation

Managed digital preservation actions include creating checksums, validating files, and extracting technical metadata upon ingest; monitoring file format obsolescence; migrate file formats; tracking and copying files to LTO tapes





# Digital Dunhuang database (subscribed/licensed)

Institutional subscription of images, video, articles, manuscripts, rare books, documents, and other publications related to Dunhuang Studies. Surrogates of master assets managed in the DAM will be pushed for external delivery to this system





# Content Categories

- Stitched/composite cave images
- Raw cave images
- Cave QTVRs
- Historical photos
- Videos
- Digital restorations
- Manuscripts from Cave 17 (ca. 400 mss)





- Artifacts (approximately 10,000 objects)
- Reproductions (copies) of images in caves
- Microfilm of manuscripts (digitized)
- Interactive panoramic of caves

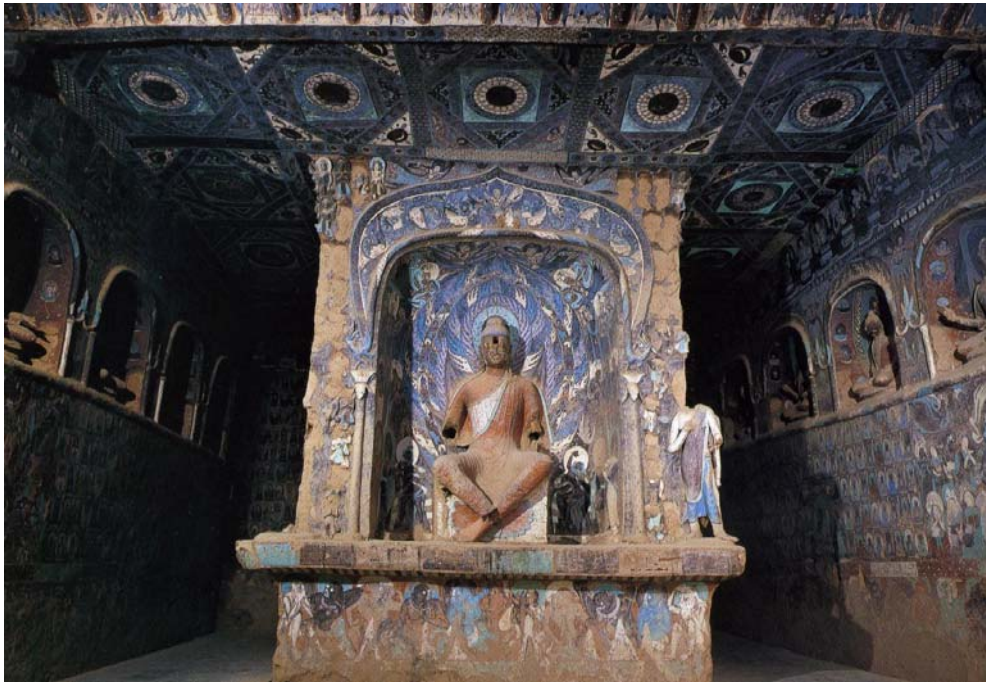




- Research created by members of the Dunhuang Academy
- Scholars' research publications: current
- Scholars' research publications: previously published
- Bibliographies, indices, glossaries, and finding aids created by staff
- Conservation data



- Climate monitoring data
- Conservation photography
- Conservation photography (legacy analog and current digital)
- UV digital conservation photography



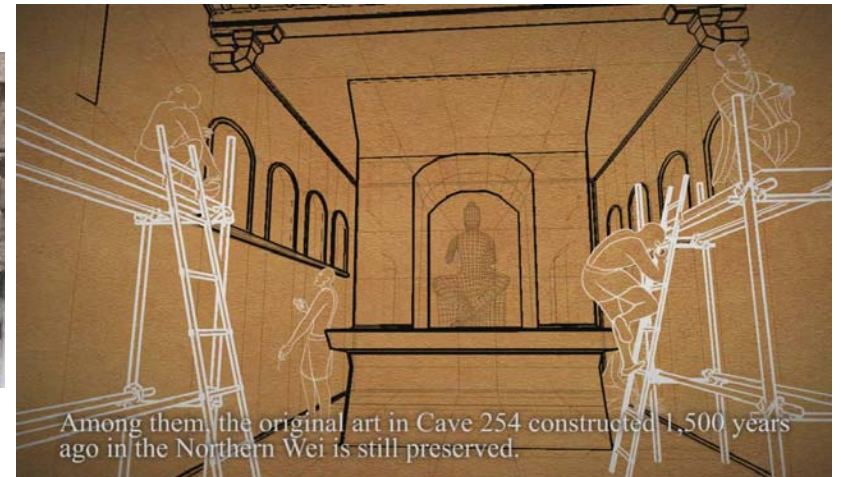


- Conservation materials
- Archaeological reports
- Archaeological drawings: hand-drawn
- and CAD drawings
- Archaeological reports: 3D laser cloud data points



# File formats

TIFF, JPEG, JPEG2000 (still image), PSD, BMP, PSB (Photoshop large file format), CR2 (Canon raw format), DCR (Kodak raw format), DNG (Adobe/universal raw format), other RAW camera formats (list), CDR (Corel Draw), CAD, PTX (original 3D cloud points), DGN (Microstation Design File), PDF, CAJ, MOV (QTVR), MPEG2/35 Mbps (AVI wrapper), HD video files (format TBD), DPG (Ninetendo video file format), Word, Excel, txt, MPEG4, H.264, FLV (Flash)





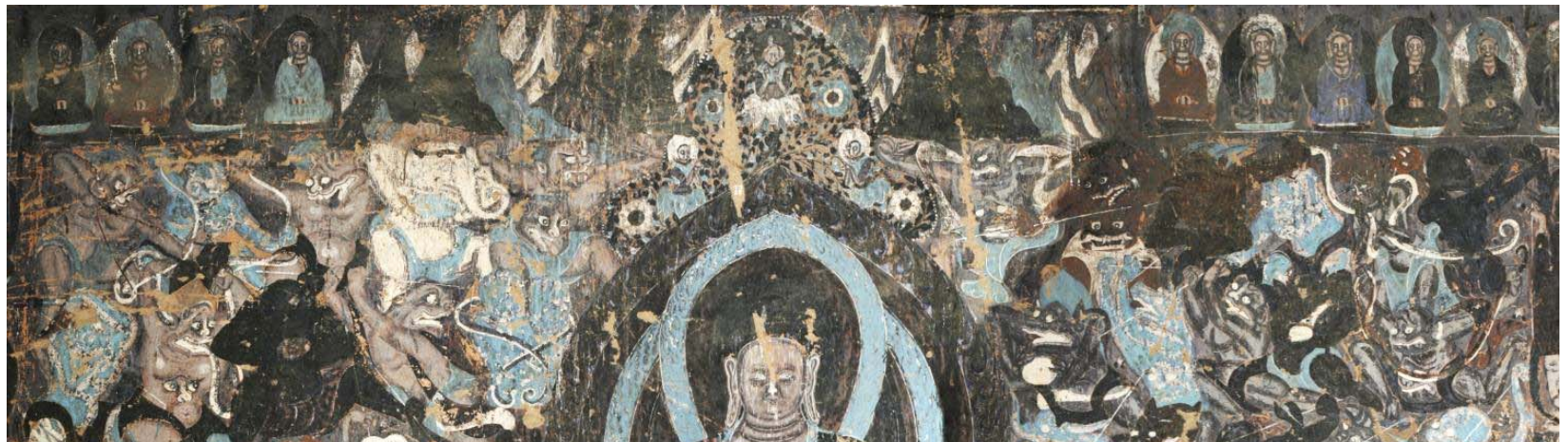
# Controlled vocabularies

- Thesauri (controlled vocabulary/thesauri ingest (hierarchical; not just a picklist)
- Specific thesauri are assigned to specific fields (subjects, names)
- Related terms
- Picklists/dropdowns



# Searching

- Browse & faceted search based on controlled vocabularies
- Keyword search in metadata records
- Keyword search within documents (not just metadata records)
- Boolean search
- Related keyword search, e.g. cave photos and articles about the cave
- Non-textual search, for example, searching for images based on color, or visual recognition





Digital Dunhuang's entry portal will include a panoramic view of the caves. Users should be able to click on a cave, and be offered options to find images, manuscripts, documents, publications, conservation data, archeological data, as related to that cave. The user could also zoom into the cave entrance, and see a QTVR of the cave for an interactive experience.



# Cross-linking

- Content and metadata linking. All content (images, documents, etc.) must be displayed with metadata. For example, an initial search result might display thumbnails with basic data. The user can then select to see a large image and fuller data.





# Major challenges

- Creation of a huge DAM as the backend file management system supported by a sophisticated metadata structure for workflow control and data ingestion
- Display of high resolution images at the front end
- Digital preservation of millions of files- duplicate storage, version control, error checking and data migration

# Thank you and Q&A

