

; 04/04/13 2:33 PM
;
;;;LEARNINGTIMES WEBIINAR 4/4/2013

>> HELLO, EVERYBODY, THIS IS MIKE FROM LEARNING TIMES. PLEASURE TO BE WITH YOU HERE AGAIN TODAY AND WE WILL GET STARTED IN ABOUT 30 SECONDS A REMIND, THE WINDOW YOU SEE ON THE LEFT WILL DISAPPEAR IN A FEW MOMENTS SO FEEL FREE TO GO AHEAD AND SAY HELLO WHILE YOU STILL CAN, IT WILL BE REPLACED BY A MODERATED CHAT WHICH MEANS THAT AS YOU TYPE YOUR QUESTION YOU WILL SEE THE QUESTION TWICE ON YOUR SCREEN. IT'S ONLY VISIBLE TWICE TO YOU. EVERYBODY ELSE WILL ONLY SEE IT ONCE. ONCE WE ANSWER THE QUESTION AND RELEASE IT. SO PLEASE DON'T GET CONCERNED IF YOU DO SEE IT TWICE. THE SESSION IS BEING RECORDED. THE RECORDING OF WHICH WILL BECOME AVAILABLE IN A DAY OR TWO ON THE COLLECTIONS WEB SITE. AS WELL THE WGBH CLOSED CAPTIONING TRANSCRIPT WILL ALSO BECOME AVAILABLE AT THAT TIME. IF ANYBODY HAS QUESTIONS OF A TECHNICAL NATURE REGARDING THE AUDIO OR THE CONNECTION FEEL FREE TO SEND US A CHAT ONCE THE MODERATED WINDOW BECOMES AVAILABLE. WE WILL RESPOND AS QUICKLY AS WE CAN. IF YOU ARE ON A WIRELESS CONNECTION PLEASE NOTE THAT IS IT POSSIBLE THAT YOU MAY HAVE INTERMITTENT SOUND. IF THAT IS THE CASE TRY

CONNECTING WITH AN EITHER NET
CABLE.

THAT TYPICALLY RESULTS IN BETTER
AUDIO QUALITY WITHOUT FURTHER
DELAY, I'M GOING TO PASS THE
AUDIO OVER TO OUR HOST KRISTIN
RAYS OF HERITAGE PRISON
SEPARATION.

KRISTEN, IF YOU'D LIKE TO GO
AHEAD

>> HELLO, EVERYONE, THIS IS
KRISTEN LAISE FROM HERITAGE
PRESERVATION AND WELCOME TO OUR
SECOND CLASS, CARING FOR DIGITAL
MATERIALS, PREVENTING A DIGITAL
DARK AGE, THIS IS PART OF OUR
SERIES: CARING FOR YESTERDAY'S
TREASURES TODAY WHICH HAS BEEN
MADE POSSIBLE THROUGH GENEROUS
SUPPORT THROUGH THE INSTITUTE OF
MUSEUM AND LIBRARY SERVICES AND
WE WANT TO THANK LEARNING TIMES
FOR THEIR HELP IN PRODUCING
THESE WEB SITES AND THIS WEBINAR
FOR US TODAY.

IF YOU ARE HEARING AN ECHO, YOU
MAY HAVE LOGGED IN TWICE SO YOU
MIGHT WANT TO CHECK THAT.

HOPEFULLY YOUR SOUND IS
OTHERWISE GOOD.

IF NOT FEEL FREE TO DROP US A
NOTE IN THE CHAT.

I WANTED TO APOLOGIZE AGAIN IF
YOU HAD TROUBLE ON TUESDAY
LOGGING IN.

WE WILL BE DOING MORE CLASSES
NEXT WEEK.

TODAY WE HAVE JACOB NADAL
TALKING ABOUT FILE CONVERSION
AND NEXT TUESDAY WE'LL BE
TALKING ABOUT METADATA.

NEXT WEDNESDAY SAFE ARCHIVING
THROUGH BACKUP COPIES AND THE
FOLLOWING WEEK ON MONDAY WE'LL
FINISH OUR COURSE TALKING ABOUT
NETWORKS FOR PRESERVATION.

JUST A QUICK FEW NOTES OF WHAT

YOU CAN FIND ON OUR COURSE WEB PAGE.

I PUT THE URL UP HERE.

WE HAVE A LINK TO THE POWERPOINT SLIDES YOU'LL SEE TODAY SO YOU DON'T HAVE TO RUSH AND TAKE NOTES, YOU CAN PRINT THEM OUT AND HAVE THEM HANDY.

A LINK TO OUR HOME WORK ASSIGNMENT FOR THIS AND A LINK TO ANY RESOURCES THAT HAVE BEEN REFERENCED IN THIS CLASS.

AND IF MORE ARE TALKED ABOUT DURING OUR CHAT TODAY WE'LL MAKE SURE TO GET THAT UP AS WELL. WE HAD SOME GREAT SUGGESTIONS FOR LINKS ON TUESDAY'S CLASS AND THAT IS GOING TO BE PUT UP HOPEFULLY TODAY OR FIRST THING TOMORROW.

JUST TO NOTE, THE RECORDINGS WON'T BE POSTED ON A PAGE AS A RULE OF THUMB.

YOU'LL GET THESE BY E-MAIL ONCE THE ENTIRE COURSE CONCLUDES WE'LL PUT UP ALL THE RECORDINGS. SO THANK YOU FOR LOOKING THAT THE SITE, FOR ALL THE INFORMATION THAT YOU MIGHT NEED. SO YOU HAVE THE OPTION IN COMING TO THESE COURSES TO WORK TOWARDS A CERTIFICATE OF COMPLETION AND A DIGITAL CREDENTIAL.

SO JUST MAKE SURE YOU'VE REGISTERED THROUGH OUR SITE, THAT YOU COMPLETE ALL HOME WORK ASSIGNMENTS AND WHILE WE DON'T PROVIDE INDIVIDUAL FEEDBACK ON HOME WORK ASSIGNMENTS WE'LL GIVE YOU A LITTLE BIT GENERAL FEEDBACK TO THE GROUP.

AND MAKE SURE YOU'VE DONE ALL OF THIS BY MONDAY, APRIL 22.

SO A WEEK AFTER THE VERY LAST WEBINAR AIRED.

OKAY, ALSO WANT TOO OLD MAKE SURE THAT YOU KNOW ABOUT THE

ONLINE COMMUNITY.

YOU PROBABLY ALREADY HAVE JOINED.

WE'RE ACTUALLY VERY CLOSE TO BREAKING 3,000 PARTICIPANTS LIMIT ON THAT SO WE HAVE ALMOST 3,000 PEOPLE THAT HAVE JOINED AND WE'RE REALLY HAPPY IT'S GROWING SO MUCH AND WE HOPE IT CONTINUES TO GROW.

SO IF YOU HAD QUESTIONS THAT DIDN'T GET ADDRESSED TODAY OR OTHER QUESTIONS ABOUT COLLECTIONS CARE, THIS IS A GREAT PLACE WHERE YOU CAN ASK YOUR PEERS AND A NUMBER OF CONSERVATORS HAVE BEEN ACTIVE PARTICIPANTS ON THIS LIST.

IF YOU HAVE ANY OTHER QUESTIONS ABOUT THE COURSE LOGISTICS OR ANYTHING WE TALK ABOUT TODAY, HERE'S OUR CONTACT INFORMATION SO PLEASE FEEL FREE TO SEND US AN E-MAIL OR GIVE US A CALL BY PHONE AND WE'LL MAKE SURE TO RESPOND TO YOU.

BEFORE WE GET TOO FAR GOING INTO THIS I WANTED TO FIND OUT A LITTLE BIT ABOUT YOU GUYS SO I WANT TO ASK A POLL QUESTION TO SEE WHAT TYPE OF INSTITUTION YOU'RE COMING FROM.

SO IF YOU COULD JUST CLICK YOUR RESPONSE HERE, THAT WOULD BE GREAT.

IT WOULD HELP OUR SPEAKERS GATE HANDLE ON WHO THE AUDIENCE IS. WE KNOW NOT ALL OF YOU WORK AT AN INSTITUTION.

SO IF YOU HAVE QUESTIONS OR QUESTIONS ABOUT HOME WORK, LET US KNOW.

HOME WORK ASSIGNMENTS DISCUSS INSTITUTIONAL COLLECTIONS BUT YOU CAN THINK ABOUT YOUR OWN PERSONAL DIGITAL COLLECTIONS WHEN YOU LOOK AT THE HOME WORK

QUESTIONS.

LOOKS LIKE A NICE MIX, NICE
BALANCE.

I'M GOING TO CLOSE THIS POLL AND
I WANT TO ASK ONE MORE QUESTION
IF I CAN AND THAT'S JUST TO GET
A SENSE OF THE SIZE OF YOUR
INSTITUTION.

WE TYPICALLY DO THIS WITH A
BUDGET -- WITH THE BUDGET OF
YOUR INSTITUTION.

>>

WE KNOW THAT MANY OF YOU ARE ON
SMALLER BUDGETS AND MANY OF
THESE SPEAKERS HAVE DONE THEIR
BEST TO SCALE THE ADVICE SO IT'S
APPLICABLE TO PEOPLE OF ALL
TYPES AND SIZES OF INSTITUTIONS.
SO GIVE IT JUST A FEW MORE
MINUTES FOR YOU TO LOG IN YOUR
REPLY.

I'M GOING TO GO AHEAD AND CLOSE
THIS POLL.

AND THEN I WANTED TO INTRODUCE
YOU TO DANIELLE PLUMBER.

YOU MAY HAVE SEEN HER RESPONDING
TO QUESTIONS IN TUESDAY'S
WEBINAR AND SHE IS REALLY
HELPING US BY COORDINATING ALL
OF THE CONTENT FOR THIS COURSE,
HELPING ANSWER YOUR QUESTIONS SO
WE CAN GET YOU GREAT FEEDBACK IN
THE CHAT AND SHE'S ASKED ALL OUR
SPEAKERS TO PARTICIPATE.

SHE'S GOING TO GIVE YOU A QUICK
OVERVIEW AND WHILE SHE DOES THAT
I'M GOING TO ACTUALLY DRAG AWAY
THE "HELLO" BOX.

NOT TO CUT ANYBODY OFF BUT WE'RE
GOING TO START THE MODERATED
CHAT NOW.

SO DANIELLE, I'LL TURN IT OVER
TO YOU.

>> THANK YOU, KRISTEN.

AND WE REALLY THANK HERITAGE
PRESERVATION FOR MAKING THIS
OPPORTUNITY POSSIBLE FOR US TO

WORK WITH YOU GUYS A LITTLE BIT ABOUT DIGITAL PRESERVATION. BECAUSE OF SOME OF THE PROBLEMS WE HAD ON TUESDAY WE WANTED TO BACK UP A LITTLE BIT AND JUST GIVE YOU A LITTLE BACKGROUND ABOUT WHY THIS CLASS WAS PUT TOGETHER AND THE SORTS OF THINGS WE'RE HOPING TO GET OUT OF THIS. SO THE FIRST QUESTION-- AND A FEW PEOPLE ASKED THIS ON TUESDAY-- "WHAT IS DIGITAL PRESERVATION?"

THIS IS A DEFINITION AND OUR MAIN PRESENTOR TODAY, JACOB NADAL, ACTUALLY WORKED ON HELPING TO COME UP WITH THESE DEFINITIONS.

BUT THE SHORT DEFINITION IS DIGITAL PRESERVATION COMBINES POLICIES, STRATEGIES AND ACTIONS THAT ENSURE ACCESS TO DIGITAL CONTENT OVER TIME."

MANY OF YOU WORK IN INSTITUTIONS THAT HAVE COLLECTIONS OF PHYSICAL MATERIALS: PAPER, ARTIFACTS, PERHAPS AUDIO/VIDEO MATERIALS.

YOU PRESERVE THEM.

DIGITAL HAS TO BE TREATED THE SAME WAY.

WE HAVE TO THINK ABOUT OUR DIGITAL CONTENT GOING FORWARD SO THAT'S WHY THIS CLASS WAS PUT TOGETHER.

THERE ARE REASONS TO DO IT.

THE FIRST IS COST.

IF YOU'VE GONE OUT AND DONE A DIGITIZATION PROJECT YOU KNOW THAT IT IS NOT FREE.

IF YOU PAY A VENDOR YOU HAVE A REALLY GOOD SENSE OF HOW MUCH IT COST YOU, BUT IF YO YOU DO IT IN HOUSE WITH YOUR OWN STAFF YOU MAY NOT EVEN THINK ABOUT, WELL, THAT TOOK 400 THUNDERSHOWERS OF STAFF TIME TO DO THE PROJECT.

BUT THE POINT IS, IF YOU PRESERVE THE DIGITAL CONTENT YOU CREATED YOU DON'T TO DO THAT WORK OVER AGAIN.

SO COST IS A BIG REASON WHY GOING FORWARD WE WANT TO KEEP THE THINGS WE'VE DONE.

THERE ARE SOME THINGS YOU'D LIKE THAT DO OVER BECAUSE YOU DON'T LIKE THE ORIGINAL RESULTS.

SLIGHTLY DIFFERENT CASE.

BUT WE DON'T WANT YOU TO WASTE YOUR MONEY TO HAVE SOMETHING DONE AND THEN LOSE IT.

AND THE SECOND REASON IS INSTITUTIONAL MISSION.

THIS IS ESPECIALLY TRUE FOR BORN DIGITAL MATERIALS.

THINGS THAT YOU DON'T HAVE ANY KIND OF PHYSICAL COUNTERPART FOR SO A DATABASE IS A REALLY GOOD EXAMPLE OF SOMETHING THAT YOU CAN'T PRINT THAT JUST DOESN'T WORK.

E-MAIL TO SOME EXTENT AND THEN ALSO THINGS THAT YOU KNOW YOU'RE NOT GOING TO BE ABLE TO PRESERVE THE ORIGINAL PHYSICAL FORM OF SOMETHING INDEFINITELY.

AND THIS IS RIGHT NOW TRUE OF AUDIO AND VIDEOTAPES WHERE THE MAGNETIC MEDIA THAT WE'RE USED FOR THESE IS DEGRADING AND WE HAVE TO FIGURE OUT HOW TO PRESERVE THE INTELLECTUAL CONTENT EVEN IF WE CAN'T PRESERVE THING A ACTUAL TAPE.

SO THIS IS PART OF YOUR INSTITUTIONAL MISSION TO PRESERVE THESE MATERIALS AND YOU HAVE TO THINK ABOUT THAT.

I'VE GOT LINKS, YOU CAN FIND THEM YOURSELF PRETTY EASILY.

BUT THE FIRST IS THE 1960 U.S. CENSUS.

IN 1960 THE CENSUS WAS DONE BY COMPUTER SO THE MATERIALS WERE

SENT TO PEOPLE, THEY WERE SENT BACK, SCANNED, AND COMPUTER TAPES WERE USED TO PROCESS ALL OF THE CENSUS RETURNS.

THE NATIONAL ARCHIVES MISSION IS TO PRESERVE THE RECORDS OF THE UNITED STATES GOVERNMENT.

SO THEY HAD TO FIGURE OUT HOW DO WE PRESERVE THIS COMPUTER, THIS DIGITAL VERSION OF THE U.S. CENSUS?

THERE WERE SOME PROBLEMS AND I THEY HAVE WRITTEN AN ARTICLE ABOUT IT.

THE GOOD NEWS IS THAT MOST OF THE DATA IS STILL THERE.

THEY WERE SUCCESSFUL WITH THEIR DIGITAL PRESERVATION PROJECT. THEY'RE ESTIMATED THAT 5% OF THE DIGITAL DATA WAS CORRECTED AND WON'T BE RECOVERABLE.

WE WON'T FIND OUT EXACTLY WHAT THAT DATA LOSS WAS UNTIL 2032 WHEN THAT CENSUS BLB LB AVAILABLE TO THE PUBLIC.

IN 1986, THE BBC IN ENGLAND DID A PROJECT CALLED THE DOMES DAY PROJECT WHERE THEY COLLECTED STORIES FROM ALL OVER ENGLAND. AND THIS WAS TO COMMEMORATE THE DOMES DAY BOOK WHICH WAS DONE IN AFTER WILLIAM THE CONQUERER INVADED ENGLAND IN 1066.

THEY PUT ALL THIS TOGETHER AND PUT IT ON TO TWO LASER DISKS. MOST OF YOU HAVE NEVER SEEN A LASER DISK.

I SAW THEM BUT I NEVER OWNED A LASER DISK PLAYER.

LONG STORY SHORT, ONLY ABOUT TEN YEARS, MAYBE 13 YEARS AFTER THIS PROJECT WAS DONE THEY REALIZED LASER DISK WAS A FORMAT THAT HAD NOT CAUGHT ON AND THEY NEEDED TO PRESERVE THE INFORMATION IN A DEVICE INDEPENDENT STATE.

SO THEY SPENT ALMOST TEN YEARS

GETTING THE DATA BACK OFF OF THE LASER DISK IS WHERE THEY'VE STORED IT AND FIGURE OUT HOW TO PRESERVE THAT INFORMATION AND MAKE IT ACCESSIBLE SO IT WAS A SUCCESS STORY BUT IT OPENED A LOT OF EYES.

SO I'M SURE YOU CAN THINK OF A LOT OF THINGS IN YOUR PERSONAL LIVES, FILES THAT YOU HAVE THAT YOU NO LONGER HAVE ACCESS TO, MATERIALS THAT HAVE BEEN DAMAGED OR DESTROYED EITHER BECAUSE OF DISASTERS OR JUST BECAUSE NOBODY THOUGHT ABOUT TRYING TO PRESERVE IT.

SO, AGAIN, THAT'S WHY WE'RE DOING THIS CLASS.

I WANT TO MAKE A NOTE ABOUT JARGON.

THIS IS ANOTHER COMMENT THAT CAME UP ON TUESDAY.

SOME OF THE TERMS AND CONCEPTS USED IN THIS MAY NOT BE FAMILIAR TO YOU BECAUSE THIS IS A NEW FIELD.

WE'VE ONLY BEEN DOING DIGITAL PRESERVATION -- SOME PEOPLE HAVE BEEN DOING IT FOR PERHAPS 20 YEARS BUT FOR MOST OF US IT'S BEEN WITHIN THE PAST FIVE YEARS OR SO.

SO WE'RE STILL TRYING TO FIGURE OUT WHAT TO DO WITH IT.

SOME PEOPLE ARE CALLING IT DIGITAL CURATION, DIGITAL PRESERVATION AND VARIOUS OTHER TERMS FOR THE SAME IDEA.

AND PEOPLE FROM DIFFERENT COMMUNITIES OF TRACK DISTECHNICAL ASSISTANCE-- AND LIBRARIES, ARCHIVES, MUSEUMS, HISTORICAL SOCIETIES USE DIFFERENT TERMS FOR THE SAME THING.

THAT MAKES IT REALLY HARD FOR US SOMETIMES TO SHARE INFORMATION

EFFECTIVELY.

I PUT A LINK UP TO THE NATIONAL DIGITAL STEWARD SHIP ALLIANCE, N.D.S.A., WHICH IS AN ORGANIZATION RUN BY THE LIBRARY OF CONGRESS.

THEY HAVE A GLOSSARY OF DIGITAL PRESERVATION TERMS UP AT

www.digital

preservation.gov/NDSA.

IT'S NOT COMPLETE BUT YOU CAN FIND A LOT OF TERMS THERE.

IF YOU DON'T A TERM OR CONCEPT THAT WE USE, PLEASE ASK US IN THE CHAT TO EXPLAIN IT AND WE'LL BE HAPPY TO DO THAT.

SO SERIOUS GOALS.

AT THE END OF THE SERIES WE WANT YOU TO HAVE A BETTER

UNDERSTANDING OF THE INHERENT FRAGILITY OF DIGITAL OBJECTS.

WE WANT TO HELP YOU SELECT PRESERVATION FORMATS, METADATA AND BACKUP SYSTEMS FOR YOUR DIGITAL OBJECTS.

AND WE WANT YOU TO BE ABLE TO IDENTIFY ONE OR MORE ACTIONS THAT YOU CAN TAKE TO IMPROVE YOUR INSTITUTION'S DIGITAL PRESERVATION EFFORTS.

AS KRISTEN MENTIONED, THIS IS THE SECOND SESSION, CONVERT IT TO PRESERVE IT, DIGITIZATION AND FILE CONVERSION.

IT WILL COVER BOTH CREATING DIGITAL MATERIALS AND ALSO SOME CONVERSION TIPS TO MAKE SURE IT CAN BE PRESERVED.

ON TUESDAY WE DID THE FIRST SESSION WHICH WAS AN OVERVIEW OF DIGITAL AND WE'VE MENTIONED WE GOT A LOT OF HOME WORK RESULTS FROM THAT ONE AND I WANTED TO PASS ON A NOTE FROM THE INSTRUCTOR WHO SAID SHE WAS VERY PLEASED.

SHE THOUGHT THE RESPONSES WERE

VERY REALISTIC.

SOME PEOPLE PUT ROLLS, LIKE ARCHIVISTS, FOR THEIR DREAM TEAM OF WHO THEY'RE HELPING TO PRESERVE MATERIALS AND STUFF.

PEOPLE PUT THE NAMES OF ACTUAL INDIVIDUALS IN THEIR ORGANIZATION, WHICH WAS GREAT SO THE IDEA IS GETTING YOU TO THINK ABOUT HOW TO REALISTICALLY APPROACH YOUR PROJECTS AND THAT WAS A BONDORFUL SUCCESS STORY FROM THE HOME WORK.

IF YOU HAVE QUESTIONS ABOUT THE HOME WORKS FEEL FREE TO CONTACT US ABOUT IT.

AFTER TODAY WE HAVE TWO MORE SESSIONS NEXT WEEK AND ONE THE FOLLOWING WEEK AS KRISTEN SAID SO I HOPE YOU WILL COME BACK AND JOIN US FOR THOSE.

WITH THAT, I WANT TO TURN IT OVER TO TODAY'S INSTRUCTOR.

JACOB NADAL.

>> HI, EVERYONE -- OH, SORRY.

>>

>> LET ME JUST INTRODUCE YOU. HE'S THE DIRECTOR OF THE LIBRARY AND ARCHIVES AT THE BROOKLYN HISTORICAL SOCIETY.

AND THAT'S A NEW JOB FOR HIM. HE JUST STARTED THAT IN APRIL. FROM 2008 TO TWELVE HE SERVED AS THE PRESERVATION OFFICER FOR U.C.L.A.'S LIBRARY AND FROM 2005 TO 2008 HE WORKED AS THE FIELD SERVICES LIBRARIAN AND ACTING HEAD OF COLLECTION CARE IN THE NEW YORK PUBLIC LIBRARY.

HE HAS A MASTER'S IN LIBRARY INFORMATION SCIENCE AND HE'S BEEN WIDELY RECOGNIZED FOR HIS ROLE ON VARIOUS PAST COURSES AND GROUPS WITHIN THE AMERICAN LIBRARY ASSOCIATION PRESERVATION AND REFORMATTING SECTION AND IN OTHER AREAS.

SO WE'RE VERY GLAD TO HAVE HIM
TODAY TO TALK TO YOU ABOUT
DIGITIZATION AND FILE CONVERSION
GOOD DAY, EVERYONE, I'M
OBVIOUSLY EAGER TO TALK TO YOU
ABOUT THIS.

THANK YOU, DANIELLE, FOR THE
INTRODUCTION AND KRISTEN,
HERITAGE PRESERVATION, EVERYONE
AT LEARNING TIMES, THANKS FOR
PROVIDING THE TECHNICAL SUPPORT
REQUIRED TO GET THIS PROGRAM
GOING.

TODAY WE'RE REALLY GOING TO TALK
ABOUT HOW TO MAKE SOMETHING
DIGITAL IF IT'S NOT DIGITAL IN
THE FIRST PLACE.

THEN ON TOP OF THAT TALK A
LITTLE BIT ABOUT THE
CHARACTERISTICS OF DIGITAL
INFORMATION, WHAT IS IT THAT
MAKES A FILE THE SORT OF BASIC
UNIT OF DIGITAL INFORMATION LONG
LASTING, LOVE LINGED AND
EVENTUALLY PRESERVEABLE.

TODAY WE'LL TALK ABOUT A COUPLE
MAJOR TYPES OF DIGITAL MATERIALS
TEXT AND IMAGES WILL BE OUR
PRIMARY FOCUS AND THIS IS THE
CORE OF WHAT EXISTS IN MOST LIE
LIBRARY AND ARCHIVES COLLECTION.
WE'LL ALSO SPEND TIME TALKING
ABOUT AUDIO.

MORE AND MORE PEOPLE ARE DOING
AUDIO NOT JUST FOR MUSIC BUT FOR
ORAL HISTORY PROJECTS, WEBINARS
LIKE THIS ONE.

SO AUDIO I THINK IS A FORMAT
GROWING IN IN IMPORTANCE IN OUR
COMMUNITY.

THESE THREE FORMATS-- TEXTS,
IMAGES, AND AUDIO-- ARE ALSO THE
FORMATS FOR WHICH WE HAVE PRETTY
GOOD PRESERVATION ANSWERS AT
THIS POINT.

TOWARDS THE END OF THE WEBINAR
I'LL TALK BRIEFLY ABOUT VIDEO

DATA AND INTERACTIVE SYSTEMS.
WE DE-EMPHASIZE THOSE A LITTLE
BIT PARTLY BECAUSE OF THE TYPES
OF INSTITUTIONS WHO ARE INVOLVED
IN.

THIS IT'S LESS COMMON FOR THEM
TO HAVE THESE TYPES OF
COLLECTIONS.

ALSO BECAUSE THIS IS AN AREA
WHERE PRESERVATION CONSENSUS IS
STILL EMERGING.

THERE ARE LOTS OF VERY TECHNICAL
QUESTIONS BEING DEBATED AND VERY
FEW GUARANTEES ABOUT HOW TO MAKE
THE RIGHT CHOICE HERE.

FOR EACH OF THE FORMATS WE TALK
ABOUT TODAY I'LL TELL YOU A
LITTLE BIT ABOUT HOW THE FORMAT
IS DESIGNED.

SO A DIGITAL FORMAT HAS A
PARTICULAR DECIDED-UPON
STRUCTURE AND ENGINEER HERE
INING BEHIND IT.

WE'LL TALK ABOUT THE RISKS AND
ADVANTAGES OF THOSE FORMATS FROM
A PRESERVATION POINT OF VIEW.
AND WE'LL TALK ABOUT THE KEY
SPECIFICATIONS FOR CREATING OR
WORKING WITH VENDORS TO CREATE
FILES IN THOSE FORMATS.

I WANT TO JUST TOUCH AGAIN ON
SORT OF THE DEFINITION OF
DIGITAL PRESERVATION AND SHOW
YOU THE MEDIUM FORM VERSION OF
THAT DEFINITION.

THERE'S ACTUALLY A LONG FORM AS
WELL UP ON THAT A.L.A. WEB SITE.
THE MEDIUM FORM DEFINITION
EXPANDS ON WHAT DANIELLE
PRESENTED TO YOU BY MENTIONING
BOTH REFORMATTED AND BORN
DIGITAL CONTENT.

AND TODAY WE'RE GOING TO FOCUS
LARGELY ON REFORMATTED, BUT ALSO
ADDRESS ISSUES OF BORN DIGITAL
CONTENT.

SO WE'RE REALLY AT THIS POINT IN

THE WEBINAR AT THE BEGINNING OF THIS PROCESS TALKING ABOUT HOW YOU GET THE CONTENT.

OTHER SPEAKERS WILL FOLLOW UP ON PIECES OF THIS RELATED TO MEDIA FAILURE, CHANGES IN TECHNOLOGY AND TALK ABOUT RENDERING AND AUTHENTICATING CONTENT OVER TIME.

SO WITH THAT SAID, LET'S START IN ON TEXT.

IS THIS ONE OF OUR POLL QUESTIONS, KRISTEN?

>> YUP, IT'S ONE OF OUR POLL QUESTIONS.

LET ME BRING THIS OVER.

>>

AND JUST TO NOTE, IF YOU DON'T CURRENTLY DO THIS, EACH OF THESE POLL QUESTIONS WILL GIVE YOU A PLACE TO SAY "NOT CURRENTLY BUT PLAN TO."

IF YOU DON'T KNOW, BE HONEST ABOUT THAT, TOO, BECAUSE IT TELLS US A LOT.

>> YEAH, I SUSPECT, ACTUALLY, IF WE DID THIS POLL AT THE BEGINNING AND END WE MIGHT GET DIFFERENT RESULTS.

>> I'M HOPING WE WILL!

>> OKAY, GOOD, SO IT LOOKS LIKE A LOT OF ATTENTION WITH HAVING BEEN INVOLVED IN TEXT DIGITIZATION OR HAVING PLANS FOR IT.

>>

THAT'S GREAT.

ALL RIGHT, TEXT IN THE DIGITAL WORLD REALLY JUST MEANS TEXT.

IT'S JUST LETTERS.

AND THE ENCODING OF TEXT WE LIKE FOR DIGITAL PRESERVATION IS WHICH IS A SUBSET OF UNICODE. THERE'S SOMETHING CALLED ASCII TEXT WHICH WE'LL TALK ABOUT BRIEFLY.

IT'S LIMITED.

IT'S 128 CHARACTERS, A LOWER
CASE A AND A CAPITAL A ARE TWO
DIFFERENT CHARACTERS,
PUNCTUATION MARKS, NUMBERS AND
SOME CONTROL CHARACTERS, THINGS
LIKE CARRIAGE RETURN OR DELETE.
USEFUL IF YOU ARE SPEAKING
ENGLISH, DECREASINGLY LESS
USEFUL IF YOU'RE SPEAKING, SAY,
FRENCH OR SPANISH.
OF ALMOST NO USE TO YOU IF
YOU'RE SPEAKING CHINESE OR
JAPANESE OR HINDI.
YUAN CODE IS A MUCH BROADER
CHARACTER SET.
WE'LL LOOK AT IN THE A MOMENT.
AND UTF-8, THE UNICODE
TRANSMISSION FORMAT IS THE BREAD
AND BUTTER OF DIGITAL TEXT
PRESERVATION.
AND I'LL SAY AGAIN, IT'S
IMPORTANT TO REMEMBER THIS HAS
NO FONT FACE ASSOCIATED WITH IT.
VERY LITTLE IN THE WAY OF LAYOUT
NOTHING MUCH MORE THAN A
CARRIAGE RETURN.
BUT TEXT IS WHAT'S CRITICAL FOR
SEARCHING AND MANIPULATING DATA.
IF YOU TYPE A SEARCH INTO THE
GOOGLE SEARCH BOX, IT'S GOING
OUT AND IT'S LOOKING THROUGH THE
RAW TEXT OF WEB PAGES.
IT DOESN'T PARTICULARLY CARE IF
THE TEXT IS PRESENTED IN TIMES
NEW ROMAN OR AERIAL OR HELL HELL
VET A.
IT'S LOOKING JUST AT THE
CHARACTER DATA.
UTF IS IMPORTANT BECAUSE THAT'S
THE DEFAULT ENCODING OF XML.
YOU'LL HEAR IN MY SESSION AND
THE SESSIONS TO COME PEOPLE TALK
ABOUT XML OVER AND OVER
SO THIS IS THE ASCII CODE CHART
FOR THE MAR FAT I WAS TALKING
ABOUT.
ANS A SKI CHARACTER HAS SEVEN

BITS TO IT.

THESE ARE THE ZEROS AND ONES
TWRIN THE HARD DISK.

SO IF YOU WANT TO WRITE CAPITAL
LETTER A AND YOU LOOK IN COLUMN
4 ROW 2 YOU WOULD KNOW THE BITS
THAT WOULD BE ON THE DISK.

A 1, 0, 0, 0.

A 1, 0, 0.

AND THAT'S AN ASCII CHARACTER OF
THE CAPITAL LETTER A.

AS I SAID EARLIER THIS IS USEFUL
FOR CERTAIN LANGUAGES, NOT FOR
OTHERS.

IT'S A LIMITED CHARACTER SET.
SO IT'S REALLY MOVED ON TO
UNICODE.

UNICODE IS AN IDEA OR A
SPECIFICATION OR STANDARD.
IT'S A WAY OF REPRESENTING TEXT.
UTF-8 IS THE FORMAT.

UTF-8 IS THE ACTUAL MECHANISM
FOR RECORDING UNICODE TO A DISK.
IT'S IMPORTANT TO KNOW THIS,
YOU'LL HEAR PEOPLE USE THE TWO
TERMS ESSENTIALLY
INTERCHANGEABLY IN CONVERSATION
WHEN IT COMES DOWN TO THE RUBBER
ON THE ROADTOR BITS ON THE DISK
IN THIS CASE UTF-8 IS WHAT
YOU'LL ACTUALLY ENCOUNTER.

UTF-8 CHARACTERS ARE LONGER.
INSTEAD OF EIGHT LITTLE BITS
THEY CAN BE FROM ONE TO FOUR
OCTETS, ONE TO FOUR EIGHT BIT
BITES.

THE FIRST CHARACTERS ARE THAT
US-ASCII CHARACTER SET WE LOOKED
AT SO IT'S BACKWARDS COMPATIBLE
WITH THE US-ASCII CHARACTERS.
AFTER THAT THERE ARE LOTS OF
OTHER THINGS.

WE'LL LOOK AT A UTF-8 CHART IN A
SECOND.

THE VIRTUES OF UTF-8, ASIDE FROM
DOING A BETTER JOB OF
REPRESENTING LANGUAGE, EASY TO

IDENTIFY.

SO IF THERE'S AN UNKNOWN STRING OF TEXT ON A DISK, THERE ARE SIMPLE SEARCH PATTERNS THAT WILL CORRECTLY FIND UTF 8 MORE THAN 99.5% OF THE TIME.

THIS IS IMPORTANT IF YOU'RE CONSIDERING THAT YOU HAVE MEDIA THAT YOU DON'T KNOW THE SOURCE OR CONTENTS.

SOMETHING LIKE UTF-8 IS EASIER TO FIND.

IT'S THE DEFAULT FOR XML AND HAS MULTILANGUAGE SUPPORT.

SO THIS IS SOME OF THE UTF-8 CHARACTER SET.

IT GOES ON AND ON AND ON AND ON. I CAN SEE HERE SOME ARMENIAN CAPITAL LETTERS, SOME BENGALI LETTERS AND SOME THAI CHARACTERS AND UTF-8 CODE CHARTS INCLUDE MANY LANGUAGES.

THEY INCLUDE MANY SOME BOLL I CAN SETS SUCH AS MATHEMATICS.

IF YOU CAN IMAGINE THERE ARE SOME DISCIPLINES THAT RELY ON THIS VERY HEAVILY.

IF YOU'RE LOOKING FOR CONVERGE AND YOU'RE WORKING WITH SOMEONE TO DO O.C.R. OR TEXT PRESCRIPTION FOR YOU YOU AT THIS POINT IN THE GAME IN THE YEAR 2013 SHOULD BE REQUESTING UTF-8 ENCODED TEXT FILES OF THOSE PRESCRIPTIONS AND OCR.

IF YOU'RE RUNNING AN OCR SOFTWARE PROGRAM, GET INTO THE SETTINGS AND MAKE SURE YOU'RE CREATING UTF-8 AND THAT WILL GIVE YOU THE MOST LONG-LIVED TEXT.

AND

THAT'S SOMETHING YOU WOULD DO FOR CONVERSION AND CREATING NEW TEXT.

IF YOU ARE CREATING TEXT DOCUMENTS, WEB PAGES OR XML YOU

WANT TO MAKE SURE YOU'RE
CREATING THEM UTF-8.
SO THOSE THINGS WE JUST SAW,
THOSE UNICODE CHARACTERS AND
THAT ASCII CHART WERE IMAGES
THAT I PUT ON MY SLIDE.
COMPUTERS DON'T READ, THEY
ENCODE AND DECODE.
WE HAPPEN TO KNOW SOMETHING IS A
LETTER.
THE COMPUTER SEES 1001000.
FOR US TO HAVE A DIGITIZED BOOK
WE NEED PAGE IMAGES, A PICTURE
THAT THE PAGE LOOKS AS WELL AS A
TEXT TRANSCRIPTION OF THAT PAGE
AND METADATA THAT BINDS ALL OF
THOSE THINGS TOGETHERING IN THE
PROPER ORDER.
SO WHEN WE'RE DOING REFORMATTING
WHEN WE'RE DIGITIZING MATERIAL
IT IS WAY WE GENERALLY DO THAT
IS TO EITHER REKEY IT OR TO USE
OPTICAL CHARACTER RECOGNITION.
OCR.
SOME OF YOU I KNOW HAVE
ENCOUNTERED THIS AND SOME OF YOU
PROBABLY WILL IN THE NEAR
FUTURE.
IT'S WORTH HAVING A LITTLE BUYER
BEWARE MOMENT TO SAY THAT OCR
ACCURACY IS USUALLY REPORTED AS
CHARACTER-LEVEL ACCURACY FROM
IDEAL SOURCES.
SO THAT'S A RELATIVELY MODERN
CLEANLY PRINTED TYPESET DOCUMENT
IT'S IMPORTANT TO NOTE THAT THE
ACTUAL OUTCOMES YOU GET IN TERMS
OF NUMBER OF WORDS ACCURATELY
TRANSCRIBED AND NUMBER OF
ACCURATE CHARACTERS FROM
LESS-THAN-PERFECT SOURCES IS
LOWER THAN THAT.
OCR HAS PRETTY HIGH CHARACTER
LEVEL ACCURACY AT THIS POINT.
ITS WORD-LEVEL ACCURACY AND ITS
ACCURACY FOR SUBPAR SOURCES IS A
LITTLE BIT LOWER.

BUT IT OFTEN GIVES YOU A PRETTY
READY TEXT.

THE PROBLEM YOU WILL ENCOUNTER
HERE IS THAT A SINGLE CHARACTER
MISSPELLED IN A WORD OR
MISIDENTIFIED WILL RUIN A TEXT
SEARCH.

REMEMBER COMPUTER SEARCHING USES
THAT RAW TEXT AND IF THAT
CHARACTER DATA IS BAD THE SEARCH
QUERY I TYPE IN MAY NOT RETURN
ACCURATE RESULTS.

SO DEPENDING ON THE INTENTIONS
OF YOUR PROJECT, IT'S WORTH
PAYING ATTENTION TO WHETHER OCR
OR REKEYING IS BEST.

TEXT ALSO HAS SORT OF TWO USE
CASES FOR IT.

WE'VE BEEN TALKING A LOT ABOUT
TEXT AS A SOURCE OF INFORMATION,
TEXT AS A DOCUMENT.

IT'S ALSO IMPORTANT TO RECOGNIZE
THAT TEXT IS THE BUILDING BLOCK
OF THE DIGITAL LIBRARY SO XML,
HAITI MILLION AND CASCADING
STYLE SHEETS THE CODE ALL OF
THOSE THINGS ARE WRITTEN AND
RECORDED IN TEXT.

SO IN ADDITION TO THE DOCUMENTS
OF THE DIGITAL LIBRARY, THE
DIGITAL LIBRARY ITSELF IS BEING
LARGELY BUILT UP OF TEXT FILES
THAT ARE APPROPRIATED IN A
CERTAIN WAY.

SO A WEB PAGE, FOR INSTANCE, IS
A TEXT DOCUMENT AND WHEN YOU
POINT A BROWSER AT THAT TEXT
DOCUMENT IT KNOWS TO RENDER
COLORS AND POSITIONS AND LAYOUTS
FOR DOCUMENTS IN A DIGITAL
LIBRARY, OFTEN YOU WILL GET
THESE AS TEXT PLUS, SO A
MICROSOFT WORD DOCUMENT HAS TEXT
IN IT BUT HAS A LOT OF OTHER
THINGS THAT CAN CONTROL STYLE
AND FORMATTING AND LAYOUT.

I WANT TO TOUCH ON THIS LIGHTLY.

WE COULD GO WAY DOWN THE RABBIT HOLE HERE BUT JUST TO SAY THAT MICROSOFT WORD HAS BECOME A SORT OF DE FACTO STANDARD.

ESPECIALLY MOST RECENT VERSION IS XML BASED BECAUSE MICROSOFT WORD WANTS TO DO DOCUMENT SHARING ON THE WEB.

THIS DOESN'T MAKE MICROSOFT WORD A GREAT PRESERVATION FORMAT FROM AN ENGINEERING POINT OF VIEW.

NO ONE IS WHO IS TRYING TO CREATE A LONG-LIVED DOCUMENT FORMAT WOULD CREATE MICROSOFT WORD.

IT MAKES IT A GOOD PRESERVATION FORMAT FROM A RISK MANAGEMENT POINT OF VIEW, HOWEVER.

IF THERE'S EVER A PRESERVATION PROBLEM WITH MICROSOFT WORD YOU WILL NOT BE THE ONLY PERSON HAVING IT.

HUNDREDS AND THOUSANDS AND MILLIONS OF PEOPLE WILL BE RIGHT IN THERE WITH YOU AND YOU CAN LOOK TO A COMMUNITY SOLUTION. P.D.F. IS ANOTHER WAY THAT WE GET TEXT IN OUR COLLECTIONS, ESPECIALLY BORN DIGITAL COLLECTIONS.

P.D.F. IS A FORMAT WITH AN OPEN LICENSE.

IT CAN CONTAIN TEXT, IMAGES, AUDIO, VIDEO, FILL-IN FORMS, A WIDE VARIETY OF THINGS.

SO KNOW THAT THERE'S A SUBSET OF P.D.F., P.D.F./A FOR ARCHIVAL AND IT'S BASED ON -- I HAVE A MISPRINT IN MY SLIDE-- P.D.F.

1.4.

PEED JEFF CURRENTLY I THINK VERSION 1.7.

THIS P.D.F. 1.4, P.D.F./A HAS A SUBSET OF FEATURES THAT ARE CONSIDERED PRESERVEABLE.

SO IF YOU HAVE P.D.F. AND YOU CAN CHECK THEIR VERSION OR

CONVERT THEM TO P.D.F./A
SUCCESSFULLY AND SEE NOTHING
CONTENT IS CHANGED YOU HAVE A
LONG LIVED FILE FORMAT.
LATER VERSIONS OF P.D.F., .5, .6
AND P.D.F. WITH COMPLEX OBJECTS
IN IT CAN PRESENT PROBLEMS DOWN
THE ROAD.
SO, AGAIN, JUST TO REVIEW THE
KEY SPECIFICATIONS HERE.
UTF-8 ENCODED UNICODE TEXT IS
THE -- ONE OF THE FACTORS WE'RE
LOOKING FOR.
WE'RE LOOKING FOR USING
XML-BASED FORMATS FOR MARKUP.
SO FAR MARKING UP A WEB PAGE,
FOR INCLUDING METADATA IN
DOCUMENTS.
XML SHOULD BE THAT WE'RE
CARRYING IT, SUPPORTING OUR
DATA.
BE CLEAR ABOUT HOW YOU'RE
GETTING TEXT OUT OF DOCUMENTS.
SO IF YOU'RE CONVERTED TEXT --
PRINTED ANALOG TEXT TO DIGITAL
LEARN A LITTLE BIT ABOUT OCR,
INVESTIGATE REKEYING.
AND MAKE SURE THAT YOU HAVE
METADATA THAT CARRIES THOSE
DETAILS SO THAT THE ARCHIVIST
WHO FOLLOWS YOU UNDERSTANDS THAT
THIS TEXT CAME FROM, FOR
INSTANCE, AN OCR ENGINE THAT WAS
RUN IN 2013 AND PRODUCED THIS
TEXT.
THAT WILL HELP LATER GENERATIONS
UNDERSTAND WHEN WHAT THEY'RE
WORKING WITH.
FOR DOCUMENTS, BORN DIGITAL
DOCUMENTS LIKE WORD PROCESSING
OR REPORTS, KNOW YOUR VERSIONS.
P.D.F./A WHEREVER POSSIBLE AND
THE .DOCX OR XLS VERSIONS
WHEREVER POSSIBLE FOR OFFICE
DOCUMENTS.
SO WE'LL PAUSE NOW AND TAKE SOME
QUESTIONS ABOUT TEXT.

THEN WE'LL MOVE ON TO OUR POLL
FOR IMAGE FORMATS.

DANIELLE, DO YOU WANT TO START
US GOING?

>> OKAY.

WE DID HAVE A NUMBER OF
QUESTIONS THAT CAME IN AS YOU
WERE GOING OVER THAT.

A COUPLE OF THEM ARE FORWARD
LOOKING AND WE MAY ADDRESS THEM
LATER OR IN OTHER SESSIONS SO
THE FIRST ONE WAS WILL
INFORMATION BE GIVEN ON THE
PROBLEM OF COMPUTER SYSTEMS
THEMSELVES GOING OUT OF DATE,
I.E. WINDOWS 98, WINDOWS XP, WIN
DOTS 7.

DO YOU HAVE ANYTHING YOU'D LIKE
TO SHARE ABOUT THAT?

>> THIS MAY GET TOUCHED NONLATER
SESSIONS.

THERE'S ACTUALLY -- LET ME
REPHRASE YOUR QUESTIONS A LITTLE
BIT.

THERE ARE ACTUALLY TWO FACTORS
HERE, ONE IS THE OPERATING
SYSTEM ITSELF, THESE VERSIONS OF
WINDOWS ARE DIFFERENT OPERATING
SYSTEMS AND THEN THERE ARE THE
PIECES OF APPLICATION SOFTWARE,
THE PROGRAMS THAT RUN ON THAT
OPERATING SYSTEM.

THIS IS AN AREA WHERE THERE'S A
SPREAD OF POSSIBILITIES SO SOME
FORMATS WORKED GRACEFULLY
BETWEEN DIFFERENT PROGRAMS AND
THE FORMATS WE TALKED ABOUT
TODAY ARE IN SOME SENSES
PREFERRED BY THE PRESERVATION
COMMUNITY BECAUSE THEY HAVE A
HIGH DEGREE OF INDEPENDENCE FROM
EITHER -- BOTH AN OPERATING
SYSTEM AND A PARTICULAR PROGRAM.
YOUR CHOICES, YOUR OPTIONS WHEN
YOU HAVE A -- AN OS AND PROGRAM
MIX RUPP FAIRLY LIMITED.

YOU REALLY HAVE TO EITHER RUN

THAT PROGRAM ON A VIRTUAL
COMPUTER, WHAT'S CALLED
EMULATION OR MAINTAIN A WORKING
VERSION OF ANOTHER SYSTEM OR
CONVERT.

FORTUNATELY, I THINK AS WE MOVED
INTO THE NETWORK ERA THE PROBLEM
OF OPERATING SYSTEMS AND
PROPRIETARY FORMATS IS IN SOME
WAYS KIND OF BLOCKED INTO THIS
SORT OF TURN OF THE 21st CENTURY
MOMENT.

GOING FORWARD WITH SEE EASE AND
COMPATIBILITY SO TO A CERTAIN
EXTENT PROBLEM CAN SOLVE ITSELF
IN THE INTERNET.

BUT, YEAH, IT CERTAINLY CAN
NEVER BE A FACTOR THAT WE NEED
TO WATCH.

NEXT ONE, DANIELLE?

>> OKAY.

THIS ONE I'M PRETTY SURE WE LOOK
AT TWO IN A BIT BUT WHAT ARE THE
BEST DATA FORMATS TO SAVE DATA
IN THAT WILL CARRY OVER THROUGH
TIME AND CHANGES IN COMPUTER
SYSTEMS?

>> YEAH.

SO WE'LL ACTUALLY TALK ABOUT
THIS OVER AND OVER TODAY.

THE TWO GOOD PIECES OF THE
DEVICE AS WE GO THROUGH EACH OF
THESE FORMATS WE'LL TALK ABOUT
THE PREFERRED PRESERVATION
FORMATS.

THOSE AREN'T ALWAYS THE ONES YOU
GET FOR BORN DIGITAL
COLLECTIONS.

SO -- NOR ARE THEY ALWAYS THE
ONES THAT ARE AVAILABLE TO YOU.
SO THE OTHER PIECE OF ADVICE, IF
YOU CAN'T HAVE THE OPTIMAL
STANDARD, RUN WITH THE HERD.
SO IF YOU CHOOSE ANY OF THE
MICROSOFT FORMATS, AGAIN, NO
PRESERVATION WOULD HAVE CREATED
THOSE FORMATS FOR THE PURPOSES

OF PRESERVATION BUT THE FACT THAT MILLIONS OF OTHER PEOPLE HAVE THE SAME PROBLEM YOU DO MEANS THAT YOU'RE PRETTY SAFE FROM A RISK MANAGEMENT PERSPECTIVE.

YOU KNOW KNOW WHEN A CHANGE IS COMING AND WHEN A CHANGE DOES HAPPEN THERE WILL BE LOTS OF TOOLS TO HELP YOU MIGRATE FILES FORWARD.

SO CHOOSE THE BEST STANDARD AND FAILING THAT RUN WITH THE CROWD.

>> OKAY.

WE HAD A SERIES OF QUESTIONS ABOUT P.D.F., OCR, P.D.F./A.. I'M GOING TO KIND OF CONDENSE SOME OF THESE INTO A COUPLE DIFFERENT QUESTIONS.

WE DID HAVE SOME QUESTIONS ABOUT OCR AND THE CHALLENGE OF CHARACTER RECOGNITION OF ALPHABETS THAT CONTAIN THINGS LIKE FRENCH LETTERS, HANDWRITING OTHER THINGS, SOMEONE THEN ASKED ABOUT REKEYING AND WONDERED IF YOU CAN EXPLAIN THAT MORE.

>> SO OCR IS PRETTY SOLID FOR ENGLISH AND ROMANCE LANGUAGES AND PROGRESSIVELY LESS AVAILABLE AS YOU MOVE INTO OTHER CHARACTER SETS OUTSIDE OF THE KIND OF LATIN CHARACTERS.

GOOGLE NOTABLY HAS OCR BOOKS IN DOZENS OF DIFFERENT CHARACTER SETS IN THEIR GOOGLE LIBRARY PROJECT.

THEY HAVEN'T NECESSARILY MADE THAT SOFTWARE AVAILABLE TO THE REST OF US AND ACCURACY DECLINES FOR NON-LATIN SCRIPT.

SO THAT'S STILL A TECHNOLOGICAL PROBLEM RESOLVED.

THE WAY REKEYING WORKS -- THERE'S A VARIETY OF WAYS.

THE SIMPLEST ONE IS TO HAVE A PERSON SIT AND TYPE WHAT THEY

SEE ON THE SCREEN OR PAGE.
MANY PEOPLE HAVE OUTSOURCED THIS
FUNCTION.

OFTEN OVERSEAS TO SHOPS WHERE
TWO OPERATORS TYPE THE DOCUMENT
KIND OF BLIND FROM EACH OTHER
AND THEN THEIR TWO
TRANSCRIPTIONS ARE COMPARED.
AND THIS IS ACTUALLY HIGHLY
EFFECTIVE.

WHAT HAPPENS -- EVEN IF THE
OPERATORS AREN'T NATIVE LANGUAGE
SPEAKERS-- IN FACT, SOMETIMES
THE RESULTS ARE BETTER IF
THEY'RE NOT-- THE CHANCES OF
THEM BOTH MISTYPING THE SAME
WORD ARE INFINITESIMALLY SMALL
AND FOR LARGE CONVERSION
PROJECTS THIS IS SOMETIMES EVEN
MORE COST EFFECTIVE THAN DOING
OCR AND CLEANING UP THE OCR
TRANSCRIPT.

THIS IS ALSO SOMETHING THAT
DOESN'T REQUIRE YOU TO
PHYSICALLY SEND THE ORIGINALS IF
YOU HAVE GOOD IMAGE CAPTURE.
IT CAN BE DONE REMOTELY.
SO THERE ARE A NUMBER OF VENDORS
THAT WILL DO REKEYING FOR YOU.
JUST WHAT YOU WOULD THINK IT
WOULD BE.

SO LET ME ALSO SAY A LITTLE
THING ABOUT P.D.F.
WE THINK ABOUT P.D.F. AS USERS
AS A FILE FORMAT.

IT'S BETTER FOR HAVING YOUR SORT
OF PRESERVATION HAT ON TO THINK
OF P.D.F. AS A METADATA FORMAT.
P.D.F. IS A SET OF INSTRUCTIONS
FOR HOLDING DIFFERENT TYPES OF
MEDIA.

THE MOST COMMON IS TO CREATE
PORTABLE DOCUMENTS BUT THINK OF
PDF AS A METADATA WRAP AND THINK
OF SOMETHING THAT SAYS HERE IS A
PAGE, IT'S PAGE ONE, IT HAS THIS
TEXT ON IT AND THIS TEXT IS

FORMATED TO LOOK LIKE THIS AND
IN THE MIDDLE OF THE TEXT THERE
IS AN IMAGE AND THE NAME OF THE
IMPAJ IS FIRSTIMAGE.JPG.

PDF WILL MAKE MORE SENSE IN
TERMS OF HOW IT FUNCTIONS AND
FAILS TO FUNCTION AS A
PRESERVATION MEDIUM.

SO PDF/A AND THE EARLIER
VERSIONS OF PDF ARE PRETTY
SIMPLE.

THEY HAVE LIMITED NUMBER OF
IMAGE FORMATS THEY SUPPORT.
A LIMITED NUMBER OF THINGS THEY
DO WITH TEXT AND SO THEIR
COMPLEXITY IS LOW AND THE TYPES
OF IMAGES INCLUDED IN THEM ARE
LIMITED TO A PRESERVEABLE SUBSET
OTHER VERSIONS OF PDF ARE --
INCLUDE MORE TYPES OF CONTENT
AND CONTENT IN -- THAT HAVE K
HAVE MORE COMPLEX INTERACTIONS.
PDF IS ALSO A FAIRLY OPEN
STANDARD SO MANUFACTURERS ADD
THEIR OWN PARTS TO IT.

PDF SOFTWARE LIKE THE PRO
VERSIONS OF ADOBE, ACROBAT, PDF
DISTILLERS WILL LET YOU INSPECT
FILES, UNDERSTAND THE PARTS OF
THEM.

IT'S IMPORTANT TO UNDERSTAND
THAT PDF IS A PRETTY BROAD
CATEGORY IN COMPUTING TERMS.
USUALLY WE'RE THINKING ABOUT
THIS FAIRLY SIMPLE SUBSET OF
PDFS WHICH IS BY AND LARGE THE
MAJORITY OF PDF THAT'S OUT
THERE.

>> GREAT, THANK YOU VERY MUCH.
I DON'T WANT TO HANG US UP TOO
LONG WITH THIS.

PEOPLE HAD A LOT OF QUESTIONS.
WE MAY COME BACK TO SOME OF THEM
AT THE END, I THINK.

>> YEAH, GREAT.

LET'S DO THE IMAGING POLL AND
THEN WE'LL MOVE INTO TALK

TALKING ABOUT DIGITAL IMAGE
DATA.

>>

OKAY, WE HAVE LOTS OF USE OF
IMAGING.

ALL RIGHT, SO TWO MAIN TYPES OF
IMAGES.

RASTER IMAGES WHICH WE'LL FOCUS
ON TODAY AND VECTOR IMAGES WHICH
WE WON'T TALK ABOUT.

RASTER IMAGES ARE A GRID OF
COLORED DOT WHICH IS WE CALL
PIXELS.

VECTOR GRAPHICS YOU SEE QUITE A
LOT.

SO MOST CLIP ART AND SHAPES YOU
PUT ON A POWERPOINT SLIDE ARE
VECTOR GRAPHICS AND THEY ARE
MATHEMATICALLY DEFINED.

SO VECTOR GRAPHICS ARE THINGS
LIKE DRAW A CIRCLE, FILL IT WITH
RED, GIVE IT A BLACK BORDER AND
MAKE IT -- GIVE IT THIS RADIUS.
SO YOU CAN CHANGE AND ALTER SIZE
AND CHARACTERISTICS OF VECTOR
GRAPHICS FAIRLY FREELY WHICH YOU
CAN'T -- SOMETHING YOU CAN'T DO
WITH RASTER GRAPHICS.

VECTOR GRAPHICS ARE ALMOST
ALWAYS BORN DIGITAL.

YOU'VE HEARD IT'S PRETTY
LABORIOUS TO CONVERT ANALOG
SOURCES INTO VECTOR GRAPHIC
EQUIVALENTS.

IT CAN BE DONE BUT IT'S A RARE
USE CASE SO WE WON'T SPEND TIME
ON IT TODAY.

FOR IMAGING, SCANNING
PHOTOGRAPHS, TAKING PICTURES OF
ART OR ARTIFACTS AND GETTING
DIGITAL PHOTOGRAPHY THE STANDARD
FORMAT, THE ONE THAT MOST PEOPLE
GO TO MOST OF THE TIME THEIFF,
THE TAG IMAGE FILE FORMAT.

THIS FORMAT HAS BEEN AROUND
SINCE THE 1980s, WAS DWOUNPED BY
A COMPANY CALLED ALBUS AND

TURNED OVER TO ADOBE.

I HAVE THEF HAS BEEN I HAVE TOF
HAS BEEN STABLE SINCE THE 1990s
WITH DECADES OF CONSISTENT USE
BEHIND IT.

IT'S USUALLY USED AS THE PLACE
TO STORE UNCOMPRESSED IMAGE DATA
JPEG 2001 ALSO DISCUSSED IN THE
COMMUNITY AS A NEW ALTERNATIVE
TO JPEG.

IT HAS A LOT TO RECOMMEND IT.
THERE'S A REAL SHORTAGE OF TOOLS
TO CREATE AND WORK WITH JPEG
2000 IMAGES.

TIFF IS THE SAFE EFFECTIVE BET
RIGHT NOW.

IN YOUR TIFF FILES OR YOUR JPEG
2000 FILES, THERE'S SOMETHING
WE'RE LOOKING FOR.

THEY SHOULD HAVE UNCOMPRESSED
IMAGE DATA.

SO TIFF AND JPEG 2000 ARE BOTH
ABLE TO SUPPORT COMPRESS IT HAD
DATA.

TIFF GIVES YOU THE OPTION OF
UNCOMPRESSED.

WE GENERALLY LOOK WHEN YOU'RE
DIGITIZING IMAGES TO GET 300 OR
MORE PIXELS PER INCH.

THAT'S AT READING DISTANCES OF
ABOUT A FOOT OR TWO.

THAT'S ABOUT THE POINT AT WHICH
THE HUMAN EYE CAN'T DISTINGUISH
THE INDIVIDUAL PIXELS, THE
INDIVIDUAL THOUGHTS AND IMAGE.

AND WE LOOK FOR 24-BIT COLOR.

THAT MEANS THERE ARE EIGHT BITS,
EIGHT POINTS OF DATA ASSIGNED TO
EACH OF THE COLOR CHANNELS, RED,
GREEN, AND BLUE THAT A COMPUTER
USES.

IF YOU HAVE MORE THAN 300 PIXELS
PER INCH-- AND MOST PEOPLE AIM
FOR 600-- THAT EFFECTIVELY GIVES
YOU THAT ABILITY TO ZOOM IN
WITHOUT SEEING PIXEL LAGS IN AN
IMAGE.

THE OTHER THING YOU NEED IS
COLOR CALIBRATED AND PROFILED.
TO CAPTURE GOOD IMAGES-- AND
THIS IS IMPORTANT IF YOU'RE
THINKING ABOUT HOW TO SET UP
YOUR OWN WORK FLOW AS WELL AS
WHAT TO LOOK FOR IN A RELIABLE
VENDOR'S WORK FLOW IS THAT
EQUIPMENT SHOULD BE SET UP,
PROFILED AND OTHER SOFTWARE
PROFILING SUITES THAT ARE USED
TO DO THIS AND THEN IT SHOULD BE
LEFT ALONE.

THERE SHOULD BE LITTLE
INTERVENTION.

THE MASTER FILE YOU CREATE, THE
PRESERVATION FILE YOU CREATE
FROM THAT SHOULD BE UNALTERED
IT SHOULD HAVE A COLOR PROFILE
ATTACHED.

AND WE'LL TALK ABOUT WHAT'S IN
THAT.

BUT IT SHOULDN'T BE SOMETHING
YOU COLOR CORRECT.

EDITING, RETOUCHING, COLOR
CORRECTION SHOULD ALWAYS BE DONE
ON A SECONDARY COPY FOR A
PARTICULAR USE.

SO YOU MIGHT CREATE A SECOND
COPY FOR PRINTING A POSTER OR
PRINTING A BILLBOARD OR PUTTING
SOMETHING ON THE WEB.

THE ORIGINAL FORMAT SHOULD BE
UNCORRECTED YOU CAN SEE RIGHT
HERE PROBABLY WHAT'S WRONG WITH
THIS IMAGE.

IT'S GOT A YELLOW AND BLUE CAST
TO IT.

I'M GOING TO SHOW YOU A COLORED
PANE IN JUST A MOMENT AND YOU'LL
SEE THERE'S A BLACK DOT IN THE
CENTER OF THE SCREEN.

AND I'M GOING ASK YOU TO LOOK AT
THAT BLACK DOT FOR A SECOND AND
THERE WILL BE A LITTLE ANIMATION
THAT COMES ALONG TO HELP YOU DO
THAT.

WE'RE GOING TO LOOK AT THIS FOR 20 OR 30 SECONDS AND JUST KEEP STARING IN THE MIDDLE OF THE SCREEN WITH RESPECT THE LITTLE SUN IS SPINNING AND BLINKING IN AND OUT AND TRY TO KEEP YOUR EYES FOCUSED IN THE VERY MIDDLE HERE.

AND HERE'S THE ORIGINAL IMAGE AGAIN.

YOU'RE PROBABLY SEEING IT AT FIRST LOOK APPROPRIATE, LOOK LIKE IT DOESN'T HAVE A COCOLOR CAST AND AS YOU LOOK AT IT MORE YOU SHOULD SEE THE SIDES FADING BACK TOWARD THAT YELLOW AND BLUE TINGE.

SEEING AN IMAGE BEFORE AND AFTER YOUR EYES HAVING EXPOSED TO THE CONTRASTING COLOR DO COLOR CORRECTION AUTOMATICALLY AND ONE OF THE MISTAKES PEOPLE MAKE IS TO TRY TO DO THAT COLOR DIRECTION TO THE DIGITAL IMAGE. THE COLOR OF THE SHIRT YOU'RE WEARING, THE CLOUDINESS OF THE SKY OUTSIDE WHETHER YOU HAD YOUR CUP OF COFFEE CAN AFFECT WHAT YOUR EYES ARE GOING TO SEE.

LET THE COMPUTER DO ITS JOB IN ANY NUMBER OF LABORATORY STUDIES OF PEOPLE TRYING TO COLOR CORRECT IMAGES AND ESPECIALLY PROFESSIONAL PHOTOGRAPHERS THE RESULT IS THAT THEY HAD NOISE TO THE FILE.

THEY DON'T GET CLOSER TO ACCURATE ORIGINAL COLOR.

THIS SLIDE IS SHOWING YOU -- PROVIDED YOU HAVE A 24-BIT MONITOR, ALL OF THE COLORS THAT CAN BE SHOWN IN A 24-BIT RGB FILE.

THERE ARE ABOUT 16.7 MILLION OF THEM.

THAT'S ENOUGH TO CAPTURE MOST OF THE VISIBLE SPECTRUM YOU CAN

PERCEIVE A DIFFERENCE FOR.
THIS SLIDE SHOWS YOU THE VISIBLE
SPECTRUM.

THE YELLOW ZONE SHOW WHAT IS'S
CALLED THE RGB GAMUT.

THOSE ARE THE COLORS THAT
COMPUTERS AND DIGITAL IMAGING
THE DEVICES CAN DISPLAY.

YOU'LL SEE IT DOESN'T OVERLOOP
WITH THE CMYK GAMUT.

NONE OF THOSE COVER THE ENTIRE
VISUAL SPECTRUM.

I THINK IF YOU LOOK CLOSELY AT
THIS, THOUGH, YOU'LL SEE THAT
MOST OF THE DRAMATIC COLOR
CHANGES FROM WHITE TO FAIRLY
PURE RED GREEN AND BLEW ARE
COVERED BY THAT GAMUT SO THE
THINGS THAT ARE EXCLUDED ARE
DIFFERENT BUT MINIMALLY SO SO
THE RGB GAMUT COVERS THE MOST
IMPORTANT PART OF THE VISUAL
SPECTRUM AND THE MOST IMPORTANT
PARTS OF THE SPECTRUM.

WHAT HAPPENS IN COLOR
CALIBRATION IS ESSENTIAL THEY
THE YELLOW ZONE, THE GAMUT
THAT'S CAPTURED MOVES SOMEWHERE
ELSE ON THE VISIBLE SPECTRUM THE
POINT OF CAL INAUGURATION
SOFTWARE IS TO MAKE SURE YOU'RE
IMAGING THE VISIBLE SPECTRUM
EVERY TIME YOU TAKE A PICTURE OR
MAKE A SCAN.

THE WAY THIS WORKS IS THAT A
TARGET IS SCANNED ON YOUR
COMPUTER DISPLAYED ON YOUR
MONITOR THEN A CROSSWALK IS
CREATED, THAT'S WHAT THE COLOR
PROFILE IS THAT THE SCANNER SEES
RED, A LITTLE BIT OFF FROM THE
WAY IT SHOULD AND IT MAKES THE
COLOR CORRECTION POSSIBLE.

THESE THINGS ARE ALL COVERED IN
GREATER DEPTH IN SOME OF THE
ASSOCIATED LINKS.

WHAT'S IMPORTANT FOR YOU TO KNOW

NOW IS THAT YOUR IMAGING DEVICES
AND YOUR VENDOR'S IMAGING
DEVICES GIVE YOU A WINDOW INTO
THE VISIBLE SPECTRUM AND IT'S
IMPORTANT THAT YOU AND THEY HAVE
A PLAN IN PLACE TO MAKE SURE
THAT THEY ALWAYS LOOK THROUGH
THE SAME WINDOW AT THE SAME
PORTION OF THE SPECTRUM.
OR THAT THEY ALWAYS GET COLOR
THE SAME WAY.

SO THAT YOU HAVE CONSISTENCY
FROM IMAGE TO IMAGE.

SO UNCOMPRESSED 24-BIT RGB, A
COLOR MANAGED PLAN THAT USUALLY
SHOWS UP IN AN ICC PROFILE,
YOU'LL BE LOOKING FOR TIFF
FORMAT AND YOU WANT TO GET MORE
THAN 300 DOTS PER IMAGE IN TERMS
OF VISUAL RESOLUTION.

SO WITH THAT, BEFORE WE MOVE ON
TO AUDIO, LET'S TAKE A LITTLE
TIME AND DO SOME IMAGING
QUESTIONS.

>> OKAY, ONCE AGAIN WE HAD A LOT
OF QUESTIONS COMING IN.

SO I -- AND SOME OF THESE YOU
TOUCHED ON BRIEFLY BUT OTHERS
WE'LL TRY TO GO OVER.

YOUR LAST -- THE LAST BIT THAT
THERE YOU TALKED ABOUT A 24-BIT
COLOR AND A NUMBER OF PEOPLE
WERE ASKING ABOUT WHETHER EVEN
BLACK AND WHITE DOCUMENTS SHOULD
BE SCANNED AS 24-BIT COLOR OR
WHETHER THERE ARE LOWER STEPS
YOU SHOULD BE USING AND
BASICALLY SOME OF THOSE
TECHNICAL DETAILS.

>> SURE, SO IT'S IMPORTANT TO
DIFFERENTIATE BETWEEN BLACK AND
WHITE AS WE THINK OF IT.

SAY, BLACK-AND-WHITE PHOTOGRAPHY
OR BLOCK PRINTING IN A BOOK AND
BLACK-AND-WHITE AS THE COMPUTER
UNDERSTANDS IT.

SO FOR A COMPUTER BLACK AND

WHITE MEANS ONE BIT.
ZERO OR ONE, BLACK OR WHITE, NO
EXCEPTIONS.

PRINTED BLACK-AND-WHITE
DOCUMENTS AND PHOTOS ARE NOT
BLACK AND WHITE.

THEY HAVE A WHOLE RANGE OF
TONALITYS IN BETWEEN THOSE IN
THE GRAY SPECTRUM.

THERE IS A GRAY SCALE COLOR
FORMATING THAT GIVES YOU GRAY
TONALITYS, MIXING OF BLACK AND
WHITE.

I GENERALLY DON'T RECOMMEND
THEM.

THERE MAY BE IN PARTICULAR USE
CASES WHERE THEY'RE VALUABLE SO
FOR CAPTURING SOURCE DOCUMENTS
GENERALLY EVEN WHAT WE THINK OF
AS BLACK AND WHITE IS COLOR OR
GRAY TO A CERTAIN EXTENT SO IN
GENERAL 24-BIT COLOR IS
RECOMMENDED AND STORAGE IS CHEAP
ENOUGH NOW AND DEVICES ARE
ROBUST ENOUGH THAT THAT'S A SAFE
RECOMMENDATION.

YOU MAY CERTAINLY IF YOU HAVE A
PHOTO ARCHIVE THAT'S ALL
METICULOUSLY CARED FOR
BLACK-AND-WHITE PHOTOGRAPHY, YOU
MAY BE FINE WITH GRAY SCALE
IMAGING.

LIKE WISE IF YOU'RE JUST DOING
TEXT IMAGING OF NEWSPAPERS SO
THAT YOU CAN DO OCR ON THEM YOU
MAY BE FINE WITH JUST A GRAY
SCALE IMAGING.

>> AND THERE WAS A SERIES OF
QUESTIONS THAT ARE RELATED TO
ISSUES OF FILE COMPRESSION.

>> YEAH.

>> SO SOME PEOPLE WERE ASKING
ABOUT IF YOU HAVE A JPEG IS IT
OKAY TO JUST CONVERT IT TO A
TIFF, IF YOU HAVE PDF IMAGES CAN
YOU TAKE THE PICTURES OUT OF
THEM.

WHAT ARE YOUR RECOMMENDATIONS?

>> SO LET ME START WITH THE PDF AND SAY I DON'T THINK I WOULD BOTHER.

SO THE PDF PICTURE ITSELF PROBABLY IS ENCODED AS A JPEG OR A TIFF.

SO PDF SPORTS BOTH OF THOSE FORMATS.

USUALLY BETTER TO LEAVE THINGS IN THEIR NATIVE FORMAT UNLESS THEIR NATIVE FORMAT IS HIGHLY PROBLEMATIC.

SO IF YOU GET A BORN DIGITAL J PEG PHOTOGRAPHY COLLECTION IT DOESN'T PRESENT A PARTICULAR PROBLEM FOR YOU, THE CHANCES OF NOT HAVING A J PEG DECODER DOWN THE ROAD ARE VERY LOW.

I WOULD GIVE YOU A DIFFERENT ANSWER FOR A PROPRIETARY COMPRESSED FORMAT.

SOMETHING LIKE PHOTO C.D. WOULD BE AN EXAMPLE WHERE YOU WOULD PROBABLY WANT TO SAVE THOSE PHOTOS INTO A DIFFERENT FORMAT AN UNCOMPRESSED TIFT FILE WOULD BE THE NATURAL CHOICE SO TIFT SO WITH BORN DIGITAL RESOURCES WE TAKE WHAT WE'RE GIVEN AND A LOT OF PHOTOGRAPHY COMES AS JPEG. SO THAT'S YOUR STARTING PLACE FOR A FORMAT.

>> GREAT, AND THAT BRINGS UP ANOTHER QUESTION SOME PEOPLE HAD WHICH IS THAT THEY'RE GETTING DIGITAL IMAGES AND IT SOUNDS LIKE THEY MAY GET SOME IN DNG OR RAW FORMAT?

>> THE DNG FORMAT HAS A REAL STRONG CASE TO BE MADE FOR IT. DNG IS ESSENTIALLY A RAPPER FOR RAW CAMERA DATA.

SO RAW DAY FOR FOR THOSE NOT FAMILIAR WITH IT IS JUST A READOUT OF ALL OF THE VOLTAGES THAT ARE PART OF A CAMERA'S

LIGHT IMAGE SENSOR.
SO IT READS THOSE AS A
PARTICULAR VOLTAGE AND THE
SOFTWARE TRANSLATES THOSE INTO
COLOR.

EACH MANAGER HAS A WAY OF
MANAGING THAT RAW DATA.
DIGITAL NATIVE IS SOMETHING THAT
ADOBE DEVELOPED TO CREATE A
PLATFORM INDEPENDENT WAY OF
SHARING RAW DATA.

SO THE DIGITAL NEGATIVE FORMAT,
THE DNG FORMAT IS A SMART WAY TO
PRESERVE BORN DIGITAL PHOTOGRAPH
THE RAW FORMAT HAS A LOT OF
RISKS BUILT INTO IT.

IF CANNON CHANGES ITS SOFTWARE
AND GOES OUT OF BUSINESS YOUR
RAW FILES GO OUT OF BUSINESS
WITH IT.

SO THE DNG IS A GREAT WAY TO GET
BORN DIGITAL PARAGRAPHY.

I DON'T KNOW THAT I WOULD
RECOMMEND IT FOR REFORMATTING
FORMATS.

I DON'T -- I HAVEN'T ENCOUNTERED
PEOPLE WHO ARE, FOR INSTANCE,
SCANNING AN IMAGE AND SAVING IT
AS A DNG INSTEAD OF A TIFF.
YOU COULD MAKE A GOOD ARGUMENT
FOR DOING THAT BUT IT HASN'T
BECOME THE CONSENSUS OF THE
FIELD YET.

THE DIGITAL NEGATIVE IS A VIABLE
FORMAT.

>> VERY GOOD.

>> I'M GOING TO MENTION THAT WE
DO HAVE SOME RESOURCES LINKED ON
THE COURSE PAGE AND SOME OF
THESE WILL ADDRESS SPECIFIC
QUESTIONS PEOPLE ARE HAVING
ABOUT DIFFERENT FORMATS AND
SCANNING STANDARDS.

SO EVERYONE IF YOU HAVE A CHANCE
TO LOOK THROUGH SOME OF THESE, I
DON'T KNOW IF YOU WANT TO SAY
ANYTHING ABOUT THOSE NOW.

>> SO YOU'LL SEE A NUMBER OF RESOURCES AT THE END OF MY PRESENTATION AND ON THE WEB SITE THAT TALK ABOUT THESE DIFFERENT FORMATS AND WALK YOU THROUGH HOW TO LOOK FOR A VENDOR, HOW TO SORT OF GO THROUGH THE DETAILS OF SETUP.

WE COULD HAVE AN EIGHT-HOUR WEBINAR ON HOW TO SET UP AND COLOR PROFILE A SCANNER.

SO WE -- RATHER THAN DO THAT WE -- IN THIS SERIES WE'RE TRYING TO GIVE YOU THE SMART QUESTIONS TO ASK AND POINT YOU TO THE RESOURCES THAT WILL HELP YOU DO THE STEP BY STEP DETAILS OF THAT PROCESS.

>> GREAT.

SO, AGAIN, I THINK WE'LL JUST HOLD US HERE AT THESE QUESTIONS AND PERHAPS -- WE'LL HOPE THAT WE HAVE TIME TO GET BACK TO SOME OF IT BUT LET'S GO ON AND TALK ABOUT AUDIO.

>> SOUNDS VERY GOOD SO YOU CAN GUESS THE POLL QUESTION. HOW DO YOU DIGITIZE AUDIO MATERIALS?

>>

YEAH, AND A LITTLE DIFFERENT MAKEUP HERE.

INTERESTING TO ME THAT HALF OR MORE HAVE OR PLAN TO DIGITIZE AUDIO MATERIALS I THINK EVEN A COUPLE YEARS AGO WE WOULDN'T HAVE SEEN THOSE NUMBERS IN A SERIES LIKE THIS.

IT'S BECOMING A MORE HEAVILY USED FORMAT FOR LIBRARIES AND ARCHIVES.

SO WE'RE GOING TO START AGAIN WITH SORT OF THE MAGIC WORDS AND THEN GET A LITTLE BIT INTO WHY THOSE THINGS MATTER AND HOW WE GO ABOUT CREATING THEM.

SO THE TONGUE TWISTER FOR THE

DAY FOR ME AND ALL OF YOU IS UNCOMPRESSED PULSE CODE MODULATION WHICH I CAN'T SAY TWICE EVEN VERY SLOWLY AND CONSEQUENTLY WE HAVE AN ACRONYM FOR IT, UNCOMPRESSED PCM. WE'RE LOOKING FOR UNCOMPRESSED DATA, THE RAW CAPTURE OF A DEVICE AND THIS UNCOMPRESSED PCM, THIS METHOD OF CAPTURING SOUND DATA ACTUALLY DATES ALL THE WAY BACK TO TELEGRAPH CODES AND EARLY TELEPHONIC SYSTEM. SO IT PREDATES THE COMPUTING ERA.

AND THE THINKING BEHIND THIS IS A VERY LONG LIVED OF REPRESENTING AUDIO DATA. THE FORMAT THAT'S GENERALLY USED FOR AUDIO PRESERVATION IS THE BROADCAST WAVE FILE.

B-WAVE WHICH IS JUST THE WAVE FILE THAT YOU MAY HAVE ENCOUNTERED ELSEWHERE IN YOUR LIFE WITH A METADATA HEADER ATTACHED TO IT.

AND I KNOW SOME AUDIO ENGINEERS WHO GIVE IT THE PITHY NAME OF CATASTROPHE METADATA.

IT'S FAIRLY SCANT.

IT USUALLY SAYS SOMETHING TO THE EFFECT OF THE TITLE OF THIS FILE IS THIS, ITS AUTHOR IS THAT AND IT'S THIS LONG.

IT'S ENOUGH THAT SHOULD OFF DISASTER AND LOSE YOUR OTHER METADATA YOU CAN AT LEAST DETERMINE WHAT IS THE BROADCAST WAVE FILES, ASIDE FROM THAT METADATA HEADER IT'S JUST A WAV FILE, AN UNCOMPRESSED AUDIO STREAM.

JUST LIKE IMAGES, THERE'S RESOLUTION AND BIT DEPTH TO PAY ATTENTION TO AN AUDIO SAMPLE. THE RESOLUTIONS WHICH THE NUMBER OF TIMES THE AUDIO IS SAMPLED IN

ANY GIVEN SECOND THE
RECOMMENDATION IS THAT AT LEAST
44.1 KILOHERTZ.

THAT'S 4411 THOUSAND TIMES PER
SECOND.

44,100 TIMES PER SECOND AND
THAT'S THE QUALITY THAT'S
RECORDED ON C.D.s.

MOST AUDIO ENGINEERS WHO ARE
WORKING IN PRESERVATION PREFER
96 KILN HERTZ, 96,000 SAMPLES
EVERY SECOND.

THE BIT DEPTH THAT'S RECOMMENDED
IS AT LEAST 16 BITS, THAT'S C.D.
QUALITY AND PREFERABLY 24 BITS.
TO YOU'LL SEE THESE PAIRED UP SO
A 44/16 RECORDING IS C.D.
QUALITY AUDIO, A 96/24 CORDING
IS THE PREFERRED PRESERVATION
AUDIO QUALITY.

THESE IMTWO IMAGES ARE INTENDED
TO START GIVING YOU A SENSE OF
WHY THESE THINGS MATTER.

SO THERE IS SOMETHING CALLED
QUANTITYIZATION WHERE THEY'RE
TRYING TO MAKE SURE YOUR
SAMPLING RATE CATCHES ALL THE
DIFFERENT PARTS OF THE SIGNAL.
AND SO YOU CAN SEE THE BLUE LINE
HERE REPRESENTS THE ORIGINAL
SIGNAL, THE ACTUAL AUDIO SIGNAL.
AND THE RED JAGGED LINE SHOWS
WHERE SAMPLES ARE TAKEN.

AND THIS SHOWS WHAT HAPPENS WITH
A AN ERROR.

SO BECAUSE OF THE POINTS AT
WHICH SAMPLES ARE TAKEN, IF YOU
LOOK AT THE IMAGE BELOW YOU SEE
THE RESULT OF SOME CLIPPINGS, SO
THE PEAK OF THAT FIRST UPSWING
IS CLIPPED OFF AND THE TROUGH AT
.5 IS LOPED AWAY.

AND YOU END UP MISSING PARTS OF
THAT SIGNAL.

ONE OF THE WAYS TO AVOID THAT IS
TO SAMPLE MORE FREQUENTLY AND
ALSO TO SAMPLE WITH GREATER BIT

DEPTH.

SO A MORE FREQUENT SAMPLING WILL CATCH MORE OF THE CHANGES IN THE AUDIO FILE.

A WIDER SAMPLING TOP TO BOTTOM FROM 16 TO 24 BIT LETS YOU CATCH MORE GRADATIONS OF THE FILE.

SO YOU'LL SEE IN THIS NEXT SLIDE WHERE PARTICULARLY IN THIS EXAMPLE SOUND WAS LOPED OFF.

4416 IS ALMOST THE LIMIT OF HUMAN HEARING.

9624 CAPTURES ANYTHING YOU AND KY HEAR AS WELL AS THINGS DOGS AND BIRDS WOULD ENJOY LISTENING TO. 9624 IS INTENDED REALLY TO GIVE US SOME HEAD ROOM TO MAKE SURE WE'VE MISSED NONE OF THE SAMPLE FOR MOST OF THE AUDIO WE WORK YOU MAY HEAR OH, WELL IT'S JUST A VOICE RECORDING BUT SOUND ENGINEERS WOULD ARGUE THE OPPOSITE IS TRUE.

BECAUSE WE HAVE EVOLVED TO LISTEN TO THE HUMAN VOICE, YOU NEED MUCH MORE PHI DAHLTY RECORDING THAN LISTENING TO AN ORCHESTRA.

9624 HAS COME -- HAD VERY STRONG CASES MADE FOR IT BY THE PRESERVATION ENGINEERING COMMUNITY.

ALSO IMPORTANT TO REALIZE THAT IN MOST RECORDINGS WE HAVE SEVERAL CHANNELS SO MONO AUDIO JUST HAS ONE CHANNEL.

MOST OF THE AUDIO WE LISTEN TO NOW DAYS-- THIS COULD BE TRUE FOR EVEN THIS WEBINAR-- IS STEREO.

THE AUDIO IS BROKEN INTO A RIGHT AND LEFT CHANNEL.

SO EACH ONE OF THOSE CHANNELS WOULD BE 96 KILOHERTZ, 24 BITS PER SECOND.

THIS IS WHY AUDIO FILES START TO GET FAIRLY LARGE COMPARED TO

TEXT AND IMAGE FILES.

IT'S ALSO INCREASINGLY COMMON, ESPECIALLY, FOR INSTANCE, MOTION PICTURE SOUND TRACKS TO HAVE 5.1 WHICH IS ACTUALLY SIX CHANNELS CENTER FRONT RIGHT AND LEFT STEREO, LOWER BASS TRACK, SIX CHANNELS OF AUDIO.

IN ADDITION TO THOSE WAVE FORMS, IT WERE UNCOMPRESSED WAV FORMS OF THE AUDIO FILE, THERE MAY BE MET DA METADATA IN BROADCAST FORMS RELATIVELY SMALL COMPARED TO THE AUDIO AND, AGAIN, SEE THE AUDIO 44.1 MOST DIGITAL PRESERVATION IS 96 KILOHERTZ TO GET MORE ACCURATE REPRODUCTION AND HELP AVOID ERRORS.

THIS BIT DEPTH, THE C.D. AUDIO 16 BIT PROVIDES 6565,536 LEVELS OF AMPLITUDE.

THAT'S THE AM PLU TUESDAY FROM ZERO, THE BOTTOM OF THE GRAPH, TO THE TOP OF IT.

THAT LETS YOU RANGE FROM ZERO TO 96 DECIBELS.

THAT'S MEASURING THE INTENSITY OF THE SOUND.

ROCK CONCERTS GET UP INTO THE HUNDREDS OF -- 110, 120 DECIBELS.

AND T REASON YOU CAN STILL HEAR ROCK MUSIC ON A C.D. IS BECAUSE VERY RARELY DO THEY EVER GET TO ZERO.

SO LIKE THAT RGB GAMUT, AN AUDIO ENGINEER PLACES THE SAMPLING ON THE ACTUAL HEARD PART OF THE SPECTRUM.

SO VERY FEW C.D.s CAN MAINTAIN ZERO DECIBELS WHICH IS SOMETHING CLOSE TO ABSOLUTE ZERO WHEN THINGS JUST DON'T VIBRATE AT ALL.

24 BIT AUDIO HAS A THEORETICAL MAXIMUM OF 16.7 MILLION LEVELS FROM ZERO TO 144 DECIBELS SO

THAT'S FROM ABSOLUTE ZERO, NO VIBRATION AT ALL WELL BEYOND THE LIMITS OF HUMAN TOLERANCE AT THE TOP.

IT'S ACTUALLY MORE AUDIO DATA THAN CURRENT CIRCUITS ALLOW US TO TRANSMIT.

MOST CIRCUITS, EVEN THOUGH BEST ONES ALLOW MAYBE 100, 120 DECIBELS OF AMPLITUDE TO BE RECORDED.

THAT SAID, SOME PEOPLE ARE RECORDING FOR PURPOSES BEYOND HUMAN HEARING SO ANYONE WHO'S TUNING IN FROM THE CORNELL ORNITHOLOGY LAB OR IS WORKING ON WHALE SONG, THERE'S SOME INFORMATION BEYOND THIS PEREZ STATION AND YOU MAY NEED TO ACQUIRE IT.

REALLY IN CAPTURING ANYTHING THERE ARE THREE COMPONENTS. ONE IS THE SOURCE, RIGHT? SO FOR AUDIO REFORMATTING THIS IS MAGNETIC TAPE, IT COULD BE L.P. OR QUITE OFTEN IT'S A MICROPHONE.

THAT IS ANALOGOUS TO THE PHOTO OR DOCUMENT FOR THE BOOKS YOU WOULD BE SCANNING ON IMAGING WORK FLOW THE TRANSFORMATION HAPPENS WITH WHAT'S CALLED A DIGITAL AUDIO CONVERTER AND YOU'LL SEE TWO TYPES OF THESE ANALOG-TO-DIGITAL CONVERTERS, A.D.C.s AND DIGITAL TO ANALOG CONVERTERS.

SO THE DIGITAL AUDIO CONVERTER IS WHAT DETERMINES THE BIT DEPTH AND RESOLUTION AND BASIC QUALITY OF YOUR CAPTURE.

SO A BETTER DIGITAL AUDIO CONVERTER HAS LESS NOISE, SAMPLES MORE ACCURATELY AND LIMITS THE AMOUNT OF AMPLITUDE AND FREQUENTLY OF SAMPLING THAT YOU CAN MANAGE.

THIS IS WHERE YOU WANT TO QUIZ YOUR VENDORS ON WHAT KIND OF D ARE YOU USING, WHY DID YOU CHOOSE THEM?

HOW MUCH ACTUAL DATA DO THEY CAPTURE.

THIS IS ALSO WHERE YOU AS A PERSON WHO MIGHT BE SETTING UP AN AUDIO REFORMATTING WORK WANT TO SPEND A LOT OF TIME COMPARING MANUFACTURERS AND READING SPECIFICATIONS THIS IS THE EQUIVALENT OF THE SCANNER OR CAMERA IN YOUR DIGITAL IMAGING WORK FLOWS.

AT THE TAIL END IS AN AUDIO MASTERING OR EDITING SOFTWARE SIMILAR TO THE IMAGE EDITING SOFTWARE YOU MIGHT USE, ADOBE PHOTO SHOP OR WHAT HAVE YOU. YOU NEED SOME SORT OF SOFTWARE TO MANAGE AND CREATE AUDIO FILES AFTER THEY'RE CAPTURED IN FROM THAT DAC.

SO, AGAIN, JUST TO REVIEW OUR KEY SPECKS, THE BROADCAST WAV FILE WHICH IS A WAVE FILE WITH A METADATA HEADER IS THE BEST FORMAT FOR STORING DIGITAL AUDIO AT PRESENT.

THE WAV AUDIO HAS AN UNCOMPRESSED PCM AUDIO CHANNEL, ONE PER CHANNEL OF THE FILE. THAT'S SORT OF OUR UNIVERSAL FORMAT FOR UNCOMPRESSED AUDIO. WE WANT TO GET RESOLUTION OF AT LEAST 44.1 KILOHERTZ, PREFERABLY MORE, 96 KILOHERTZ AND A BIT DECHT OF 16 BITS BUT PREFERABLY 24.

SO WITH THAT WE'LL PAUSE BEFORE WE TALK ABOUT OTHER DATA FORMATS AND TAKE QUESTIONS.

>>

>> WELL, WE HAD SOME MORE REALLY GOOD QUESTIONS COMING IN. SOME OF THEM WERE TECHNICAL

QUESTIONS AND SOME OF THEM WERE MORE GENERAL SORTS OF QUESTIONS ABOUT AUDIO MATERIALS.

OR ONE OF THE QUESTIONS THAT I THINK WE SHOULD START WITH, IT WAS JUST A JARGON QUESTION. IS THERE A SIMPLE DEFINITION OF SAMPLING?

>> SAMPLING IS ACTUALLY IDENTICAL TO WHAT YOU DO WHEN YOU SCAN AN IMAGE. SO FOR EVERY SECOND OF THE AUDIO FILE YOU RECORD THE AMPLITUDE OF THAT WAVE.

SO AUDIO FILE COMES TO YOU AS A WAV FORM, A SERIES OF VIBRATIONS.

AND EACH ONE OF THOSE HAS AN AMPLITUDE, ESSENTIALLY A HEIGHT. IT'S ZERO DECIBELS OR 96 DECIBELS ALL THE OR TEN DECIBELS TALL AND YOU SAMPLE BY RECORDING THOSE HEIGHTS OF WITHAV FORM A NUMBER OF TIMES PER SECOND.

THAT'S THE FREQUENCY ASPECT. AND SO A 44.1 C.D. RECORDING AS 44,100 SAMPLES IN A GIVEN SECOND IT SAYS.

THE WAV FORMS 17, NOW 18, NOW 19 NOW 13, NOW 12, NOW 9, NOW 23. RAPIDLY.

44,000 TIMES PER SECOND AND SO WHAT A DIGITAL AUDIO CONVERTER DOES IS MEASURES THE INTENSITY OF THAT WAV SEVERAL THOUSAND TIMES PER SECOND AND STHERZ AS AN ADDITIONAL FILE.

>> GREAT.

THERE WAS A QUESTION ABOUT A WAV VERSUS V WAV, BROADCAST WAVE. AND OBVIOUSLY FILE EXTENSIONS ARE DIFFERENT BUT OTHERWISE WHAT ARE THE DIFFERENCES?

>> SO YOU CAN THINK OF -- IT'S NOT TERRIBLY INING A TROOT SAY THAT BWAV IS A WRAPPER FOR WAV FILES.

SO THEY BOTH HAVE THE SAME
ESSENTIAL DATA INSIDE THEM.
THESE UNCOMPRESSED PCM AUDIO
SAMPLES.

WHAT WAV DOES IS ADD METADATA
TO THAT THAT TELLS YOU USEFUL
THINGS BOTH DESCRIPTIVELY-- THE
TITLE OF THE TRACK FOR
INSTANCE-- AS WELL AS
TECHNICALLY.

THAT THE TRACK IS X MINUTES LONG
OR CONTAIN THIS IS MANY
MEGABYTES OF DATA.

AND SO THE BROADCAST PART OF
THAT IS METADATA THAT'S ADDED TO
WITHAV FILE AND IT COMES FROM
BROADCASTING.

IT'S THE SORT OF THING THAT AS A
RADIO STATION YOU MIGHT WANT TO
KNOW TITLE, AUTHOR, AIRTIME.

AND SO WAV ADDS SOME OF THOSE
KEY PIECES OF MET A DATA TO THE
FILE OTHERWISE THE SOUND SAMPLES
ARE IDENTICAL BETWEEN THE TWO.

>> OKAY.

AND THE GENERAL EXTENSION FOR
BROADCAST WAV BY THE WAY IS BWS.

>> YEAH, SORRY, I THINK I --

>> I WAS JUST -- BECAUSE I THINK
MY RESPONSE TO ONE OF THE
QUESTIONS WAS WRONG.

ONE QUESTION THAT CAME UP,
PEOPLE ARE ASKING ABOUT HOW TO
CONVERT DIFFERENT TYPES OF AUDIO
SOURCES TO A MORE PRESERVATION
AND FROM 44 KILOHERTZ TO 96 OR
16 BITS TO 25.

>> THIS IS SIMILAR IN SOME WAYS
TO THE QUESTION ABOUT WHAT IF I
GET A JPEG FILE AS MY -- IF IT'S
BORN DIGITAL.

UPCONVERTING DOES YOU NO GOOD.
SO IF YOU TAKE A C.D. AND
CONVERT IT TO 96/24 YOU JUST
MAKE IT BIGGER.

THERE'S NO WAY TO ADD BACK THAT
DATA.

THIS IS ONE OF THE REASONS THAT PRESERVATIONS YOU THINK NEARS STRESS DOING A VERY HIGH RESOLUTION CAPTURE.

YOU CAN GO FROM 96/24 DOWN TO 44/16 TO PRODUCE A SMALLER BUT STILL VISIBLE FILE FOR USE IN THE READING ROOM.

YOU CAN'T GO THE OTHER WAY.

NOW, YOU MAY GET DATA INTO YOUR COLLECTION THAT'S RECORDED IN MP 3 FORMAT OR AS A C.D.

YOU JUST KEEP WHAT YOU'VE GOT BUT YOU CAN'T ADD DATA BACK IN. WHAT HAPPENS WITH AN MP 3 FILE IS LIKE WHAT HAPPENS WITH THE J JPEG FILE.

SO THERE'S AN ALGORITHM, A WAY OF GOING MATH THAT EXPANDS IT TO GUESS AT WHAT THE FULL DATA OUGHT TO LOOK LIKE.

AND WHEN YOU GET BORN DIGITAL FILES YOU KEEP WHAT YOU'RE GIVEN AND IF IT'S A PROPRIETARY FORMAT YOU MIGHT WANT TO CONVERT IT TO A WAV FILE BUT THERE'S NOTHING GAINED BY CONVERTING IT TO A LARGER WAV FILE THAN IT IS TO BEGIN WITH.

>> GREAT, I THINK WE'LL HOLD THE REST OF THE QUESTIONS.

THERE ARE A FEW LINKS THAT HAVE BEEN POSTED IN VARIOUS OTHER THINGS.

THEY'RE GREAT.

BUT WE WANT TO MAKE SURE JACOB CAN GET THROUGH ALL OF HIS MATERIAL TODAY.

>> GREAT, SO WE'LL TOUCH JUST BRIEFLY ON VIDEO AND MOVING IMAGES AND SOME OTHER THINGS. SO YOU CAN GUESS THE POLL QUESTION.

HAVE YOU DIGITIZED YOUR IMAGE MATERIALS?

>>

OKAY.

SO SIMILAR TO AUDIO WE'RE ACTUALLY TALKING ABOUT TWO DIFFERENT SOURCES EVEN THOUGH WE END UP IN THE SAME PLACE WITH DIGITAL.

MOTION PICTURES ARE ALMOST ALWAYS RECORDED OPTICALLY ON THE FILM.

VIDEO IS A SERIES OF MAGNETICALLY RECORDED SIGNALS THAT ARE WAV FORMS SIMILAR TO AUDIO THAT DEPICT IMAGES AND SOUND.

WAV.

VIDEO HAS A SPECIFIC REVOLUTION FOR EACH FRAME THAT PROVIDES A FIXED NUMBER OF SCAN LINES.

SO IF YOU START GETTING INVOLVED IN VIDEO YOU'LL START SEEING THINGS LIKE 720 X 480I60.

WHICH MEANS 486 LINES FROM TOP TO BOTTOM.

THAT'S ISECAM STANDARD THAT WE'RE USED TO HERE IN NORTH AMERICA AND 720 PIXELS ACROSS.

THE I60 MEANS INTERLACED, WHICH WE'LL TOUCH ON IN A MOMENT.

AND NOTABLY THERE ARE SIX LINES FOR METADATA IN THAT FORMAT AND IT'S GRAPHICAL METADATA NOT TEXT WALL METADATA SO THERE ARE THINGS THAT TELL THE ENGINEER ABOUT THE VIDEO FILE.

AND WHAT'S TREMENDOUSLY CLEVER AND TO US NOW HAVING TO NOT HAVING TO REFORMAT THE MATERIALS LATER ON TREMENDOUSLY IRRITATING.

THIS IS WHAT ACTUALLY EXISTS IN A MAGNETIC VIDEO FILE.

SO IT'S A SERIES OF WAV FILES THE FIRST PART, HORIZONTAL BLANKING AND THE SYNC TIP HELPS MAKE SURE VIDEO FRAMES ARE CREATED AT THE RIGHT RATE.

YOU'LL SEE A GRAY BOX THAT HAS A BLOWUP DOWN BELOW AND THAT'S THE COLOR BURST.

IT'S A LOW RESOLUTION COLOR WASH
THAT IS THEN GIVEN DEFINITION BY
THE ACTIVE VIDEO PORTION.

THAT ACTIVE VIDEO IS ESSENTIALLY
LIKE A BLACK-AND-WHITE RENDERING
OF THE IMAGE AND WHEN THE COVER
IS OVERLAID ON THAT
BLACK-AND-WHITE REBDERING YOU
GET A CRISP COLOR IMAGE.

THIS IS SIMILAR TO THE WAY
PRINTING IS DONE WHEN YOU DO
COLOR LITHOGRAPHY, ACTUALLY.
BUT THIS IS JUST A SERIES OF
WAYS TO DEPICT HIGH AND LOW IN
DIFFERENT HUES AND SATURATION OF
COLOR.

SO MAGNETIC VIDEO IS A WAV AND
WHEN WE DIGITIZE IT WE TURN IT
INTO A RASTER FRAME, WE TURN IT
INTO A PICTURE WHEN YOU PAINT
VIDEO IN THE SCREEN IN THE
ANALOG WORLD YOU PAINT
HORIZONTALLY ACROSS THE SCREEN
FROM LEFT TO RIGHT BY SENDING A
BEAM OF ELECTRONS VERY QUICKLY,
FASTER THAN THE EYE CAN PERCEIVE
IT.

YOU ESSENTIALLY FREEZE THAT
FINAL IMAGE WHEN YOU DIGITIZE
IT.

SO VIDEO CONVERSION IS A
DRAMATIC CHANGE FROM VIDEO TO
DIGITAL.

FROM MOTION PICTURE TO DIGITAL
IS ESSENTIALLY GOING FROM ONE
WAY OF REPRESENTING A STILL
IMAGE TO ANOTHER WAY OF
REPRESENTING A STILL IMAGE.

IN BOTH AREAS, STANDARDS AND
PRACTICES ARE STILL DEVELOPING.
UNUNCOMPRESSED DIGITAL VIDEO
DATA IS DESIRABLE FOR THE USUAL
REASONS.

THIS IS SOMETHING THAT REALLY
BUMPS UP AGAINST THE PROBLEM OF
STORAGE COSTS.

THE -- FOR TEXT AND IMAGES THE

STORAGE HAS BECOME FAIRLY
INEXPENSIVE IT'S NOT AS
SIGNIFICANT AN OBSTACLE.
FOR SOUND IT CAN BE AN OBSTACLE
CERTAINLY.
SOUND CAN TAKE UP A LOT OF DATA.
FOR VIDEO AND MOVING IMAGE IT
ALMOST CERTAINLY IS AN OBSTACLE
BECAUSE OF THE COST OF STORAGE
AND THE SIZE OF THESE FILES.
COMPRESSION IS PRETTY NORMAL IN
VIDEO.
IT MAY CAUSE PRESERVATION
PROBLEMS.
THIS IS AN AREA WHERE THE MATH
HAS JUST NOT QUITE WORKED ITSELF
OUT YET.
STORAGE ISN'T CHEAP ENOUGH.
FILES ARE FAIRLY BIG AND FORMATS
ARE NOT AT A PLACE WHERE WE FEEL
THEY'VE STABILIZED.
THE CURRENT SORT OF SAFE BET IS
UNCOME PERES PRESSED .AVI BUT
THERE ARE PLENTY OF COMPETING
FORMATS.
JPEG 2000 AND M PEG 21.
THE QUICK TIME .MOV CORPS FORMAT
SO THERE ARE LOTS OF COMPETING
FORMATS.
UP WITH THING IS THAT IS
POSITIVE IS THAT THIS CODEX HAS
BECOME THE STANDARD WAY.
H.264 IS SUITABLE FOR WEB
DELIVERY BUT PART OF THE GOOD
NEWS IS EVEN IF YOU HAVE A
MASSIVE VIDEO ARCHIVE IN MPEG21
AND ENDS UP BEING A DEAD END AS
A PRESERVATION MASTER FORMAT, IF
YOU'VE RECREATED H.264
DELIVERABLES EVEN IF YOU HAVE TO
REFORMAT YOUR MASTERS YOU WON'T
HAVE TO RECREATE NEW DELIVERABLE
FILES FROM THOSE.
THE H.264 LOOKS LIKE IT'S HERE
TO STAY.
UNFORTUNATELY THE BOTTOM LINE
HERE IS THAT A AT THE END OF THE

DAY YOU PICK A VIDEO FORMAT THAT MAKES SENSE FOR YOU AND YOU EXPECT THAT YOU HAVE TO DO A MIGRATION.

IT COULD BE FIVE, TEN YEARS FROM NOW SO THAT'S THE NATIVE VIDEO FORMAT IN THE HAITI MILLION 5 SPECIFICATION AND IT'S OFTEN THE FORMAT USED IN FLASH VIDEO THERE ARE OTHER OPTIONS, QUICK TIME, REAL, WINDOWS MEDIA.

I WOULD NOT RECOMMEND ANY OF THEM.

IF YOU'RE DOING VIDEO DELIVERY NOW.

DO H.264.

THAT'S THE SAFEST BET.

SO YOU CAN CERTAINLY PICK A PROPRIETARY OPTION BUT A MIGRATION IS LIKELY IN YOUR FUTURE.

I ALSO JUST BRIEFLY BEFORE WE DO FINAL QUESTIONS WANT TO GIVE A NOD TOWARDS INTERACT ACTIVITY. WHEN YOU'RE DEALING WITH DATA I THINK SOME USEFUL QUESTIONS ARE TO DECIDE IF YOU ARE TRYING TO STORE FIXED POINTS IN TIME.

THAT IS ARE YOU TRYING TO STORE THE CONTENTS OF THE DATABASE OR A DATA COLLECTION AT A PARTICULAR POINT IN TIME IN WHICH CASE YOU CAN RENDER THINGS OUT TO THE EQUIVALENT OF AN XL FILE OR AN XML FILE.

ARE YOU PLANNING TO KEEP COLLECTING AND SHARING DATA THAT A LIVING SYSTEM IN WHICH CASE YOU HAVE A SYSTEMS MAINTENANCE QUESTION TO ANSWER.

SOMETIMES YOU MAY BE DOING BOTH. SOME EXAMPLES OF HOW THIS AFFECTS YOU, PROBABLY ALL OF YOU WOULD BE E-MAIL AND SOCIAL MEDIA.

SO E-MAIL IS A WIDELY KNOWN BUT LOOSELY DEFINED SET OF STANDARDS

AND THE WAY PEOPLE INTERACT WITH E-MAIL IS TIGHTLY COUPLED TO THE PROGRAM THEY USE.

YOU CAN READ THE SAME E-MAIL MESSAGE IN OUTLOOK AND THUNDERBIRD AND APPLE MAIL AND ON YOUR CELL PHONE AND SEE A DIFFERENT VERSION OF IT DO DIFFERENT THINGS WITH IT IN EACH INSTANCE.

EVEN THOUGH THE UNDERLYING DATA WILL BE THE SAME.

LIKE WISE, SOCIAL MEDIA, THINGS LIKE FACEBOOK AND TWIT CHER ARE WHAT SOME PEOPLE CALL CLOSED BUT FREE SO IT DOESN'T COST YOU ANYTHING USE THEM BUT THERE'S NO PROVISION TO MOVE DATA TO OTHER SYSTEMS, EITHER.

THIS IS AS YOU CAN IMAGINE BECOMING A VERY IMPORTANT ISSUE IN ARCHIVING AND LOTS OF RESEARCH, DISCUSSION AND HAND WRINGING IS GOING INTO THIS SPACE.

I'VE INCLUDED A COUPLE LINKS FOR WHERE TO LEARN MORE AND WHERE TO KEEP TRACK OF ISSUES ATTACHED TO THIS LIKE VIDEO THIS IS AN AREA WHERE IF YOU CAN WAIT A LITTLE WHILE BEFORE GETTING INVOLVED IT BEHOOVES YOU TO DO SO.

IF YOU HAVE TO BE INVOLVED RIGHT NOW YOU BUILD A CASE FOR THE WORK YOU'RE DOING AND EXPECT THE WINDS OF TECHNOLOGY MAY BLOW IN A DIFFERENT DIRECTION SO THIS APPEARS ON THE WEB SITE AS WELL. AND I OFFER TO PUT A PAGE ON MY OWN PERSONAL WEB SITE JUST JACOBNADAL.COM/342.

THAT'S A PERSISTENT URL.

THERE'S A CONTACT FORM YOU CAN USE TO SEND ME A MESSAGE AND IF THERE ARE BIG UPDATES OR THING TO POST WE CAN HAVE THEM THERE MOST WILL BE ON THE CONNECTING

TO COLLECTIONS WEB SITE.

SO WITH THAT LET'S DO WRAPUP
QUESTIONS.

>> OKAY.

SO WE ARE WE ARE GETTING CLOSE
TO THE END OF TODAY'S SESSION
AND DON'T HAVE A LOT OF TIME FOR
THINGS.

SOME OF THIS I'M SORRY WE WON'T
BE ABLE TO GET TO.

ISSUES ABOUT CODEX AND SOME OF
THE PROBLEMS YOU MAY RUN INTO
WITH VIDEO THAT YOU CAN'T READ
OR PLAY BACK?

>> YEAH, SO, AGAIN, WE TOUCHED
LIGHTLY ON VIDEO.

VIDEO IS -- EACH VIDEO CO-DECK
IS SOMETHING LIKE WHAT WE TALKED
ABOUT WITH UTF-8.

SO THERE'S VIDEO, THE IDEA, AND
THEN THERE'S CO-DECK, THE WAY OF
WRITING THAT DATA TO A DISK SO
FOR VIDEO OFTEN TIMES THE TOP
LEVEL FILE FORMAT THAT YOU SEE
.WMV OR .MOV OR .AV ZISHGS A
WRAPPER AND INSIDE THAT WRAPPER
IT'S SOMETHING LIKE HERE'S SOME
VIDEO DATA AND YOU SHOULD DECODE
THIS VIDEO DATA USING THE H.264
CO-DECK AND YOU NEED THESE VIDEO
WRAPER TO HOLD THE PIECES
TOGETHER, THE VIDEO AND AUDIO
DATA ITSELF AS WELL AS THOSE
CO-DECKS TO INTERPRET THEM.

IF YOU THINK ABOUT WHAT GOES
INTO A MOTION PICTURE THERE'S A
SERIES OF IMAGES THAT YOU WATCH
AS FRAMES THAT SHOW THE VISUAL
PART OF THE MOTION PICTURE.

THERE ARE BETWEEN ONE AND SIX
AUDIO FILES.

THERE MAY BE SUBTITLES AND
CAPTIONS.

AND SO AT THE TOP LEVEL VIDEO
FORMAT DOES, LIKE THE PDF FORMAT
FOR EXAMPLE, OR AS BY WAY OF
COMPARISON IS TO SHOW HOW THOSE

THINGS RELATE TO ONE ANOTHER AND FOR EACH ONE OF THOSE IT SAYS RENDER THE VIDEO USING THIS SET OF TOOLS UNDER THE AUDIO USING THIS OTHER SET OF TOOLS, DISPLAY THE CLOSED CAPTIONING DATA IN THIS WAY.

AND SO THE VIDEO FILES ARE THEMSELVES SORT OF VERY COMPLEX MULTIMEDIA OBJECTS.

>> WELL, THANK YOU VERY MUCH FOR ALL OF THAT INFORMATION.

WE'RE GETTING THANKS AND CONGRATULATIONS AND WHAT NOT.

I WANT TO TURN IT BACK OVER TO KRISTIN SO SHE CAN GIVE YOU SOME FINAL INSTRUCTIONS

>> THANKS SO MUCH, YOU GUYS, THAT WAS USEFUL AND HELPFUL. IT'S A LOT TO COVER IN A SHORT AMOUNT OF TIME SO I THINK EVERYONE WILL FIND YOUR LINKS HELPFUL.

PEOPLE HAVE BEEN SHARING GREAT LINKS IN THE CHAT AND WE'LL GET THOSE UP AND WE JUST WANT TO JUST QUICKLY GIVE YOU THE LINK TO THE HOME WORK ASSIGNMENT, ALSO LINK THROUGH-TO-THROUGH THE COURSE PAGE AND JAKE HAD JUST ASKED A QUESTION ABOUT FIRST OF ALL WHAT ARE THEY -- THE FACTORS THAT YOU WOULD NEED TO CONSIDER. BASED ON -- THERE'S SOME SUSTAINABLE FACTORS, YOU TALKED ABOUT HOW DIFFERENT FORMATS ARE SORT OF BEEN DETERMINED TO BE MORE STABLE AND THERE ARE OTHER THINGS THAT ARE STILL TO BE DETERMINED BUT THERE'S A GREAT LIBRARY OF CONGRESS WEB SITE YOU'LL BE ASKED TO LOOK AT AND TO LIST THREE POTENTIAL FACTORS YOU MIGHT NEED TO CONSIDER WHEN CHOOSING A FORMAT FOR DEPENDING ON WHAT ITEMS YOU HAVE IN YOUR INSTITUTION.

THEN A LITTLE CHANCE FOR YOU TO EXPRESS YOUR OWN OPINION.

THIS IS, AS YOU KNOW, AN EVOLVING SCIENCE AND THERE'S A LOT OF DECISIONS YOU HAVE TO MAKE SO WE'VE ASKED FOR YOUR OPINION ON SOMETHING.

SO NO WRONG ANSWER BUT INTERESTED TO HEAR WHAT YOU HAVE TO SAY.

AND WE'VE ASKED FOR YOU TO LET US KNOW IF YOU'RE WATCHING WITH COLLEAGUES TODAY, IF YOU WATCHED BY YOURSELF DON'T WORRY.

WE SAW WHEN YOU LOGGED IN AND YOU WILL BE TAKEN AS PRESENT BUT IF YOU WATCHED WITH COLLEAGUES LET US KNOW SO WE CAN COUNT THAT.

WE HAD OVER 400 PARTICIPANTS TODAY, WHICH IS OUTSTANDING. AND WE REALLY THANK EVERYONE FOR THEIR TIME.

WE'RE JUST AT 3:30, DANIELLE, DID YOU WANT TO ASK ANY MORE OUTSTANDING QUESTIONS OR WE'LL TRY TO MAKE GET THEM ADDRESSED AM N SOME OTHER MANNER.

>> JAKE, CAN YOU STICK ON FOR A LITTLE BIT LONGER?

IN NEW YORK, DROP BY BROOKLYN CIRCLE FOR A TIME.

>> SO WE HAD A QUESTION ABOUT CAD DATA.

>> CAD IS A PROBLEM FOR THE DIGITAL PRESERVATION COMMUNITY. OBVIOUSLY IT'S CRITICALLY IMPORTANT EVERYTHING WE WALK ON, SIT ON, DRIVE, FLY IN EVERYDAY HAS A A CAD FILE SOMEWHERE IN ITS PAST.

CAD FILE IN PRINCIPLE IS PRESERVEABLE, SO CAD FILE SAYS THINGS LIKE THERE ARE TWO LINES THAT CONNECT AT A CERTAIN POINT AND THEY GO OFF IN THESE DIFFERENT DIRECTIONS AND THEY

SHOULD BE MADE OF MAGNESIUM.
THE ISSUE WE ARE GRAPPLING WITH
IS THAT MOST CAD FORMATS ARE
PROPRIETARY SO THEY'RE PRODUCED
BY A CERTAIN MANUFACTURER AS
PART OF THEIR SOFTWARE PACKAGE
AND THEY'RE BOUND PRETTY TIGHTLY
BY I.P. RESTRICTIONS.

SO THIS IS AN AREA WHERE I THINK
YOU NEED TO BE ALERT.

AND RUN WITH THE HERD, RIGHT?
SO ANYTHING PRODUCED BY THE SORT
OF AUTOCAD, AUTODESK SERIES OF
FAMILIES HAS A LARGE USER BASE
AND A LOT OF INVESTMENT IN IT.
YOU'RE LIKELY TO KNOW QUICKLY
WHEN CHANGES ARE COMING AND HAVE
PLENTY OF TIME AND TOOLS TO
SUPPORT YOU IN MIGRATION BUT YOU
ARE GOING TO BE TIGHTLY COUPLED
TO A PROPRIETARY MANUFACTURER.
I DON'T SEE OR KNOW THAT THERE
ARE REALLY COMPELLING KIND OF
FREE AND OPEN ALTERNATIVES AT
THIS POINT.

BUT THIS IS SOMETHING THAT
THERE'S A LOT OF DISCUSSION
ABOUT IN DIGITAL PRESERVATION
COMMUNITIES.

I THINK IN SOME WAYS THE BEST
ANSWER RIGHT NOW IS TO STICK
WITH THE DE FACTO STANDARD AND
MAKE SURE YOU'RE WORKING -- THAT
YOU'RE NOT USING SORT OF A SMALL
MARKET LOW SUPPORT CAD PROGRAMS,
THAT YOU'RE USING WELL-SUPPORTED
PROGRAMS THAT HAVE A BIG USER
BASE.

>> GREAT.

WELL, I THINK PROBABLY IT'S BEST
THAT SOME OF THE REMAINING
QUESTIONS WE DIDN'T GET TO IS TO
GO AHEAD AND JUST POST SOMETHING
LATER ON AND JUST TYPE OF SOME
OF THE RESPONSES AND POST THAT.
SO PLEASE CHECK BACK ON THE WEB
SITE AND WE'LL TRY TO GET THOSE

ANSWERS UP.
BUT THANK YOU ALL.

>>

>> THANK YOU, SO MUCH, DANIELLE
AND TO JACOB, REALLY APPRECIATE
THIS GREAT CONTENT WE HAD TODAY
AND AS DANIELLE SAID, KEEP
CHECKING THAT WEB SITE.
YOU'LL BE GETTING AN E-MAIL FROM
US LATER TODAY WITH A RECORDING
OF TODAY'S WEBINARS AND LINK TO
THE WEB SITE AND OTHER
REMINDERS.

SO WE LOOK FORWARD TO SEEING
EVERYBODY BACK HERE NEXT TUESDAY
2:00 P.M. EASTERN TIME AND THAT
IS APRIL 9.

WE'LL BE HEARING FROM DANIELLE
ABOUT METADATA WHICH CAME UP
TODAY.

SO I KNOW IT WILL BE GREAT.
THANKS, EVERYONE, AND HAVE A
GREAT REST OF THE DAY.

>>