Family History Technology

Family and Church History Interim Digital Preservation Process

Rick Laxman

Family & Church History Department

IMAGING/PRESERVATION PROCESS

Images captured by digital cameras

→ Metadata converted to XML

→ Image and metadata files renamed

→ Preservation images converted to JPEG 20000 Lossless

TIFF bitmap compared to JPEG2000 bitmap

→ MD5 hash created for each image and metadata file

→ Preservation images and metadata written to DVD *AND* DLT tape

→ MD5 hash run on preservation DVD files and compared to original MD5 hash

Sample of DVDs run through analyzer—results saved as CSV for future comparison

→ Sample DVDs re-run through analyzer on a periodic basis & compared to original results to determine refresh needs

→ Test DLT tapes periodically

Family History Technology

© 2005, Intellectual

IMAGING LAB COMPUTER APPLICATIONS

- AMS
- Image Processing Utility (IPU)
- Audit
- Make Copy

Family History Technology

© 2005, Intellectual

IMAGE LAB COMPUTER APPLICATIONS

- AMS
  - Project setup
  - Process tracking
  - Assigns GS numbers
  - Records quality results
  - Generates production reports through Access and retake requests via email
  - Records media content for preservation DVDs and DLT tapes

Family History Technology

© 2005, Intellectual

IMAGE LAB COMPUTER APPLICATIONS

- Image Processing Utility (IPU)
  - Loads folders with images and log files (metadata) to a server
  - Creates the donor copy, preservation copy, and distribution copy of each image
  - Compares original image to preservation image
  - Renames images and folders
  - Creates a hash signature of all files in all folders
  - Converts all metadata files to the XML format

Family History Technology

© 2005, Intellectual

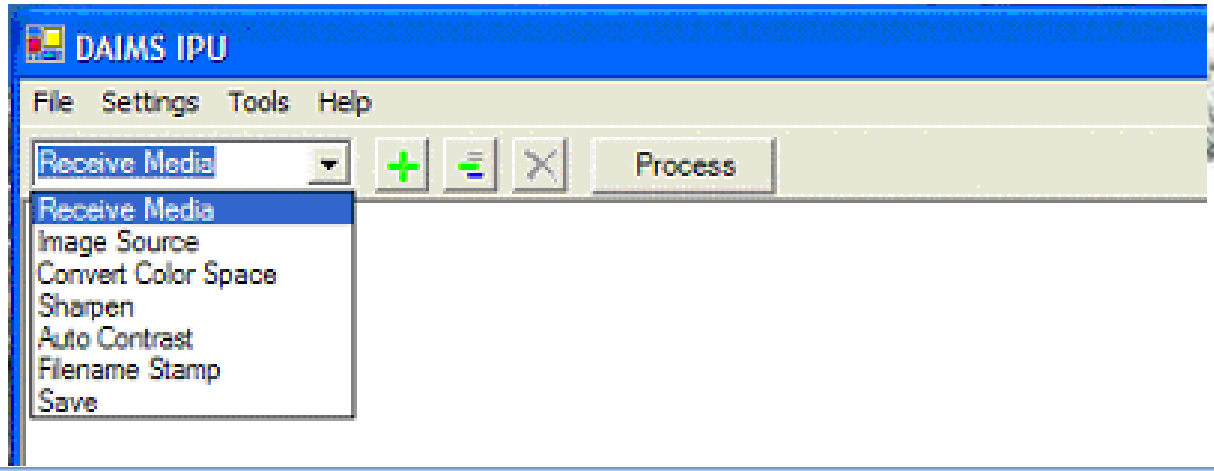
Digital Capture in Archives

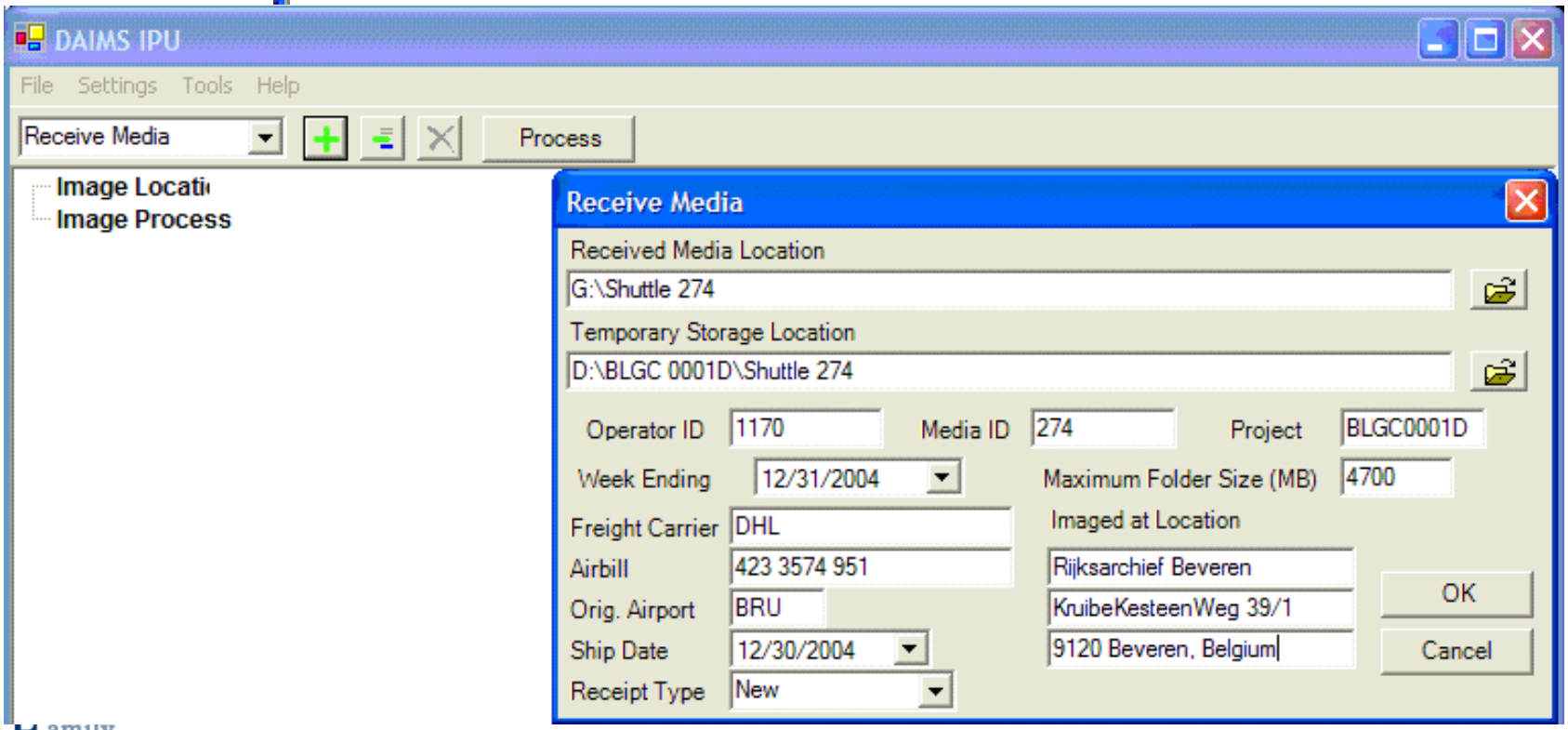
© 2005, Intellectual

IPU—Receive Transfer Media

© 2005, Intellectual

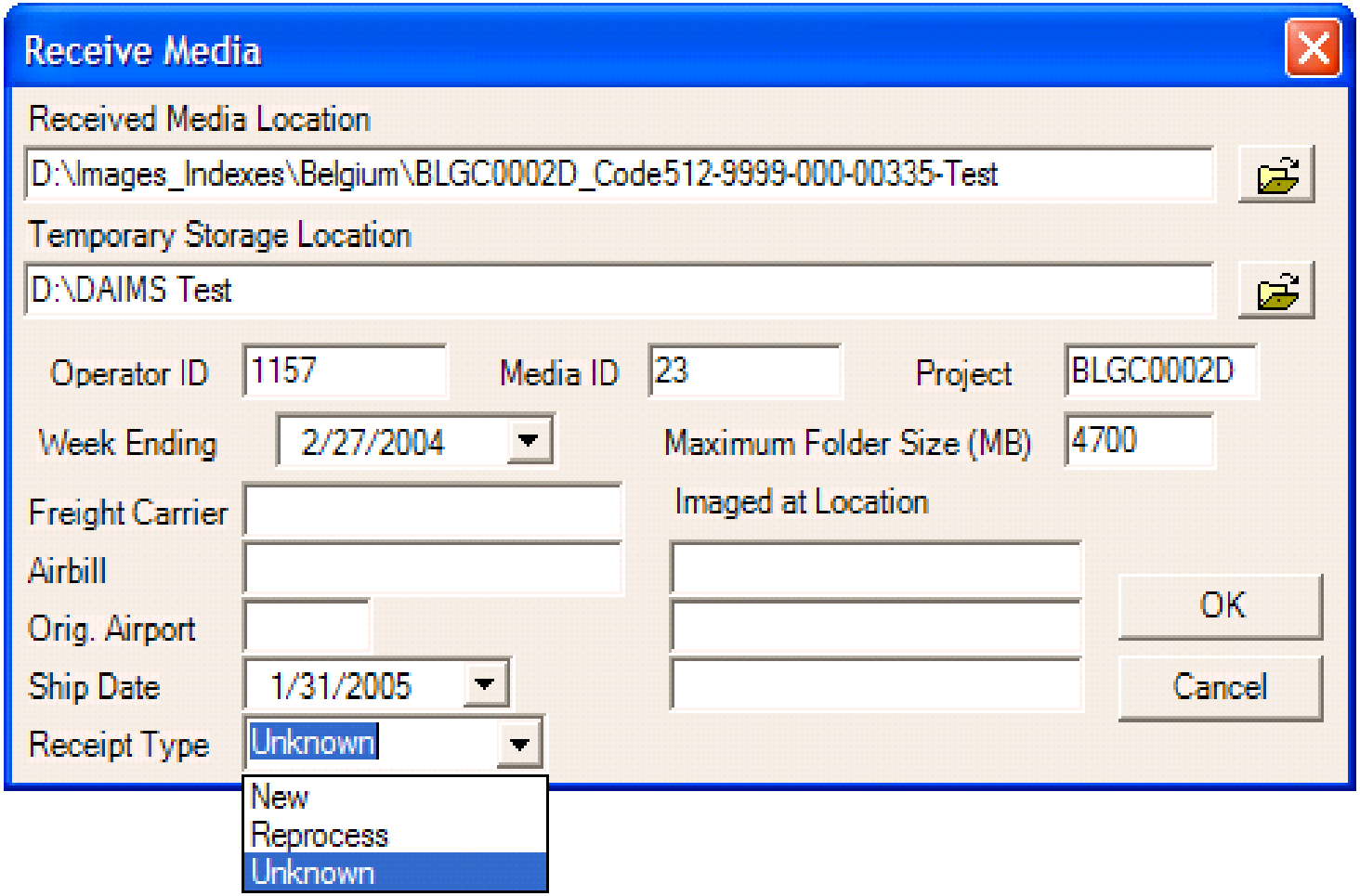
IPU





© 2005, Intellectual

IPU—Receive Media



© 2005, Intellectual

IPU

DAIMS IPU

File Settings Tools Help

Save Process

- ReceiveMedia
  - Image Source = D:\Images_Indexes\Belgium\BLGC0002D_Code512-9999-000-00335-Test
  - Files To Process = *.tif
  - Include Sub Folders = False
  - Temporary Staging Directory = D:\DAIMS Test
  - Operator = 1157
  - Week Ending = 1/28/2005
  - Project = BLGC0002D
  - Media ID = 23
  - FreightCarrier =
  - Airbill =
  - OriginatingAirport =
  - ShippedDate = 1/31/2005
  - ImagedAtLocation1 = Rijksarchief Beveren
  - ImagedAtLocation2 = Kruibekesteenweg 39/1
  - ImagedAtLocation3 = 9120 Beveren, Belgium
  - Split Large Folders = Yes
  - Maximum Folder Size = 4700
  - Receipt Type = New
- **Image Process**
  - Save
    - Location = PRESERVATION
    - Location Relative to Source = No
    - Format = JPEG 2000 Lossless
    - Compression = 0
    - CompressBasedOnSize = No
    - Rename = Yes
    - Verify Lossless = Yes
  - AutoContrast
    - Pixel Significance = 0.15%
  - Sharpen
    - Sharpen Amount = 800
  - Save
    - Location = DISTRIBUTION
    - Location Relative to Source = No
    - Format = JPEG
    - Compression = 60
    - CompressBasedOnSize = No
    - Rename = Yes
    - Verify Lossless = No

Family History Technology

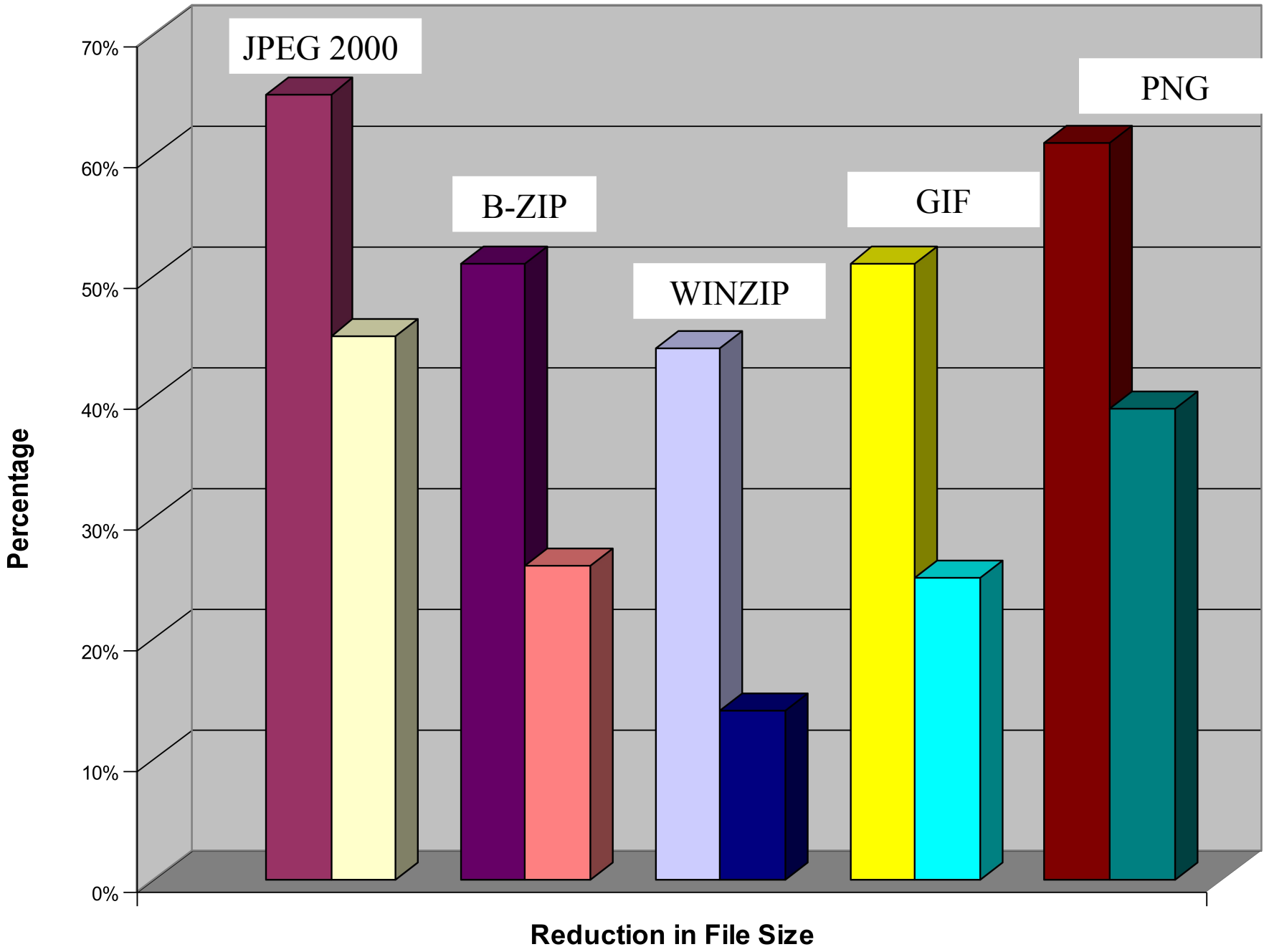
Compression Testing

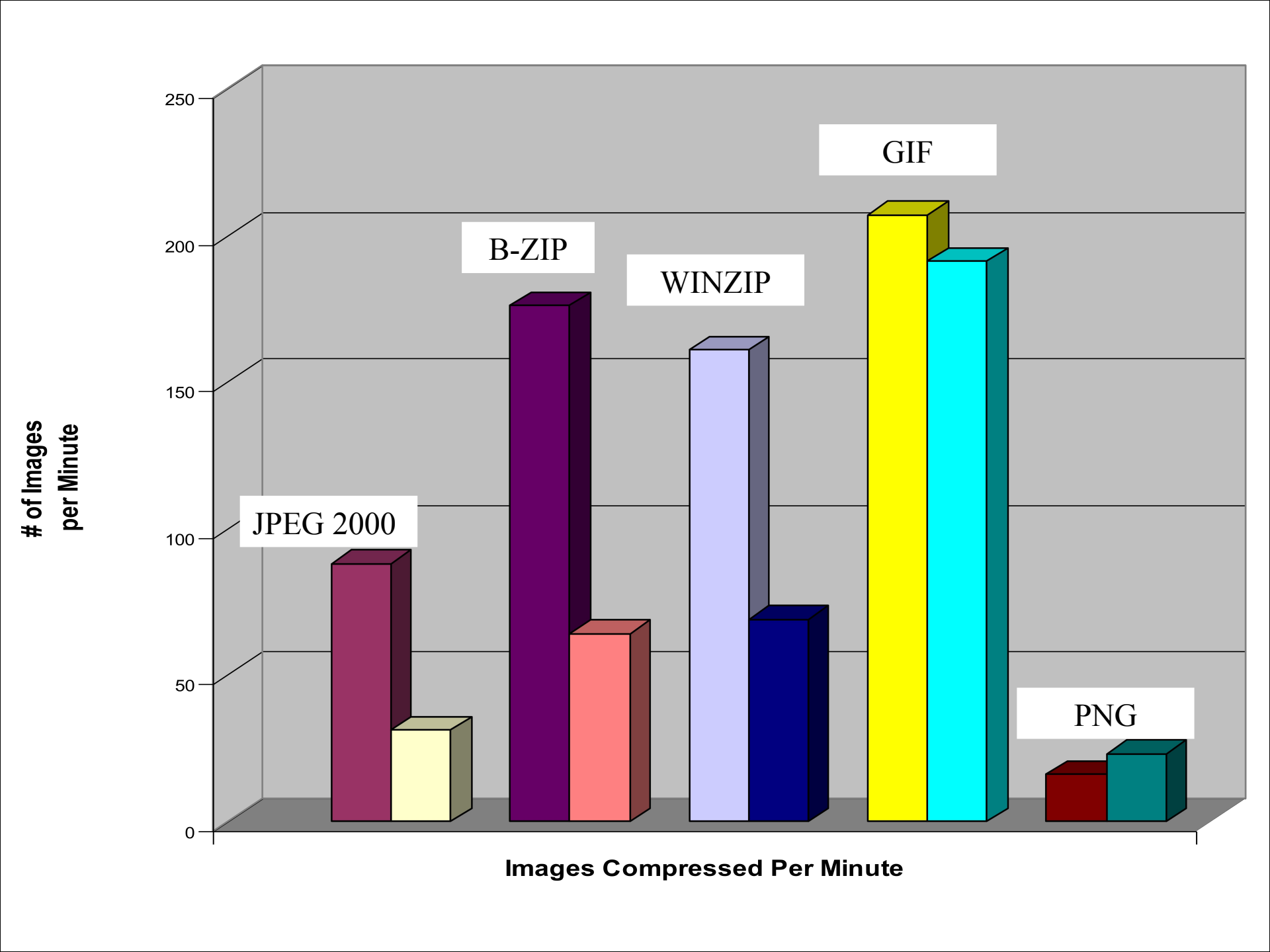
Source	Byte Size	Time to Write in Minutes	Compressed % of Original	Reduction in File Size	Images Compressed Per Minute
J2K/ USING IPU – 200 DPI Microfilm Scanning	142,112,830	510	35%	65%	88
Digital Camera Capture	54,574,760	714	55%	45%	31
PNG/ USING IPU – 200 DPI Microfilm Scanning	201,785,603	510	49%	51%	176
Digital Camera Capture	72,469,653	346	74%	26%	64
GIF/ USING IPU – 200 DPI Microfilm Scanning	228,861,373	510	56%	44%	161
Digital Camera Capture	84,565,129	335	86%	14%	69
WINZIP/ USING IPU – 200 DPI Microfilm Scanning	200,134,094	510	49%	51%	207
Digital Camera Capture	73,467,078	247	75%	25%	191
B-ZIP/ USING IPU – 200 DPI Microfilm Scanning	65,595,354	510	39%	61%	16
Digital Camera Capture	59,956,671	1,030	61%	39%	23

Image Compression Testing—Microfilm Scanning & Digital Camera Images[1]

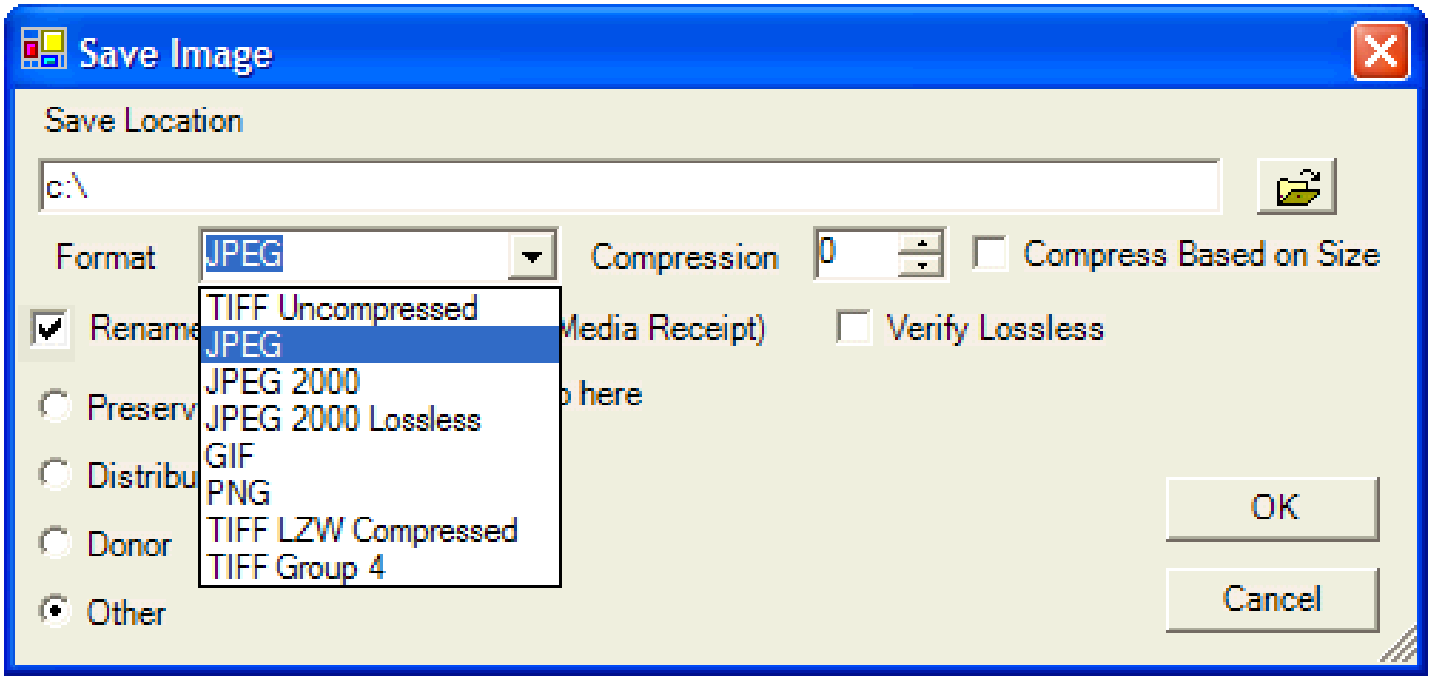
History Technology

© 2005, Intellectual


JPEG 2000
B-ZIP
WINZIP
GIF
PNG
70%
60%
50%
40%
30%
20%
10%
0%
Percentage
Reduction in File Size

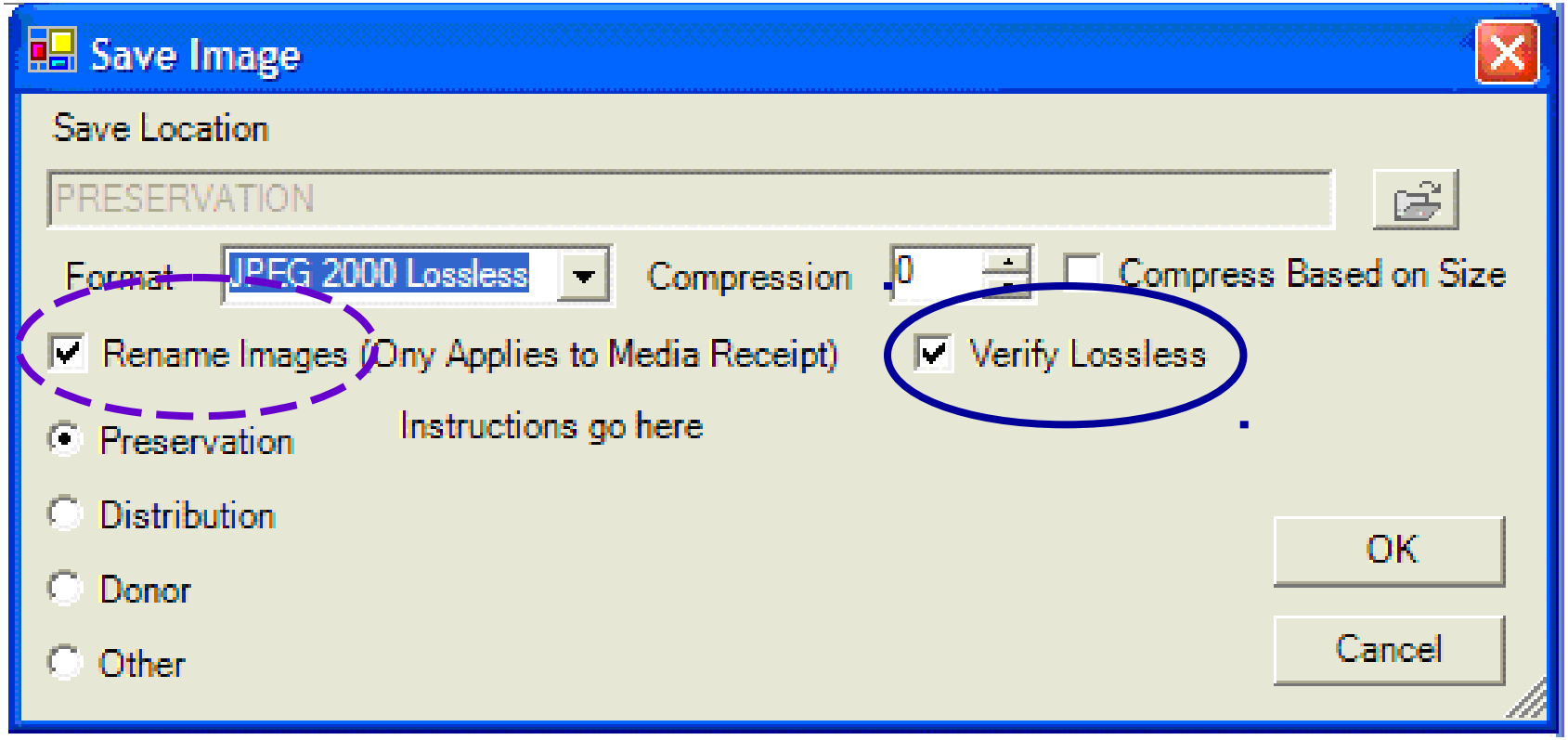

250
200
150
100
50
0
# of Images per Minute
JPEG 2000
B-ZIP
WINZIP
GIF
PNG
Images Compressed Per Minute

IPU—Output File Formats



© 2005, Intellectual

IPU—Verify Lossless (Compare TIF & JPEG 2000)



© 2005, Intellectual

IPU—dCam Log File (Metadata)

GA--060AD_Death Certificates-1990-File-Number-007041.log - Notepad

File Edit Format View Help

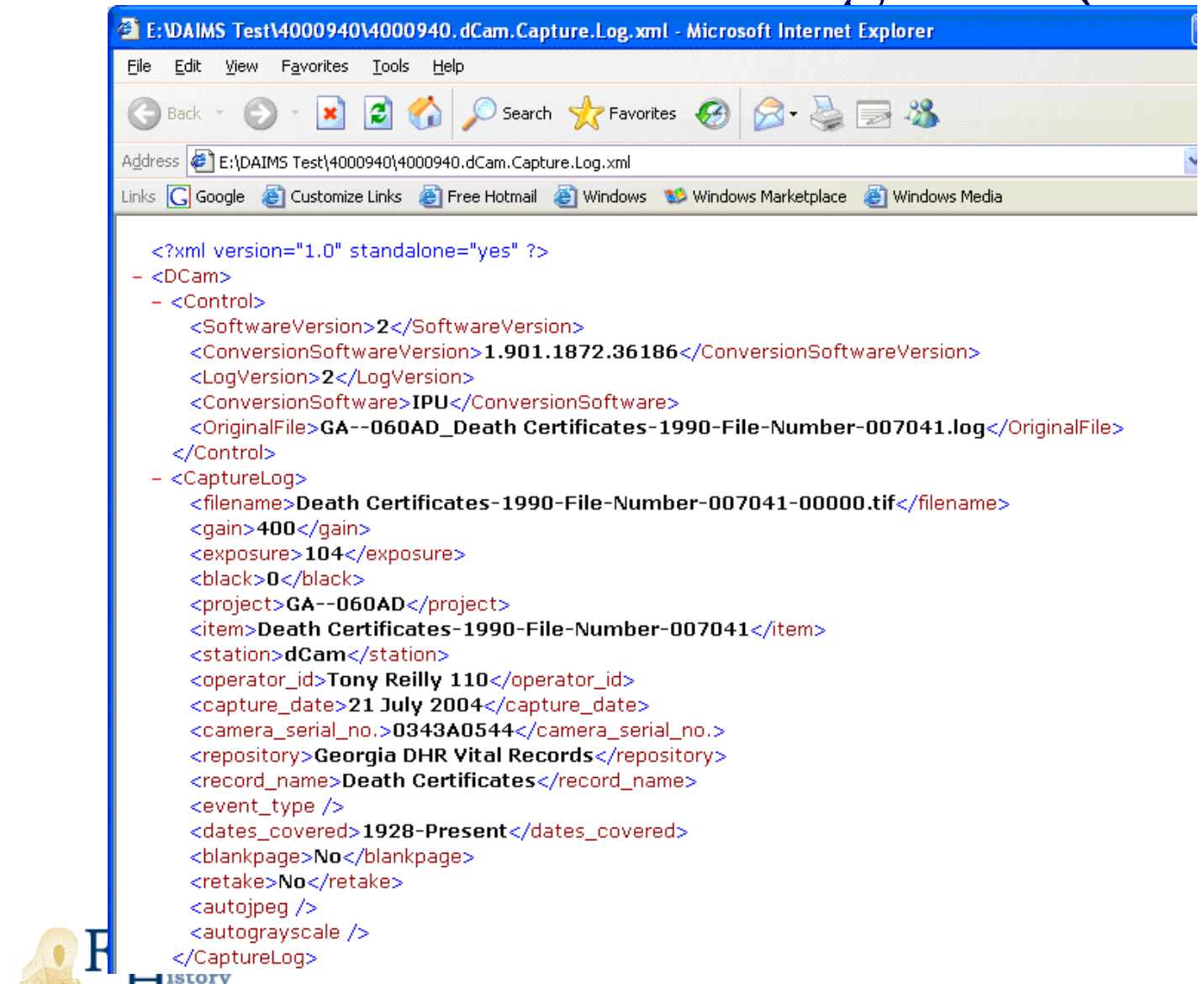
```
"FileName","Gain","Exposure","Black","Project","Item","Station","Operator ID","Capture date","Camera Serial
No.","Repository","Record Name","Event Type","Dates Covered","BlankPage","Retake","AutoJPEG","AutoGrayscale"
"Death Certificates-1990-File-Number-007041-00000.tif","400","104","0","GA--060AD","Death
Certificates-1990-File-Number-007041","dCam","Tony Reilly 110","21 July 2004","0343A0544","Georgia DHR Vital
Records","Death Certificates","","1928-Present","No","No","",""

Log settings changed: New Column header follows.
"FileName","Gain","Exposure","Black","Project","Item","Station","Operator ID","Capture date","Camera Serial
No.","Repository","Record Name","Event Type","Dates Covered","BlankPage","Retake","AutoJPEG","AutoGrayscale"
"Death Certificates-1990-File-Number-007041-00001.tif","400","104","0","GA--060AD","Death
Certificates-1990-File-Number-007041","dCam","Tony Reilly 110","21 July 2004","0343A0544","Georgia DHR Vital
Records","Death Certificates","","1928-Present","No","No","",""
"Death Certificates-1990-File-Number-007041-00001.tif","400","104","0","GA--060AD","Death
Certificates-1990-File-Number-007041","dCam","Tony Reilly 110","21 July 2004","0343A0544","Georgia DHR Vital
Records","Death Certificates","","1928-Present","No","Yes","",""
"Death Certificates-1990-File-Number-007041-00002.tif","400","104","0","GA--060AD","Death
Certificates-1990-File-Number-007041","dCam","Tony Reilly 110","21 July 2004","0343A0544","Georgia DHR Vital
Records","Death Certificates","","1928-Present","No","No","",""
"Death Certificates-1990-File-Number-007041-00003.tif","400","104","0","GA--060AD","Death
Certificates-1990-File-Number-007041","dCam","Tony Reilly 110","21 July 2004","0343A0544","Georgia DHR Vital
Records","Death Certificates","","1928-Present","No","No","",""
"Death Certificates-1990-File-Number-007041-00004.tif","400","104","0","GA--060AD","Death
Certificates-1990-File-Number-007041","dCam","Tony Reilly 110","21 July 2004","0343A0544","Georgia DHR Vital
Records","Death Certificates","","1928-Present","No","No","",""
"Death Certificates-1990-File-Number-007041-00005.tif","400","104","0","GA--060AD","Death
Certificates-1990-File-Number-007041","dCam","Tony Reilly 110","21 July 2004","0343A0544","Georgia DHR Vital
Records","Death Certificates","","1928-Present","No","No","",""
"Death Certificates-1990-File-Number-007041-00006.tif","400","104","0","GA--060AD","Death
Certificates-1990-File-Number-007041","dCam","Tony Reilly 110","21 July 2004","0343A0544","Georgia DHR Vital
Records","Death Certificates","","1928-Present","No","No","",""
```

Family History Technology

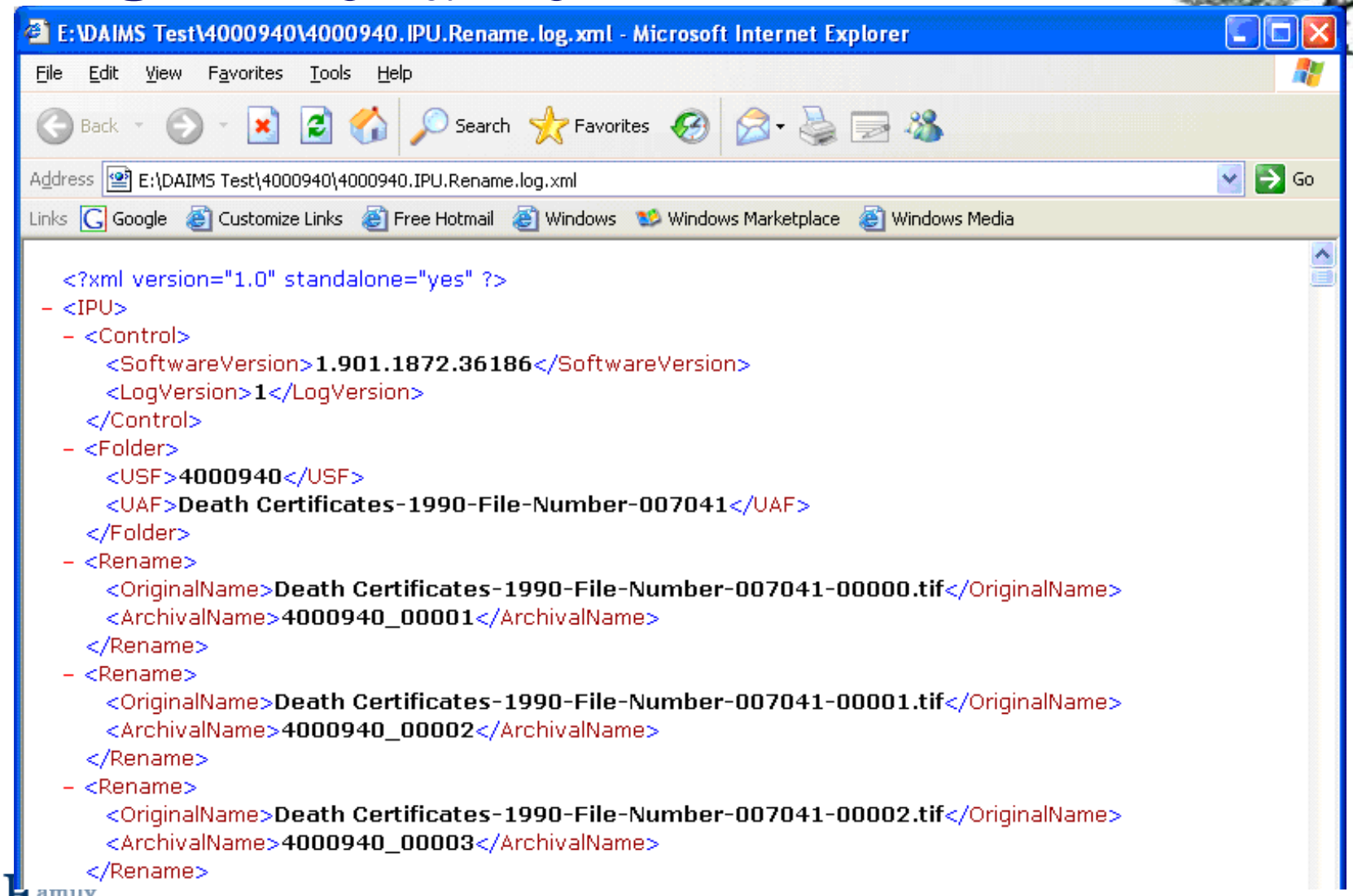
© 2005, Intellectual

IPU—XML dCam Log File (converted)



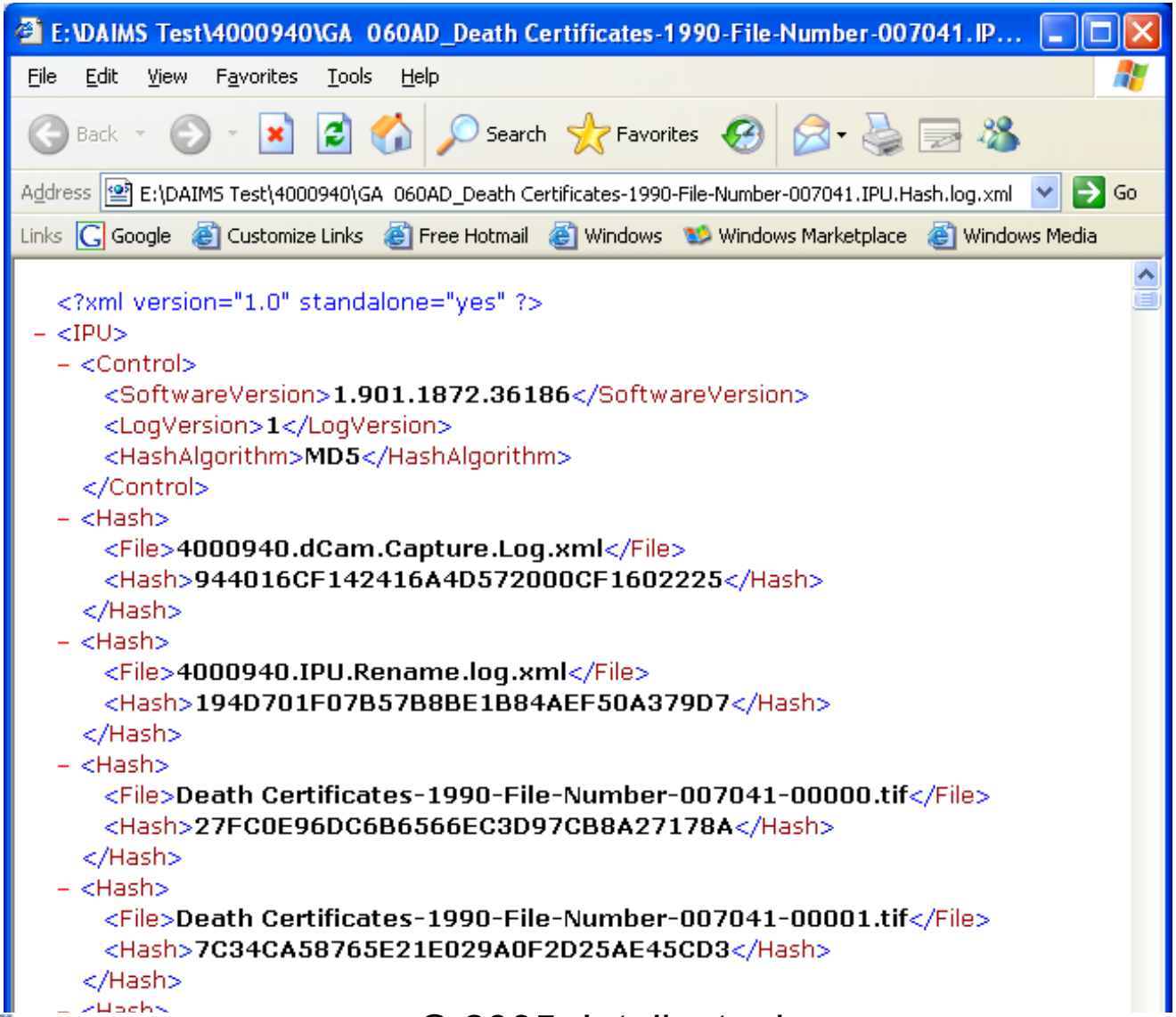
© 2005, Intellectual

IPU—Rename



© 2005, Intellectual

IPU—Hash



© 2005, Intellectual

HASH FORMATS

Algorithm	Message Size (bits)	Block Size (bits)	Word Size (bits)	Message Digest Size (bits)	Security (bits)
MD5		512	32	128	60
SHA-1	< 264	512	32	160	80
SHA-256	< 264	512	32	256	128
SHA-384	< 2128	1024	64	384	192
SHA-512	< 2128	1024	64	512	256

Table 1: Secure Hash Algorithm Properties [1]
Note: MD5 information added by the author of this paper[2]

[1] National Institute of Standards and Technology. *Secure Hash Signature Standard (SHS) (FIPS PUB 180-2).* http://csrc.nist.gov/publications/fips/fips180-2/fips180-2.pdf

[2] Touch, Joseph D. Performance Analysis of MD5. (Sigcomm '95, Boston MA). http://www.isi.edu/touch/pubs/sigcomm95.html

Family History Technology

© 2005, Intellectual

IMAGE LAB COMPUTER APPLICATIONS

- Audit
  - Selects a statistical sample set of images from a folder(s) for auditing
  - Note: Quality Evaluation takes place at the time of capture by the digital camera operator
  - Allows auditor to select viewing speed
  - Automatically presents image to the auditor
  - Records the audit results in metadata and in AMS

© 2005, Intellectual

Audit

- Statistical Sample

© 2005, Intellectual

Folders Ready For Audit

OPERATOR	WORKSTATION	CREATION_	GS_NUMBE	PROJECT_N	IMAGE_COU	DRIVE	PROCESS_S	RETAKE_FL	STATUS	DCDID
123	RUBY	19-DEC-2004	4000605	BLGC0001D	18	d:\DAIMSWor	NEW IMAGE		UNAUDITED	BLGC0001D_432DISK270
123	RUBY	19-DEC-2004	4000606	BLGC0001D	13	d:\DAIMSWor	NEW IMAGE		UNAUDITED	BLGC0001D_432DISK270
123	RUBY	19-DEC-2004	4000607	BLGC0001D	9	d:\DAIMSWor	NEW IMAGE		UNAUDITED	BLGC0001D_432DISK270
123	RUBY	19-DEC-2004	4000608	BLGC0001D	18	d:\DAIMSWor	NEW IMAGE		UNAUDITED	BLGC0001D_12345DISK2
123	RUBY	19-DEC-2004	4000609	BLGC0001D	13	d:\DAIMSWor	NEW IMAGE		UNAUDITED	BLGC0001D_12345DISK2
123	RUBY	19-DEC-2004	4000610	BLGC0001D	9	d:\DAIMSWor	NEW IMAGE		UNAUDITED	BLGC0001D_12345DISK2
123	RUBY	20-DEC-2004	4000642	BLGC0001D	13	d:\DAIMSWor	NEW IMAGE		UNAUDITED	BLGC0001D_738DISK282
123	RUBY	20-DEC-2004	4000643	BLGC0001D	9	d:\DAIMSWor	NEW IMAGE		UNAUDITED	BLGC0001D_738DISK282
123	RUBY	20-DEC-2004	4000644	BLGC0001D	18	d:\DAIMSWor	NEW IMAGE		UNAUDITED	BLGC0001D_738DISK283
123	RUBY	20-DEC-2004	4000645	BLGC0001D	13	d:\DAIMSWor	NEW IMAGE		UNAUDITED	BLGC0001D_738DISK283
123	RUBY	20-DEC-2004	4000646	BLGC0001D	9	d:\DAIMSWor	NEW IMAGE		UNAUDITED	BLGC0001D_738DISK283
123	TRICOLOR	20-DEC-2004	4000626	BLGC0001D	18	c:\DAIMSWor	NEW IMAGE		UNAUDITED	BLGC0001D_9836DISK27
123	TRICOLOR	20-DEC-2004	4000627	BLGC0001D	13	c:\DAIMSWor	NEW IMAGE		UNAUDITED	BLGC0001D_9836DISK27
1157	WS-C2-IMGLAB4	04-NOV-2004	4000512	POL EVA1	291	E:\DAIMS Te	NEW IMAGE		UNAUDITED	POL EVA1_26DISK7
1157	WS-C2-IMGLAB4	05-NOV-2004	4000513	POL EVA1	240	E:\DAIMS Te	NEW IMAGE		UNAUDITED	POL EVA1_26DISK7
1157	WS-C2-IMGLAB4	05-NOV-2004	4000514	POL EVA1	96	E:\DAIMS Te	NEW IMAGE		UNAUDITED	POL EVA1_26DISK7
1157	WS-C2-IMGLAB4	05-NOV-2004	4000515	POL EVA1	73	E:\DAIMS Te	NEW IMAGE		UNAUDITED	POL EVA1_26DISK8
1157	WS-C2-IMGLAB4	05-NOV-2004	4000516	POL EVA1	103	E:\DAIMS Te	NEW IMAGE		UNAUDITED	POL EVA1_26DISK8
1157	WS-C2-IMGLAB4	05-NOV-2004	4000517	POL EVA1	169	E:\DAIMS Te	NEW IMAGE		UNAUDITED	POL EVA1_26DISK8
1157	WS-C2-IMGLAB4	05-NOV-2004	4000518	POL EVA1	90	E:\DAIMS Te	NEW IMAGE		UNAUDITED	POL EVA1_26DISK8
1157	WS-C2-IMGLAB4	05-NOV-2004	4000519	POL EVA1	221	E:\DAIMS Te	NEW IMAGE		UNAUDITED	POL EVA1_26DISK8
1157	WS-C2-IMGLAB4	05-NOV-2004	4000520	POL EVA1	96	E:\DAIMS Te	NEW IMAGE		UNAUDITED	POL EVA1_26DISK9
1157	WS-C2-IMGLAB4	05-NOV-2004	4000521	POL EVA1	94	E:\DAIMS Te	NEW IMAGE		UNAUDITED	POL EVA1_26DISK9
16	WS-COB2-IMAGELB	21-DEC-2004	4000647	QUEB0570D	0	J:\DAIMSTest	NEW IMAGE		UNAUDITED	QUEB0570D_234DISK234
863	WS-COB2-IMAGELB	21-DEC-2004	4000649	BLGC0002D	31	J:\DAIMSTest	NEW IMAGE		UNAUDITED	BLGC0002D_99DISK3

Confidence Level (Z score) 2.575 PRC (Between .5 and .99) 0.98 MoE (Between .001 and .25) 0.05 Quality % 0.98

Images Selected = 531 Auditor RJL Audit View Speed (Secs.) 0.5 OK Cancel

DAIMS Image Auditing - [Audit - J:\DAIMSTest\4000649\distribution\4000649_00023.jpg]

File Edit Image View Window Help

274% Reject Code D0 Comments

4000649_00001.jpg
4000649_00005.jpg
4000649_00006.jpg
4000649_00007.jpg
4000649_00010.jpg
4000649_00011.jpg
4000649_00013.jpg
4000649_00014.jpg
4000649_00015.jpg
4000649_00017.jpg
4000649_00018.jpg
4000649_00019.jpg
4000649_00020.jpg
4000649_00021.jpg
4000649_00023.jpg
4000649_00025.jpg
4000649_00027.jpg
4000649_00028.jpg
4000649_00029.jpg
4000649_00031.jpg

an mil huit cent septante neu
Victor Verlinden, fils de Jean
en marge de l'acte
légitimé par le
te 51) entre: Jean
rlinden. _____
balines, le 1 Mars 1800 quatre
Echevin, officier de l'Etat-civ

PanWindow

Image 15 of 20 70% Complete 0 items Rejected: Threshold=1 Gray Value=228 INS CAPS NUM 12/21/2004 6:46 PM

IMAGE LAB COMPUTER APPLICATIONS

- Make Copy
  - Prepares images and metadata files for writing donor copies to DVD, preservation copies to DVD and DLT tapes, and distribution copies to a network server
  - Sends copies to DVD and DLT writers/drives and to the network server
  - Labels donor and preservation DVDs

© 2005, Intellectual

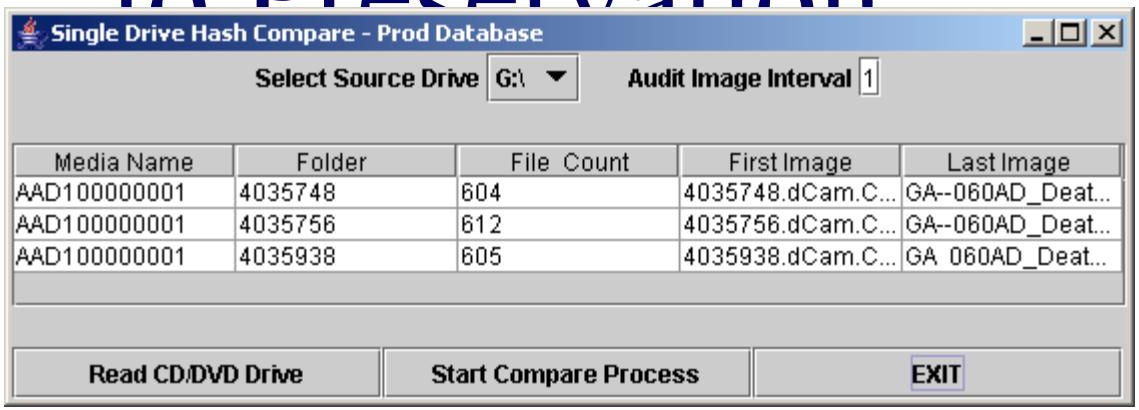
Make Copy—Writing DVDs

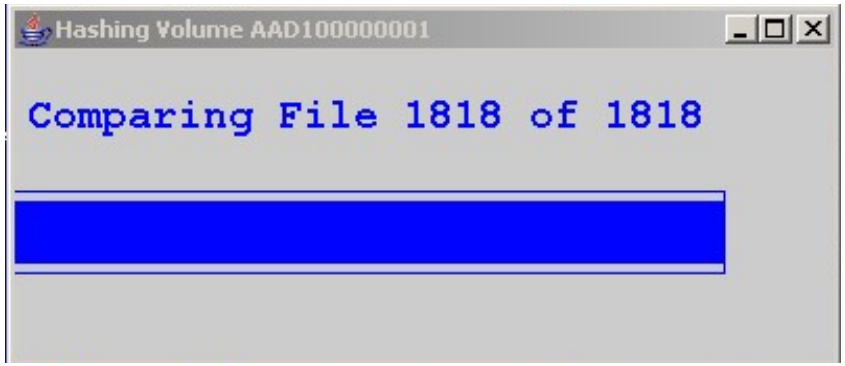
© 2005, Intellectual

Make Copy—Preservation DVD Copy Label

© 2005, Intellectual

Hash Comparison—Original to Preservation





Family History Technology

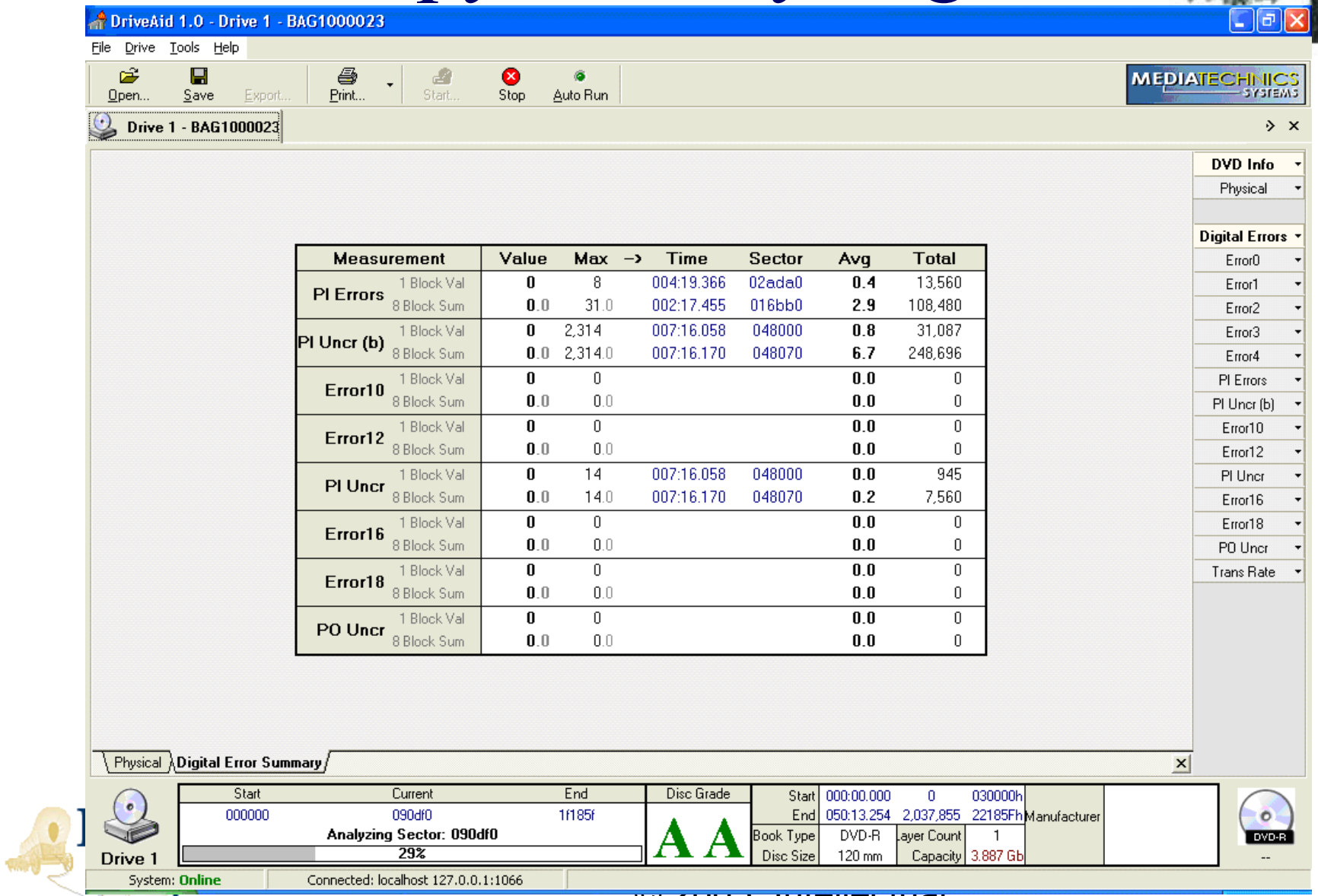
© 2005, Intellectual

Make Copy—Writing DLT Tapes

Family History Technology

© 2005, Intellectual

Make Copy—Analyzing DVDs



Measurement		Value	Max →	Time	Sector	Avg	Total
PI Errors	1 Block Val	0	8	004:19.366	02ada0	0.4	13,560
	8 Block Sum	0.0	31.0	002:17.455	016bb0	2.9	108,480
PI Uncr (b)	1 Block Val	0	2,314	007:16.058	048000	0.8	31,087
	8 Block Sum	0.0	2,314.0	007:16.170	048070	6.7	248,696
Error10	1 Block Val	0	0			0.0	0
	8 Block Sum	0.0	0.0			0.0	0
Error12	1 Block Val	0	0			0.0	0
	8 Block Sum	0.0	0.0			0.0	0
PI Uncr	1 Block Val	0	14	007:16.058	048000	0.0	945
	8 Block Sum	0.0	14.0	007:16.170	048070	0.2	7,560
Error16	1 Block Val	0	0			0.0	0
	8 Block Sum	0.0	0.0			0.0	0
Error18	1 Block Val	0	0			0.0	0
	8 Block Sum	0.0	0.0			0.0	0
PO Uncr	1 Block Val	0	0			0.0	0
	8 Block Sum	0.0	0.0			0.0	0

© 2005, Intellectual

Make Copy—Analyzing DVDs: CSV Files

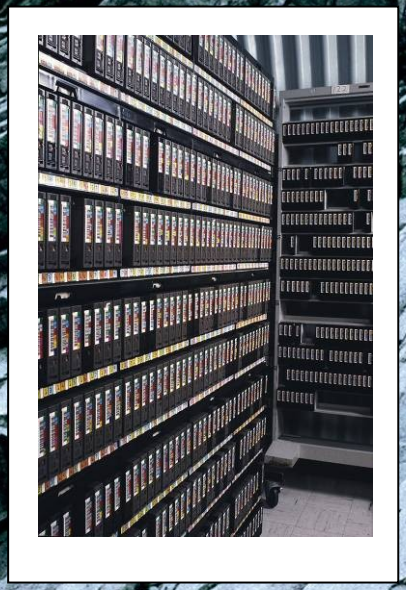
BlemishTest_10_cobfusion2_mitsuitestdisc#2_2Mar05.csv - Notepad

File Edit Format View Help

```
File ID,Title:,BAG1000023,,▯,File path:,.\2005-03-02_174003.dap,,▯,Creation Time:, Wednesday
2341,0.000000,0.000000,0.000000,11927145.000000,,▯,,,,Total,87489,18716,0,0,0,355251,2245601
0,0,0,4,0,0,0,0,0,0,0,6050476,,▯,192,,00:00:00.00,0.000000,,1,0,0,0,0,8,0,0,0,0,0,0,0,606016
,0,0,0,0,7,0,0,0,0,0,0,0,6214609,,▯,448,,00:00:00.00,0.000000,,1,0,0,0,0,9,0,0,0,0,0,0,0,615
,,1,0,0,0,0,12,0,0,0,0,0,0,0,6093222,,▯,704,,00:00:00.00,0.000000,,1,0,0,0,0,4,0,0,0,0,0,0,0
0,,1,0,0,0,0,3,0,0,0,0,0,0,0,6505881,,▯,960,,00:00:00.00,0.000000,,1,0,0,0,0,4,0,0,0,0,0,0,0
:00.00,0.000000,,1,0,0,0,0,7,0,0,0,0,0,0,0,6644453,,▯,1216,,00:00:00.00,0.000000,,1,0,0,0,0,
546,,▯,1456,,00:00:00.00,0.000000,,1,0,0,0,0,9,0,0,0,0,0,0,0,6739515,,▯,1472,,00:00:00.00,0.
0,0,0,0,0,0,0,6753874,,▯,1712,,00:00:00.00,0.000000,,1,0,0,0,0,8,0,0,0,0,0,0,0,6601448,,▯,17
0000,,1,0,0,0,0,6,0,0,0,0,0,0,0,6750764,,▯,1968,,00:00:00.00,0.000000,,1,0,0,0,0,4,0,0,0,0,0
,,00:00:00.00,0.000000,,1,0,0,0,0,4,0,0,0,0,0,0,0,6935990,,▯,2224,,00:00:00.00,0.000000,,1,0
,0,7023624,,▯,2464,,00:00:00.00,0.000000,,1,0,0,0,0,2,0,0,0,0,0,0,0,6960686,,▯,2480,,00:00:0
,0,0,0,3,0,0,0,0,0,0,0,6911061,,▯,2720,,00:00:00.00,0.000000,,1,0,0,0,0,8,0,0,0,0,0,0,0,7129
0.00,0.000000,,1,0,0,0,0,2,0,0,0,0,0,0,0,7108759,,▯,2976,,00:00:00.00,0.000000,,1,0,0,0,0,5,
87640,,▯,3216,,00:00:00.00,0.000000,,1,0,0,0,0,7,0,0,0,0,0,0,0,6949551,,▯,3232,,00:00:00.00,
9,0,0,0,0,0,0,0,0,7026148,,▯,3472,,00:00:00.00,0.000000,,1,0,0,0,0,12,0,0,0,0,0,0,0,6839331,,▯
.000000,,1,0,0,0,0,12,0,0,0,0,0,0,0,6969786,,▯,3728,,00:00:00.00,0.000000,,1,0,0,0,0,3,0,0,0
3968,,00:00:00.00,0.000000,,1,0,0,0,0,8,0,0,0,0,0,0,0,7037108,,▯,3984,,00:00:00.00,0.000000,
0,0,0,0,6977248,,▯,4224,,00:00:00.00,0.000000,,1,0,0,0,0,7,0,0,0,0,0,0,0,6984727,,▯,4240,,00
,0,0,0,0,3,0,0,0,0,0,0,0,6966474,,▯,4480,,00:00:00.00,0.000000,,1,0,0,0,0,7,0,0,0,0,0,0,0,69
0:00.00,0.000000,,1,0,0,0,0,3,0,0,0,0,0,0,0,6785521,,▯,4736,,00:00:00.00,0.000000,,1,0,0,0,0
372,,▯,4976,,00:00:00.00,0.000000,,1,0,0,0,0,8,0,0,0,0,0,0,0,7092854,,▯,4992,,00:00:00.00,0.
,0,0,0,0,0,0,7017741,,▯,5232,,00:00:00.00,0.000000,,1,0,0,0,0,8,0,0,0,0,0,0,0,6950375,,▯,524
000,,1,0,0,0,0,5,0,0,0,0,0,0,0,6969786,,▯,5488,,00:00:00.00,0.000000,,1,0,0,0,0,2,0,0,0,0,0,
8,,00:00:00.00,0.000000,,1,0,0,0,0,10,0,0,0,0,0,0,0,6849715,,▯,5744,,00:00:00.00,0.000000,,1
,0,0,6958209,,▯,5984,,00:00:00.00,0.000000,,1,0,0,0,0,13,0,0,0,0,0,0,0,6986391,,▯,6000,,00:0
0,0,0,0,4,0,0,0,0,0,0,0,6950786,,▯,6240,,00:00:00.00,0.000000,,1,0,0,0,0,4,0,0,0,0,0,0,0,705
00.00,0.000000,,1,0,0,0,0,5,0,0,0,0,0,0,0,6917174,,▯,6496,,00:00:00.00,0.000000,,1,0,0,0,0,5
17741,,▯,6736,,00:00:00.00,0.000000,,1,0,0,0,0,11,0,0,0,0,0,0,0,6852917,,▯,6752,,00:00:00.00
0,5,0,0,0,0,0,0,0,7080865,,▯,6992,,00:00:00.00,0.000000,,1,0,0,0,0,7,0,0,0,0,0,0,0,7021102,,
0,0.000000,,1,0,0,0,0,11,0,0,0,0,0,0,0,6985143,,▯,7248,,00:00:00.00,0.000000,,1,0,0,0,0,7,0,
6,,▯,7488,,00:00:00.00,0.000000,,1,0,0,0,0,8,0,0,0,0,0,0,0,7129933,,▯,7504,,00:00:00.00,0.00
0,0,0,0,0,0,0,0,6894406,,▯,7744,,00:00:00.00,0.000000,,1,0,0,0,0,7,0,0,0,0,0,0,0,7116522,,▯,77
```

Family History Technology

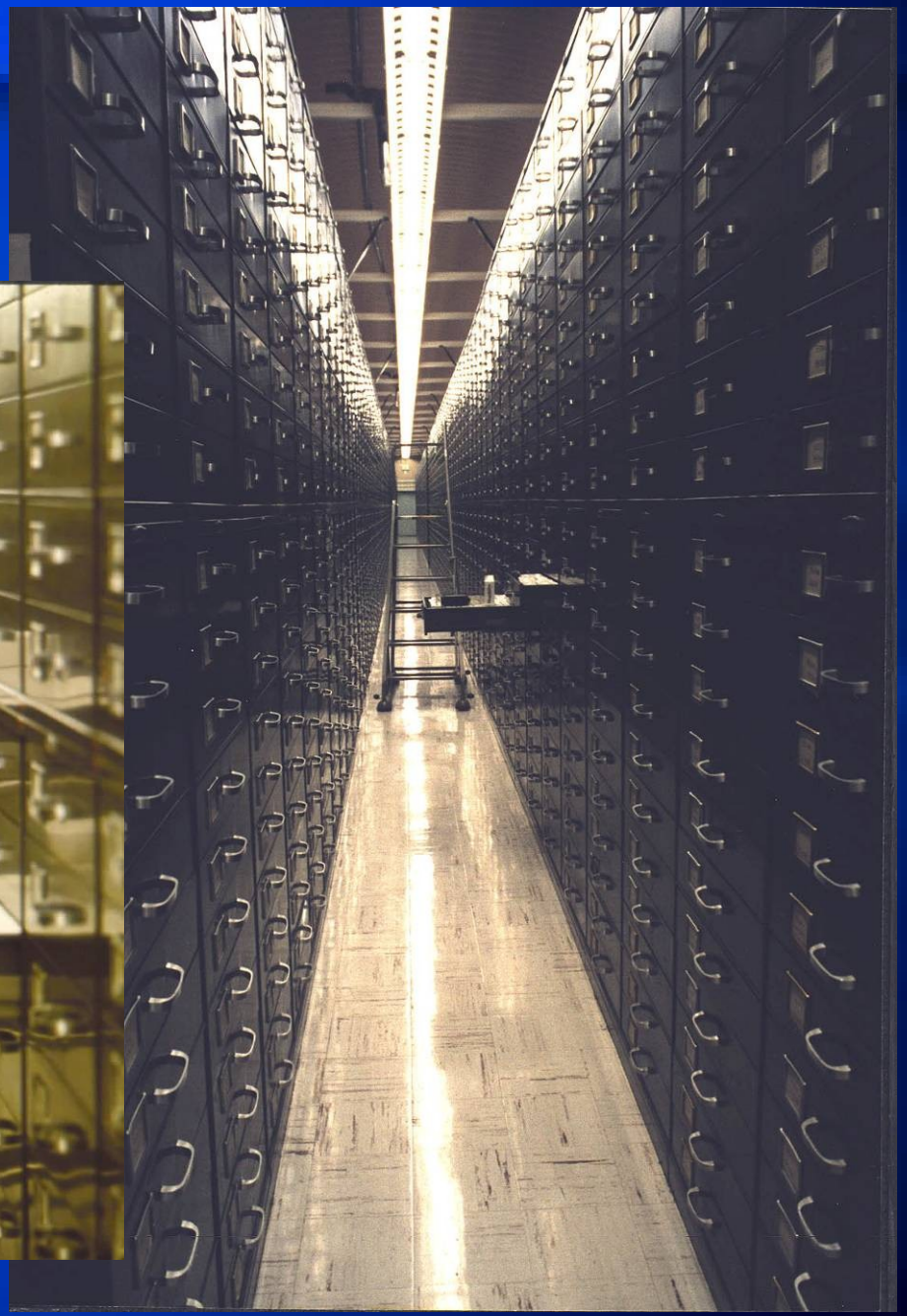
© 2005, Intellectual



Preservation Collection Monitoring

Granite Mountain Records Vault

Microfilm Records Preservation



FUTURE PRESERVATION
SPINNING HARD DRIVE?

- Long-term cost will be lower
- Migration is achieved by adding new technology (servers, disks, etc.) to existing equipment
- Testing the viability of the files and media (disks and supporting components) is done in real time and with built-in monitoring
- The reloading of files in unnecessary to fill requests from other users or to recover, reprocess, and restore data to the hosting system

© 2005, Intellectual

Imaging Lab Process

Questions?

© 2005, Intellectual