DataTrade
DIVISION

GOLDLEAF
FINANCIAL SOLUTIONS

White Paper on Digital-to-Digital Conversion Services

Digital-to-digital conversion services entail converting document images, checks, statements and computer reports that are already in a specific digital format and change them to a different format. Digital conversion should not be confused with migration, which involves moving data from one media to another without changing the format. While digital conversions are required for numerous reasons, most stem from companies moving to a new technology that is incompatible with the one they are currently using. This document further explains the reasons conversion services are needed and necessary, the process that takes place, and how the use of old legacy information can be maximized regardless of origin or format.

The Evolution of Document Imaging and Report Management (COLD) Technology

Digital storage of images and reports has been available to the business community for about twenty years. In that relatively short period of time, the technology has come a long way. The slow, big iron hardware with low density media and rudimentary software has now morphed into $249 desk top systems on the one hand to an integral element in today's Enterprise Content Management (ECM) systems, on the other.

The evolution of digital storage technology can be categorized into three stages:

Stage 1
In the early 1980s, document imaging and report management was generally developed by large corporations for large corporations – with a corresponding price tag. Most systems were standalone technologies that promised a paperless office. Not only did this not come to pass, but in order to maintain their perceived competitive edge and to lock in customer loyalty, vendors systematically employed complex compression routines and placed proprietary "wrappers" around their images ***(see sidebar next page)*)** so that the images could not be read by any other system. For example, IBM used MODCA (*Mixed Object Document Content Architecture)* and IOCA (*Image Object Content Architecture)* instead of a standard format such as TIFF (*Tag Image File Format)* that uses well defined "industry standard" and published compression routines, headers and wrappers. In the check imaging world, an example would be Bisys/DSI that actually used a TIFF compression routine but also employed a proprietary wrapper that rendered the images unreadable by anything other than their proprietary software.

Stage 2
In the late 1980s through the late 1990s, a middle tier of document management companies evolved that developed and sold systems for middle and lower tier companies. These vendors started out developing proprietary formatted systems, but later, as more industry standards evolved, eventually moved to embracing these formats. During this time hardware performance increased as prices decreased; and storage media density, mainly optical disks, also increased in capacity.

Stage 3
This stage started in the late 1990s and continues today. During this period, imaging technology became a commodity item and while some implementations are still standalone, most are integrated with other information repositories or used as a component in a larger Enterprise Content Management (ECM) system that pools images, data, sound video, email and any other information asset into repositories with a single access portal. ECM has its own integration problems when pooling many existing disparate proprietary technologies.

Problems Associated With Legacy Archive Systems

1. Proprietary images formats

As technology evolves, each stage has a finite lifecycle. Once a lifecycle is over, it is replaced with the latest technology, which invariably is better, faster, smaller and cheaper than the previous one. In the case of imaging and COLD technology, the problems with proprietary wrappers and compression routines emerged once stage one and the early years of stage two mentioned above, were over. Vendors that used proprietary formats to store images and reports locked users into the old system. **The bottom line was a new system could not read an old system's images or reports**. This resulted in end users being forced to maintain both the new and old systems which were undesirable for the following reasons:

1. Keeping both systems means paying hardware and software maintenance on both systems.
2. Users have to launch and access two systems to get the information they need.
3. The media that contains the old images becomes unreliable as it ages.
4. The old hardware that reads the old media becomes obsolete, unsupported and eventually permanently inoperable.

2. Effect of recent legislation

With the passing of recent legislation such as Gramm-Leach-Bliley, HIPAA, Sarbanes-Oxley and SEC 17a-4, records management was finally forced from its dusty basement to a central place in business strategy. With particular regard to information stored in old legacy systems, the new laws caused corporations to take a harder look at the security, reliability and accessibility of that information. In addition to facing the strong probability of inoperable hardware and unreadable media mentioned in items 3 and 4 above, other problems corporation faced include:

1. The inability to meet the lifecycle and archive standards mandated by current legislation.
2. The high level of personal risk that face C level executives and corporate boards for ignoring the archiving integrity of corporate information.

"Standard" and "Proprietary" formats explained

An image is created using pixels that are collectively called raster data. Raster data must be assembled (formatted) to create a viewable image. The information on how data is formatted for a particular image is contained in a "header" or "wrapper". The formatting information contained in the header may include compression algorithms and methodology, bits per pixel, image width and height, image aspects and other such specifications. Without this information, it is virtually impossible to "interpret" the data and create an image from it.

Headers fall into two basic categories, "Standard" and "Proprietary". Examples of "standard" headers would be "TIF", "JPG", and "BMP", etc. With published, consistent formatting information in the header, standard generic image viewers are able to interpret the raster data and easily display images.

With proprietary headers, the situation is different. The roadmap in the header to interpret the raster data is known only to the vendor or programmer. Only the vendor can assemble the raster data to view the image. Accordingly, a TIF viewer (or any other "standard" viewer) will be unable to interpret the raster data or display a proprietary image.

Digital-to-digital conversion involves "cracking" the codes and keys on the image headers to turn the image from a proprietary format to a standard format.

Digital-to-digital Image Conversion Can Solve These Problems

Digital-to-digital conversion solves the above problems by taking proprietary formatted images and reports and converting them into standard formatted data. In the case of images, all indexes/MICR are extracted from the old system's database and associated with its corresponding image. The resulting output is prepared to enable end users to use one of the four archive/retrieval scenarios mentioned below.

Archive/retrieval methods of converted images and reports

1. The images and indexes are placed in files utilizing a naming convention that allows them to be easily ingested into the client's new system. This process results in users accessing all images and reports, old and new, from the new system.
2. Content can be loaded into a new archive using the storage technology such as RAID, coupled with search and retrieval software. This system is then placed at the client site and users can access the content using network or web based workstations.
3. Content can be archived on new media as in (2) above but hosted at the conversion company's site in an ASP environment with users accessing content via web connectivity.
4. Content can be loaded onto individual digital media such as CD or DVDs along with retrieval software so that each media is a self contained archive.

Digital-to-digital Conversion Process

1. The client sends the conversion company a sample of the images or reports that require conversion and an approximate volume count.
2. The conversion company processes a test sample and returns it on digital media along with an estimated cost to complete the services.
3. After signing off on the sample all the media that requires converting in sent in.
4. The actual job is run with rigorous Q/A and the content returned on the media of choice.

Benefits of the Conversion Process

The benefits of the conversion process are numerous and include:

1. The risk of legacy hardware failing or not being supported is eliminated.
2. The risk of the old media becoming unreliable is eliminated.
3. The cost of maintenance on legacy hardware and software is eliminated.
4. All reports from day one are archived and accessed from one system using the latest technology, thereby improving usability and user effectiveness.
5. Combining all content into one repository streamlines document retention policies and compliance requirements.
6. The need to re-scan and re-index old documents into a new system is eliminated.
7. The risk of non-compliance with legislated lifecycle requirements and archive standards is mitigated.
8. A roadblock to achieving a complete Enterprise Content Management System (ECM) is eliminated.

Other Items to Note in Data Conversions

Auditable quality control procedures

Accuracy should be a conversion vendor's number one priority, and proving it should be number two. All conversion work should pass through a rigorous quality control process that provides auditable reporting of the content on (1) the media received, (2) the conversion process and (3) the target media produced. This will result in clients having the assurance that all content intended for conversion was actually converted. And conversely, any specific source content not converted is reported as exceptions.

The difference between conversion and migration of data

Conversion involves taking data in one format and converting it into another format.
Migration involves moving data from one type of media to another type of media without changing format. Migration services are usually used when archive media or hardware is old and unreliable and clients wish to move the data to new higher density media that can be read by the latest hardware technology. Most conversion vendors should support a long list of hardware and media when offering migration services.

Test conversions

Prior to selecting a vendor to perform conversion services, companies requiring digital conversion should request that the vendor convert a sample set of images or COLD reports for test purposes. The test should include converted images and reports along with, as necessary, all corresponding indexes that have been extracted from the legacy system's database. This test process is particularly relevant if the converted output is to be imported into a new system. The fact that a conversion vendor has previously converted specific source content has nothing to do with preparing the content so that the target system can properly import it. When importing images and indexes into a new system, that system usually requires that the files have a specific naming convention and indexes are placed in a specific format. This should be tested by whoever is responsible for the import, which is typically the new system vendor.

Imaging system audit service

Some conversion vendors will offer image system audit services. This is important in industries affected by recent legislation such as Sarbanes-Oxley and Check/21 that mandate document archive and lifecycle standards. A conversion services vendor can help in compliance issues by offering audit/quality control services that can mitigate the risk of loss and liability from irretrievable customer information.
To accomplish this, a copy of the image archive and corresponding index database is sent to the conversion services vendor. The vendor runs these files through a process that will identify all images that are irretrievable due to:

1. Bad scans
2. Corrupt files
3. Images that have partial or no database records
4. Database records with no corresponding images
5. Duplicate records

Whenever possible, the vendor should extract all exception items on to separate media for the client's investigation and resolution.

Data encryption and confidentiality and non-disclosure agreement

It is often desirable that conversion source data be encrypted before sending it to the conversion vendor and also when the target data is returned by the vendor to the client.. Clients should also ensure that prior to sending any data to a conversion vendor, confidentiality and non-disclosure agreements, appropriate to the client's situation and industry mandated requirements, are signed and in place.

Vendor experience in converting images, reports and data

Even though they are few in number, it is usually best to select a company whose core competency is digital-to-digital conversion services. Data conversion work is often coined the "dirty end" of the content storage business. This is probably why it is not just end users that use conversion vendors but many ECM and document management companies, outsource service bureaus and ASPs, as well as systems integrators and resellers, all subcontract out conversion work.

Conversion companies should have the in-house expertise and talent to "crack" proprietary image formats. Cracking files is both an art and science.

Conversion vendors should be able to handle conversions that range in size from a few hundred megabytes to ten terabytes or more. This means high capacity hardware, wide bandwidth connectivity and a large inventory of old obsolete hardware to enable them to read old and long forgotten archive media.

Quality control is of utmost importance. Clients should receive a report from the conversion vendor that demonstrates that all content intended for conversion is actually converted. And conversely, any specific source content not converted, is categorized in directories and reported as exceptions.