

**Additional Material for the Congressional Record from Dr. Richard Lysakowski,
Executive Director of the Global Electronic Records Association (GERA),
President of the Collaborative Electronic Records Association (CENSA)**

Witness on Expert Panel

**House Government Reform Committee, Subcommittee on Technology, Information
Policy, Intergovernmental Relations and the Census, Meeting on July 8, 2002**

Submitted July 28, 2003

1 Followup Written Points for the Congressional Record:

I was asked to provide additional clarification points, research results and further recommendations to the Subcommittee. I respectfully submit the following points for consideration by Congress.

**1.1 NARA Alone Cannot and Should Not Be Expected to Solve the USA's
Problems with Electronic Records.**

Electronic records are not uniquely the problem of the US National Archives and Records Administration (NARA). Electronic records are integral to all eGovernment work processes in all federal, state, and local agencies. Several times over the past five years, private industry has approached OMB to develop or mandate common government-wide and industry-wide solutions and standards for electronic records. Everytime, OMB has pushed all responsibility for electronic records onto NARA's shoulders, now calling NARA the "Managing Partner for the Electronic Records Management (ERM) e-Gov Initiative." This has slowed badly-needed progress.

NARA's ERM Initiative's stated goals are to provide a government-wide policy framework and to provide guidance for electronic records management. NARA's mission, scope, focus, size, budget (\$268M FY2003) do not permit it to solve all electronic records problems of federal, state, industrial, and nonprofit organizations. NARA's primary mission remains to "ensure continuing access to essential evidence that documents: 1) the rights of American citizens, 2) the actions of federal officials, 3) the national experience." NARA takes stewardship of only a very small percentage of federal and other historical records for long-term preservation and access.

NARA is underfunded and understaffed to develop scalable solutions and standards for electronic records for all federal, state, and local government agencies. Thus industry is left to wait for open or de facto standards to emerge or to create its own. More often than not it is a waiting game because industry standards must be compatible with government regulations for recordkeeping. At times, industry creates de facto standards that governments adopt, like AT&T's UNIX, IBM's PC, and Adobe's Portable Document Format (PDF). However, at this point IT innovation is being stifled across the board by the lack of standards for electronic records and data formats. If federal agencies were to simply adopt or drive archival formats and interoperability standards, industry would adopt them, implement them, and drive faster creation of robust markets and technical services to support them. It does not matter who leads, but it is important that someone does.

A great example of what the US Government can do for itself and industry with a moderate investment is the US DoD's 5015.2 Standard for Records Management Applications. Once DOD

created and mandated this standard, it was adopted and used as a baseline for standards throughout many industries, including pharmaceutical, chemical, food, medical devices, aerospace, automotive, and others. It remains a baseline against which all good records management technology systems are measured and certified. The total budget to create DoD 5015.2 was estimated to be under \$5M USD and it took less than three years time to develop. Industry spent at least ten times that amount of money to develop DoD-compliant solutions. Such a large industrial investment would have been needed to meet any specification where technology changes were needed.

DoD 5015.2 standard did not go far enough and specify standards for the entire lifecycle of electronic records, including requirements and specifications for records creation, records management, archival functions, preservation functions, retrieval and reproduction functions, and file format standards for individual document-based records and collection of records. This is excusable because the scope had to be limited so that DoD could deal with known problems for technological solutions existed in 1995.

We should take advantage of the fact that The Department of Defense has the power to create and mandate standards faster than any other government body and there is frequently spillover of benefits directly to the civilian sector and the commercial markets.

1.2 Government Administrative and Regulatory Bodies Drive Industry, Not NARA.

Industrial and private sector problems with electronic records are driven by factors not directly related to NARA's mission. For example, the private sector is driven by many concerns simultaneously, e.g., the USPTO drives concerns for intellectual property protection, the EPA drives concerns for environmental protection and quality, the HHS / FDA drive concerns for healthcare product safety and quality, the SEC drives concerns for accurate financial reporting and integrity. For transportation, the FAA flight safety regulations mandate long-term availability of design, test, calibration and maintenance records for commercial and military aircraft in service for longer than 20 years. For automotive transportation safety, records are needed for product design, testing, maintenance and recalls for two or more decades. In the life, medical, disability insurance industries records retention periods are anywhere from 15 to 90+ years. Mortgage and banking industry mandate retention 15, 30 years (or longer). DOL ERISA-compliance and Social Security Administration pension benefits plan record retention periods cover the working life of all citizens and beyond. These are just a few of many immediate concerns with long-term impact on the private sector. All these industries are trying now to use electronic records whenever possible.

Problems of electronic records are multifaceted – driven by administrative, legislative, legal, regulatory and business concerns – but solved via well-designed and implemented organizational programs, qualified people, policies, quality standards and procedures, and technology systems. CENSA and GERA have partnered industry and government agencies to create the set of standards called “the Quality Electronic Recordkeeping Practices” (QERP) standards that articulate all the required program and technology elements to design, staff, operate, audit, and maintain electronic evidence over any time frame. While the need for keeping evidence of activities is universal throughout the public and private sectors, the scope, scale, values and accountability reasons for recordkeeping vary so widely that NARA's solution and approach will not work for all.

1.3 NARA's ERM eGov Initiative – Where's the eArchive in each eGov Agency?

NARA's current strategy focuses on assisting other federal agencies with electronic records management initiatives by providing a methodology for determining agency-unique requirements on top of the US DoD 5015.2 RMA Standard. While this approach takes good advantage of the excellent work of the DoD standard – which NARA provided significant help to develop – it does not directly address the need for each federal agency to set up and run their own electronic records archive. It does not lay down the standard for electronic records archives to successfully deal with the 99% of agency records that NARA never receives.

NARA's Electronic Records Archive (ERA) solution is being designed to solve NARA's immediate technological needs for a system. The NARA ERA is a "custom" system designed to NARA's specifications. NARA's ERA is not being designed to be a general-purpose, "Configurable-Off-The-Shelf" (COTS) product designed to scale up to large agencies like the Department of Defense or Homeland Security, or scale down to much smaller agencies at federal, state, and city levels. From my point of view, this is a wasted opportunity to mandate standards and build markets.

1.4 A More Coordinated, Multi-Agency and Industry Approach is Needed

For eGovernment to work, a more universal, distributed approach is needed that does not put NARA at the center of everything. Sole reliance on NARA for guidance and standards could cause a bottleneck in moving to electronic recordkeeping in eGovernment. A universal, large-scale approach can be developed with NARA as a leader, follower, and participant. However, some fundamental strategy shifts must occur so NARA does not constrain processes it seeks to facilitate.

I recommend that a small coordinated team of US agencies and industry with a large stake in preservation of digital information assets be assembled to make progress much faster than thus far. I recommend NARA, LOC, The Smithsonian Institution, GAO, NASA, DoD, HHS, and industry work together. NARA's expertise is crucial and it cannot solve everyone's problems.

NARA is too small to cause the kind of change that we need within a reasonable timeframe. Its small budget (est. \$289M FY-2004) is seriously constraining. Other US agencies have resources that greatly exceed those of NARA. For example:

- The Library of Congress (est. \$416M FY-2004)
- The Smithsonian Institution (est. \$478M FY-2004)
- General Accounting Office (est. \$474M FY-2004)
- NASA (est. \$15B FY2004)
- DoD (est. \$389B FY2004)
- HHS (est. \$539B FY-2004)

These other agencies have their own internal needs for total lifecycle solutions for electronic records (including long-term preservation and access) that are similar to NARA's internal needs.

Several agencies working together would push for faster and better solutions to the digital records problem much better than NARA alone. A focused team of large US government departments and agencies with the most at stake should be assembled to solve this problem. The Library of Congress' Digital Library program includes digital preservation and is succeeding. NASA has preservation needs for massive digital data sets that record the weather, the condition of the earth,

space science, and defense-related issues. HHS scientific and medical research, development, and regulation are being slowed by the lack of standards for digital data and record interchange. *These few, but powerful agencies and departments could collaborate as a focused team to create and mandate solutions in three to five years, rather than 20.*

1.5 Leverage The Success of Other Countries

Other countries have succeeded by working on a scale appropriate to the size of their agencies.

Australia in particular has much we can learn from and adopt. The Australian success has led US experts to acknowledge that scalable, long-term solutions can be built using today's technologies. They solved electronic records management and archives problems on a smaller scale than the US's problems, but one that can be extrapolated for the USA. It too is meant to serve as a core building block of their eGovernment initiative. The system was developed to first fit the State of Victoria (11M citizens) and then fit the Whole of Australia (26M citizens). The solution is commercially available now and can adapted to fit a wide range of needs of agencies, states, and industries. US agencies and industry worldwide should look to leverage as much as possible from Australia. I have studied their system in detail and visited Australia twice to see it in action. It is clear that their system is fully compatible with US NARA's proposed ERA system.

The State of Victoria Electronic Records Strategy (VERS) Program started in about 1995. The Victoria State Department of Infrastructure led a cross-disciplinary team of Australia's leading government archivists, records managers, computer scientists and software engineers, and leading software vendors. It uses XML and the Adobe Portable Document Format (PDF) for electronic record archival collection storage and retrieval. Their initial goal is to retrieve and reproduce authentic electronic records in an XML/PDF format for 100 years minimum (with or without Adobe Systems, Inc. in existence.) The VERS program includes organizational program, policy, technology system design standards, training, and government-wide dissemination projects. An important point to note is that they reproduced the PDF Viewer software completely on their own (with no help from Adobe) from the written specification as a "digital archaeology experiment" to prove to themselves that could reliably reproduce PDF-based documentary records.

The VERS team of experts researched requirements, prototyped systems using available commercial software as the starting point, tested the solution in pilot projects for many months, and put the final system into production in 2001. The Victoria Government has now mandated compliance by law for all state, city, and local agencies for the software system functionality, PDF documentary record file format, and XML electronic record collection format. File formats other than PDF can be embedded with the PDF preservation copy if full reprocessability is required.

The VERS Program Office has established a Centre for Excellence for Electronic Records that is charged with disseminating the technology solution and training on programs, policies, and procedures to all the other government departments. They have budgeted money to certify enough competing vendors so that they have a set of flexible, scalable technology solutions that will fit different government department needs.

An important point is that the Australian Government removed a barrier to market development (and their own success) by paying for vendor and product compliance testing and certification themselves. In the US, to get a product certified for DoD 5015.2 standard compliance each vendor

must pay \$10,000 or more for each major product version. Many vendors have complained that this is excessive when it must be repeated for each major product release. In the case of Australia they can keep costs of re-testing compliance and recertification by evolving their record and file collection archive formats prudently using XML's schema versioning and version management tools.

By 2007, the plan is to disseminate the program and technology solution to all departments of the State of Victoria, and likely to the Whole of Australian Government. The total program budget from 2001 through 2007 is expected to be \$90M Australian Dollars. It is startling that Australia could make so much progress so fast, starting after but finishing before the USA, and using less than 1/10 the money. Australia is about 1/10 the population of the USA.

1.6 Avoiding Great Waste and Expense to Society

We must create (or adopt) and mandate standardized solutions soon. We will continue to buy and sow the seeds of our own destruction until then. In a December 1999 research report¹, subtitled "Titanic 2020" I gave examples of the hidden costs to society for the lack of standards. This report is available at "<http://www.censa.org/html/Publications/Titanic2020.htm>". Using the example of one generation of word processing documents alone, I *conservatively* estimated the cost of the lack of electronic document interchange standards to society was \$52 billion USD. I did not assess economic impact for other documentary formats. Without standards for information asset preservation across technology generations, every new system that we install becomes the nucleus of an iceberg that we will either hit and destroy assets, or pay excessive costs to avoid asset loss.

1.7 Subscription-Based Software Is A Huge Threat to Record Security and Assets

Software products are tools used to generate personal and business property (data, information, and knowledge). The usual model of software product sales is a perpetual, right-to-use license on a limited number of computers for a fixed price. Like all tools, software depends on having quality product manufacturers. Software vendors would love to change the way that software is purchased to transform software into a service or utility like water, electricity, gas or oil that is consumed and must be re-paid for periodically – either monthly, annually, or some other period. Microsoft was the first major large software vendor to push for subscription-based pricing models. Advocates for subscription-based software licensing (mostly vendors) argue that it eliminates hassles with software updates, because updates can be done automatically as part of the subscription.

The problem with this is method of software purchasing is that if you do not pay your bill, your software can be "deactivated", "time out", shut off, or somehow made unusable, thereby denying you access to data and documents that you created.

There are federal and state laws that prevent denial of access to one's own property. I question the legality and ethics of this method of selling software without giving buyers a no-cost exit strategy in the form of easy migration. Having to pay a subscription to access your own property is similar to extortionary tactics used by the Mafia to shut down businesses or take people's livelihood and property hostage until regular "payments" are made. Records are corporate, government, or

¹ "Looming Information Age Crisis Expected To Cause Trillion-Dollar Losses Over Next 20 Years" by Dr. Richard Lysakowski and Zahava Leibowitz, CENSA, Inc., December 1999.

individual property – one’s access to them must never be at risk of “deactivation.” Subscription-based pricing models being foisted upon unwary consumers give vendors freedom to use abusive pricing or unethical business practices.

The best way around this potential loss of access to one’s property is to require and buy open data and record portability standards with perpetual, right-to-use licenses for data schemas and file formats. There are too many cases of lost data assets because undocumented, proprietary file formats were not migrated before they became obsolete, the vendor “upgraded” the file format to include incompatible features, the vendor dropped that product line, the vendor’s business failed, or the business was taken over by another company that stopped supporting the format. Buying perpetual rights to use data schemas and file formats avoids all of these problems because it permits easy migration by the property owner, the original vendor, or any third party vendor.

We do not need privatized “software utilities” or software sold as a service. We need more high-quality software product manufacturers that do not use file formats as a way to control their customers’ property. We are in this quandary because buyers have not insisted on open, published formats as an absolute condition of purchase of software.

1.8 Lack of Government-Mandated Record Storage Format Standards is Holding Back Innovation and Costing Taxpayers Big Money

Many US government bodies adopted PDF as a document transmission and preservation standard years ago (FDA, the US Courts, EPA, FAA, IRS, and others). Slow formal acceptance of PDF as an official record storage format by the US Federal Government as a whole harms innovation.

Some compelling statistics about PDF are:²

- 2400+ government agencies worldwide use PDF for document interchange.
- 500 million freely-licensed copies of PDF Reader have been downloaded by public and private sector individuals. Last year over 50,000 copies were being downloaded per day.
- Over 2,000,000 official publications exist on US government external websites.
- Over 675,000 links exist worldwide to download the PDF Reader from sites all over the world.
- Over 1800 PDF tools developers are creating add-ons and tools that compete directly with Adobe.
- Adobe Reader comes preloaded on PCs from the top 10 PC manufacturers.
- The Government of Germany just signed a country-wide deal for PDF and Acrobat and has mandated the use of PDF for interchange format for interagency communication.
- Holland, Italy, and Australia have written PDF and XML into law as mandatory for archival submissions of high quality documentary records.
- Japan and the European Union (EU-EMEA) have specified PDF for all electronic Common Technical Document (eCTD) supporting documents; an XML backbone is used to relate all files together for new applications for approval of new medicines and medical devices.

² Obtained the market research department at Adobe Systems, Inc.

While the numbers are large, they make an important point, that when a technology that is necessary and sufficient exists, adoption as a de facto standard proceeds quickly. Adoption speeds innovation, saving government and industry huge amounts of money.

Money-savings tied to the adoption of PDF is well illustrated by the US Bankruptcy Courts. The US Bankruptcy courts used PDF for all court records for the Enron, Worldcom, and Global Crossing bankruptcies. The US Courts admitted that they did not have to add one new clerk, but estimated that if they had to use paper records, they would have had to add 50 new court clerks.

1.9 Risk Assessment in Section 300 of OMB Circular No. A-11 (2002) Does Not Detail or Score the Asset Protection Requirements

OMB must specify a procurement process for eRecords systems that requires demonstration of interoperability and complete independence of supplier data and file formats before purchasing. Product independence means full capabilities to migrate one's property away from one vendor's system to another without additional expense. With electronic records and data formats, this is a good way to provide an insurance policy against rapid planned obsolescence by computer hardware and software vendors.

OMB Circular No. A-11 Section 300 (2002) Planning, Budgeting, Acquisition, and Management of Capital Assets, is incomplete and insufficient for information asset protection. This OMB Circular A-11 must separately articulate and quantitatively score all US government IT system procurement processes for "interoperability" and "preservation capabilities." These critical aspects of IT systems are buried in the discussion on "Risk Inventory and Assessment (All Assets)" in Part 1, Section F on page 20 and 21. These should be separately scored as "Asset Protection and Preservation" and "Interoperability and Information Interchange."

Without an explicit scoring system for information technology system capabilities for "asset protection and preservation" and "interoperability", US Government procurement processes will not protect government, industry, or taxpayer property (records). We continue to put everyone's assets at high risk, and pay more than necessary to retain ownership and access to them.

1.10 The US Needs Better Procurement Processes for Protecting Record Assets

Given that OMB's A-11 Circular does not explicitly measure and quantitatively score IT system information asset preservation, protection, or interoperability, one can conclude that procurement processes will not sufficiently protect agency or taxpayer assets over long periods of time. This is not unique to OMB, but is a prevalent buying practice throughout government and industry.

All IT system buyers' behavior must change to require as a pre-condition of purchase:

- 1) full portability and reuse of records, and
- 2) permanent access to data and records generated by the buyer from all vendor's software.

As long as government and industry do not insist on these pre-conditions for all purchases, they will continue to lose valuable assets. This has been the status quo for decades.

See slides in Appendix for more information on why the market status quo is the way it is now.

1.11 A Simple Buying Strategy Exists to Protect Business Assets

From watching some of the smartest companies buying software products, we recommend a simple, logical strategy and process for buying IT systems so that business owners' assets are protected. The process is explained here and illustrated in the next section. The Appendix is a handout that explains why the status quo exists and what to do about it.

The strategy is simply “Require suppliers to demonstrate full data interoperability before you buy.” Put another way “only buy software products from suppliers that will ensure your access to your information assets without them in the future.”

This buying strategy is not very popular with suppliers that like to use closed file formats as a way to control customer migration. However, this brute force method of account control results in many lost customer assets, because migrations are much more difficult. More responsive vendors are competing at levels above file formats, i.e., on product quality, functionality, and customer service.

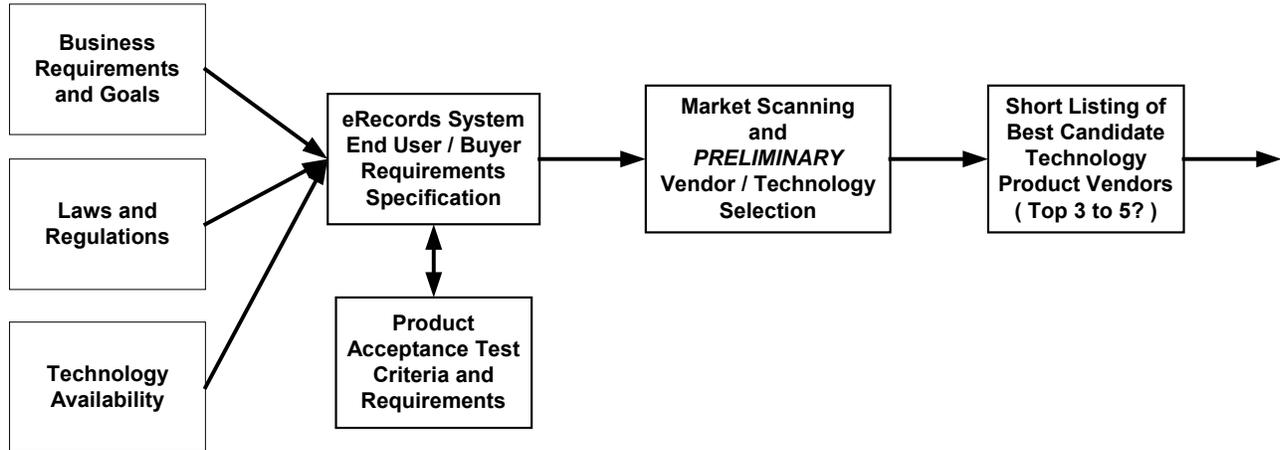
It is important to distinguish between closed, proprietary formats and open, documented proprietary formats. Closed formats are not freely available for anyone to read and write. Proprietary formats are not bad to use for records as long as they are open, documented, perpetually and freely available for anyone to use without license fees. With the advent of XML, and widespread usage of PDF for documentary records, there is no longer any excuse for buying software that does not come with open, documented formats.

A well-designed electronic records archive is a fixed collection of information structured, stored, and secured in such a way that permits easy reconstruction and retrieval in another records management system, whether current or future system. If a system cannot do this, then the system designers not include long-term asset preservation and access as a design center. This decision by designers is either an oversight or made for competitive reasons. To achieve full electronic records ownership and access one must be able to export “Archival Information Packets” that preserve the full content, structure, context, and presentation of records. One must also be able to verify the authenticity of the entire collection of records and individual records within the Archival Information Packet after it has been migrated to another system.

1.12 The Process for Buying IT Systems To Protect Business Assets

The four phases of this recommended process are illustrated below. “Supplier candidates” must pass each phase in succession. This process is particularly critical for Electronic Recordkeeping Systems, but is also recommended for buying any type of data or information system.

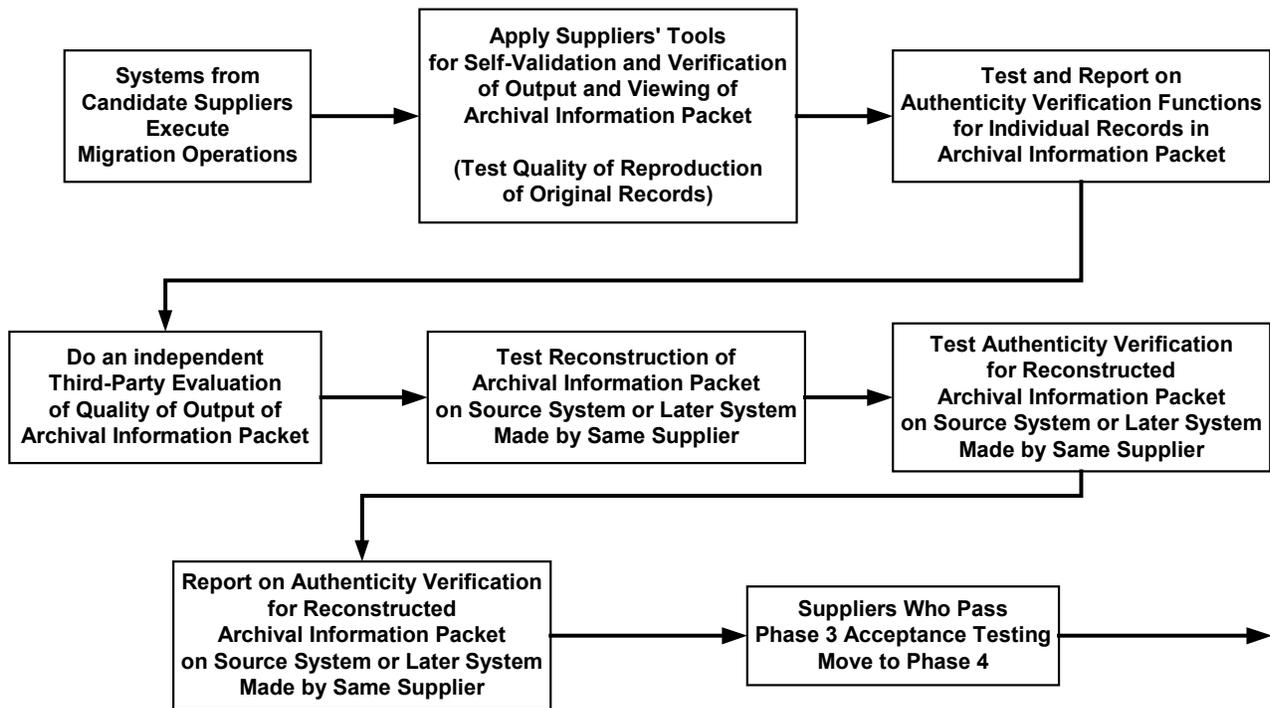
Phase 1 - Articulate Requirements and Find Qualified Suppliers. Be sure to use a User/Buyer Driven Process with Input From Suppliers.



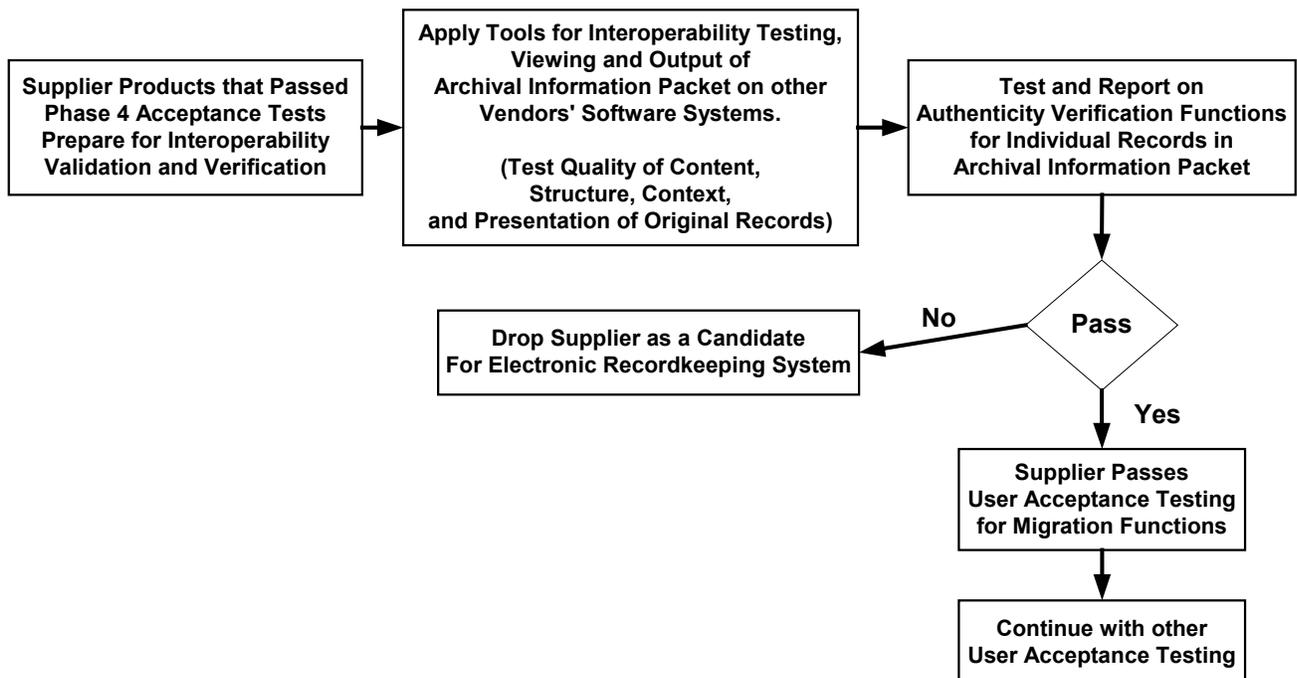
Phase 2 - Work With Qualified Suppliers to Prepare User Acceptance Testing Systems / Environment



Phase 3 - Test Quality of Output of Migrated Archival Information Packets and Authenticity of Records



Phase 4 - Test Quality of Interoperability of Migrated Archival Information Packets



Appendix – Business Asset Protection Program

We have studied the problem of electronic business asset protection and have found it to be primarily the problem that buyers:

- do not make it a priority to own permanent access to file formats for all software product licenses they purchase,
- do not exercise their purchasing power,
- do not take responsibility for migrating their information assets.

The status quo is caused by other reasons too. However, changing procurement practices to include testing and scoring for asset preservation, migration, and interoperability will fix the problems. In some cases, they may be required or decide to pay extra to purchase the rights to permanent access to their assets. In the long run, however, it is worth the cost, because problems with asset loss will be greatly reduced.

It is up to buyers to change the status quo and set up their own Business Asset Protection Program.



The Electronic Business Asset Protection Program

Opening the Road
to Long-term Asset Protection



CENSA © Copyright 2003 – All Rights Reserved Internationally



What's Wrong with the Status Quo?

- Software buyers / users don't fully "own" the data assets they pay for...
- "Asset Access" expires due to:
 - Closed data formats used as a primary method for vendor "account control strategy"
 - Rapid planned obsolescence / change to keep revenue stream flowing to vendors
 - Product retirements or failures
 - Vendor business failures
 - Lack of diligence in migration (symptom of lack of open formats and design for preservation)
 - Subscription license to use software "expires"

CENSA © Copyright 2003 – All Rights Reserved Internationally

CENSA © Copyright 2003 – All Rights Reserved Internationally



What's Causes the Status Quo?

- Computers and software have been designed only to increase productivity ... not protect information assets
- Buyers do not exercise their control before purchasing software and equipment
- Government or Corporate buyers do not think they control evolution of markets
- If no standard format specifications exist, vendors use that as an excuse for closed, proprietary formats

CENSA © Copyright 2003 – All Rights Reserved Internationally



The Goals of using PDF and XML

- PDF provides strong insurance against information asset decay or loss
 - PDF / A gives high-quality, full fidelity rendering
 - Many agencies, laws, countries mandate PDF for recordkeeping and document submissions
- XML provides reusable knowledge components
 - XML DTDs, XML Schemas, specifications, and tools
 - Component, information, semantic models(ontologies)
- XML provides portable information objects
 - moved across applications over space and time
 - easily imported into other suppliers' software
 - Allow full reprocessing of data where needed
 - faster software design and implementation

CENSA © Copyright 2003 – All Rights Reserved Internationally

CENSA © Copyright 2003 – All Rights Reserved Internationally



Buying Behaviors Must Change !!

- Don't buy systems that don't properly comply with ISO PDF / Archive and XML / XSD Standards
- Our laws must mandate e-records system interoperability demonstrations in procurement process
- Do not submit to "urgent" need to buy a system
- Withhold purchasing just long enough to secure interoperability (multiple sourcing insurance)
- Buyers must take back their purchasing power in this renewed "age of accountability"

CENSA © Copyright 2003 – All Rights Reserved Internationally



Planning for The Entire Record Lifespan

- Electronic Records Systems Are Different from Traditional Systems
- If you don't plan (and budget) to preserve it... you won't be able to access it !!
- Thus...
"Design for Preservation"
is a new critical systems design center

CENSA © Copyright 2003 – All Rights Reserved Internationally