
Global Industry Research Team Report

Prepared by Rich Lysakowski, CENSA, and Bill Rhind, Pharmacia Corporation

Acronym Glossary

AIIM	Association for Information and Image Management
ARMA	Association of Records Managers and Administrators
CENSA	Collaborative Electronic Notebook Systems Association
CIPO	Canadian Intellectual Property Office
DoJ	U.S. Department of Justice
EMEA	European Medicines Evaluation Association
EPA	U.S. Environmental Protection Agency
ERLS	Electronic Records Life-cycle Specification
FDA	U.S. Food and Drug Administration
GALPs	Good Automated Laboratory Practices promulgated by EPA
GCPs	Good Clinical Practices promulgated by FDA
GERA	Global Electronic Records Association
GLPs	Good Laboratory Practices—promulgated by both EPA and FDA
GMPs	Good Manufacturing Practices promulgated by FDA
cGMPs	current Good Manufacturing Practices promulgated by FDA
GxPs	refers collectively to GCPs, GLPs, GMPs
InterPARES	International Research on Permanent Authentic Records in Electronic Systems
IT	Information Technology
JPO	Japanese Patent and Trademark Office
LAGER	Legal Acceptability Guide for Electronic Records
NARA	U.S. National Archives and Records Administration
NHPRC	National Historical Publications and Records Commission
NRC	Nuclear Regulatory Commission
OECD	European Office of Economic and Cooperative Development
OJT	on-the-job training
OMB	U.S. White House Office of Management and Budget
Predicate Rule	a regulatory agency rule governing business area(s) to which a subsequent rule applies (e.g., Good Manufacturing Practices take precedence over US FDA 21 CFR Part 11)
PTO	U.S. Patent and Trademark Office
QERPs	Quality Electronic Records Practices
RIM	Records and Information Management
SAA	Society of American Archivists
SLA	Special Libraries Association

UBC

University of British Columbia

UCLA

University of California Los Angeles

Executive Summary

Scope of this Report

This report looks at the situation of the global industry companies in the United States, Canada, Europe, and Australasia. Our discussion of the various record contexts defined by the InterPARES Project generally relates to the regulated industries where intellectual property protection (patents, trademarks, copyrights, trade secrets, etc.) and regulatory compliance (high-consequence health, safety, and potential liability records) are important. The analysis of the individual InterPARES principles focuses primarily on the corporate or R&D records of management and archival groups within the global companies. We reviewed the processes of continuum and life-cycle approaches of records creation, indexing, management, appraisal, preservation, long-term access, and reuse as these practices exist within the industries we surveyed. Our surveys were informal and based largely upon the companies within the Collaborative Electronic Notebook Systems Association (CENSA) and others in the same industries, not just CENSA association members. Specific company identities and survey results are confidential and results and conclusions were made anonymous in this summary report.

For the Global Industry Research Team's work, the initial findings of InterPARES Project 1 relating to authenticity, appraisal, and preservation (as they became available) were constantly assessed against pre-existing juridical (legal), administrative (regulatory), provenancial/procedural/documentary, and technological contexts that are relatively uniform throughout the developed world.

We studied the applicability of the InterPARES 1 findings in more than twelve global industries, including high technology, software, pharmaceutical, chemical, biotechnology, hospital products, medical devices, clinical diagnostics, food, beverage, nutritionals, and oil and gas. They all have common ties within the context of R&D record keeping. They also all have an environmental impact or a direct impact on human health, safety, comfort, and cleanliness, and quality of human or animal life (summarized by the slogan "better living through chemistry").

The State of the Art at the Beginning of InterPARES

The Global Industry Research Team needed requirements, tools, techniques, and strategies from InterPARES to assess authentic electronic records preservation capabilities of current and future systems. The global industry team's primary focus was on very high commercial value records for intellectual property protection, demonstrating regulatory compliance, and historically significant records. However, when we started, no good examples existed of any organization using fully electronic records systems—where no paper was used.

Progress of the Global Industry Team Research

In our investigations, we first spent considerable time understanding the rigorous theoretical background of the global library and archival science communities. We were already familiar with contemporary records management as practised in U.S.- and European-headquartered global companies. We also had to spend a significant amount of time understanding how to adapt and apply the new concepts of *contemporary archival diplomatics* as initially defined by the University of British Columbia group, and modified and codified by the InterPARES 1 group.

Within approximately eighteen months we were able to use the concepts and terminology of contemporary archival diplomatics because we were knowledgeable enough about the new concepts, vocabulary, and processes required for assessing authenticity, doing appraisal, and long-term preservation and access. However, we could not use the formal knowledge from the detailed process models for appraisal and preservation in initial assessments because they were not close enough to completion until the third year of the InterPARES Project.

The global industry team would have liked to do case studies within industrial settings, which would have provided the InterPARES Project some “real-world” examples of industrial systems. However, we found it was simply not possible to find able and willing test sites to be subjects for the case studies. This is despite the fact that CENSA was willing to completely fund the first case study as carried out by InterPARES-trained graduate students with the assistance of Bill Rhind, Rich Lysakowski, or other trained InterPARES investigators. Initial estimates of the time it took academic and government archives to do case studies ran from at least two full weeks to in some case several person-months of work. Frankly, this scared industrial companies and the return on investment was perceived as very small.

Even when we had refined the case study analysis tools to require only approximately two days of industrial company contact time, the time constraints of this amount of interviewing time were prohibitively expensive given the other responsibilities. Rich Lysakowski made serious requests at least four or five times at CENSA meetings for volunteer companies to step forward and offer case study sites. Lysakowski also drafted a “case study marketing brochure” for the case studies in order to outline “What’s In It For Me? (WIIFM)” if an industrial company in CENSA wanted to do a case study. The benefits of the case studies appeared to be simply too small to justify the investment of time, when most of the companies already knew they did not have formal record-keeping systems as defined by InterPARES or CENSA. The answers expected from the InterPARES case studies were in most cases already known as a result of comparison with the more than twenty-six case studies done by academic and governmental institutions.

Executive Summary Conclusions

From extensive surveys within and outside the CENSA membership, we concluded that *in virtually all cases, global industrial companies are not doing fully electronic record keeping where no paper is required*. Fewer than 5 percent of our members claim to be using fully electronic records where no paper is required. The 5 percent that are doing so are in non-regulated areas of the companies. The rest of the industrial population relies on paper, microfilm, or optical image storage for archiving records. In all cases, “hybrid systems” are still being used for the most valuable or consequential records. We found this to be true across the majority of regulated industry, in Europe, Australia, Canada, and the United States.

In general, industry exhibits enormous levels of ignorance and confusion about the formal concepts and definitions of contemporary archival diplomatics or archival science. Most key issues are understood and articulated in colloquial terms, but not formal, rigorous, or consistent terms. All except two of the corporations in the more than fourteen industries we studied did not even employ a corporate archivist, and those two do so only because they are about one hundred and fifty years old and have a large number of physical artifacts of historical significance to the company. These corporate archivists are beginning to worry about what they need to do and are going to do about electronic records.

Benefits to the Global Industrial Team from InterPARES 1 Project

We found the knowledge from the InterPARES deliverables to be very useful in educating industry and even the government regulatory and administrative agencies (besides NARA) on the proper usage of concepts and terminology. The InterPARES deliverables offer a preciseness and rigour that clearly distinguish the concepts and terminology from an archival perspective; they also define what is meant by the legal perspectives quite clearly, and help to eradicate the confusion of terminology that information technology has introduced into the general technical vocabulary (e.g., a record as a legal document, rather than as “database object” or “collection of database fields”).

We incorporated our findings from the InterPARES Project into conference presentations, CENSA documents, specifications, and papers in the various scientific and technical literature, and the Legal Acceptability Guide for Electronic Records (LAGER). The production of the LAGER would have been very difficult had we not acquired a lot of the rigorous foundations from InterPARES 1 Project deliverables and especially Heather MacNeil's book, which served to document many of the key diplomatic principles succinctly for laypersons.¹ InterPARES provided the formality of terms, concepts, and processes that were necessary for people to think more completely and accurately about what needs to be done for long-term preservation and access to authenticated records. These included the InterPARES glossary, the *Requirements for Assessing and Maintaining the Authenticity of Electronic Records*, the initial and refined *Template for Analysis*, the case study analysis tools (CSIP and TEDGI), the many case studies done in other non-industrial domains, and other deliverables from the project.

Our own case study—using the InterPARES methodology for the current paper notebook-keeping process and its extension into future electronic record books—helped us formalize and document our understanding of the process. It also gave InterPARES a concrete high-value example of an electronic system that is used to assemble documents, and creates records from them, but in nearly all cases does not keep them in authentic electronic form for their full retention period. The final electronic data are assembled and either bound in a book or printed to special security paper, signed and date-stamped by an author and witness, and then managed the traditional way, although their location and contents may be tracked electronically within a commercial database or, typically, one developed in-house.

We hosted a special “Legal and Regulatory Symposium” at a CENSA quarterly meeting where Heather MacNeil from UBC/InterPARES attended as an academic expert, and was available to provide commentary on a “mock trial” that we held at the symposium to illustrate the state of unreadiness of industry to deal with the complexities of fully electronic records systems.

As a result of InterPARES participation and the deliverables being shared with these global industrial companies via the various meetings and conferences that CENSA is involved in, the level of ignorance or misunderstanding of terms is beginning to drop sharply for those CENSA member companies that participate. Knowledge is diffusing rapidly out from CENSA members to industry at large. This is an important first step that will facilitate the introduction and usage of the remaining output from InterPARES 1 and InterPARES 2.

True electronic records management systems as defined by the *CSIP* and *TEDGI* are simply not found outside of industries that are not heavily regulated. Although many companies have employed electronic document management systems (EDMS), most corporations still do not use commercially available electronic records management (ERM) software systems. Some EDMS systems vendors have used third-party EDMS vendors' software (PC-DOCS, iRIMS/OpenText, Documentum, Provenance, and others) and ported their ERM application on top of the EDMS and had it DoD 5015.2 certified.

Many EDMS systems in the pharmaceutical industry have been rigorously validated against a predefined system functional requirements document. This includes a rigorous internal audit of the vendor's internal development and documentation practices. Such validation is an expensive process, estimated by the Society of Quality Assurance to cost more than U.S.\$300,000 when all labour and documentation costs are included. Once validated against a rigorous requirements set, such as those produced by CENSA or the U.S. Department of Defense, these systems have been found to be good for the creation and online maintenance of records. However, these validated or certified systems still do not archive records in the formal sense (by achieving long-term preservation and access regardless of the underlying technology changes). Most frequently documentary records are either printed from these systems and signed and witnessed, or the plan is to simply keep “the records” online indefinitely until a good long-term archiving technology

¹ H. MacNeil, *Trusting Records—Legal Historical and Diplomatic Perspectives* (Dordrecht, Netherlands: Kluwer Academic Publishers, 2000).

system product will be developed. There is significant risk inherent in this strategy, because it does not follow the findings of planning for preservation before creating records.

In general, all the industries we studied are heavily regulated by the U.S. Food and Drug Administration, the Environmental Protection Agency, the Nuclear Regulatory Commission, or their international counterparts in Canada, Europe, Japan, and elsewhere. We studied the processes for assembly and submission of product licensing applications (e.g., new drug applications, new medical applications, new pesticide applications) to the FDA, EPA, and Health Canada. The companies we studied are impacted heavily by the records creation, records management, and records retention requirements specified by these regulatory agencies.

We also studied records systems in the same companies from the perspective of submission of patent and trademark applications to government administrative agencies such as the U.S. Patent and Trademark Office (PTO), the World Intellectual Property Organization (WIPO), or the European Patent Office (EPO). All companies considered are impacted, and held back by their desire to protect and sell their products in the U.S. marketplace first (by far the largest and most profitable market in the world for most of their products). As a result, European and Asian divisions of these companies still feel bound by the U.S. rules of “first to invent” that is the patent approval practice of the U.S. Patent Office, rather than the “first to file” rules of the rest of the world. This means that the original records of invention (usually held on paper) must be kept and produced in the case of a patent interference or patent infringement.

Global Industrial Team Findings

We found in our investigations that, except for a few rare exceptions, there are still too many things missing in many contexts (juridical context, regulatory and administrative law, technological context and others) that negatively affect the implementation of the InterPARES Project’s findings regarding long-term preservation and access right now. There is nothing wrong with the InterPARES 1 findings. In fact, they will be very useful for influencing and eventually making or lobbying for specific changes of these contexts themselves. If we had had the full InterPARES results many years ago, the pharmaceutical and chemical industries could have saved billions of dollars in complying with regulations that are now or soon to be law and ignore many basic archival science principles and findings of InterPARES. However, these are now sunken costs that cannot be recovered. At best, we can learn from the mistakes of the past.

One important overall conclusion is that nothing in the InterPARES intellectual framework is at odds with the various contextual aspects affecting record keeping and archiving in global industry. However, currently there are significant gaps between government and industrial policies, laws, regulations, and practices (which are largely inconsistent with InterPARES requirements) and the state of the art of technology and program implementation—as implemented “on the ground” in current in-house technology systems. This means that implementation of InterPARES requirements will be possible only once some remedial measures have been taken by getting regulatory or administrative law changes or additional guidance from key government agencies. For example, 1) we must get additional guidance or implementation guidelines from the relevant administrative or regulatory agencies (e.g., PTO, OMB, OECD, EPA, and EMEA), and 2) where they are at odds with the InterPARES intellectual framework, we must get changes made in various laws or regulations to achieve or allow consistency with the InterPARES framework. Knowledge gained in InterPARES and via our own observations has clarified the best avenues to pursue to start making such changes.

There is nothing in the InterPARES model of the preservation process that industry either doesn’t already do in best cases or could not begin doing across the board, though this will take many years to be commonplace. In general the work of the global industry team in InterPARES has focused on records of value for the purposes of intellectual property protection and regulatory compliance, but the principles will apply equally well to records for product liability, commercial or civil litigation, or preserving corporate or organizational history.

One important realization that the Global Industry Team working on the Legal Acceptability Guide for Electronic Records made was that the organization must plan and budget for, and technologically be ready to convert the record to a long-term preservation format at the time of creation, or else it will have difficulty migrating records to future records formats. This plan and commitment must be done before the first record is created, or the organization will likely run into problems when it comes time to do migration sometime in the future.

The changes required to comply with the InterPARES requirements are starting to take place now as new government laws and regulations are introduced and implemented. For example,

- U.S. *Government Paperwork Elimination Act (GPEA)*.
- U.S. FDA 21 CFR Part 11 Rule on “Electronic Signatures and Electronic Records.”
- U.S. White House’s Office of Management and Budget “E-sign Act”—*The Electronic Signatures in Global and National Commerce Act*, 15 U.S.C. Sections 7001–7031 covers electronic signatures for contracts and documents involving interstate and foreign commerce. In recent implementation guidance advanced by the U.S. Department of Justice, the records retention requirements are included as important to be considered when creating the records.
- Various “e-government” initiatives in the developed world (United Kingdom, Australia, New Zealand, Ireland, United States, and others).

The various e-government laws and initiatives are forcing the hand of government agencies to plan, budget for, and start to implement systems that permit citizen-to-government commerce.

As these new laws and regulations eventually take full effect across the private and public sectors, organizations are grappling to understand and build first-, second-, and even third-generation systems that will be capable of integrated creation, management, and archiving (long-term preservation and access) for authenticated electronic records.

The electronic records software systems marketplace is also beginning to recognize and address the needs of industry for long-term archival preservation and access for electronic records. The Australian government has helped to catalyze the creation of this software market by investing more than four million dollars in developing a conceptual model, software system functional specification for an entire electronic records archive, and an XML/PDF-based file format for permanent electronic records collections. Several commercial software companies and end-users (e.g., CSIRO) collaborated in Australia to build a prototype. The prototype was tested for four to six months, modified, and then put in place as the production version of The Electronic Records Archive for the Victoria State Archives. The Victoria government has declared and mandated that its XML/PDF-based file format for permanent electronic records collections will be relied upon for a minimum of one hundred years. This hundred-year long-term accessibility requirement was one of Victoria’s essential business requirements and design criteria for its system.

CENSA has decided recently that it will piggy-back onto this Victoria government standard approach for the near term, meaning at least the next twenty-five years. The exact approach for “piggy-backing” is still being decided upon, but we expect to finalize our technical strategy and announce it by the end of 2002. We are working with U.S. NARA and Australia’s Victoria State Archives to ensure that their long-term strategies are functionally compatible.

Assessment of InterPARES Principles Against Industrial

Contexts

In this section, the InterPARES principles as defined in the report of the Strategy Task Force are first considered as a set, and assessed against the various contexts that define the record-keeping (creation, management, and preservation) regimes within regulated industries where

intellectual property protection is important. Where appropriate, individual principles are then assessed against those contexts as well.

In many cases we found that industry has identical concerns as government and academic institutions. However, in a few cases, the emphasis of concerns of industry mandate faster for less expensive solutions to an immediate problem, sometimes because of a regulatory or legal mandate with a deadline that industry must meet regardless of whether government or academia has the same deadline. In government and academia, it is not a primary concern to meet environmental, food, drug, or medical-device product licensing regulations, or achieve rock-solid intellectual property or patent protection. In the case of government or academia, other government or institutional acts or policies—such as right-to-privacy, freedom-of-information, the *Government Paperwork Elimination Act (GPEA)*, or e-government initiatives—become primary concerns the institution must pay attention to implementing.

Principles Overall—Findings in Critical Contexts

Juridical/Administrative Context

This is the key context defining the possibilities for successful implementation of the InterPARES findings. The juridical and administrative elements must be dealt with separately, as their effects and impact are not uniform.

Juridical Context

U.S. information legislation (The U.S. courts' *Federal Rules of Evidence*, the U.S. government *Paperwork Reduction and Paperwork Elimination Acts*, the *E-Sign Act*, *Freedom of Information Act (FOIA)*, the *Privacy Act*, and regulatory rules by the FDA, EPA, NRC, PTO, and other agencies) provides a strong overall framework within which the InterPARES principles could be implemented.

A huge problem that exists is that U.S. regulatory agencies, the U.S. Code of Federal Regulations, and U.S. congressional code define *records* and *electronic records* to be *any* information in *any* format that is stored for later evidential, business, or historical purposes. They thus equate with records, *all* evidence or data of any type created by anyone anywhere within the business. They also do not associate records with the business processes they relate to, nor do they include the archival requirement of the record to be “fixed and set aside under the care of a qualified custodian with the responsibility of ensuring the ongoing authenticity of the record.” There is no measure of the quality of the evidence that defines whether something is a record, such as author’s and/or witness signature(s), or the custodial care or evidence about authenticity that is generated during custodial care.

Equating all data into “records” turns retrieval for litigation and criminal prosecution purposes into an evidentiary discovery “circus” whereby intellectual property attorneys, regulatory inspectors, product liability litigators, and other attorneys can treat any data or information as though they were all records created and stored within a record-keeping system, when in reality they are simply just evidence used to satisfy the data collection needs of the business. There is so much uncontrolled data/evidence floating all over corporations and government agencies that it is nearly impossible to control it all right now. E-mail is the worst/best-case example that gets cited repeatedly as poorly managed evidence that the creators don’t consider as records, but litigators absolutely have a field day using.

The U.S. *Federal Rules of Evidence* provide specific guidance about how electronic records may be admitted as “hearsay” into court proceedings. There are industry practices and standards promulgated over the years by the Association of Records Managers and Administrators (ARMA), Association for Information and Image Management (AIIM), Society of American Archivists (SAA), and Special Libraries Association (SLA), but no firm single set of rules exists in the United

States like the Canadian and Australian governments' Requirements for Management of IT, Security, and Record Keeping Metadata Requirements.

Because of this lack of common detailed guidance from either the U.S. executive branch of government (OMB), or detailed NARA directives, various administrative and regulatory agencies create their own regulations to fill in the gaps. This is not a fault of NARA, which has simply chosen to work at higher levels of policy setting and let individual regulatory and administrative agencies develop detailed directives that work best in their own agency. Some of these agency regulations are performance-based and less murky about specific requirements.

Because of the inconsistencies in scope, policies, standards, and knowledge across the set of regulations, laws, guidances, and recommendations issued by the OMB, NARA, the U.S. Department of Justice (DoJ), FDA, EPA, PTO, and other agencies, and the actual record management and archival practice in corporations—at least in the global industry team members represented in CENSA—CENSA had to create its own standard guides and specification and knowledge transfer symposia to educate at least eight audiences that explicitly have electronic records among their concerns and responsibilities. These audiences are:

- end users
- attorneys
- regulatory affairs
- quality assurance
- executive and R&D managers
- records managers
- archivists (where they exist)
- information technologists.

Other functions within the corporations are just now beginning to be examined (administrative, financial, personnel, and other types of records).

Administrative Context

The overall administrative context of industrial work is conducive to the implementation of the InterPARES findings, given the presence of record management policies of NARA, though they do not apply directly to industry. Nearly all government regulatory agencies have recommended retention schedules. Most industries have specific records retention schedules imposed upon them by industry regulators, or they have general guidelines on how to determine these retention schedules. Good businesses have implemented and regularly update their retention schedules to achieve better business record keeping in support of managing their intellectual asset portfolios.

Other than the FDA and EPA regulated industries that implement the Good Laboratory Practices, Good Manufacturing Practices, or Good Clinical Practices, and the Nuclear Regulatory Commission (NRC) Recordkeeping Guidelines, global industry does not currently have effective standards for the entire records management and archival regimes. *Many* initiatives and policy creation efforts are now under way to address this problem, but it will be many months to years before widespread creation occurs for management policies, information technology, and standards operating procedures (SOPs) to support a general adherence to the principles enunciated in the InterPARES framework. That is one reason the InterPARES work is so important to CENSA from both industrial and international governmental perspective.

CENSA and the Global Electronic Records Association (GERA) wish to base their standards, guides, and tools upon the most rigorous and best time-tested principles of archival science and records management. The worst that could happen is that uninformed and ignorant regulatory or administrative agencies promulgate inconsistent laws and regulations without this rigorous knowledge, resulting in massive efforts to interpret what poorly written regulations really mean, filling in the gaps to include what they *think* is missing—without direct specification and input from the regulatory or administrative agency—and then building systems that may or may not comply with a rule that was poorly written in the first place. The net negative economic result is that it costs business hundreds of millions or billions of dollars to comply with such rules, and the costs get passed along to either consumers or taxpayers.

In order to address the needs for rigorous, consistent tools for electronic records program implementers “on the ground” (not simply corporate-level policy makers), CENSA spun off the non-profit research institute GERA. Its mandate includes facilitating interaction and collaborative true research, authoring, and standards work with government regulatory and administrative agencies, academia, and non-profit organizations that are stakeholders in establishing quality electronic records programs. The purpose of this institute for electronic records is to produce standards called the Quality Electronic Records Practices (QERPs) to assist implementers in knowing the detailed requirements (we call them “80/20 blueprints or templates”) of what electronic records programs, systems, and components to build, as well as how to validate and audit them; and legal defence procedures and case law with recommendations on how to effectively litigate using electronic records. The “core” deliverables include professional education and training materials to teach professionals about the QERPs. In addition, the goal is to put these deliverables in an online, integrated knowledge base that supports records managers, archivists, cyberterrorist response teams, and others who must use electronic records over any stage of their life cycle. We will continue to update and feed key knowledge from InterPARES and other sources into these standards as it becomes available, fully acknowledging the origin as we integrate content from InterPARES and other sources. The work of GERA centres on the R&D, creation, publication, and dissemination of these standards (and maintenance when the time comes.)

GERA (via a jointly funded CENSA/NHPRC project) is developing a general-purpose life-cycle standard that includes a framework model covering all basic program design, policy, procedural, personnel, technological, and other programmatic and technological aspects of QERPs across government agencies and industries. The QERPs are a set of integrated documents that include how to validate and litigate using records drawn from a system implemented against the requirements identified in the Electronic Records Standard Lifecycle Specification. However, it will take at least three more years from the time of writing before all of the standards are done. The e-Records Lifecycle Specification (ERLS), the Validation Guide for Electronic Records Systems, and the Legal Acceptability Guide for Electronic Records (LAGER) were to be finished by October 2002. The Auditing Guide for Electronic Records; the Certification Guide for Electronic Records Programs, Systems, and Archivists; and some other companion guides and tools were to be created after October 2002 as funding is secured from various federal and corporate sources. These QERP deliverables—as they are applied—will help to prove out the concepts, terminologies, and process models defined in InterPARES 1.

As a point of emphasis, at the time this report was written the global industry team badly needed InterPARES 2 to produce an integrated life-cycle model spanning the entire electronic records continuum of electronic records creation, appraisal, maintenance, long-term preservation, access, and reuse. It was especially hoped that this would be ready in time for version 1.0 of the ERLS global standard. If the integrated InterPARES model was not ready by mid-2002, we planned to use what was available as part of the ERLS draft global standard.

Provenancial/Procedural/Documentary Contexts

In regulated industries, the provenancial context is established by default, rather than by design, as a result of the implementing regulations and complying with law, for example, complying with FDA 21 CFR Part 11 on electronic records and signatures. The EPA Good Automated Laboratory Practices, the FDA current Good Laboratory, Clinical, and Manufacturing Practices.

Detailed documented procedures or performance criteria must be specified and tested before putting data, information management, and record-keeping systems for labs, clinics, and manufacturing processes into “production” usage. The systems must be “validated” to ensure compliance with the predicate rules of the agency. In these cases, the procedural context is linked to the record creation process that is linked closely to the business functions it supports. This appears to establish the *provenancial/ procedural/documentary contexts*.

The White House’s Office of Management and Budget—*The Electronic Signatures in Global and National Commerce Act*, 15 U.S.C. Sections 7001–7031 “The E-Sign Act” covers electronic

signatures for contracts and documents involving interstate and foreign commerce. In recent E-Sign "Implementation Guidance" advanced by the U.S. Department of Justice, records retention requirements are clearly spelled out to be included as important when creating the records, and the guidance makes it quite clear that it does not just cover e-commerce records but all records resulting from government transactions with citizens and businesses. The implementation of E-Sign will affect all government agencies and their interactions with industrial companies, especially regulatory and administrative agencies, and commercial transactions.

Technological Context

A technological context appropriate to the realization of InterPARES principles exists (the many commercially available records, documents and information management system products—an integrated suite of softwares available to all corporate, government departments and agencies). Many of these systems have been implemented and validated, but these are not formal record-keeping systems because they are missing the capacity for long-term record keeping. These commercial systems' records and information management (RIM) needs to be more fully implemented across corporate and government enterprises. Our survey of thirty-eight corporations and at least sixty-five federal agencies indicate that EDMS systems are starting to be implemented widely, but in nearly every case, records for archiving are being kept on paper.

Separation between the information technology and the records, and RIM responsibilities and reporting structures within global corporate institutions, has caused many problems for records managers and archivists. This frequently leads to RIM (especially archival considerations) *not* being inadequately reflected in IT implementations. This also raises a more important concern: the view that records managers or archivists are filing clerks, and should not be considered in the design of all new IT systems. This is a huge concern, particularly with what are proposed are "the full implementations of RDIMS [records, documents, and information management systems]."

The only government that has accomplished the research, engineering, construction, prototype, fixing problem, validation, and production implementation of technology systems for full electronic records management and archiving (ERMA) systems is the Australian government with the Victoria Electronic Records Strategy (VERS). They are also constructing a policy framework to permit the cross-government users of the VERS software, which will round out and complete the fully electronic records archiving system, by providing the requisite policy and program guidance. Various other state governments within Australia have produced policy frameworks that support full implementation of electronic records programs and systems.

Individual Principles Extracted from the Research

We generally concur with the Canadian team's principles extracted from the research. We made some minor modifications to fit some broader, immediate needs for quality electronic records practices.

However, in general, any records preservation policy, strategy, or standard should:

- address records specifically rather than digital objects generally.

This principle is not explicitly reflected in the overall global industry or government context, but is directly supported by various local agency rules such as FDA 21 CFR Part 11, SEC rules, EPA GLPs and GALPs, OECD Rules for Electronic Records. In the United States, the most comprehensive support is the *Government Paperwork Elimination Act (GPEA)* and the E-Sign Rule, which will have sweeping effects throughout industry and government.

- focus on authentic electronic records.

Several industrial initiatives are being driven by regulatory and legal concerns on product and environmental quality and liability. The long-term preservation and access to electronic records is emerging as a by-product of new industry regulations, e-government initiatives, and last, but not least, the need to improve business productivity overall in all types of businesses, including

government, industrial, and academic. Regulatory agencies needed to speed up the product application safety review process, which could take as long as seven to ten years, cutting years off the useful patent protection period at the end of a product's life.

This resulted in regulatory agencies starting to specify electronic records business requirements and regulations, but *without the restraint that comes from informed knowledge of the current state of the art of information technology and the total lack of design of IT systems to meet even near-term (25+ years) records retention requirements*. Thus the agencies specified the full length of time of product research and development studies plus the full amount of time that products are sold on the market.

Among the best-known U.S. industry regulations that affect industrial companies worldwide that wish to sell their products in the United States is the FDA 21 Code of Federal Regulation (21 CFR Part 11) rule on electronic signatures and electronic records.

Originally, the pharmaceutical industry asked the US FDA for guidance and regulations that would permit fully electronic submission and review processes in lieu of paper applications processing. The FDA's response is that it also wanted to be able to inspect original data and records to ensure that required quality levels are met. This requires companies to retain all supporting records and source (raw) data and make them available to inspectors during audits. The FDA equated all data to records, whether or not they were of a quality capable of being preserved or accessed in authentic form for long periods of time, and whether or not they were given to a trusted party with a duty of care for preservation.

In fact, though the FDA 21 CFR Part 11 rule is about retention of authenticated electronic evidence, it is written in a language that emphasizes "electronic signatures" and "electronic records." The FDA took licence and mistakenly equated all computerized data to "records" as defined by the U.S. Congressional Code (44 U.S.C. 3301), which reads "All books, papers, maps, photographs, machine readable materials, or other materials, regardless of physical form or characteristics, made or received by an agency of the United States under federal law in connection with the transaction of public business and preserved or for preservation by that agency or its legitimate successor as evidence of the organization, functions, policies, decisions, procedures, or other activities of the Government or because of the value of the data in them." Building upon this U.S. Congressional Code was the U.S. Code of Federal Regulations 36 CFR 1234.1, which defines *electronic record* as "Information on any electronic medium that can be read by using a computer or any other electro-magnetic device that satisfies the definition of a Federal record in 44 U.S.C. 3301," which in turn served to equate all electronic evidence to electronic records. In the broadest sense this may be true, because any data may serve as a record and be used as evidence in court or for historical purposes; in practical everyday terms, however, it erroneously gives data and record creators the duty of care for long-term preservation and access normally given to trained records managers and archivists. Most data and record creators are improperly trained, and are not given the responsibility, budgets, or the job spans to take on the long-term duty of care for their data and records. Strictly speaking, even if records creators were given responsibility for long-term care, this violates the principle of good accounting practices that one should give responsibility for records maintenance and auditing to a trusted third party. Thus, a few oversights made in this historic rule that affected FDA-regulated industries.

The archival community and records management was not given the opportunity to intervene, or even have significant influence to develop criteria or standards for archival electronic records that include "capable of being preserved in authentic form for as long as needed by the business." There was also the practical matter that not enough technology standards existed yet to make long-term preservation via migration mechanisms facile or straightforward.

FDA-regulated companies were unsuccessful in presenting an effective financial justification case against the FDA demonstrating that the FDA rules, as written, simply could not be implemented without extreme cost and undue burden. FDA-regulated industry spent several years just to interpret the rule, and is still working hard to comply. The FDA Part 11 rule simply left too many details of records management and archival science unclearly specified or simply unstated. From

the time the rule was proposed (early 1992) until it was finally implemented as law (early 1997), industry *and government* were just beginning to understand the full cost of implementing basic electronic records creation, records management, and the implications of long-term records preservation and access.

Many people in industry believe the FDA regulation simply violated the *Government Paperwork Elimination Act (GPEA)*'s goals for reducing paperwork burdens, and that it caused undue financial burden on the companies forced to comply with record-keeping requirements. However, the FDA-regulated industries failed to respond fast enough or early enough with compelling financial research results to show to the contrary that, while the goals to accelerate electronic submission and review process requirements were reasonable and achieved, the records retention burdens required retrofitting or in many case total replacements of otherwise modern equipment. By late 2001, it was estimated that the 21 CFR Part 11 compliance costs for large FDA-regulated companies was already between US\$100 million and US\$200 million *per company*. This translates into billions of dollars industry-wide, just to get compliant, and then additional costs for maintaining more complex record-keeping systems, in addition to information management systems used to gain competitive advantage.

NARA guidance on electronic signatures and electronic records retention for GPEA purposes did not appear until October 2001, and by then the pharmaceutical industry had already spent huge sums of money to comply with FDA 21 CFR Part 11. It was not until 25 September 2000 that the U.S. Department of Justice provided guidance to help with all federal agencies' interpretation on GPEA with regards to long-term electronic commerce transactions.

The U.S. Patent and Trademark Office is a branch of the Department of Commerce. The Patent and Trademark Office applications can be viewed as long-standing commercial licensing application transactions (records) that are covered under E-Sign.

The recently proposed EPA Cross-Media Reporting and Recordkeeping Regulation (CROMMERR) may also be in danger of trying to do too much too fast, trying to combine *GPEA* reporting requirements, while at the same time imposing internal records retention requirements on the entities that it regulates (estimated to be between 1.2 million to 1.7 million entities.) The public comment period for this rule ended 27 February 2002. Many companies in industry have asked that the CROMMERR rule be withdrawn and evaluated against the current state of readiness of regulated entities and technologies to actually implement the rule.

By 2002, the Japanese Patent Office, the European Patent Office, and the U.S. Patent and Trademark Office (PTO) online electronic submissions processes had been in place for more than ten, exactly five, and less than one year, respectively. All of these patent offices recognize that they do not yet have supportable and supported long-term retention and access systems for fully electronic records, even though they accept the submissions in fully electronic form, complete with the patent agent's digital signature. Patent and trademark submission regulations cover all parts of the submission form. By 2003, the PTO wanted to have a fully electronic application submission process; it has already achieved that goal. The challenge remains on the back-end long-term records retention and access systems that the PTO will use in-house.

Challenges exist for both the sending and receiving parties involved in the patent application submission process. For example, the PTO is required to retain submissions online as records, in secure and confidential format, *for a minimum of forty years*, before final disposition occurs, where the records are either submitted to NARA or destroyed. NARA has admitted openly that it does not know how it will meet these retention and access requirements, but that it is working on a solution. On the other hand, the submitting organization must keep records from the moment of invention and reduction to practice that can go back as far as two decades earlier. Thus, the submitting organization in an FDA-regulated industry must preserve evidentiary records of invention and reduction to practice that may need to be kept for as long as twenty-seven years or longer. Drugs can be granted up to twenty years of patent protection; some records of R&D must be kept for seven years beyond the last date of manufacture for pharmaceutical products. For medical devices embedded in a patient at an early age, the retention periods can last from the

time of invention, and then for the life of the patient and several years beyond in case of class-action product liability lawsuits, potentially many decades.

The FDA, PTO, and EPA initiatives are being watched closely by equivalent agencies outside the United States because frequently these agencies set the pace for product safety and environmental protection in many other parts of the world, and the United States represents such a profitable market that many companies outside the country want to meet FDA or EPA requirements in order to be able to sell products within the United States.

Overall, the problems of long-term electronic record preservation and access remains unsolved for both parties involved in submitting and receiving patent applications, and in trying to defend or attack intellectual property rights if infringement or interferences occur; those entities submitting or receiving product licensing applications; or those demonstrating that they “did the right thing” with potentially hazardous chemical or biological materials.

- recognize and provide for the fact that authenticity is most at risk when records are transmitted across space or time.

This fact is certainly recognized in the policies of the PTO, FDA, EPA, Canada’s public key infrastructure (PKI), and other regulations worldwide. However, appropriate PKI initiatives are underway to help guarantee the transmission of records between agencies and to national archives in many countries. Industry will be forced to follow the lead of local national governments in adopting the rules and practices for digital signatures, identity cards, and public key infrastructures. There is still much resistance in the United States and elsewhere to universal electronic identity cards that can be used for electronic signatures.

- recognize that preservation of authentic electronic records is a continuous process that begins with the process of records creation.

This requirement is reflected in the record-keeping guidance that CENSA provided to its members in the “Legal Acceptability Guide for Electronic Records.” This view is consistent throughout global industry, the National Archives of Canada, NARA, Australia, and elsewhere. This unites the life-cycle and continuum viewpoints of electronic records, rather than invalidating either one. It also provides guidance to government and industry that govern the transfer of electronic records to the custody of an institution’s records managers or archivists (commonly called “terms and conditions of transfer”).

Industry eventually came to realize that records must be converted *at the time of creation* to a format that facilitates preservation—but only after observing the inability of many projects to access “records” in systems that were never designed for preservation in the first place. An enormous expense is required if you do not plan for preservation before you create the record. We found a significant number of cases where IT groups designed systems or vendors sold systems that did not provide sufficient documentation of file formats, software interfaces, or export processes to permit the migration of critical assets. All too often this fact was not discovered until a system was obsolete, really ready for the graveyard, and its records were needed for pending litigation. The result is that a lot of money and time were spent resurrecting systems to extract their records and then trying to prove those records’ authenticity.

The Global Industry Research Team findings strongly support this recognition that one does best to plan for preservation before the first record is created. CENSA has been promulgating this view through its many meetings with industry software implementers.

- be based on the . . . concepts of a trusted record-keeping system and the role of the preserver as a trusted custodian.

CENSA and its global industry members achieved consensus that “custodianship” is the central theoretical concept that fulfills the requirement for a trusted record-keeping system, but actual current implementations routinely fall short of fully ensuring records’ authenticity—sometimes because records are left under the control of their creators, sometimes because trained *electronic* records managers and archivists are not consulted when systems expected to keep records are

being specified and procured. These gaps will be addressed in the coming years with increased training throughout industry, government, and the archival and records management communities. The role of trusted custodianship will grow as organizations learn more about it, but also experience failures and losses from not implementing it properly.

- be predicated on the understanding that it is not possible to preserve an electronic record as a stored physical object: it is only possible to preserve the ability to reproduce the record.

The InterPARES researchers fully understand this principle, and its preservation strategy focuses on maintaining the ability to reproduce records in authentic form (while still accounting for the need to maintain the various physical carriers on which records are stored over time).

- recognize that the physical and intellectual components of an electronic record do not necessarily coincide and that the concept of digital component is distinct from the concept of element of documentary form.

The former point is generally recognized among those who deal with electronic records, but it is highly unlikely that the diplomatic concept of “element of documentary form,” *per se*, will figure in industrial or government policies, strategies, or standards unless or until these diplomatic concepts become commonplace via more widespread adoption and education.

- specify the requirements a copy of a record should satisfy to be considered equivalent to an original.

In the case of the industry and the court systems with which industry deals, the requirements to be satisfied will relate primarily not to the copy itself but to the system used to transfer, store, and maintain the record’s digital components and to reproduce the record. But the system used will only be selected once it has demonstrated that it can be used to reproduce accurately both the content and documentary form of the record. Courts must evolve their preference for paper as the “best original form” to include reproduced authentic electronic records as the best original form.

- integrate records appraisal in the continuous process of preservation.

Appraisal for both economic or historical value is done formally in only the best industrial organizations. Appraisal methodologies must play an increasingly important role in industrial records and archives management. This would be facilitated if somehow processes for record classification, archival descriptions, and appraisal could become more closely integrated and more automated, because archival description and appraisal are all time-consuming processes. Again, these must be viewed as continuous processes that are integrated into a continuum operating model for preservation.

- explicitly recognize that the traditional principle that records all relied upon in the usual and ordinary course of business can be presumed to be authentic needs to be supplemented . . . by evidence that the records have not been inappropriately altered.

The information management policies do not always explicitly include such recognition, nor recognition of the purpose and value of audit trails in proving both reliability and ongoing authenticity of electronic records. It must be put into system and policy requirements that archiving audit trail records (in preservable, authentic, accessible form, of course) should become an institutional practice to document the continuing authenticity of the institution’s electronic record collections over time.

- recognize that the preserver is concerned with both the assessment and the maintenance of the authenticity of electronic records.

An archival group must assess the authenticity of records before they are transferred, and be responsible for maintenance of their authenticity as a critical part of preservation activities following transfer into the archive.

- draw a clear distinction between the preservation of authenticity . . . and . . . authentication.

It is not certain that this distinction is well understood in industry or government generally, but it is recognized by the groups participating in InterPARES. It is reflected in some local country guidelines for records created in a PKI environment, but not globally in countries or companies that do not rely on PKI or electronic national identity cards yet.

Effects on Stakeholders

In short, the primary benefit will be that meeting the benchmark authenticity requirements will substantially improve the quality of record-keeping processes throughout government and industrial institutions, whether local, state, federal, or global. The principal direct effect of the implementation of the InterPARES requirements in the global industry context will be much stronger archival records in electronic form, with a more explicit focus on determining the reliability and authenticity of the records being created and transferred to the corporate archival repositories, and greater confidence by companies in asserting their authenticity when reproduced. The principal indirect effects will be considerable, both industry-wide and government-wide. The costs and the benefits of implementation of the record-keeping infrastructure necessary to meet the benchmark authenticity requirements must be better understood. Right now these are understood qualitatively as “better current and future records, improved accountability, and increased access and reuse of records.” Some administrations, agencies, or corporations that wish to keep their operations covert view this heightened quality and accountability as a liability. However, this must not impede those entities with a desire for better quality and accountability.

Conclusion

Everything in the InterPARES principles can be incorporated into policies, strategies and standards at the international level at which the global industry team works. There are still significant administrative, legal, regulatory, and technological barriers, but these are not insurmountable. However, it will require completion of the InterPARES life cycle or continuum process models and their incorporation into the Quality Electronic Records Practices and other standards, their worldwide promulgation, followed by extensive training, implementation, and certification. Many of the remaining administrative, legal, regulatory, and technological barriers will need to be removed by legal, regulatory, government administrative law, and technology market development processes. These activities are simply outside the scope of an international collaborative research project. InterPARES should retain its focus on research, codification, and documentation of key concepts, definitions, and processes from an integrated viewpoint of records management and archival science for electronic records.

In general, the current situation and “digital gap” is because of the poor understanding of formal records management and archival science by the masses, the subservient position in which records managers and archivists are put relative to information technologists, and their strong dependence on information technologists for system implementation. There is also a serious lack of training of information technologists and computer scientists to include “design for preservation” for authenticated information as an important design centre for current and future systems. Dependencies on IT may force records managers and archivists to simultaneously become “IT-savvy,” become good team players and negotiators on IT systems requirements analysis and specification teams, diplomatically bring the attorneys to their rescue when necessary, become good electronic records systems project implementation managers, and become good at financial justification and business management. They must become truly multidisciplinary themselves. This is not unique to any one country, but is a worldwide phenomenon. As the need for “bridging the digital gap” continually grows, multi- or cross-disciplinary professionals will emerge via on-the-job training or via formal university or post-professional training and certification programs. This process may take another five years to complete, because it is truly required to complete the paradigm shift from paper to electronic records.