SAA. *63rd Annual Meeting*, held August 23-29 1999, Pittsburgh, Pennsylvania.

Rich Lysakovski. No title on the tape recording

My presentation illustrates one the important things about modern technology, is you have to have gurus and experts on hand and all times to ensure that they work and continue to work. I will talk about industry's participation in InterPARES. It is a pleasure for us to work with the InterPARES group. This is the problem that we are up against and this problem exists in many places, in academia, in government, and in industrial facilities you find examples like this everywhere. It is a major problem. This is the traditional data and records management system for research and development; it contains records for intellectual property protection, demonstrating regulatory compliance, protecting trade secrets. If you look inside of them, the system is completely broken. You can see the book itself is broken, the spine iis completely broken, you open up the pages to microfilm it, and it falls apart. You will find structures, photographs, spreadsheets, script charts, printouts, and even chemical samples in little buggies stuck between the pages. Now electronic records are not going to fix that, but for that we have automated sample handling systems. There are so many problems with this picture that it would take me a half an hour to explain them all to you. But it gives you the right image. What we want to do is move to something that on a work station or on a desktop PC looks like this. It is a page or entered metaphor for an interface you interact with it like you would on a normal page. What you see on that page are digital objects that have been pulled in from many different sources, literally hundreds of different sources, they are by directional connections; that is until you go and commit this compilation as a record and then the by directionality may still be there, but it is only for pulling data out and possibly reusing that data to create yet another record. But there are connections to... I have only shown six different sources of information here including the authentication tools, but if you add up, Park- Davis Warren Lambert gave us a great example, they said every year in research alone they support four hundred different applications and every year two hundred of them change. So maintaining ongoing integration of all those data sources with the notebook is a major challenge. But the facility it affords you to have it fully automated is very great. We see an average of about $33.000 recoverable time alone pro notebook from our

workflow analysis and cost accounting studies what time could be saved. That does not include the collaborative benefits of the electronic form of the notebook. I get asked, when people ask me what I do for living I say electronic notebook systems and often think of hardware, like these things. These are great steps in the right direction, but at most these will only ever be data collection systems. It is that backend record keeping and archiving systems that really make the results credible in a court of law. But, as you can see, this was a great commercial success, because it has sold two million in the first year and a half, it is evolving to systems like this that include microphones, cameras, wireless connections, touch screens, pen computing input and so on. And this is what we will see in the future, where these things are one pound or two pounds and they can do everything that a paper notebook like this can do and more. If you go near future, we are talking next year or the following year out through the end of the next decade, we will see wireless hand-held pen computers with voice video, full handwriting to text translation, voice to text translation, infored data transfers to informed instruments, intelligent agents doing things for the researcher and probably to the researcher, ubiquitous record keeping, so that you can go anywhere on the planet and work on something and commit it to backend archive that will keep it for as long as the business needs it; and then finally full interoperability between our video recorders, our electronic notebooks, cell phones, you name it. It is an exciting future.

So, what is CENSA? CENSA is a new type of industry association, or trade association that is focused on market development. We are not a professional or standards organization, we really just focus on getting the right products on the market, much faster than by normal market processes, but we are not vendor ourselves, we work with end users and suppliers to have the end users articulate their needs and the suppliers to build the right products; and the end users work with the suppliers to ensure that the right products are indeed being build and tested before they hit the market. We have three problem domains that we focus on: intellectual property protection, demonstrating regulatory compliance with these records, and improving the markets themselves in terms of making component based architectures that allow us to rapidly reconfigure the R&D automation environment to support new type of research and development. Current end user members listed here, we are constantly recruiting new members and there are some other multinational companies that will be joining us very soon. Suppliers, we have about nine different suppliers, some of them you will recognize. We do need more records

management and archiving suppliers to join us.
So why InterPARES? Well, we have particular problems we want to deal with, our focus is on the notebook itself. The notebook, if you look at the attributes of a paper notebook, is a very high speed multimedia device; you can flip through it very quickly, you can put just about any data type, except dynamical ones into it; it is a very sophisticated device that has evolved over the last four hundred years, replicating all the convenience and use of views and portability of the notebook, but preserving its utility as a recordkeeping device; it is difficult, it is a challenge, major challenge; it is also the R&D reports, which themselves are multimedia compound documents; and then, dealing with more general multimedia databases. These are the three types of records and systems that we are focusing on in InterPARES.

Our reasons for working in InterPARES are two: work as closely as possible with the leading specialists and experts in electronic recordkeeping and archiving systems; and to ensure that research coming from it can be applied as soon as possible; to make sure that the theories and practices that are developed are as grounded as possible so that we can put them into commercial systems as soon as possible, I mean that is really an urgency for us. You saw the overstuffed notebook, we can live with that anymore. The other thing is that commercial software systems often do not have the rigor. They were not designed for recordkeeping, they are designed to be information systems and that leads to the last bullet here, which is to transform IT itself into technology that is designed for a long term, designed for preservation; it has the necessary subsystems for migration, for audit trailing, for authentication on an ongoing basis. And another nice feature, you push a button and all the records come out in a vendor neutral format that are still authenticated and can be pumped into the next generation system with no programming required. That is when we will know our job is done as far as translating the results from InterPARES and other projects into commercial systems.
What we will be providing to InterPARES is a set of test beds to test hypotheses, theories, models, and other things that come from InterPARES. Also to do proof of concept, models and systems. We have to start with existing commercial technologies, because that is where we have already invested in, in industry. So, things like XML, PDF, digital signatures, PKI, how do you combine them in such a way as to achieve the objective of long term preservation and access? And then different models and models specifications for the metadata, the objects, the business

processes and broader definitions of systems that include the policies, the procedures and the technologies together. We want to make sure that the InterPARES group is as informed as possible about the state of the art in industry, because it is really important, and one thing that they will tell you is that we have been trying to drive the group to meet our priorities. We need deliverables yesterday, and it changes the nature of the project a little bit when you have got people that are constantly nipping at your heels to get deliverables; that is, we in industry have managers constantly nipping at our heels: "All right; why are you going to this meeting? where are the deliverables from it?" You know, we have to address those concerns. We have a broad program to apply the results. First of all make sure that industry understands the vocabulary and the language of records management and archiving. Broadly through our IT and R&D management; they do not. When we first started, they really did not understand the language at all. We would say record, they would say database. So to both develop new terminology and concepts and clarify and translate for people that are not in the specializations. Collaborating with the government organizations that are in InterPARES is also a key benefit and taking the learnings that are going on in places like San Diego Super Computer Center, on behalf of the Patent and Trade Mark Office and NARA, these are invaluable to us, because the super computing problems that they are facing today will be industry's desktop computing problems tomorrow. And then, ultimately we hope to apply the learnings from InterPARES into commercial systems. And the last thing is to take the various learnings and turn them into training programs, that can help bring up the existing generation of people in industry to understand and interact with the next generation. We also see the InterPARES expanded group it is probably close to hundred, hundred twenty people, that this is a talented pull of individuals, and some of them want to go to work in industry. There are a lot of different professions within industry, so we have many openings, and we have more and more. The projected shortage of IT professionals by two thousand and five is over a million and we do not want just IT professionals; we want multi-disciplinary people, who know IT, they know science, they know records management and archival science. And that is the real beauty of InterPARES project for us; is that is multiple disciplines and there are indeed many multidisciplinary individuals working within InterPARES. People who have been employed in records management and archiving positions for decades and they have been using computers for much of that time. This people are really hard to find. So, when we looked at InterPARES to supply a steady stream of

educating individuals who understand the, I call this, the next grant challenge of science. It is as big as a grant challenge and is at least a hundred to a thousand times larger than the Y2K problem. How are we going to achieve one term preservation and access society wide when the systems we create today using information technology are designed to be obsolete tomorrow and not even five years from now? More often they are obsolete the day you buy them, so the design center has to change.

So we are looking for a lot of different things from InterPARES and we are happy to provide things to InterPARES. If you need more information on CENSA, this is the contact information.