

Document Imaging Report

Business Trends on Converting Paper Processes to Electronic Format

Table Of Contents

Back to the Grindstone	2
The Capture Conference Will Go On	3
DocuWare Continues Focus on Ease-of-Use in Version 7.4	5
Surveys Say - the Capture Industry Will Grow	9
Ephesoft Updates Ephesoft Transact	14

Back to the Grindstone

Summer vacation is over, at least for most of us, and it's time to buckle down and return to work. A few items popped up and pushed this issue a bit further than I would have liked, but, and perhaps this is related, I did enjoy my time away and visiting family (and demolishing a good bit of boudin as well).



As you no doubt know by now, all of us here at Infosource are excited about the upcoming Capture Conference; September 8 & 9 in Chicago. As the first conference I'll be going to since 2013 (HubSpot's Inbound event), I'm looking forward to the learning experience.

In the coming months, you'll be seeing more online content from us. Our analysts will be sharing their thoughts and insights online. I will be referencing them in these pages too, so there's minimal risk you'll miss them.

Two are in the editing process and I think you'll like what you'll be seeing.

Thanks for reading,

Bryant Duhon
Editor-in-Chief, Document Imaging Report
bdu@info-source.com
+1(301) 275-7496

Comments, criticisms, and witticisms welcomed.

The Capture Conference Will Go On

We all watch the news, so it's no surprise that there's concern about the ongoing evolution of COVID. However, as they say, the show must go on. Our Capture Conference is still planned as scheduled for September 8 and 9 (reminder: should the event be cancelled due to new COVID restrictions, we'll provide a full refund on your conference fee).

We have a great group of speakers (see below for two of our speakers) and attendees coming to the Eaglewood Resort and Spa in Chicago and look forward to learning from each other in a safe, relaxing environment.

Safety Measures

To enable social distancing, many hotels aren't selling all of their rooms -- including Eaglewood. Tour cut-off date to ensure our attendee have a room at the hotel was this past Friday, August 13. To our knowledge, there ARE rooms available, but delaying registration risks having to stay in an off-site hotel.

That's one safety measure, another is that the hotel is spacious enough for social distancing during sessions and meals. And everyone is, of course, welcome to wear a mask (my looks are generally improved with one).

I know it's an uncertain time, but we look forward to hosting a safe, fun, and educational event. I know I'm looking forward to getting out of my house and meeting many of you for the first time .[Click here to register today.](#)

Secure Capture for Cloud Applications: Critically important cybersecurity technologies and concepts

Bill Edwards, Researcher and Cryptographer, Deep Blue Krypto.

As the Capture industry evolves, cloud computing presents a great opportunity for software and hardware companies to deploy solutions more rapidly. However, a rush to embrace cloud computing without careful consideration of critically important cybersecurity technologies and concepts are putting your business, your customers and your career at great risk. In this session, we will discuss the important considerations, and big mistakes to avoid, that Capture software and hardware companies make when moving to the cloud.



Healthcare and the Shift to Telemedicine -- What Are the Capture Opportunities?

Suleima Salgado, Director of Telehealth, Telemedicine, and Rural Health Initiatives, and **Jackie Woodard**, Telehealth Special Projects Manager, Georgia Department of Health.

Traditionally Capture has been mostly related to administration and records management. More recently it has been deployed to extract data for analytics. But with Capture 2.0 services, it can be deployed to support the shift to telemedicine with such technologies as voice stress analysis, sentiment analysis, photographic analysis of ailments, and video analysis of areas such as walking. When combined with sensors, this provides a set of very exciting opportunities. This session will explore the potential use cases for these Capture technologies.



Suleima Salgado

Jackie Woodard

DocuWare Continues Focus on Ease-of-Use in Version 7.4

As remote work continues to expand, companies are looking for easy-to-use and implement solutions for their automation challenges. DocuWare continues to build on its momentum from 2020 and 2021 as we begin to round the corner into 2022.

I've included a few stats about cloud adoption to place DocuWare's growth into some overall context in the wider world at the end of this article. TL;DR version: they're surfing a pretty damn big wave. On April 29 of this this year, the company reported a 40% growth in cloud revenues. Of all new customers that chose DocuWare from 4/1/2020 to 3/31/2021, 75% of them chose the cloud-based version. More on the cloud after a conversation about DocuWare 7.4 with Michael Berger, President, DocuWare Group.

DocuWare Version 7.4

DocuWare continues to build upon it's cloud-first momentum with a new release. Here's a quick overview of the updates:

- The ability to collaborate live on Microsoft Office documents stored within DocuWare.
- Highlight search -- highlight text and search for relevant documents.
- The ability to copy workflows from one cabinet to another, from one system to another, and clone them, which saves time by allowing workflows to be tweaked for different uses rather than built from scratch.
- Improved user authentication for security.
- Preconfigured solutions can be combined (invoice processing and smart document control).

Regarding security, DocuWare announced it is SOC 2 Type 2 certified as of August 5. <https://start.docuware.com/press/soc2type2-certification>

Working at the Speed of Partners

Don't take the above header as derogatory. The first question I had for Berger centered around the release schedule. My main direct cloud experience has been with HubSpot, with major updates at intervals, but also ongoing tweaks and "improvements" (which sometimes boil down to "where the hell did they move that button this time"). I asked Berger why DocuWare still seems to be adhering to an on-prem software upgrade schedule.

His answer made a lot of sense.

Firstly, they (and partners) continue to sell on prem and have legacy on prem users. By slowing down that release schedule, they provide a more straightforward upgrade path and the same look and feel and functionality if someone moves to the cloud. Basically, fewer things can break.

The second reason is because of their partner programs. This was not a dig at partners, but a statement of reality. Berger pointed out that partners don't have time to continually stay up-to-date on multiple, frequent releases. They are out selling/installing/supporting their customers. A slower schedule helps partners focus on delivering to customers, rather than, as I said earlier, wondering "where did that button move?"

The third reason is sales-related. Says Berger, "We wanted to give our sales forces two different versions, a public cloud and an on-premise option." By keeping the versions synced, customers know they can "choose the level they want to buy. They can buy on premise now and migrate to the cloud and have the same features and the same look and feel. They are flexible."

This isn't to say that the product stands still between releases. There are updates to the cloud release "here and there" throughout the year.

Why the Focus on Microsoft Documents?

I was initially puzzled that they led the release announcement with the MS Office functionality. Berger's answer made me understand why. DocuWare focuses on mid-sized companies and they focus on the MS product line because most of their customers are in the Office environment. "We integrated the editing of MS Word so you can natively open it out of our software and edit it. With the increase of remote work and the demand for better remote collaboration, we think that this feature is highlighted for these customers because it's improving their business productivity," said Berger.

How Preconfigured Is Preconfigured?

DocuWare has a number of solutions, including for finance & accounting, Invoice processing, employee management, and secure document archiving.

Says Berger, "They are designed from 30 years of experience -- and designed in a way to be up and running in under three days. That's what we see in the preconfigured solutions we've sold so far. They also sell much faster -- a 6 week average as of now. And then 3 days to implement and begin using. It's really a great achievement. We have several different solutions and you install several in your environment."

Berger pointed out that this is a fork in the sales road for companies. Many customers take and use the preconfigured solutions as-is, only adding company-specific fields and minor tweaks like that. For customers who do need something more customizable, that becomes a deeper conversation that on-prem software might be the right way for them to go.

Berger says that DocuWare will continue to focus on horizontal solutions while encouraging partners to develop their own pre-configured solutions, especially related to particular industries. While DocuWare has quite a few customers in financial services, you won't see them develop "invoice processing for financial services," for example.

He also pointed out that the preconfigured solutions are available for any DocuWare customer to use at no additional cost. Says Berger, "It's easy, you don't have to think. You can just use it, all the features and functions are there."

Given the platform approach many vendors are taking in this space (Ike Kavas of Ephesoft and I discussed this in this issue), I asked about plans for a DocuWare marketplace. Berger said that while they are definitely taking a platform approach to the market and encouraging partners to create solutions based on the software, they are seeing how the partner ecosystem develops before making concrete plans for an official marketplace. As he said, "There's nothing stopping partners from exchanging ideas, but there's no formal program in place at the moment to encourage this."

The Cloud

There are more cloud reports than one can read (at least if there are other topics you're interested in). I took a quick spin around Google to see what some of the numbers behind the "COVID-19 has driven cloud adoption" assumptions we all have.

From [Flexera 2021 State of the Cloud Report](#) which focuses on IaaS and PaaS:

- 90% of enterprises expect cloud usage to exceed prior plans due to COVID-19
- 36% of enterprises spend more than \$12 million/year on public clouds

[Gartner projects public cloud services](#) to grow from \$314B in 2020 to \$396 billion in 2021, with an additional growth of 21.7% in 2022 to hit \$482 billion. That breaks down to:

- \$51B in cloud business process services (BPaaS)
- \$80B in Cloud application infrastructure services (PaaS)
- \$145B in cloud application services (SaaS)

The balance is from cloud security, infrastructure, and desktop cloud services.

It's clearly a good time to be a cloud service provider.

Surveys Say - the Capture Industry Will Grow

Maybe because I caught a few minutes of Family Feud on the Game Show Network a month or so ago (old school with Richard Dawson), I thought it'd be interesting to run through a couple of the various research reports around capture.

For the sake of simplicity, we'll just go with "capture" when I'm referring to the industry; between intelligent process automation, intelligent document capture, intelligent information management, capture 2.0, and old school document imaging; the goal is the same: take a container (paper or digital) of information and extract the relevant data into your system for further action or storage. And possibly saving the image as well.

I had thought about going on a mini-definition rant, but decided to punt that to a potential blog post and focus on the research data.

What's interesting is that when you look at industry research from a variety of sources, a common story emerges.

It's a good time to be in the industry. Rather than citing study after study, a quick Google search on variations of "capture market size" shows:

1. A surprisingly large number of companies that track the market
2. A brief skim of the results shows expected year-over-year growth at healthy percentages

I'll make one quick note on the pandemic before turning to the stats. While much has been made of the role of the pandemic in pushing the growth of cloud-based products (like DocuWare, see page 5), that accelerated existing growth. I think we should be wary of contributing all growth to COVID as the industry was expanding at a steady rate before. It's obvious that the pandemic accelerated the need for tools that enabled remote work.

Context and the Ongoing Explosion of Content

Every couple of years it seems that there's a new meaning of "content explosion." They all tend to look like hockey sticks -- and to be out of date after a few years. Forrester, Gartner, etc. all predict massive data growth (I even read someone who suggested storage may not be able to keep up -- I didn't bookmark that. If anyone read the same thing, please send me a link at bdu@info-source.com). A good portion of that will be driven by video, but unstructured content will continue to be a major part of this growth.

AIIM's 2021 State of the Intelligent Information Management Industry uses survey results to put this in context.

"On average, organizations expect the volume of information coming into their organizations to grow from X to 4.5X over the next two years. They expect more than 57% of this information to be unstructured or semi-structured." Now, it's not clear how much of this content is digital versus paper. Regardless, both need to be captured.

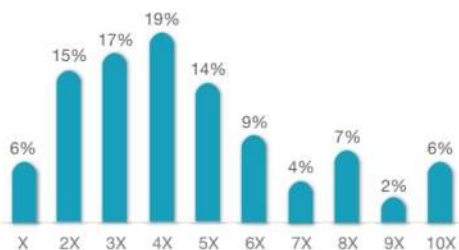
The graphic below shows the growth volume many industry practitioners expect.

2020 – A Wake-Up Call for Organization Leaders

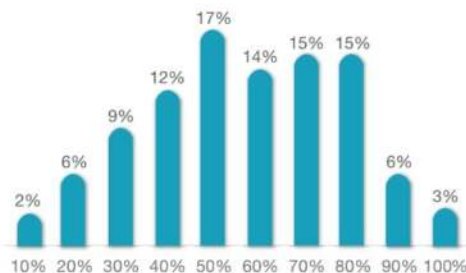
The Context for the 2021 Report – 1

A well-documented rising tide of INFORMATION CHAOS...

Assume the current volume of incoming information is "X" – what do you think it will be in 2 years?
Average = 4.5X



What would be your best guess for the percentage of UNSTRUCTURED information in your organization?
Average = 57%



©2020, AIIM – redistribution with attribution permitted – Overall N = 482

I won't belabor the "there's a lot of unmanaged content out there" point. You know it as well or better than I do. Let's move on to how poorly companies are responding/prepared to respond to this content avalanche.

Businesses Generally Don't Know What They're Doing

Not only is information growing, but companies are doing a poor job of managing this content, particularly manual approaches to managing information.

Again from the AIIM research, organizations were given an average C minus grade in addressing "information chaos." The research reveals a split in how senior managers think they are doing versus how the worker bees see things, as shown in the graphic below.

Role	Grade (4-point scale)	% ranking: "B-good" or "A-excellent"
Senior executive – i.e., VP and above	2.38	55%
Mid-level executive – i.e., director or manager	1.57	23%
Non-managerial staff position	1.39	11%

Those who describe their role as "executive level with broad organizational responsibilities" again are more optimistic about progress in the information chaos battle (average grade = 2.18) than are those with RM/DM/CM roles (1.47), IT responsibilities (1.59), or in line of business functions (1.58).

In fact, nearly half of respondents identified organizational response to managing this chaos as "needs improvement" (31%) or "poor" (15%). This is an opportunity for every vendor in this space, at least once you break down the misperceptions held by those making the decisions!

And the most common information chaos challenges? The two most-cited challenges are the same as they ever were:

Document Imaging Report August 16, 2021

- Managing information throughout its lifecycle: 24%
- Digitizing, automating, and integrating processes: 25%

These results are similar to the initial AIIM Industry Watches from the mid-2000s.

Research sponsored by Singularity from the Shared Services and Outsourcing Network (SSON) (*How to Become Truly Digitized Through Intelligent Document Processing*) echos the AIIM findings:

86% of survey respondents acknowledge the problem of unstructured data in their enterprises, half of them don't have a strategy in place to address the problem.”

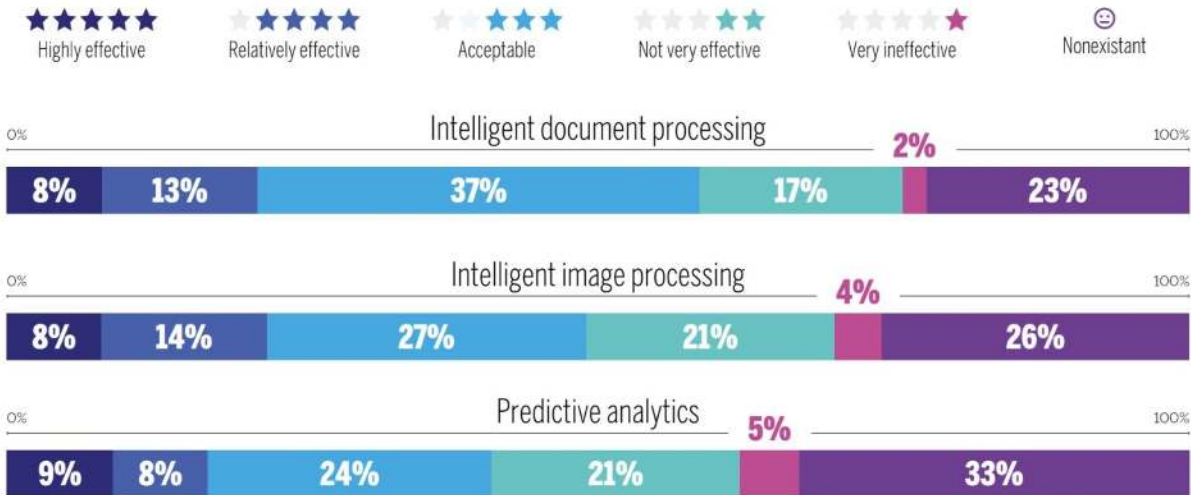


Additionally, the SSON research reveals that 60% of organizations describe their ability to be data-ready as basic, at best.

For vendors in the capture space, the following stats from the SSON research should be music to your ears: “Only two out of 10 survey respondents described their current intelligent document processing/intelligent image processing solutions as effective.”

Here's what that looks like graphically.

Please rank your current solution effectiveness in terms of:



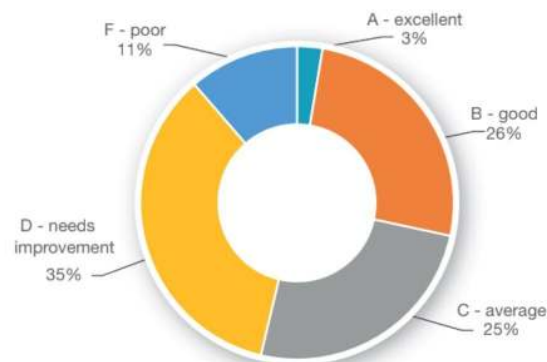
Now, unless one of the solutions in-place is one of yours, 80% of the companies in the SSON survey need help with automation.

Success Isn't Guaranteed -- For End Users

A common failure across IT is for executives to throw technology at a problem without proper planning. The AIIM research identifies the failure of the C-suite to align business and technology strategies as a problem; giving an average grade of C minus.

Average Business/Technology alignment grade = C minus (1.73)

Think about how well your organization aligns BUSINESS strategy with INFORMATION MANAGEMENT strategy. What grade would you give your organization?



©2021, AIIM - redistribution with attribution permitted Overall N = 222

What's This All Mean

As I wrote in the beginning, it's good to be in the capture space. Beyond digitization opportunities, companies know they are facing a deluge of data. Most of them are at least vaguely aware they aren't prepared to handle that incoming data effectively.

I've never been an industry analyst, and I didn't even stay at a Holiday Inn last night, but there continues to be a wide-open opportunity to digitize and ingest digital documents.

When you begin to think about the need to act on the data contained in these documents via workflow/process automation tools, the future looks even brighter, especially as the lines between the capture and workflow/automation vendor communities become ever blurrier.

Look for thoughts on the process side of automation in a future issue of DIR.

Ephesoft Updates Ephesoft Transact

Ike Kavas, CEO, Co-Founder, and Chief Productivity Office, and his team at **Ephesoft** have been busy this year. The company announced updates to Transact on July 21. Ephesoft Transact 2020. 0.05 has two major updates:

1. Cloud native, AI-enabled handwriting recognition
2. A complex table rule builder

Here's what Kavas said in the release, "In our quest for freeing up valuable resources for high-value tasks and maximizing enterprise productivity, Ephesoft continues to push the envelope of what's possible in document processing automation. Our latest Transact features fully automate two of the most complex challenges in document automation, eliminating manual data entry and errors. The extracted data will be readily available within seconds, redefining our customers' speed of work."

We spoke with Kavas about this recent update.

Before we got to the release, Kavas reminded me of the company's Cloud Hyperextender because the handwriting recognition is a part of that strategy. You can read about it here, <https://ephesoft.com/products/transact/cloud-hyperextender/>. Basically, Hyperextender is Ephesoft's cloud approach to allowing customers to plug apps into their capture process (handwriting recognition, in this case) as well as to scale image processing and OCR through the cloud. It's a part of customers' Transact implementation to allow them to accommodate unexpected spikes in processing demands.

Any customer can use the handwriting recognition capabilities as-needed. The handwriting recognition component of the new release is largely GoogleVision. Kavas' belief is that if there's something out there that already does an excellent job, incorporate it. Likewise, customers can also pull in Amazon's Textract as needed. Kavas alluded to more coming in the future.

For instance, for invoices, it's usually Textract. GoogleVision for handwriting recognition. Kavas points out that "the way they DON'T STORE information is very important to our customers for security and the extensions have our own AI tech and security around it."

It works like this, says Kavas, "It actually allows us to recognize the entire page. So if a customer receives an invoice and there are some notes at the bottom, we extract all of it. In the old days, you'd have to draw zones. This isn't that. We analyze the entire page, finding all of the characters, and recognize all of them."

VERY High Handwriting Accuracy

Given the inability of anyone to process cursive with any degree of accuracy, I was a little surprised when Kavas said, "We can achieve 88% accuracy or more with any doctor's writing."

I enjoy Kavas' occasional tweaking of the industry, but I'm still a little dubious about this one -- and said so.

Kavas didn't back off after I said that's a bold statement. "Yes, but we've done a lot of testing and I'm confident that we can do this based on cursive writing," he said. When asked if there's any element of a human looking at this behind the scenes, he said no, software only.

Because this is based on Google and they've been getting better and better at semantics and identifying intent, some of that knowledge and expertise carries over into this product. Kavas explains, "The beauty of the system is that Google has been collecting data from the world for years and they've put that data to use to create this engine. And it's not just English, it's 50 languages. They've built the engine to recognize not only the characters, but the words. So that's why I think it's very accurate. AI learns from the examples what the commonalities are. That's why I think the accuracy is high, even with cursive."

I asked him why they're taking this approach. Kavas replied, "Our job as tech company is to bring value to our customers based on how much automation we can create. If Google is developing the handwriting recognition part, we can focus on how to deliver an enterprise-grade document processing system. That's where we can classify, sort, and extract the data they really need. And we can leave the low-level tech, like handwriting recognition [lower level, not simple], to Google. It helps us to increase our value."

Left unsaid is that, practically, you're not likely to create something better than Google in this area. Why try.

He thinks the ROI can be extremely high. He's saying that they are offering something ten times better than before. To keep things simple, "What this means is that if you're processing insurance claims or you're reading the doctors' notes and you're receiving these as a company, you had to employ people to enter the data. If there are privacy concerns, you couldn't offshore that process. So if you had 20 people, at 50k each, that's a million annual expense. Now you need two, so that's a 900k annual savings."

This isn't a new product, it's part of the Transact product now. Pricing is based on pages you process per price tier; the handwriting is simply included there.

Kavas summed up this functionality addition, "With the handwriting now available, nothing changes, but we can now do all of that on the handwriting as well as the machine print. That's it. Customers now don't have to worry if their docs have handwriting combined with machine print, just give it to Ephesoft and we'll classify and extract the data you need."

Setting the Table

Table recognition is one of the most difficult challenges for document capture. Ephesoft has developed an AI algorithm that can detect where the table is and where the lines and columns are that, says Kavas, takes care of 80% of the problems. He adds, "This AI module is 100% ours, and we're in the process of applying for a patent for it."

"This new feature helps users to identify the tables. If a table isn't recognized when a user opens a screen, the AI-based system will recognize the columns and rows and the fields that need to be captured and then a few clicks, it's done. They don't need to see that vendor/invoice again, because we can automatically extract it in the future."

It's an AI-user assisted functionality to allow customers two big benefits. Says Kavas, "First, 70% to 80% accuracy on day one, not after three months of training and, two, because it's AI plus the rule builder, they can improve the system on their own without help from Ephesoft engineers or support."

Patent-Pending

While prepping for our conversation, I had noticed that Ephesoft has a patent incentive for employees. I asked Kavas what was the thinking behind that. He noted that he wanted Ephesoft to continue to innovate as he believes vendors in the space were too complacent when he started the company. Said Kavas, "It goes back to our core values. We're going to innovate and invent new things in this industry every day. As a result of this focus, we have some patentable ideas. I want ephesoftians to be rewarded and encouraged to contribute."

Building a Platform

Like most vendors in this space, Ephesoft is taking a platform approach to their software. Companies like Zia Consulting are partnering with Ephesoft and building applications on top of Ephesoft. For instance, Ephesoft Mortgage Powered by Zia was announced July 7 of this year.

Says Kavas, "We are in the process of building a platform." Kavas says that Ephesoft needs to be smart about where it spends its time creating product, using existing handwriting recognition from Google and allowing companies like Zia to build on the Ephesoft platform allows Ephesoft to focus on adding value.

Kavas notes there are benefits for customers as well. With the Zia integration, he notes that "on day 1 our customer is up and running with accuracy so they don't have to wait 3 months for a return."

The company doesn't yet have a marketplace, but it's on their roadmap. Currently, they are managing the process with Github Labs. Kavas explains, "Our partners can put their solutions there, license free or for a fee and then our ecosystem can use it from there. The next evolution of that is to create a marketplace."

When I spoke with Kavas at the beginning of the year, he said it would be a busy year for the company. He's certainly held true to that statement.

DOCUMENT IMAGING REPORT

Business Trends on Converting Paper Processes to Electronic Format

DIR is the leading executive report on managing documents for e-business.
Areas we cover include:

1. Document Capture
2. OCR/ICR, AI, and Machine Learning
3. RPA
4. ECM
5. Records Management
6. Document Output
7. BPM

DIR brings you the inside story behind the deals and decisions that affect your business.

Vol. 31, No. 8

Managing Editor: Ralph Gammon
Email: rg@info-source.com

Editor-in-Chief: Bryant Duhon
Email: bdu@info-source.com
Phone: +1 (301) 275-7496

DIR is published approximately 15 times per year by:

Infosource
SAAvenues des Grande-Communes 8, 1213
Petit-Lancy, Geneva, Switzerland
<http://www.info-source.com>

Copyright © 2021 by InfoSource SA. Federal copyright law prohibits unauthorized reproduction by any means including photocopying or facsimile distribution of this copyrighted newsletter. Such copyright infringement is subject to fines of up to \$25,000. Because subscriptions are our main source of income, newsletter publishers take copyright violations seriously. Some publishers have prosecuted and won enormous settlements for infringement. To encourage you to adhere to the law, we make multiple-copy subscriptions available at a substantially reduced price.

Subscriptions:

Single user: \$597 or \$670 (paper) per year.

Two user: \$897

From 3 to 5 every extra user: +\$100

10 user: \$1497

Company wide subscription: \$2500

Enterprise wide subscription: \$4500

