

Overview – Data Management Services

In today's unpredictable and fast-changing marketplace, entrepreneurs find that running a business consumes the bulk of their energy. They lack the time and resources to manage non-core, resource-intensive functions properly. In this scenario global organizations are pursuing a business process outsourcing (BPO) strategy in which data management services manage these non-strategic business processes.

Word processing, Mailing lists, Legal documents, Yellow page listings, E-Books, White Pages, Manuscript typing, Business Card data entry, Online data entry of URL links for books with their categorization, health reports, whatever be your Textual Data Entry requirements, Data Entry Outsourcing accomplish these efficiently through a combination of basic text entry, coded phrases and templates or through sophisticated word-processing packages.

Organizations big or small require accurate data for completing transactions, making timely decisions and supporting customers. Double data entry is a time consuming process that require costly managerial, clerical and technical staff as well as hardware, software and maintenance.

OCR and ICR processing services in a wide range of their data capture requirements as follows

- Medical forms
- Insurance forms
- Banking forms
- Bill Remittance forms
- Survey forms
- Bounded books
- Directories
- Magazines and Periodicals
- Ancient documents
- Airline Tickets

List of Services for Data Management

- Data Entry Services
- Data Processing Services
- Data Conversion Services
- Online Catalog Services
- OCR services
- In dexing Services

- **Data Entry Services:** Data Entry services are the most indispensable and internal function of Invensis and has exceptionally excelled in this vital area of expertise. Our extensive work force is made up of experienced, professional and reliable. The experts in data entry services are trained in online data entry of e-books, online data entry of catalogs and online data entry of databases, textual data capturing from printed matter, manuscripts, scanned images, web research, on-line data entry, offline data entry work, data conversion etc
- **Data Processing Services:** Invensis offers a data processing service that adds immense value to your business and generate revenue streams, which were hitherto untapped due to lack of cohesive information. Collection and logical processing of data inputs using data processing services has become an imperative function of any organization in the modern age irrespective of its size of operations. Invensis with its extensive knowledge, multi-industry experience, and resources provide flexible and reliable services that assure accuracy and timely execution.
- **Data Conversion Services:** Invensis is a one stop solution for quality, cost effective and time bound document conversion services. Source material may be in an electronic format (PDF, Word, etc.) or a non-electronic format (paper documents). Our dedicated data conversion team is equipped with the latest document conversion software that is capable of handling any level of customer demands in the field of document conversion.
- **Online catalog Services:** Invensis catalog processing services for online stores include converting suppliers' paper catalogs to the web, updating online catalogs and a building web-based catalog which helps you expand your business via Internet through an e-commerce store. Well defined and elegant images have always been a priority to Invensis. With the help of our aesthetic and quality images, the customers are bound to visit us over and again. Our catalog processing / web content production services include graphic image work, PDF conversion of catalogs & database updating. We get the catalogs from a variety of trade sources, like website, physical catalogs, digitally shot images etc.
- **OCR Services:** Invensis offers Optical Character Recognition or OCR services where the machines read text from paper and convert them into soft copies. We convert your scanned paper documents, books, PDF files in to editable and searchable text format. Our OCR experts then store, retrieve and edit accordingly. The main purpose is to convert the softcopies later into various formats in order to meet the client requirement and effectively use it for various business needs. Invensis is capable in handling any type of complex and huge voluminous document OCR conversion project.
- **Indexing Services:** We maintain the documents in a categorical manner leading to well organized documentation of the important documents. We deal with different documents like Manuals, Journals, Library Books Indexing, Catalogs,

Medical Records, Legal document indexing etc. Our team can analyze, plan and execute a project to satisfy any client requirement, no matter how simple or complex.

The list of services utilizing OCR and ICR processing technology:

- Invoice Processing
- Forms Scanning
- Document Imaging
- Document Automation
- OCR document conversion
- OCR scanning
- Medical Claims Billing
- Claims Processing
- Order Processing
- Survey Processing
- Tax Forms Processing
- Credit Cards Payment Processing
- Cheques Payment Processing

The List of Common Services in Data Management

- Data Entry / Data Processing
- Data / File Conversion
- Data Capturing
- Data Cleansing/ Data Enrichment
- Database Development and Migration
- Online Data Entry
- Offline Data Entry
- Image Entry
- HTML / XML/ SGML Conversion
- Data Extracting from Web
- Adobe PDF Conversion
- E-Book Publishing/ Conversion
- Scanning/ OCR/ Editing and Indexing/OCR cleaning, ICR, OMR
- Document Management
- Writing, proof reading and editing

Trends in data Management Services for 2011

As the global economy recovers, the technology sector has seen a welcome lift in business conditions so far in 2010, setting the stage for new design opportunities for engineers across a diverse set of expanding markets in 2011. Demand for a wide array of digital devices means that advancements in analog technology – in particular data conversion – is rapidly changing to meet new market requirements. In its mid-year 2010 data converter report, research firm Databeans Inc. predicts the analog/digital and digital/analog converter sector (ADC, DAC) will experience annual growth of nine percent through 2015, most notably in emerging markets that require more portability and longer battery life, as well as traditional applications in which performance is defined by faster data sampling rates and/or increased dynamic range.

Whatever 2011 may bring, engineers will find that data conversion continues in many ways to define the performance of today's digital devices by regulating the quality, quantity and speed of information that passes between the analog and digital domains. From audio and video functions in consumer electronics to wireless base stations to deep-space instrumentation, data conversion is playing a key role in the success of these devices and missions.

For example, data conversion is the key to enabling higher fidelity audio, improved on-camera video displays and a variety of capacitive (touch) sensing options--developments that are being fueled by the continuing boom of the iPad, iPhone, and a host of other personalized, handheld gadgets and smartphone-enabled devices.

Rather than increasing power budgets, however, such new and improved features will be accompanied by a continued reduction in energy consumption. This seemingly dichotomous trend will play out for the foreseeable future driven not only by battery-powered consumer goods but also by the trend toward portability in devices as diverse as handheld ultrasound equipment and smart water meters. Even non-portable, power-line devices are not immune to this seismic shift as today's market environment emphasizes "green" designs that address the global mandate for energy conservation, lower overall system power, and reduced cost of operation and ownership.

The wireless communications market will remain another key driver of data converter performance, power efficiency and calculated integration. Emerging 4G cellular networks are expected to support greater call and data volume and higher data conversion rates in order to keep pace with the exploding popularity of wireless video transmission.

At the same time, OEMs are calling for wireless infrastructure equipment that simultaneously supports multiple carriers and air-interface standards, which places a new set of demands on data converter speed, flexibility, dynamic range, noise reduction and size. Add to that the proliferation of smaller cellular networks

comprised of picocell and femtocell wireless base stations, and it's clear that the future of high-speed converters in this market will be defined by lower power consumption combined with faster sampling rates and more usable bandwidth at higher intermediate frequencies.

While the power efficiency mandate is asserting itself in markets that historically cared much more about raw data converter performance, OEMs serving the industrial, instrumentation, aerospace and defense electronics industries additionally continue to require rugged, robust converters for applications that must withstand harsh operating environments. On one end of the spectrum, designers are looking for high-performance, innovative data conversion solutions for medical, energy, communications and other devices.

On the other, engineers are unwilling to trade off on reliability, especially in industrial and scientific applications where system repairs may be difficult or impossible. The ability to provide not only the converters but to understand and recommend solutions to meet these diverse needs is increasingly important. Analog Devices, for example, collaborates with customers to provide exceptionally high data conversion accuracy in monitoring and measurement equipment that may be used in extraterrestrial exploration, as in the case of the Mars Phoenix Lander, or that are permanently buried under kilometers of South Polar ice, as with the University of Wisconsin's IceCube Neutrino Observatory project IceCube Neutrino Observatory.

Regardless of the end application whether performance is defined by speed, power efficiency or precision sensing and measurement data conversion technology will remain a core ingredient for defining the quality of the end user's digital experience.

The Changes Taking Place and the Technology Usage in Data Management Services

In this digital world, financial institutions move towards an electronic and image exchange environment which permits them to accept a wide range of image processing files. By using cutting edge technology, It assists global organizations to transform paper and electronic documents into a wide variety of output forms tailored to both traditional and e-commerce marketplaces.

In the changing scenario of manual labor-intensive **data conversion** being replaced by machines, using Optical/Intelligent Character Recognition (OCR / ICR Processing) technology has become a viable option to meet most of the medium and large-scale business requirements. The technology that allows to transform paper documents into editable computer files has been proved as the best alternative to manual data

capture/typing in terms of cost, time, accuracy, perfection, and quality.

Text-based contents of a digital library can be stored and presented as i) simple text or ASCII (American Standard Code for Information Interchange); ii) unstructured text; and iii) structured text (SGML or HTML or XML).

Simple text or ASCII

Simple text or ASCII is the most commonly used encoding scheme used for facilitating exchange of data from one software to another or from one platform to another. “Full-text” of articles from many journals has been available electronically through online vendors like Dialog and STN in this format for over two decades. Typically what is stored in the text of each article, broken into paragraphs, along with bibliographic information is a simple tagged information.

Simple text or ASCII is compact, economic to capture and store, searchable, interoperable and is malleable with other text-based services. On the other hand, the simple text or ASCII can not be used for displaying complex tables or mathematical formulae. Photographs, diagrams, graphics, special characters cannot be displayed in ASCII. ASCII format does not store text formatting information, i.e. italics, bold, font type, font size or paragraph justification information. Simple text or ASCII in many ways is inadequate to represent many journal articles because of the reasons mentioned above. Although simple text or ASCII is extremely useful for searching and selection, its inability to capture the richness of the original makes it an interim step to structured text formats.

Structured text format

Structured text attempt to capture the essence of documents by “marking-up” the text so that the original form could be recreated or even produce other forms. Structured text format have provision to imbed images, graphics and other multimedia formats in the text. SGML (Standard Generalized Markup Language) is one of the most important and popular structured text format. ODA (Office Document Architecture) is a similar and competing standard. SGML is an international standard (ISO, 1986) around which several related standards are built. SGML is flexible language that gave birth to HTML (Hyper-text Markup Language), *de facto* markup language of the World Wide Web, to control the display and appearance of documents. Like simple text or ASCII, structured text can be searched or manipulated. It is highly flexible and suitable both for electronic and paper production. Well-formated text increases visual presentation of textual, graphical and pictorial value of information. Structured formats can easily display complex tables and equations. Moreover, the structured text is compact in comparison to the image-based formats, even after including imbedded graphics and pictures.

Standardized General Markup Language (SGML)

SGML is application independent, non-proprietary and extremely flexible mark-up language. It was first developed in 1970 as GML (Generalized Markup Language) and evolved into both a national and International Standard (ISO, 1986). SGML is frequently referred to as a meta-language, which means that SGML is not a single language but a language that describes a family of markup languages. In other words, SGML is the framework for defining particular markup languages. SGML is an effective solution for handling complexity of electronic publishing because of its powerful and flexible structuring capabilities, as well as for its capacity to capture and organize information about the publications (“metadata”). It provides for descriptive, as opposed to procedural markup. That is, it simply, states names to categorize parts of a document instead of specifying process to be carried out (Kardorf, 1998). SGML uses text characters both for the text as well as for mark up that describes that text. It has no proprietary codes; instead each user (or group of users) may create whatever codes are necessary and meaningful for what is being published. A publisher can define his own set of codes for books and journal publishing. The key to self-defined codes in an SGML document is called DTD (Document type Definition). Codes sets or DTDs can be specific to a single book or journals or can span to a group of related books or journals. An SGML document consists of three distinct parts namely:

Declaration: It gives fundamental information like language of document and code set being used (i.e. English/ASCII)

DTD: Details of codes and rules restricting their use.

Instance: The text being published, marked up with the codes described in the DTD.

SGML concerns itself with the structural features of a document while the appearance and display features are left to the ultimate presentation system to determine how those features appear on display or print. Resultantly, when documents move from system to system, or portions of one document are used in another, they don't need to be recoded. Because of its powerful and flexible structuring capabilities, as well as its capacity to capture and organize information about the publications, SGML-coded documents can be used effectively to search information contents of documents based on the structure and content of the information. Many SGML depositories are considered as “text databases”. Since they enable a publisher to organize the published information in different ways for different contexts.

Contents of an SGML documents are stored separately from its format, resultantly contents on parts of contents can be rendered in different ways for different needs, platforms and display methods. SGML is often used as an archival format and for document reuse and repurposing. Richly-coded SGML documents also facilitate more complex searching than unstructured, word-processed text. For fully marked-up documents, searches can be made on bibliographic citation marked <site> or such citations can be extracted from each document to create a citation database as a secondary product.

SGML liberates documents from the cumbersome and costly process of conversion from system to system. It does not require any special hardware or software. It is possible to create a valid SGML file in any word processor or text editor although there are number of SGML-based system available in the market. SGML preserves the document and its coding from obsolescence as well. Owing to the fact that an SGML document incorporates the key to its own codes (Declaration, DTD), it is possible to validate SGML codes by parsing the SGML file. Parsing is a process by which the document instance is checked against the declaration and the DTD to make sure all the codes in a file are legal and used properly.

Extensible Markup Language (XML)

XML is subset of the Standard Generalized Markup Language (SGML). It is designed to make it easy to interchange structured documents on the web. Like SGML, XML also deals with the structure of document and not its formatting. The Cascading Style Sheet (CSS) developed for HTML would also function for XML to take care of formatting and appearance. Unlike HTML, XML allows for the invention of new codes. XML files are not only consistent and compatible with SGML, it also simplifies SGML in many ways. For example, while SGML allows "tag minimization", enabling the omission of end tags, XML always requires explicit end tags that make it a lot easier to write tools and browsers. XML introduces the concept of a "well-formed" document, one in which the tags used are nested correctly and proper XML syntax is followed. In addition, like SGML, XML allows for "valid" documents too, which go a step beyond "well formed" status by using an explicit structure defined in a DTD. "Well-formedness" is a very appealing feature of XML, because it allows publishers to tag what they are publishing in whatever way is meaningful, without being confined to a specific set of tags (as with HTML) or needing to write a DTD.

XML document may require companion XSL (Extensible Style Language) to reformat in into RTF, LaTeX or any other format. XSL also makes it possible to offer database functionality from XML documents with no actual database needed. XML also defines how Internet Uniform Resource Locators can be used to identify component parts of XML data streams. Akin to an SGML document, XML documents can also be verified to ensure that each component of document occurs in a valid place within the interchanged data stream by defining the role of each element of text in a formal model, known as a Document Type Definition (DTD). An XML DTD allows computers to check, for example, that users do not accidentally enter a third-level heading without first having entered a second-level heading, something that cannot be checked using the Hypertext Markup Language (HTML). However, unlike in SGML, DTD is not a necessity in XML. If no DTD is available, either because all or part of it is not accessible over the Internet or because the user failed to create it, an XML system can assign a default definition for undeclared components of the markup. (Kardorf, 1998)

XML allows users to:

- ✓ . Bring multiple files together to form compound documents
- ✓ Identify where illustrations are to be incorporated into text files, and the format used to encode each illustration
- ✓ . Provide processing control information to supporting programs, such as document validators and browsers
- ✓ Add editorial comments to a file.

Like SGML, XML does not have a predefined set of tags of the type defined for HTML that can be used to markup documents in a standardized template for producing particular types of documents. XML is formal language that can be used to pass information about the component parts of a document to another computer system. XML is flexible enough to be able to describe any logical text structure, whether it is a form, memo, letter, report, book, encyclopaedia, dictionary or database. XML is based on the concept of documents composed of a series of entities or objects. Each entity or object can contain one or more logical elements. Each of these elements can have certain attributes (properties) that describe the way in which it is to be processed. XML provides a formal syntax for describing the relationships between the entities, elements and attributes that make up an XML document, which can be used to tell the computer how it can recognize the component parts of each document. XML differs from other markup languages in that it does not simply indicate where a change of appearance occurs, or where a new element starts. XML sets out to clearly identify the boundaries of every part of a document, whether it is a new chapter, a piece of boilerplate text, or a reference to another publication. The structure of a document can be checked if the user provides a document type definition that declares each of the permitted entities, elements and attributes, and the relationships between them.

Hypertext Markup Language (HTML)

Hypertext Markup Language (HTML) is an SGML application complete with DTD. It is designed to tell a browser how to format documents on the web. HTML is the *de facto* language of the web and is largely responsible for resurgence of interest in SGML in the past few years. Unlike SGML, HTML has a pre-defined set of codes, that are easy to learn and use and build tools for writing HTML pages. HTML codes are imbedded into the text that communicate to a web browser such as Netscape Navigator or Microsoft Internet Explorer. Like SGML, it also uses simple text or ASCII for text as well as for the HTML codes. An HTML page can thus be built using a word processing package or a text editor. There are several HTML editors and conversion programme that act similar to a word processing package. These editors typically show the codes as they are inserted. In a What You See Is What You Get (WYSIWYG) environment, such as MS Word or other MS Windows packages, the user never see these codes. Web browsers are similar to WYSIWYG word processors because it reads the imbedded codes and then applies them to the specified text.

HTML is competent at presenting text, graphics, images in a reasonably decent layout. Web browsers readily accommodate a multitude of plug-ins that allow inclusion of audio, video, 3-D, and other specialized files. Any of these can also be included as a link in a standard HTML page. Clicking the link loads the plug-in to view or play the file. HTML files are tiny since they are simple text files. Further, the static HTML web pages can be transformed into vibrant, dynamic and interactive web creations using ever evolving web technologies like CGI Script, Perl, Java, Javascript, ASP, DHML, XML and open database connectivity (ODBC) for incorporating interactivity on a web site. Simplicity of HTML is also its serious limitation for books and journals. HTML does not provide enough codes to present complexities of a scientific text. It does not provide for Greek and maths characters that are important to scientific text. Moreover, HTML is all about presentation and not for structure or contents. The only contents that it describes is in Metadata codes or in its title. Furthermore, an HTML file can be derived from an SGML file any time but the reverse is not possible. The competency of HTML at presenting text has further been enhanced with use of Cascaded Style Sheet (CSS) in its version four.

Portable Document Format (PDF)

HTML cannot maintain all the formatting and presentation layouts of the document as it appears in original. Portable Document Format (PDF), a by-product of PostScript, is used to maintain page integrity and layout of the page. The PDF format maintains the look and presentation of the original document. However, PDF is a proprietary item that a browser cannot read on its own. PDF files require Acrobat Reader, a free reader available from Adobe's site for different platforms. A PDF file is a self-contained cross-platform document and can be viewed across multiple platforms using the appropriate reader for that platform. PDF files are page independent, i.e. a user can extract any given page for printing. Every page of a PDF document has information to display or print. PDF files preserve all of the fonts, formatting, colors, and graphics of any source document, regardless of the application and platform used to create it or the platform or device to which they are downloaded. Windows, Macintosh, UNIX, and DOS versions of Acrobat Reader are available for free from Adobe Systems, Inc. In plain language, it is a file that will look the same on the screen and in print, regardless of what kind of computer or printer someone is using and regardless of what software package was originally used to create it. Moreover, since PDF uses vector technology rather than raster, the resulting files are very compact. (Kardorf, 1998).

PDF is the *de facto* standard for electronic document distribution worldwide although it is proprietary of Adobe. PDF files are compact and can be shared, viewed, navigated, and printed exactly as intended by anyone with a free Adobe Acrobat Reader. Any document can be converted to Adobe PDF, even scanned paper, using Adobe Acrobat 4.0 software. PostScript files can be converted into PDF using PostScript interpreter called Distiller.

PDF carries little or no structural information although Acrobat does offer some navigational features. Full text in a PDF document can be indexed using Acrobat's

Catalog software enabling Boolean searching on collection of PDF files. Hyperlinks can also be created within or between Acrobat documents using Acrobat Exchange. Moreover, PDF files also allow incorporation of keywords, author's name, etc.

Some of the form fillings

- Coupon redemption forms
- Credit card application
- Health claim forms
- Insurance claim forms
- Legal forms
- Medical claim forms
- Market research forms
- Patient record forms
- Product registration forms
- Questionnaires
- Restaurant Menus
- Rebate forms
- Shipping documents
- Subscriptions forms
- Survey forms
- Tax forms

The Sources

<http://www.dataentryoutsourcing.co.uk/>

<http://www.dataentrymonster.com>

<http://www.eetimes.com>

www.streambase.com