

## ZyLAB's Deployment at the Historic UN War Crimes Tribunals



A ZyLAB Case Study

*January 2015*

# Contents

Introduction.....	3
Issues of Regulated Access & Controlled Document Processing .....	5
Mandatory Search Capabilities.....	6
Scale .....	6
Connector architecture for disparate system data .....	6
Recall .....	6
Text analytics.....	7
Taxonomy .....	7
Multilanguage operations.....	7
Multimedia operations .....	7
Selection and Deployment.....	8
Advice to Other CIOs and CTOs .....	9
Prepare for multi-media operations.....	9
Understand the demands of recall in search.....	9
Embrace the challenge of scale.....	10
Expand to true multi-language processing.....	10
Build for repeatability .....	10
Conclusion .....	11
About ZyLAB .....	12

# Summary

In 1994, the UN International Criminal Tribunal for the former Yugoslavia (ICTY) turned to ZyLAB in order to get help to manage the accumulation, organization, and access to evidence relating to war crimes. The collection consisted of millions of pages, electronic documents, and multi-media recordings in a variety of languages and from many different sources. As the court started, this collection was expanded with thousands of additional documents such as witness testimonies, transcripts of court sessions, reports of special investigations and depositions from defense teams. The management of this massive document collection required the deployment of a specialized system which could combine advanced multi-lingual search, eDiscovery, content analytics, content management and support the specific workflow required by the court to assess and disclose the evidence.

This case study describes the depth of the technical and operational issues addressed by the UN Information Management team in the Office of the Prosecutor and explores how professionals and software from ZyLAB literally partnered with the UN Information Management group to achieve a groundbreaking new system in a very condensed timeframe. It also suggests best practice areas revealed by the work of the overall team in responding to the challenges presented by this major eDiscovery project.

Now, 21 years later, in 2015, ZyLAB's products are still used, by not only the ICTY, but also by several other war-crimes tribunals such as the courts for Rwanda (ICTR), Cambodia, and the Special Court for Sierra Leone (SCSL).

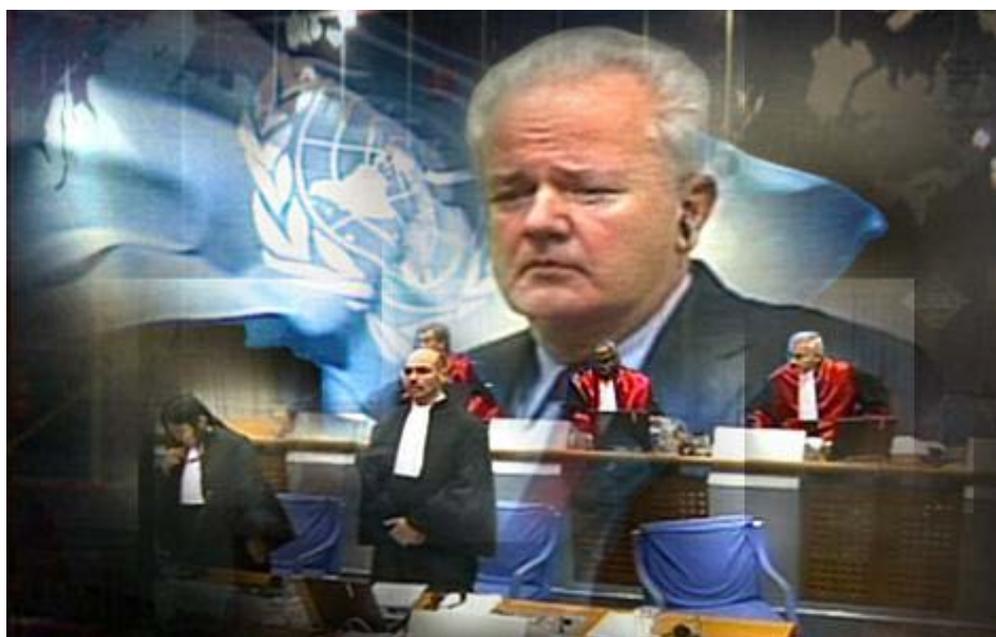
ZyLAB is proud to have contributed to these historic war crimes tribunals and continues to be a technology provider of eDiscovery and Enterprise Information Management solutions to other organizations supporting international law and justice such as the International Court of Justice (ICJ) and the International Tribunal of the Sea.



# Introduction

This case study describes how the United Nations faced the need to manage the accumulation, organization, and access to evidence relating to war crimes in troubled areas around the world, including trials in the former Yugoslavia, Rwanda, Cambodia, and Sierra Leone. The UN plays a unique role in establishing and supporting an increasing number of international tribunals whose role is to fairly and promptly provide legal process and to conduct trials for individuals charged with crimes against humanity.

The UN team that is responsible for gathering and handling the information to be used in such trials faced the challenge of making millions of documents in many formats and many languages available to prosecutors, defense attorneys, judges, and other court stakeholders. This war crimes evidence originates in multiple formats from disparate sources, for example: TV program tapes, radio broadcasts, news and military photographs, home movies, home photos, recorded telephone communications and other rich media formats, in addition to masses of paper documents and the standard electronic text of



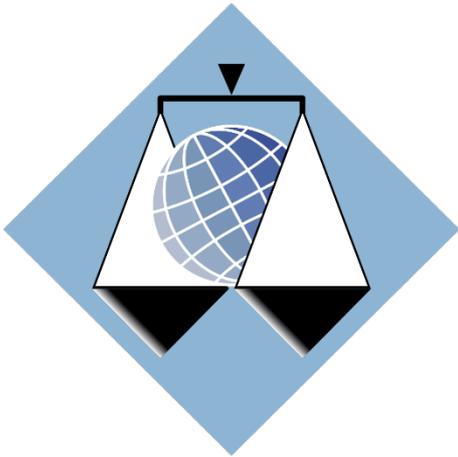
emails and other natively electronic documents.

The first tribunal that the team needed to enable with a new automated system was the International Criminal Tribunal for the former Yugoslavia (ICTY). Specifically, they needed to provide a system to support the conduct of the trial of Slobodan Milošević, the former President of the Serbian Republic and the Federal Republic of Yugoslavia. As their project

developed, they came to create a ground-up definition of an electronic disclosure capability that could and did serve as a model for conducting similar kinds of trials (e.g. for the continuing trial of Radovan Karadžić and the forthcoming trial of recently-apprehended Ratko Mladić in Yugoslavia). This system of evidence handling and disclosure has been proven to operate successfully not merely in UN work related to the Yugoslavian tragedy but also in other areas of the world where the UN faced similar crises.

This case study describes the depth of the technical and operational issues addressed by the UN Information Management team in the Office of the Prosecutor and explores how professionals and software from ZyLAB literally partnered with the UN Information Management group to achieve a groundbreaking new system in a very condensed timeframe. It also suggests best practice areas revealed by the work of the overall team in responding to the challenges presented by this major eDiscovery project.

# Issues of Regulated Access & Controlled Document Processing



The mission for the Information Management function in the Office of the Prosecutor is to ensure full and equitable access to all evidentiary information by all parties to the trial. Such access requires that all parties with legitimate “right to know” have to receive complete, accurate, and timely production of requested documents or for topics within documents. The process typically involves multiple professions, such as digital forensics specialists, lawyers, and IT professionals, all with slightly different objectives and requirements, which must ultimately ensure system operations protocols that can be certified by the governing authority, in this case by the UN tribunal itself.

System operations must at a minimum comply with overarching regulatory standards like the Federal Rules of Civil Procedure in the US or, in the case of ICTY, the UN’s Rules of Procedure and Evidence for International Tribunals. These regulations provide the parameters that evidence discovery and production processes must adhere to.

When the UN team began the project, the de facto discovery process was still based primarily on hard copy procedures, with multiple-copy (twelve copies per request) production typical. Given the amount of manual effort to process the quantities of information in the ICTY using these traditional approaches, the team turned immediately to a strategy of document capture and subsequent digital production of paper documents. But even with evidence available in electronic format, it often was physically impossible for the team to print the multiple copies of documents in the discovery requests quickly enough. There was simply too much data to be able to attempt to process it with manual methods. With native digital items, the information forensics professionals have interest in reviewing documents in their original file formats, and this added an additional challenge for the UN team to be able to open many kinds of application files and archived documents from a large number of disparate systems.

In order to accomplish its goal, the team had to develop a system design to process both paper and digital documents in a manner that streamlined the work enough to meet both the stipulated deadlines and the quality parameters in the procedural guidelines. In both hardcopy and digital operations, the team needed to apply advanced software conversion and analytics processes in order to support the needs of the discovery system.

# Mandatory Search Capabilities

Search and discovery technologies lie at the heart of the system created by the UN Information Management team. In this section, we review briefly distinctive search capabilities that became mandatory given the tribunal's requirements for the eDiscovery environment for the ICTY, and later in other trouble spots around the globe.

## Scale

The first capability required of the UN system was scalability. Given the massive quantity of documents and the disparity in formats, the search system could have no limitations on quantities of data processed, size of index, and other raw quantity measures. At the same time, it had to accomplish discovery work at a level of processing performance that would effectively support the human-driven direction of the discovery investigations. This requirement meant that sophisticated analytics on captured and original documents had to be able to run fast and not create bottlenecks in the overall system. The system also came to support a potentially large number of simultaneous users, all needing to conduct explorations at peak system speeds.

## Connector architecture for disparate system data

Search systems increasingly must play a central role in what IDC refers to as Unified Access platforms. "Unified Information Access Platforms provide a single point of access to multiple heterogeneous sources of information. These platforms are highly scalable and built to accommodate quickly changing information through real time or near real time updating and analytics. They combine elements of database, business intelligence, and search technologies in order to make information access dynamic and ad hoc for business users. They are capable of indexing and integrating large volumes of unstructured, semi-structured, and structured information into a unified environment for analysis or decision support. Like portals and search applications, they offer rich and interactive user interface designs. Like BI applications, they integrate visual analysis tools. These platforms promise to boost information worker productivity because they enable ad hoc queries and quick, easy access to both data and content for business decision makers as well as trained data analysts." See Unified Access to Information: Less Seeking, More Finding (IDC #227780, April, 2011)

In the UN system, the mandate was to provide discovery system users with a customized unified access environment that broke down application and format barriers and unified silos of data into a single index to support exploration and analysis.

## Recall

Discovery brings the Recall capabilities of search to the forefront. Many search systems prioritize processing in the direction not of recall, but precision. Ecommerce site search, for example, wants to provide each user with a very few product offerings that appear to be relevant to the search term, not overwhelm the users with every conceivable match for that term. In discovery, the opposite is true. The whole discovery process is grounded on the idea that every relevant scrap of data about a particular topic, person, event, or relationship must be brought to light. Occurrences from any information source and in any format could be equally critical to the legal framework of a case. A strong recall capability, based on such capabilities

as fuzzy matching, text analytics, and entity, relationship, and event extractions must be engineered into the basic discovery search system design.

## Text analytics

Text analytics broadly refers to a wide set of technologies that have been developed over the past decades to analyze text at linguistic and structural levels that operate at more fundamental levels than the matching of keywords. In the discovery application, text analytics is critical to recognizing instances of multiple languages, for example, or differentiating whether a term is being used in an object or verbal context, or differentiating between a named individual of interest in a legal case and an unrelated person who happens to share the same name. Text analytics also makes fundamental contributions to automated generation of metadata across a document collection. A capable discovery system needs to incorporate a strong set of text analytics functions to provide legal professionals with the levels of granularity of analysis they need to perform their jobs.

## Taxonomy

Classification, auto-categorization, structured metadata management—all these capabilities and more fall into the general area of taxonomy software. Taxonomy capabilities also form much of the foundation for electronic records management systems and underlie systems of compliance for retention and production of regulated information. An ability to leverage subject-based and other forms of hierarchical organization of information is a key tool for the discovery professional. Beginning at the simplest level of providing navigation aides in exploration sessions for document collections, taxonomies can also boost productivity dramatically within analytic processes by exposing relationships that other tools obscure.

## Multilanguage operations

The ability of the UN's system for tribunal information management to work in locations as diverse as Yugoslavia, Rwanda, and Cambodia speaks to the core requirement for a full suite of language analysis and processing capabilities, both for indexing and text analytics operations, and for query processing and results matching in the discovery environment. The ICTY represents a particularly challenging case, with its 13+ languages at use in the evidence data and its users fluent in different languages themselves, but it will be increasingly common for true multi-language operations to be required across enterprises and industries, as globalization brings the languages of the world into close proximity and more user-friendly authoring systems encourage increased localization of content.

## Multimedia operations

With wars now routinely carried out on local and global television, with the latest communications from the front provided by embedded journalists with camera teams, the importance of rich media in offering evidence of what took place in particular times and places has never been higher. The UN team was able to call on audio and video analytics capabilities to bring these forms of rich media into the review environments for analysts on both the prosecution and defense teams. While in many search systems, this multimedia analysis capability remains rudimentary at best, discovery search must provide a state-

of-the-art suite of audio and video indexing and retrieval functions in order to support the mandate for a comprehensive discovery environment.

## Selection and Deployment

The UN team selected ZyLAB among a number of possible vendor options for reasons that encompass both distinctive business relationship experience and technical capabilities and expertise.

When the need to create the application for the International Criminal Tribunal for the former Yugoslavia arose, The Office of the Prosecutor's Information Management office had already established a relationship with ZyLAB. The company was located nearby, had done a number of smaller projects with the prosecutor's office previously, and was willing to commit highly skilled professionals to on-site support. Such support was particularly important because of the complexity of the problems the UN team was facing and because of the foreshortened time frames within which the team needed to make the evidentiary information available to the prosecution and defense teams. The overall schedule of the trials depended on the success of the project.

ZyLAB's software also offered a broad range of features and functionality critical to the eDiscovery process. Document capture and sophisticated OCR, for example, had been a foundation capability in developing the company's business. The ability to recognize documents from different languages and to differentiate and process languages notated in multiple character sets was a system necessity (for example, evidence in the Milošević trial incorporated over thirteen languages, including some written in Cyrillic character sets). The multi-language query parsing capability supported system users working in one native language to achieve the same results as users working in a different native language. Sophisticated indexing technology was required for the project which enabled the UN team to use the ZyLAB system to classify documents, exhaustively identify people, features, and entities referenced in documents, and create a text and metadata repository that normalized millions of evidence items while supporting legal review operations that could be executed within the timeframes stipulated in the UN tribunal's procedural guidelines. The software also offered a portal-like interface that provided an intelligent discovery workspace for the set of varied professionals at work on the trial.

The system grew over the course of the ICTY trial work and the global implementations from a team of 30 professionals to over 150 people operating in four countries. This spread of supported staff, the increase in their productivity, and the export of the core capabilities to African and Southeast Asian tribunals underscores the effectiveness of the design and development work on the project.

# Advice to Other CIOs and CTOs

The challenges faced by the UN Information Management team in the Office of the Prosecutor, while exceptional in scale and often gruesome in content, can shed light on more day-to-day challenges facing many businesses and public organizations today.

The following (minimum) best practice considerations, will be discussed in more detail below:

- ❖  Prepare for multi-media operations
- ❖  Understand the demands of recall in search
- ❖  Embrace the challenge of scale
- ❖  Expand to true multi-language processing
- ❖  Build for repeatability

## Prepare for multi-media operations

Multimedia operations are now the norm. Simply put, today's increasingly digital world is bringing all kinds of digital information formats into the mainstream of business operations. At the same time, the prevalence and long-term inertia of paper-based processes that continue to service businesses effectively will insure that use of paper will actually continue to grow, and IT's challenge in helping the business manage that hard copy flow will continue to confront both legal departments and other functions of the business. For natively digital content, the text formats that have long been the focus of enterprise search systems will be joined by equally important audio and video formats, as digital communications raises the bar for production quality and user expectations shift to richer formats, an established clearly in evidence in the newspaper business online, for example. For discovery operations, rich formats will need to be as searchable as text, and this capability will distinguish advanced technology suppliers from the rest over the next 3-5-year period. For hardcopy content, digitization processes will continue to be a keystone of operations for years to come, particularly focused in operations involving regulated information. In this area, world-class capture and OCR capabilities are a foundation requirement for any software evaluation.

## Understand the demands of recall in search

The two overarching goals of digital search systems, constantly in tension, are precision and recall, as any student of search learns from Day 1. No one system so far developed balances the two perfectly, and the considerations that tip the balance in one direction or another are ultimately rooted in the business problems in which search is deployed and the intentions of the audience(s) of human searchers interacting with the system. In the area of discovery, the truly dominant capability must always be the capability of recall, as the cost and risk to the prosecution or defense of missing an instance or reference of a key piece of evidence is exceptionally high. In evaluating discovery search, then, it is fundamental to understand the functionality of the engine in regards to recall, and gain comfort in the ability of a particular software system not to lose references to key people, places, things, and relationships in the process of indexing and producing content.

## Embrace the challenge of scale

Scale has become an essential ingredient in discovery operations. Although not all applications require the kinds of extensive repositories created for the evidence of war crimes, almost any organization faces unpredictable demands, under the Federal Rules for Civil Procedure, for example, to provide discovery against arbitrarily large sets of data within days of notification. In these circumstances, the risk of being unable to respond because of the failure of a system to scale up sufficiently is not acceptable. The potential cost to the organization of failure in this area far outweighs the cost of design and operations of a discovery software capability that will handle whatever “big data” load that it may be required to process. There are many innovations in hardware and software now being deployed to handle these applications with extreme scale. Discovery systems are a strong candidate for deploying these new capabilities.

## Expand to true multi-language processing

Multi-language search facilities have been a “nice to have” for many organizations in the past. In search, however, as global business expands, it has become increasingly important for many organizations to have multi-language processing across the capture and indexing of the content, the content analytic routines (e.g. identification of entities and relationship references), and in the interpretation of the query. For legal cases whose evidence crosses borders and languages and alphabets, as in the UN’s work, the need is pressing and extreme. But in any multinational organization, and in many single-nation organizations who are doing increasing amounts of global business, these same capabilities are taking on a critical level of importance. This is another area where search technology offerings differ substantially, and discovery teams will be wise to consider multi-language capabilities particularly closely in product evaluations.

## Build for repeatability

Any discovery system must operate in an atmosphere which balances “surprise” and readiness. Organizations often are not able to anticipate future litigation or new regulatory requirements. Operating under a scenario in which discovery responses need to be designed on a custom basis for each new risk, regulatory regime, or court order is simply not economical or sensible business practice. The UN team could see from the beginning of their project that the investigations in Yugoslavia would spread over multiple trials involving multiple sectors of a regional war. Their situation dictated that a central system design that provided the modularity to be adapted to other tribunals and trials around the globe would be the most effective investment in the medium to long term. Their project enabled them create such a system and to extend adaptations of the software and operational processes quickly to additional genocidal trouble spots as Rwanda, Cambodia, and Sierra Leone. Organizations should consider this best practice in designing systems for enterprise scale governance, risk, and compliance systems in addition to discovery operations.

# Conclusion

The UN Information Management team supporting the war crimes tribunals was able to bring into place a complex system with many innovative features in search processing in a short timeframe through establishing a strong partnership with their selected vendor, ZyLAB, and through taking advantage of highly capable processing capabilities within the software to reduce or eliminate the bottlenecks and delays inevitably associated with manual discovery operations. As a result, the prosecutor within the ICTY tribunal was able to carry out trials with evidence that had been transformed into digital formats and analyzed and produced with full accessibility and consistency for prosecutors, defense attorneys, and judges. This is a significant accomplishment for a steady-state commercial enterprise. The UN work is all the more impressive growing as it did out of the horror of the prosecution of war crimes and the need of the international community to respond quickly and fairly in executing justice.

# About ZyLAB

ZyLAB's industry-leading, modular eDiscovery and Intelligent Information Governance technology puts organizations in command of boundless enterprise data in order to increase productivity, mitigate risk, reduce costs, investigate matters and elicit business knowledge and intelligence.

The company's products and services are used on an enterprise level by corporations, government agencies, courts, and law firms, as well as on specific projects for legal services, auditing, and accounting providers. ZyLAB systems are also available in a Software-as-a-Service (SaaS) model.

ZyLAB is positioned by Gartner as one of the strongest "Visionaries" in the 2013 and 2014 Magic Quadrants for eDiscovery Software and has received numerous other industry accolades over the last 3 decades.

Headquartered in Amsterdam, the Netherlands and McLean, Virginia, ZyLAB also serves local markets from regional offices in New York, Barcelona, Frankfurt, London, Paris, and Singapore. To learn more about ZyLAB visit [www.zylab.com](http://www.zylab.com).

## **COPYRIGHT**

No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written consent of ZyLAB Technologies B.V. The information contained in this document is subject to change without notice. ZyLAB Technologies B.V. assumes no responsibility for any errors that may appear. All computer software programs, including but not limited to microcode, described in this document are furnished under a license, and may be used or copied only in accordance with the terms of such license. ZyLAB Technologies B.V. either owns or has the right to license the computer software programs described in this document. ZyLAB Technologies B.V. retains all rights, title and interest in the computer software programs.

## **DISCLAIMER**

This Case Study is educational in nature and is not legal advice for an organization's particular circumstance. This White Paper is for informational purposes only. ZyLAB Technologies B.V. makes no warranties, expressed or implied, by operation of law or otherwise, relating to this document, the products or the computer software programs described herein. ZYLAB TECHNOLOGIES B.V. DISCLAIMS ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE. In no event shall ZyLAB Technologies B.V. be liable for (a) incidental, indirect, special, or consequential damages or (b) any damages whatsoever resulting from the loss of use, data or profits, arising out of this document, even if advised of the possibility of such damages.