A Case for Judicial Data Warehousing and Data Mining in Kenya

Christopher Moturi[*], Rahab Mburu, Njeri Ngaruiya

School of Computing and Informatics, University of Nairobi, Nairobi, Kenya
*Corresponding author: moturi@uonbi.ac.ke

Abstract This aim of this study was to demonstrate how the Extract, Transform and Load (ETL) process can be utilized to assist the Kenyan Judiciary address challenges of data integration in its operational systems and hence provide better mechanisms for extracting data to allow easier reporting and generating judicial intelligence. The research determined the common data sources and operational systems, demonstrated, using case returns data, how the ETL process can be used to migrate data from sources to a data warehouse, proposed a framework for an ETL environment, and developed guidelines for creating a data warehouse for the Kenya Judiciary. This is in line with the Kenya Judiciary Transformation Framework that seeks to harness Information and Communications Technology as an enabler in the justice system in order to achieve expeditious delivery of justice. The practical implication of this work is the better preparation of judiciaries with limited adoption and utilization of ICT in laying the groundwork for judicial knowledge discovery.

Cite This Article: Christopher Moturi, Rahab Mburu, and Njeri Ngaruiya, "A Case for Judicial Data Warehousing and Data Mining in Kenya." *American Journal of Computing Research Repository*, vol. 4, no. 1 (2016): 7-14. doi: 10.12691/ajcrr-4-1-2.

1. Introduction

The Kenya Judiciary Transformation Framework [1] calls for quick administration of justice to the Republic of Kenya. As mandated by the constitution of Kenya, the Judiciary is the custodian of justice. The Judiciary is expected to make major reforms set out in this transformative framework premised on four key pillars: people focused delivery of service; transformative leadership, organization culture and professional, motivated staff; adequate financial resources and physical infrastructure; and harnessing ICT as an enabler for justice. These pillars have been adopted to boost efforts geared towards yielding a certain level of performance. The proper adoption of ICT in the justice system will boost all the departmental operations with the end result being acceleration towards achieving expeditious delivery of justice.

The Kenya Judiciary has had very limited adoption and utilization of ICT with the result being inefficiency and ineffectiveness in the administration of justice [1]. There is great need to improve on how information is being collected and processed in various aspects including case management, court records, recording of proceedings and transcriptions, integration of systems, and public access. The nature of the data sprouts a challenge on data warehousing and data mining because of the difficulties encountered during the process of capturing, documenting and storage in operational databases. Court data collection needs to be improved and a repository of all vital information from various departmental databases needs to be considered as a way of providing a basis for decision making and generating judicial intelligence. Varying efforts have been initiated in the Kenyan Judiciary to provide standardized data collection across all the courts but they are not giving a sustainable solution.

There is need for a way of extracting and storing data to allow easier reporting as well as perform predictive analytics. This process of collecting and transforming data once improved would boost the efforts geared towards achieving high levels of efficiency as outlined in the JTF, without slowing down or interfering with the operational systems.

A number of initiatives in data collection have yielded creation of relational databases such as the Case Management System which prompts the Judiciary in building a data warehouse in order to boost information storage, backup processes and get new patterns which can lead to better decision making. Increasingly large amounts of judicial data have been stored and this volume is ever increasing, yet despite this wealth of data, the judicial system has been unable to fully capitalize on its value. Data mining can be used effectively in judiciary to extract knowledge from a repository of information or a knowledge base with an aim of discovering trends, improving decision making, developing performance indexes, understanding the judicial operations, and providing judicial data for integration with other organizations. Data mining as a technique consists of different tools can be used appropriately in the area of data

modeling, data surveying and data preparation (building the right model) for the mining process. During data preparation, the data miner is able to gain proper understanding of its contents, use and applicability, limitations of the data as well as come up with quality models that can be adopted for the mining process. A three step process known as ETL (Extract, Transform and Load) will be efficient and effective in carrying out mining in judicial data. ETL entails data to be extracted from an identified source system, loaded into a staging area where various transformations are done to ensure that data is in a format that is acceptable, and finally loading it into a destination ready for exploration with a data mining tool.

This aim of this study was to demonstrate how the ETL process can be utilized to assist the Kenyan Judiciary address its problem of data integration and hence prepare for Judicial Data Warehousing and Data Mining. The research sought to:

a) Determine common data sources and operational systems in judicial environment and display the kind of data available and the pre-processing needs

b) Demonstrate, using case returns data, how the ETL process can be used to migrate data from sources to a data warehouse

c) Propose a framework for an ETL environment with its associated architectures

d) Develop a guideline for creating a data warehouse for the Kenya Judiciary.

2. Related Work

Various studies have demonstrated successful use of data mining in judicial systems. Reference [2] proposed a conceptual model for Decision Support Systems based on Data Mining and Case-Based Reasoning. The model has capacity of self-learning, identify association between data, classifying and clustering of the data based on the characteristics and suggest recommended actions to users through incorporation of intelligence to the system thus increased capacity in problem solving and accuracy. This model demonstrates a relation between the Data Mining, Case-Based Reasoning System and Decision Support Systems.

Criminology is one of the most important fields where the application of data mining techniques can produce important results. Hussain [3] proposed a criminal behavior analysis model using data mining. The model interlinks the common set of universal principles with the attributes of the individuals for profiling the criminal behavior including personality, criminal activity, extrinsic factors and the attributes of individuals.

Kufandirimbwa and Kuranga [4] argue for the importance of adopting mining of judicial data by developing an online analytical processing and data mining model for judicial system. The study illustrates how data mining can be used to perform automatic summaries, help discover hidden patterns and display the results to the end user. Their approach adopted use of Microsoft Structured Query Language server to create the OLAP and Microsoft Excel to present the information to the end user.

Singh and Singh [5] highlights the dimensions of data quality in data warehousing that includes accuracy, reliability, importance, consistency, precision, timeliness, fineness, understandability, conciseness and usefulness. Data quality problems may occur when there are poor handling processes, lack of well-defined data entry and maintenance procedures, errors emanating from the ETL process and external data that does not fit the expected organization standards.

Thammaboosadee and Silparcha [6] proposed a framework for criminal judicial reasoning system using data mining to determine reasons from court verdicts. Judicial reasoning is one of the most complex legal activities in court. The study developed a system with modules based on legal rules and principles that are used to construct knowledge system that used data from the Thailand Supreme Court verdicts on criminal cases as training data set. They demonstrate the use of XML standard in document structuring and in supporting judicial decision support system that guides in judgment supported by law theories and principles.

Xie [7] analyzed case guiding system and the necessity of setting up such a system with the aim of saving judicial resources, promoting judicial justice, and improving judicial efficiency. The paper outlines the management of the overall design of case guiding system including the outline of the system, the composition of every subsystem, each function module, and the basic technical requirement of the information database and systematic operation characteristics.

Annapoorani and Vijaya [8] developed a tool for sorting and summarizing judicial data simultaneously into various groups based on the title and thus making it simpler and easier in searching for the required data.

Ticom et al. [9] present a labor-related judicial application of unstructured data mining techniques using probabilistic methodologies, linear regression and rule induction in text categorization. This text mining is essential in discovering previously unknown patterns useful for judicial applications.

An analysis of the main systems of data collection of the Brazilian Judiciary point out the future prospects of new computing technologies such as data mining and cloud computing [10]

Hamin et.,al [11] outlined the technical applications available in the Malaysian courts and highlight how the growing adoption of ICT by the courts in many jurisdictions raises several issues of security risks. The paper argues that despite the common notion that law will always lag behind technology, the courts would need to understand these legal and non-legal security risks and manage them in the most efficient and effective manner if judicial business were to be continually advanced by the ICT.

Ward et al. [12] explored the challenges confronting organizations in responding to the electronic discovery amendments to the US federal rules of civil procedure, considering that the vast majority of information and data is electronic and stored in numerous files and on a variety of media, and concluded that electronic information and data are discoverable and therefore requires protection. The recommend an enterprise-wide multi-functional electronically stored information Discovery Team to

develop, implement, and periodically review electronic records management policy and procedures.

Zernik [13] discusses how data mining was conducted through the online public access system to examine the validity and integrity of records and of the system as a whole. They report how many records were not verified, records were universally missing their authentication counterparts as required by law to render them valid and effectual. The system as a whole was deemed invalid. The public was unable to discern the difference. The author asserts that case management systems of the courts must be subjected to certified, functional logic verification.

By developing an android application for retrieving law information using data mining methods, [14] have demonstrated the advantage of reducing time lawyers India take to search for material for their cases.

Using a two-stage classifier according to the concept of machine learning, [15] proposed a framework to identify the relevant law articles consisting of sentences and range of punishments, given facts discovered in the criminal case of interest. The first stage determines a set of case diagnostic issues, while the second stage determines the relevant legal elements which lead to legal charges identification.

Bianchi et al. [16] have developed ICT tools that combine statistical and the semantic approaches to support legal professionals in exploring a complex corpus of norms and documents in the legal domain. They demonstrate the use ICT to facilitate search and retrieval of information in large archives in the legal domain.

These studied affirms the dire need for the Kenya Judiciary to adopt use of Data Warehousing and Data Mining to aggregate data in court files to promote better management of the judicial system and ultimately this will result in acceleration towards achieving expeditious delivery of justice as envisaged in its Judicial Transformation Framework. Information management policy should focus on encouraging public participation in the judicial process and discouraging practices which undermine the administration of justice [23]. Potential solutions are multifaceted, interrelated, require new and creative ways, and must be coordinated by using all the tools available such as rules, training, and technology, in order to further the twin goals of public access and privacy within judicial information management.

3. Methodology

3.1. Research Design

This study adopted a descriptive study that entailed generating primary data through use of interviews as well as analyzing existing data collected overtime in form of court case details. The structure and content of the existing data sources was done with an intentional mapping to the common data warehouse model.

3.2. Data Source & Collection

Court registries are charged with the responsibility of collecting and storing data as they register the court cases details on daily basis. Judicial employees working in the registries were targeted for interview: court clerks, registrars and executive officers. In addition pre-determined questions were put to senior ICT staff. The key elements of the interviews focused on the documentation on the existing systems. These include registers used to enter case application on daily basis, user manuals and system documentation of operational systems. Observations were made on how registers are filled and court proceedings in order to identify the possibility of recording data in real time by the court clerks.

3.3. Modeling the ETL Process

Extract, Transform, and Load (ETL) processes physically integrate data from multiple, heterogeneous sources in a data warehouse. ETL tools are pieces of software responsible for the extraction of data from several sources, cleansing, customization and insertion of the data into a data warehouse.

Consolidating data into a single physical repository has proven to be the most effective approach to provide fast, highly available, and integrated access to relevant information [17]. The extraction phase largely deals with the technical heterogeneity of the different sources and imports relevant data into a staging area. The transformation phase is the heart of an ETL process whereby all the data is brought into a common data model and schema using mapping technology, standardized consolidated to a single representation. In the load phase, the integrated, consolidated, and cleaned data from the staging area is loaded into the appropriate databases of the warehouse.

In large organizations such as the Kenya Judiciary there are many data sources from different departments, many ETL processes, which may cover shared data sources, same data targets, common sub-processes and stages configured in an equal or similar way.

The generic model for ETL scenarios [18] and the methodology for conceptual modeling of ETL process [19] were used in analyzing the contents of the existing data sources and how they will map into the warehouse model developed, while considering the existing constraints originating from data attributes. Precautions were taken in the data integration process, particularly with regard to privacy-sensitive data [20]. The ETL process requires high level management to enable their flexible re-use, optimization, and rapid development. The proposed ETL environment is shown in Figure 1
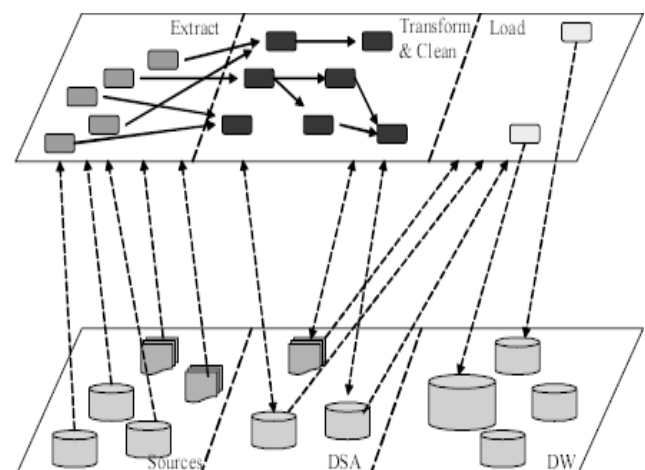

Figure 1. ETL Environment

4. Results and Discussion

Our results reflected the objectives of the study which included identification of data sources available in the judiciary that can provide a platform for data extraction, determination of data pre-processing needs in preparation for mining, and identification of an ETL tool for collecting data and analyzing it. This section describes the results obtained from demonstrating how the ETL process can be utilized to solve the Kenyan Judiciary problem. The section starts with a description of the technology used and then shows how data can be collected using a standardized tool and loaded into a technology platform. This data is then extracted using a sample ETL tool, transformed and loaded into the final repository which is the data warehouse. This paper proposes to demonstrate how ETL process can be utilized as a vital tool for data preparation as well as for populating a data warehouse.

4.1. Technology Description

a) Data Warehouse Architecture

The data warehouse architecture described in [21] was adopted (Figure 2). This includes tools for extracting data from multiple operational databases and external sources; for cleaning, transforming and integrating this data; for loading data into the data warehouse; and for periodically refreshing the warehouse to reflect updates at the sources and to purge data from the warehouse, perhaps onto slower archival storage. In addition to the main warehouse, there may be several departmental data marts. Data in the warehouse and data marts is stored and managed by one or more warehouse servers, which present multidimensional views of data to a variety of front end tools: query tools, report writers, analysis tools, and data mining tools. Finally, there is a repository for storing and managing metadata, and tools for monitoring and administering the warehousing system.
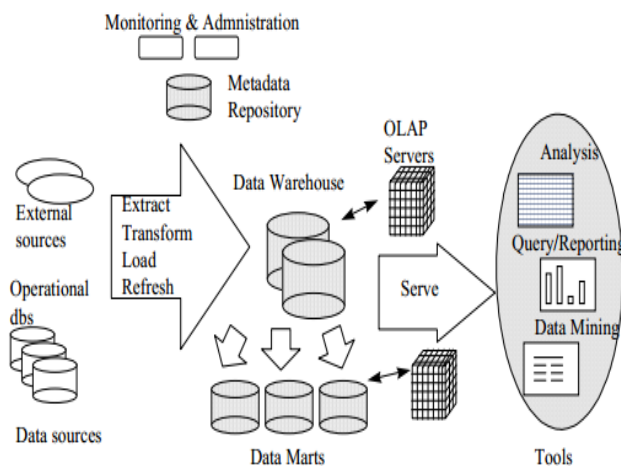

Figure 2. Data warehouse Architecture (Source: Chaudhuri and Dayal, [21]

b) Software Platform

The judicial data warehouse requires use of a strong database that has the capacity to accommodate integration and analysis services. The integration services are necessary for performance of the ETL process. The analysis services engine is necessary for performance of the data mining and reporting tasks. For this project the software that was adopted for use is Microsoft SQL Server 2008 as the DBMS and Rapid Miner as the ETL tool. Rapid Miner is popular data analytics software. Microsoft Excel was used in uploading data into a centralized database. Ubuntu operating system and Windows Operating system were the preferred platform for operation. Although these were the preferred solutions, several other platforms could be tried to experiment which will work optimally. Such will include Oracle DBMS with Rapid Miner as the ETL tool.

4.2. Data Sources

The main sources of data in the Kenya judiciary include Court Management System (CMS), Judicial Help Desk (JHD), Integrated Financial Information System (IFMIS) and flat files. To extract data from the above sources the following key steps were followed:-

a. Collect acceptable data that is in a standardized format: This was achieved by creating a template for the manual records (flat files) and standard reports from the CMS, JHD and IFMIS.

b. Collected data channeled to a staging area (for cleansing and manipulation).

c. Data was loaded into the warehouse which is a database created in SQL server.

This study concentrated on the handwritten register that is usually stored in the courts registry. Part of this data is captured differently by various court registries and entered in Excel file on daily basis in form of case returns. However, vital information often goes uncollected thus remains in the registry files. As part of computerization efforts the data has been collected and filled in Excel sheets.

4.3. Data Collection and Analysis Challenges

The results who that the existing system has its flaws which include:

a. Data Collection

1. Lack of a standard way of collecting data across all registries and thus the process has been rejected by some court stations.

2. Some field computations are done manually thus a possibility of producing erroneous information.

3. Physical files pertaining to a case application, which are usually stored in the registry department, contain a lot of information that is often not captured in the registers. This information includes financial aspect of the cases e.g. fines, bails and charge sheet information.

4. Any judicial staff is given the responsibility of updating the data on daily basis. This data is hand written as it is acquired over the counter meaning it is subject to modification without trace.

5. The data collection chain is too long because of the many people involved.

b. Data Collection Template

A template for use across the registries commonly known as STAT 1 and STAT 2 that had been initiated has faced some challenges in terms of being tedious to use, complicated to users, not capturing all case details, and no existing database.

c. Data Pre-Processing

Data preparation is a key aspect of transforming this data into a manner that can be transformed using an

appropriate ETL tool. Since there is no agreed standard of collecting and recording the data, it becomes hard to export or import the data because of omitted fields, blank fields, repetition, and typingspelling errors.

4.4. Proposed Data Collection and Analysis Solution

In order to curb the above challenges the following approaches may be adopted in attempting to solve the problems associated with data collection and analysis.

1. Acquisition of a standardized data collection tool that can be used to collect case details across all the registries. The use of this will avoid delays in sending data especially when the email server is not operational.

2. Acquisition of a proper ETL tool that is applicable to the scenario e.g. Rapid Miner server software.

3. Procurement of a comprehensive database management system that comes with Business Intelligence tools.

a. Standardized Data Collection Tool

The collection tool is an Excel worksheet whose fields are organized to reflect the contents of the documents as they appear in the original files. Vital information was extracted and filled in this worksheet. To avoid errors data validation was enhanced at input so that there is a drop down list that acts as a guideline to selecting fields that are pre-known. This application will ease the data collection and retrieval process. Once filed the worksheet is shared using share point over the intranet architecture to the end-user who is the performance analyzer.

b. The ETL Tool

The ETL tool allow the user to receive files or other data from vendors or other third parties which needs to manipulation in some way and then insert into your the database. In the context of this project, the main role of this tool is to help extract the data that was collected using the standardized Excel sheet, perform possible manipulations to the data and thereafter save the work in a database for a future miner. Rapid Miner has been adopted as the ETL tool of choice. The standardized template can be extracted directed using ETL process operators available in Rapid Miner. These include processes such as Read Excel, Read csv, Read Database.

c. Business Intelligence Tool

Business intelligence platforms will allow the judiciary to monitor data, analyze judicial trends, and generate judicial intelligence. The importance of using business intelligence tools is crucial for the success of judicial data mining. It is not an easy task to decide which business intelligence tool to select and use. An evaluation of major business intelligence tools has been in terms of provide by [22]. The tools that can be adopted include commercial products such as IBM Cognos, Microsoft BI, MicroStrategy and Oracle BI as well as open source i.e. JasperSoft, Pentaho, SpagoBI and Vanilla.

4.5. The ETL Process

The ETL tasks comprised of a three-step process as described in the Table 1:

a. Data Extraction and Preparation

The task involved gathering data from external sources and bringing this data to the target systems and databases. The data in question is located in spreadsheets that are not integrated with any master database. The goal of this task was to understand the format of the data, assess the overall quality of the data and to extract the data from its source so it can be manipulated in the next task. Data required for this study was obtained from registers that are entered daily from the courts registries. This data was then recorded in Excel worksheets. This data was collected, merged using available Excel tools and SQL queries and where applicable the data was merged using Rapid Miner. Prior to extracting, the data needed to be formed in such a way that it became acceptable for extraction using any preferred ETL tool. Figure 3 illustrates a proposed solution where there is no existence of a central repository as data is collected through an email system.
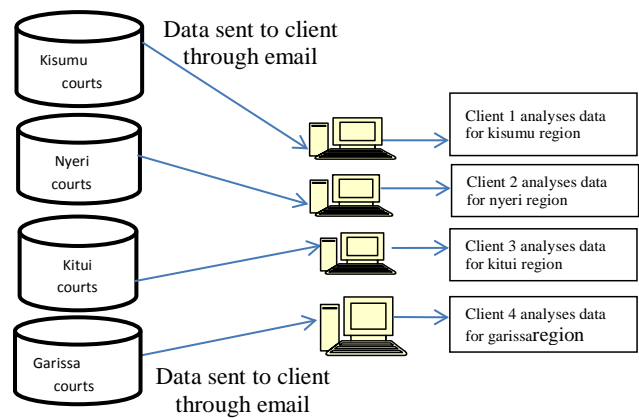

Figure 3. Proposed solution with no central data repository

Table 1. The ETL Process

Stage	Description	Tasks involved	Operations
Data Extraction/Data Preparation	This stage involved getting the data and describing it using metadata. It also called for further exploration with various tools as one attempts to verify its integrity.	Getting the data from the identified source systems. Creating a metadata reference table.	Read Excel Read CSV Read Database
Data Transformation	This stage involved collecting and categorizing data into respective regions they came from, cleaning the data and reconstructing it such that it is ready for exploration with an ETL tool. The data was formatted using appropriate criteria	Inserting and removing unnecessary columns and rows. Filling in the gaps with the missing attributes. Introducing new attributes that will contribute to the completeness of data. Refining the metadata reference tables. Creating	Data cleansing Replace missing values Fill Data Gaps Remove unused/unnecessary values Filtering Remove duplicates Sorting Sort Ascending/Descending Shuffle
Data Loading onto DW	This involved transferring data	Write Database Write Excel Write CSV	Exporting the transformed data to an a different destination.

b. Data Transformation

This stage calls for use of a variety of software tools and even custom programming to manipulate the data so that it integrates with data you already have. For example, consider data that is collected with errors such as missing attributes, data gaps, unnecessary values, and duplicates among others. It is necessary to transform this data and make the corrections so that it matches up with the data that currently resides in databases and systems. Data that is collected from the registry can be referred to as "dirty". This is because it is subject to a lot of modification before it is loaded into the staging area. The following are error issues relating to the data:

i. Missing values/attributes - When the Excel files are sent, they are found to be lacking some values such that it makes it difficult to even import them into the database. For example some fields like date are left blank and such fields need to be filled with the correct data probably from the physical register.

ii. Introduction of new attributes - This was introduced to help censor the data so that the information is not displayed as it is originally from the register.

iii. A possible modification of data such as unique application numbers, case determinant id assigned and displayed instead of judge names, grouping entities, and format type such as for dates.

c. Loading Data

The transformed data was moved from the staging area to the warehouse which is a database that is in SQL server ready for mining. Once data is successfully transformed it

is important to load it into the database. However, before data loaded, it is important to have a backup of current system in case of failure. After loading the data, it is common to run audit reports so one can review the results of the merged databases and systems to make sure the new data has not i-ntroduced any errors.

4.6. Implementation

The implementation of the ETL was in form of tables which are created using SQL Server database. Data can also be stored in XML format which is text based. It should have the capability to split data, manipulate, cleansing data through rules that enforce the integrity of data. Cleansing entails removing unwanted fields, merging fields, and inserting new fields to achieve standardization. Figure 4 demonstrates architecture of Kenya judiciary data warehouse architecture illustrating how the ETL process can be carried out in the judiciary using the existing operational systems as data sources. It further includes all other possible sources from systems that are being implemented or are in the process of implementation. The extraction, transformation and loading tools was chosen based on the user preference from a list available. The data once transformed is deposited in a knowledge repository or a data warehouse that will house all the different types of data. Thereafter, data mining and analysis tools will be used to generate further insight or gather more intelligence.
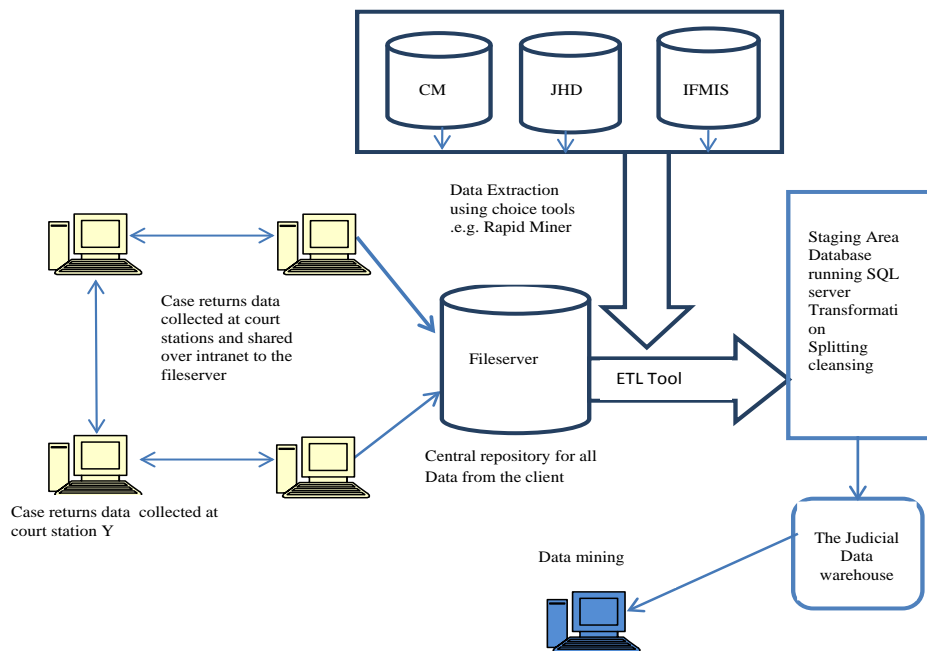

Figure 4. Kenya Judiciary Data Warehouse Architecture with a Staging Area

The Staging Area

Before running an ETL process it is important to have an intermediate storage area whose main purpose is to help in the data processing. This sits between the data sources and the final destination for the data which is the warehouse. This area keeps on being refreshed by new updates of files and data before, during and after a successful ETL process. There are staging area architectures which are designed to hold data for extended periods of time for archival or troubleshooting purposes.

These architectures can be implemented in form of tables in relational databases, text-based flat files e.g. XML files stored in file systems or proprietary formatted binary files stored in file systems. The staging area used for this study is SQL server and Ms Excel application. Data collected from the registers and recorded in Excel worksheets have a lot of inconsistencies because the person in charge has to count the case of a given instance manually. Therefore it generates possible errors such as: modification either intentionally or no intentional by a third party; sending of

these data is through email system and when the email system is down there is a lot of delay in sending data; and the data collected goes through many third parties thus possible introduction of chances of modification.

4.7. Security and Privacy Issues

The security, trust and privacy of the envisaged huge information base is crucial for the sustainability and reliability of data warehouse. An assessment will be required on existing data security mechanisms, focusing on specific issues and requirements concerning data warehousing environments, challenges and opportunities. Security of data should be highly prioritized. This can be done by implementing a standard security policy across the 147 court stations.

Reference [11] raises several issues of security risks from the adoption of ICT in the Malaysian courts and suggests basic security controls such as authentication, non-repudiation, confidentiality, data integrity and privacy encroachment. The Kenya Judiciary would need to understand the legal and non-legal security risks and manage them in the most efficient and effective manner in order for the implementation of the data warehouse to succeed. The data warehouse should be used with caution, to avoid incorrect conclusions and to prevent discrimination and stigmatization of certain groups of individuals [20].

The Judiciary would need to assess the data quality, develop design & build standards for quality assurance. This may entail defining requirements for data quality, data management procedures, report validation, performing tests on the ETL logic and business rules.

4.8. Training

As assessment of the training needs determined that a comprehensive training is required for the various cadres of staff who will use the data warehouse. This is particularly so for those working in the registries who are charged with the responsibility recording the data: court clerks, registrars and data entry clerks.

As part of the rollout plan, the Judiciary will require defining, developing and documenting an incremental training plan for technical and end user staff. A training strategy will include identification of specifics on requirements, roles, objectives, materials, and creation of training databases.

4.9. Guide for Creating a Data Warehouse

A guide to design of data warehouse for the Kenya Judiciary was prepared. The guide covers detailed description the importance of data warehouse and data mining in the Kenya Judiciary, the various data sources available, the ETL architecture for modern data warehouses in which all or most data transformation is performed on the database that hosts the data warehouse. The guide describes the proposed data collection and analysis solution and the typical day to day activities to be executed. The issues cover hardware, network and software platforms, as well as security requirements. A step-by-step rollout map for the data warehouse covers design, technical requirements, data acquisition, quality assurance, meta data management, data access, building, documentation, testing, training, installation, transition,

and post implementation support. The suitable tools for building, managing and using the data warehouse have been identified. This guide would be beneficial as the Kenyan Judiciary works toward s judicial data warehousing and data mining its transformation framework of harnessing use of ICT as an enabler in the justice system.

5. Conclusion

With advances in ICT, sophisticated data mining and business intelligence tools are increasingly accessible to the law enforcement community. Data mining is increasingly used for analysis to give out new knowledge which can assist in decision making. This study sought to demonstrate how the ETL process can be utilized to assist the Kenyan Judiciary prepare for Judicial Data Warehousing and Data Mining. We have shown the major areas of focus towards creation of a data warehouse. The study has managed to automate the data collection process through use of a standardized data collection tool and demonstrated how to use ETL process to prepare data for loading into a data warehouse. The data sources available in the Kenya Judiciary have been clearly illustrated. The conceptual framework for the ETL process has been developed. A guide to developing a data warehouse for the Kenya Judiciary has been prepared and submitted. The guide proposes the setting up of an Intranet, setting of a backup storage and centralized data, comprehensive user training, building of the Data Warehouse, Use of Automation in Populating the Warehouse and an enterprise-wide multi-functional team to implement.

The real benefits once a Kenyan Judiciary Data Warehousing includes use of data mining in searching and sorting information [2,7,10] judicial judgment [6,7,15] Knowledge Discovery [9], and criminal behavioral analysis [3].

Acknowledgement

References

[1] JTF (2012). The Kenya Judicial Transformation Framework 2012-2016 available at www.judiciary.go.ke.

[2] Verma, L., Srinivasan, S., Sapra, V. (2014). Integration of rule based and case based reasoning system to support decision making, Proceedings of International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT),pp106-108.

[3] Hussain, K.Z., Durairaj, M., Farzana, G.R.J. (2012). Criminal behavior analysis by using data mining techniques, Proceedings of International Conference on Advances in Engineering, Science and Management (ICAESM), pp656- 658.

[4] Kufandirimbwa, O., & Kuranga, C. (2012). Towards Judicial Data Mining: Arguing for Adoption in the Judicial System. Online Journal of Physical and Environmental Science Research, 1(2), 15-21.

[5] Singh, R., & Singh, K. (2010). A descriptive classification of causes of data quality problems in data warehousing. International Journal of Computer Science Issues, 7(3), 41-50.

[6] Thammaboosadee, S., Silparcha, U., (2008). A framework for criminal judicial reasoning system using data mining techniques, Proceedings of the International Conference on Digital Ecosystems and Technologies, DEST 2008. 2nd IEEE, pp 518-523.

[7] Xie, R., (2008). The Application of Data Mining in Judicial Judgment, Proceedings of the 4th International Conference on on Wireless Communications, Networking and Mobile Computing, WiCOM '08, pp 1-4.

[8] Annapoorani, V., Vijaya, A., (2013). A prevailing *judicial* package for clustering and sorting information extraction, International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME), pp241-244.

[9] Ticom, A.A.M, de Souza, B., de Lima, L.P., (2007). Text Mining and Expert Systems Applied in Labor Laws, Seventh International Conference on Intelligent Systems Design and Applications, ISDA 2007, pp788-792.

[10] Serbena, C. A. (2013). Actual interfaces between Q-Justice and E-Justice in Brazil. Revista de Sociologia e Política, 21(45), 47-56.

[11] Hamin, Z., Othman, M.B., Mohamad, A.M., (2012). ICT adoption by the Malaysian high courts: Exploring the security risks involved, International Conference onInnovation Management and Technology Research (ICIMTR), pp285-289.

[12] Ward, B. T., Sipior, J. C., Volonino, L., & Purwin, C. (2011). A United States perspective on electronic discovery rules and electronic evidence. Transforming Government: People, Process and Policy, 5(3), 268-279.

[13] Zernik, J., (2010). Data Mining of Online Judicial Records of the Networked US Federal Courts, International Journal on Social Media, MMM: Monitoring, Measurement, and Mining, pp 69-83.

[14] Poonkuzhali, S., Kumar, R. K., & Viswanathan, C. (2015). Law Reckoner for Indian Judiciary: An Android Application for Retrieving Law Information Using Data Mining Methods. In Advanced Computer and Communication Engineering Technology (pp. 585-593). Springer International Publishing.

[15] Thammaboosadee, S., Watanapa, B., & Charoenkitkarn, N. (2012). A framework of multi-stage classifier for identifying criminal law sentences. Procedia Computer Science, 13, 53-59.

[16] Bianchi, M., Draoli, M., Gambosi, G., Pazienza, M. T., Scarpato, N., & Stellato, A. (2009). ICT tools for the discovery of semantic relations in legal documents. In Proceedings of the 2nd International Conference on ICT Solutions for Justice (ICT4Justice).

[17] Albrecht, A., & Naumann, F. (2008). Managing ETL Processes. NTII, 8, 12-15.

[18] Vassiliadis, P., Simitsis, A., Georgantas, P., Terrovitis, M., & Skiadopoulos, S. (2005). A generic and customizable framework for the design of ETL scenarios. Information Systems, 30(7), 492-525.

[19] Simitsis, A., & Vassiliadis, P. (2008). A method for the mapping of conceptual designs to logical blueprints for ETL processes. Decision Support Systems, 45(1), 22-40.

[20] van den Braak, S., Choenni, S., & Verwer, S. (2013). Combining and analyzing judicial databases. In Discrimination and Privacy in the Information Society (pp. 191-206). Springer Berlin Heidelberg

[21] Chaudhuri, S. and Dayal, U (1997). An overview of data warehousing and OLAP technology. ACM Sigmond, 26(1) pp 65-74.

[22] Bernardino, J., & Tereso, M. (2013). Business Intelligence Tools. In Computational Intelligence and Decision Making (pp. 267-276). Springer Netherlands.

[23] Winn, P.A., (2009). Judicial Information Management in an Electronic Age: Old Standards, New Challenges, The Federal Courts Law Review, Vol 3, Issue 2, pp 135-176.