Capstone Project Cover Sheet

Capstone Project Name: Visualizing Lesson Plans

Student Name: David Andrew Nichols

Degree Program: Software Development

Student Mentor Name: Rita Kemner Salyer

Table of Contents

Introduction

The world is flooded with data, but too often that data is unused, deemed too difficult to analyze, or misunderstood. Schools are not unusual in this sense: each producing large amounts of data for student and teacher alike. Especially in an era where curriculums and educational priorities are shifting to emphasizing *written standards*, there are ample reasons that schools need to now use the data they have been collecting: to ensure they are fulfilling their own mission, to verify to independent standards committees that they meet said standards, and to help students understand their own education.

This project aimed to assist a local private school, Ipeka IICS, to understand their own data. The school teaches based on standards and is rapidly moving towards getting all departments and teachers on board with this education practice. They have been implementing requirements for teachers to produce daily lesson plans to track what they are teaching. Critical to convincing students, parents, and accreditors that what they are teaching is what they *say* they are teaching is analyzing and understanding the data that they already have.

To achieve this goal, the author set up an automated workflow for processing existing and future date into a persistent database. This database will be queried for desired information as needed, and the results *visualized* in a simple-to-understand way. Administrators, coordinators, teachers, parents, and students will all be able to see immediate and relevant data about how they are being educated.

In this report, existing work is reviewed, both from an educational and a technological perspective. A rationale for why this particular approach is suggested is examined. The project was the work of a single developer, so an agile software approach was preferred as one accommodating potentially rapid shifts in requirements. The established goal, objectives for how to achieve that goal, and milestones on the road to those objectives are discussed.

For the implementation of these goals and objectives, the Elastic software stack was selected, comprised of Beats, Elasticsearch, and Kibana. Despite some difficulties in communication and a somewhat reduced scope of dataset, a custom Beats software utility was written to perform the extraction and transformation of data. An Elasticsearch database was configured to hold this data. Visualizations of this data were prepared using a Kibana webpage, seen in Appendix B.

Review of Other Work

Educational Work

Numerical data may once have been irrelevant in the halls of academia and the bustling classrooms full of students, but in an era that increasingly emphasizes written *standards* over tradition, this is no longer desirable. (Carpinelli 2008) If standards were designed to create equitable student outcomes across different schools or areas, how are educators to know whether their goal has been met?

Evaluation of standards-based programs may be desirable but impossible due to lack of data availability. Although great effort has been spent on preparing standards-based curriculum and standards-based testing strategies, Carpinelli questions whether this effort has translated into lesson plans that teachers generate on a daily basis. Thus, the aim of this project is to examine these ground-level documents that may represent the closest data to the act of teaching itself. (Carpinelli 2008)

But why is evaluation necessary? In previous studies, two primary reasons for evaluation occur: to identify teacher efficacy in educating students, and to identify learning outcomes of students themselves. Torff puts plainly why it is necessary to evaluate teacher efficacy: the increasing demand for teachers is not being supplied by traditional avenues, so a variety of alternative pipelines for producing certified teachers have emerged. Whether these pipelines are producing effective teachers is unclear. (Torff 2009)

There are two main approaches to assessing teacher effectiveness: quantitative and qualitative. The quantitative approach has generally meant reducing teachers to a collection of student test grades. The qualitative approach gives principals and other supervisors the task of assessing teacher competence, but potential bias reigns. Torff leaves undiscussed the idea of quantitatively assessing teachers on the basis of their actual output rather than student success. (Torff 2009)

At least as important as evaluating teachers is evaluating student outcomes, but in some instances the desired accountability is merely theoretical. Avery laments the simple "absence of data on student learning" in the context of gifted learning programs. While it is suggested that states may simply be focusing on the larger body of students, if states cannot make data-driven decisions in the context of a relatively small program, how can they hope to do so as scale dramatically increases? And this is not a forgotten issue: in interviews with various stakeholders, a consistent complaint was the lack of available data. (Avery 2001)

One of the few attempts to look at bottom-up data is Carpinelli. In their study, they developed a rubric-based grading model for lesson plans and recruited and trained a panel to evaluate 200 lesson plans from a variety of sources. They identified several common problems across lesson plans: the use of too many or incorrect standards, lesson objectives that simply do not correlate to stated standards, and assessments that do not correlate it either standards or objectives. Unfortunately, the inter-rater agreement on which lesson plans were poor was weak. (Carpinelli 2008)

What are schools to make of the lack of data, the lack of good ways to interpret it, or the lack of willingness to use it? As it turns out, how to use data to drive decision making has long been a focus of the corporate world, where it is often discussed as "Business Intelligence."

Visualization Work

It appears that data visualization has had limited penetration into educational circles; similar and illuminating is a self-published article by Lisa Spiro, of Rice University. Ultimately expressing skepticism that a submission to an academic journal full of charts and tables of text analysis would be given much interest or even understood, Spiro strongly argues that such analysis can "inform the kind of arguments critics make". In other words, using analytical tools as a complement to expert judgement and experience can provoke entirely new kinds of discussions. Very quickly into her exploration of visualization technologies, Spiro becomes frustrated with the relative lack of power in the tools at her disposal. But even limited tools can help confirm or deny the more common qualitative analysis. Spiro also makes the divide between text analysis and text visualization stark, finding that while hard numbers were more trustworthy, visual impressions "opened my eyes so that I could see the stats more meaningfully." (Spiro 2008)

Nualart-Vilaplana's study of the variety of approaches to text visualization makes this same point with the backing of data: the evidence indicates that information-retention is increased by a "combination of text and illustration." Their study explores the history of the field, attempting to catalogue these techniques. Noting that organizations are in fact flooded with data, they argue for several key factors when selecting a visualization technique. First, because of the amount of data involved, searching for the relevant data must inevitably be a central concern. Second, even when search turns up the correct data, the search results must be presented in such a way as to simplify selecting the right data. Done correctly, searching through massive data can allow for the discovery of new information, possibly even information that could not have been identified without aggregation techniques. (Nualart-Vilaplana 2014)

As a practical matter, most data is simple unstructured text, rather than highly structured JSON or XML that is already easily searchable. This unstructured text must have its information extracted to be usable. Miner discusses the two primary approaches: *rule-based* analysis or *statistical* analysis. For semi-structured texts such as templated Microsoft Word documents, a rule-based approach seems superior, since the structural elements that do exist can be converted into rules for extraction. (Miner 2012)

A final insight from Nualart-Vilaplana's study is highly relevant when examining the various tools and techniques that can be used to implement data visualization. For software solutions that dated more than 5 years old, the vast majority of software was simple no longer available. As no organization should want to either continually change their technology solutions or to be stuck self-supporting legacy solutions, this remains a serious question. No particular cause is offered up, but it is indicated that they found a correlation between commercial software and abandoned software. (Nualart-Vilaplana 2014)

Implementation Work

For implementation of the project, the open-source Elastic stack was chosen. The Elastic stack as implemented consists of Beats, Elasticsearch, and Kibana. These three pieces of software work together by design, each handling a separate type of work, neatly mapping onto the project objectives. A *Beat* is a small binary application that executes on the same server as the raw data. The job of this program is simple, to collect data, perform some basic processing of the data, and then send along events to the Elasticsearch index. *Elasticsearch* is a searchable database that holds the data as loaded by Beats, but can also manipulate this data. Finally, *Kibana* is a visualization platform that can be configured to help end-users understand the data that has been collected and manipulated in Elasticsearch. (Gormley)

Creating the *Beats* implementation so that raw word documents could be extracted and sent to the database was key. While no existing implementation existed for the particular use case, extensive documentation was provided by Elastic. Used as a guide was a blog post detailing the process of building a custom *Beats* product, from start to finish. (Sarbu 2017) Once the skeleton of the software existed and was verified to work, the business logic of extracting and transforming data could be executed.

Another key task was preparing visualizations of the data for end users. While Kibana itself, by design, is easy to experiment with produce pleasant looking charts, an Elastic produced webinar demoing various features of the software was also used to consider possible ways of presenting the data. Reviewed were timeline charts, pie/bar charts, and the ability for users to interactively “drill down” or “slice and dice” data. (Bragin n.d.)

Project Rationale

Leadership of Ipeka IICS has expressed a need for aggregate information on several questions: ESLR (expected schoolwide learning results) coverage, biblical integration coverage.

It was also desired to know how well the mandated NESA curricular requirements are being implemented. These data were needed for two major reasons. First, to validate the accreditation requirements for the various accreditation bodies that govern the school. Second, for the internal direction of the school, to validate that what is being taught in classrooms accurately corresponds to organizational vision and goals.

Historically only the IICS high school has engaged in standards-based education, drawing the standards from the nearby international NSW Board of Studies as an attempt to provide added value to students and parents from the national curriculum. Teacher lesson planning has lagged behind, which is not surprising as this phenomenon is discussed in some detail in Torff. But in recent years an emphasis on gathering data has led to several departments producing consistent and detailed daily lesson plans that ideally conform to the written curriculum maps.

At the same time, a greater awareness of the benefits of using standards to measure education has permeated the rest of the school and parental community. It was demanded that a consistent set of standards were applied not only to the high school but also the middle and primary schools. The process of implementing these practices at other levels began in 2014.

But gathering of data is not the same as understanding data or having actionable data. Last year, when school leadership demanded to know how well biblical integration coverage was progressing (a key organizational value), the only solution offered to collate this data was reading through a sampling of lesson plans by hand. Although teachers were using digital templates to build lesson plans, the produced documents were currently stored in a shared school folder. Data on the drive is useless.

Thus the project was to build an automated system to perform a standard ETL (extract, transform, load) process on the existing digital documents, sending them into a text analysis engine. Further, automated and on-demand visual reporting on the analyzed data was configured. This system uses the existing data with minimal changes to employee workflow, but provides the opportunity for school leadership to easily answer the present questions about aggregate data, as well as being flexible enough to inform future data-driven decisionmaking.

Considered and set aside were ideas that would more radically alter the workflow of teachers by having them build lesson plans with a new software solution. While such an approach has its advantages, the size of work required exceeds the scope of the WGU capstone project and potentially duplicates other planned IT solutions.

Systems Analysis and Methodology

Systems Analysis: Original State

The original workflow consisted of teachers preparing lesson plans by copying and modifying a Google Document template file, which has a variety of fields listed relevant to the goals of data extraction. Lesson plan writing has been standardized in this fashion for the past 1-2 years. Completed plans were uploaded to a departmental shared Google Drive folder, where they could be reviewed by subject coordinators and administrators. A single school department was estimated to generate 500-1000 such lesson plans each year.

Both the English and Math departments operated in this manner, albeit with slightly different template files, meaning that either the data extraction process must flexibly identify correct information or the rules for extraction must be modified for each such template. When the project was proposed, it was unclear to what extent standards-based lesson plans were being produced by other school departments, though it was the desire of administrators that all departments should do so.

After being uploaded to the shared Google Drive, lesson plans were rarely referred to again but were stored. The project proposed to draw on these stored lesson plans to do the work of analysis. A key question for implementation was whether the software solution can/should interface directly with Google Drive via the provided APIs or whether it was simpler to interact with documents that are mirrored on a local hard drive.

Systems Analysis: Current State

After project completion, the previous workflow was largely not altered. However, instead of documents gathering digital dust, the developed software utility “Wordbeat” automatically scans mirrored Word documents on a physical hard drive. The data in the documents is extracted and transformed, loaded into a database, and visualized in real-time on a website portal accessible to subject coordinators and school administrators.

Methodology

The present project was being driven by a single developer, the author, so highly structured software engineering methodologies such as the SDLC were quickly rejected as impractical. The requirements for the final project were also not well understood, with some stakeholders arguing that what leadership wants may not be what they need. In view of these realities, an Agile approach was adopted.

Emphasized in this approach are flexibility, iterative design, and continuous stakeholder involvement. It was believed that involvement from relevant stakeholders on a weekly basis was practical, which was considered critical to project success. Each iterative step was intended to conclude with a concrete deliverable and these steps: Requirements Gathering, Analysis/Design, Implementation, and Validation.

Requirements Gathering is a core methodological step which will be focused on heavily at the project outset but repeated as necessary. This step requires that the developer and stakeholders collaborate to try to understand the desired solution and write up requirements in the form of short user stories. A user story represents a concrete narrative of a way in which a user will interact with the system. "A user can login from a central webpage." is a simple example.

The Analysis/Design step consists of analyzing user stories and identifying useful architectural patterns to the desired functionality. Repeated ideas or themes in the user stories are highly relevant to this goal. Importantly, the recommended design needs to be shared with stakeholders early to ensure that the requirements have been properly understood and interpreted.

Implementation consists of actually producing, through code and documentation, the realization of the user stories and design. Typically only one or two user stories will be implemented at any one time, to ensure that effort is not unnecessarily complicated.

Validation is a critical step is this process. Simply, it consists of allowing the stakeholders to use the developed solution to prove that the user stories have been successfully implemented. It is expected that conflict here will be common and require that user stories be rewritten or added, causing the iterative approach outlined to begin again.

Goals and Objectives

Goals

This project aimed to analyze and present school lesson plan data, but to what end? The primary goal behind this project was one concerning organizational strategy. The strategic goal is this: that the school be able to measure how well their teaching matches their educational goals. This is not an unusual organizational goal, since it asks the question: are we doing what we say we will do?

A fundamental problem identified by Torff for schools that implement standards-education. Administrators can proclaim that the school follows certain standards, books matching those standards can be purchased, and curriculum coordinators can draw up maps specifying how the standards will be met. But none of this guarantees that when it comes to the environment of the classroom that teachers will use and implement these standards.

Verification requires data. While analyzing lesson plans is still a proxy for what happens in the classroom, at least this documentation is being produced by the teachers themselves.

This strategic goal was motivated not only by the desire of Ipeka IICS to ensure their educational goals. It was also driven by the need to verify ongoing improvements to the various accrediting institutions that the school works with. These institutions frequently identify school weaknesses and requiring reporting on efforts to strengthen these weaknesses.

This goal is wide reaching and can perhaps never be fully met. By measuring the ESLRs (Expected Schoolwide Learning Results) that teachers implement in their lesson plans, the completed project has measured one aspect of whether the school's educational goals are reflected in its teaching.

Objectives

The goal of this project was to determine whether teaching matches educational goals, and this goal was answered with three objectives. Existing data was extracted and transformed. The prepared data was loaded into a database, and that database was configured to quickly search for complex queries across the available data. These queries are then presented in a simple and pleasing visual way.

First, existing data needed to be made usable. A standard ETL (extract, transform, load) process was applied to existing lesson plan data. Extraction consisted of reading the information in whatever format it already existed. Transformation, a key step, consisted of adding structure to the data in order to make it more easily searchable. These two steps were accomplished by implementation of an *Beats* program called “Wordbeat” that scans for lesson plan documents and transforms their text into structured information.

Second, the data must be loaded into a database configured to allow for fast, complex queries on the dataset that has been loaded into it. Loading was accomplished by deploying an *Elasticsearch* database and configuring the “Wordbeat” program publishing events to this database. This database of lesson plans can be queried to determine how many hours each teacher has spent in the classroom in a given school year, and which topics have been most frequently addressed by each teacher. Because the project continues with visualization, the returned results are not especially human-readable.

Third, the query/search results are presented to users in a friendly and easy to understand visual way. This last step is of utmost importance, since it is critical to not just have data, but to be able to understand it. Users are provided with a web page that allows for reports on current data to easily be seen and downloaded. A *Kibana* service was deployed, configured to connect to the *Elasticsearch* database, and interactive visualizations of the lesson plan data were designed. A dashboard of these visualizations was then prepared for embedding into a normal webpage for the school to use.

Completion

Goal: Measure how well the teaching of school matches the educational goals of the school.

This goal has been met.

Objective: Extract and transform the raw data.

This objective has been accomplished.

Objective: Load the data into a searchable database.

This objective has been accomplished.

Objective: Visualize the data for easy presentation to users and school administrators.

This objective has been accomplished.

Project Deliverables

The project was planned to operate according to an Agile methodology, so no comprehensive listing of deliverables was possible, since user stories were gathered both at project opening throughout the project lifecycle. However, the following milestones were proposed to provide some structure to the anticipated timeline.

At each milestone, certain user stories were implemented in order to add functionality to the product, but a usable product was presented for stakeholder feedback as early as the first milestone. In addition to serving the iterative method, this ensured that important design decisions about UI are not ignored until the last minute. These milestones required some adjustment in practice, any such deviations are discussed below under "Deviations."

1. Using preconfigured data in the database, expose some basic visualization reports.

This milestone will focus on configuration of the chosen technology stack in the development environment. Since it is anticipated that most of the work of the project will take place in the ETL (extract, transform, load) of data and visualization of data, these will sidestepped by using preconfigured "dummy" data and simple reports. The primary goal of this milestone is to ensure that the chosen technology stack works without major error and that the basics of stakeholders viewing reports is satisfactory.

2. Implement a software utility to perform the ETL procedure on real data.

This milestone will focus on the building of a software tool to read from existing data documents and perform the ETL process on those documents, finally loading them into the database. This is a large task and the bulk of the software development work in the project, and the software utility will almost certainly undergo further iterations of design and functionality. At a minimum, at this milestone the utility must be able to extract all the text of documents for full text search.

Stakeholders should be given the software utility so they can exercise the workflow for adding documents into the database.

3. Configure more advanced visualization reports.

This milestone will focus on answering some of the key questions posed by project animus. For example: what is the ESLR coverage across all lesson plans. Alongside developing more advanced reports, the data schema will need to be refined to represent more detailed information, and the software utility will need to support this refined schema.

At this time, stakeholders should be able to interact with the generated reports in an attempt to determine usability for business and accreditation purposes.

4. Automate the ETL software utility so that new documents can easily be added to the database as they become available.

This milestone will focus on simplifying the user workflow for adding documents to the database. Ideally, user workflow will be reduced to simply specifying a directory to watch for new documents, however this may be of limited feasibility.

To ensure that the workflow is working correctly, the existing database will be deleted and stakeholders allowed to setup the automated ETL process.

5. Deploy the database and reporting portal.

This milestone will focus on studying, recommending, and implementing a deployment strategy for the organization. Possible deployments are to cloud based storage or an on-site server. To be studied are questions about cost, data privacy, and availability.

Stakeholders will be apprised of various options and be allowed to use the deployed product.

Project Plan and Timeline

Milestone Plan

For each milestone iteration, the following steps should be performed:

1. Meet with stakeholders to create/refine user stories.
2. Prioritize user stories, select 1 or more to sketch design ideas.
3. Implement selected user stories.
4. Meet with stakeholders to exercise selected user stories.

Timeline

Project Milestone	Planned Duration	Actual Duration	Actual Start Date	Actual End Date
Basic visualization reports with preconfigured data	2 weeks	1 week	11 April 2017	17 April 2017
Software utility to perform ETL process on lesson plans	2 weeks	5 days	18 April 2017	22 April 2017
Advanced visualization reports and transformed data schema	2 weeks	4 days	23 April 2017	26 April 2017
Automate software utility to perform ETL on new files	2 weeks	5 days	18 April 2017	22 April 2017
Deploy database and reporting portal	2 weeks	5 days	26 April 2017	30 April 2017

Deviations

The original schedule was drafted to fit with the schedules of key school administrators who were intended to regularly meet to discuss the project progress, exercise user stories, and identify additional requirements. Because this type of oversight was largely absent from the project, all of the actual work was able to be accelerated at a faster pace.

The second and fourth milestones, intended to separate the concerns of the work into more easily accomplished parts, proved duplicative. Because of implementation choices, it proved relatively trivial to ensure that the ETL software utility was automated to discover new documents in its path from the beginning. Therefore, these two milestones were effectively done at the same time.

Project Development

After spending time exploring the capabilities of Elasticsearch/Kibana with a fake dataset, work commenced on creating a Beat software utility that could continuously or periodically run on a server, identifying new documents, performing the extraction and transformation on the data in those documents, and then load the prepared data into an Elasticsearch database. The basics of doing this went smoothly, although the process of correctly extracting data proved tricky. Over the course of the project, given the same dataset, the number of extracted documents increased from 451 to 1075.

Similarly the efforts on visualization were simple to get started with and time-consuming to correct and iterate on, each time making sure that the presentation was more pleasing and simple to view. Ultimately, an easily understood Kibana dashboard was produced and embedded into a web page, allowing stakeholders to experiment with the data in several different fashions.

Problems Encountered

By far the biggest problem that was faced was communication. After a successfully pitching the project idea to the school and taking the time to write out a detailed proposal, later efforts to schedule meetings to discuss requirements or do end user testing were continually delayed or ignored. No indication was given that the project was low priority or undesired; instead, communication simply slowed to a trickle with little explanation.

Eventually the decision was made to press forward with the project regardless, leaning on other, non-administrative, contacts at the school. The dataset of lesson plans, previously approved, was acquired and project work commenced. It was determined that perhaps a working prototype was needed to reignite interest: however this was not a great deviation anyway, since the intent was always that successful project completion would enable and further similar work in the future.

Unanticipated Requirements

Based on the sample data given at project onset, it was assumed that although the dataset in question was not strictly structured, that the textual data inside the Word documents was relatively regular. This proven a deep error. While teachers were all working from a single template file, many had their own common quirks of formatting or presentation, making parsing the raw data difficult. For instance, when listing ESLRs, some started the list on the same line as the heading, others on the next line. ESLRs were separated by line, comma, semicolon, or period. Some teachers included numeric codes, others did not.

Unfortunately in order to capture as much data as possible and enable good decision-making, all of these disparate notions of how the textual data was represented needed to be taken into account and parsed appropriately, without prior categorization. This was accomplished by checking for key indicators of data format and occasionally rewriting the data to match the assumed format.

Reasons for Change

From the many individual pieces of data present on the lesson plans, the original scope was only to extract a few key details. However, because the project as a whole was re-understood as a prototype, the decision was made to focus on a single key indicator, the school ESLRs, which most represented the project goal and were the genesis of the project in the first place. The learned information about correctly extracting and visualizing these data points can be applied to further extraction in later expansions of scope.

Actual and Potential Effects

With clear visual data in hand, the school can now easily answer questions about ESLR coverage--that is, how well what they *say* they value matches what their teachers spend time on in the classroom. While developing business intelligence on the data was not in scope for the present project, it was instantly clear that teachers spend a great deal of time teaching certain kinds of stated organizational values but not others. More precise questions can also be asked and answered, such as examining trends by individual teachers, specific school years or grade levels.

These answers could stimulate change in many different ways. It could allow for employee development sessions to be directly targeted at real areas of weakness. It would also be another piece of data informing the complex question of teacher effectiveness. It could even be used by senior administrative officials when evaluating the organizational goals themselves.

Conclusion

While data driven decision making is often emphasized in managerial literature, actually using data properly to this end is perhaps rarer, especially in an educational system that has far more data than can easily be comprehended. This project aimed and successfully executed an effort to help one particular organizational understand the data they already had. It is hoped this effort will also bring at least a little more clarity to the decisionmaking process. While the basic technical implementation of setting up an extract/transform/load process for unstructured or semi-structured data was not especially challenging, the major stumbling block is the messiness of real data as written down by individuals. A balance must be struck between fragility of implementation and accuracy of the dataset in question.

The author has heard many anecdotal stories of organizations that when given understandable, *visual,* data have opened up even more of their forgotten storage devices to this kind of analysis. By the same token, it is hoped that the presentation of even one small but important metric might help spur the school on to desiring better data, better information, and ultimately better wisdom. It has often been the domain of the humanities, but perhaps in this way computer science can help answer a long burning human question: “What is truth?”

Works Cited

Avery, L. D., & VanTassel-Baska, J. (2001). Investigating the impact of gifted education evaluation at state and local levels: problems with traction. Journal For The Education Of The Gifted, 25(2), 153-176.

Bragin, T., & Samant, A. Getting Started with Kibana. Webinar. Retrieved April 29, 2017, from https://www.elastic.co/webinars/getting-started-kibana.

Carpinelli, J., Kimmel, H., Hirsch, L., Burr-Alexander, L., Rockland, R., & O'Shea, M. (2008). A Rubric to Evaluate Standards-Based Lesson Plans and Students' Achievement of the Standards. In American Society for Engineering Education Annual Conference and Exposition (Vol. 12, pp. 13.98.1-13.98.10). American Society for Engineering Education. Retrieved March 5, 2017, from http://www.asee.org/

Gormley, C., & Tong, Z. (2015). Elasticsearch: The Definitive Guide. Retrieved April 27, 2017, from https://www.elastic.co/guide/en/elasticsearch/guide/current/index.html

Miner, G. (2012). Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications. Amsterdam: Academic Press.

Nualart-Vilaplana, J., Pérez-Montoro, M., & Whitelaw, M. (2014). HOW WE DRAW TEXTS: A REVIEW OF APPROACHES TO TEXT VISUALIZATION AND EXPLORATION. El Profesional De La Información , 23(3), 221-235. doi:10.3145/epi.2014.may .02

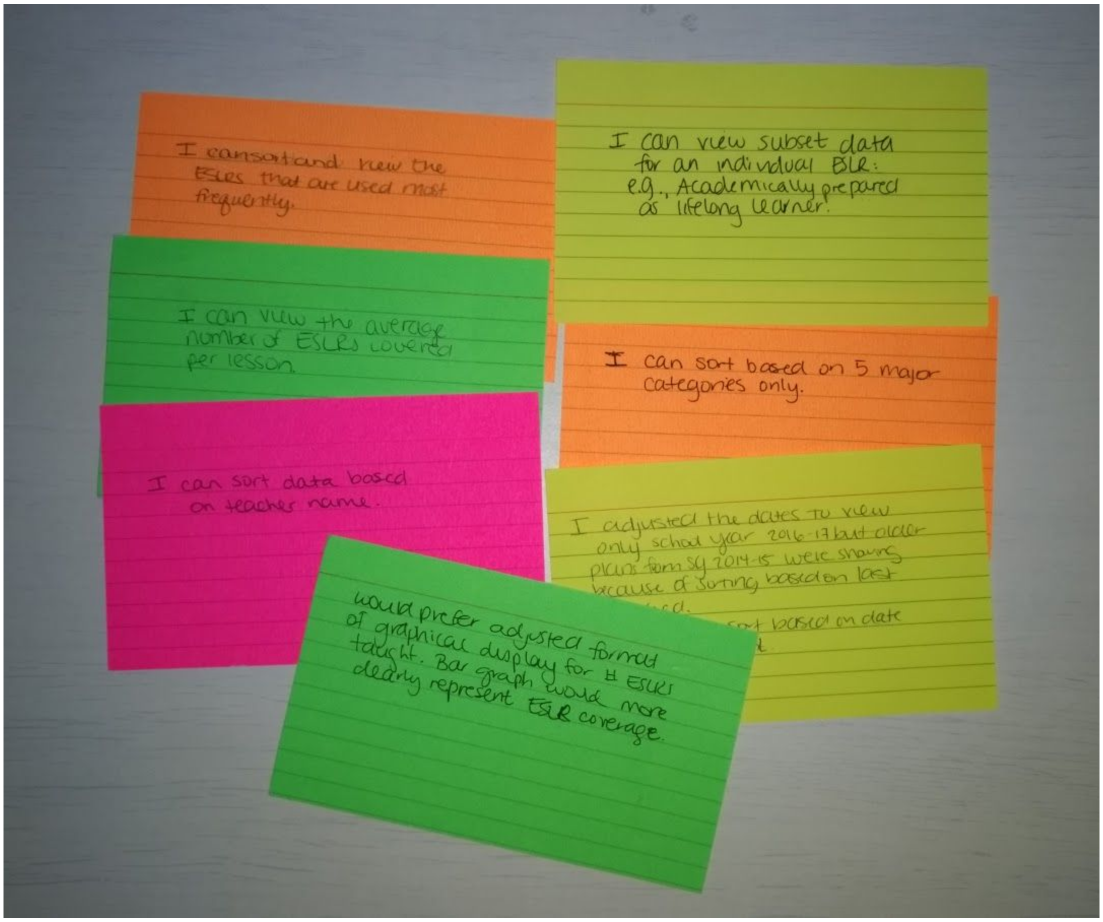
Sarbu, J. K. (2017, March 22). Build your own Beat. Retrieved April 27, 2017, from https://www.elastic.co/blog/build-your-own-beat

Spiro, L. (2008, June 22). Using Text Analysis Tools for Comparison: Mole & Chocolate Cake. Retrieved March 06, 2017, from https://digitalscholarship.wordpress.com/2008/06/22/using-text-analysis-tools-for-comparison-mole-chocolate-cake/

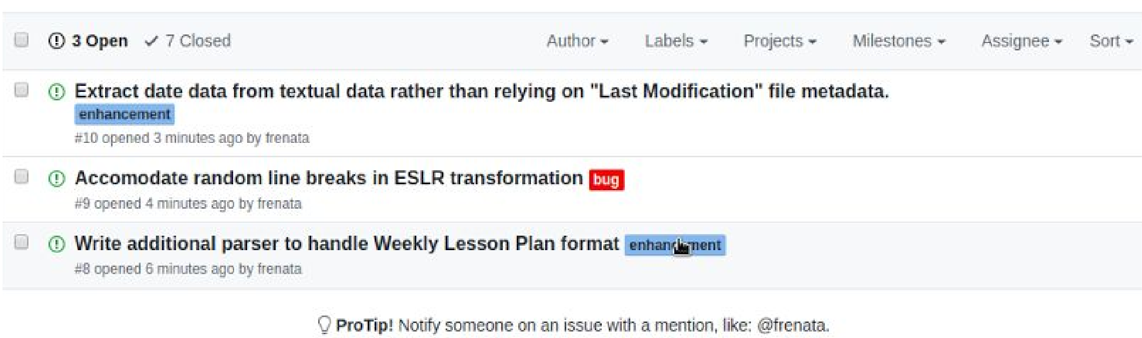
Torff, B., & Sessions, D. (2009). Principals' Perceptions of the Causes of Teacher Ineffectiveness in Different Secondary Subjects. Teacher Education Quarterly, 36(3), 127-148.

Appendix A: Evidence of Process

Early Stage of Project Planning



Sampling of User Stories

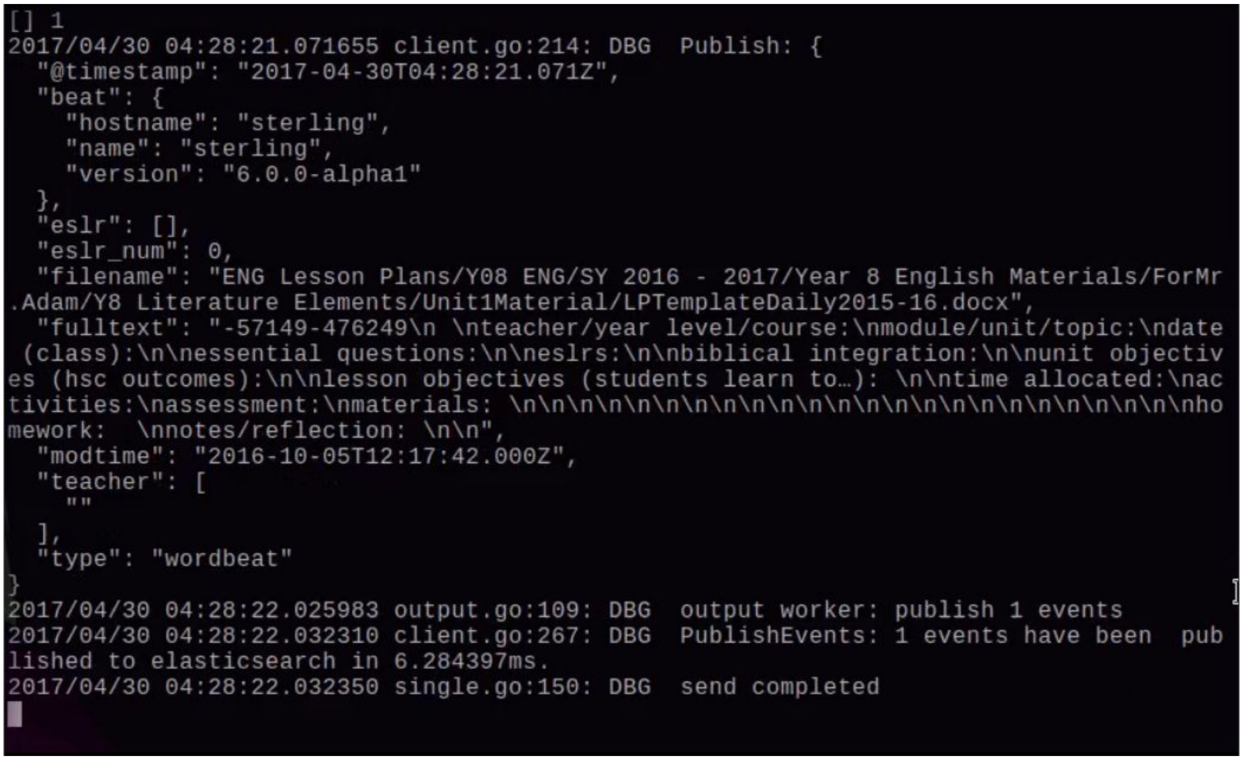


***Issues* page detailing future improvements to be made**

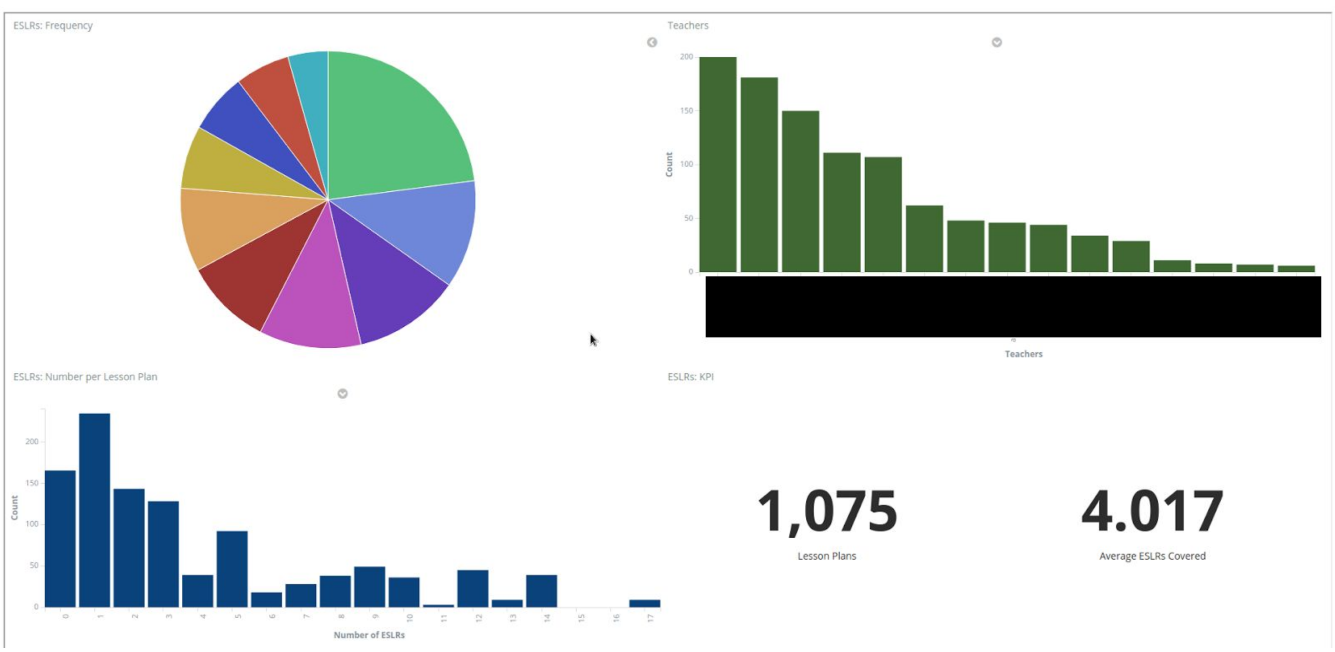
Appendix B: Evidence of Deliverables

```
}

// Recursively walk the directory structure and parse all Lesson Plan files.
func (bt *Wordbeat) listDir(dirFile string) {
    files, _ := ioutil.ReadDir(dirFile)
    for _, f := range files {
        t := f.ModTime()
        path := filepath.Join(dirFile, f.Name())

        if !f.IsDir() &&
            strings.HasSuffix(path, ".docx") &&
            t.After(bt.lastIndexTime) {

            filename := strings.TrimPrefix(path, bt.config.Path)
            fulltext := strings.ToLower(extractText(path))

            plans := strings.Split(fulltext, "daily lesson plan")
            for i := 1; i < len(plans); i++ {
                if i > 1 {
                    filename = filename + strconv.Itoa(i)
                }
                event, err := parseLessonPlan(plans[i], filename, t)
                if err == nil {
                    bt.client.PublishEvent(event)
                }
            }

        }
        if f.IsDir() {
            bt.listDir(path)
        }
    }
}

// Parse a Lesson Plan file, which may have multiple daily lesson plan sections in it.
func parseLessonPlan(fulltext, filename string, modTime time.Time) (common.MapStr, error) {
    lines := strings.Split(fulltext, "\n")

    teachers := extractTeacher(lines)
    eslrs := extractESLR(lines)

    event := common.MapStr{
        "@timestamp": common.Time(time.Now()),
        "type":       "wordbeat",
        "modtime":    common.Time(modTime),
        "filename":   filename,
        "fulltext":   fulltext,
        "eslr":       eslrs,
        "eslr_num":   len(eslrs),
        "teacher":    teachers,
    }
    return event, nil
}

// Check a portion of a file to verify if it is a Daily Lesson Plan
func isDailyPlan(lines []string) bool {
beater/wordbeat.go                                              122,67         39%
"beater/wordbeat.go" 224L, 5333C written
```

Sample of code from the Beats ETL Software Utility



Beats ETL utility in the process of scanning and extracting documents



Kibana Website Dashboard showing Lesson Plan data