Rutgers AICPA Data Analytics Research (RADAR) Initiative
In-person meeting
October 5, 2016

Participants:

Mike Leonardson	Jason Guthrie
Brian Miller	Erica Nelson
Eric Cohen	Katie Greehan
Brian Collins	Brian Wolohan
Evan DeFord	Dorothy McQuilken
Brian Foster	Trevor Stewart
Amy Pawlicki	Miklos Vasarhelyi
Dave Dauksas	Sue Coffey
Greg Shields	Nicole Deschamps
Dan Balla	Al Anderson
Alex Kogan	

Meeting minutes:

- The Board approved the August meeting minutes for posting to the RADAR website
- The group discussed data needs, options for obtaining sample data sets, and the methods that the research teams could use to obtain data. These methods included:
    - Receiving data already anonymized (i.e. not including identifiable information)
    - Receiving through a private server located at the client using a VPN
    - Working from the client site/facility and running the programs from there.
    - Receiving a few records (not all files) and developing programs, based on these records, then bringing them back to the client site and showing them how to run the programs there.
    - Creating simulated data and theories around the simulated data. The Board reps could then take the programs to clients to run on their data. Research teams could go to client sites as well to assist.

  The group discussed additional ways to obtain sample data, suggestions included:

    - Reaching out to former clients, friends to discuss the initiative and data needs
    - Utilizing the data discussed at the AAA meeting by a professor from North Eastern. This data is mainly retail data that is freely available to support work on analytics. Contact information was shared with the researchers.
    - Reaching out and working with internal audit to work with data that they have
    - Reaching out to other universities to see if they have sample data sets, as well as reaching out to alma maters to see if they have data sets that could be used
    - Utilizing governmental data that is available

It was noted that once possible donors are identified, the research teams can meet with them to explain the data needs, projects and ways to anonymize and receive the data. Research teams can also help with anonymizing the data sets, if needed.

It was also suggested that the techniques developed by the researchers could be taken and used by the Board reps on their 2016/2017 audits.

- The research team presented an update on the MADs project:
    - The purpose of this project is to improve the effectiveness and efficiency of the audit sampling process.
    - The research team developed a framework that will guide auditors through the process of prioritizing outliers.
    - The framework is made up of several components:
        - Overall audit objectives
        - Objective and criteria identification
        - Data collection and understanding
        - Data preparation (i.e. formatting, cleaning)
        - Applying the MADs process to identify more suspicious items
            - This process includes the following steps:
                1. ***General techniques*** – applying outlier detection techniques to evaluate the entire population. This will filter down to a population of potential outliers.
                2. ***Data specific techniques*** - applying another set of data analytic techniques to the outliers to identify outlier candidates that are more likely to be problematic
                3. ***Prioritization –*** rank based on a prioritization method
                4. ***Optional procedures –*** additional procedures that could be performed over the population to ensure that the process is accurate.
            - The following suggested edits were noted:
                1. Research teams should not focus too much of their efforts on AI techniques, but more on what techniques and practices are currently being used today on audits.
                2. Step 1 should be the filters that auditors use during their audits today (i.e. data specific filters should be moved up). Step 2 should be the current step 1 "general techniques" and would focus more on statistical techniques. This would then filter down the population to "potential" candidates. Then the prioritization would start. Once there is a listing of prioritized items, it would be up to the auditor's judgement to determine how many of the candidates need to be looked at/tested.
                3. Although the "optional" procedure should be done each time the model is built, it is something that should be tested by the research teams and not an additional procedure that the audit team would need to do.
                4. Research teams will need to measure the effectiveness of the framework/process. One potential way of doing this is comparing the framework to how this is currently being done in

practice. Researchers could also randomly produce errors in the data set to see if they are caught.

- **Next Steps/Items to Consider Going Forward:**
  - **The researchers need to compare this framework to what is currently being done in practice. As the project moves along, Board participants should work with the researchers to give them a process to follow and apply. Board will advise during an upcoming meeting.**
  - **Data sets are still needed for this project**

- The research team provided an update on the process mining project:
  - Process mining is a technique that uses information from event logs to analyze business processes.
  - In order for logs to be analyzed they need to include a purchase order, activity, event type, timestamp, and originator.
  - By applying process mining techniques to logs, auditors are able to understand standard process flows and identify non-standard process flows that may require additional analysis.

  - **Next Steps/Items to consider Going Forward:**
    - **The research team is looking for data sets that are linked to transactional data.**

- The research team provided an update on the visualization project:
  - The purpose of this project is to show how visualization can be used as audit evidence.
  - Visualization can be used during the risk assessment, planning, reviewing and reporting phases of the audit.
  - Some suggestions from the board included:
    - The examples should also layer in the thresholds, to clearly show what items may have fallen above or below the intended boundary. These parameters should be clearly explained. If these visualization are designed within the parameters of traditional substantive analytical procedure it could potentially become audit evidence.
    - It may be helpful to include a description of how an auditor can obtain substantive evidence from the visualization, and what, if any, additional procedures were performed.
    - Researchers should compare these techniques to the current techniques being used on audits

  - **Next steps and items to consider going forward:**
    - **The research teams will be working with the board to obtain additional information and insight into the insurance data. The researchers are looking to speak with auditors who have a background in insurance.**
    - **The research team is requesting additional data sets beyond insurance**