

JUDICIAL COUNCIL OF CALIFORNIA

QUESTIONS AND ANSWERS

RFI# BS-2020-11-DM

Data Mining Software

May 12, 2020

1 **QUESTION:** Section A.3 – Detecting and reporting errors – Please explain if functionality desired is in detecting data quality errors in the data itself OR detecting errors on how the data is mined (i.e. Did this model mine the data correctly and as intended?)

ANSWER: Both. Detecting data quality issues in the data should be essential to provide accurate data for mining. Detecting errors in how the data is mined is also important and it will help in understanding underlying the information accuracy. [REVISED FROM PREVIOUS ANSWER POSTED ON MAY 5, 2020]

2 **QUESTION:** Section A.6 – Forecasting tool – Please explain if functionality desired is for financial forecasting, predictive forecasting or other; Please provide details of either.

ANSWER: The purpose of this question is to find out what forecasting capabilities exist with data mining software. We forecast based on prior year expenditure history; current month expenditures, and additional information provided by the JCC program re planned expenditures.

3. **QUESTION:** Section D.6 – User count will frame RFI response and what solution (desktop or server). Please identify if there will be less than 15 end users or more than 15 end users. If more than 15 end users, please identify if the end user count will be greater than 50.

ANSWER: Currently 17 anticipated users. [REVISED FROM PREVIOUS ANSWER POSTED ON MAY 5, 2020]

4. **QUESTON:** What is the volume of data the tool/framework need to deal with?

ANSWER: Both individual data dump that need to be processed and uploaded to framework's database and historical data that will be used for analysis.

5. **QUESTION:** Currently does your organization have assets in MS Azure or AWS?

ANSWER: Yes, we have assets in both MS Azure and AWS.

6. QUESTION: Does the software need to use data from JCC's Legacy system as well for Analysis purposes? If so, what is the volume of this Legacy Data?

ANSWER: Possibly, but it would be the same file format as FI\$Cal, either PDF, Excel or CSV. The software would not need to interface with the JCC Legacy System.

7. QUESTON: Does this solution need to be built completely with on-site resources or can offsite resources be engaged for building this solution?

ANSWER: Solution can be either on premise or in the cloud, but solution must be secure. Off-site resources located within US territories are able to support the effort. Access by off-shore resources from foreign shores will not have access to branch resources and data.

8. QUESTION: To assist the Judicial Council with its oversight activities, in addition to forecasting, is there an interest in other analytical capabilities such as trend analysis, anomaly detection, etc.?

ANSWER: While not the primary need, additional analytical capabilities would be beneficial.

9. QUESTION: Users:

How many users will be accessing the tool?

ANSWER: Currently 17 anticipated users.

What types of users will be accessing the tool (e.g., analysts, executives, non-technical users)?

ANSWER: Analysts in our budget and accounting offices.

10. QUESTION: Please provide more information on the data including:

a. Number of data sources:

ANSWER: Varies

b. Size

ANSWER: Varies

- c. Complexity (e.g., number of columns, number of rows, etc.)

ANSWER: Varies

- d. Other than SQL-relational databases, are there other databases from which JCC is extracting data?

ANSWER: Other data sources may include data extracts provided by other entities or systems which could include structured and unstructured data files. NoSQL DB may also be a data source in the future.

- 11. QUESTION:** What is the JCC's current operating system?

ANSWER: Our current Standard OS are Windows 10 for desktops, Windows Server and Linux RHEL.

- 12. QUESTION:** Does the JCC have an existing/preferred cloud vendor? Or is the JCC cloud vendor agnostic?

ANSWER: JCC has two preferred cloud vendors (AWS and Azure). Other cloud vendors can be considered, but need to demonstrate value in cost, security and operational support.

- 13. QUESTION:** Does this project have an incumbent vendor and, if so, which vendor?

ANSWER: No

- 14. QUESTION:** If a formal solicitation (i.e. RFP, RFQ) for this project were to be released, what is the estimated timeline for that release?

ANSWER: RFIs are for investigative purposes only. Information gained through an RFI assists in future planning and does not mean that an RFP is forthcoming. At this time the RFI is only being used to gather information.

- 15. QUESTION:** Are there any Data Size requirements you can share for this initiative?

- Initial: What is the initial data size requirement?

ANSWER: No

- On-Going:
 - Do you want to retain historical data? If so, what approximate data growth / month do you anticipate?

ANSWER: Yes. Unknown

- Do you want to retain historical documents?
 - If so, what approximate document growth / month do you anticipate?

ANSWER: Yes, unknown.

- Do you want to scrub the unstructured data and create structured data to be stored in the database?

ANSWER: Yes

- Are there any requirements to extract data from other data sources (e.g. other ERP systems)?

ANSWER: None identified at this time but expect that it will need to extract data from other data sources in the future.

- Are there any requirements to reconcile data from other data sources (e.g. other ERP systems)?

ANSWER: None identified at this time but expect that it will need to extract data from other data sources in the future.

- Does data need to be extracted out of this solution and sent back to Fi\$cal or other systems?

ANSWER: No

16. QUESTION: Reporting / Analytics

- What total number of reports do you expect?

ANSWER: Since the purpose of obtaining the software is to create custom reports, the total number of reports is unknown and dependent on the capabilities of the system that is ultimately selected.

- How many users will run basic formatted reports?

ANSWER: Currently 17 users anticipated.

- How many users will run self-service reports?

ANSWER: Currently 17 users anticipated.

- How many users will build forecasting models?

ANSWER: Since this is not the primary focus of the solicitation, this is unknown.

- How many users will run reports based on forecasting models?

ANSWER: If forecasting models are used, it would likely be 6-7 users.

- Do you have existing reporting tools that you plan to allocate to this project?

ANSWER: No

- Are there any requirements to publish data to the public?

ANSWER: No

- Sample Reports listed in RFI attachments appear to be pre-formatted static reports – Are there any requirements for self-service / discovery type of analytics?

ANSWER: The purpose of acquiring data mining software is to extract data from the pre-formatted static reports to create custom reports.

17. QUESTION: Is there a preference for cloud vs. on-prem?

ANSWER: No preference, but depends on the product, services, security and operational support requirements.

18. QUESTION: Are you open to solutions that offer hot/warm/cold data storage?

ANSWER: Yes. Warm and cold data storage may be used to retain less frequently used data and/or archiving.

19. QUESTION: What is the expected timeline for the solution to be fully implemented?

ANSWER: The goal is contract with a vendor by June 30, 2020.

20. QUESTION: What are the most commonly utilized forecasting algorithms utilized by the Judicial Council?

ANSWER: None. Forecasting is not the primary focus; data extraction to create custom reports is the focus, forecasting is secondary.

21. QUESTION: In addition to Fi\$Cal, are there other data systems or sources that the Judicial Council would like to utilize for reporting purposes?

ANSWER: None at this time. The software does not need to interface with any system, current or legacy.

22. QUESTION: Have you received any demonstrations from vendors regarding this opportunity? If so, which vendors?

ANSWER: Yes, Monarch.

23. QUESTION: Are you open to 2 or more software tools that we could propose to meet your requirements?

ANSWER: It would depend on the complexity of using multiple tools and the cost.

24. QUESTION: Can you offer more detail around the volume of data we should anticipate?

ANSWER: Difficult to answer. FI\$Cal uses reporting structures. Each reporting structure can have one or more associated projects and activities; and one or more fund sources depending on the size and complexity of the office. When FI\$Cal was implemented, there were approximately 866 lines that were some combination of reporting structure, projects, activities and funds. In addition, there are 200+ account codes. All of which represent data in FI\$Cal. For example, one of our largest JCC offices utilizes 19 reporting structures, 15 projects, 41 activities and 4 fund sources in various combinations. The FI\$Cal report included in Attachment B is one of the smaller Judicial Council offices.

25. QUESTION: Can you describe your data users and number of users with regards to licenses?

ANSWER: Refer to questions #6 and 9.

26. QUESTION: What are common data errors that you foresee needing to detect?

ANSWER: Unknown

27. QUESTION: Are there specific graphical features/capabilities you interested in seeing/utilizing?

ANSWER: None

28. QUESTION: Can you give an example of mapping from source to target exports?
(Expand on target exports)

ANSWER: No specific source and target examples.

Data from Fi\$Cal data files can come in multiple file formats. (CSV, Excel, PDF, XML). The data from these files will need to be organized and organized into usable data structures that will enable mining, correlating and reporting. Target results may be an Excel spreadsheet or database.

29. QUESTION: Can you expand on what the granularity in controls you are looking for? (RFP Section B.3 on page 4)

ANSWER:

Describe what security level of security can be supported by your solution. Does it apply a coarse grain or fine grain control? Or describe the security controls within system to support access to functions and to data.