RFI 02 /2025: Request for Information for Master Data Management and Data Governance Tool

Communication #2

Date of Issue: 26 June 2025

Questions and Answers:

No	Question	Answer
1.	Please confirm as to whether the South African Revenue Service is seeking: •A Master Data Management (MDM) solution, encompassing capabilities such as master record creation, data consolidation, entity resolution, and stewardship workflows, or	SARS is requiring A Metadata Management and Data Governance tool, with an emphasis on capabilities such as metadata discovery, data lineage, data quality profiling, policy enforcement, and cataloguing.

No	Question	Answer
	•A Metadata Management and Data Governance tool, with an emphasis on capabilities such as metadata discovery, data lineage, data quality profiling, policy enforcement, and cataloguing.	
2.	Kindly advise if a non-submission of the RFI will lead to bidders not being invited to the RFP or if the RFP (once advertised) will be open to all service providers?	RFP will be open to all service providers. However, it is encouraged to submit the RFI
3.	We kindly request clarification on the following points: Could SARS provide insights into the current and projected data volume growth over the next 3 – 5 years?	SARS need to get internal guidance on how much is allowed to share about the data sizes and additional information.
4.	What are the most critical domains and sources that require governance (e.g., Customer data)?	The most critical domains and sources that require governance at SARS are related to taxpayer information. This includes all processes that govern taxpayer data.
5.	Are there existing data governance policies or standards in place that the solution must align with?	SARS has a data governance framework and is implementing it. Data classification is catered for in other tools.

No	Question	Answer
6.	What are the most pressing data quality issues (accuracy, completeness, timeliness) currently faced?	The most pressing data quality issues include accuracy, validity, completeness, and timeliness.
7.	Does SARS require automated data quality checks, profiling, cleansing, and enrichment for all data sources or only selected ones?	SARS requires automated data quality checks, profiling, cleansing, and enrichment for all data sources.
8.	Is there an existing metadata repository, data dictionaries or glossary that the solution needs to be integrated with?	Currently, metadata is captured manually, and SARS is looking for an automated solution with a comprehensive enterprise-wide metadata repository.
9.	Is there an established data governance framework including DG committees and data stewardship?	SARS has a data governance framework in place and is implementing it. There are data governance policies and standards that the solution must align with.

No	Question	Answer
10.	What resources and skills will SARS provide versus what is expected from the vendor?	SARS is looking for a partner to support the full implementation initially. The discussion on adopting the tool to become theirs will come later.
11.	Are there specific change management or training required as part of the implementation?	Specific change management or training requirements were not discussed in detail during the meeting.
12.	What are the primary regulatory requirements (e.g., POPIA) that the solution must support?	The solution must support regulatory requirements such as POPIA.
13.	Are there specific requirements on data protection e.g., masking, end-to-end encryption etc.?	Data masking and redaction are part of the requirements.
14.	What is the expected timeline for implementation?	The timeline for the RFP issuance is not explicitly defined, but it is desired to be before the end of the financial year, depending on the information received from the RFI and subsequent funding approval.
15.	Can you provide a list of data source applications?	The main data sources include Natural Adabas, DB2, SQL Server, and the enterprise data warehouse running on Microsoft SQL Server.

No	Question	Answer
16.	Please inform how does mainframe expose its data? DB2, flat files (VSAM), FTP, or APIs?	The mainframe data is exposed through a mixture of methods, including flat files, VSAM, FTP, and APIs
17.	How is SAP data exposed or extracted today for reporting?	Please consider SAP as a whole due to limitations on sharing detailed information.
18.	What products of SAP are being used today? ECC or S/4 Hana?	The specific SAP products in use are not disclosed due to sharing limitation therefore considering SAP as a whole is ideal when providing the solution capabilities.
19.	What technologies are the other data sources / applications running on?	The technologies include Natural Adabas, DB2, SQL Server, and Microsoft SQL Server for the enterprise data warehouse
20.	Is this RFI only about master data (Entities like Customers, Products, Policies, Jurisdictions) or transactional data as well?	The RFI covers both master data (e.g., Customers, Products, Policies, Jurisdictions) and transactional data
21.	Do you receive external data? How many external data sources are there?	Yes, SARS receives external data.
22.	In what file format do you receive external data?	External data will not directly impact the solution as SARS has processes in place to manage external data.
23.	What is the frequency of receiving external data? daily, weekly, monthly?	External data will not directly impact the solution as SARS has processes in place to manage external data.

No	Question	Answer
24.	What is the size of this data?	External data will not directly impact the solution as SARS has processes in place to manage external data.
25.	Are there reporting use cases where you need data in real time from data sources? Can you name some?	The need for real-time notifications and alerts are indeed a requirement for timely data updates.
26.	What is size of historical data collected so far from different applications?	The size of historical data collected from different applications was not disclosed.
27.	Can you provide some examples of current data quality issues and challenges?	Current data quality issues include the labor-intensive nature of in-house developed tools and the need for automated solutions to provide accurate, valid, complete, and timely data
28.	What are the current challenges faced by business users?	Current data quality issues include the labor-intensive nature of in-house developed tools and the need for automated solutions to provide accurate, valid, complete, and timely data
29.	What are the key use cases that you plan to address using this solution?	Key use cases include enhancing data governance, improving data quality, automating metadata management, and ensuring compliance with data privacy and security policies
30.	Which business functions or departments are the primary beneficiaries or stakeholders of this initiative?	The primary beneficiaries or stakeholders include the data governance and master data management teams, as well as business units relying on accurate and timely data for decision-making

No	Question	Answer
31.	Are there specific business processes that must remain uninterrupted during tool integration and rollout?	Critical data governance and quality management processes should continue smoothly.
32.	Is there a phased rollout envisioned – by business unit or data domain?	A comprehensive and integrated solution suggests that a structured implementation plan may be necessary.
33.	Are there existing documentation standards or formats we need to align with during delivery (e.g., SOPs, data policy docs)?	Specific documentation standards or formats (e.g., SOPs, data policy docs) were not mentioned in the meeting content.
34.	Is SARS seeking a full implementation partner (including change management, training, and support), or only a product solution with limited advisory support?	SARS is seeking a full implementation partner to support the entire implementation process, including change management, training, and support.
35.	Can the RFI closing date be extended?	The RFI closing date is set for the 8th of July 2025 and not be extended.
36.	How many endpoints does SARS have?	The specific operating systems used for endpoints were not mentioned.
37.	What OS'es are being used for endpoints?	The specific operating systems used for endpoints were not mentioned.
38.	How many data stores does SARS have?	The number of data stores SARS cannot be disclosed
39.	What data stores are currently being used (databases, file systems, object storages, directory services, email storage systems)?	The data stores currently in use include Natural Adabas, DB2, SQL Server, and Microsoft SQL Server for the enterprise data warehouse.

No	Question	Answer
40.	How many locations does SARS have?	Will mainly be at Head Office
41.	Is SARS data storage hosted on-premises, private data center, or cloud?	SARS data storage is hosted on-premises and in the cloud, indicating a hybrid model
42.	What is the current technology stack for data storage?	The technology stack includes Natural Adabas, DB2, SQL Server, and Microsoft SQL Server
43.	What are the desired storage locations (on-premises, cloud, private data center or hybrid)?	The desired storage locations include a hybrid model with both on-premises and cloud storage