Frequently Asked Questions (FAQ) - Durban Data Exchange (SHANA)

1.   What are we building and why?

eTheKwini is building a municipal data exchange named SHANA to support water services with improved ability to collect, store, standardize, analyze and access data. SHANA is short for "ukushintshana", which means "exchange" in Zulu - the primary language in eThekwini. The data exchange is a cloud based data lake that provides a common location to ingest, store, analyze and use both public and private data sets. On top of the data lake are tools for access and identity management, the ability to subscribe to data sets, a customizable dashboard that allows users to get access to analytics needed in support of their mission.

SHANA is being built to create a process for staff to access the data and analytics needed to manage water operations in the municipality. Today water data lives in different agencies, in different versions and different levels of accuracy. Staff does not have a full understanding of the existing data inventory, how to access the data sets or confidence in the accuracy, completeness and timeliness of the data. As a result, there are multiple copies of data sets and different numbers used for benchmarking, modeling and reporting - which impacts trust in data being used by decision makers.

SHANA is also a test case for a municipal data exchange built on the water use case, but scalable to other use cases.

2.   Who are we building it for?

eThekwini is building SHANA for internal and external customers involved in all aspects of water supply, demand, operations and policy. The primary customers are Municipality staff, including the GIS, reservoir, finance, planning, and maintenance staff.

3.   What is the value to the customer?

The primary value is to have one reliable source of data for water services that can be depended on for accurate and up to date information for benchmarking, modeling, and reporting.

4.   How does SHANA work?

SHANA is a cloud based data exchange controlled and operated by the municipality. Users authorized by the municipality can upload and access data through the SHANA portal.

SHANA combines several capabilities into one tool. A robust data lake to store data, access and management tools to authorize users, data standardization tools to improve data integrity and usability, customizable dashboards to assemble desired reporting and analytics and an exchange that allows for the access and sharing of data.

The SHANA data exchange works by creating a central data lake for the municipality where data owners can upload data sets that can then be used by other municipal staff and authorized stakeholders.

Access and authorization is managed through an Identity and Access Management portal where users can both make and respond to access requests. Using the home portal, users can see the full data inventory in SHANA with all of the related meta data.

SHANA offers base and customizable dashboards for users to access data, build reports and analytics to support specific needs.

5.   Who will be the publisher(s) of the data?

One of the benefits of SHANA is it allows authorized users to both publish and access data on the exchange. SHANA is the common platform and everyone gets to use it who has been given permission to the system.

6.   Who will be subscribing to/consuming the data?

SHANA allows users to subscribe to datasets they want to access. Data owners receive requests and authorize users to access the system. The prototype use case is based on improving water operations. Initial users will be Municipal Water Department staff. As additional use cases are deployed, it is expected that users internal and external to the Municipality will want to subscribe and consume data.

7.   How much data will be added daily/weekly/yearly?

The amount of data that will be added is unclear. As SHANA is developed and becomes operational the projections for the amount of data use will become clearer. The Municipality may add sensors like water meters which will increase the amount of data that could be added. The platform is built on AWS so it can scale to meet a "virtually" unlimited data set size.

8.   Will this data be real time data? If so, what type of files, size, format, update cadence?

SHANA supports all types of data, including real time data. For the prototype, one of the data sets will be real time and the others will be flat files. This capability to accept all types of files is essential to the ability to serve as a centralized data exchange.

Data sources are expected to include GIS, sensor, revenue, customer, modeling, and other data sets from across municipal agencies. Today, the water staff has most of the data they need to do their jobs, the data is just hard to access and lacks data integrity.

Data sets will remain in control of data owners in SHANA and the owners will control who has access to the data.

9.   What types of data (images, excel, databases, etc) can be used with SHANA?

Any type of digital data can be used with SHANA after potentially going through an extract transform load (ETL) process. Data is expected to include GIS files, spreadsheets and sensor data.

10.   Does the Municipality want any portion of this data to be available open and free to everyone?

Owners of various data sets will have the ability to determine who the data is available to and if it comes with a cost. Data owners can also choose a subset of the entire data set. Data owners for the prototype are internal staff from the Municipal Water Department, including GIS Department, the Water Reservoir managers, the Finance Department and Maintenance Department.

11. Does the availability of this data need to be hidden from the public catalog for security or other reasons?

Data owners have the ability to determine who can access data sets and if the data is included in the public datalog of available data sets. Sensitive data could include the locations of secure equipment that the municipality wants to protect.

12. Where will SHANA live? Who is responsible for managing?

The SHANA data exchange will be managed by the eThekwini Municipality IMU - Information Management Unit. SHANA will be a cloud hosted solution on the AWS cloud. All data will remain within the AWS South African Cloud Region.

13. Who can access SHANA?

Access to SHANA will be controlled by the Municipality. Each data set on the system will have a level of access from public data to secured data. Municipal users are expected to include: -

- political/administrative leadership
- EWS Unit Head
- EWS Technical
- Research and Policy
- Community (viewing accessing dashboards)

15. Will the data be secured and how?

The data SHALL be secured by using the well-architected framework for securing workloads on AWS. The storage of data will also comply with city data storage policy and national legislation. Encryption in transit and at rest will be used

16. How much will it cost to build and maintain?

This is still TBD. After the Solutions Workshop the TCO - Total Cost of Ownership will become clearer. The ASU CIC and the Municipality are building a prototype and modeling out the cost structure of the scaled out architecture of the prototype. The total cost will be highly dependent on the amount of data stored within the system.

17. Can data owners retain control and access to data loaded into SHANA?

Yes. When data owners upload data sets to SHANA they will retain ownership and be able to control who has access to the data and what they can do with the data. Users will use Identity and Access Management to control who access data sets and what levels of access they have.

18.   How will data be uploaded?

The initial users (municipal water staff) will create accounts on SHANA and login. Once in SHANA users will be able to create files and upload data. A desired function is to allow for the automatic uploading of data to ensure it remains accurate and users are all working off the same information. Automatic uploading of data will reduce the overhead required of the municipality water staff.

19.   Will analytics tools be available and if so which ones?

The Municipality is identifying the analytics tools currently in use and the tools or functions desired. As part of the prototyping project the CIC is working to identify interoperability and facilitate when possible.

20.   Does SHANA have a dashboard, what's on it, who controls and is it customizable?

Yes. The dashboard is a critical function and is customizable. The initial version of SHANA comes with four stock dashboard versions for different user roles: executive, supply, use, and revenue.

These dashboards, include, but are not limited to analytics on;

- non-revenue water
- response time related to leaks and bursts
- consumptions patterns
- revenue collection patterns
- expenditure patterns
- water demand

The eThekwini Water Service technical office will control the dashboard maintenance and will have ongoing responsibility for ensuring operations, continuity, and refinement of analytics.

21.   Will SHANA allow self-service report creation and customizable dashboards?

Yes. SHANA will have the ability to create customized reports. The required skills to build reporting will be varied, between the SHANA application skills to build the reports and the subject matter knowledge for the topics reporting on.

22.   How will SHANA standardize data?

Data standardization is one of the most important customer benefits. Currently data is not standardized and this impacts data integrity and the ability to aggregate data. SHANA will include a multi-faceted plan for data standardization, quality, and integrity. A first step will be the development of a plan that provides guidance on data standards. The plan will be used to get everyone on the same page. SHANA will use data standardization and quality testing tools like Amazon Deepqu. This combination of a comprehensive plan and tools will be the primary ways SHANA standardizes data.

23.   How do you upload a new data set?

The data will be uploaded manually to an S3 bucket or a-live connection to the data will be made to automate the process. This can be done within the AWS console. The data would then be visualized with AWS quicksight.

24.   Can reports be authorized and signed off efficiently and securely on the system?

Yes. The governance model for this in development. The use of pre-approved report templates is one part of the process to streamline reporting.

25.   Will this eventually cover the rest of the municipality?

SHANA is scalable and can be expanded to cover additional use cases as desired.

26.   Is it possible to eventually include citizen data like reports on leaks etc?

The prototype will be focused on internal data, but the Municipality expects a final version to include public access to data and reporting.

27.   Will we be able to receive information like meter readings from households for more accurate regular billing?

SHANA has the potential to support meter reading and billing if smart meters are deployed and connected to SHANA. This is not an initial feature, but one for later development.

28.   Would it eventually be possible to send out bills or billing information on this system?

Yes. SHANA has the potential to serve as the supporting database for client billing. An additional application will be required. Further research will be conducted to fully understand customer needs and explore viable solutions. The customer has also identified a desired function to allow water customers a download option for their bills would be useful.

29.   What are some of the advantages of using IoT reading of electricity and water meters that is then automatically collated from the householder to the SHANA system

IoT devices have the potential to generate significant amounts of data that can be used for multiple purposes. Ingesting the data into SHANA allows the data to be used for multiple purposes from one location. Storing IoT data in the cloud has an added advantage of allowing SHANA to scale in or out to accommodate the storage for the amount of data collected as needed on a pay as you go model.

31.   What laws, regulations, and policies does SHANA need to be compliant with?

SHANA is customizable and able to be deployed fully compliant with all applicable laws, regulations, and policies and be adjusted as needed as requirements change.

Known requirements include the City Data Storage policy, the Protection of Personal Information Act, and the Promotion of access to Information Act.

SHANA's governance model will require an annual data compliance audit to ensure compliance with all applicable requirements and laws.

32.   What applications does SHANA need to be interoperable with and do those applications have existing APIs?

The Municipality currently uses multiple applications in support of water operations that SHANA will need to be interoperable with, including;

Faultman (CRM)

Arc GIS

Laboratory Information Management System

J D Edwards

Performance Management

Revenue Management System

Field Data

Key Performance Indicator System

Excel Spreadsheets on local Drives

Municipal Benchmarking Initiative

SCADA

Telemetry

IoT

33.   Can SHANA be programmed with alerts and if so who develops them?

Yes. Alerts are one of the most important functions identified by the customer and SHANA will allow users to set their own alerts. The initial configuration of SHANA will come with a basic set of alerts identified and developed during design.

34.   What benchmarking and modeling tools will SHANA have?

This is yet to be determined and will not be static. Benchmarking is another essential function identified by the customer. Like alerts, SHANA will allow users to customize their own benchmarking metrics.

35.   How does the Municipality ensure that no one misuses the data?

There are a number of tools and processes for ensuring that data is not misused.

A primary way is Access RIghts. The AWS Identity and Access Management tool allows eThekwini to control who has access both to SHANA and individual data sets. By controlling who has access, eThekwini will limit the people who have access and could misuse the data.

Data Sharing Agreements are a way to set parameters on how data can be used and what will constitute misuse so there are clear definitions and understandings.

Other tools like CloudTrail can be used to identify activity and can alert on possible misuse and provide a clear understanding of who did what.

36.   How will the Municipality ensure that the quality of the models and benchmarking used is at acceptable standards?

Consult and Benchmark with Water Research Commission, South African Local Government Association, National Treasury and eThekwini Performance Management

###