



APPENDIX A: MARKET DATA ANALYTICS SPECIFICATIONS

DATA WAREHOUSE / DATA MANAGEMENT

1	DATA WAREHOUSE / DATA MANAGEMENT	Comply	Do not Comply
1.1	General Requirements		
1.1.1	The DWH platform can be deployed on-premises at the SIU, however supports cloud and/or a hybrid deployment.		
1.1.2	All data storage and processing must be onsite, however the platform supports offsite/cloud backup options.		
1.1.3	The platform must include all technology to enable data ingestion.		
1.1.4	The platform architecture supports the concept of enterprise data warehousing and independent data marts.		
1.1.5	The platform leverages a normalized logical data model for creation of the enterprise data warehousing subject areas and support dimensional modelling for dependent data marts used for analysis and reporting.		
1.1.6	The platform must support an agile data warehousing approach and iterative development and the rapid creation of subject-specific data marts.		
1.1.7	The platform must support graphical data modelling tools, logical and physical data modelling and source-to-target mapping.		

1.1.8	The platform must support multiple types of slowly changing dimensions, enable conversion from one type to another, and generate and manage dimension surrogate keys as necessary.		
1.1.9	The platform must support the agile implementation of the following types of changes.		
	○ <i>Business relationships (changes to the relationships between existing data)</i>		
	○ <i>Business rules (changes to existing model structures)</i>		
	○ <i>New information (addition of new model elements and rules)</i>		
	○ <i>Retired information (the "expiration" of model elements and rules)</i>		
1.1.10	The platform support integration with business intelligence and data analytics tools at the semantic level. Changes to the data warehousing structures and semantics must be seamlessly provided to the BI tools.		
1.1.11	The platform generates metadata layers for reporting tools, such as SAP Business Objects Universes, IBM Cognos Framework Packages and Microsoft Analysis Services cube definitions.		
1.1.12	The platform support Information Lifecycle Management best practices.		
1.1.13	The platform supports a graphical model-driven approach to master data management.		
1.1.14	The platform must use Microsoft SQL Server as its relational database management system.		
1.1.15	The vendor should provide installation, configuration, and performance tuning of the SQL Server database as required for successful implementation of the platform.		

1.1.16	The vendor should use staff with current SQL Server certification and sufficient experience with installation and performance tuning to ensure successful implementation using SQL Server operational, administration and maintenance best practices.		
1.1.17	The vendor should be able to provide complete operational support services that, in addition to database administration, include platform configuration and administration, data warehousing and data management.		
1.2	Specific Requirements		
1.2.1	Platform Capability		
	<ul style="list-style-type: none"> The solution support 64-bit, and virtualized deployment options. 		
	<ul style="list-style-type: none"> The platform should capture changes to data entities over time, both in data and structure, to enable seamless reporting across periods regardless of changes in format. 		
	<ul style="list-style-type: none"> The platform must support model-driven development to be able to use a logical data model to create a physical data warehousing environment. It must support source-to-target mapping of data from multiple sources into this environment. 		
	<ul style="list-style-type: none"> The platform should support structure and model version history in addition to data history. 		
	<ul style="list-style-type: none"> The platform should support reporting at any point in time under the current data structure or under any previous data structure. This will also enable the data warehouse to be restored to any previous structural version. 		
	<ul style="list-style-type: none"> The platform should integrate reference data from disparate sources. 		

	<ul style="list-style-type: none"> The platform should support reference data management to resolve multiple reference code representations for the same description. 		
	<ul style="list-style-type: none"> The platform should support ragged hierarchies in reference data as well as cross-dimensional relationships. 		
	<ul style="list-style-type: none"> The platform should support building and managing summary datasets graphically. 		
	<ul style="list-style-type: none"> The platform should support additive, semi-additive and non-additive measures, and must be able to differentiate measures that can be aggregated over time from those that cannot or those that can be aggregated against certain dimensions but not others. 		
	<ul style="list-style-type: none"> The platform should support incremental loading of dimension and fact tables, including management of fact-dimension relationships. 		
	<ul style="list-style-type: none"> The platform should support data that is available at multiple levels of granularity and be able to intelligently aggregate data at the appropriate grain. 		
	<ul style="list-style-type: none"> The platform should support modification to data warehouse structures without requiring that data be unloaded and reloaded. 		
	<ul style="list-style-type: none"> The platform should have a graphical user interface that can provide information and access to business data stewards and technologists in support of master data management. 		
1.2.2	Platform Administration, Security and Architecture		
	<ul style="list-style-type: none"> The platform must have a robust, role-based security model that leverages user authentication technologies such as Active Directory. 		

	<ul style="list-style-type: none"> Enabling the authentication of users in real time via LDAP and/or Active Directory as an individual user, roles based, and as a member of a group. 		
	<ul style="list-style-type: none"> Allowing multiple concurrent users to log on to the same application with the same or different security privileges. (Granularity of privileges down to the row level or cell level is a key requirement.) 		
	<ul style="list-style-type: none"> Allowing a business author, of a report or analytic view, assign security permissions to the content they created, and have those permissions perpetually associated to the content as it is promoted to a server. 		
	<ul style="list-style-type: none"> Creating and maintaining an audit trail regarding user transactions, failed log-on attempts, and provide for protection of the logs against modification and unauthorized use – or any other business over-ride or event. 		
	<ul style="list-style-type: none"> Enabling Administrators to track user access/activity to/on various data objects (dimensions and measures), by date, including whether the information was exported outside the DA platform. 		
	<ul style="list-style-type: none"> Allowing for encrypted transmissions across a distributed client server architecture (supports HTTP-S, SFTP). 		
	<ul style="list-style-type: none"> Backup and recovery and reports/dashboards and flexibility in controls available to manage the number of historical instances to be maintained. 		
	<ul style="list-style-type: none"> High availability, failover, database cluster-awareness, and event capturing and viewing. 		
	<ul style="list-style-type: none"> Enabling monitoring of performance, throttle and manage workload and/or prioritize user requests, workload balancing options to divide calculations across all available processors, cores, clusters, etc.? 		

	<ul style="list-style-type: none"> Enabling the SIU to conduct capacity assessments to determine future needs. 		
1.2.3	Metadata Management		
	<ul style="list-style-type: none"> The platform should support data lineage to determine which source platforms contributed to the report regardless of which extraction, transformation, and loading (ETL) tool is used. 		
	<ul style="list-style-type: none"> The platform should highlight changes in the source data (down to which fields are impacted within a report) that affect downstream reports. 		
	<ul style="list-style-type: none"> The platform should enable a report consumer and a report author to view common definitions for metadata objects. 		
	<ul style="list-style-type: none"> The platform should provide for a single repository for ALL metadata, such as mappings of business concepts to underlying data structures (e.g., dimensions, measures), as well as layouts AND report configurations (e.g., prompts, filters). 		
1.2.4	Data Source Connectivity and Ingestion		
	<ul style="list-style-type: none"> The platform should allow for reports, dashboards, ad hoc queries, and analysis to be created "out of the box" from at least one multidimensional online analytical processing (MOLAP) data source, such as Hyperion Essbase, SQL Server Analysis Services - traditional OLAP and/or tabular models, SAP BW or Oracle OLAP. 		
	<ul style="list-style-type: none"> The platform should provide connectivity capabilities to other data sources, such as web-based platforms, XML, RSS feeds, JSON, flat file, spreadsheets, etc. 		
	<ul style="list-style-type: none"> The platform should provide for the analyses of structured and unstructured data sources (text, video, voice, log files) and list each one and how. 		

	<ul style="list-style-type: none"> The platform should provide for native connectivity to enterprise applications either in the cloud or on-premises (Microsoft Dynamics NAV, OpenText Vibe (Maria DB/SQL), Salesforce, NetSuite etc.). 		
	<ul style="list-style-type: none"> The platform must enable the continuous, scheduled and ad-hoc ingestion of data from multiple sources into the data warehouse in Batch, Near Real time and Real time. 		
1.2.5	ETL and Data Storage		
	<ul style="list-style-type: none"> The platform should enable Data Engineers to perform Extract Transform Load (ETL) or extract Load Transform (ELT) / data integration capabilities. 		
	<ul style="list-style-type: none"> The platform should enable storing and managing data in the Enterprise Data Warehouse (EDW). 		
	<ul style="list-style-type: none"> The platform should provide for incremental data loads. 		
	<ul style="list-style-type: none"> The platform should provide for the performing of parallel/multithread loading of data, breaking the source file(s) into multiple streams. 		
	<ul style="list-style-type: none"> The platform should enable the performing of refresh scheduling by source type that are available within the platform. 		
	<ul style="list-style-type: none"> The platform should enable the leveraging of existing analytic storage (e.g., X-IO Storage, data warehouse, third-party in-memory engines) 		
	<ul style="list-style-type: none"> The platform should allow for the managing and monitoring active data loads. 		

DATA ANALYTICS

The Data Analytics (DA) System must provide functionality for:

2	DATA ANALYTICS	Comply	Do not Comply
2.1	General Requirements		
2.1.1	Configurable workflow management by an administrator.		
2.1.2	Collaboration Workspaces (including having multiple users working on the same project or workflow task at the same time, having due consideration to user roles and permissions).		
2.1.3	Include full audit trail capability from data collection, processing, analysis, and output.		
2.1.4	Provide for the consistent application of security controls throughout the data collection, processing, analysis and output functions (e.g. search results, reports and workflow information must only be visible to a user that has the required clearance).		
2.2	Specific Requirements		
2.2.1	Data Collection		
	<ul style="list-style-type: none"> Data sources will comprise both quantitative and qualitative data in structured and unstructured format: 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> <i>Primary sources</i> 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> <i>SIU Forensic Investigations including performance information.</i> 		

	▪ <i>Programme and Time Management information</i>		
	▪ <i>SIU Civil Litigation</i>		
	▪ <i>Internal Business performance</i>		
	▪ <i>Market surveys</i>		
	▪ <i>Impact assessments</i>		
	▪ <i>Financial Information</i>		
	○ <i>Secondary sources such as:</i>		
	▪ <i>Open Sources: Public and social media</i>		
	▪ <i>Grey Sources: Available by way of agreements or subscriptions from other public and private entities</i>		
	▪ <i>External Relevant Sources</i>		
2.2.2	Data Processing		
	The platform must enable the following activities:		
	• Data review and inspection		
	• Cleaning of data		
	• Enrichment of data		
	• Transformation of data that will allow the modelling of data in order to identify useful information.		
	• Classification of data in preparation of data analytics processes.		

2.2.3	Data Analysis and Outputs		
	<ul style="list-style-type: none"> The platform must enable access by multiple concurrent users via a web-based GUI, based on a hierarchy of security classifications (such as user roles, users, workflow, operational clusters, etc...). 		
2.2.4	Intelligence Analytics		
	<ul style="list-style-type: none"> The platform must provide a variety of intelligence analytics tools such as Statistical analysis, Predictive analysis, Forecasting, Social Network Analysis (visual display of relationships between entities); Pattern-matching Analytics (a capability for performing advanced queries against entities and relationships); and Entity Identity Resolution (automated identification and merging of incoming entity records 		
	<ul style="list-style-type: none"> The analytical tools must allow multiple and dynamic views of information (e.g. timelines/temporal view, splitting and grouping of data, collapsing, and expanding results, structure and style of view such as cluster or hierarchies in order to find patterns and trends, explore complex relationships or work through multiple records of information. 		
	<ul style="list-style-type: none"> The platform must cater for the security classification of a user of the analytics capability (e.g. a user with "secret" access, may not view "top secret" records in the visualisation tool). 		
2.2.5	Search Engine		
	<ul style="list-style-type: none"> The platform must provide a dynamic capability that searches and retrieves different types of data (structured and unstructured data) from the various sources (internal and external), and generates access-controlled, meaningful, and relevance-scored search results. 		

	<ul style="list-style-type: none"> • The platform must provide for 'fuzzy' and 'elastic' search functionality. 		
	<ul style="list-style-type: none"> • The search and query results must provide for different views, eg tables, grid, map etc 		
	<ul style="list-style-type: none"> • The platform must support the sharing of search results to other users/analysts. 		
2.2.6	Rules Engine		
	<ul style="list-style-type: none"> • The platform must enable configurable rules engine functionality to allow the setting of triggers, alerts, notifications, workflow processes and business rules that can be easily modified. 		
	<ul style="list-style-type: none"> • The platform must cater for various levels of configuration by a user (e.g. end-user level, business administrator level, system administrator level, etc). 		
2.2.7	Reporting & Dashboard		
	<ul style="list-style-type: none"> • The platform must provide for the flexible creation, by administrative users of configurable dashboards, and by users for reports that cater for different views and customisations. 		
	<ul style="list-style-type: none"> • Users must be enabled to produce and schedule both standard/production and ad-hoc reports. 		
	<ul style="list-style-type: none"> • The platform must also cater for the auto-population of standard/production reports (eg a status report automatically updated from the required structured data) 		
	<ul style="list-style-type: none"> • Dash boarding/reporting must be as close to real-time as possible. 		
	<ul style="list-style-type: none"> • The proposed solution must provide for basic, intermediate, and advanced Analytic Visualisations 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> ○ <i>Creating basic chart types</i> 		

	▪ <i>Table;</i>		
	▪ <i>Bar chart;</i>		
	▪ <i>Line chart;</i>		
	▪ <i>Area chart;</i>		
	▪ <i>Pie chart;</i>		
	▪ <i>Sparklines;</i>		
	▪ <i>Candlestick;</i>		
	▪ <i>Scatter.</i>		
	○ <i>Creating intermediate chart types:</i>		
	▪ <i>Combo chart;</i>		
	▪ <i>Heat map;</i>		
	▪ <i>Box plot;</i>		
	▪ <i>Histogram;</i>		
	▪ <i>Bubble;</i>		
	▪ <i>Bullet;</i>		
	▪ <i>Pareto;</i>		
	▪ <i>Tree map;</i>		
	▪ <i>Trellis;</i>		

	▪ <i>Word cloud:</i>		
	▪ <i>Geo-mapping.</i>		
	○ <i>Creating advanced chart types:</i>		
	▪ <i>Marimekko;</i>		
	▪ <i>Network and Link Analysis;</i>		
	• Displaying values as percentages and variances via toggles, buttons versus creation of new metrics.		
	• Supporting the ability for selections in one chart to filter related chart objects.		
	• Providing color-coded summary of the state of a particular metric compared to a goal or threshold target, without programming.		
2.2.8	Self-Service Data Preparation		
	• The platform should enable end-user with an interface from within various types of analytic content to view sources, calculations, and manipulations in the underlying business user data mash-up.		
	• The platform should enable end-user data mashup from multiple data sources,		
	• The platform should provide for end-user data modelling, including custom groups, hierarchies, filtering, and calculations (e.g., combining multiple data sources and applying logic/transformations to create a dataset ready for analysis).		
	• The platform should enable end-user to join data sources (structured and/or unstructured) and can access a variety of joins between data sources (i.e., full outer, inner, left, and right, etc.).		

	<ul style="list-style-type: none"> The platform should auto infer and suggest relationships between different data sources for joining; infer and auto formats data types such as date, measurement, geography and hierarchy; and provides capabilities to preview inferences before accepting. 		
	<ul style="list-style-type: none"> The platform should enable user data enhancement, such as renaming, combining or splitting columns, automatic date transforms, replacing values, etc. 		
	<ul style="list-style-type: none"> The platform should allow the user to mask/redact/encrypt data within the workflow to hide protected/sensitive information? 		
	<ul style="list-style-type: none"> The platform should allow for a data model and individual metadata objects (e.g., dimensions, measures, calculations, parameters) to be shared across applications, reports and dashboards. For example, a common dimension such as ID Number be modelled once and shared. 		
	<ul style="list-style-type: none"> The platform should provide for end-user viewing of statistics on the overall quality and distribution of the data, potential issues with the data and facilities that auto recommend actions to fix identified issues. 		
2.2.9	Embedded Advanced Analytics		
	<ul style="list-style-type: none"> The platform should allow for building decision trees via a menu-driven interface. 		
	<ul style="list-style-type: none"> The platform should allow for providing forecasting (AI, trends, predictions, etc) and text analytics (including Word Cloud and sentiment analysis) via a menu-driven interface. 		
	<ul style="list-style-type: none"> The platform should allow for performing k-means clustering via a menu-driven interface, then allows the user to use a cluster as a filter in subsequent analysis. 		
	<ul style="list-style-type: none"> The platform should allow for analytical functions, such as time series analysis, clustering, estimation, classification, affinity analysis and attribute importance. 		

	<ul style="list-style-type: none"> The platform should provide for user-defined algorithms in a plug-and-play fashion, and import/export via Predictive Model Mark-up Language (PMML), R-based models or Python. 		
	<ul style="list-style-type: none"> A library of commonly used descriptive statistical functions, including mean, min., max., standard deviation, confidence interval and hypothesis testing using basic statistical capabilities (e.g., t-test, chi-square). 		
2.2.10	Interactive Visual Exploration		
	<ul style="list-style-type: none"> Supporting the interactive capabilities, such as: 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> <i>Excluding data from views;</i> 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> <i>Cascading filters;</i> 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> <i>Drill down and up for different chart types;</i> 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> <i>Drill down and up for dashboards;</i> 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> <i>Sort;</i> 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> <i>Zoom;</i> 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> <i>Panning and Brushing;</i> 		
	<ul style="list-style-type: none"> Show detailed behind a visualization 		
	<ul style="list-style-type: none"> Interactively create custom groupings for use as a new dimension or filter or within a hierarchy or offer similar functionality. 		
	<ul style="list-style-type: none"> Allowing users to apply filters globally across all content within an analysis or across selected content, as determined by the user. 		

	<ul style="list-style-type: none"> • Binning capabilities. 		
	<ul style="list-style-type: none"> • Interactive formatting options for charts, such as: <ul style="list-style-type: none"> ○ <i>Category colour consistently displayed across all charts;</i> ○ <i>Ability to modify colours;</i> ○ <i>Legend placement and toggling;</i> ○ <i>Set scale.</i> 		
	<ul style="list-style-type: none"> • Advanced visualization features (without coding): <ul style="list-style-type: none"> ○ <i>Multidimensional rendering;</i> ○ <i>Representation of isosurfaces and contours;</i> ○ <i>Asymmetric reporting.</i> 		
2.2.11	Smart Data Discovery		
	<ul style="list-style-type: none"> • Automatically generating advanced analytic visualizations (such as the ability to visualize correlations or clusters in a dataset, display decision trees, etc.). 		
	<ul style="list-style-type: none"> • Menu-driven capabilities for forecasting, trends, predictions, clustering, segments, correlations, factors analysis, etc. Including types of models and algorithms supported (linear, Winters, etc.). 		
	<ul style="list-style-type: none"> • Automatically generating advanced analytics models that can be viewed and modified by specialist data scientists. 		
	<ul style="list-style-type: none"> • Executing or filtering reports through queries entered into a search bar in natural language 		

	<ul style="list-style-type: none"> Automatically generating advanced analytics models to be operationalized into recurring analysis and processes. 		
2.2.12	Analytic Dashboards		
	<ul style="list-style-type: none"> Autonomously, and code-free business-user authored data discovery, and dashboards, as well as "WYSIWYG" design to users without expertise in data models, SQL, ETL, application development, MDA reports design or advanced analytics. 		
	<ul style="list-style-type: none"> Users must be able to create and edit metrics and dimensions; 		
	<ul style="list-style-type: none"> Users must be able to create and edit hierarchies and groupings of data items; 		
	<ul style="list-style-type: none"> Users must be able to create parameters to control visualizations? 		
	<ul style="list-style-type: none"> Users must be able to to change Dashboard design formatting options for charts, such as 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> Category colour consistently displayed across all charts; 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> Ability to modify colours; 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> Legend placement and toggling; 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> Set scale. 		
	<ul style="list-style-type: none"> <ul style="list-style-type: none"> Placing multiple visualizations on a page, multiple tabs, formatting of borders, titles, etc. 		
	<ul style="list-style-type: none"> Users to customize dashboard themes and save as a style sheet for future use on a new dashboard. 		

	<ul style="list-style-type: none"> Google Maps, OpenStreet, MapBox or similar maps being supported out of the box. 		
	<ul style="list-style-type: none"> Building custom map layers and shape files; 		
	<ul style="list-style-type: none"> Mapping with open data and external data service providers (e.g., demographics, population). 		
2.2.13	Publish, Share and Collaborate		
	<ul style="list-style-type: none"> Publishing and exporting of reports of analytic content to other Enterprise Portals and Platforms in the following formats: HTML, PDF, XML, Excel, comma-separated values (CSV), RSS feed and PPT. 		
	<ul style="list-style-type: none"> Layered PDF creation (the ability to create layers that display while printing but not on the screen, and visa versa - renaming and merging layers; changing the properties of layers; and/or locking layers to prevent them from being hidden. 		
	<ul style="list-style-type: none"> for users to search the report repository and analytic content for keywords contained in the MDA metadata. 		
	<ul style="list-style-type: none"> allow for users to have a code-free means to blend data visualizations, text, multimedia content and links to external content, to create a live infographic-type object. 		
	<ul style="list-style-type: none"> content to be disseminated to end users on a scheduled basis and in a variety of output formats. 		
	<ul style="list-style-type: none"> allow for the scheduling of reports and dashboards to be distributed based on user-defined business events, such as inventory below threshold. 		
	<ul style="list-style-type: none"> allow for the delivery of alerts to users based on defining thresholds in a report - being notified if a specific condition exists. 		

	<ul style="list-style-type: none"> allow for pixel-perfect or highly formatted reports. 		
	<ul style="list-style-type: none"> enable users to print either a portion of, or an entire dashboard, or report. 		
	<ul style="list-style-type: none"> allow users to print to "fit to page," with headers on each page. 		
	<ul style="list-style-type: none"> allow users to post reports, graphs and analysis into public folders for others to access. 		
	<ul style="list-style-type: none"> allow for the creation of a storyboard to present a series of analytic content. 		
	<ul style="list-style-type: none"> allow users to enable discussion threads and commentary on shared MDA content (reports, dashboards and analysis). 		
	<ul style="list-style-type: none"> allow users to see a snapshot of the data at the time of the comment in the discussion thread? 		
	<ul style="list-style-type: none"> allow users to see their collaboration, annotations and comments on a timeline. 		
2.2.14	Workflow Integration		
	<ul style="list-style-type: none"> Provide workflow capabilities, and 		
	<ul style="list-style-type: none"> Provide the ability to design custom workflows 		
2.2.15	Ease of Use and Visual Appeal		
	<ul style="list-style-type: none"> Ability to incorporate the SIU Logo for ownership of the report. 		
	<ul style="list-style-type: none"> Easy to use layout for end users. 		