**Date:**      January 19, 2022

**RFI Title:**   METADATA GENERATION AND DIGITAL ASSET MANAGEMENT (PHL)
**Department:** Commerce Department, Division of Aviation

### Addendum ONE Cover Sheet

**TO ALL APPLICANTS:**

**You are hereby notified of the following information pertaining to the above RFI:**

1) Attendee sign-in sheet from the Informational Meeting held on January 6, 2022.
2) Aviation's response to vendor questions.


_____                    _____
**AUTHORIZED SIGNATURE**                                 **NAME (Print)**


_____                    _____
**COMPANY NAME**                                         **TITLE (Print)**


_____
**DATE**

**Metadata RFI informational meeting attendees**

1. Carol Bee Ab Initio Software 201 Spring Street Lexington Massachusetts 02421 Cell: 617-828-7509 email: cbee@abinitio.com www.abinitio.com

Not M/W/DBE

2. Stephen Reed Senior Account Manager Smartlogic, Inc. (although we were recently acquired by Marklogic, Inc.)

111 N Market St, Suite 300 San Jose, CA 95113

Phone: 703-608-5943 E-Mail: stephen.reed@marklogic.com; Stephen.reed@smartlogic.com

We do *not* fall under any of the M/W/DSBE business designations.

3. Bob Parker

Sr. Director

Cambridge Semantics

1 Beacon St.

Boston, MA 02108

781-760-2813 bob.parker@cambridgesemantics.com

CSI is a Small Business but does not have any other designation (W/M/DSBE).

4. Clemens Burmeister

Customer Success Manager

Semantic Web Company GmbH PoolParty Software Ltd.

Semantic Web Company GmbH

Mariahilfer Straße 70, Neubaugasse 1/8

1070, Vienna, Austria

Frontdesk: +43-1-4021235

Semantic Web Company, Inc.

One Boston Place, Suite 2600

Boston, MA 02108

Phone: (857) 400-0183 clemens.burmeister@semantic-web.com

5. Tommy DeAngelo

Account Executive

Boomi

1400 Liberty Ridge Dr., Suite 200, Chesterbrook, PA 19087

(862) 881-1977 Tommy.deangelo@dell.com

Laurence Ross

Account Executive

Boomi

1400 Liberty Ridge Dr., Suite 200, Chesterbrook, PA 19087

(617) 584-9239 Laurence.ross@dell.com

Boomi, LP is not an M/W/DSBE

6. Elizabeth Squire

Business Development Director (Government and Public Sector)

Solidatus

30 Stamford Street

London

SE1 9LQ

United Kingdom elizabeth.squire@solidatus.com

+44 7957 366812

We are not a M/W/DSBE company

## Metadata RFI Questions and Answers

| Question Number | RFI Section # (if applicable) | Questions | PHL Response |
|---|---|---|---|
| 1 | Users (Authors & Analytical Users) | How many Authors and Analytical Users will PHI require Enterprise wide (meaning across all business units). | Authors/Stewards: 15 ; Analytical Users: 45 |
| 2 | Users (Business Users/Readers) | How many Business users/Readers will PHI require Enterprise wide (meaning across all Business units). | Business Users/Readers: 90 |
| 3 | Data Base Connections (Standard Databases) | How many Enterprise wide Standard Databases will need to ne connected to. Standard databases - SAP S4 HANA, Oracle, SQL Server, Salesforce, Workday - each physical server is a connection. | Enterprise wide Standard Databases: 6 PRODUCTION |
| 4 | Data Base Connections (Clustered Databases) | How many Enterprise wide Clustered Databases will need to me connected to. Ex: Snowflake, Hive, Exadata, Terradata | Enterprise wide Clustered Databases: 1 PRODUCTION + SharePoint |
| 5 | Data Base Connections (Visualization tools & applications) | How many Enterprise wide Viualization tools & applications will need to be connected. Visualization tools & Applications - Tableau, PowerBI, QlikView, SISENSE, SAP Business Objects, - each instance is a Database. | Enterprise wide Visualization tools: 4 (2 PRODUCTION, 2 non-PRODUCTION) |
| 6 | Auto - Lineage & Impact Analysis | The number of connections that will need to establish and show upstream and downtream affected data objects. (These auto-lineage connections allow for impact analysis to be conducted) EX: ETL/ELT, BI/Analytic platforms/tools | Auto-lineage connections: 10 |
| 7 | Support & Services | 24X7 support comes with the Enterprise package - will PHI prefer onshore or offshore support or a mix determined by tiers and urgency level. | 24X7 support: PHL prefers a mix determined by tiers and urgency level. |
| 8 | Support & Services | The Enterprise package has 20 workshops(1 or 2 hour sessions) built into the package - will PHI require more than 20 workshops - How many PHI employees be attending (Authors/stewards & Readers/Business users). | PHL requires enough training for users to become proficient. If the selected software is easy enough to use the number of workshops required will be fewer, if more difficult, more workshops. For the number of users attending, see answers to questions 1 and 2. |
| 9 | Support & Services | Will PHI prefer to have a dedicated OvalEdge Governance engineer assisting with implementing PHI's Governance best practices. | PHL has strong data governance experience. The need for high level assistance will depend on the complexity of the selected product. If it is easy enough to use without a high level engineer, none will be needed. |
| 10 | Implementation & Deployment | Will PHI prefer to have OvalEdge handle the entirety of the implementation/deployment. | The selected vendor must provide technical resources to assist with the implementation/deployment of their software. It is expected that the work will be a collaborative effort between PHL personnel and vendor personnel. |
| 11 | Implementation & Deployment | What is PHI's current expectation and timeline for Go-Live. | PHL hopes to go live with the metadata management solution during calendar year 2022. |
| 12 | Implementation & Deployment | How many internal Resources will PHI be devoting to this project. | For the implementation, deployment and roll-out of the metadata management solution, PHL plans for seven (7) people: PM, Data Architect, DBM/DBA, Infrastructure Engineer, Software Engineer, QA Engineer, Analyst/Trainer. |
| 13 | Pricing/License Model | Would a test, staging and production server environment (three tier) be your preferred architecture? | PHL prefers separate QA, Pre-Production Staging and Production environments. |
| 14 | Data governance | What is the type of (basis) basic documents for metadata generation? | PHL has documents in multiple formats including all MS Office formats, Adobe PDF and ASCII/ANSI text |

| | | | |
|---|---|---|---|
| 15 | Data governance | What does it exactly mean - generating metadata for databases? Do you expect to generate the data catalog (metadata model of connected technologies) from the connected databases, or do you expect some kind of export functionality which would create a metadata model from e.g. a conceptual model? | PHL expects concept metadata generation describing the contents of database tables and columns. This implies functionality that would create a conceptual metadata model from scanning the physical model and inferring the conceptual meaning, where possible, based on the table and column names. This may require a strong a-priori predictive model of enterprise data concepts. The first several "expected benefits" clarify the outcomes:<br>•The data catalog serves as an index to all content in the data lake.<br>•Can help reduce data duplication, especially in new implementations<br>•Can help reduce manual scrubbing/cleansing (requires good UI...)<br>•Semantic layer to point Self-service BI/report users to the right data<br>•Master Data Management to control KPI and CSF definitions and formulas<br>•Searchability across structured, semi-structured, and unstructured digital data |
| 16 | Data governance | You are stating in the section "Data Governance" that the tool should properly classify information and generate metadata including similar data from different systems. What does it mean similar data? Are we at the level of harmonised meaning of the metadata, or is it just connected to the same classes (clusters) of the data? What kind of similarities are you looking for? | Aggregating information from different systems/databases that contain overlapping content, such as vendors, contracts, invoices and payments, requires metadata that correctly classifies information conceptually. Associating quantitative information from databases with qualitative information from documents to provide qualitative analytics/intelligence requires the ability to make associations between the conceptual meaning of the associated database content and document content. This includes "harmonised meaning of the metadata" where possible and connections between classes of data where that is the best that can be achieved. We anticipate a combination of both. |
| 17 | General | Is part of the scope to define the organizational processes connected to e.g. creation of the data inventory? Or some other project related tasks? | If the selected tool provides process modeling and metadata capablities that are intrinsically tied to content metadata, this would be a plus, but it is not a requirement. |
| 18 | Data governance | The metadata model retrieved from the database is a blueprint of a technical world. On the other side there can be a business model, which is standardly created based on harmonization meetings with different stakeholders. How do you intend to retrieve the semantic/conceptual layer (the business model) over your databases (the technical world)? | PHL Business SMEs, Data Stewards, Business Analysts, Technical Engineers and Analysts will all participate in the metadata curation process, as needed to prime the metadata generators prior to scanning and fine-tune the output to ensure the catalog accurately reflects the conjunction between the business and technical models. |
| 19 | Data governance | Can you provide more details how the real time synchronization (mentioned in the point relevant to transaction management standards) is associated with the metadata/data inventory tool? | Engineering transactional and bulk data exchange between different systems requires accurate mapping. The catalog is expected to significantly help data synchronization engineers in defining and implementing such interconnections, and may provide technology that becomes part of the synchronization process itself (a plus). |
| 20 | Data governance | Can you clarify a bit how the streaming data management standards are associated with the metadata/inventory tool? We understand metadata/inventory tool should collect just the metadata, not referring to real data itself. Is there a use case you can share with us? | The receiving endpoints of streaming data are the easiest objects from which to collect the metadata and understand the conceptual categories of information in the stream. "Define streaming data management standards and tools" implies an ability for data governance personnel to document streaming standards. The "in-flight data scanning and filtering, to ensure Internet of Things (IoT) and other streaming data sources can be properly managed and monitored" implies the ability to scan data as it arrives at the endpoints for anomalies and data of interest based on conceptual markers, usually specific words or combinations of words, possibly including numbers and ranges. |
| 21 | Data governance | What kind of data integration do you want to support? Which systems do you want to integrate the data catalog with? | Relational databases such as SQLServer and Oracle with te ability to integrate data with future implementations of graph databases, document data stores or other "Big Data" sources. |
| 22 | Data governance/Data Integration | Do you want to do the data integration with our tool, or do you want to just document which data integrations you have in your processes? | See answer to question 19. Mostly metadata management supporting other integration tools. |
| 23 | General | Which technologies do you want to connect to? Which databases do you want to have data catalog | See answer to question 21. |
| 24 | Self Service BI | Do you expect that the complete conceptual/semantic layer will be created completely automatically without any business alignment (harmonization) workshops? | No |
| 25 | Self Service BI | Do you want to cover reporting with this tool as well? Or do you want to focus only on data governance and description of your metadata in different kinds of models? | Metadata generation and Data Governance. See answer to question 57. |
| 26 | Self Service BI | How do you expect that inclusion of salient unstructured data should work? Can you elaborate this? | By making conceptual associations between narrative text in unstructured content and quantitave results |
| 27 | Cognitive search | Do you want in the search to search for metadata only or data itself as well? | The metadata should point to the data itself. This implies FTS (full text search) indexing. |

| 28 | Data Integration | How should a mapping for transactional and bulk data synchronization look like? What should it capture? Can you give some examples? | Mapping should enable technical users to see the column names and types with the conceptual association. For example: Vendor:Contract may be mapped to Invoice:Contract may be mapped to Payment:Invoice to associate payments with vendors. |
|---|---|---|---|
| 29 | Data Integration | Which relevancy do you see in the mentioned data mover and creation of data inventory (metadata model)? Can you describe the use cases for this? Which data do you want to move from where to where? | See answers to questions 19 and 55. |
| 30 | Data Governance |  but we are developing an ontology that contains a taxonomy of classes that will be used as part of this implementation | All conceptual classes to which our data applies. The list is too large for this response but PHL is developing an ontology that contains a taxonomy of classes and associations that will be used as part of this implementation. |
| 31 | | What are the data volumes in scope? What databases? What file formats? | See answers to questions 49, 57. |
| 32 | | What is the diversity of file types, number of documents, total volume of data? | Much diversity and volumes consistent with a medium-sized enterprise. |
| 33 | Data Governance | What does IoT mean in your context? Is this structured or unstructured data? | IoT refers to streamed structured and unstructured data from external sources and sensors deployed throughout the airfield and the terminals. |
| 34 | | What are the different types of data sources. Are they integrated? Are they located on the same space? Are you looking to virtualize the access to your data source? | See answers to questions 49, 57. They are located in the cloud and on premise. Data access is not part of this RFI. |
| 35 | | Do you want to discover relationship between your datasets from different sources? | The conceptual relationships between similar data in different sources will support such discovery, automated or manual. |
| 36 | II.B | For the semantic search, "What are some examples of the questions you want to ask?" | Any questions that interest airport executives such as "What is the average time between passenger ticketing and boarding for each air carrier?" and "What is the longest time between touchdown and gate arrival for each gate?" |
| 37 | II.B, II.D, VII. | Can you please provide some examples of KPIs and/or CSFs? | KPIs may include "Elapsed time from badge application to Issuance" and "Ratio of Federal grants to commercial loans for capital projects". CSFs may include "NPS based on direct Passenger feedback" and "Average time from reported custodial issue to resolution." |
| 38 | II.E | What Business Intelligence tools are you currently using? | Primarily PowerBI |
| 39 | II.B | What technologies and formats are you using in your Data Lake? | Any unstructured document format and primarily SQLServer. |
| 40 | | Do you already have Taxonomies/Ontologies? Or, are you expecting the vendor to create them? | PHL already has Taxonomies/Ontologies and may seek vendor assistance in expanding, updating and reformatting them based on solution requirements. |
| 41 | II.C | Can you please elaborate on your request to provide metadata at the document and paragraph level? Do you mean to generate "features" of the document and paragraphs to support semantic search? | Yes. Features include keywords and key concepts, as well as markers for tone and associations with categories in the ontology. |
| 42 | II.B | Is the RFI asking for "chatbox" semantic functionality? | No |
| 43 | | Moving forward to improving airport operations, are you interested in the following areas?<br>· Automating and streamlining the passenger journey<br>· Passenger processing and flow management<br>· Flight Operations<br>· Baggage Processing<br>If so, what data do you currently collect and what additional data would you like to collect to improve real-time airport operations? | This is not applicable to this RFI. |
| 44 | | Do you require one vender to do everything or work with other vendors to provide a complete solution or are you willing to tender to different suppliers to complete different parts of the project? | PHL is open to single-vendor and multi-vendor solutions, but sensitive to cost and complexity. |
| 45 | | How will the software be procured, will it be a direct procurement or will you be using any specific frameworks? | It may be direct procurement but will be determined at a later time. |
| 46 | | Do you have preferred implementation partners? | PHL has internal personnel and external contractors who could contribute the technical and other work, but PHL is open to other partners. |
| 47 | II.B | "scope of this project will begin with two (2) or three (3) core systems", have you identified the initial 2 or 3 core systems? If so, what are they? | Yes. One internally developed system based on SQLServer data and Maximo based on DB2 and Sciforma with a proprietary data structure, all with industry standard APIs. |

| | | | |
|---|---|---|---|
| 48 | | How many systems would be involved once you have established frameworks and practices? | Up to 300 systems |
| 49 | | What are the different data sources in use by PHL? Any planned future ones? | Examples include SharePoint, Azure data lake, PowerBI, commercial software such as Maximo and Amadeus, internally developed software, mostly Dot.net stack, all MS Office formats, Adobe PDF and ASCII/ANSI text and future implementation of COTS systems such as Workday or Oracle. |
| 50 | | What are the security requirements for personnel involved in the project? | PHL would prefer remote support and would prefer vendor personnel to not have direct access to PHL systems or infrastructure other than that directly associated with the solution itself. If onsite support or access to internal systems is required, vendor personnel would be subject to PHL security access requirements. No other special secuirty clearances would be required. |
| 51 | | What are the timelines for feedback on the RFI? | Please see Section V. Anticipated Timeline for initial feedback. If PHL has an interest in continuing discussions with any vendors it is likely that they will be contacted during March 2022. |
| 52 | | What are the next steps on the process after the RFI? | PHL will review and score the responses, then decide whether we will proceed to the RFP phase. |
| 53 | | How many power users and read-only users do you envisage using the solution? | See answer to questions 1 and 2. |
| 54 | II.B | Regarding the scope referred to in the Background section of the RFI, what are the names of the "two (2) or three (3) core systems" PHL plans to integrate first? | See Answer to question 47. |
| 55 | | Can you please further describe the ETL processes you are looking to be completed? What are the source and target systems for the loads? For bulk data integrations, can you please provide estimates for the volume of records in a bulk data integration (i.e. total records per day, peak records per hour, etc.)? | See answers to questions 16 and 19. PHL's data integrations are typical for mid-sized organizations such as ours. PHL will not provide details as we use standard data integration tools and we do not consider these details relevant to the solution selection process. |
| 56 | | How many end users does PHL anticipate would be working within a data catalog and preparations tool? Can you describe the anticipated user roles and responsibilities? | End users: See answer to question 1. Roles and responsibilities: Authors are data stewards/curators and technically proficient data architects and modelers assisted by business analysts. They are responsible for identifying sources to include in the catalog, priming the system with information to improve the scanners' ability to correctly infer conceptual metadata, and reviewing/curating the output to fine-tune the metadata in the catalog. |
| 57 | | Can you please provide more details on the number of unique endpoints (endpoints can be an application, database, file, datastore, etc.) that PHL needs to integrate with? Roughly how many total endpoints (source and target systems) are in scope? What data warehousing tools are currently being used? Do you have any existing tools around governance? What reporting tool(s) us currently being used? | Total endpoints in scope: 30 applications and their databases Data Warehousing tools: mostly Azure Data Lake, SQLServer, some Oracle, some DB2 Governance Tools: Azure DevOps Reporting Tools: SSRS, PowerBI |
| 58 | | How many test environments are required? | One test environment is required. |
| 59 | | Our company offers several implementation models with varied levels of consulting engagement: Full implementation, Enablement, and Self-Implementation.  Can PHL please clarify which implementation model it would prefer to execute? | PHL's preferred model is "enablement and self-implementation" but our experience has shown that the meaning of "enablement" profoundly differs from one solution to another based on the maturity, complexity and reliability of the solution's deployment model and facilities. |