

4050 Esplanade Way Tallahassee, FL 32399-0950 850-488-2786

Ron DeSantis, Governor Jonathan R. Satter, Secretary

Request for Information (RFI)

Enterprise Data Catalog Solutions

DMS-19/20-078

Addendum No. 4

The Department's responses to timely submitted questions are below.

Question Number	Question	Response
1	What are the data sources (including total data size in TB) for data ingestion/Consolidation?	Part of the expected work contemplated by this RFI is for vendors to obtain the size of state databases.
2	Does the Data sources include both Structured and Unstructured data for Data ingestion/Consolidation?	Yes.
3	Number of End-users Accessing the Data from the Proposed System?	Unknown.
4	Are there any existing BI/reporting tools? If so, is there any requirement to integrate with existing BI tools?	Existing Bl/reporting tools are Qlik PowerBl, Tableau and Alteryx.
		Yes, there is a requirement to integrate with existing BI tools.

5	Is there any existing Big Data Environment (Ex: Hadoop)? if yes, what is the distribution used? If No, is there any plan to get up Big Data Repository for Data	Not currently. There may be plans to set up a Big Data Repository as we move the data center to the cloud.
	there any plan to set-up Big Data Repository for Data Consolidation, Access and Reporting?	
6	Does DST has any preferred deployment models On- Premise/On Cloud/SaaS ?	On cloud/on-premises at first but will transition to on cloud.
7	Can DMS provide a list of sources for metadata? i.e. Oracle, SQL Server, Snowflake, SAP, Salesforce, etc?	Oracle and SQL Server are known but multiple others may exist outside of these two sources for metadata.
8	Does DMS use any cloud platforms (or plan to) such as AWS, Azure, GCP, etc. If so, which ones and what products on those platforms?	The future goal is that there will be no limitation. Consequently, all may be available.
9	What tools does DMS use for analytics and reporting? i.e. Tableau, PowerBI, Cognos, Microstrategy, etc	The Department currently uses PowerBI, Tableau, Smartsheets and Alteryx.
10	Would cataloging and tagging data based on column headers be sufficient for your needs, or will DMS want to leverage AI/ML to look for patterns in the data and tag the data based on those patterns?	Both.
11	To encourage enterprise-wide use of the catalog, does DMS prefer a pricing structure not tied to users?	Yes.
12	Given that some cloud-based catalogs could create security issues (employee backdoors, treasure maps in the cloud, requiring inbound ports for scanning, etc), does DMS prefer an on-prem solution or is the "to-be" architecture cloud-based?	Cloud-based.
13	Regarding Question 14, could you provide more details regarding regarding the question about general pricing structure? Are you trying to understand how it is priced? Or do you need a quote for services, EDC solution, training, support? If so, such a quote would require additional information.	In section IV. Response, subsection B, item 14, the Department is seeking to understand vendors' general pricing structures and how each vendor determines pricing for its solution. No price quote is being requested with this RFI.
14	What are the higher level business imperatives behind the state wanting a statewide EDC? Self Service? Analytics?	Please see <u>Florida House Bill 1391 (2020)</u> for more information on the legislative requirements of the state data catalog.

45	Depending Question Question states quellity functionality is	All DELevie with the new investigation of the sector of the meteorical
15	Regarding Question 6, what data quality functionality is	All RFI submissions will be reviewed to understand the potential
	needed to for this project ? E.G; Data Profiling,	cost vs. benefit of different solution types and features.
	Standardization, Prebuilt Reusable Rules,	
	Deduplication? Cleansing?Scorecarding data?	
16	Regarding Section III, Goal, You stated there is a need	Any state and federally regulated data sets, sensitive data, or
	to identify sensitive and regulated data. Once this data is	Personally Identifiable Information (PII) needs to be handled
	discovered, a common best practice to provide safe	and governed appropriately, and in accordance with the
	analytics is to automate data masking for PII data. Is this	applicable state and federal laws and regulations.
	a requirement?	
17	How many datasets are there in total?	Unknown at this time. Depending on adoption by agencies, the
		number of datasets could range widely.
18	Is the department currently maintaining a DWH/DB?	Yes, multiple instances.
19	Any other reporting to be done after documenting Data	Yes.
	Lineage?	
20	Please provide details on the input data information -	The Department is seeking cloud native; otherwise, please see
	size, data dictionary, layouts, connection type, where is	Florida House Bill 1391 (2020) for more information on the
	it hosted, etc.,	legislative requirements of the state data catalog.
21	What format files can be provided in (Excel, SPSS, CSV,	Excel, CSV.
	etc.)?	
22	How much data will be anticipated annually in MBs, GBs,	Depending on adoption by agencies, data size could range
	or TBs?	widely.
23	Does the department expect to deploy any dashboards	Yes.
	or BI tools for visualization?	
24	What is the hosting preference of the department? (i.e.,	Cloud.
	Cloud or On-prem)	
25	Can we expect the department to release a solicitation	The Department will use the responses submitted for this RFI to
20	for this project soon? If yes, could you provide us the	determine the type of solution that will best meet its needs and
	estimated date?	the best procurement method for obtaining a solution. Any
		resulting procurement will occur during Fiscal Year 2020-2021.
26	If the department is satisfied with our POC, is there any	Future purchases will be conducted in accordance with Chapter
20	possibility to enter into a contract directly without	•
	releasing a solicitation?	207, 110100 3000000000000000000000000000000
	ו דובאווע א געווטומווטוו	

27	1What will be the overall budget to be associated with	Please see Florida House Bill 1391 (2020) for more information
	the project?	on the legislative requirements of the state data catalog.
28	Can you please provide the data stores or data sources that will be ingested in the EDC? RFP Location III	No.
29	How is the Department currently carrying out data cataloging? RFP Location III	There is currently no statewide data catalog managed by the Department. However, multiple agencies have independently implemented data catalogs. The Department is looking to implement an all-inclusive data catalog.
30	What data sources will the data be collected from?	Assume any and all data sources commonly used in government agencies.
31	Can you list the make and model of storage arrays in the DMS environment?	The Department does not believe this question is relevant to the RFI.
32	What is the backup platform used at DMS?	The Department does not believe this question is relevant to the RFI.
33	How many data sources will data be collected from?	Depending on adoption by agencies, the number of data sources could range widely.
34	Is data being sent to an archive for long term retention?	The Department does not believe this question is relevant to the RFI.
35	Are data lifecycle policies in place today? If so, what is the retention policy?	Depending on adoption by agencies, the retention policies could range widely as these are determined by the agency customer.
36	Is there any requirements to perform eDiscovery searches?	Please see <u>Florida House Bill 1391 (2020)</u> for more information on the legislative requirements of the state data catalog.
37	Are there any reporting or alerting requirements for the proposed solution?	Yes, both are needed.
38	When is the Ent Data Catalog project planned or estimated to start?	It is estimated to start during Fiscal Year 2020-2021, following the procurement process.
39	What solutions if any have been tried in the past by the Florida Department of Management to stand up an Enterprise Data Catalog?	See response to question 29.
40	What components are currently in place or are anticipated to be leveraged, i.e DQ tool, Data Glossary, ETL lineage, etc.?	Please see <u>Florida House Bill 1391 (2020)</u> for more information on the legislative requirements of the state data catalog.

41	What are considered to be the top 3-5	1. Agency adoption
	challenges/obstacles to overcome to successfully	2. Security and role access
	implement the Enterprise Data Catalog?	3. Reportability of catalog data
42	How many users do you anticipate for the 1st year of	Depending on adoption by agencies, the number of users could
	the Data Catalog? How many do you anticipate by the	range widely.
	3rd year?	
43	What Departments are likely to be the early users of the	Depending on adoption by agencies, departments that are likely
	Enterprise Data Catalog?	users are presently unknown.
44	Are there Data Governance committees in place at the	Yes, various Data Governance committees are in place.
	State and/or Department levels? If so, which ones?	Depending on adoption by agencies, the data governance
		committees could range widely.
45	Has Florida Department of Management identified Data	Yes, various Data Stewards are identified at both agency and
	Stewards for key data sets? If so, which data sets?	state level.
46	What are the top 3-5 uses cases for the enterprise data	This information is not available.
	catalog in business operations terms (non-IT)?	
47	Has the Florida Department of Management identified	Yes.
	the top 5-10 systems that you anticipate collecting	
10	metadata from for the catalog?	
48	Can you share any examples of data domains for the	Depending upon adoption by agencies, the number of data
	data catalog (ex: assets, citizens, employees)?	domains may vary widely.
49	Can you share examples of workflows you would like in	This information is not available.
	place to govern the data catalog?	
50	Can you share examples of data policies you would	Many policies will be implemented based on adoption by
	want to enforce?	agencies.
51	What are the appx sizes of your various data	Depending on adoption by agencies, data repository sizes could
	repositories/assets?	range widely.
52	Are all datasets predefined or will you expect this	Predefined, yes, but new data sets will be created.
	solution to also manage creation of datasets	
	(indexing)?	

53	What applications, databases, or data warehouses will	Please see Florida House Bill 1391 (2020) for more information
55	be considered for indexing? Or which will be	on the legislative requirements of the state data catalog.
	contributing data to the Enterprise Data Catalog (EDC)	Depending on adoption by agencies, the number of
	Solution.	applications, databases or data warehouses could vary widely.
		applications, databases of data warehouses could vary widely.
	 How many sources How many datasets 	
	5	
E A	How many datasets yet to be created	The evenested outcome from the date estales implementation is
54	How many datasets do you currently maintain? How do	The expected outcome from the data catalog implementation is
	you currently maintain information related to	data lineage.
	relationships or links between datasets?	
55	What is the current categorization method/policies,	Various, depending on agency involvement.
	etc.?	
56	Will you provide DB schemas and other background	Depending on agency adoption, documentation will be provided
	information not limited to data manuals, procedures,	as needed if available.
	dictionaries, etc. to enhance vendor understanding of	
	your solutions and sources of your data?	
57	Will you provide for the duration of an implementation	Depending on agency resource availability, technical experts
	access to technical experts that can help facilitate	will be provided as needed if available.
	changes or modifications on the client side required for	
	successful implementation?	
58	Do you have a clear understanding of where your	Communication of guidelines of sensitive and regulated data will
	sensitive and regulated data currently resides? Will you	be provided based on agency resource availability.
	provide clear guidelines on how to manage visibility of	
	sensitive or regulated information datasets?	
59	Will permissioning of datasets containing sensitive or	The most granular level of security available would be desired.
	regulated information be required for the whole dataset	
	or on the field level?	
60	Do you currently have a clear set of defined use cases	Please see Florida House Bill 1391 (2020) for more information
	for the outcome/interface of the project and what types	on the legislative requirements of the state data catalog.
	of problems need to be solved or insights conveyed?	

		T
61	How do you envision this solution staying up to date on data policies and how to manage specific datasets or types of data? Should the solution be able to be configured by your user to manage this?	The solution must stay up to date on data policies and datasets metadata. Yes, the solution should be able to be configured by our users.
62	What types of permissions would each of the users have?	Depending on agency adoption, permissions will vary.
63	Do you envision all relevant users being onboarded at the same time? In phases?	Users will be onboarded in phases.
64	What are the requirements for visualization resulting from a combination of datasets?	Integration with multiple BI solutions, including Tableau, Power BI and Qlik.
65	What are the requirements for distribution of information resulting from modeling or visualization of the datasets?	Integration with multiple BI solutions, including Tableau, Power BI and Qlik.
66	Is the website for public consumption? Is there a component of the website that you would need to verify clearance for sensitive and regulated data?	Please see <u>Florida House Bill 1391 (2020)</u> for more information on the legislative requirements of the state data catalog. Open data for access by citizens with clean and anonymized data sets is included in future plans.
67	How do you currently collect and maintain metadata on your datasets? How do you currently leverage that information? Would that information be available for use during this implementation?	See answer to question 29.
68	What types of metadata are not currently associated with your datasets that you would like to have in the future?	We are seeking to consolidate disparate data resources from all state agencies including all associated metadata.
69	Why there is a need to implement EDC? What current and future challenges are you trying to solve with EDC?	Please see <u>Florida House Bill 1391 (2020)</u> for more information on the legislative requirements of the state data catalog.
70	What will be the consequences if EDC is not be implemented in the current/next year?	Please see Florida House Bill 1391 (2020) for more information on the legislative requirements of the state data catalog.
71	Are there any time constraints, milestones?	Please see Florida House Bill 1391 (2020) for more information on the legislative requirements of the state data catalog. It is the intention that the data cataloging starts in Fiscal Year 2020- 2021.

72	How you will measure the success of EDC implementation?	Successful implementation will have multiple complimentary benefits to all state agencies. Agency adoption and systemic
	•	data quality will be monitored by the FL CDO office.
73	What are the main criteria for choosing the solution	Please see Florida House Bill 1391 (2020) for more information
	(cost, functionality, support, etc.)?	on the legislative requirements of the state data catalog.
74	Are there any other data related projects ongoing? Are	There will always be many simultaneous data centric projects
	those delegated to other vendor(s)?	going on at the state level by numerous vendors.
75	Do you have any technical team to be part of the	Resources will be made available as needed.
	implementation/maintenance?	
76	Data from how many organizations are to be covered	Between 30 - 40 agencies.
	with the EDC? (rough number)	
77	What are the criteria for choosing the vendor? Specific	Hybrid or cloud-based data catalog as well as ease of adoption
	requirements like locations/onsite presence, etc.?	and implementation, at a minimum.
78	Do you already have a preferred	We are open to collaboration/contracting models that make
	collaboration/contracting model?	sense for the future state of the EDC.
79	Has your organization currently adopted remote work	In the current COVID-19 pandemic we are working with many of
	practices?	our staff and contractors/consultants in a remote working
	Who will be the and years of EDC2	capacity.
80	Who will be the end users of EDC?	Please see <u>Florida House Bill 1391 (2020)</u> for more information on the legislative requirements of the state data catalog. End
		users will be based on adoption by state agencies.
81	Which data sources is expected to be supported?	The Department is seeking a solution that will be able to work
01	o Rough number	with all data types as many are unknown (outside of RDBMS,
	o Type variety (RDBMS, NOSQL, particular file	structured and unstructured) at this time prior to collection by all
	extensions, cloud specific data sources, exotic field-	state agencies.
	specific data sources)	
	o Data variety (structured, unstructured	
82	Are there any existing data quality\profiling rules or	Depending on agency adoption, various levels of data quality
	instruments in place, which future implementation of	and documentation exist.
	EDC can rely on ?	
83	Are there any hosting preferences ?	Hybrid and cloud.

84	Can we clarify last point from GOAL section - is it about	Data governance and security as well as role management wil
	fine grained security for metadata access based on	be incorporated into the future state of EDC, dependent upor
	classification associated with entities ? e.g. who can	agency adoption.
	access data classified as PII	
85	What are the data volume requirements?	Depending on adoption by agencies, the number of users and
		the data volume could range widely.
86	Where are you planning on storing the actual data	Hybrid and cloud.
	(associated with the metadata)?	
87	How many data sources exist today?	Assume any/all data sources commonly used in governmen
		agencies, scale depends on adoption by agencies.
88	Is there currently a Data Warehouse or Data Lake in	Dependent upon agency adoption there will be multiple
	place? If not, is there a plan?	instances of both that need to be incorporated.
89	On what systems do they exist today (Mainframe, Db2,	Proposed EDC solution should be able to work with all available
	Oracle, SQL Server, etc.)	systems.
90	What type of data? Structured(transactional), semi-	The Department is seeking a solution that will be able to work
	structured XML, JSON, etc.) and un-structured (emails,	with all data types as many are unknown (outside of RDBMS
	audio, video, sensors, documents, texts, etc.) - all types	structured and unstructured) at this time prior to collection by al
		state agencies.
91	What source system(s) do you use today?	The Department is seeking a solution that will be able to work
		with all data source systems as many are unknown at this time
		prior to collection by all state agencies.
92	What monitoring tools are used today?	This information is not available.
93	Do product systems interact with (read/write) to tape for	This information is not available.
	log files, master files, etc?	
94	What tools are used for dev/test?	This information is not available.
95	What are the archival requirements?	Depending on adoption by agencies, the archival requirements
		could range widely as these are determined by the agency customer.
96	What are the names of applications to be migrated?	Depending on adoption by agencies, the number of users and
		their applications could range widely.

97	• Section IV Part B. 10 – Does the timeline need to include implementation of security and compliance controls?	The timeline should include all integral parts of the EDC solution.
98	 Section IV Part C. 1 – Some potential problems may arise from new builds vs integration of existing solutions, in order to better address benefits and problems what is the current state of EDC for the Department of Management Services at DST? 	Certain agencies have data catalogs, but the Department does not currently have an EDC solution in place.
99	How many data sources exist today?	Assume any/all data sources commonly used in government agencies; scale depends on adoption by agencies.
100	On what systems do they exist today (Mainframe, Db2, Oracle, SQL Server, etc.)	Oracle and SQL Server are known but multiple others may exist outside of these two sources for metadata.
101	Is there currently a Data Warehouse or Data Lake in place? If not, is there a plan?	Dependent upon agency adoption there will be multiple instances of both that need to be incorporated.
102	What type of data? Structured(transactional), semi- structured XML, JSON, etc.) and un-structured (emails, audio, video, sensors, documents, texts, etc.) - all types?	The Department is seeking a solution that will be able to work with all data types as many are unknown (outside of RDBMS, structured and unstructured) at this time prior to collection by all state agencies.
103	How many users are expected to access the platform	Depending on adoption by agencies, the number of users could range widely.
104	Application Information relevant to the Data Catalog below questions:	The Department does not understand the question.
105	"What are the names of applications to be migrated?	See response to question 96.
106	This will be determined together after further analysis."	The Department does not understand the question.
107	File counts (by type, including the number of database tables, stored procedures, database triggers)	Depending on adoption by agencies, the number of database tables, stored procedures and database triggers could range widely.
108	Will there be language codes feeding the Data Catalog (Cobol, Fortran, etc.)>	Depending on adoption by agencies, the language codes may vary widely.
109	What are the language codes?	Depending on adoption by agencies, the language codes may vary widely.

110	What is the estimated total of lines of (Cobol, Fortran,	Depending on adoption by agencies, the number of lines of code
	etc.) code that will be integrated?	may vary widely.
111	What is the estimated number of a) online and b) batch	Depending on adoption by agencies, the number of online and
	programs?	batch programs may vary widely.
112	What is the estimated total of lines of 4GL code? And	Depending on adoption by agencies, the total lines of 4GL code
	what type?	may vary widely.
113	What is the source Transaction System used, ex?	Depending on adoption by agencies, the source transaction
	(CICS, IMS, etc.)	systems may vary widely.
114	Does UI support Double Byte Characters? (Japanese	Yes.
	etc.)	
115	What data storage do apps use? (VSAM, DB2, IMS,	Oracle and SQL Server are known but multiple others may exist
	etc.)	outside of these two sources for metadata.
116	What is the estimated number of transactions during	Expectation is that the EDC will adjust automatically to scale
	your peak period?	needs.
117	What is the duration of your longest batch process?	Expectation is that the EDC will adjust automatically to scale
		needs.
118	What is your available batch window in hours?	This information is not available.
119	What is the job control language used?	Depending on adoption by agencies, the JCL language used
		may vary widely.
120	How many JCL job streams and total lines of JCL	Depending on adoption by agencies, the total number of JCL
	code?	job streams and total lines of JCL code may vary widely.
121	What security system does application use? (RACF,	Depending on adoption by agencies, the security system used
	TOP Secret, etc.)	may vary widely.
122	Are any special printing facilities utilized?	Depending on adoption by agencies, the printing facilities may
		vary widely.
123	What other 3rd party s/w is utilized by application or is	This information is not available.
	used to develop, test, or operate the application?	
	(SyncSort, MQSeries, FileAid, Omegamon, etc.)	
124	What are your current SLAs?	Depending on adoption by agencies, the service level
		agreements may vary widely.

125	Do the applications have related items that need to be migrated (3270, VPlus, Java, etc.), and if so, how many screens?	Depending on adoption by agencies, the number of users and their applications could range widely.
126	What is the maximum number of concurrent transactions?	Depending on adoption by agencies, the maximum number of concurrent transactions may vary widely.
127	What are the SLA's for the online transactions?	Depending on adoption by agencies, the SLAs may vary widely.
128	How many JCL job streams are there and the total number of steps per job stream?	See response to question 120.
129	What is the maximum concurrent of batch threads?	Depending on adoption by agencies, the maximum number of concurrent batch threads may vary widely.
130	Do you use tape information as part of the production cycle?	No.
131	Print Management – do you have requirements for online viewing, batch printing, unique forms, specialized hardware?	Reportability function is assumed in the future state of the EDC.
132	BI / ETL BREAKDOWN App-name # of Jobs / Reports / Mappings	Depending on adoption by agencies, the BI/ETL may vary widely.

133	 APPLICATION 1. Number of Applications 2. # and type of Current Screen/Display Type(s) 3. # and type of Current Data Interfaces 4. # and type of Current External Calls (Support Libraries, System Utilities, etc.) 5. # and type of Current Interfaces to other Systems 6. Target Language (Version and compiler) 7. Are Database and Programmer's Manuals Available? 8. Are Source Application Architecture and Design details available? 	Depending on adoption by agencies, the number of applications may vary widely. Documentation and resources will be provided as needed.
134	 Data Profile ("Mix") - % 1. Sensors 2. Transactions 3. Email/SMS 4. Documents 5. Audio/Video 	The Department is seeking a solution that will be able to work with all data types as many are unknown (outside of RDBMS, structured and unstructured) at this time prior to collection by all state agencies.
135	QUERIES - %1.Light queries2.Medium Queries3.Complex queries	Depending on adoption by agencies, the percentages by type of query may vary widely. Documentation and resources will be provided as needed.
136	System Information What is the model of your mainframe?	The Department does not believe this question is relevant to the RFI.
137	What are the OS and version?	The Department does not believe this question is relevant to the RFI.
138	Total MIPS rating	The Department does not believe this question is relevant to the RFI.
139	What are the MIPS under consideration to migrate?	Depending on adoption by agencies, the number of MIPS to be migrated may vary widely.

140	Details on each LPAR: a. Purpose (Production, Staging, Test, Backup) b. Size of LPAR, including memory, # channels?	Depending on adoption by agencies, the LPARs may vary widely.
	c. MIPS d. IOs per second or minute	
141	 § Average amount of data transferred per IO a. # Applications b. Which LPARS do you want to combine 	Depending on adoption by agencies, the data transferred may vary widely.
142	How many developers work on the application?	Depending on adoption by agencies, the number of developers may vary widely.
143	Do you have a backup installation (DR)?	Depending on adoption by agencies, the DR plans may vary widely.
144	Do you have a high-availability system?	Incorporation of high availability would be expected.
145	Do you offload information for a data warehouse?	Dependent upon agency adoption there may be multiple instances offloading information for a data warehouse that need to be incorporated.
146	What other systems interface with this application?	Depending on adoption by agencies, the system interfaces may vary widely.
147	Please enumerate by data type and by interface type."	The Department is seeking a solution that will be able to work with all data types as many are unknown (outside of RDBMS, structured and unstructured) at this time prior to collection by all state agencies.
148	Is there an existing DR plan, if so please define	Depending on adoption by agencies, the DR plans may vary widely.
149	Provide a summarized RMF report	This information is not available.
150	Within mainframe, are there any Linux partitions? If yes the details of it	The Department does not believe this question is relevant to the RFI.

151	 Size & Speed Requirements in Terabytes 1. Data Capacity (TBs) 2. Ingest Speed (TB/Day) 3. Queries per Day (million) 4. Data Retention (Number of Years) 5. Data in Retention (% Total) 6. # of Concurrent Users 	The state is seeking a solution that will be able to work with all size and speed types as many are unknown (outside of RDBMS, structured and unstructured) at this time prior to collection by all state agencies.
152	UAT, Cutover, Post Migration and Others What QA/SIT duration would you like to have?	This information is not available.
153	What UAT duration would you like to have?	This information is not available.
154	How much time would you like us to dedicate to post- production deployment support?	This information is not available.
155	Will you be providing Use Cases and Test Cases for Applications?	This information is not available.
156	Will you require Database Redesign, Modeling, or Re- modeling?	Depending on adoption by agencies, database modeling/remodeling may vary widely.
157	Will you require Backup / Restore and Disaster Recovery configuration and testing, including batch jobs, scripts, and scheduling?	The Department does not believe this question is relevant to the RFI.
158	Will you require a re-write of database management and monitoring scripts?	The Department does not believe this question is relevant to the RFI.
159	Will you require load and/or stress testing on upgraded databases?	The Department does not believe this question is relevant to the RFI.
160	Will you require performance and tuning of scripts, procedures, tools?	The Department does not believe this question is relevant to the RFI.
161	Will you require performance and tuning of reports, queries, or application code?	The Department does not believe this question is relevant to the RFI.

End of Questions and Answers