

CESSDA MTNA's RDS Webinar 2021-05-05

Questions and Answers

Description

This document was used for collecting questions asked before and during the webinar. It is intended to be used as a reference document for attendees and others.

Presenters:

Presenters (all Metadata Technology North America):

Pascal Heus - Vice President

Andrew DeCarlo - Lead Systems Developer

Carson Hunter - Developer

Document info

Status	Public
Author	John Shepherdson
Date	05/05/2021
License	CC BY 4.0
Version	2.0
PID	Not applicable

Version history

Version	Date	Comment	Revised by
0.1	27/04/2021	First version	John Shepherdson (CESSDA)
0.2	05/05/2021	Added Q&A text from the webinar	John Shepherdson (CESSDA)
0.3	05/05/2021	Reviewed and expanded Q&A from webinar	Pascal Heus (MTNA)
1.0	05/05/2021	Finalised for dissemination to attendees	John Shepherdson (CESSDA)
2.0	11/05/2021	Changed status to 'Public'	John Shepherdson (CESSDA)

Contents

Description	1
Document info	1
Version history	1
Questions asked	3
Classification of questions	3
Seeded questions	3
Questions asked during the event (via Questions window, 28 participants)	4

1. Questions asked

This document serves as a reference, so that attendees/non-attendees can view the questions asked during the webinar and see the answers given.

1.1. Classification of questions

Questions that relate to the generic aspects of RDS are marked as GEN.

Questions that relate to CESSDA-specific usage of RDS are marked as SPEC.

1.2. Seeded questions

Questions that were prepared in advance by the organisers, to make sure the presenters have something to respond to.

Question	Class
<p>Do you have a different RDS pricing model for commercial and non-commercial customers?</p> <p>A: No, as no two Use Cases are the same. The RDS website just gives an indicative price, not THE price. The price depends on the Customer's needs, resources, use cases, licensing model (e.g. subscription vs perpetual license), support needed from MTNA, infrastructure, and potentially other factors. MTNA's approach is to collaborate with a Customer and find a pricing model that is affordable for the Customer, and sustainable for MTNA, so we can keep strengthening and growing the platform for the greater benefits of all its users.</p>	GEN
<p>How can RDS be made to enforce/comply with the CESSDA Metadata Model?</p> <p>A: We actually do not feel it is necessarily the role of RDS to enforce this (it should be done at the metadata preparation stage, before pushing to RDS. We could however certainly add support for metadata profile if this is a hard requirement (e.g. add a validation before a data product can be made public or published).</p> <p>It's important to know that the RDS underlying metadata model is not DDI, but a more flexible model (more RDF like) that can be used to generically store various 'properties' around 'resources' (such as data products, variables, records, classifications/codes/categories). It's inspired by GSIM/DDI/SDMX and the likes, so we can be compliant with these standards, but are not constrained by them. So basically we map DDI elements to these internal model properties. Not all the elements you have in the CESSDA model are currently represented, but these could be added accordingly.</p> <p>We are currently also looking at this from a broader perspective, and are interested in support properties that are common to many standards, such as schema.org, DCAT, DataCite, and the likes.</p>	SPEC

1.3. Questions asked during the event (via Questions window, 28 participants)

Question	Class
<p>What are the input formats to RDS?</p> <p>A: From a data perspective, any queryable database. This includes most SQL servers, as well as platforms such as Google BigQuery or Socrata. If your data is not readily in the database, RDS comes with utilities to automatically load data from SAS, Stata, SPSS and other formats. And many options and software are available to load data into SQL.</p> <p>From a metadata perspective, we currently support JSON, and DDI-Codebook. Other specifications could be added. We do use a variety of tools and techniques to generate or capture metadata (such as DDI) if not available.</p>	GEN
<p>Are the original data migrated to RDS?</p> <p>A: Data is not migrated into RDS. RDS relies on your backend SQL server(s) to query and retrieve data. This provides a wide range of options in terms of infrastructure, licensing, performance/scalability. Data may be cached by RDS. RDS uses OpenLink Virtuoso to store the metadata.</p>	GEN
<p>Do you provide citation information for the data extract (the package) or the tabulations? Or some general guidance on how to cite data packages or tabulations?</p> <p>A: Citation can be stored as a property of the metadata. It would also be possible to generate citations in various formats by retrieving relevant metadata elements through the API, and compose the text.</p>	GEN
<p>Can an spss.sav file be downloaded?</p> <p>A: No. We prefer to generate a CSV+script approach which is open and not bound to proprietary formats. An SPSS user for example would only need to generate the SPSS syntax script to create such a .sav file. Other advantages of having a script is that it can be customized by the user.</p>	GEN
<p>Analysis - can more than crosstabs be done without downloading, eg., regression?</p> <p>A: Not regression, because SQL databases don't typically support it natively. So at this time, RDS supports common SQL operations such as count, sum, average, min/max. Some SQL servers can support regression or other statistical aggregation functions (some can also embed R or Python). If there is a solid need for this, we could look into supporting SQL server specific features.</p>	GEN

<p>Additional question: it is about the PID. The user need to know which data object it is working with</p> <p>A: At this time RDS does not enforce any restriction on data outputs. Variable-level metadata could be used to hide sensitive variables (custom use case), but at this point it would be easier to create a public/safe view of the data in the database, and use it as input for the RDS data product.</p> <p>For the tabulation service, we could integrate with a disclosure control engine (e.g. TauArgus). This is a feature we have on the roadmap, but will only be implemented if we have a use case or Customer need.</p>	GEN
<p>I suppose you have Unicode support. Is it working correctly? I am asking this because I have experience with tools that declare they have Unicode support, but if I try to extract metadata from SPSS or Stata, it actually doesn't work. Do you have any examples with non-English (meta)data?</p> <p>A: Yes, supported intrinsically as the platform is by nature UTF-8. When it comes to reading data files such as SPSS or Stata, we're pretty sure our reader would properly convert, but this might need further testing. Having some files to test with would be useful.</p>	GEN
<p>There is a difference between Dataverse and NESSTAR. The variable information is not available in Dataverse, how is this solved?</p> <p>A: This is not an RDS specific issue. If the metadata (e.g XML file) is available from somewhere, then can combine it. If it doesn't exist, then we have to rely on various metadata creation tools or manual capture. We do have various tools and techniques to address this.</p>	SPEC
<p>How does RDS get permission from the data providers prior to harvesting their data from e.g. Nesstar or Dataverse?</p> <p>A: This is a management, rather than technical issue. Not enforced in the software.</p>	SPEC
<p>Is there also a search function, that you can search for a variable in multiple datasets?</p> <p>A: Yes, you can search for variables and codes within data products. We are also further enhancing our search services.</p>	GEN