Reuse Readiness Assessment of Data Quality Software Products

Status of this Memo

This memo is a Technical Note that provides information to the NASA Earth Science Data Systems (ESDS) community about the assessment of the reuse readiness of data quality software products that were identified by the Earth Science Data System Working Groups (ESDSWG) on Data Quality (DQWG). This memo does not specify an ESDS standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

Abstract

Recognizing the need to evaluate data quality software products that were identified by the NASA Earth Science Data System Working Group (ESDSWG) on Data Quality (DQWG), a reuse readiness assessment was conducted to provide guidance for potential adopters of the identified data quality software products. The assessment found significant variation in the readiness, for potential reuse and adoption, among the identified software products.

Table of Contents

1 Introduction

Software can be used throughout the data lifecycle to improve the quality of data products and services. Software products for improving the quality of Earth science data that had been identified by the NASA ESDSWG DQWG, during 2016, were assessed during 2017 and 2018 to determine their readiness as software for potential reuse by the Earth science community. The reusability assessment of the identified data quality software products was conducted by three members of the DQWG using the Reuse Readiness Levels (RRL) instrument that had been developed previously by the ESDSWG on Software Reuse [1].

2 RFC Content

During 2015-2017, the DQWG identified and created a "master list" of 26 solutions for improving the quality of Earth science data products and services [2]. The process for identifying the data quality solutions is described in [3]. Recognizing the value of providing additional information for potential adopters to determine the viability of implementing any of the identified solutions for improving data quality and to guide their decision-making regarding adoption of a particular solution, the DQWG conducted an assessment of the solutions that were identified as software products. The RRL instrument was selected for conducting the assessment due to the availability of the RRL as an open access resource that contained the assessment criteria for evaluating the potential reusability of software products. Members of the DQWG were familiar with the RRLs, including the assessment criteria, since the RRL had been developed by the NASA ESDSWG on Software Reuse to enable the Earth science community, as well as others, to assess the reuse readiness of software products.

Prior to using the RRL to assess the potential for the Earth science community to adopt the solutions for data quality that had been identified by the DQWG, each identified solution was reviewed to determine whether it was within the scope of the planned reuse readiness assessment. The review of the identified data quality solutions categorized each of the solutions in terms of their functionality and identified whether each solution was a software product or some other type of resource. Fourteen of the identified data quality solutions were determined to be software products and were selected for the reuse readiness assessment. The categories of these software products and the number of software products within each category are presented in Table 1. The identified data quality solutions that were determined to be non-software resources were excluded from the assessment. These non-software resources represented two categories – committees (2 counts) and guidance documents (10 counts).

Table 1. Data Quality Software Solutions: Software Product Categories and Counts

Software Product Categories	Relationship of this Category of Software to Data Quality	Number of Software Products
Data Analyzer	Enables inspection of data values	1

Data Filter	Filters data values based on quality flags	1
Data Convertor	Migrates data between formats	1
Helpdesk Software	Allows data users to better understand quality and report quality concerns	2
Metadata Checker	Inspects the quality of metadata	1
Metadata Convertor	Migrates metadata between schemas	1
Metadata Editor	Enables collection and update of metadata	5
Quality Search Service	Enables finding data matching specific criteria	1
User Forum Service	Enables users to review data quality issues	1

2.1 Reuse Readiness Assessment of Data Quality Software Products

The reuse readiness assessment was conducted on fourteen data quality software products, which are listed in Table 2.

Table 2. Data Quality Software Products and Product Categories Selected for Assessment

Data Quality Software Products	Product Categories
Collaboratory for quAlity Metadata Preservation (CAMP) (2017 version)	Metadata Editor
Advanced Tracking and Resource tool for Archive Collections, Rel 3.2	Metadata Editor
DAAC-Ingest Dashboard	Metadata Editor
Ocean CO_2 Metadata Collection Form	Metadata Editor
EUFAR Metadata Creator, 2016	Metadata Editor
NCO version 4.6.8	Data Analyzer
Kayako *	Help Desk Software
UserVoice *	Help Desk Software
Data Quality Screening Service	Data Checker
Metadata Compliance Checker	Metadata Checker
Metadata XML Conversion Script	Metadata Convertor
PO.DAAC Forums	User Forum Service
Virtual Quality Screening Service	Quality Search Service
ArcGIS MODIS Python Toolbox, V1.0 *	Data Convertor

* Products marked with an asterisk require purchase of a subscription or license.

The authors conducted the reuse readiness assessment of the data quality software products by following a Reuse Readiness Assessment Procedure that was prepared and reviewed by the authors prior to beginning the reuse readiness assessment. The Reuse Readiness Assessment Procedure is summarized in Appendix B. The locations of the software products are listed in the Solutions Master List [2]. The reuse readiness assessment of the data quality software products was conducted by assessing each software product using the criteria of the nine topic areas of the RRLs [1], which are listed in Table 3.

Table 3. Topic Areas of the Reuse Readiness Levels

(Source: NASA Earth Science Data Systems Software Reuse Working Group, 2010 [1])

RRL Topic Areas
Documentation
Extensibility
Intellectual Property Issues
Modularity
Packaging
Portability
Standards Compliance
Support
Verification and Testing

Each of the topic areas of the RRLs contains up to nine levels that were developed as criteria for assessing the reusability of a particular software product. Each topic area level was defined so that the definition for a particular topic area can be used to assess whether the software product meets the criteria for that level [1]. The definition for each of the topic area levels generally corresponds to the general RRL summary for that level. The general summaries of the nine RRLs are presented in Table 4.

Table 4. General Summaries of the Reuse Readiness Levels

(Source: NASA Earth Science Data Systems Software Reuse Working Group, 2010 [1])

RRL	RRL Summary

RRL 1	Limited reusability; the software is not recommended for reuse.
RRL 2	Initial reusability; software reuse is not practical.
RRL 3	Basic reusability; the software might be reusable by skilled users at substantial effort, cost, and risk.
RRL 4	Reuse is possible; the software might be reused by most users with some effort, cost, and risk.
RRL 5	Reuse is practical; the software could be reused by most users with reasonable cost and risk.
RRL 6	Software is reusable; the software can be reused by most users although there may be some cost and risk.
RRL 7	Software is highly reusable; the software can be reused by most users with minimum cost and risk.
RRL 8	Demonstrated local reusability; the software has been reused by multiple users.
RRL 9	Demonstrated extensive reusability; the software is being reused by many classes of users over a wide range of systems.

Each data quality software product was initially reviewed by one of the authors to determine the possible level of the software product for each topic area of the RRLs. During the initial assessment of one data product, the Ocean CO_2 Metadata Collection Form, the product was found to be no longer accessible online and that product was excluded from the assessment exercise. Following the initial assessment of each of the data quality software products, a second review was conducted by another author to determine whether the initial assessment for each topic area was correct or whether an additional review of the topic area level was needed to address a particular issue that was identified. During the subsequent review of a data quality software product, any issue that was identified as having the potential to affect the assessed topic area level for the product was described to enable a successive review and assessment of the software product in light of the described issues. For each described issue that could affect the initial assessment of the topic area level for a particular data quality software product, a re-assessment of the product was conducted and subsequently described by the reviewing authors to attain consensus on the topic area level that was being assessed. Upon attaining consensus on all of the topic area levels assessed for a particular software product, the topic area level scores for the product were averaged to determine the overall RRL score for the product.

2.2 Results of the Reuse Readiness Assessment of Data Quality Software Solutions

The reuse readiness assessment of the data quality software product revealed considerable variation among the scores of the assessed software products. The scores range from 1.7 to 9.0, where the higher scores indicate that a particular data quality software product has been assessed as being more ready for reuse. One software product received a perfect score of 9.0, indicating that the product could be implemented by an adopter if the capabilities of the software product address the needs for software adoption. The RRL scores for each of the assessed data quality software products are presented in Table 5. The scores for each topic area of the assessed software products are listed in Appendix C.

Table 5. Reuse Readiness Levels (RRLs) of Assessed Data Quality Software Products

Data Quality Software Product	Product Category	RRL Score
Collaboratory for quAlity Metadata Preservation (CAMP) (2017 version)	Metadata Editor	1.9
Advanced Tracking and Resource tool for Archive Collections, Rel 3.2	Metadata Editor	2.8
DAAC-Ingest Dashboard	Metadata Editor	3.6
EUFAR Metadata Creator, 2016	Metadata Editor	4.2
NCO version 4.6.8	Data Analyzer	8.8
Kayako	Help Desk Software	3.6
UserVoice (Hosted solution, version # not available externally)	Help Desk Software	4.4
Data Quality Screening Service	Data Filter	4.8
Metadata Compliance Checker	Metadata Checker	2.0
Metadata XML Conversion Script	Metadata Convertor	4.9
PO.DAAC Forums	User Forum Service	9.0
Virtual Quality Screening Service	Quality Search Service	1.7

ArcGIS MODIS Python Toolbox, V1.0	Data Convertor	3.9

3 Discussion

Readers should note that the assessment of the identified data quality software solutions only examined the software products in terms of the criteria of the RRLs, based on the accessible on-line documentation. The reuse readiness assessment did not address the functionality of the software that was examined. Prior to adoption, potential adopters should consider testing the use of the software to determine the potential of the candidate software for meeting the needs for which the software is being considered for adoption. Software developers also could use the RRLs to assess the potential reusability of software, prior to releasing such software. In addition, software developers should ensure that assessors and potential users will be able to determine the potential usability of the software in terms of how the software meets the criteria of the RRLs. Subsequent reviews of the software products that were assessed for this study could offer additional insight into the potential reusability of these products and should cite this study so that such reviews can be compared. Future reviewers of the assessed software products could contact ESDIS through the Solutions Master List to submit comments.

4 Summary

The RRL assessment of the identified data quality software solutions offers insight into some of the implementation issues that should be considered when planning to adopt software products and, in particular, the software products that were examined as part of this assessment. The assessment also offers insight into the use of the RRLs for assessing the potential reusability of software products that are being considered for adoption, as well as a guide for software developers to produce reusable software.

5 Acknowledgements

The reported work was supported by NASA. Robert R. Downs was supported by NASA under Contract 80GSFC18C0111 for the Socioeconomic Data and Applications Center (SEDAC) Distributed Active Archive Center (DAAC). Hampapuram K. Ramapriyan was supported by NASA under Contract NNG15HQ01C with Science Systems and Applications, Inc. Yaxing Wei was supported by NASA under Contract NNG14HH39I for the Oak Ridge National Laboratory (ORNL) DAAC.

6 References

[1] NASA Earth Science Data Systems Software Reuse Working Group (2010). *Reuse Readiness Levels (RRLs), Version 1.0*. April 30, 2010. Available: https://earthdata.nasa.gov/files/RRLs_v1.0.pdf

[2] NASA Earth Science Data System Data Quality Working Group. (2018). *Solutions Master List for Earth Science Data Quality*. https://wiki.earthdata.nasa.gov/x/2pASBg

[3] NASA Earth Science Data Quality Working Group (2019). *Comprehensive Data Quality Recommendations for Data Producers and Distributors, ESDS-RFC-033*. https://earthdata.nasa.gov/user-resources/standards-and-references/recommendations-from-the-data-quality-working-group

[4] Downs, R. R., Ramapriyan, H. K., Wei, Y. (2018, December 11). *A Reusability Assessment of Recommended Software Solutions for Improving the Quality of Earth Science Data Products and Services*. eLightning paper (IN21B-05) presented at the American Geophysical Union (AGU) 2018 Fall Meeting. Session: IN21B: Free and Open-Source Technologies for Advancing Earth and Space Sciences I eLightning. Washington, DC, 10-14 Dec 2018.

7 Authors' Addresses

Robert R. Downs, Center for International Earth Science Information Network (CIESIN), Columbia University, Palisades, New York USA 10964, Tel: 845-365-8985, email: rdowns@ciesin.columbia.edu

Hampapuram K. Ramapriyan, Science Systems and Applications, Inc., Hampapuram.Ramapriyan@ssaihq.com

Yaxing Wei, Oak Ridge National Laboratory DAAC, weiy@ornl.gov

8 Appendix A

Glossary of acronyms

The glossary is required. All acronyms used in the RFC should be included.

Example

Acronym	*Description*
CAMP	Collaboratory for quAlity Metadata Preservation
CIESIN	Center for International Earth Science Information Network
DAAC	Distributed Active Archive Center
DQWG	Data Quality Working Group
ESDS	Earth Science Data System
ESDSWG	Earth Science Data System Working Groups
EUFAR	EUropean Facility for Airborne Research
MODIS	Moderate Resolution Imaging Spectroradiometer
NCO	NetCDF Operators
ORNL	Oak Ridge National Laboratory
PO.DAAC	Physical Oceanography DAAC
RRL	Reuse Readiness Levels
SEDAC	Socioeconomic Data and Applications Center

9 Appendix B

Summary of Reuse Readiness Assessment Procedure*

Each assessor should use the Reuse Readiness Levels to complete multiple software assessments by following the steps below, using the assessment spreadsheet.

1. For each topic area, start with level 1 and determine the highest level at which all the specifications are met by the software. Assign that level as the score for that topic area. (If level 1 specifications are not met, assign a score of 1).
2. Enter the score for each topic area in the designated column within the row of the software being assessed.
3. Describe any assumptions or difficulties with the assignment of topic area level scores in the Notes column.
4. Modify the Software Name and Version column, Software Category column, or the URL column if the assessment of the software indicates that these attributes have been updated.
5. Review assessments completed by another assessor and use the Comments column to indicate agreement, to respond to notes or to recommend a change to the assigned scores in one or more topic areas. The purpose of this step is to ensure that the assessment of each software product is reviewed by at least one other assessor.
6. Review all the comments that pertain to the software reviewed and subsequently review and revise any Topic Area Levels previously assigned, if necessary .
7. Download a copy of the assessment spreadsheet for your own records to verify your work later, if necessary.
8. Enter suggestions for improving this procedure, at the bottom of the spreadsheet.

* Source: Reference [4]

10 Appendix C

Topic Area Scores for Assessed Software Products

Data Quality Software Product	Documentation	Extensibility	Intellectual Property Issues	Modularity	Packaging	Portability	Standards Compliance	Support	Verification and Testing	RRL Score
Collaboratory for quAlity Metadata Preservation (CAMP) (2017, V?)	1	1	1	1	1	2	1	3	6	1.9
Advanced Tracking and Resource tool for Archive Collections, Rel 3.2	3	3	1	3	1	2	3	3	6	2.8
DAAC-Ingest Dashboard (V?)	3	3	1	3	3	5	4	2	8	3.6
EUFAR Metadata Creator, 2016	2	2	9	5	3	3	4	4	6	4.2
NCO V 4.6.8	9	9	9	9	9	7	9	9	9	8.8
Kayako (V?)	1	1	5	3	1	1	3	8	9	3.6
UserVoice (V?)	1	6	6	1	1	2	5	9	9	4.4
Data Quality Screening Service (V?)	3	5	9	7	5	5	3	2	4	4.8
Metadata Compliance Checker (V?)	1	1	1	3	1	1	1	6	3	2
Metadata XML Conversion Script (V?)	3	3	9	9	1	4	8	4	3	4.9
PODAAC Forums (V?)	9	9	9	9	9	9	9	9	9	9
Virtual Quality Screening Service (V?)	1	1	1	1	1	1	1	6	2	1.7
ArcGIS MODIS Python Toolbox, V1.0	5	3	1	7	3	4	3	4	5	3.9

Note that version numbers were not available for the software products listed without version numbers and that the assessed software products were last accessed online on March 9, 2018.