



Dataset-JSON Pilot

Dataset-JSON as an alternative transport format for regulatory submissions

Pilot Instructions

Sam Hume, Stuart Malcom, & Jesse Anderson

Leads, Dataset-JSON Pilot





Pilot Resources

Dataset-JSON Schemas, Examples, and Specification

- Dataset-JSON example datasets:
 - <https://github.com/cdisc-org/DataExchange-DatasetJson/tree/master/examples>
- Dataset-JSON schema file:
 - <https://github.com/cdisc-org/DataExchange-DatasetJson/tree/master/schema>
- Dataset-JSON standard repository:
 - <https://github.com/cdisc-org/DataExchange-DatasetJson>
- The Dataset-JSON specification
 - <https://www.cdisc.org/dataset-json>
 - <https://wiki.cdisc.org/display/PUB/Dataset-JSON>



Conversion Software Tools

- SAS

- <https://github.com/lexjansen/dataset-json-sas> ([submit issues](#))
- The SAS conversion software by Lex Jansen
- Dataset-JSON example files are included in the repository
- Includes a macro for comparing libraries with SAS datasets
- Includes a Python script for validating Dataset-JSON
- Documentation is included

- R

- <https://github.com/atorus-research/datasetjson> ([submit issues](#))
- <https://atorus-research.github.io/datasetjson/index.html>
- <https://cran.r-project.org/web/packages/datasetjson/index.html>
- R conversion package by Atorus Research and Johnson & Johnson
- Documentation is included

- Python

- <https://github.com/dostiep/Dataset-JSON-Python> ([submit issues](#))
- Python conversion software by Pierre Dostie
- We will not cover the Python tooling in the workshop
- The Dataset-JSON Pilot will focus on SAS and R, but any conversion tool can be used



Dataset-JSON Pilot Questionnaire

- After completing the pilot testing, complete a short on-line questionnaire:
 - <https://forms.office.com/r/SjQ0zfhP0s>
- Review the questions in the questionnaire prior to running the tests:
 - Review the questionnaire questions in the next 3 slides

Dataset-JSON Pilot Questionnaire: Questions 1-7

Dataset-JSON Pilot Questionnaire

* Required

Respondent

1. Company *

CDISC

2. Lastname *

Hume

3. Firstname *

Sam

4. Email *

shume@cdisc.org

Next

Pilot Findings

5. CONVERSION TOOLS - Which software conversion tools did you use to convert datasets to and from Dataset-JSON? Select all that apply.

☐ SAS

☐ R

☐ Both

☐ Other

6. CONVERSION TOOLS - Use comments to expand on your software conversion tools response in Question 5 above.

Enter your answer

7. SUCCESS - Were you able to convert existing datasets to Dataset-JSON?

☐ Yes

☐ No

☐ Not Sure

Dataset-JSON Pilot Questionnaire: Questions 8-15

8. SUCCESS - Use comments to expand on your conversion success response in Question 7 above.

Enter your answer

9. FORMATS - Were you able to convert Dataset-JSON datasets into SAS datasets or R dataframes or other dataset formats?

- ☐ Yes
- ☐ No
- ☐ Not Sure

10. FORMATS - Use comments to expand on your dataset formats response in Question 9 above.

Enter your answer

11. EXPECTATIONS - Did the conversions to and from Dataset-JSON work as expected?

- ☐ Yes
- ☐ No
- ☐ Not Sure

12. EXPECTATIONS - Use comments to expand on your expectations response in Question 11 above.

Enter your answer

13. CONTENT - What types of datasets content did convert? Select all that apply.

- ☐ SEND
- ☐ SDTM
- ☐ ADaM
- ☐ Other

14. CONTENT - Use comments to expand on your content response in Question 13 above.

Enter your answer

15. ACCURACY - Based on your conversion testing, did Dataset-JSON accurately represent your data?

- ☐ Yes
- ☐ No
- ☐ Not Sure

Dataset-JSON Pilot Questionnaire: Questions 16-19

16. ACCURACY - Use comments to expand on your accuracy response in Question 15 above.

Enter your answer

17. Please provide any feedback or comments on your experience testing Dataset-JSON.

Enter your answer

18. IMPACT - Moving forward, what level of impact is needed to change your existing workflow to use Dataset-JSON instead of SAS V5 XPORT?

- ☐ High
- ☐ Medium
- ☐ Low

19. IMPACT - Use comments to expand on your impact response in Question 18 above.

Enter your answer

Back

Submit

- There are 8 unique questions on the questionnaire
 - Comment text boxes exist for most questions
 - Contact information is requested (questions 1-4) for possible follow-up questions
- The form is on-line and should be completed by each group that tests Dataset-JSON
- Anonymized results will be published in the pilot final report



Pilot Process

What testing needs to be conducted to complete the pilot?



High-level Process Overview

1. Select the software conversion tool(s) to use (one or more)
 - Select a tool(s) listed on slide 4, *Conversion Software Tools*
 - Or use a tool of your own choosing
2. Select study datasets and convert them to Dataset-JSON
 - Convert datasets from one of your own studies into Dataset-JSON
3. Convert Dataset-JSON datasets into a native dataset form
4. Compare the native datasets back to the original format and ensure there are no unexpected changes in converted dataset
5. Optionally, test other conversion scenarios (e.g., interoperability tests)
6. Complete the on-line questionnaire to record your findings
<https://forms.office.com/r/SjQ0zfhP0s>

Example Conversion Scenarios

Pilot Conversion Scenarios

1. SAS dataset -> Dataset-JSON -> SAS dataset
2. R dataframe -> Dataset-JSON -> R dataframe
3. Dataset-JSON -> native format -> Dataset-JSON

Complete at least 1 scenario from the list above or use your own scenario

Example Interoperability Scenarios

Interoperability Conversion Scenarios

1. SAS dataset -> Dataset-JSON -> R dataframe
2. R dataframe -> Dataset-JSON -> SAS dataset

Interoperability scenarios are optional and may be tested in addition to the Pilot Conversion Scenarios. Improved interoperability is a benefit of Dataset-JSON.

Pilot Testing Questions and Timelines

- For questions about the pilot, please add your question to the Q&A discussions section in the following repository:
 - <https://github.com/cdisc-org/DataExchange-DatasetJson-Tests/discussions>
- For issues or technical questions about the open-source software conversion tools, please create an issue within that tool's repository
 - SAS: <https://github.com/lexjansen/dataset-json-sas> (submit issues)
 - R: <https://github.com/atorus-research/datasetjson> (submit issues)
- Timeline:
 - All Pilot Questionnaires should be completed by **Dec. 15th, 2023**
 - Initial results presented at the 2024 PHUSE US Connect (February 25-28)
 - Final Report published by the 2024 PHUSE CSS Conference (June 3-5)

A decorative vertical bar on the left side of the slide, featuring a grid of dots in red, yellow, and blue, connected by thin lines to form a complex geometric pattern.

Additional Resources

Resources

- Dataset-JSON specification
 - <https://www.cdisc.org/dataset-json>
 - <https://wiki.cdisc.org/display/PUB/Dataset-JSON>
- Dataset-JSON GitHub repository
 - <https://github.com/cdisc-org/DataExchange-DatasetJson>
- Dataset-JSON ODM v2.0 specification
 - <https://wiki.cdisc.org/display/PUB/ODM+v2.0>
- COSA Directory Dataset-JSON Hackathon I projects
 - <https://cosa.cdisc.org/hackathons/datasetJson>
- 2022 Working with Dataset-JSON using SAS (Lex Jansen)
 - https://www.lexjansen.com/cgi-bin/xsl_transform.php?x=pharmasug2022#pharmasug2022.ad150
- Dataset-JSON schema validation in Python Jupyter Notebook:
 - <https://github.com/swhume/dataset-json-hackathon>
- JSON Standard
 - <https://www.json.org/json-en.html>