

Dataset-JSON Viewer Hackathon Kick-off

Sam Hume, CDISC

Charles Shadle, CDISC

2024-10-08





Agenda

1. Introduction
2. Objectives
3. Judging Criteria and Rules
4. Timeline
5. Resources
6. Communications
7. Questions



Introduction

Introducing the COSA Dataset-JSON Viewer Hackathon

Welcome to the COSA Dataset-JSON Viewer Hackathon

~ 200
registered
participants

~ 12 weeks in
duration

Report
outcomes Q1
COSA Spotlight

Results
presentation at
EU Interchange

Viewers are a
Must Have
requirement

A competitive
hackathon



CDISC Open-Source Alliance (COSA)

COSA Mission: The CDISC Open-Source Alliance (COSA) supports, promotes, and sometimes sponsors open-source and free software development projects that create tools for implementing or developing CDISC standards to drive innovation in the CDISC community.

- COSA has hosted several hackathons
 - This is the 3rd Dataset-JSON Hackathon
- Dataset-JSON Hackathon solutions may apply to be included in the COSA Repository Directory
 - Projects from the first Dataset-JSON Hackathon: <https://cosa.cdisc.org/hackathons/datasetJson>
- Requires an open-source license
- Requires a public repository
- There will be opportunities to demonstrate your viewer software
 - COSA Spotlight Webinar to demo solutions



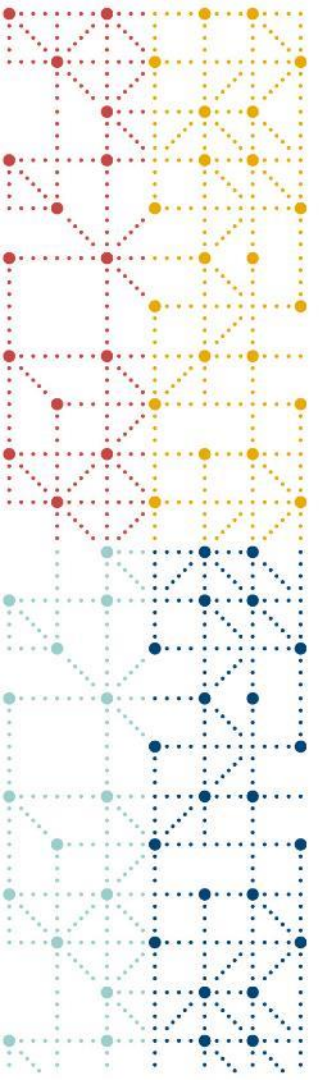
Objectives

The Hackathon Objectives



COSA Dataset-JSON Viewer Hackathon

- Primary objective: Create Dataset-JSON Viewer software
 - Projects can be new or extensions to existing software
 - Projects must support Dataset-JSON v1.1
- Significant and broad interest in viewer software for Dataset-JSON
 - A finding from the *Dataset-JSON as an Alternative Transport Format for Regulatory Submissions Pilot*
- Virtual hackathon
 - Individuals or teams will work virtually
 - Periodic meetings will be scheduled to discuss progress and answer questions
 - Project demonstrations will be scheduled for a COSA Spotlight Webinar
- Dates: Oct. 8th – Jan. 6th



Judging Criteria and Rules

The Hackathon judging criteria and rules that each project should follow and how each project will be evaluated by the judges



Hackathon Judging Criteria

- Usability
 - Overall ease of use
 - Ease of getting started
- Functionality
 - Key features
 - Basic browsing, filtering, and sorting
 - Extended features
- Deployment
 - Ease of installation
 - Deployment flexibility
 - Ease of access
- Performance
 - Responsiveness
 - Ability to handle large datasets
- Documentation
 - Documentation of viewer features
 - Instructions on using the viewer
 - Changes from baseline
- Bonus
 - NDJSON support
 - Support for the final version of Dataset-JSON v1.1 (include changes made after Public Review)
 - Design (appearance)



Hackathon Rules and Guidelines

- The project must be available in a public GitHub repository (or equivalent)
- All judged entries must be released under an open-source license
- We must be able to run the viewer in order to judge it
 - If the project is cloud-based, please provide access to a hosted version for judging
 - If the project is locally deployed, please provide installation instructions
- Provide a README.md that highlights the viewer's features and provides instructions for using the Viewer
- At a minimum, it must support the Public Review version of Dataset-JSON v1.1
- Support for Dataset-JSON v1.0 is not expected and will not be evaluated.
- If updating an existing Viewer, please report what was added or improved as part of the hackathon



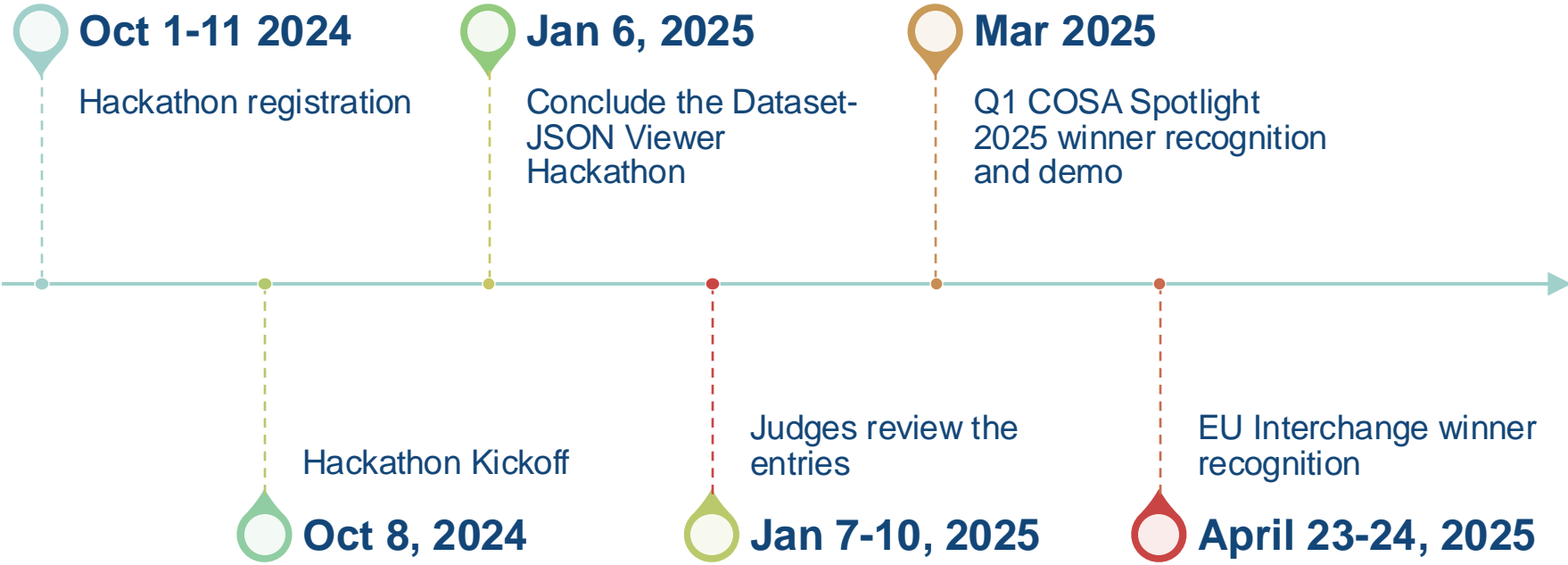
Competitive Hackathon

- 3 winners will be selected
- Winners will receive prizes
 - Swag (e.g. t-shirt)
 - Certificate
 - Bragging rights
 - Q1 COSA Spotlight recognition and demonstration + additional promotions and demos
 - Possibly some additional sweeteners
- The COSA Governance Board will serve as judges
 - Board members that are participating in the hackathon may not be a judge
- All participants will get
 - Feedback from the judges
 - Basic swag (e.g. sticker)
 - Opportunity to demo their product and become part of COSA



Timeline

Dataset-JSON Viewer Hackathon Timeline





Public Review

- Public Review starts on September 10, 2024
- Public Review lasts 30 days and closes on October 10, 2024
 - Some parties have asked for extra time
- The JIRA project is Dataset-JSON review comments (DSJSONCT) located at:
 - <https://jira.cdisc.org/projects/DSJSONCT>
- As of today, no major changes expected from the Public Review comments
 - This could change based on new comments coming in
- If minor changes are implemented, ideally Hackathon projects will update their viewers to support the final version of Dataset-JSON v1.1
 - If there are changes in Dataset-JSON v1.1 that you do not implement, make sure you make test datasets available and let the judges know when you plan to update the project to support the final v1.1 release



Resources

Links to resources that may be helpful for your Hackathon project



Dataset-JSON v1.1 Public Review Content

1. Dataset-JSON v1.1 Specification

- <https://wiki.cdisc.org/display/DSJSON1DOT1/Dataset-JSON+1.1>

2. User Guide

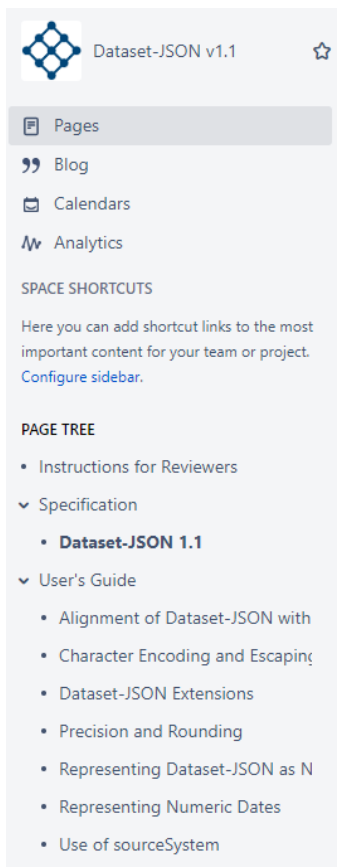
- <https://wiki.cdisc.org/display/DSJSON1DOT1/User%27s+Guide>

3. Schemas and Example Datasets

- <https://github.com/cdisc-org/DataExchange-DatasetJson>

CDISC Wiki

Contains the Dataset-JSON v1.1 Specification and a User's Guide with articles



The screenshot shows the sidebar of the CDISC Wiki page for Dataset-JSON v1.1. At the top is the CDISC logo and the page title 'Dataset-JSON v1.1' with a star icon. Below are navigation options: Pages, Blog, Calendars, and Analytics. A 'SPACE SHORTCUTS' section explains that users can add shortcut links to important content. The 'PAGE TREE' section lists the following items:

- Instructions for Reviewers
- ▼ Specification
 - **Dataset-JSON 1.1**
- ▼ User's Guide
 - Alignment of Dataset-JSON with
 - Character Encoding and Escaping
 - Dataset-JSON Extensions
 - Precision and Rounding
 - Representing Dataset-JSON as N
 - Representing Numeric Dates
 - Use of sourceSystem

Pages / Dataset-JSON v1.1 / Specification     Analytics

Dataset-JSON 1.1

Created by Omar Garcia Calderon, last modified by Lex Jansen on Aug 28, 2024

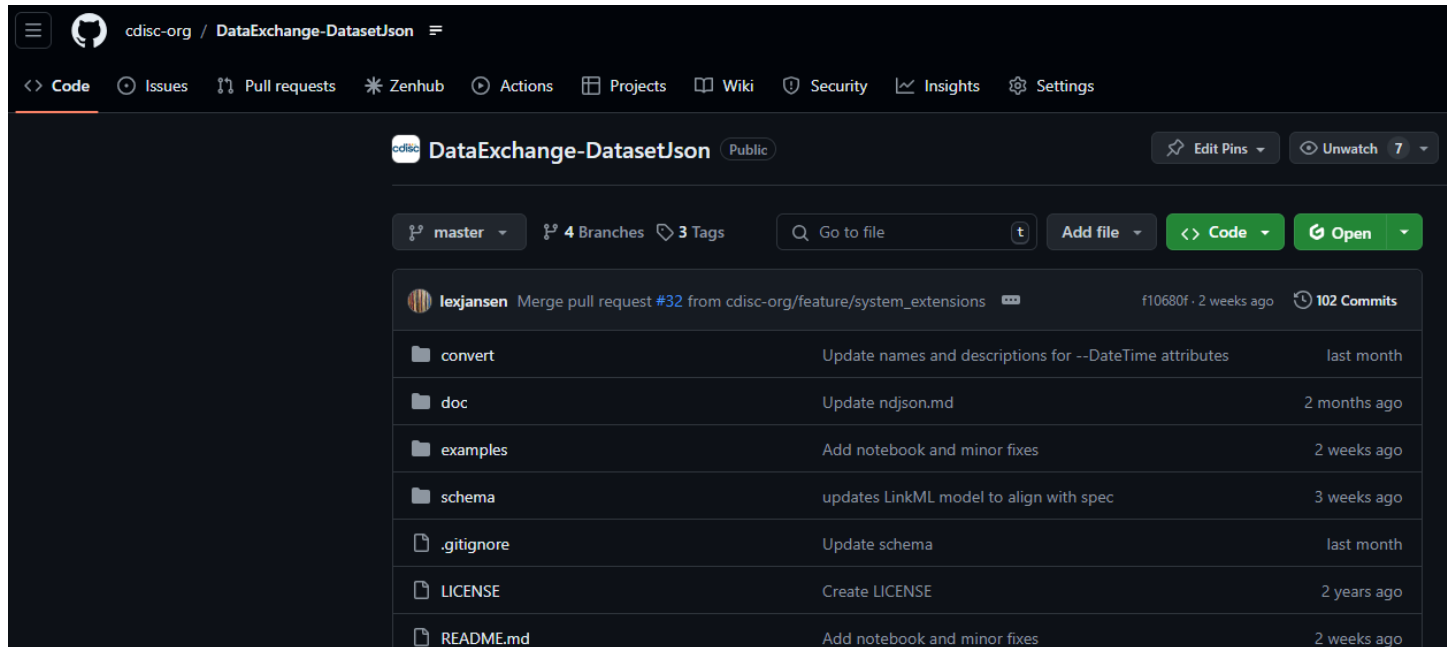
DRAFT

Title	CDISC Dataset-JSON Specification		
Version	1.1		
Prepared by	CDISC Data Exchange Standards Team		
Notes to Readers	<ul style="list-style-type: none">• This is the specification for Version 1.1 of CDISC Dataset-JSON.		
Revision History	Date	Version	Summary of Changes
	2024-09-06	1.1	Draft
	2023-08-23	1.0	Final

- [Introduction](#)
- [Top-level Metadata Attributes](#)
- [Column Metadata](#)
 - [Supported Column Data Type Combinations](#)
 - [Date/Time Variables](#)
 - [Decimal Variables](#)
- [Row Data](#)
- [A Full Example of a Dataset-JSON File](#)
- [NDJSON Representation of Dataset-JSON](#)
- [A Full Example of an NDJSON Dataset-JSON File](#)

GitHub: JSON Schema and Examples

- The JSON schema, both for JSON and NDJSON representations, and examples can be found at the GitHub repository for the Dataset-JSON Version 1.1 standard:
<https://github.com/cdisc-org/DataExchange-DatasetJson>



The screenshot shows the GitHub repository page for `cdisc-org / DataExchange-DatasetJson`. The repository is public and has 4 branches and 3 tags. The main branch is `master`. The repository contains several files and folders, including `convert`, `doc`, `examples`, `schema`, `.gitignore`, `LICENSE`, and `README.md`. The most recent commit is by `lexjansen`, titled "Merge pull request #32 from cdisc-org/feature/system_extensions", which updates names and descriptions for `--DateTime` attributes. The repository also has 102 commits in total.

File/Folder	Description	Last Commit
<code>convert</code>	Update names and descriptions for <code>--DateTime</code> attributes	last month
<code>doc</code>	Update <code>ndjson.md</code>	2 months ago
<code>examples</code>	Add notebook and minor fixes	2 weeks ago
<code>schema</code>	updates LinkML model to align with spec	3 weeks ago
<code>.gitignore</code>	Update schema	last month
<code>LICENSE</code>	Create LICENSE	2 years ago
<code>README.md</code>	Add notebook and minor fixes	2 weeks ago



Software Conversion Tools



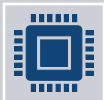
SAS

- [The SAS conversion software by Lex Jansen](#)
- Includes a macro for comparing libraries with SAS datasets
- Documentation is included



R

- [R conversion package by Atorus Research and Johnson & Johnson](#)
- Documentation is included



Python

- Multiple Python conversion software tools
- Documentation is included
- Covers multiple dataset formats, including Parquet and SAS

COSA Directory Dataset-JSON Hackathon I projects
<https://cosa.cdisc.org/hackathons/datasetJson>



Communications

Tools for communications during the Hackathon



Hackathon Communication Channels

- To receive communications and meeting notices, please register:
 - <https://www.cdisc.org/events/webinar/dataset-json-viewer-hackathon>
- Meeting date, time, cadence:
 - Meet-ups every other week (attendance is optional)
 - Date/time of meetings?
- Use existing Dataset-JSON Hackathon Slack workspace
 - Slack is the best way to have an ongoing dialog with the Dataset-JSON Hackathon community
 - Will send out invites to participants not already on it
 - https://join.slack.com/t/dataset-jsonhackathon/shared_invite/zt-2rzvlout0--e4u2QjTJZ~XNgqSla0eFA
- Wiki site for Dataset-JSON Hackathons
 - <https://wiki.cdisc.org/pages/viewpage.action?pageId=157583203>



External Communications

- COSA Spotlight Webinars
- Interchange presentations
- Social media posts
- Projects included in Dataset-JSON presentations
- Tools available for use in future pilots

Thank You!

Questions?

Sam Hume

shume@cdisc.org

<https://www.linkedin.com/in/sam-hume-dsc>

Charles Shadle

cshadle@cdisc.org

