

Dataset-JSON Update & Dataset-JSON Viewer Hackathon

COSA Spotlight Q3

Omar Garcia Calderon, CDISC

Dmitry Kolosov, Parexel



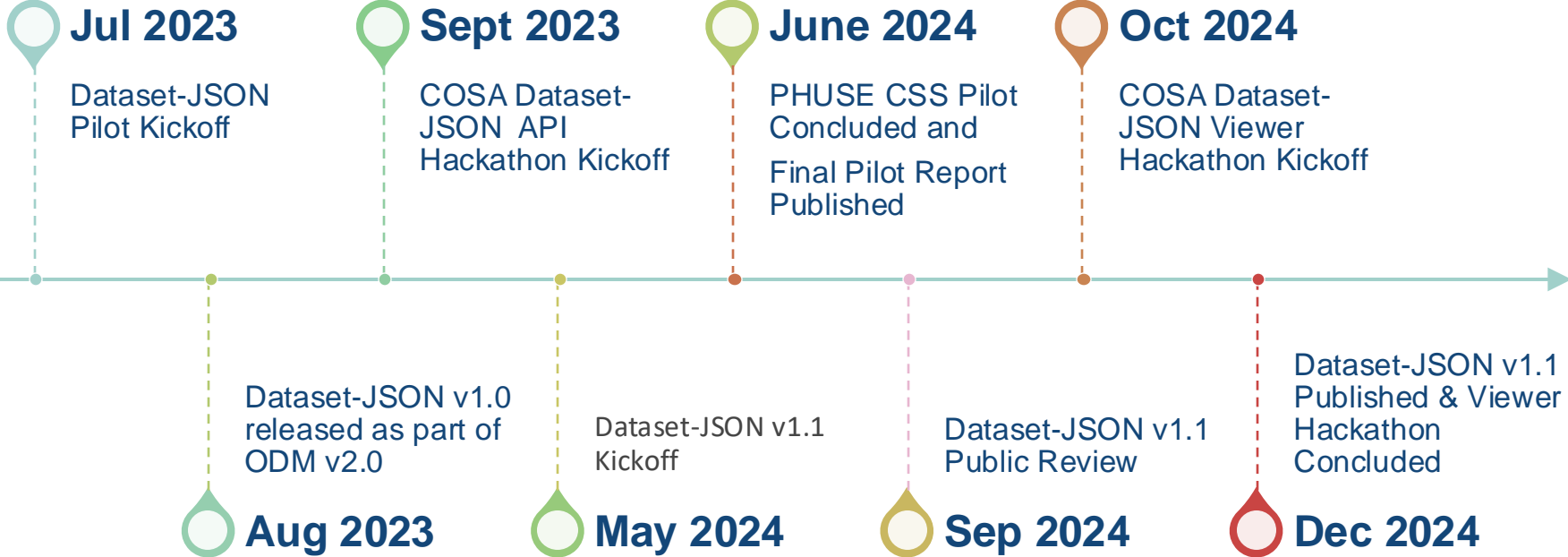
cdisc[®]



Agenda

1. Dataset-JSON Timeline
2. What is Dataset-JSON v1.1?
3. Public Review
4. COSA Dataset-JSON Viewer Hackathon

Dataset-JSON v1.1 Timeline



Features to complete

Feature	Due Date
• Review and decide on the solutions to the pilot finding	✓
• Update the standards specification	✓
• Author a User's Guide	✓
• Update the schema	✓
• Update the examples	✓
• Develop an NDJSON version based on the draft	✓
• Update NDJSON examples	✓
• Internal Review	✓
• Public Review	10 September 2024
• JSONX development - compressed archive of datasets	TBD
• Review draft API specification	TBD
• Work with conversion software developers to update tools	TBD
• Work with viewer software developers to develop/update tools	TBD



What is Dataset-JSON v1.1

A high-level overview of Dataset-JSON



Introducing Dataset-JSON

What is Dataset-JSON?

A dataset exchange standard for exchanging tabular data leveraging JSON designed to meet the regulatory submission needs and eliminate the limitations of legacy formats

Dataset-JSON is...

- Designed to support a broad range of data exchange scenarios
- Supports API and file-based data exchange
- JSON is simple to implement, very stable, and widely supported
- Open-source schema supports any tabular format
- Extensible to support new metadata and new use cases
- Linked to Define-XML for complete metadata

Changes in Dataset-JSON 1.1

Among other changes version 1.1 the main changes are:

- Flattening the overall structure
- Additional datatypes and metadata
- Support for NDJSON
- Converting integer dates to ISO 8601

```
{
  "datasetJSONCreationDateTime": "2023-03-22T11:53:27",
  "datasetJSONVersion": "1.1.0",
  "fileOID": "www.sponsor.xyz.org.project123.final",
  "dbLastModifiedDateTime": "2023-02-15T10:23:15",
  "originator": "Sponsor XYZ",
  "sourceSystem": {
    "name": "Software ABC",
    "version": "1.0.0"
  },
  "studyOID": "xxx",
  "metaDataVersionOID": "xxx",
  "metaDataRef": "https://metadata.location.org/api.link",
  "itemGroupOID": "IG.DM",
  "isReferenceData": false,
  "records": 100,
  "name": "DM",
  "label": "Demographics",
  "columns": [ ... ],
  "rows": [ ... ]
}
```



Public Review



Public Review

- Public Review started on September 10, 2024
- Public Review lasts 30 days and closes on October 10, 2024
- Reviewers are requested to provide comments via JIRA
- The JIRA project is Dataset-JSON review comments (DSJSONCT) located at:
 - <https://jira.cdisc.org/projects/DSJSONCT>
- Instructions for Reviewers:
 - <https://wiki.cdisc.org/display/DSJSON1DOT1/Instructions+for+Reviewers>
- Public Review Webinar recording and slides:
 - <https://www.cdisc.org/events/webinar/dataset-json-v1-1-public-review-webinar>



Dataset-JSON v1.1 Content to Review

1. Dataset-JSON v1.1 Specification

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

2. User Guide

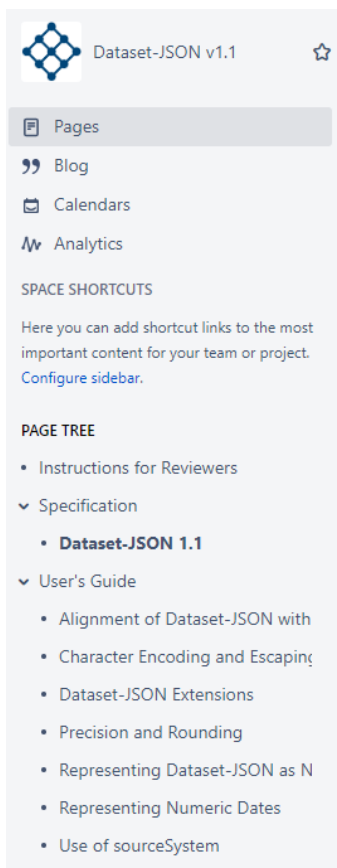
- <https://wiki.cdisc.org/display/DSJSON1DOT1/User%27s+Guide>

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 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

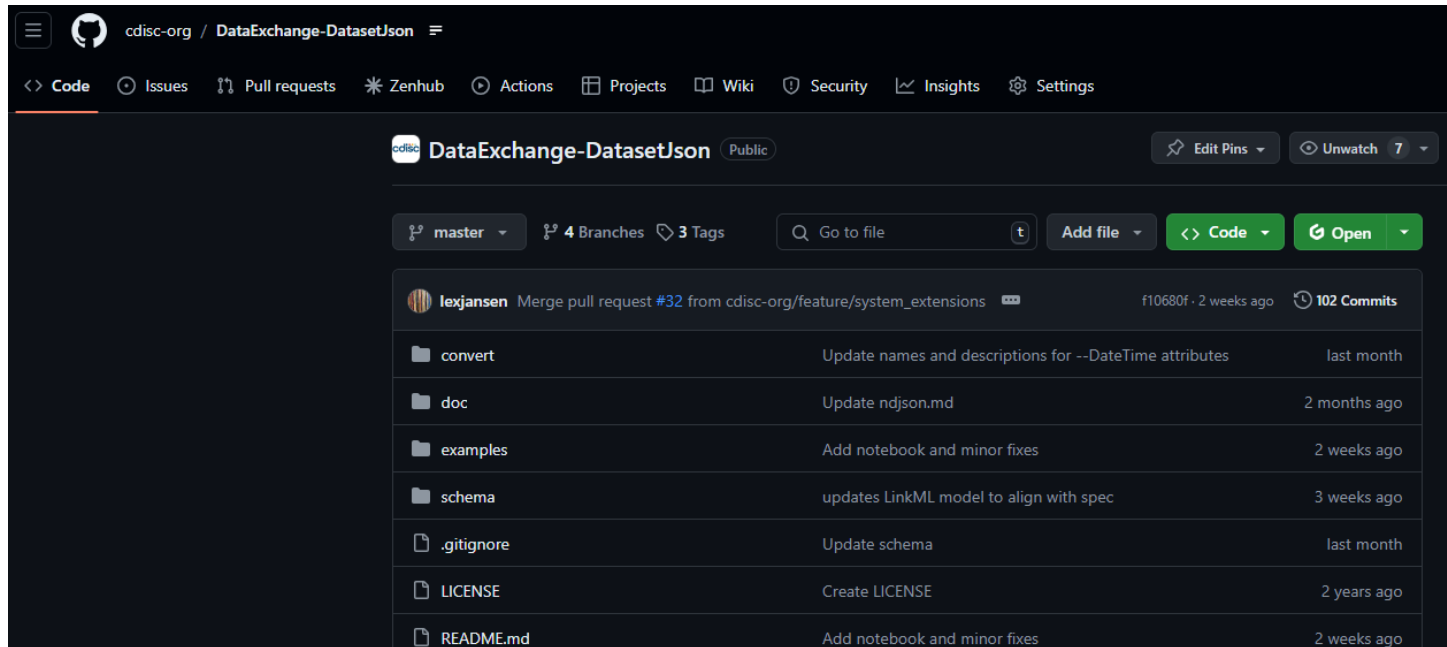
DRAFT

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 commit history is visible, showing a merge pull request #32 from `cdisc-org/feature/system_extensions` by `lexjansen` 2 weeks ago, with 102 commits. The commit history table is as follows:

Commit	Description	Time
<code>convert</code>	Update names and descriptions for --DateTime attributes	last month
<code>doc</code>	Update ndjson.md	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



COSA Dataset-JSON Viewer 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.

- 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 Webinar to demo solutions



Previous Dataset-JSON Hackathons

- COSA has hosted several hackathons
 - Dataset-JSON Hackathon I (July - October 2022)
 - Goal:** Develop open-source software tools for working with Dataset-JSON
 - Result:** 21 solutions for Dataset-JSON 1.0 available in [COSA directory](#)
 - Dataset-JSON Hackathon II (September - October 2023)
 - Goal:** Develop API specification for Dataset-JSON
 - Result:** GitHub repository with the API specification

Language	Solutions
R	5
SAS	4
Python	5
JavaScript	4
Java	1
Swift	1
XSLT	1



COSA Dataset-JSON Viewer Hackathon

- Primary objective: Create Dataset-JSON Viewer software
 - Projects can be new or extensions to existing software
- 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
 - More details will be made available before the start of the hackathon
 - Let us know if you may be interested in participating



Thank You!

Questions?

Dmitry Kolosov
dmitry.kolosov@parexel.com

Omar Garcia Calderon
ogcalderon@cdisc.org

