



SPA MDR CDR SCE

JSON for Submissions: Why You Should Care, What You Need to Know, and What To Do Now

Bill Qubeck
Sridhar Vijendra



The Clinical Data
Science Conference

EU 2025

16–19 November
CCH – Congress Center Hamburg



About the Presenters



Bill Qubeck

Senior Director of
Services



bill.qubeck@sycamoreinformatics.com



North Carolina
USA



<https://www.linkedin.com/in/williamqubeck/>



Bio

Over the past 30 years, Bill has led many digital transformations in clinical trial analysis and reporting, optimizing trials through advanced data strategies and emerging technologies. He has partnered with leading pharma, CROs, & technology providers to modernize clinical research, integrating new tech & next-gen analytics. With expertise in statistical computing, biometrics, and regulatory, he helps organizations harness technology to improve efficiency & drive innovation.



Sridhar Vijendra

Director of Services



sridhar.vijendra@sycamoreinformatics.com



Bengaluru
India



<https://www.linkedin.com/in/sridhar-vijendra-396b07b/>



Bio

Sridhar Vijendra has a Master's degree in Biomedical engineering and an inclination towards all things technical. He started his career doing programming and research in biomedical signal processing algorithms. After serving multiple roles including that of a statistical programmer, he believes he has found his calling working closely with business analysts and project managers on SCE implementations at Sycamore Informatics, India.



Agenda

- What is JSON?
- What is Dataset-JSON?
- Evolution of Submission Standards
- What does Regulatory think?
- Modern SCEs & Dataset-JSON
- Recommendations
- Q & A

Why You Should Care? It will replace XPTs and define.xml (eventually)



What is JSON (JavaScript Object Notation)?

- Lightweight, **text-based** data interchange format
- Both **human-readable** and **machine-readable**
- A key value structure
- Syntax uses braces, brackets, commas and colons
- Supports **hierarchical and nested data** therefore it can represent complex relationships and metadata

```
{ "employees": [  
  { "firstName": "John", "lastName": "Doe" },  
  { "firstName": "Anna", "lastName": "Smith" },  
  { "firstName": "Peter", "lastName": "Jones" }  
]}
```

- A JSON example defines an 'employees' object, with an array of 3 employees*



JSON vs XML (What is the Difference?)

How JSON & XML are Similar

Both file formats are:

- Human readable (aka "self describing")
- Have a hierarchical structure (values within values)

Why JSON is Better?

- JSON is faster and easier than XML to parse
- JSON files file sizes are smaller than XML

How JSON Differs

- Doesn't use end tag
- Can use arrays
- Standard JavaScript functions (and other languages too) can parse it

Why JSON?



Example JSON and Corresponding XML

Total Number of Characters 336 vs 420

JSON (array of records)

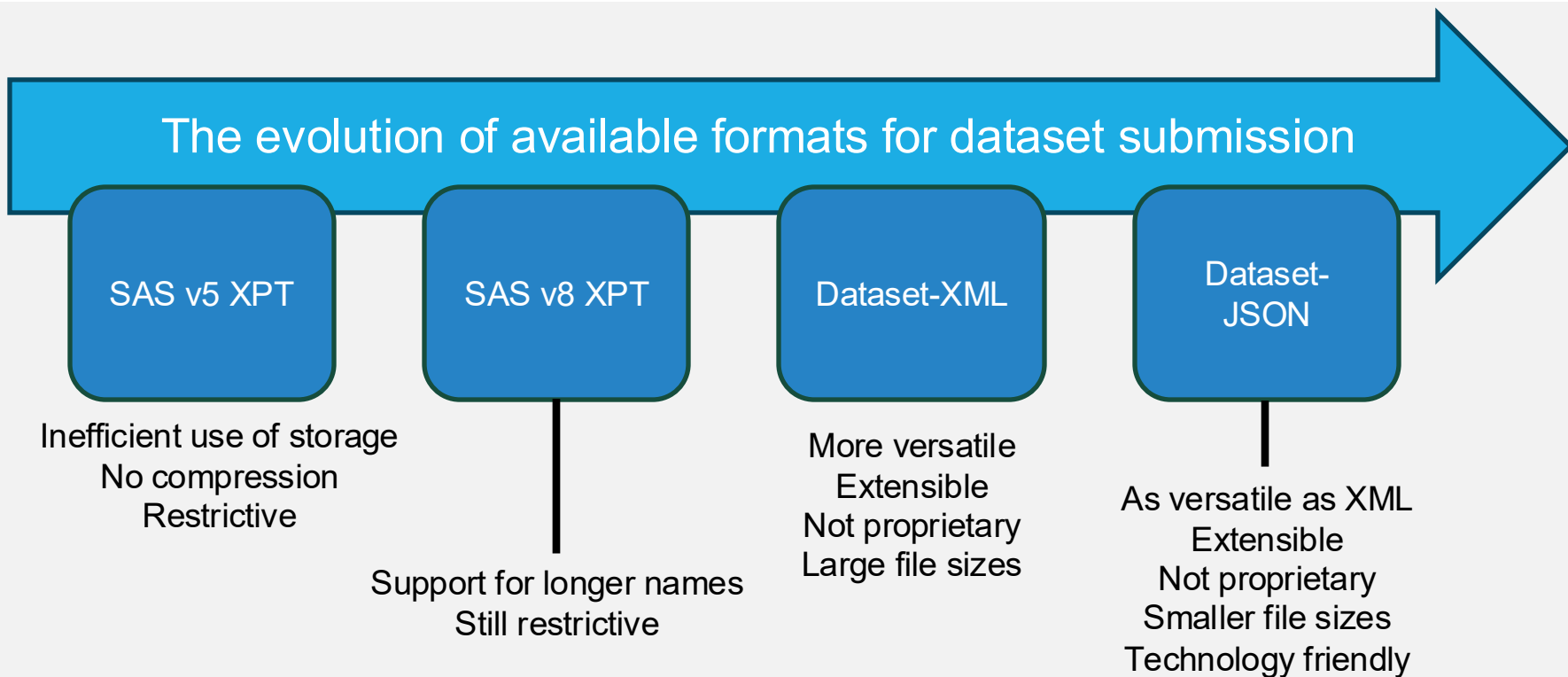
```
[
  {
    "STUDYID": "ABC123",
    "DOMAIN": "AE",
    "USUBJID": "ABC123-001",
    "AESEQ": 1,
    "AETERM": "Headache",
    "AESTDTC": "2025-10-28",
    "AEENDTC": "2025-10-29",
    "AESEV": "MILD",
    "AESER": "N",
    "AEREL": "POSSIBLE",
    "AEACN": "DOSE NOT CHANGED",
    "AEOUT": "RECOVERED/RESOLVED",
    "AETOXGR": "1"
  }
]
```

XML (simple SDTM-style)

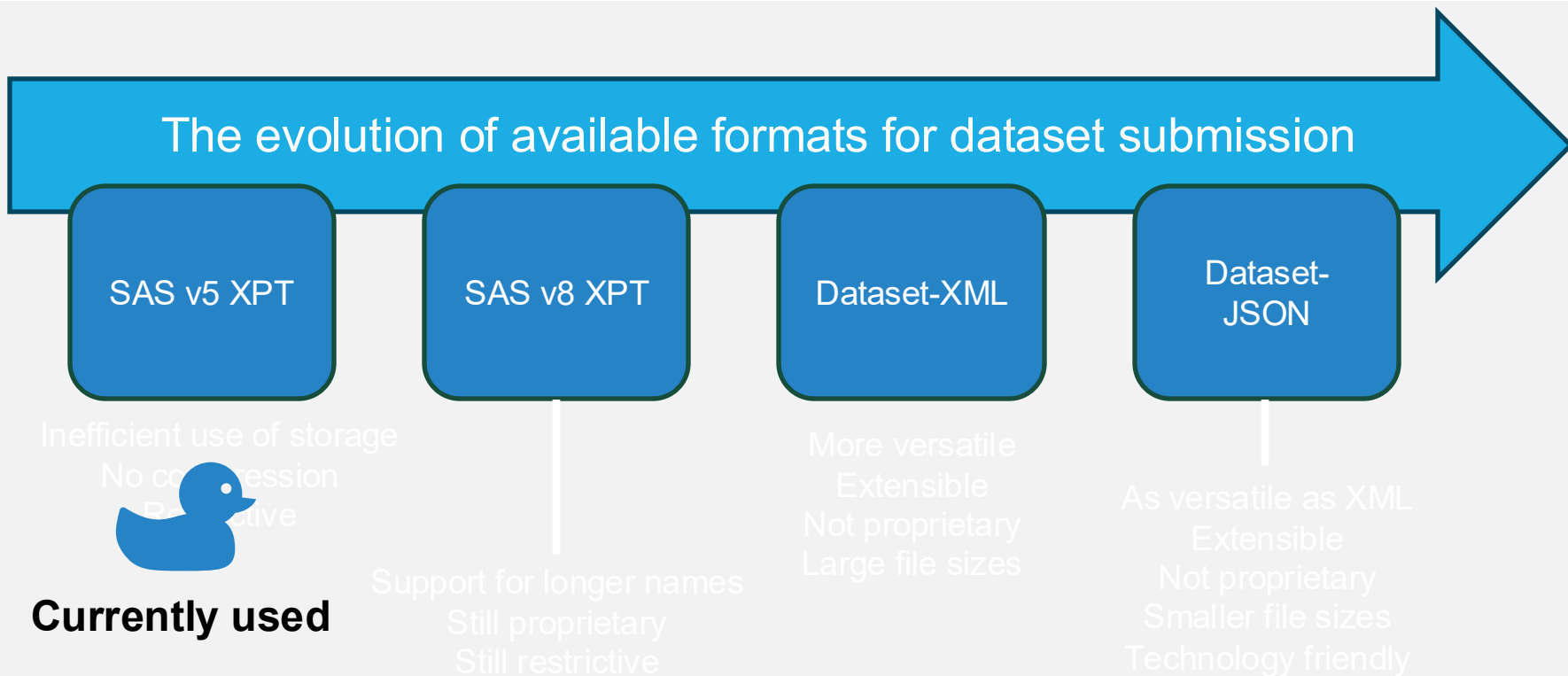
```
<AE_RECORDS>
  <AE>
    <STUDYID>ABC123</STUDYID>
    <DOMAIN>AE</DOMAIN>
    <USUBJID>ABC123-001</USUBJID>
    <AESEQ>1</AESEQ>
    <AETERM>Headache</AETERM>
    <AESTDTC>2025-10-28</AESTDTC>
    <AEENDTC>2025-10-29</AEENDTC>
    <AESEV>MILD</AESEV>
    <AESER>N</AESER>
    <AEREL>POSSIBLE</AEREL>
    <AEACN>DOSE NOT CHANGED</AEACN>
    <AEOUT>RECOVERED/RESOLVED</AEOUT>
    <AETOXGR>1</AETOXGR>
  </AE>
</AE_RECORDS>
```



Evolution of Submission Standards

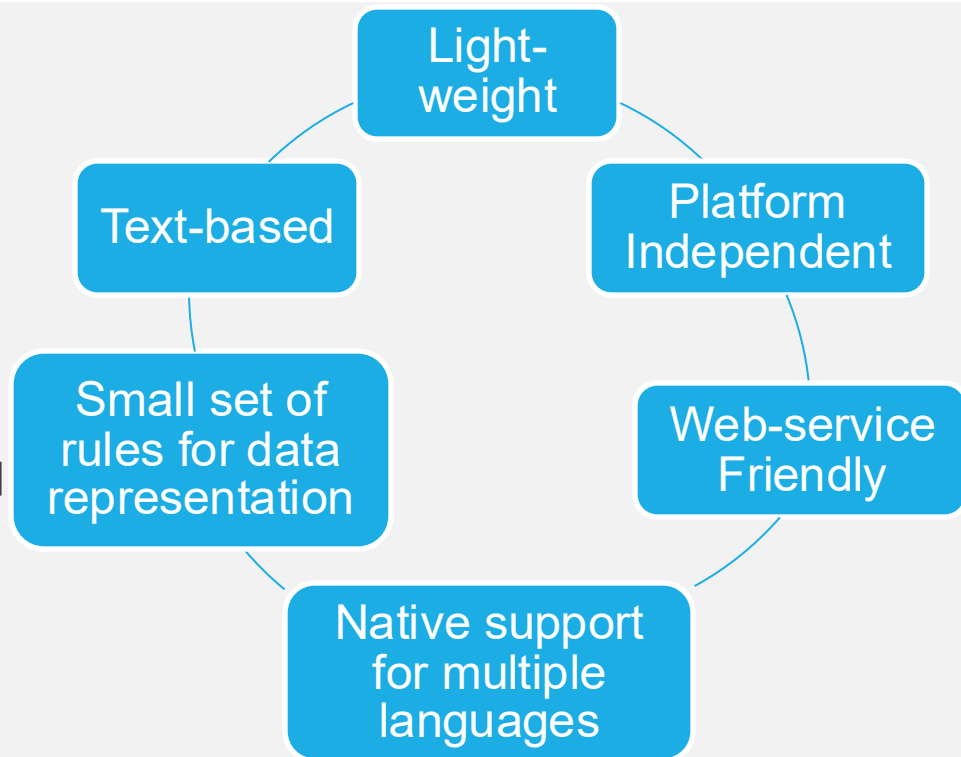


But Have We Really Evolved?



What is Dataset-JSON?

- Created by the CDISC-Dataset-JSON for Data Exchange Team
- Adapted from Dataset-XML but uses JSON format
- Linked to the Define-XML via object identifiers
- Human-readable since it is JSON
- UTF-8 Support and no name limitations
- Contains one object for each dataset



What Does Dataset-JSON Look Like (3 Parts)?

Part 1. Metadata Section (itemGroupData):

Provides high-level context, including the dataset name, label, domain, and creation details. It supports study identifiers, version information, and pointers to controlled terminologies or codelists

```
"itemGroupData": {  
  "IG.DM": {  
    "records": 254,  
    "name": "CM",  
    "label": "Concomitant Medication",  
    "items": [...  
      ],  
    "itemData": [...  
      ]  
  }  
}
```

*Dataset
metadata*

Data



Part 2: JSON Variables Section (items)

```
"items": [  
  {"name": "USUBJID", "label": "Unique Subject Identifier", "type":  
"character"},  
  {"name": "AGE", "label": "Age", "type": "numeric"},  
  {"name": "SEX", "label": "Sex", "type": "character"}  
]
```

- Defines the dataset structure, listing names, labels, data types, lengths, and formats
- Supports additional metadata like controlled terminology references, origins, and derivation details, significantly improving traceability and alignment with CDISC Define-XML



Part 3: JSON Data Section (itemData)

```
"itemData": [  
  {"USUBJID": "1001-1001", "AGE": 36, "SEX": "M"},  
  {"USUBJID": "1001-1012", "AGE": 65, "SEX": "F"}  
]
```

- Contains the actual observations
- Missing values are explicitly represented as null, avoiding the ambiguity inherent in XPT's missing value representations



Essential SAS Toolset Mastery

SAS Tool/ Procedure	Primary Function	Programmer's Mandate
PROC JSON	Writing (Export)	The primary output tool. Must be used with options (like KEYS or WRITE OPEN/CLOSE) to enforce the complex, nested structure required by the CDISC Dataset-JSON schema.
JSON LIBNAME Engine	Reading (Import)	Used for round-trip testing and reading JSON source data, allowing the JSON file to be treated as a standard SAS dataset for comparison and validation.
PROC GROOVY or PROC LUA	Advanced Handling	Used for custom, highly efficient parsing or generation for very large datasets where complex mapping rules exceed standard PROC JSON performance.



CRITICAL INTEROPERABILITY CHALLENGES FOR LOSSLESS DATA for SAS Programmers

- **Date/Time Epoch and Representation**
 - **Mitigation:** Programmers must ensure dates and datetimes are converted to the unambiguous ISO 8601 string format in the JSON output
 - **SAS Example:** Use explicit formats (e.g., IS8601DA, IS8601DT) to enforce the correct string structure before export via PROC JSON
- **Floating-Point Precision and Rounding**
 - **Risk:** Rounding differences constitute a loss of data precision & validation failure upon round-trip testing (from SAS to JSON and vice versa)
 - **Mitigation:** For high-precision variables, the best practice is to represent the value as a JSON string (text) to prevent the receiving system's JSON library from introducing unintended rounding



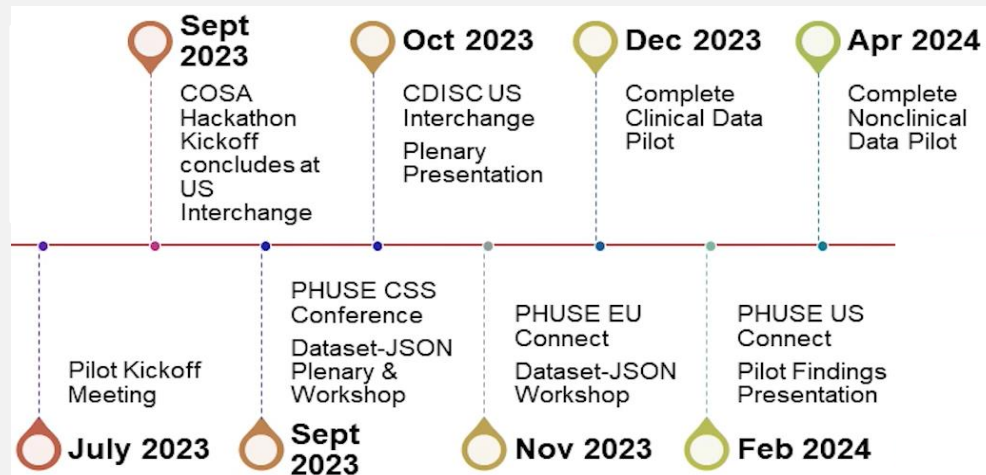
A Framework for JSON Validation: The Three-Pillar Protocol

1. **Structural Validation (Schema Check):** Ensure JSON file adheres to the physical rules defined in CDISC Dataset-JSON v1.1
2. **Metadata Conformance (Define-XML Cross-Check):** A semantic check to ensure metadata in JSON aligns with the Define-XML file. The programmer must ensure attributes, origins, and derivation details in the JSON "Variables Section" are consistent
3. **Round-Trip Integrity (Lossless Data Verification):** Read the JSON back into the native environment and performing a comparison against the original source dataset



FDA and Dataset-JSON

- FDA did a **pilot** on the use of Dataset-JSON v1.0
- Results showed **minor issues** related to conversion tools and interoperability testing
- “Dataset-JSON is **viable** as the primary transport option”
- FDA’s tools **require updates**
- **Dataset-JSON v1.1**, was released in December 2024



Modern SCEs & the Endless Possibilities

- Modern Statistical Computing Environments (SCEs) must support JSON natively along with SAS, R, Python, and other analytical tools
- Web services already speak in JSON
- Opens the door to seamless end-to-end integrations



Recommendations & Next Steps

- **Assess** current submission workflows
- **Modify** existing submission macros
- **Leverage** modern Statistical Computing Environments (SCEs)
- **Incorporate** open-source which natively support JSON, into submission pipelines
- **Pilot** Internal Proof-of-Concepts
- **Engage** with Industry Collaboratives
- **Monitor** Regulatory Developments



Summary

- Dataset-JSON is **forward-looking** and gets past all the limitations of the existing SAS v5 XPT format
- **Supported** by multiple modern languages that are rapidly being adopted by organizations
- Requires **minimal** re-work for submission macros but very little re-training of programmers
- **Seamless** integration with modern web-based SCEs

Dataset-
JSON

Define.XML

SAS v5 XPT



Q & A

References

- ANDERSON, JESSIE AND HU ME, SAM. "DATA SET-JSON A SALTERNATIVE TRANSPORT FORMAT FOR REGULATORY SUBMISSIONS." YOUTUBE VIDEO. OCTOBER 26, 2023. [HTTPS://YOUTU.BE/UjR6D4AW7NA?si=K0EGWCKPKWNYKf](https://youtu.be/UjR6D4AW7NA?si=K0EGWCKPKWNYKf).
- ANDERSON, JESSIE AND HU ME, SAM. "DATA SET-JSON PILOT REPORT AND NEXT STEPS." PRESENTATION, CDISC JAPAN INTERCHANGE. JUNE 2024. [HTTPS://WWW.CDISC.ORG/SITES/DEFAULT/FILES/2024-06/2024-JP-INTERCHANGE-SESSIONS-DATASET-JSON.PDF](https://www.cdisc.org/sites/default/files/2024-06/2024-JP-INTERCHANGE-SESSIONS-DATASET-JSON.PDF).
- ANDERSON, JESSIE AND HU ME, SAM. CDISC "DATA SET-JSON A SALTERNATIVE TRANSPORT FORMAT FOR REGULATORY SUBMISSIONS." PRESENTATION, CDISC PLENARY. OCTOBER 2023. [HTTPS://WWW.CDISC.ORG/SITES/DEFAULT/FILES/2023-10/2023-CDISC-DATASET-JSON-PLenary-V5_0.PDF](https://www.cdisc.org/sites/default/files/2023-10/2023-CDISC-DATASET-JSON-PLenary-V5_0.PDF).
- CLINICAL DATA INTERCHANGE STANDARD CONSORTIUM (CDISC). "DATASET-JSON." ACCESSED OCTOBER 28, 2025. [HTTPS://WWW.CDISC.ORG/STANDARDS/DATA-EXCHANGE/DATASET-JSON](https://www.cdisc.org/standards/data-exchange/dataset-json).
- CLINICAL DATA INTERCHANGE STANDARD CONSORTIUM (CDISC). DEFINITION XML SPECIFICATION. AUSTIN, TX: CDISC. ACCESSED OCTOBER 23, 2025. [HTTPS://WWW.CDISC.ORG/STANDARDS/DATA-EXCHANGE/DEFINEXML](https://www.cdisc.org/standards/data-exchange/definexml).
- FOOD AND DRUG ADMINISTRATION (FDA). "ELECTRONIC STUDY DATA SUBMISSION; DATA STANDARDS; CLINICAL DATA INTERCHANGE STANDARDS CONSORTIUM DATASET-JAVAScript OBJECT NOTATION; REQUEST FOR COMMENTS." FEDERAL REGISTER 90, NO. 67 (APRIL 9, 2025). [HTTPS://WWW.FDA.GOV/DOCKETS/OPA/2025-04-09-0129-0001](https://www.fda.gov/dockets/opa/2025-04-09-0129-0001).
- FOOD AND DRUG ADMINISTRATION (FDA). "PROVIDING REGULATORY SUBMISSIONS IN ELECTRONIC FORMAT — STANDARDIZED STUDY DATA GUIDANCE FOR INDUSTRY. SILVER SPRING, MD: U.S. FOOD AND DRUG ADMINISTRATION, JUNE 2023. ACCESSED OCTOBER 23, 2025. [HTTPS://WWW.FDA.GOV/REGULATORY-INFORMATION/SEARCH-FDA-GUIDANCE-DOCUMENTS/REGULATORY-SUBMISSIONS-ELECTRONIC-FORMAT-STANDARDIZED-STUDY-DATA](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/regulatory-submissions-electronic-format-standardized-study-data).
- FOOD AND DRUG ADMINISTRATION (FDA). "STUDY DATA TECHNICAL CONFORMANCE GUIDE. SILVER SPRING, MD: U.S. FOOD AND DRUG ADMINISTRATION. ACCESSED OCTOBER 23, 2025. [HTTPS://WWW.FDA.GOV/REGULATORY-INFORMATION/SEARCH-FDA-GUIDANCE-DOCUMENTS/STUDY-DATA-TECHNICAL-CONFORMANCE-GUIDE-TECHNICAL-SPECIFICATIONS-DOCUMENT](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/study-data-technical-conformance-guide-technical-specifications-document).
- FOOD AND DRUG ADMINISTRATION (FDA). "DATASET-JSON PILOT REPORT AND NEXT STEPS." YOUTUBE VIDEO, 15:47. JULY 25, 2024. [HTTPS://WWW.YOUTUBE.COM/WATCH?V=UjR6D4AW7NA](https://www.youtube.com/watch?v=UjR6D4AW7NA).
- FOOD AND DRUG ADMINISTRATION (FDA). "DATASET-JSON PILOT REPORT AND NEXT STEPS." NEWS & EVENTS FOR HUMAN DRUGS. JULY 25, 2024. [HTTPS://WWW.FDA.GOV/DRUGS/NEWS-EVENTS-HUMAN-DRUGS/DATASET-JSON-PILOT-REPORT-AND-NEXT-STEPS-07252024](https://www.fda.gov/drugs/news-events-human-drugs/dataset-json-pilot-report-and-next-steps-07252024).
- FOOD AND DRUG ADMINISTRATION (FDA). "STUDY DATA STANDARDS RESOURCES." LAST REVIEWED JULY 25, 2024. [HTTPS://WWW.FDA.GOV/INDUSTRY/FDA-DATA-STANDARDS-ADVISORY-BOARD/STUDY-DATA-STANDARDS-RESOURCES](https://www.fda.gov/industry/fda-data-standards-advisory-board/study-data-standards-resources).
- HU ME, SAM. "NO MORE XPT: PILING NEW DATASET-JSON FOR FDA SUBMISSIONS." CLINICAL LEADER, JULY 12, 2023. [HTTPS://WWW.CLINICALLEADER.COM/DOX/NO-MORE-XPT-PILOTING-NEW-DATASET-JSON-FOR-FDA-SUBMISSIONS-0001](https://www.clinicalleader.com/dox/no-more-xpt-piloting-new-dataset-json-for-fda-submissions-0001).
- JANSSEN, LEX. "DATA SET-JSON MACROS IN GIT REPO" ACCESSED OCTOBER 29, 2025. [HTTPS://GITHUB.COM/LEXJANSEN/DATASET-JSON-SAS/BLOB/MAS TER/MACROS/WRITE_DATASET_JSON_SAS](https://github.com/LexJanssen/dataset-json-sas/blob/master/macros/write_dataset_json_sas).
- JANSSEN, LEX. "WORKING WITH DATASET-JSON USING SAS." PAPER PRESENTED AT PHARMASUG 2022. [HTTPS://WWW.LEXJANSEN.COM/PHARMASUG/2022/AD/PHARMASUG-2022-AD-130_PPT.PDF](https://www.lexjanssen.com/pharmasug/2022/ad/pharmasug-2022-ad-130_ppt.pdf).
- JANSSEN, LEX. "DATASET-JSON: SAS IMPLEMENTATION" ACCESSED OCTOBER 29, 2025. [HTTPS://WWW.CDISC.ORG/SITES/DEFAULT/FILES/PDF/CDISC05A_WEBINAR_20231005_DATASET-JSON_SA SPDF](https://www.cdisc.org/sites/default/files/pdf/cdisc05a_webinar_20231005_dataset-json_sas.pdf).
- LINKER, ADA M. "CREATING AND CONTROLLING JSON OUTPUT WITH THE JSON PROCEDURE." PAPER SAS 3506-2019. ACCESSED OCTOBER 28, 2025. [HTTPS://SUPPORT.SAS.COM/RESOURCES/PAPERS/PROCEEDINGS19/3506-2019.PDF](https://support.sas.com/resources/papers/proceedings19/3506-2019.pdf).
- PHARMACEUTICAL USER SOFTWARE EXCHANGE (PHUSE). "DATASET-JSON A SALTERNATIVE TRANSPORT FORMAT FOR REGULATORY SUBMISSIONS." PHUSE ADVANCE HUB. LAST MODIFIED JUNE 10, 2025. [HTTPS://ADVANCE.HUB.PHUSE.GLOBAL/WIKI/SPACES/WEL/PAGES/2680606/DATASET-JSON-AS-ALTERNATIVE-TRANSPORT-FORMAT-FOR-REGULATORY-SUBMISSIONS](https://advance.hub.phuse.org/spaces/wel/pages/2680606/dataset-json-alternative-transport-format-for-regulatory-submissions).
- SALTWARE SOLUTIONS. "UNPACKING THE COMPLEXITIES OF CLINICAL DATA SUBMISSION STANDARDS." LAST MODIFIED AUGUST 2024. [HTTPS://WWW.SALTWARE.COM/SOLUTIONS/UNPACKING-CLINICAL-SUBMISSION-STANDARDS/](https://www.saltware.com/solutions/unpacking-clinical-submission-standards/).
- SAS GLOBAL FORUM. "CREATING AND CONTROLLING JSON OUTPUT WITH PROC JSON." SAS GLOBAL FORUM PROCEEDINGS. (YEAR AND AUTHOR VARY, FOCUSING ON PROC JSON CONTROL). ACCESSED OCTOBER 23, 2025. [HTTPS://SUPPORT.SAS.COM/](https://support.sas.com/).
- SAS. "JSON PROCEDURE." ACCESSED OCTOBER 23, 2025. [HTTPS://DOCUMENTATION.SAS.COM/DOC/EN/PGMSASCD/C/V_068/PROC/NONE1FK9QP2MNN181192QK3AH2E.HTM](https://documentation.sas.com/doc/en/pgmsascdc/V_068/proc/n0ne1fk9qp2mnn181192qk3ah2e.htm).



Connect with the Presenters!

Bill Qubeck



Sridhar Vijendra



Thank you!

