

cdisc®

2026 US Interchange

THE FUTURE IS CONNECTED: STANDARDS AND AI POWERING DIGITAL TRANSFORMATION



Dataset-JSON Is Coming: What FDA Adoption Could Mean for Standards, Systems, and Submission Readiness

Bill Qubeck, Senior Director, Sycamore Informatics

Speakers



Bill Qubeck

Senior Director

Sycamore Informatics

Over the past 25 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.

Agenda

1. Why This Matters Now?
2. What is JSON?
3. What is Dataset-JSON?
4. Tools, Interoperability, & Validation
5. FDA, Modern SCEs, & Next Steps
6. Q&A

Why This Matters Now: Dataset JSON is positioned to replace XPT and reshape how submission data is prepared, validated, exchanged, and delivered.

As of 06/17/2026

CDER and CBER, in collaboration with CDISC and PhUSE, has conducted preliminary testing of CDISC's Dataset JSON message exchange standard. Initial results indicate potential use as a replacement for XPT v5. As such, CBER and CDER **will conduct further testing** to evaluate Dataset JSON's capability to support the submission of regulatory study data. Results will be communicated, and we will engage relevant parties for input as we progress through this evaluation.

https://www.fda.gov/industry/fda-data-standards-advisory-board/study-data-standards-resources

Why This Matters Now

Takeaway: This is a readiness trigger, not a final mandate.

1. **Standard development**
   Prior work and workshops
2. **CDISC PHUSE pilot**
   Sept 2023 to Apr 2024
3. **FDA Federal Register notice**
   April 9, 2025
4. **API and compression standards**
   Dec 11, 2025
5. **Future adoption decision**
   Timing not final

What is known

FDA is requesting comment on whether to accept Dataset JSON for future electronic study data exchange.

What it means

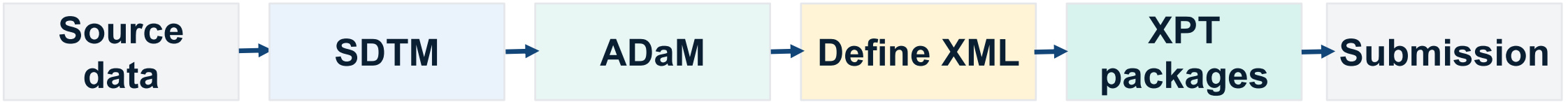
Sponsors, CROs, vendors, quality, regulatory operations, and technology groups need a shared readiness plan.

What it is not

It is not yet a regulatory requirement, and it should not be treated as a rushed conversion exercise.

The Submission Format We Inherited

XPT has lasted because it solved a real regulatory exchange problem: stability, predictability, and broad ecosystem support.



Why it worked

A common transport file helped align sponsors, CROs, vendors, and regulators around consistent delivery mechanics.

Why it became embedded

Submission pipelines, validation tools, publishing systems, archives, and reviewer expectations were built around it.

Why change is hard

The challenge is not replacing a file. It is changing a controlled operating model that touches many functions.

Limitations of XPT are no Longer Only Technical

They affect automation, interoperability, metadata discipline, and review readiness.

XPT Strengths

- Stable and broadly understood
- Deeply embedded in submission operations
- Supported by existing regulatory workflows
- Predictable for long term archival and review
- Familiar to sponsors, CROs, and vendors

Modern Exchange Needs

- Richer machine-readable structures
- Cleaner integration with R, Python, and APIs
- Stronger metadata alignment at generation time
- Automated validation and controlled services
- Better fit for scalable and interoperable pipelines

Pressure point: legacy stability versus modern exchange

What is JSON?

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

```
{
   "metadata": {
      "standard": "CDISC SDTM",
      "studyid": "ABC123",
      "domain": "DM"
   },
   "records": [
      { "USUBJID": "ABC123-001", "SEX": "M", "AGE": 34, "ARM": "Placebo" },
      { "USUBJID": "ABC123-002", "SEX": "F", "AGE": 29, "ARM": "Treatment A" }
   ]
}
```

cdisc

* Illustrative example only.

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

https://www.w3schools.com/js/js_json_xml.asp

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>
```

cdisc

* Illustrative example only.

What is Dataset-JSON?

Dataset JSON in Plain English

Dataset JSON is a structured JSON representation of tabular clinical data designed for exchange across systems and tools.

Consumption	Review tools, R, Python, SAS, validation engines, APIs
Transport	Dataset JSON files, API based exchange, compressed forms
Metadata	Define XML, controlled terminology, origins, derivations, value level metadata
Clinical standards	SDTM, ADaM, SEND where applicable, study data standards

Key point: the clinical meaning still comes from the standards and metadata. Dataset JSON changes how the controlled data asset can be exchanged and

Not Just a New File Extension

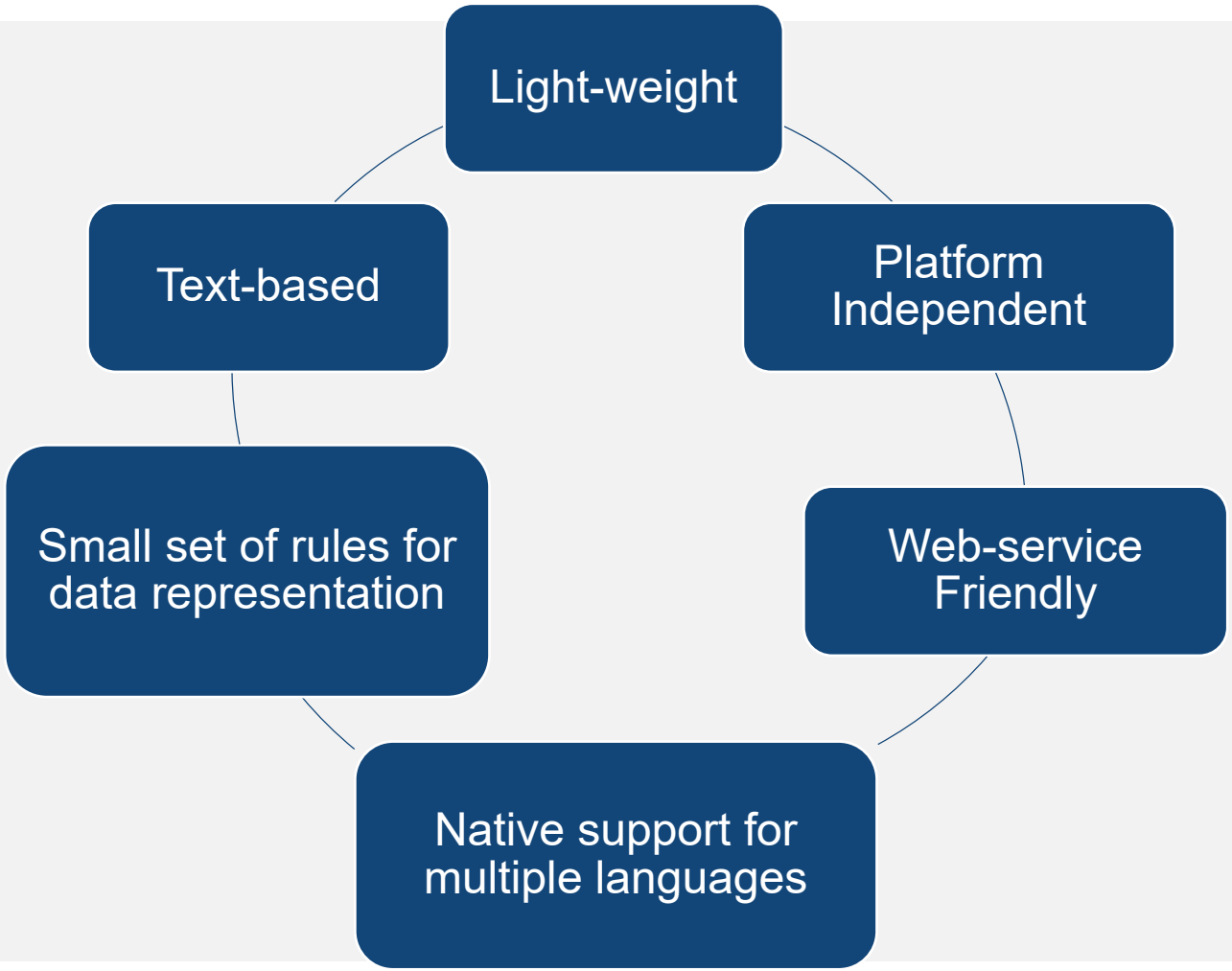
The wrong approach is to treat Dataset JSON as a late-stage conversion. The better approach is to place it inside the controlled data lifecycle.

Dimension	XPT mindset	Dataset JSON mindset
Role	Final transport package	Structured exchange asset
Metadata	Referenced through external context	Explicit alignment with Define XML is central
Validation	Transport and content checks	Schema plus content, metadata, and equivalence checks
Tooling	Mature legacy support	Modern languages, APIs, and compression need assessment
Governance	Known procedures and SOPs	Versioned standards, tools, and generation logic

Readiness question: Where should Dataset JSON enter the controlled lifecycle?

Next Level Description: 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



Tools, Interoperability, & Validation

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

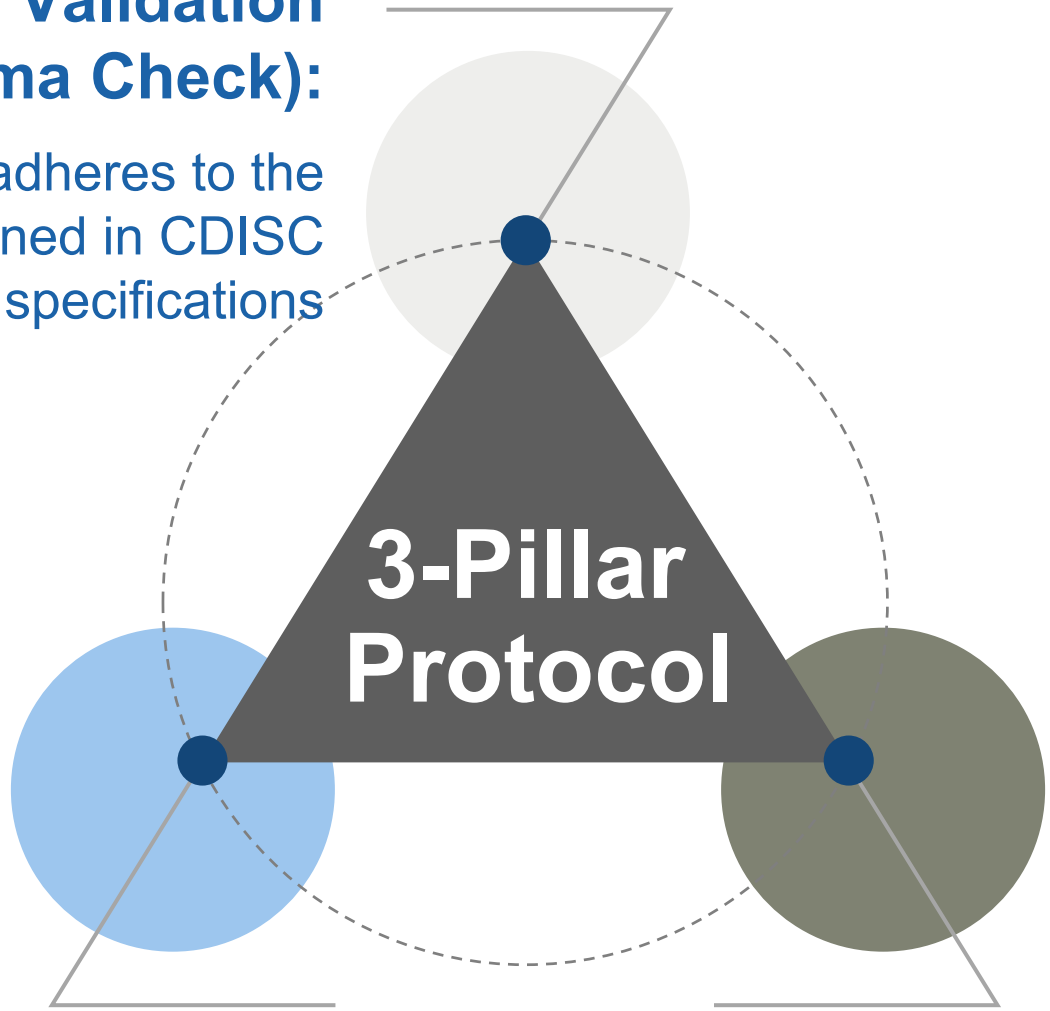
A Framework for JSON Validation: The Three-Pillar Protocol

Structural Validation (Schema Check):

Ensure JSON file adheres to the physical rules defined in CDISC Dataset-JSON specifications

Round-Trip Integrity (Lossless Data Verification):

Read the JSON back into the native environment and performing a comparison against the original source dataset



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

FDA, Modern SCEs, & Next Steps

What the FDA, CDISC, and PHUSE Pilot Demonstrated

The pilot moved Dataset JSON from concept toward feasibility. Enterprise readiness still requires organization specific evidence.

Feasibility	Transport	Data integrity	Operations	Readiness
Can it represent study data for regulatory exchange?	Can it serve as a transport format?	Can it move data without loss in tested scenarios?	Can business operations continue without major disruption?	Can the full ecosystem adopt it consistently?

Implementation lesson

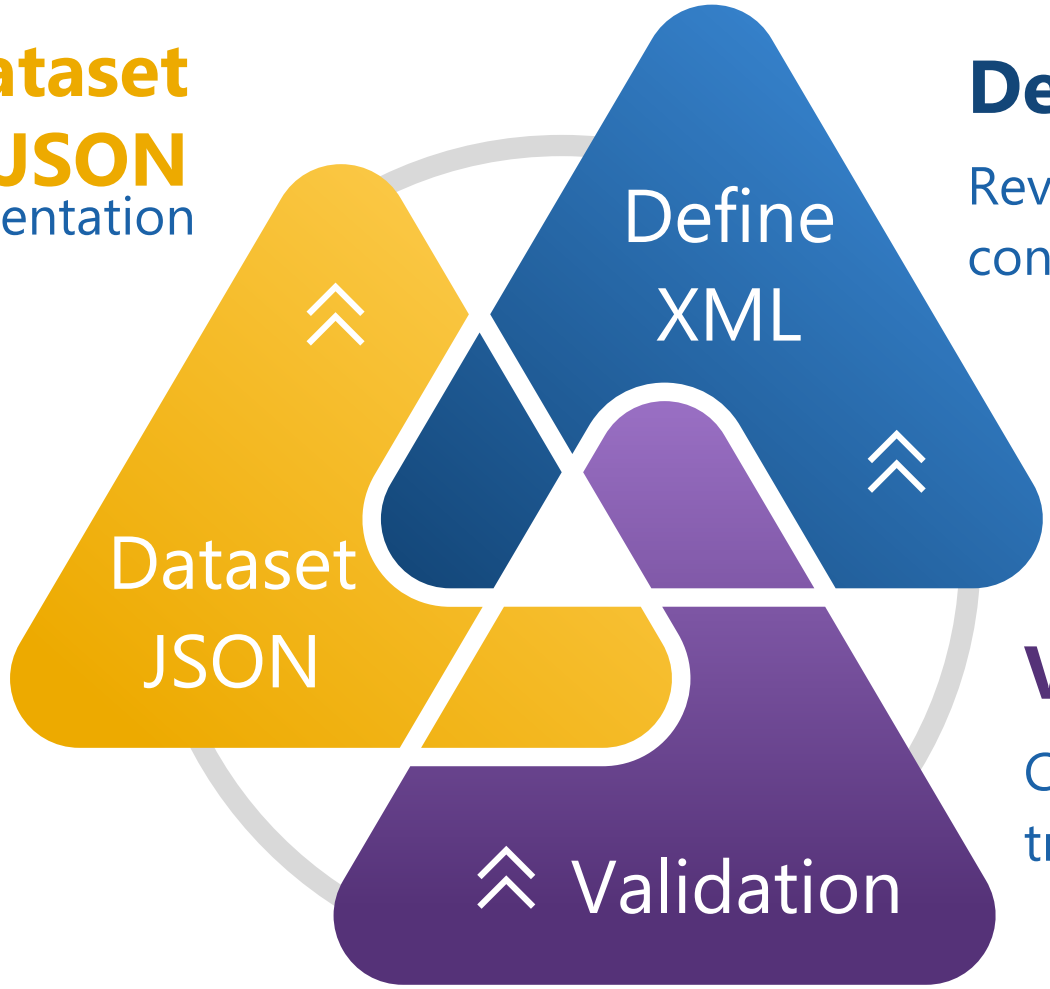
Pilot evidence reduces uncertainty, but it does not replace sponsor and vendor validation. Each organization still needs representative testing, metadata alignment checks, toolchain qualification, equivalence testing, and a clear controlled process.

The Define XML Relationship

Dataset JSON does not reduce the importance of Define XML. It increases the importance of metadata discipline.

High risk failure mode:
Dataset JSON content, Define XML metadata, and validation evidence disagree on variable attributes, controlled terminology, origins, derivations, or value level metadata.

Dataset JSON
Data representation



Define.xml
Reviewer metadata context

Validation
Consistency and traceability

The Coexistence Period

The practical transition is dual support: XPT will remain operational while Dataset JSON capability matures.

Current state	**XPT primary**	Existing submission processes, tools, archives, and SOPs remain dominant.
Transition state	**XPT plus Dataset JSON**	Parallel generation, equivalence testing, partner alignment, and controlled pilots.
Future state	**Dataset JSON primary**	Dataset JSON becomes the preferred exchange asset while XPT remains for legacy needs.

Coexistence requires clear rules: when each format is produced, which output is authoritative, how equivalence is proven, and who approves release.

Readiness Framework

Readiness should be assessed across standards, metadata, tools, validation, and operating model.

Standards intelligence

Monitor FDA, CDISC, PHUSE, implementation guidance, and standards versioning.

Metadata governance

Strengthen Define XML quality, controlled terminology, origins, derivations, and value level metadata.

Toolchain readiness

Assess generation, validation, publishing, transfer, archive, review, and API support.

Dataset JSON readiness

Validation strategy

Define intended use, risk, requirements, evidence, traceability, and release controls.

Operating model

Clarify accountability across sponsor, CRO, vendor, regulatory operations, quality, and technology.

90-Day Action Plan

The immediate goal is not a full platform rebuild. It is dependency discovery, representative testing, and governance alignment.

Days 1 to 30 Discover

- Assign accountable owner
- Inventory XPT dependencies
- Identify tools that read, write, validate, publish, transfer, or archive XPT
- Select representative prior studies

Days 31 to 60 Test

- Generate Dataset JSON from representative SDTM and ADaM datasets
- Run schema and content checks
- Compare against XPT outputs
- Test Define XML alignment

Days 61 to 90 Govern

- Draft validation approach
- Define target operating model
- Update SOP impact assessment
- Engage vendors and CROs
- Roadmap dual generation

Decision gate at day 90: define the gaps, own the roadmap, and avoid rushed compliance later.

Modern SCEs & JSON-Native Integration

From isolated programming workbenches to governed, API-ready clinical data platforms.

Outcome: lower conversion friction • stronger traceability • faster integration • readiness for Dataset-JSON

{ } JSON-native data exchange

Read, write, validate, and move Dataset-JSON without fragile late-stage conversion.

API Web services integration

Connect repositories, validators, review tools, automation services, and external systems.

R/Py Multi-language analytics

Support SAS, R, Python, and future analytical tools within the same governed environment.

GxP Governed execution

Preserve audit trails, versioned runtimes, reproducibility, and inspection-ready evidence.



A modern SCE must treat JSON as a native interoperability layer, not merely an export format.

cdisc

Final Takeaway

Dataset JSON is not just a replacement for XPT. It is a forcing function for a more modern submission operating model.

Do not wait	Do not reduce it to conversion	Use it to strengthen operations
Understand dependencies before a final adoption timeline creates urgency.	Treat Dataset JSON as part of the controlled data lifecycle, not only as a new output file.	Improve metadata, validation, automation, traceability, and partner accountability.

Organizations that treat the change as a file conversion will underprepare. Organizations that treat it as a submission operating model change will be ready.

“Dataset-JSON is the next step in our Submission Standard Evolution”



Thank You!



Questions?

