

# APAC 2026

**Next-Gen Transport Standards: Dataset-JSON from a Programmer's Lens**

**Venkata Koteswara Rao Rayala**  
**Merck Specialities Private Limited**

**Date:** 19 - 21 February 2026

**Location:** Hyderabad, India



# Disclaimer

- The views and opinions expressed in this presentation are solely those of the presenter and do not necessarily reflect the official policy or position of Merck Specialities Private Limited.
- These slides are intended for educational purposes only and are not intended for wider distribution outside the intended purpose without presenters' approval.



# Agenda



**Dataset-JSON: Background & Rational**



**XPT vs. Dataset-JSON**



**Dataset-JSON Structure**



**Dataset-JSON with SAS**



**Dataset-JSON with R**



**Dataset-JSON Exchange**



**P21 Community: Supporting Dataset-JSON Adoption**



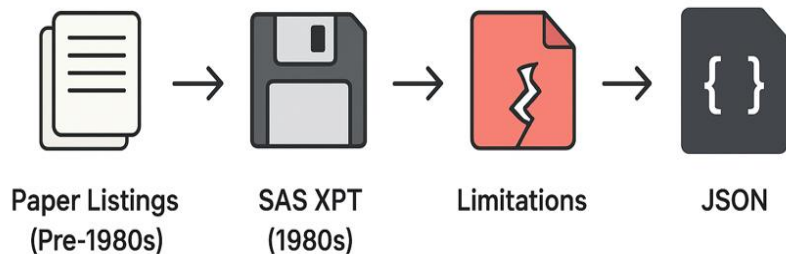
**Summary & Conclusion**



**References**



# Dataset-JSON: Background/Rationale



## Dataset-JSON Pilots (2020s)

Human-readable format with rich metadata

Built for APIs and smooth data exchange

FDA pilot – rising regulatory standard

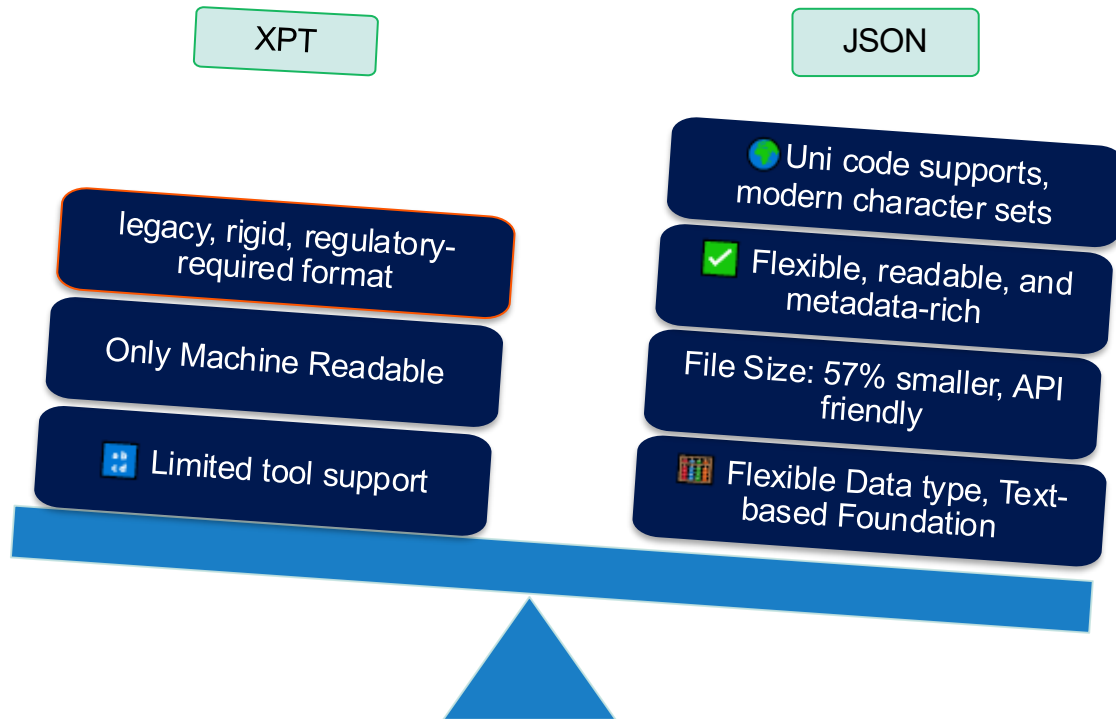
## Looking Ahead

JSON poised as the default for data sharing and interoperability

Likely successor to legacy XPT



# XPT vs. Dataset-JSON



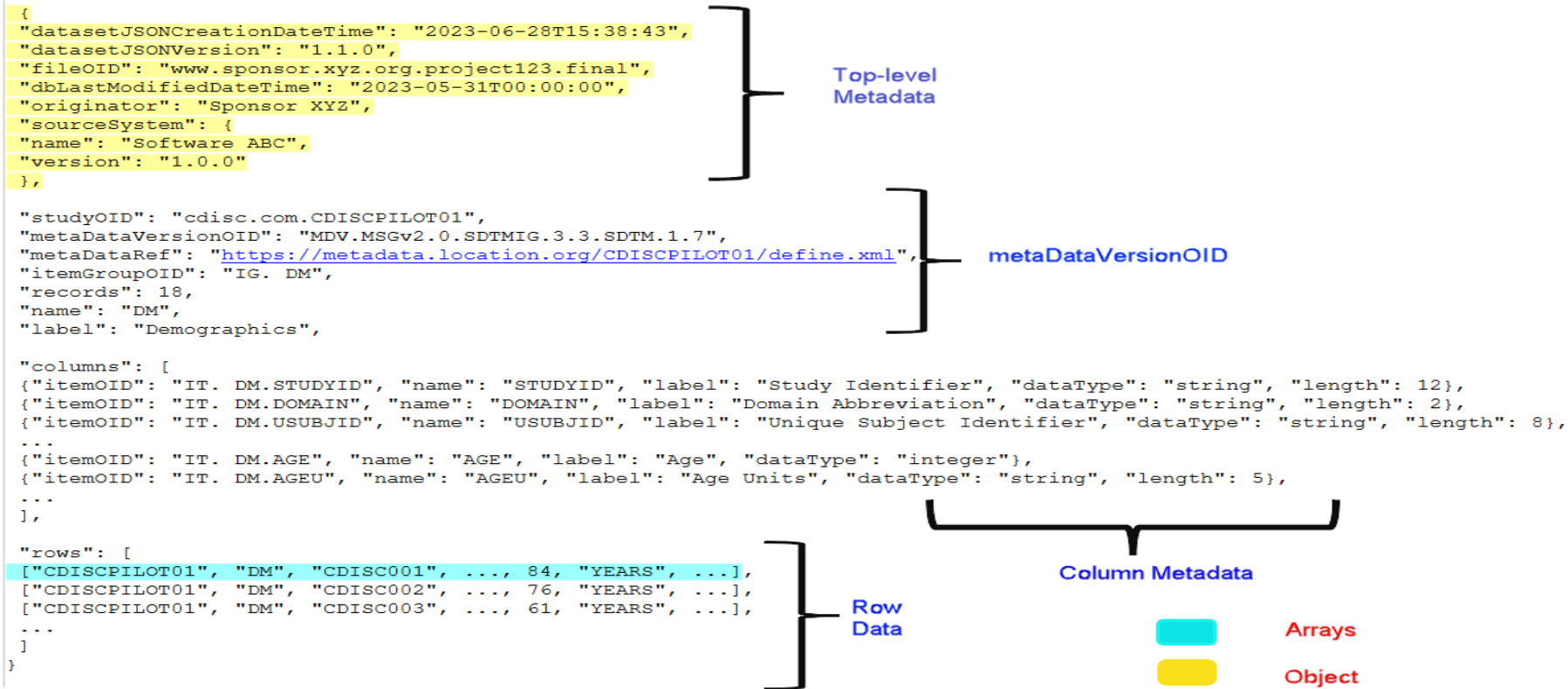


# SAS v5 XPT Size (Kb) vs Dataset-JSON

|      |      | SAS v5 XPT | Dataset-JSON | % Size Reduction |
|------|------|------------|--------------|------------------|
| SDTM | LB   | 5,917      | 837          | 85.85            |
|      | VS   | 2,699      | 624          | 76.88            |
|      | DM   | 784        | 223          | 71.55            |
| ADaM | ADAE | 4,133      | 3,204        | 22.48            |
|      | ADVS | 14,255     | 9,571        | 32.85            |
|      | ADLB | 14,150     | 13,360       | 5.58             |

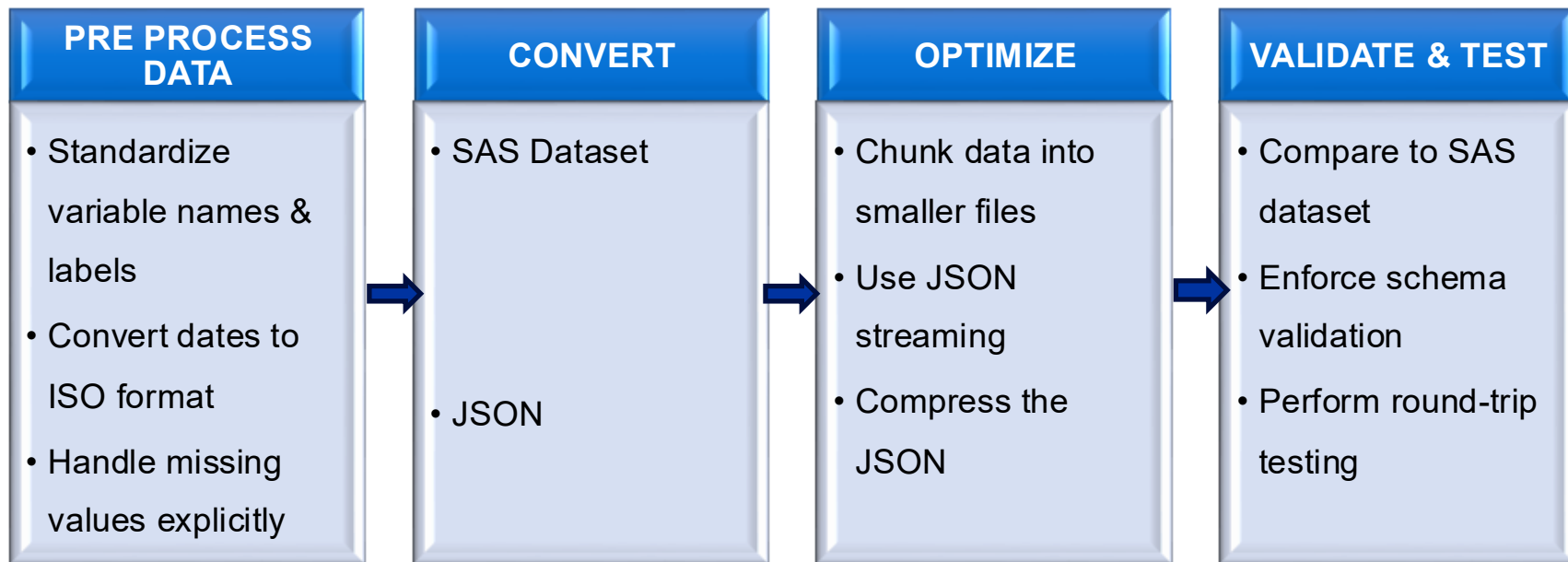


# Dataset-JSON Structure





# SAS-to-JSON Data Transformation Workflow





# Dataset-JSON with SAS

```
PROC JSON OUT=jsonfout PRETTY;
  WRITE OPEN OBJECT;
  WRITE VALUES "clinicalData";
  WRITE OPEN OBJECT;
  WRITE VALUES "studyOID" "<study OID>";
  WRITE VALUES "metaDataVersionOID" "<MDV OID>";
  WRITE VALUES "itemGroupData";
  WRITE OPEN OBJECT;
  WRITE VALUES "IG.DM";
  WRITE OPEN OBJECT;
  WRITE VALUES "records" 18;
  WRITE VALUES "name" "DM";
  WRITE VALUES "label" "Demographics";
  WRITE VALUES "items";
  WRITE OPEN ARRAY;
  EXPORT work.column_metadata / KEYS;
  WRITE CLOSE;
  WRITE VALUES "itemData";
  WRITE OPEN ARRAY;
  EXPORT work.column_data / NOKEYS;
  WRITE CLOSE;
  WRITE CLOSE;
  WRITE CLOSE;
  WRITE CLOSE;
  WRITE CLOSE;
RUN;
```

```
{
  "clinicalData": {
    "studyOID": "<study OID>",
    "metaDataVersionOID": "<MDV OID>",
    "itemGroupData": {
      "IG.DM": {
        "records": 18,
        "name": "DM",
        "label": "Demographics",
        "items": [
          ],
        "itemData": [
          ]
        }
      }
    }
  }
```

column metadata

Column data



# Dataset-JSON with R

```
library(jsonlite)
# Sample column metadata (similar to column_metadata in SAS)
column_metadata <- data.frame(
  itemOID = c("IT.DM.USUBJID", "IT.DM.AGE"),
  name = c("USUBJID", "AGE"),
  label = c("Subject ID", "Age"),
  dataType = c("string", "integer"),
  stringsAsFactors = FALSE
)
# Sample column data (similar to column_data in SAS)
column_data <- list(
  list("SUBJ001", 34),
  list("SUBJ002", 29)
)
# Construct Dataset-JSON structure
dataset_json <- list(
  clinicalData = list(
    studyOID = "<study OID>",
    metaDataVersionOID = "<MDV OID>",
    itemGroupData = list(
      IG.DM = list(
        records = length(column_data),
        name = "DM",
        label = "Demographics",
        items = column_metadata,
        itemData = column_data
      )
    )
  )
)
```

```
{
  "clinicalData": {
    "studyOID": "<study OID>",
    "metaDataVersionOID": "<MDV OID>",
    "itemGroupData": {
      "IG.DM": {
        "records": 2,
        "name": "DM",
        "label": "Demographics",
        "items": [
          {
            "itemOID": "IT.DM.USUBJID",
            "name": "USUBJID",
            "label": "Subject ID",
            "dataType": "string"
          },
          {
            "itemOID": "IT.DM.AGE",
            "name": "AGE",
            "label": "Age",
            "dataType": "integer"
          }
        ]
      }
    ],
    "itemData": [
      ["SUBJ001", 34],
      ["SUBJ002", 29]
    ]
  }
}
```



# Dataset-JSON Exchange



## Exporting Dataset-JSON

- **SAS:**  
PROC JSON OUT="path\file.json";
  - Use EXPORT dataset / KEYS; for metadata
  - Use EXPORT dataset / NOKEYS; for data only
- **R:**  
library(jsonlite)  
write\_json(data, "file.json", pretty = TRUE)
  - Structure data as nested list before export



## Importing Dataset-JSON

- **SAS:**  
libname myjson json "path\file.json";
  - Access tables: proc print data=myjson.alldata;
- **R:**  
library(jsonlite)  
data <- fromJSON("file.json", simplifyVector = FALSE)
  - Explore: data\$clinicalData\$itemGroupData\$IG.DM\$itemData



# P21 Community: Supporting Dataset-JSON Adoption



## Validator Support

- P21 now accepts and validates **Dataset-JSON** files.



## Define.xml Integration

- Datasets can be linked via **Define.xml** — no need for XPT viewers.



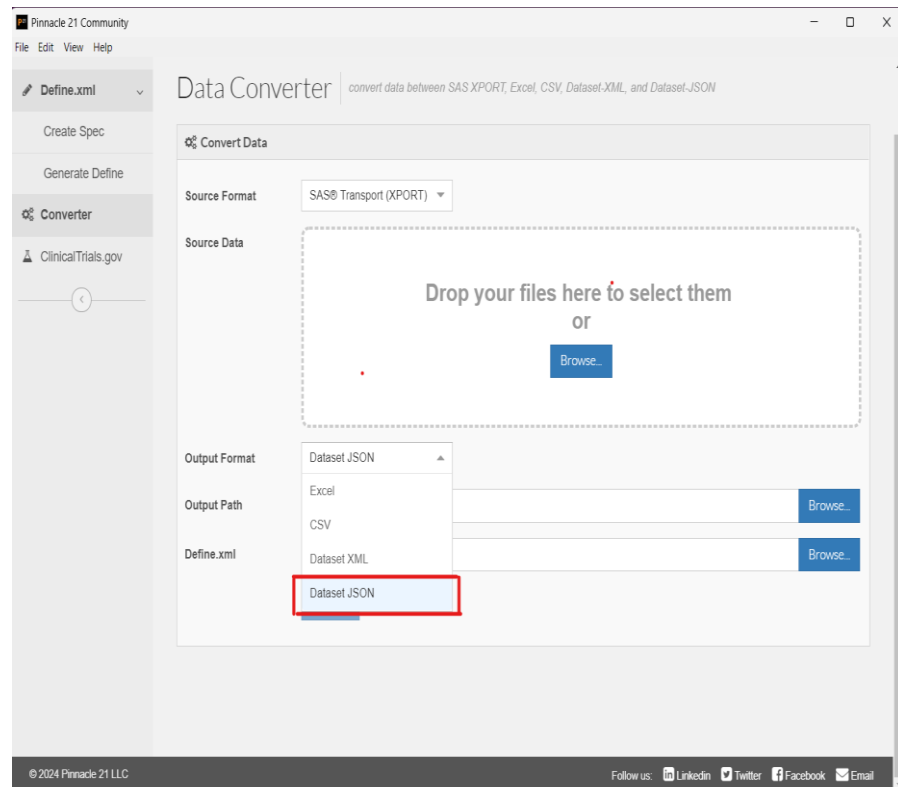
## Conversion Ready

- External tools enable **SAS XPT**  **Dataset-JSON** conversion



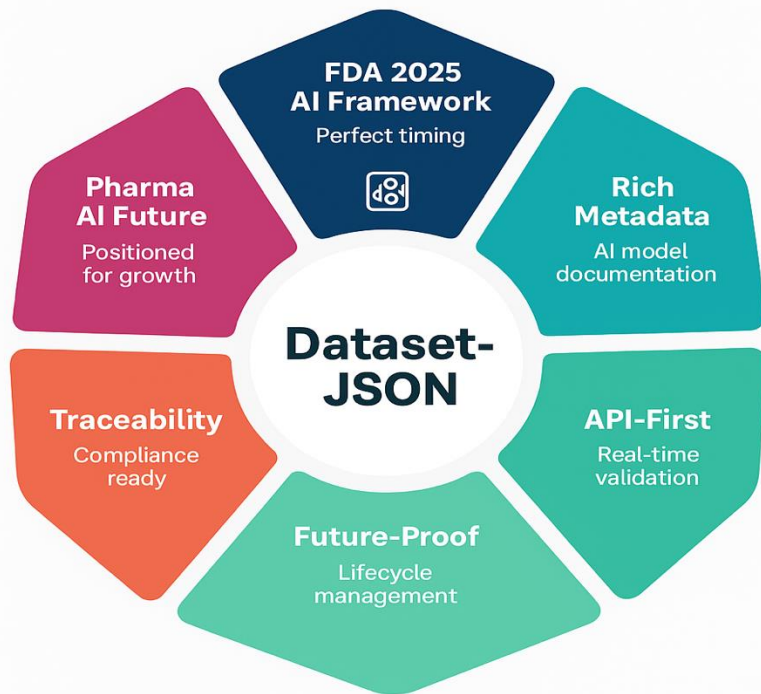
## Future-Proof Submissions

- Aligns with emerging regulatory standards for data exchange





# Dataset-JSON + FDA AI Framework



\*\*\*\* Dataset-JSON is the backbone for FDA's AI-driven regulatory future



# Summary & Conclusion

📌 **Dataset-JSON is at a pilot stage**, with early studies exploring its potential for regulatory submissions.

📌 Compared to XPT, **Dataset-JSON offers clear advantages:**

- Rich metadata support
- API-first, cloud-ready architecture
- Language-independent structure

📌 However, **challenges remain:**

- Limited compatibility with some legacy systems
- Integration with existing tools may require adaptation

📌 **Looking ahead:**

- Dataset-JSON shows promise as a **future-ready standard for data exchange**, supporting modern, flexible, and scalable workflows.



# References:

1. [www.cdisc.org/standards/data-exchange/dataset-json](http://www.cdisc.org/standards/data-exchange/dataset-json)
2. [https://www.lexjansen.com/pharmasug/2022/AD/PharmaSUG-2022-AD-150\\_ppt.pdf](https://www.lexjansen.com/pharmasug/2022/AD/PharmaSUG-2022-AD-150_ppt.pdf)
3. [https://www.lexjansen.com/css-us/2023/PRE\\_01.pdf](https://www.lexjansen.com/css-us/2023/PRE_01.pdf)
4. <https://www.fda.gov/industry/fda-data-standards-advisory-board>
5. <https://www.ecma-international.org/publications-and-standards/standards/ecma-404/>

# Thank you



**The Global Healthcare  
Data Science Community**

**Contact Channels**

📞 UK +44 1843 609600  
✉ [office@phuse.global](mailto:office@phuse.global)  
🌐 [phuse.global](http://phuse.global)

**Social Media**

🐦 [@phusetwitta](https://twitter.com/phusetwitta)  
📘 [/phusebook](https://www.facebook.com/phusebook)  
▶ [/phusetube](https://www.youtube.com/channel/UCphusetube)  
🌐 [/company/phuse](https://www.linkedin.com/company/phuse)