

APAC 2026

Optimizing Pinnacle 21 Compliance with Automated Metadata Updates

Karthik Palani
Pfizer



Abstract

Pinnacle 21 plays a pivotal role in validating clinical trial data against CDISC standards but resolving recurring SDTM and ADAM issues and manually editing Define-XML metadata remains time-intensive and error-prone. To address this, we developed a Python-based automation tool that streamlines SDTM and ADAM issue resolution and metadata population in P21 workflows. The tool leverages predefined corpora in Excel format to intelligently suggest fixes for commonly encountered SDTM and ADAM findings eliminating redundancy and improving accuracy. It also automates the completion of Define-XML fields in the Variables, Methods, and Comments tabs, ensuring regulatory compliance while reducing manual effort. Built for scalability, the tool integrates seamlessly with existing workflows and enhances consistency across studies. By automating repetitive tasks, it accelerates submission readiness, reduces human error, enables data managers to focus on higher-value review and QC activities. This solution represents a significant step toward modernizing clinical data workflows and improving regulatory submission efficiency.



Disclaimer

We confirm that, the opinions and thoughts discussed in this presentation are subject to our own independent views and are not subject to the opinions of the organization that we represent.



Program Overview



Handling Issues



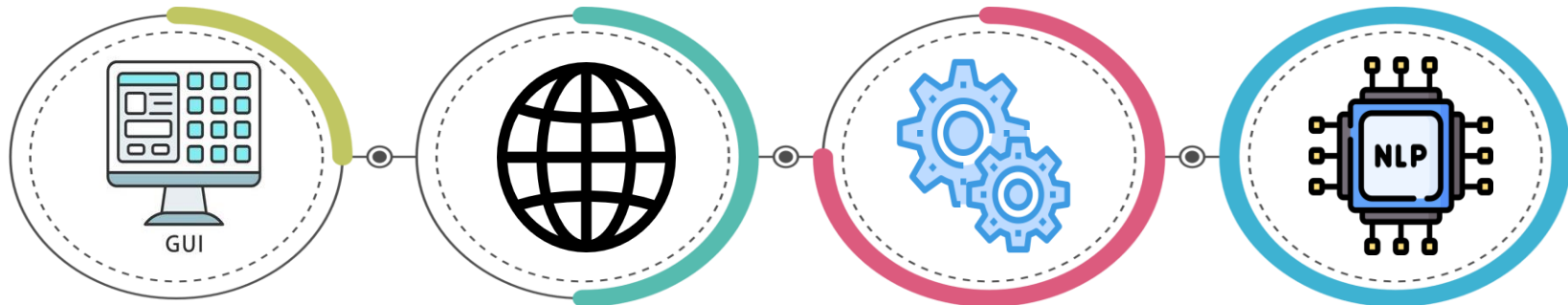
Fixing Define



Providing
explanations



P21E Tool Components



tkinter





Challenges vs. Automation

Challenges		Automation Solution
Time Intensive Process		Consistent Metadata Updates
Metadata inconsistencies		Intelligent Issue Detection & Resolution
Manual issue correction		Hands-free Issue Remediation
Repetitive Tasks		Scalable & Efficient Workflows
High Error Rates		Enhanced Accuracy & Compliance



Tool Interface

Pinnacle 21 Automation Tool

Pinnacle 21 Automation Tool

Dataset Selection & Download Define MetaData Updates (Excel) Issues Fixing Issues Automation

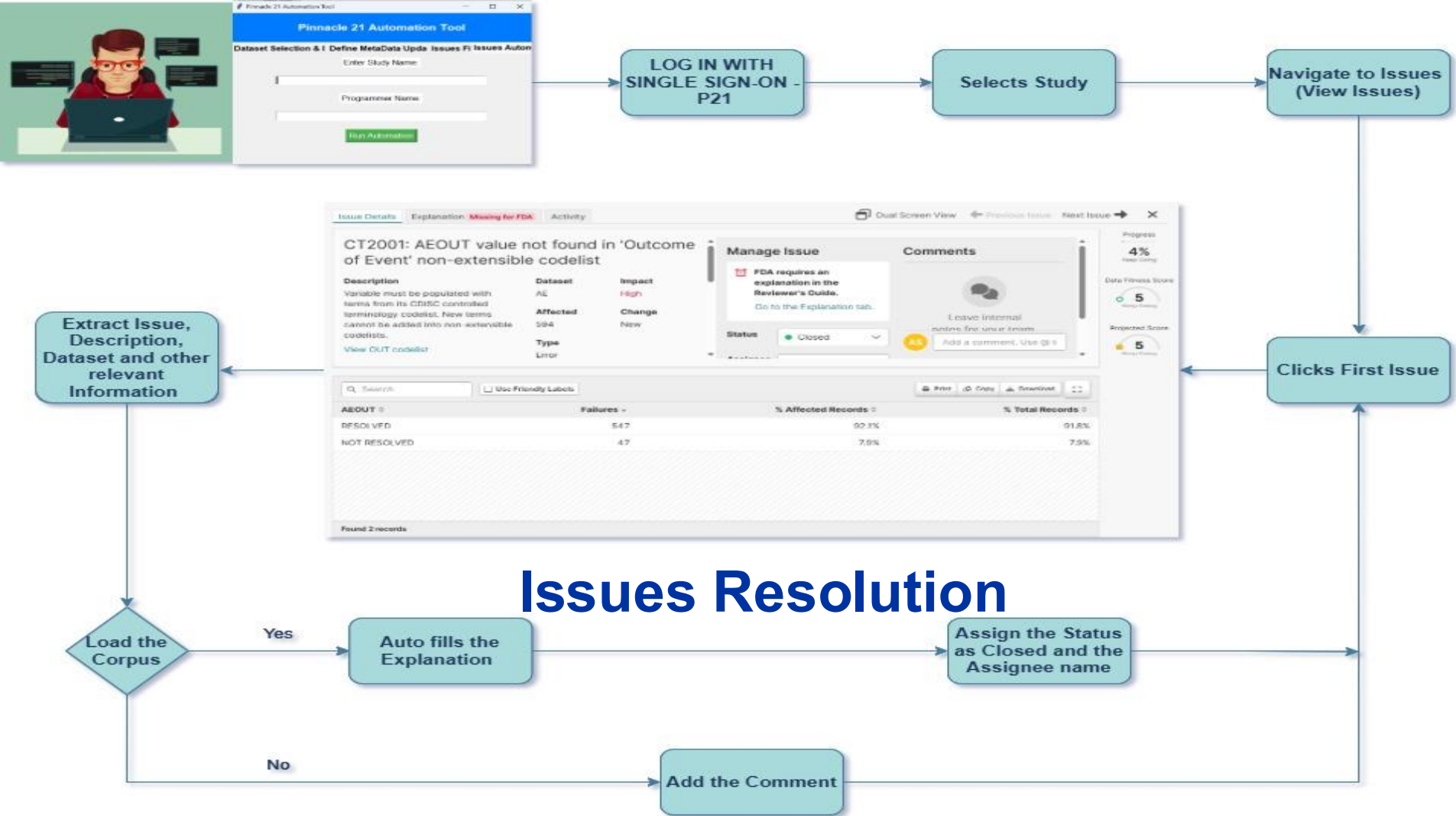
Upload Reference Excel file:

Upload Reference File

Reference File: Not uploaded

Select Datasets:

Download Selected Datasets



phuse Issues Resolution

SDTM

Project Managers

VK

RECOMMENDED TASKS

Commit 2 Reject Issues to Fix Now or Fix Later

Show More

+40 proj. points

Export Report

List ViewSplit View

Actions

Search

Status

Assignee

Source

Impact

Dataset

Explained

Change

Group By

Rule

Message

Affected

Change

Impact

Status

AE - Adverse Events

(20 Issues)

CT2001

AEOUT value not found in 'Outcome of Event' non-extensible codelist

Design

594

New

High

Closed

SD0008

Value for AEDECOD not found in MedDRA dictionary

DataMappingMetadata

46

New

High

Open

SD0021

Missing End Time-Point value

47

New

High

Closed

SD0058

Variable appears in dataset, but is not in SDTM model

Mapping

1

New

High

Open

SD1093

AEENDY variable value is imputed

Mapping

47

New

High

Open

SD1132

AESER is not 'Y'

Data

1

New

High

Open

SD1347

AEENDTC for fatal AE is not the same as DM.DTHDTC

DataMapping

2

New

High

Open

SD0038

Value of AEENDY Study Day variable equals 0

Mapping

47

New

Medium

Open

SD0055

SDTM/dataset variable type mismatch

Mapping

1

New

Medium

Open

SD0057

SDTM Expected variable AEBDSYCD not found

Mapping

1

New

Medium

Open

SD0057

SDTM Expected variable AEHLGT not found

Mapping

1

New

Medium

Open

SD0057

SDTM Expected variable AEHLGTCD not found

Mapping

1

New

Medium

Open

SD0057

SDTM Expected variable AEHLT not found

Mapping

1

New

Medium

Open

SD0057

SDTM Expected variable AEHLTCD not found

Mapping

1

New

Medium

Open

SD0057

SDTM Expected variable AELLT not found

Mapping

1

New

Medium

Open

SD0057

SDTM Expected variable AELLTCD not found

Mapping

1

New

Medium

Open

Found 63 records

Issue Handling!



Inconsistent Data Handling



Increased Workload on Programmers



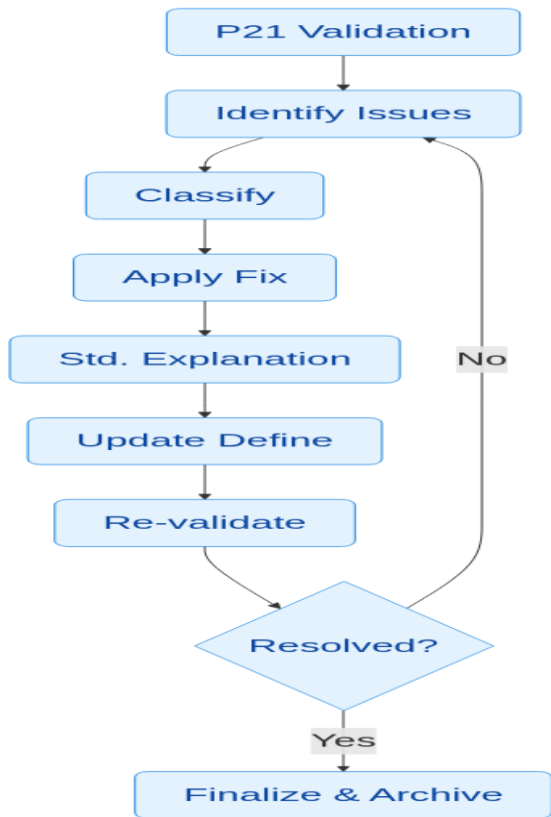
Difficulty in Ensuring Consistent Standards



Increased Time Consumption



Issues Resolution in Define - GenAI



Home > eSub Team Test > PH2026 > SDTM > Define.xml

Define

Properties Datasets Variables Value Level Codelists Terms Methods Comments **Issues 4** Versions

Search

Print Copy Download

RuleId	Variables	Values	Message	Description	Type	Actions
OD0082	Codelist ID	ETCD	Codelist 'ETCD' is missing Decoded Values	When Codelist contains at least one item with Term and Decoded Value (CodeListItem), then all other items must be of the same type.	Error	Not available
DD0021	Standard OID, Standard Type, Standard Name	SDTM IG 3.3, IG, SDTM IG	Invalid Standard Name value 'SDTM IG'	Allowable values of Standard Name are: Values from the STDNAM Codelist in Define-CT for Define-XML v2.1, 'SDTM-IG', 'SEND-IG', 'SEND-IG-AR', 'SEND-IG-DART', 'SEND-IG-GENETOX', 'ADaM-IG', or 'BIMO' for Define-XML v2.0, or 'CDISC SDTM', 'CDISC SEND', or 'CDISC ADaM' for Define-XML v1.0. Values must conform to this list of values. Note: Refer to the data standards catalog by the relevant Regulatory Agency for acceptance of standards and their versions.	Error	Not available
DD0150			Missing expected Standard	Invalid combination of def:Standard elements found in a v2.1 define.xml file. For SEND, at least two Standards must be defined: (1) where Type='IG' and Name begins with 'SENDIG' and (2) where Type='CT' and PublishingSet='SEND'. For SDTM, at least two Standards must be defined: (1) where Type='IG' and Name='SDTMIG' and (2) where Type='CT' and PublishingSet='SDTM'. For ADaM, at least two Standards must be defined: (1) where Type='IG' and Name='ADaMIG', and (2) where Type='CT' and PublishingSet='SDTM' or where Type='CT' and PublishingSet='ADaM'.	Error	Not available
DD0138	Standard OID, Standard Type, Standard Name	SDTM IG 3.3, IG, SDTM IG	Referenced Standard is missing	The referenced standard (name and version), is not valid or could not be found.	Error	Not available

Found 4 records

Save



Creation of Define

Add a New Define

Choose an option below. You can always change your mind or make edits later. [Learn about Define options](#)



Time-saver

From Scanned Data

We already scanned your uploaded datasets to **auto-fill for you ~80% of define.xml**. You can choose which variables to make Value Level Metadata for.

Use validation metadata



Import file

Upload content from *outside* P21E. Choose an Excel specification in P21E format, or an actual define.xml file.

Import XLSX

Import XML



Copy from

Reuse content from *inside* P21E. Choose a Define from another study's Data Package, or a copy of your company's Custom Standard.

Copy Study

Copy Standard

or, [start with a blank template](#)



Define - Standards

Pinnacle 21 Automation Tool

Pinnacle 21 Automation Tool

Dataset Selection & Download Define MetaData Updates (Excel Issues Fixing Issues Automation

Upload Reference Excel file:

Upload Reference File

Reference File: Not uploaded

Select Datasets:

AE
CM
CO
DM
DS
DV
EC
EG
EX
LB

Download Selected Datasets

phuse Define - Generated P21



Pinnacle 21 Automation Tool

Dataset Selection & Download Define MetaData Updates (Excel) Issues Fixing Issues Automation

Select Dataset type (SDTM/ADAM):

Select Terminology Version:

Select Standard Version:

Upload Reference Excel file:

Reference File: Not uploaded

Upload Study Define Excel file:

Study Define File: Not uploaded

Enter Dataset Name to Delete:

(Optional) Leave blank if you don't want to delete any dataset.

Select Dataset Type (SDTM/ADAM)

Select Terminology Version

Select Standard Version

Upload Reference Define.xlsx File

Upload Study Define.xlsx File

Optional: Delete Dataset

Updates Tab:
Dataset, Variables,
Methods, Comments,
ValueLevel, Codelists

Study Define.xlsx
gets updated



Define - Generated P21

Home > eSub Team Test > PH2026 > SDTM > Define.xml

Define

Actions

Properties Datasets Variables Value Level Codelists Terms Methods Comments Issues 18 Versions

Search

Clear

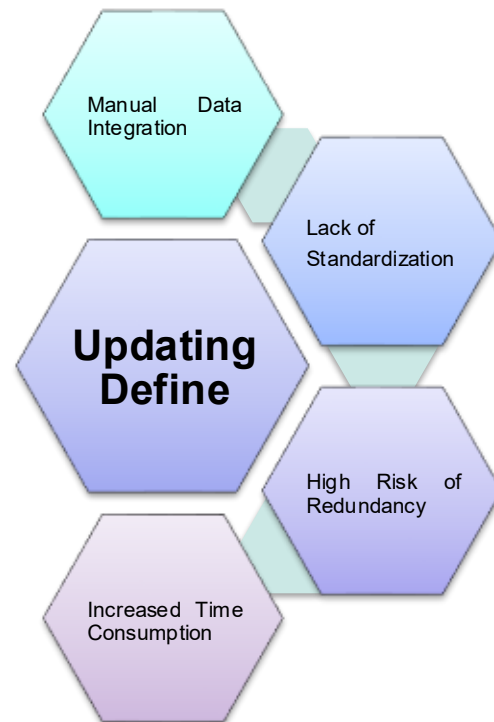
Learn about Variable columns

Columns

Dataset	Variable	Label	Type	Length	Digits	Mandatory	Assigned Value	Codelist	Origin	Source
LB	STUDYID	Study Identifier	text	12		Yes			Protocol	Sponsor
LB	DOMAIN	Domain Abbreviation	text	2		Yes			Assigned	Sponsor
LB	USUBJID	Unique Subject Identifier	text	11		Yes			Derived	Sponsor
LB	LBSEQ	Sequence Number	integer	3		Yes			Derived	Sponsor
LB	LBTESTCD	Lab Test or Examination Short Name	text	7		Yes		LBTESTCD	Assigned	Sponsor
LB	LBTEST	Lab Test or Examination Name	text	39		Yes		LBTEST	Assigned	Sponsor
LB	LBCAT	Category for Lab Test	text	10		No		LBCAT	Assigned	Sponsor
LB	LBORRES	Result or Finding in Original Units	text	5		Yes			Collected	Investigator
LB	LBORRESU	Original Units	text	8		Yes		LBORRESU	Collected	Investigator
LB	LBORNLO	Reference Range Lower Limit in Orig Unit	text	5		No			Collected	Investigator
LB	LBORNRI	Reference Range Upper Limit in Orig Unit	text	5		No			Collected	Investigator
LB	LBSTRESC	Character Result/Finding in Std Format	text	8		Yes			Derived	Sponsor
LB	LBSTRESN	Numeric Result/Finding in Standard Units	float	9	5	No			Derived	Sponsor
LB	LBSTRESU	Standard Units	text	8		No		LBSTRESU	Derived	Sponsor
LB	LBSTNRLO	Reference Range Lower Limit-Std Units	float	6	3	No			Derived	Sponsor
LB	LBSTNRHI	Reference Range Upper Limit-Std Units	float	6	3	No			Derived	Sponsor
LB	LBNRIND	Reference Range Indicator	text	8		No		LBNRIND	Assigned	Sponsor
LB	LBBLFL	Baseline Flag	text	1		No		Y	Derived	Sponsor
LB	VISITNUM	Visit Number	float	4	1	Yes		LB.VISITNUM	Assigned	Sponsor
LB	VISIT	Visit Name	text	19		Yes			Assigned	Sponsor
LB	VISITDY	Planned Study Day of Visit	integer	3		No			Protocol	Sponsor
LB	LBDTC	Date/Time of Specimen Collection	partialDate			Yes			Collected	Vendor

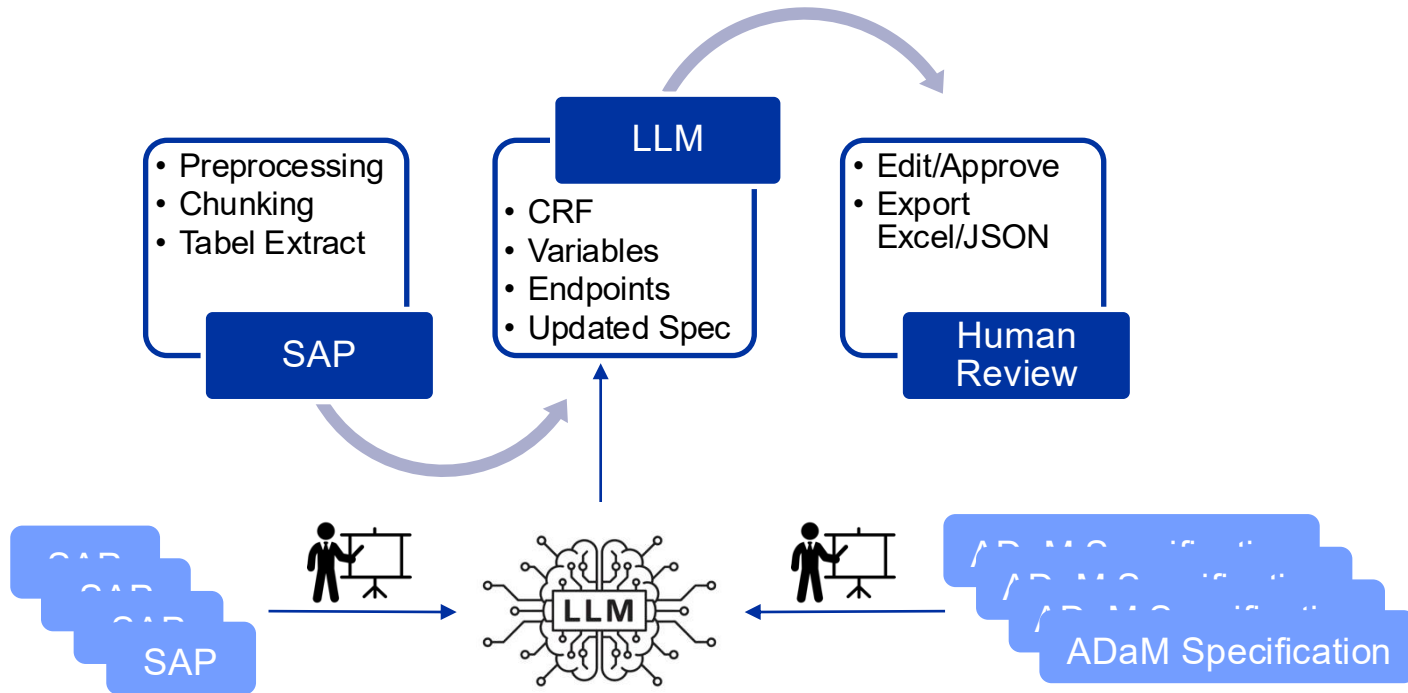
Found 258 records

Save





New Endpoint Specification - GenAI





Future Enhancements



Expand Automation using GenAI

Implement solutions to resolve more complex issues.



Multi-Condition Comments

Streamlined input for efficient data management tasks.



Automation of Fixes

Automatically resolve additional issues in Define.xml.



User Customization

Improved reporting features tailored to user needs.



Conclusion



Efficient SDTM/ADAM Issue Resolution



Automated Metadata Management and Fixing Issues in Define



Minimized Errors and Redundancy



Scalable for adding personalized comments for Issues



Future – Proof and Customizable



Reference

- <https://github.com/phuse-org/phuse-scripts/tree/master/data/sdtm/cdiscpilot01>
- <https://github.com/phuse-org/phuse-scripts/tree/master/data/adam/cdiscpilot01>
- Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. Transformers: State-of-the-art natural language processing. EMNLP: System Demonstrations. 2020.
- Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O, et al. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. ACL. 2020.
- M. Lewis et al., "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," in Proceedings of ACL, 2020.

Acknowledgement

- ❖ Priya Venkatesan Iyer
- ❖ Periasamy Karuppannan
- ❖ Surabhi Sundar
- ❖ Mayur Mishra
- ❖ Abhinav Srivastava



**The Global Healthcare
Data Science Community**

Contact Channels

- 📞 UK +44 1843 609600
- ✉ office@phuse.global
- 🌐 phuse.global

Social Media

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