

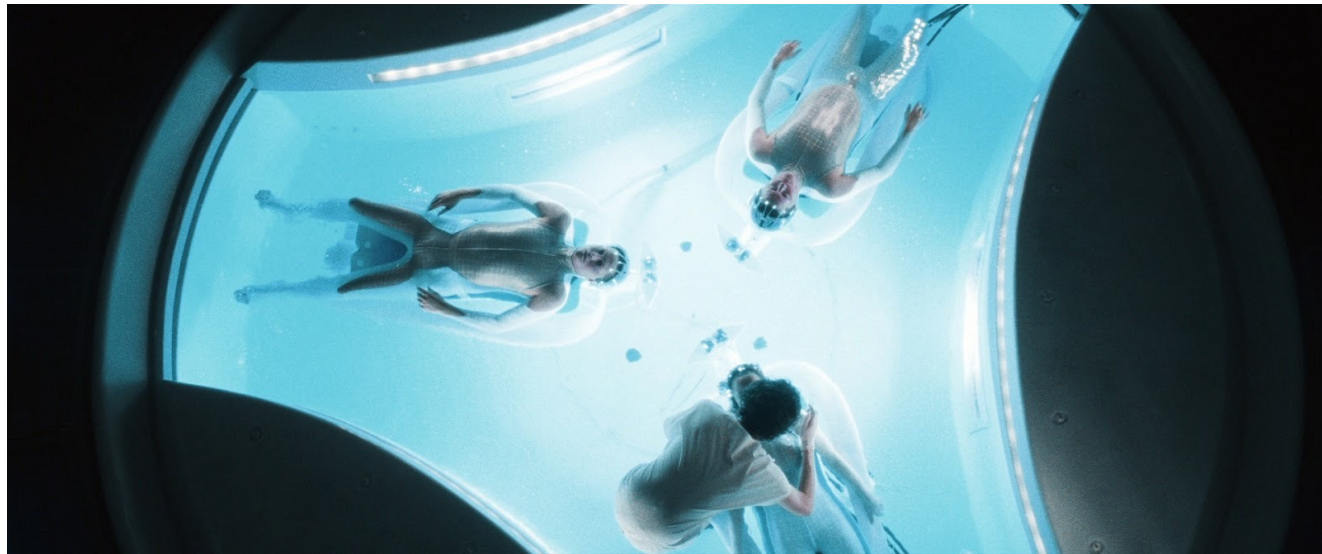
Single Day Event

Roche Automated eSub Translation

Joe Zhu, Xiaoli Duan
Roche



Guess what movie is this





Guess what movie is this

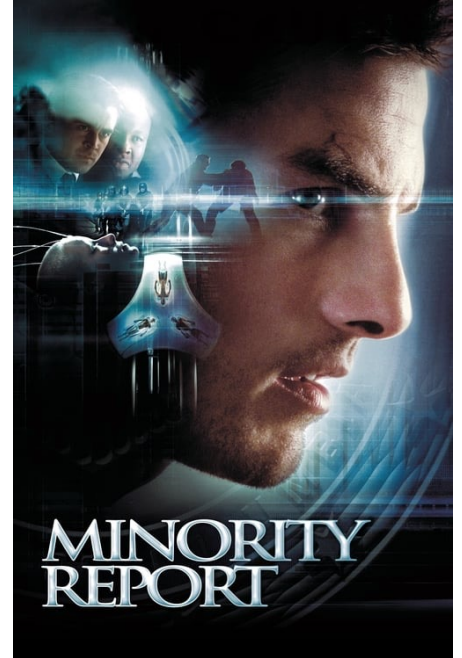
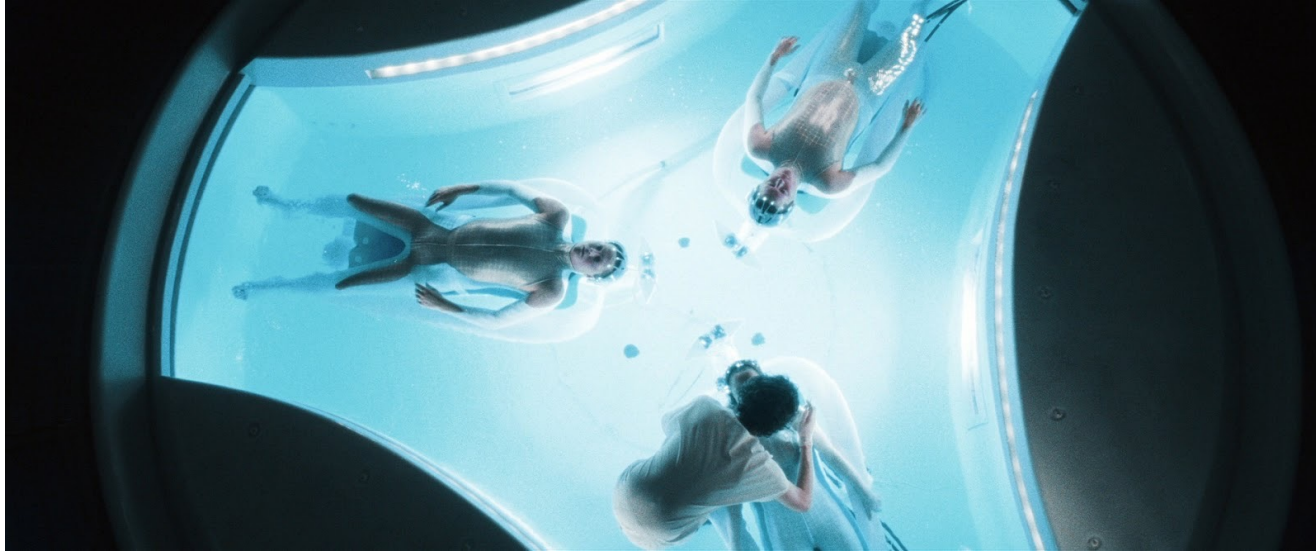


Table of contents

Context and motivations

Our goals

Three automation improvements:

1. Smart reviewing process via TOMATO
2. Source database version-controlled
3. Blank CRF translation solution - wave

New workflow and WHY?

Summary and future works

Acknowledgement

Context - China CDE translation requirements

FDA

- 1 DAP M1-M3, Protocol
- 2 SDTM annotated CRF
- 3 SDTM Define.xml
- 4 SDTM SAS xpt's
 - Conformance to FDA validation rules
- 5 SDTM Reviewers Guide (SDRG)
 - Explanation of the validation results against FDA validation rules
- 6 ADaM Define.xml
- 7 ADaM SAS xpt's
 - Conformance to FDA validation rules
- 8 SAS programs
- 9 ADaM Reviewers Guide (ADRG)
- 10 Patient Level eCRF
- 11 Patient Narratives
- 12

PMDA

- 1 DAP M1-M3, Protocol
- 2 SDTM annotated CRF
- 3 SDTM Define.xml
- 4 SDTM SAS xpt's
 - WHO Drug must be used for drug coding
 - The use of SI units is recommended for standardized result
 - Conformance to PMDA validation rules
- 5 SDTM Reviewers Guide (SDRG)
 - Unit conversion factors used
 - Explanation of the validation results against PMDA validation rules
- 6 ADaM Define.xml
- 7 ADaM SAS xpt's
 - Conformance to PMDA validation rules
- 8 SAS programs or analysis algorithm specs
- 9 ADaM Reviewers Guide (ADRG)
 - Explanation of the validation results against PMDA validation rules
- 10 Patient Level eCRF
[not required for NDA. Chugai QA function requests Roche PDQA to share "Patient Level eCRF" in the process of preparation for PMDA compliance inspection]
- 11 Patient Narratives
[not required separately from CSR. Medical writer writes "Patient Narratives" referring the description in CSR.]
- 12 Analysis Results Metadata (ARM)
[strongly recommended]

China CDE

- 1 DAP M1-M3, Protocol
- 2 SDTM annotated CRF
- 3 SDTM Define.xml
- 4 SDTM SAS xpt's
 - Conformance rule is not defined in details yet
 - We will take FDA rule in practice **for now**
- 5 SDTM Reviewers Guide (SDRG)
 - Explanation of the validation results against FDA validation rules **for now**
- 6 ADaM Define.xml
- 7 ADaM SAS xpt's
 - Conformance rule is not defined in details yet
 - We will take FDA rule in practice **for now**
- 8 SAS programs: readable
- 9 ADaM Reviewers Guide (ADRG)
- 10 Patient Level eCRF
- 11 Patient Narratives
- 12 **Translation Required**

Current translation solution and bottlenecks

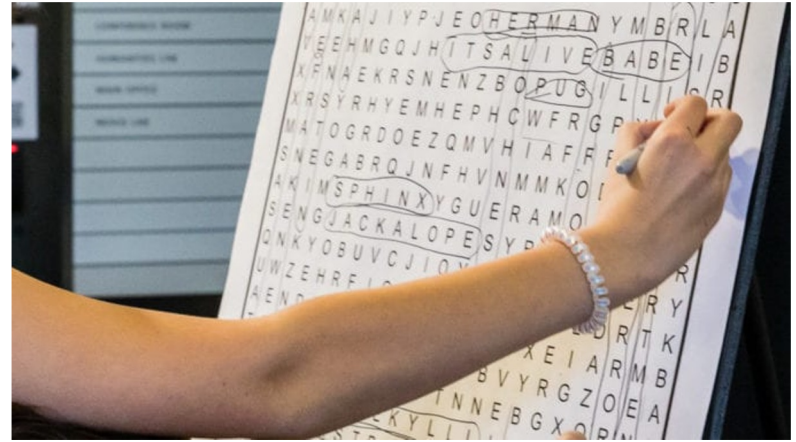
Overall translation + review (most time spent)

1.aCRF

2. Define files

1. SDTM
2. ADAM

3. Reviews guide



Bottlenecks:

1. Package preparation longer than necessary
2. Repeated reviewing process and duplicated work
3. Support limits to standard studies
4. Lack of automation and no cross-study translation evaluation

1. Smart reviewing process

Current workflow is not able to distinguish between **qualified** translation and **bad** translation.

Triage: being the worst makes you first.



bias?

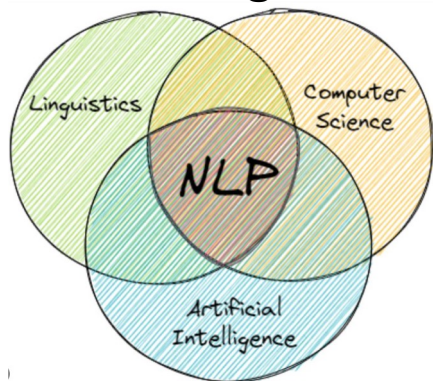
Group shared agreement: the candidate has the most supporters gets its way.



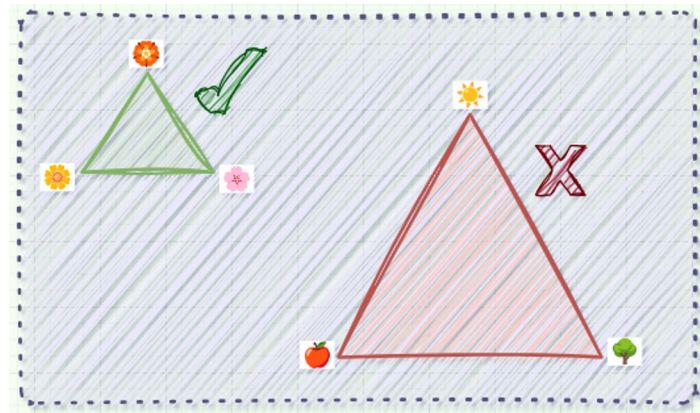
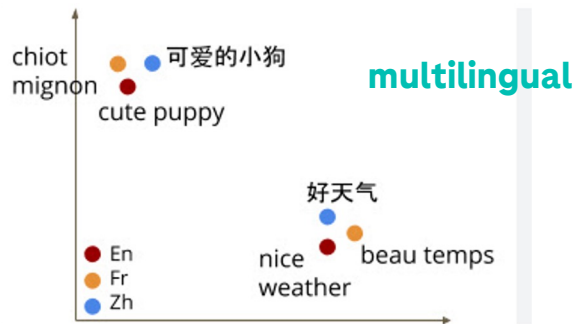
Triage of machine automated translation outcome (TOMATO)



Semantic agreement via NLP similarity measures



interdisciplinary



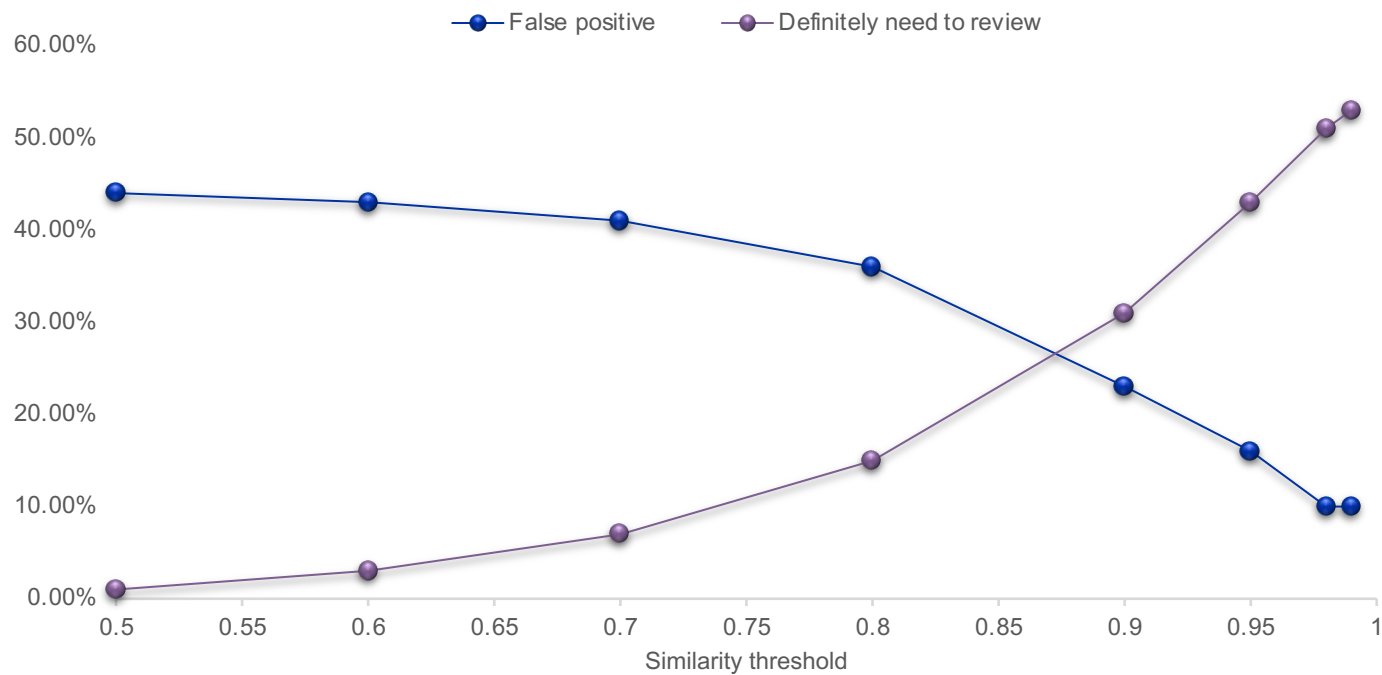
similarity vs distance

Review Interface				
EN	E1	E2	E3	Qualified?
flowers				
apple				
...	...	...	...	...

light weight on review

heavy weight on review

Choice of similarity threshold



2. Source database version-controlled

Version-controlled database := **benchmark GDSR** + **CDISC IG (SDTM + ADaM)** + **differences between two consecutive versions of GDSR** + cumulative knowledge from studies + ...,

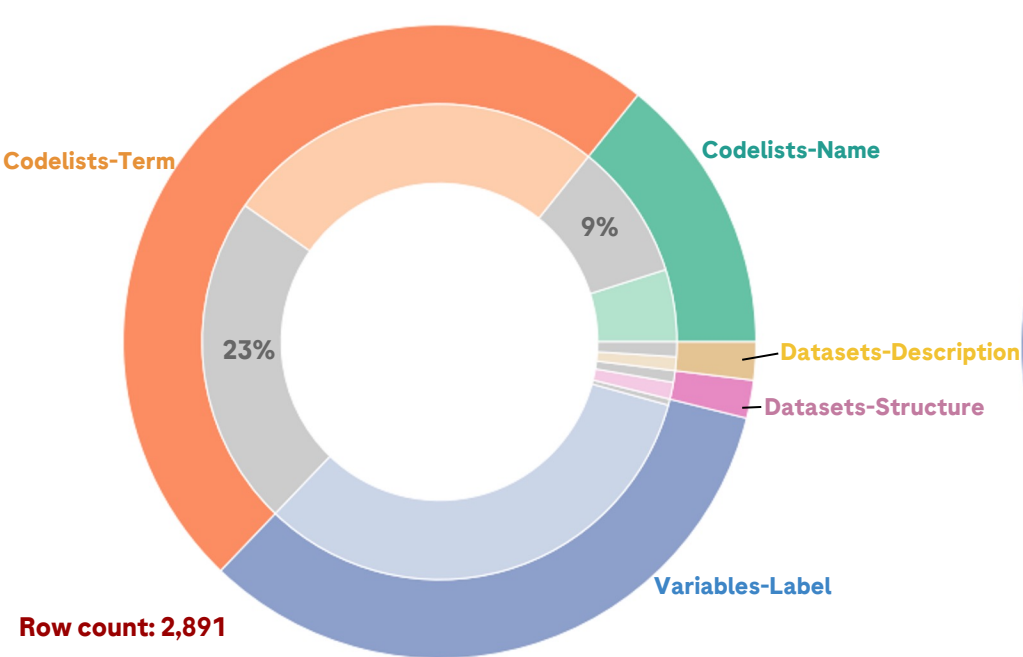
where **differences between two consecutive versions of GDSR** = *what's been “+” to the current version* - *what's been “-” from the previous version*.

Initial DB snapshot	# of entries	Newly added DB snapshots	# of entries
2022-Sep-Equinox	67,753	2022-Dec-ChicoMendes_to_2022-Sep-Equinox	2,236
CDISC IG (SDTMIG v3.2 + ADaMIG v1.1 + ADaM OCCDS v1.0)	1,494	2023-Mar-Janus_to_2022-Dec-ChicoMendes	8,887
		2023-Apr-Madiba_to_2023-Mar-Janus	916
		2023-Jun-LetsTalk_to_2023-Apr-Madiba	1,303
...	...	...	...
Total: 82,589 (as of September 2023)			

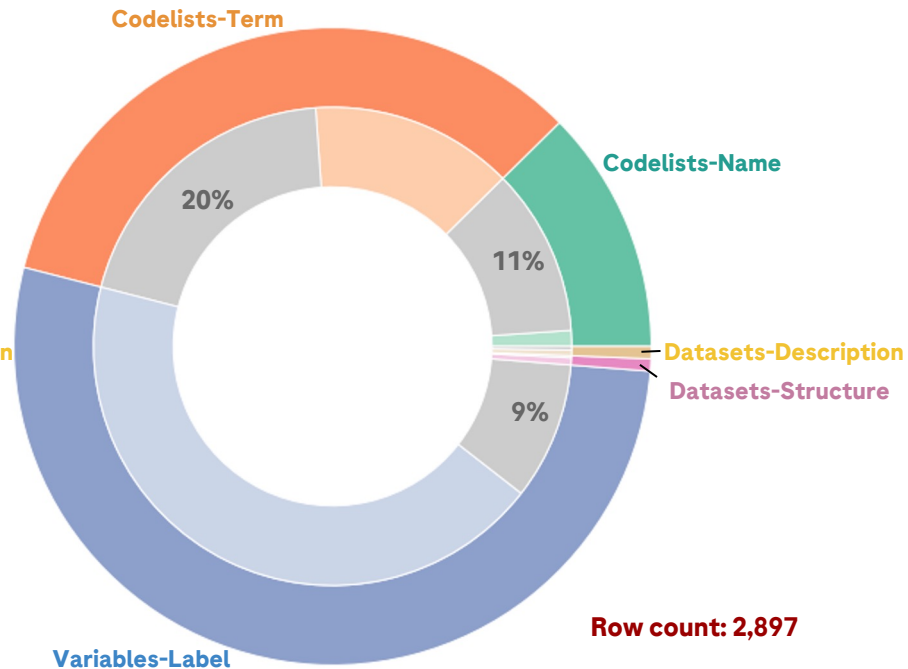
XXXXXXX study SDTM & ADaM spec coverage

> 65% content translation can be reused.

SDTM Coverage



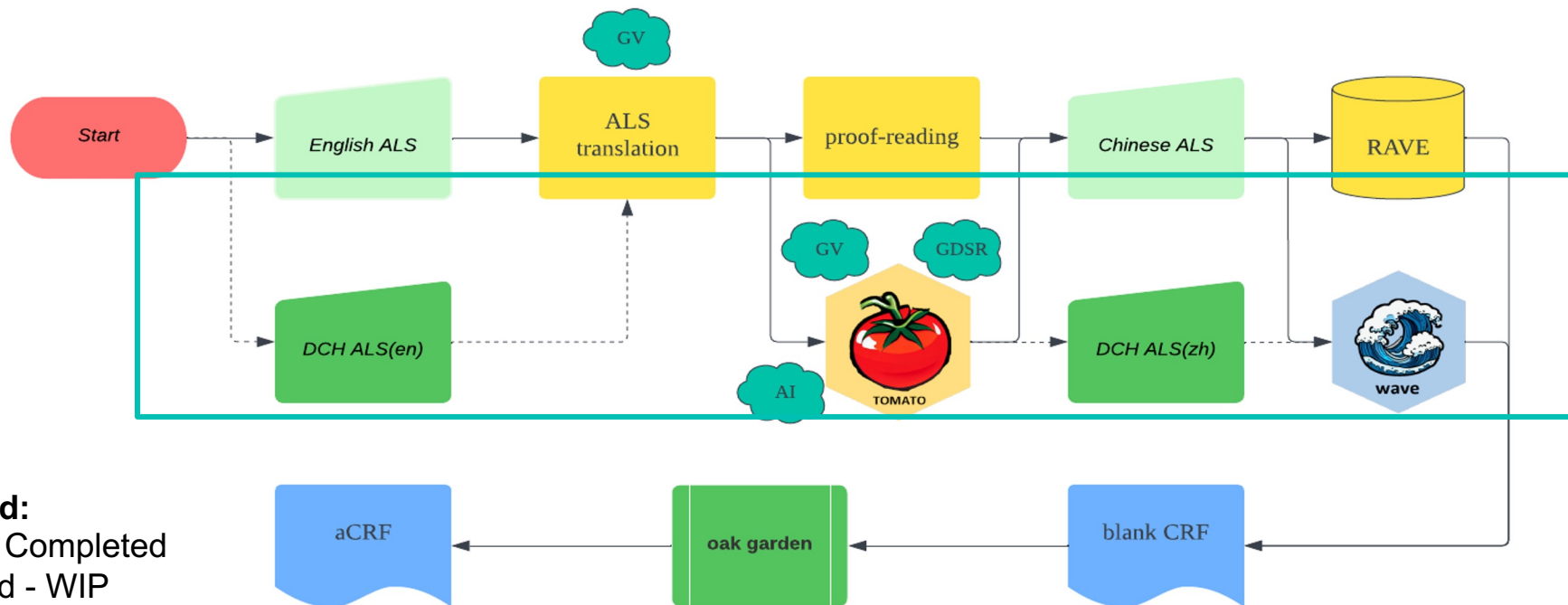
ADaM Coverage



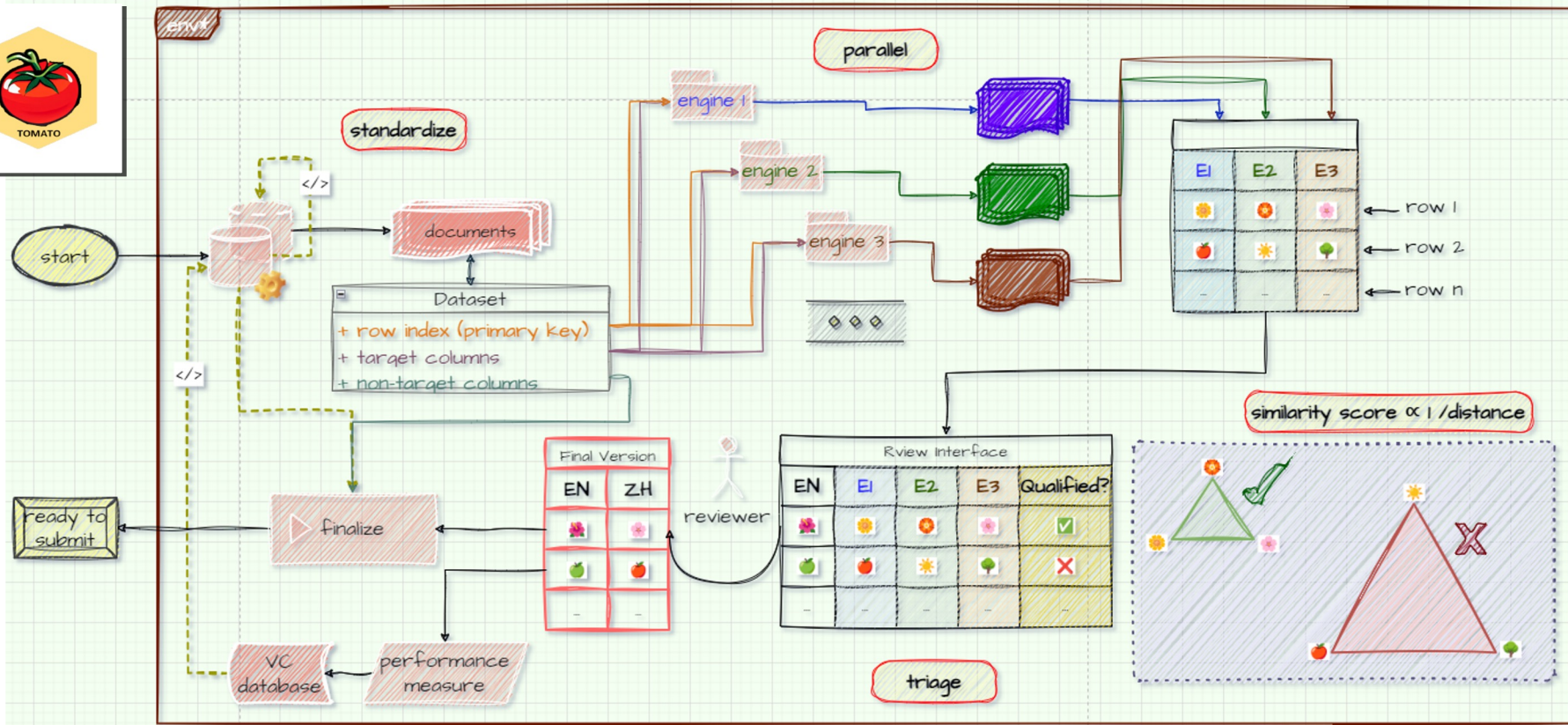
Outer - total rows per column (color-coded)

Inner - colored: covered by GDSR | gray: yet to be covered

3. Blank CRF translation: ALS translation, TOMATO & wave



New workflow



Can what we proposed be trusted?

Glance over computational analysis and study dry-run test

- **Reproducible** environment - OCEAN docker ready to use (reuse OCEAN training materials)
- **High** computational performance: 16 CPUs and 64G memory
- Prepare translation in a **central** place (version controlled)
- Effortless one-liner **bash** scripts
- User-friendly review interface (the **only step** manual effort takes place)

Study name	XXXXXXX
# of text segments fed in	22,967
# of translation machines	3 (engine1, engine2, engine3)
Elapsed time (throughout TOMATO pipeline)	<i>Approx.</i> 2hrs on OCEAN
% of low equality	37%

Summary

Require manual
update



Bypass interaction
with authentication

Repetitive translation
among studies



Translate and reuse
GDSR

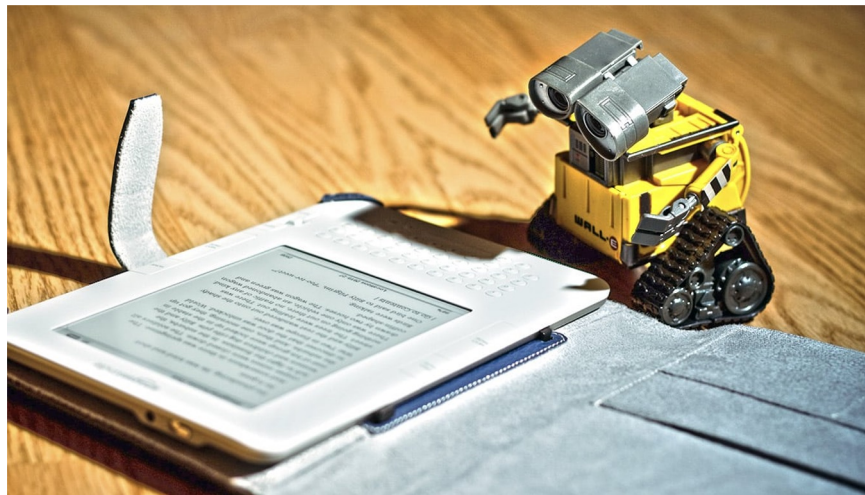
Review process is
time consuming



Automated translation
reviewing process

Future works

- Extend automation to another level – in-house translation machine APIs.
- Align GDSR translation with the upcoming CDISC Chinese translation.
- Improve as we receive user feedback.



Acknowledgement

DSS team: Xiuting Mi, Chenkai Lv and China eSub team

DSD team: Stella Gu, Emily Zhou, Samuel Zheng, Monica Qin, Lei Chen

GDSR and lynx team: Ivan Robinson, Huw Mason, Javier D. Fernandez

IT team: Yi Huang, Chantal Carina Moucouveia, Tomasz Broda, Arturo Moran Gonzalez, Gilles Prévot

And many people who have contributed to the translation and review work

Global team



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)