

# NETRIAS AI-ASSISTED ONTOLOGY HARMONIZATION FOR AKC METADATA

## THE CHALLENGE

Reduce the timing required to correct metadata, which is time consuming when curating manually, taking upwards of 3+ hours for only two target entities (disease and tissue) in large metadata spreadsheets.

### RESULTS SUMMARY

**2.5x** faster curation

**53%** lower cost

Curators stay in control

## APPROACH

Netrias conducted a pilot with three curators from the Adaptive Immune Receptor Repertoire Knowledge Commons (AKC; PMID: 40967183) to quantify the effort required to harmonize **disease\_diagnosis** and **tissue** metadata in the AKC to the Human Disease Ontology (DO) and Uberon, and to evaluate the potential impact of a human-in-the-loop, AI-assisted workflow. Using a shared Google Sheet instrumented with Apps Script timers, curators standardized 480 ontology-backed values (208 disease terms, 272 tissue terms) by mapping raw entries to final, standardized ontology labels and IDs. For each value, we logged the time from initial raw entry to final standardized value, and distinguished between first-seen terms (300 unique raw values) and repeats (180 additional occurrences).

## MANUAL CURATION

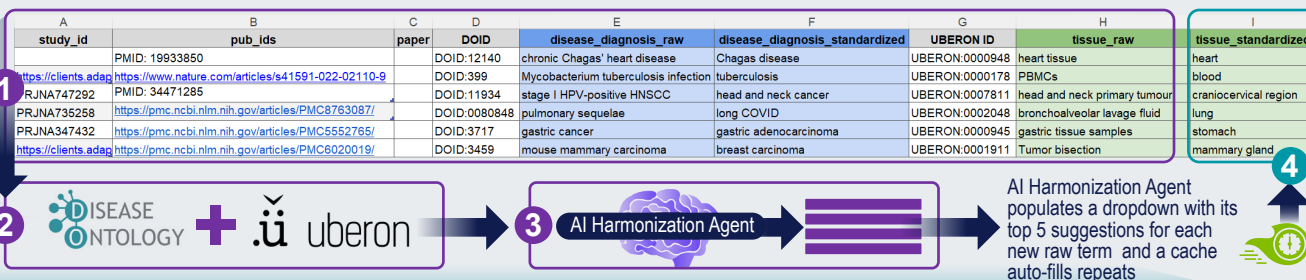
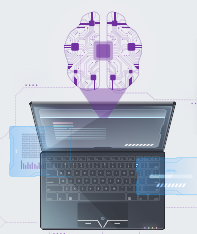
Google Sheets



A	B	C	D	E	F	G	H	I
study_id	pub_ids	paper	DOID	disease_diagnosis_raw	disease_diagnosis_standardized	UBERON ID	tissue_raw	tissue_standardized
	PMID: 19933850		DOID:12140	chronic Chagas' heart disease	Chagas disease	UBERON:0000948	heart tissue	heart
<a href="https://clients.adag">https://clients.adag</a>	<a href="https://www.nature.com/articles/s41591-022-02110-9">https://www.nature.com/articles/s41591-022-02110-9</a>		DOID:399	Mycobacterium tuberculosis infection	tuberculosis	UBERON:0000178	PBMCs	blood
PR_JNA747292	PMID: 34471285		DOID:11934	stage I HPV-positive HNSCC	head and neck cancer	UBERON:0007811	head and neck primary tumour	cranio-cervical region
PR_JNA735258	<a href="https://pmc.ncbi.nlm.nih.gov/articles/PMC8763087/">https://pmc.ncbi.nlm.nih.gov/articles/PMC8763087/</a>		DOID:0080848	pulmonary sequelae	long COVID	UBERON:0002048	bronchoalveolar lavage fluid	lung
PR_JNA347432	<a href="https://pmc.ncbi.nlm.nih.gov/articles/PMC5552785/">https://pmc.ncbi.nlm.nih.gov/articles/PMC5552785/</a>		DOID:3717	gastric cancer	gastric adenocarcinoma	UBERON:0000945	gastric tissue samples	stomach
<a href="https://clients.adag">https://clients.adag</a>	<a href="https://pmc.ncbi.nlm.nih.gov/articles/PMC6020019/">https://pmc.ncbi.nlm.nih.gov/articles/PMC6020019/</a>		DOID:3459	mouse mammary carcinoma	breast carcinoma	UBERON:0001911	Tumor bissection	mammary gland

## AI-ASSISTED WORKFLOW

Google Sheets

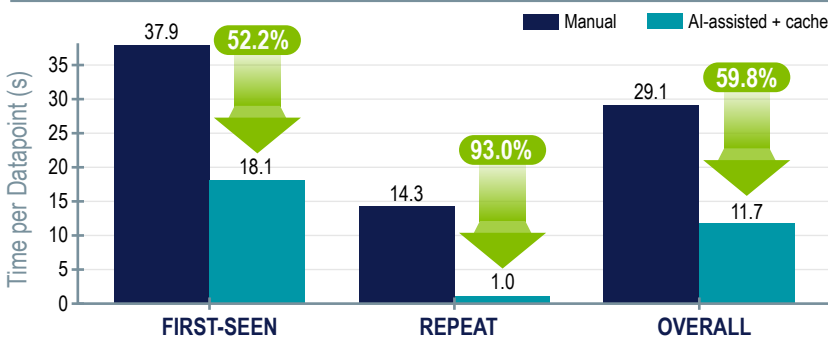


## RESULTS

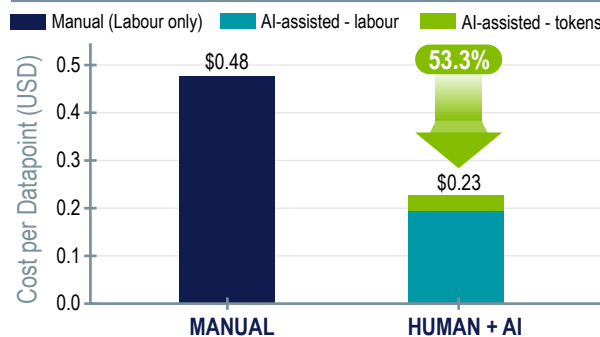
**SPEED & ACCURACY** Manual curation required on average 37.9s for first-seen terms and 14.3s for repeats, yielding an overall mean of 29.1s per datapoint (~124 datapoints/hour). A Netrias AI workflow achieved high AI-assisted candidate quality **2.5x faster**, with the correct standard appearing in the top-5 suggestions for 98% of disease values and 90% of tissue values.

**COST** At a fully loaded curator rate of \$60/hour and assuming \$0.25 per 1M input tokens plus \$2.00 per 1M output token pricing, the faster performance corresponds to a decrease from \$0.48 to \$0.23 per datapoint (**a 53% cost reduction**), with curators retaining full control over final ontology assignments.

### FIRST-SEEN, REPEAT, AND OVERALL CURATION TIME PER DATAPPOINT



### COST PER DATAPPOINT: MANUAL VS HUMAN + AI



Want to cut metadata harmonization time by 2.5x in your existing workflows – without sacrificing curator oversight?

Contact us to schedule a demo and discuss a pilot on your metadata!

[www.netrias.com/request-a-demo/](http://www.netrias.com/request-a-demo/) or [info@netrias.com](mailto:info@netrias.com)