

# Laboratory & Diagnostic Reference Dataset Acquisition Guide

This guide compiles every external dataset and knowledge base referenced in `labs_guide_1.md` and `labs_guide_2.md`.

For each resource you get:

- what it is used for in the ESAO laboratory pipeline,
- the primary link to obtain it,
- the practical access steps (including licensing hurdles), and
- whether the companion script (`scripts/download_reference_datasets.py`) can automate retrieval.

Document each download (file name, version/date, checksum) as you fetch it so future runs remain reproducible.

## Core Code Systems

Dataset	Purpose in project	Access link	Access procedure	Script
LOINC master terminology (incl. parts, panels, accessory files)	Primary vocabulary for lab/imaging codes, components, properties, units, and canonical panel definitions.	<a href="https://loinc.org/downloads/">https://loinc.org/downloads/</a> ( <a href="https://loinc.org/downloads/">https://loinc.org/downloads/</a> )	Create free LOINC account -> sign in -> accept license -> download latest release ZIP (e.g., LOINC_2.81.zip). Keep <code>Loinc.csv</code> , <code>PanelsAndForms.csv</code> , <code>LoincPartLink_*</code> and <code>LoincUnits</code> tables.	Manual (login)
SNOMED CT International RF2	Reference ontology for observations, specimens, procedures; used indirectly via mappings.	<a href="https://www.snomed.org/snomed-ct/get-snomed-ct">https://www.snomed.org/snomed-ct/get-snomed-ct</a> ( <a href="https://www.snomed.org/snomed-ct/get-snomed-ct">https://www.snomed.org/snomed-ct/get-snomed-ct</a> )	Confirm access rights (country membership or affiliate license). Register with SNOMED International or NLM UTS -> accept license -> download latest International Edition RF2 package. Log license constraints.	Manual (licensed)
OMOP Vocabulary (Athena)	Consolidated concept IDs and crosswalks (LOINC <-> SNOMED <-> OMOP) for analytics interoperability.	<a href="https://athena.ohdsi.org/">https://athena.ohdsi.org/</a> ( <a href="https://athena.ohdsi.org/">https://athena.ohdsi.org/</a> )	Register OHDSI Athena account -> sign in -> select vocabularies (LOINC, SNOMED, RxNorm, etc.) -> submit download request -> retrieve emailed link to vocabulary ZIP and unpack <code>CONCEPT</code> , <code>CONCEPT_RELATIONSHIP</code> , <code>VOCABULARY</code> tables.	Manual (login)
HL7 US Core laboratory ValueSet	FHIR value set enumerating allowable lab LOINC codes for US Core profiles.	<a href="https://build.fhir.org/ig/HL7/US-Core/package.tgz">https://build.fhir.org/ig/HL7/US-Core/package.tgz</a> ( <a href="https://build.fhir.org/ig/HL7/US-Core/package.tgz">https://build.fhir.org/ig/HL7/US-Core/package.tgz</a> )	Download <code>package.tgz</code> -> extract -> use <code>package/ValueSet-us-core-observation-lab-codes.json</code> (or bundle expansion) for validation.	Script
LABO (Clinical LABoratory Ontology)	Lightweight ontology for lab test informational entities; supplements LOINC semantics.	<a href="http://purl.obolibrary.org/obo/labo.owl">http://purl.obolibrary.org/obo/labo.owl</a> ( <a href="http://purl.obolibrary.org/obo/labo.owl">http://purl.obolibrary.org/obo/labo.owl</a> )	Fetch OWL via PURL (redirects to GitHub). Record release tag and checksum.	Script

## Panels, Policies & Local Catalogues

Dataset	Purpose in project	Access link	Access procedure	Script
WHO Essential Diagnostics List (2023)	Checklist to confirm coverage of WHO-priority diagnostics (labs & rapid tests).	<a href="https://www.who.int/teams/health-product-and-policy-standards/standards-and-specifications/essential-diagnostics-list">https://www.who.int/teams/health-product-and-policy-standards/standards-and-specifications/essential-diagnostics-list</a> ( <a href="https://www.who.int/teams/health-product-and-policy-standards/standards-and-specifications/essential-diagnostics-list">https://www.who.int/teams/health-product-and-policy-standards/standards-and-specifications/essential-diagnostics-list</a> )	Open web page -> download latest PDF or spreadsheet (via "Downloads" section) -> capture version/date metadata manually.	Manual (web)
UNC i2b2 Common Lab Panels	Real-world definitions of common lab panels for validation of component lists.	<a href="https://tracs.unc.edu/index.php/services/biomedical-informatics/i2b2">https://tracs.unc.edu/index.php/services/biomedical-informatics/i2b2</a> ( <a href="https://tracs.unc.edu/index.php/services/biomedical-informatics/i2b2">https://tracs.unc.edu/index.php/services/biomedical-informatics/i2b2</a> )	Access site -> request panel documentation (public HTML/PDF) or contact UNC TRACS if link is gated. Archive retrieved panel definitions.	Manual (request)

Dataset	Purpose in project	Access link	Access procedure	Script
NHS National Laboratory Medicine Catalogue (NLMC)	SNOMED-coded UK test catalogue; cross-check names and units.	<a href="https://isd.digital.nhs.uk/trud3/user/guest/group/0/pack/26">https://isd.digital.nhs.uk/trud3/user/guest/group/0/pack/26</a> ( <a href="https://isd.digital.nhs.uk/trud3/user/guest/group/0/pack/26">https://isd.digital.nhs.uk/trud3/user/guest/group/0/pack/26</a> ).	Register for NHS TRUD (free) -> add NLMC pack to downloads -> accept license -> download ZIP (NLMC_*.zip).	Manual (login)
Philippine institutional lab catalogs (PGH, JRRMMC, etc.)	Source of local terminology, packages ("Executive Panel"), and specimen details.	e.g., <a href="https://www.pgh.gov.ph/">https://www.pgh.gov.ph/</a> ( <a href="https://www.pgh.gov.ph/">https://www.pgh.gov.ph/</a> ) or hospital FOI portals	Collect PDFs/price lists manually (may require FOI request). Save raw files and note retrieval date/source URL.	Manual (FOI/web)
PhilHealth benefit package circulars	Identify bundled diagnostics mandated in national insurance packages.	<a href="https://www.philhealth.gov.ph/partners/providers/payments/case_rates.html">https://www.philhealth.gov.ph/partners/providers/payments/case_rates.html</a> ( <a href="https://www.philhealth.gov.ph/partners/providers/payments/case_rates.html">https://www.philhealth.gov.ph/partners/providers/payments/case_rates.html</a> )	Download relevant circular PDFs for disease/program packages. Extract included tests and map to standards.	Manual (web)

## Microbiology & Susceptibility References

Dataset	Purpose in project	Access link	Access procedure	Script
NCBI Taxonomy (new_taxdump)	Canonical organism names/IDs for culture & serology mapping.	<a href="https://ftp.ncbi.nih.gov/pub/taxonomy/new_taxdump/new_taxdump.zip">https://ftp.ncbi.nih.gov/pub/taxonomy/new_taxdump/new_taxdump.zip</a> ( <a href="https://ftp.ncbi.nih.gov/pub/taxonomy/new_taxdump/new_taxdump.zip">https://ftp.ncbi.nih.gov/pub/taxonomy/new_taxdump/new_taxdump.zip</a> )	Download ZIP via HTTPS/FTP -> extract names.dmp, nodes.dmp, citations.dmp. Record release notes.	Script
WHONET antimicrobial code list	Standard antibiotic abbreviations used in susceptibility reports.	<a href="https://whonet.org/">https://whonet.org/</a> ( <a href="https://whonet.org/">https://whonet.org/</a> ) (Documents > Antibiotic codes)	Navigate to Documents -> download latest PDF/XLS (may require enabling cookies; CloudFront sometimes blocks automated fetches). Save manually if direct download is blocked.	Manual (web)
ATC/DDD index (antibacterials J01)	Maps antibiotics to ATC classes for grouping.	<a href="https://www.whocc.no/atc_ddd_index/">https://www.whocc.no/atc_ddd_index/</a> ( <a href="https://www.whocc.no/atc_ddd_index/">https://www.whocc.no/atc_ddd_index/</a> )	Use web interface export (CSV/Excel) for J01 (systemic antibacterials). Manual export required; document date and filters used.	Manual (web)

## Imaging & Procedure Standards

Dataset	Purpose in project	Access link	Access procedure	Script
RadLex Playbook / LOINC-RSNA harmonised set	Standard IDs for imaging procedures (modality, body part, view).	<a href="https://loinc.org/download/radlex-playbook/">https://loinc.org/download/radlex-playbook/</a> ( <a href="https://loinc.org/download/radlex-playbook/">https://loinc.org/download/radlex-playbook/</a> )	Sign in with LOINC credentials -> accept RSNA/LOINC terms -> download Playbook CSV.	Manual (login)
DICOM Controlled Terminology (Part 16)	Modality, view, contrast codes for imaging metadata.	<a href="https://dicom.nema.org/medical/dicom/current/output/chtml/part16/PS3.16.html">https://dicom.nema.org/medical/dicom/current/output/chtml/part16/PS3.16.html</a> ( <a href="https://dicom.nema.org/medical/dicom/current/output/chtml/part16/PS3.16.html">https://dicom.nema.org/medical/dicom/current/output/chtml/part16/PS3.16.html</a> )	Access HTML or download ZIP/PDF from DICOM site. No account required. Save relevant tables (CID 29, etc.).	No account required. Save relevant tables (CID 29, etc.).

Dataset	Purpose in project	Access link	Access procedure	Script
WHO International Classification of Health Interventions (ICHI)	Open procedure codes for diagnostics/imaging (Target-Action-Means).	<a href="https://www.who.int/standards/classifications/international-classification-of-health-interventions">https://www.who.int/standards/classifications/international-classification-of-health-interventions</a> ( <a href="https://www.who.int/standards/classifications/international-classification-of-health-interventions">https://www.who.int/standards/classifications/international-classification-of-health-interventions</a> )	Download latest release (Excel/PDF) from WHO page. Manual acceptance of terms may be required.	Manual (w/)
HL7 FHIR ImagingStudy & Procedure specs	Reference for structuring imaging data (fields, expected code systems).	<a href="https://hl7.org/fhir/imagingstudy.html">https://hl7.org/fhir/imagingstudy.html</a> ( <a href="https://hl7.org/fhir/imagingstudy.html">https://hl7.org/fhir/imagingstudy.html</a> )	Download HTML/PDF directly or mirror using script. Keep version tag (currently R5). CPT: purchase/license via AMA Store. HCPCS: download quarterly ZIPs from CMS website. Respect licensing constraints.	Script
CPT / HCPCS	Claims codes for procedures; optional mapping layer.	<a href="https://www.ama-assn.org/practice-management/cpt">https://www.ama-assn.org/practice-management/cpt</a> ( <a href="https://www.ama-assn.org/practice-management/cpt">https://www.ama-assn.org/practice-management/cpt</a> ) / <a href="https://www.cms.gov/medicare/coding/medhcpcsgeninfo">https://www.cms.gov/medicare/coding/medhcpcsgeninfo</a> ( <a href="https://www.cms.gov/medicare/coding/medhcpcsgeninfo">https://www.cms.gov/medicare/coding/medhcpcsgeninfo</a> )		Manual (lic)

## Units, Specimens & Anatomy

Dataset	Purpose in project	Access link	Access procedure	Script
UCUM specification & XML tables	Canonical unit strings and conversions for all quantitative observations.	<a href="https://github.com/ucum-org/ucum">https://github.com/ucum-org/ucum</a> ( <a href="https://github.com/ucum-org/ucum">https://github.com/ucum-org/ucum</a> )	Clone repository or download release ZIP ( <code>ucum-main.zip</code> ). Extract <code>ucum-essence.xml</code> and documentation.	Script
HL7 v2 specimen code system (Tables 0070 & 0487)	Open specimen codes when SNOMED license is unavailable.	<a href="https://terminology.hl7.org/5.5.0/CodeSystem-v2-0487.json">https://terminology.hl7.org/5.5.0/CodeSystem-v2-0487.json</a> ( <a href="https://terminology.hl7.org/5.5.0/CodeSystem-v2-0487.json">https://terminology.hl7.org/5.5.0/CodeSystem-v2-0487.json</a> )	Download JSON directly; optionally also grab <code>CodeSystem-v2-0070.json</code> .	Script
QUDT quantities & units	Semantic annotations for unit dimensions and quantity kinds.	<a href="https://github.com/qudt/qudt-public-repo">https://github.com/qudt/qudt-public-repo</a> ( <a href="https://github.com/qudt/qudt-public-repo">https://github.com/qudt/qudt-public-repo</a> )	Clone or download release ZIP. Use <code>/vocab/quantitykind</code> and <code>/vocab/unit</code> files for reasoning.	Script
Uberon anatomy ontology	Standard anatomy terms for specimen/body site tagging.	<a href="http://purl.obolibrary.org/obo/uberont.owl">http://purl.obolibrary.org/obo/uberont.owl</a> ( <a href="http://purl.obolibrary.org/obo/uberont.owl">http://purl.obolibrary.org/obo/uberont.owl</a> )	Download OWL via PURL (redirects to GitHub release). Track version tag.	Script
OBI (Ontology for Biomedical Investigations)	Method/instrument ontology to tag assays (ELISA, PCR, etc.).	<a href="http://purl.obolibrary.org/obo/obi.owl">http://purl.obolibrary.org/obo/obi.owl</a> ( <a href="http://purl.obolibrary.org/obo/obi.owl">http://purl.obolibrary.org/obo/obi.owl</a> )	Download OWL via PURL (GitHub redirect). Capture release version.	Script

## Cross-Mapping & Synonym Resources

Dataset	Purpose in project	Access link	Access procedure	Script
OMOP concept relationship tables	Ready-made mappings across vocabularies (incl. LOINC <-> SNOMED).	Provided inside Athena vocabulary download	Included in OMOP vocabulary ZIP ( <code>CONCEPT_RELATIONSHIP.csv</code> , etc.). Requires same Athena login steps as above.	Manual (login)
LOINC-SNOMED CT hierarchy (OMOP GitHub)	Supplemental SQL scripts mapping LOINC measurements to SNOMED measurement hierarchy.	<a href="https://github.com/OHDSI/Vocabulary-v5.0">https://github.com/OHDSI/Vocabulary-v5.0</a> ( <a href="https://github.com/OHDSI/Vocabulary-v5.0">https://github.com/OHDSI/Vocabulary-v5.0</a> )	Clone repository or download specific release; review <code>extras/SQL</code> for measurement hierarchy.	Script
UMLS Metathesaurus	Aggregated synonyms and CUIs linking multiple vocabularies.	<a href="https://uts.nlm.nih.gov/uts/">https://uts.nlm.nih.gov/uts/</a> ( <a href="https://uts.nlm.nih.gov/uts/">https://uts.nlm.nih.gov/uts/</a> )	Create UMLS account -> sign annual license -> download Metathesaurus bundle (takes hours). Respect redistribution limits.	Manual (licensed)

## Locale-Specific Policy & Terminology

Resource	Purpose in project	Access link	Access procedure	Script
DOH Philippines clinical laboratory rules (AO 2021-0037 et al.)	Defines required test categories and capabilities; ensures local compliance.	<a href="https://doh.gov.ph/clinical-laboratory">https://doh.gov.ph/clinical-laboratory</a> ( <a href="https://doh.gov.ph/clinical-laboratory">https://doh.gov.ph/clinical-laboratory</a> )	Download accessible PDFs from DOH site. For restricted annexes, submit FOI request. (FOI/web)	Manual

Resource	Purpose in project	Access link	Access procedure	Script
Local terminology & abbreviation glossary	Maps colloquial PH terms (UTZ, ECU, Fecalysis, etc.) to standard descriptors.	Internal curation	Build and maintain CSV manually from hospital catalogs, stakeholder input, and field notes. Store under version control.	Manual (internal)

## Automation Coverage

The Python helper (`scripts/download_reference_datasets.py`) automates downloads only for open files that do not require authentication.

Currently automated: HL7 US Core package, LABO, NCBI Taxonomy, DICOM Part 16 HTML snapshot, HL7 specimen codes, UCUM GitHub ZIP, QUDT repo ZIP, Uberon OWL, OBI OWL, HL7 ImagingStudy spec, OMOP GitHub crosswalk.

Resources flagged as "Manual" or "Manual (licensed)" require interactive steps (accounts, license acceptance, FOI request). Capture those artifacts separately and update the script configuration once programmatic access becomes available.