

Structural alignment of LOINC hierarchy for standardized representation

Background

LOINC is the world’s most widely used terminology standard for health measurements, observations, and clinical documents¹, playing a critical role in medical data analysis. LOINC concepts are organized in a hierarchy based on multiple axes for each term. The resulting multiaxial hierarchy is extensive and well-established, but it suffers from several systemic problems. In selected cases, it allows links between measurements in different body fluids, which is often counterintuitive and confusing (e.g., the LOINC concept 26464-8 ‘Leukocytes [#/Volume] in Blood’ is connected to LP7851-1 ‘Urinalysis’)² (Figure 1). Other issues include the presence of pseudo-hierarchies and hierarchies that terminate at different levels. Since OHDSI Standardized vocabularies inherit the LOINC’s internal structure and hierarchy, the mentioned issues have led to usability challenges and multiple difficulties for researchers.

To address these challenges, we undertook an effort to align hierarchical paths and correct misleading structures. Our goal was to develop a more robust hierarchy that resolves the above-mentioned inconsistencies. “Explore on Demand”, a cohort creation tool, utilizes this curated version of LOINC to streamline the user search experience. This improvement enables more accurate data analysis and enhances the interoperability of healthcare data. By establishing a unified hierarchical approach, we aim to improve the consistency and reliability of health data for clinical decision-making, research, and policy development. We intend to share this work as a contribution to the OHDSI community in the upcoming vocabulary releases.

Methods

We utilized the OMOP-mapped versions of LOINC as a source for alignment and hierarchy development. The vocabulary server infrastructure was used to perform the transformations.³ As a first step, we prioritized codes that are broadly used to represent real clinical events, focusing on the LOINC classes “Lab Test” and “Clinical Observation.” After assembling the raw hierarchy, we introduced a pipeline to identify the longest hierarchy chains, detect mismatches and shared segments across hierarchy paths, and properly position the common components. Next, the LOINC class was repositioned to the top level of the hierarchy to preserve the test category. Finally, duplicate hierarchy paths were removed, retaining a single canonical version.

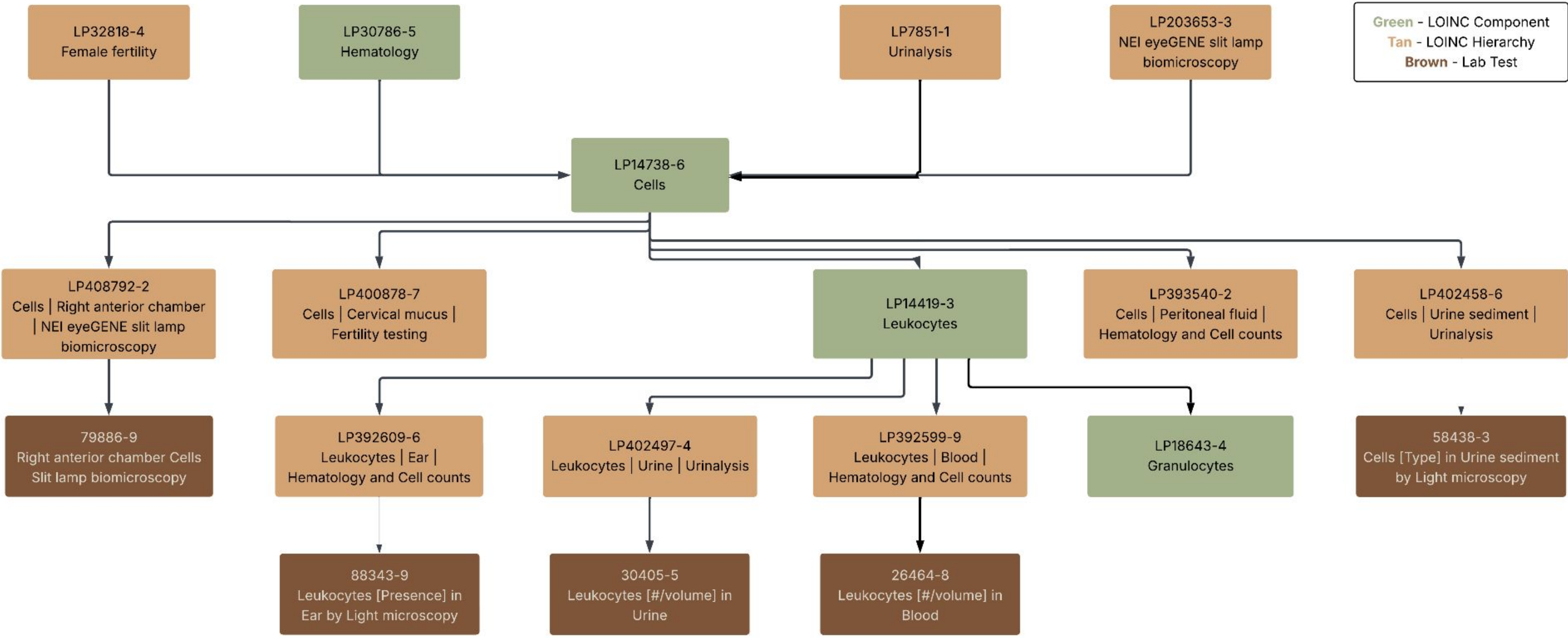


Figure 1. An example of an existing multiaxial hierarchy in LOINC.

Results

A new hierarchy was constructed for 54,114 LOINC codes, spanning a wide range of measurements and representing 59.4% of all Lab Tests and Clinical Observations in LOINC. Initially, all included LOINC codes were associated with at least two hierarchical paths, resulting in 126,283 hierarchy lines. After processing, a single hierarchical path was retained for each code (Figure 2, Figure 3).

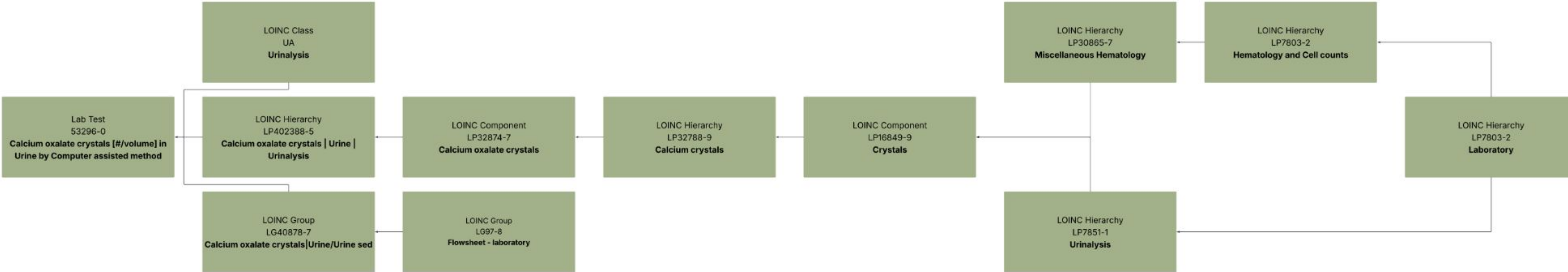


Figure 2. An example of hierarchy path before align

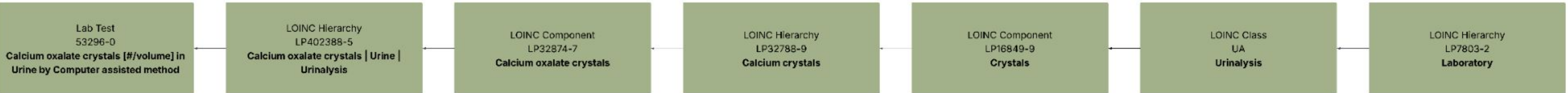


Figure 3. An example of hierarchy path after align

Discussion

The alignment process presented several challenges, primarily due to differences in hierarchical structures, overall complexity, and the presence of incomplete or immature concepts (e.g., “Lab terms not yet categorized”). Despite these challenges, the process successfully aligned 61.2% of LOINC codes related to laboratory tests and clinical observations. However, the remaining codes, as well as codes from other LOINC classes, could not be processed using the same approach. This highlights the need for further methodological improvements. Future work may focus on enhancing automation through the development of more advanced algorithms capable of handling complex alignment cases.