

THE UNDERGROUND LITHOLOGY GRID · AUGUST 2026 · DRILLERDB.COM/RESEARCH

The Underground Lithology Grid is the largest standardized, record-level compilation of driller-described subsurface lithology ever assembled for North America - and the first to apply one lithologic vocabulary across state and provincial boundaries at continental scale. DrillerDB processed more than 10 million water well drillers' logs, including scanned and handwritten pages, filed as public records with state agencies over decades. The dataset is free to academic and institutional researchers for non-commercial use.

KEY FIGURES

53 million+ STANDARDIZED LITHOLOGY INTERVALS	10 million+ WELL LOGS PROCESSED	39 + 6 US STATES + CANADIAN PROVINCES
1.5 billion ft OF DESCRIBED SUBSURFACE (~295,000 MILES - FARTHER THAN THE MOON)	18.4 million+ WELL RECORDS IN THE PARENT DATASET (US + CANADA)	96.5% OF INTERVALS CODED BY A VERSIONED DETERMINISTIC DICTIONARY; 1.7% LEFT UNCODED RATHER THAN GUESSED

Counts are live and still growing; figures verified against production Aug 3, 2026.

WHY IT IS NEW

- Driller logs are the only direct observational record of the shallow subsurface at continental scale. State agencies have preserved them for decades, each in its own format - "blue slate" in one county and "br. clay & gvl" in the next were never comparable across state lines.
- The USGS has begun a federal National Water-Well Database with harmonized lithology; it has publicly released one state (Minnesota) to date.
- The largest previously published standardized well-log compilation covers 1.57 million records (USGS, 2017, regional). The Grid is roughly 33x larger and continental.
- Geologic maps and stratigraphic databases infer the subsurface; every interval in the Grid was observed by a driller standing over the borehole.

HOW IT WAS BUILT

- AI-assisted extraction reads each state's filed well logs, including handwritten and scanned reports.
- 53 million depth intervals parsed; 51.3 million assigned a lithology from one controlled vocabulary. Color and grain size are captured where the driller recorded them (about a third and a tenth of intervals, respectively).
- Coding is overwhelmingly deterministic - 96.5% of intervals are matched by a curated, versioned dictionary, not a language model; 1.7% are left explicitly uncoded rather than guessed. The driller's original words are preserved beside more than 99% of intervals, so every coding decision is checkable at the source.
- In an independent audit of 1,000 randomly sampled coded intervals, the assigned lithology matched the driller's original description in 96% of cases; fewer than 1% were judged outright wrong.

THE RESEARCH PROGRAM

Built for water research of every kind. Academic and institutional researchers get free record-level access for non-commercial work under a one-page data use agreement; the only price is a citation. The dataset supports work on groundwater levels and long-term water-level trends, water scarcity and drought response, well depth and construction trends, groundwater development and supply planning, aquifer characterization, hydrogeology, geothermal siting, and public health for the 40+ million Americans on private wells. Apply at drillerdb.com/research.

ASSETS AVAILABLE

High-resolution stills and animated maps of well, lithology, and water-level coverage; a 60-second cinematic data visualization; founder interviews and custom data cuts for specific regions or story angles - all on request.

ABOUT DRILLERDB

DrillerDB builds business management software for water well drilling contractors and maintains one of the largest databases of water well records in North America. Founded 2024, based in Chicago, Illinois.