ELID — Easily Legible Identifiers for Wine

A Universal Coding System

Rijnard van Tonder, PhD & Certified Sommelier

Version 0.1
January 7, 2025

The wine industry lacks a universal, human-friendly system for precisely identifying wines across databases, applications, and platforms. This document describes a system of unique, compact, and human-readable identification codes for wines. An Easily Legible Identifiers (ELID) accommodates regions, producers, their wines, vintage, and grape varieties. The key principle behind its design is a terse and legible format, fit for both human consumption and automated data processing. It is a system purpose-built for applications to easily identify, cross-reference, analyze, and visualize wine data. ELIDs comprise a first-of-its-kind wine identification system made uniquely possible by leveraging AI assistance.

1 Format and Components

Example. The Easily Legible ID (ELID) for 2012 Dom Perignon Brut from Champagne, France is

FR-CHA-DOMP01-2012

FR — Country Code

2 letters conforming to ISO 3166-1 alpha-2 standard.

FR stands for France.

CHA — Region Code

3 characters describing a region within a country.

CHA stands for Champagne.

The region refers to the origin of the wine. See §2.1 for more details.

DOMP — Producer Code

4 characters describing a producer.

DOMP stands for Dom Perignon.

See §2.2 for more.

01 — Wine Identifier Code

2 characters to uniquely identify distinct wines made by this producer.

`01` is reserved for the Brut Champagne by Dom Perignon.

Wine IDs range from 01–ZZ using an alphanumeric numbering system. See §2.3.

2012 — Vintage / Non-vintage Code

4 characters to identify wine vintage, non-vintage, or non-vintage edition.

`2012` refers to the vintage year.

Non-vintage wines use code `NVXX`. See §2.4 for more details.

Grape varieties. ELIDs permit an optional extension to identify the grape variety or varieties of a wine. For example, Dom Perignon is a blend of Chardonnay and Pinot Noir encoded as

`FR-CHA-DOMP01-2012+CHP`

CHP — Grape Varieties Code

3 characters to identify grape variety or common varieties in a wine.

Indicated by adding a `+` to the ID.

Never required to uniquely identify a wine—it is convenience data for processing. See §2.5.

Key Features

— Compact representation to uniquely identify regions, producers, wines of producers, and vintage.

— Abbreviations are chosen to be phonetically and visually similar to their respective regions and producers.

— Hyphenated components to easily identify wines by ID.

— Consistent length: 15 significant characters for all wines, 18 characters including hyphens.

— Fully alphanumeric, ensuring compatibility across various systems.

— Supports intuitive sorting: primarily by country, then region, then producer and wine, then vintage.

— Systematic approach to encode wine origin with specificity (small sites or broader regions depending on origin).

— Efficient reuse of vintage component to support numbering non-vintage releases.

— Grape variety extension for data and recognition: 3 significant characters, appended to the ID. E.g., `+CAB`.

2 Detailed Code Specifications

2.1 Region Code

— 3-letter codes abbreviate well-established wine regions (BDX for Bordeaux).

— If a wine originates from a specific appellation or subregion, *the most specific code* in the system is used to identify the region. For example, a wine originating from the Stags Leap District takes region STG instead of Napa NAP.

— US state codes and associated wine appellations add a trailing X. For example, a wine from California takes region CAX. This in order to preserve both the existing 2-letter convention for US state codes and a three letter length.

— XXX is reserved for indicating wines at the country level. For example, FR-XXX indicates French table wines, or Vin de Table.

— 3-letter region codes are not universally unique on their own: They are uniquely qualified by country code. For example, DE-MOS refers to Germany's Mosel region, while FR-MOS refers to France's Moselle region.

— A country-region code like US-STG can interpreted as the start and end point of a path through subregions. For example, a data processing tool with knowledge of all regions may resolve US-STG to the region path US-CAX-NAP-STG.

2.2 Producer Code

— Producer codes are automatically generated and emphasize human recognizability and readability. For example, the Champagne producer Krug has producer code KRUG. See §3.1 for more.

— 4-letter producer codes are not universally unique on their own: They are uniquely qualified by country code. For example, KRUG is an abbreviation for distinct producers in different countries:

FR-...-KRUG for Krug Champagne, France.

AT-...-KRUG for Weingut Krug, Austria.

ZA-...-KRUG for Krug Family Wines, South Africa.

— If a producer's name is shorter than 4 letters, trailing X's are used.

2.3 Wine ID

— The Wine Identifier uses an intuitive two-digit numbering system to uniquely represent the wines of a given producer. The range starts at 01 through 99.

— Most producers produce fewer than 99 distinct wine according to the LWIN dataset [1] and requires only two numeric digits. However, some producers (e.g., Louis Jadot) produce more than 99 unique wines.

— When a producer produces more than 99 wines, the numbering continues using alphanumeric ranges according to the following sequences:

`0A, 0B, 0C ... 0X, 0Y, 0Z,`
`1A, 1B, 1C ... 1X, 1Y, 1Z,`
⋮
`9A, 9B, 9C ... 9X, 9Y, 9Z,`
	260 additional combinations (10 digits × 26 letters)

`A0, A1, A2 ... A7, A8, A9,`
`B0, B1, B2 ... B7, B8, B9,`
⋮
`Z0, Z1, Z2 ... Z7, Z8, Z9,`
	260 additional combinations (26 letters × 10 digits)

`AA, AB, AC ... AX, AY, AZ,`
`BA, BB, BC ... BX, BY, BZ,`
⋮
`ZA, ZB, ZC ... ZX, ZY, ZZ`
	676 additional combinations (26 × 26 letters)

— The alphanumeric encoding allows for 1,295 unique wines per producer while keeping the ID length at two characters for brevity, instead of three. This scheme is sufficient since all producers in the LWIN dataset produce fewer than 1,000 unique wines [1] and is likely to remain that way in perpetuity.

2.4 Vintage / Non-vintage Code

— `NVXX` is used for standard non-vintage wines without any known edition.

— `N171` is used for a non-vintage wine with a known release edition, such as Krug Champagne 171ème Édition. The encoding starts with `N` followed by 3 alphanumeric characters for the edition. When the edition cannot be accomodated due to length or format, the standard non-vintage encoding `NVXX` is used.

— Editions are pertinent to the notion of vintage because even though such wines don't state an explicit vintage, an edition suggest advancement of time and release cadence (e.g., an annual cadence for Krug Champagne). Another example is non-vintage Cappellano Barolo Chinato that references three digit edition like `LK120` based on a 2015 vintage.

— `XXXX` is a placeholder code used when the vintage of a wine is not specified or known. E.g., `FR-CHA-DOMP01-XXXX` refers to a Dom Perignon Brut without specifying the vintage.

2.5 Grape Varieties Code

— Encoding grape varieties enriches IDs directly for common use cases:

Ease of data operations such as search, filtering, and aggregation.

Provides immediate variety recognition for consumers. E.g., `+SBL` shows Sauvignon Blanc wines.

Documents both single-variety and traditional blends. E.g., `+BDB` for Blanc de Blancs (100% Chardonnay).

— Optional because a wine can be uniquely identified by ID without supplemental information about grape varieties. IDs that do not specify grape varieties may be looked up in secondary information sources.

— See Appendix B for common codes.

3 Technical Implementation and Details

3.1 AI-assisted Code Generation Method

ELID is a first-of-its-kind wine identification system that is made uniquely possible by AI assistance. It addresses the primary challenge of ensuring IDs are *succinct enough to be human-recognizable* yet expressive enough to *encode the breadth of all essential wine information*.

The core of encoding essential wine information lies in the unique assignment of producers and wine regions. It is known that there are at least 37,000 wine producers and up to 14,000 from France alone. It would be an onerous task to manually assign unique abbreviations to all producers. At the same time, it is important that an automated approach preserves phonetically and representative "human-like" abbreviations for legibility. AI was leveraged to help achieve the task: given a set of producers and regions, generate unique abbreviations for human readability.

Large Language Models (LLMs) `Claude Sonnet` and `GPT-4o mini` were used generate initial producer and region codes. The input prompt provides guidelines for code generation (human readability, phonetic resemblance). Producers were processed in order by the number of wines they produce, such that more "obvious" or favorable abbreviations are initially assigned to producers that are likely to be well-known. LLM generation is best-effort: it produced the bulk of unique codes, but was still prone to generating duplicate or short abbreviations due to the size of the inputs. To create truly unique codes across the data, remaining duplicates in the set of AI-generated codes were post-processed programmatically using various strategies of letter combinations to guarantee uniqueness.

The coding system remains flexible: In future versions, codes can be manually modified to improve clarity and better align with producer preferences, when reasonable to do so.

3.2 ID Pattern Matching: Rules and Limits

— All IDs contain 15 significant characters and maintain a consistent length.

— Alphabetic characters are capitalized.

— 2 alphabetic characters for country codes. There are 249 recognized country codes according to the ISO 3166-1 alpha-2 standard.

— 3 alphabetic characters for regions. At most 17,576 possibilities per country.

— 4 alphanumeric characters for producer codes. At most 1,679,616 possibilities per country.

— 2 alphanumeric characters for wine IDs. At most 1,295 possibilities with a hybrid alphanumeric encoding as described in §2.3.

— 4 alphanumeric characters for vintage / non-vintage.

If vintage, 4 numeric characters.

If non-vintage without edition, the alphabetic constant `NVXX`.

If non-vintage with edition, `N` followed by 3 alphanumeric characters.

Components of a valid ID may be matched by the regular expression [3]:

```
{[A-Z]{2}-[A-Z]{3}-[A-Z0-9]{4}[A-Z0-9]{2}-[0-9]{4}|XXXX|NVXX|N[A-Z0-9]{3}(\+[A-Z]{4})?}
```

4 Online Resources

The ELID system and its associated resources is a public good project and made freely available, created to serve wine professionals and enthusiasts. This commitment to open access will remain a core principle of the project.

elid.git.wine public access to the most recent ELID database.

elid.git.wine/docs for PDFs of the most recent region and producer codes.

hi@git.wine contact e-mail for questions or feedback.

5 Related Wine ID Systems

Of bibliographic note—

The Liv-ex Wine Identification Number (LWIN) standard is the most complete data set for uniquely identifying wines. In its simplest form, LWINs are seven-digit numeric identifiers that serve as a look up key for wine details in the LWIN database [2]. Numeric LWINs have little meaningful information about a wine except to identify it uniquely. Extensions of the LWIN standard allow appending vintage year, case sizes, and bottle capacity in the interest of wine trade and exchange. Unlike LWIN, the coding system in this document emphasizes meaningful information encoded in every part of the wine ID itself. This enables natural sorting, recognizable names and regions, and basic data analysis using only a list of IDs without the need of a external database.

The LWIN database (CC BY 4.0) was used extensively to generate codes described in this document and gratefully acknowledged. The LWIN database is not complete and does not contain all known producers and their wines. As an operational database, it is also normal to discover dirty data: duplicate producers due to misspellings or variations, for example. ELIDs are developed to deliver a hiqh quality identification experience and uniquely positioned to support a clean, growing, and universal wine coding system.

German Amtliche Pruüfungsnummer (AP)

German wines labeled with AP numbers uniquely identify wine bottlings. A wine that is the same in all respects (producer, origin, and name) may be in some cases be distinguished by its AP number. This is useful to indicate if a bottling is of a certain style (for example, different levels of botrytis) [4]. AP numbers use numeric identifiers for producers and region data, along with other information. AP numbers are not necessarily consistent across vintages, making unique wine identification (e.g., Fritz Haag Brauneberger Juffer Riesling Kabinett) challenging across years.

Unlike AP numbers, ELIDs stress legible characters to identify details about a wine's producer and region. It does not attempt to uniquely identify individual bottlings in the way that AP numbers do. The benefits of distinguishing bottlings in the AP numbering system is useful to consumers and wine reviewers. It is appealing to associate AP lot numbers for German wines with ELIDs where possible (e.g., using a secondary lookup). Unfortunately, there exists no public database or system to know which AP numbers exist for all bottlings of German wines.

Compatibility and Limitations

— ELIDs are largely compatible with 7-digit and 11-digit LWINs. An ELID exists for most LWINs and an LWIN can be converted to an ELID using our database. An ELID can be converted to an LWIN, provided an LWIN exists for it. Exceptions include cases where we've corrected or merged duplicate entries in the LWIN database, generated ELIDs not in the LWIN database, or LWIN entries that don't correspond to single wines (e.g., assortment cases). The current design does not attempt to accommodate case sizes or bottle capacity in other LWINs, but may be amenable to extensions. 16-digit and 18-digit LWINs can be stripped down to fewer digits to convert to an ELID.

— German wines will use the same ELID code for bottlings with different AP numbers. If AP numbers are made available publicly or by joint coordination from producers, ELIDs can incorporate aspects of AP lot numbers in a secondary database.

— One consideration for ELIDs was to include the broad style of wine (red, white, fortified) but ultimately elided to keep identifiers relatively short. Wine styles may be determined with a secondary database.

A Appendix

A.1 Example IDs

- FR-BDX-MARX01-2012

  French wine, Bordeaux region, Château Margaux, Premier Cru Classe (ID 01), 2012

- FR-CHA-MOET06-NVXX

  French Champange by Moët & Chandon, Ice Impérial (ID 06), NVXX non-vintage

- US-OAK-OPUS01-2019+CAB

  United States wine, Oakville, Opus One, Red (ID 01), 2019, Cabernet Sauvignon

A.2 Sample Producer and Region Codes for Germany

Code	Region
DE-XXX	Germany
DE-AHR	Ahr
DE-BAD	Baden
DE-FRK	Franken (Franconia)
DE-HBS	Hessische Bergstrasse
DE-MTR	Mittelrhein
DE-MOS	Mosel (Mosel-Saar-Ruwer)
DE-NAH	Nahe
DE-PFZ	Pfalz (Palatinate)
DE-RHE	Rhein
DE-RHG	Rheingau
DE-RHH	Rheinhessen
DE-SAU	Saale-Unstrut
DE-SAC	Sachsen (Saxony)
DE-WUR	Württemberg

Code	Producer
DE-MMOL	Markus Molitor
DE-KELL	Keller
DE-JJPR	Joh. Jos. Prüm
DE-VOSB	Von Schubert
DE-DONN	Dönnhoff
DE-KLEB	Kloster Eberbach
DE-RESS	Balthasar Ress
DE-VONW	von Winning
DE-LAUE	Peter Lauer
...	...

B Common Varieties Codes

Appendix B: Common Grape Varieties

Code	Variety Name and Aliases
CAB	Cabernet Sauvignon
CHD	Chardonnay
MER	Merlot
PNR	Pinot Noir, Spätburgunder, Blauer Spätburgunder
RIS	Riesling
SBL	Sauvignon Blanc
SYR	Syrah
GRE	Grenache
TMP	Tempranillo
SNG	Sangiovese, Nielluccio
NEB	Nebbiolo
MAL	Malbec
PGR	Pinot Gris
VIO	Viognier
GWT	Gewürztraminer, Roter Traminer
ZIN	Zinfandel
CHB	Chenin Blanc, Steen
SEM	Sémillon
ALB	Albariño
VRM	Vermentino, Favorita, Pigato, Rolle, Vermentinu
...	...

References

[1] Liv-ex, "LWIN – a common language for fine wine," 2023. Available: https://www.liv-ex.com/lwin-3/lwin-common-language-fine-wine/. Accessed 2 Jan 2025.

[2] "London International Vintners Exchange," Wikipedia. Available: https://en.wikipedia.org/wiki/London_International_Vintners_Exchange. Accessed 2 Jan 2025.

[3] "Regular expression," Wikipedia. Available: https://en.wikipedia.org/wiki/Regular_expression. Accessed 2 Jan 2025.

[4] Mosel Fine Wines, "AP Number | The Traceability of German Wines." Available: https://www.moselfinewines.com/understanding-ap-numbers.php. Accessed 2 Jan 2025.