

Harmonisation without Homogenisation

*An Interdisciplinary Study of the Hierarchical Location Codes
(HLC) Operational Framework as a Conflict-Preserving Crosswalk
Architecture for Geopolitical, Hydrographic and Ecological
Reference Systems*

Doctoral Monograph

submitted in partial fulfilment of the requirements for the degree of Doctor of
Philosophy (Ph.D.)

Interdisciplinary Programme in Geoinformatics, Information Science and Biogeography
— *Interdisciplinary Monograph · Independent Scholarly Edition* —

Author: Jens Weynans

Krefeld, Germany

Framework version 7.8.2 · Monograph edition 2026 · with figures and appendices

Public derivative distribution: <https://weynans.info/hlc>

Abstract

Reference systems that name and order the surface of the Earth are foundational infrastructure for the sciences, for administration and for commerce, yet they are neither unified nor mutually consistent. Countries are enumerated differently by the International Organization for Standardization (ISO 3166), by the United Nations Statistics Division (UN M49) and by the United States Geopolitical Entities, Names, and Codes standard (GENC); the oceans are delimited by the International Hydrographic Organization's ageing publication S-23 in a third edition of 1953 whose successor has never been ratified; and the living surface of the planet is partitioned by a family of biogeographic classifications — terrestrial, freshwater, marine benthic and marine pelagic — that were designed for conservation planning rather than for interoperability. Each of these systems is internally coherent and externally incompatible. The practical consequence is a persistent translation problem: an analyst who must move a single datum between a political register, a hydrographic register and an ecological register has no authoritative crosswalk to rely upon.

This thesis presents, formalises and critically evaluates the **Hierarchical Location Codes (HLC) Operational Framework**, a single-author reference architecture that responds to this translation problem with a deliberately unusual design commitment: it **harmonises without homogenising**. Rather than resolving the disagreements between source standards into one canonical answer, HLC records those disagreements as first-class information, binding heterogeneous classifications together through a shared notation, a partitioned master code space and a system of inline attribute tags while allowing each source to retain its own logic and its own authority. The framework is realised as five coordinated but never merged files — a shared front matter, a namespace-governance layer, a land-based master integrator of some four thousand entries, a hydrographic layer, and an ecoregional layer assembling twelve biogeographic classifications — and is disseminated exclusively as derived, openly readable artefacts.

The study makes four contributions. First, it locates HLC within the intellectual histories of standardisation, ontology engineering, geographic information science and biogeography, and argues that its conflict-preserving stance is a coherent epistemological position rather than an engineering compromise. Second, it reconstructs the formal architecture of the framework — its notation, its mereology, its namespace discipline and its code-space partitioning — as an object of analysis in its own right. Third, it evaluates the framework against established criteria for interoperability, provenance and sustainability, drawing on the FAIR principles and on the standards-convergence literature. Fourth, it states plainly the framework's structural limitations — single-maintainer governance, the absence of spatial geometry, interpretive assignment and snapshot currency — and distinguishes those that are contingent

from those that are constitutive. The thesis concludes that HLC occupies a genuine and under-served niche as a specialist crosswalk and reference instrument, that its principal scholarly value lies in the disciplined preservation of contested classification, and that its principal risks are matters of governance rather than of design.

Keywords: reference data; crosswalk; semantic interoperability; ISO 3166; UN M49; DIAM 65-18; IHO S-23; ecoregions; conflict preservation; controlled vocabularies; geographic information science; biogeography.

Declaration and Note on Evidentiary Basis

This thesis analyses the HLC Operational Framework as described and specified by its author, and situates it within the peer-reviewed and institutional literature on the reference systems it integrates. The structural and philosophical description of the framework derives from its own front matter and file specifications. Statements about the external standards — ISO 3166, UN M49, GENC, the Defense Intelligence Agency Manual 65-18, the IHO publication S-23, and the terrestrial, freshwater, marine and pelagic ecoregion classifications — are grounded in their primary and canonical sources, which are listed in the References. Where the argument concerns the internal counts or specific code assignments of a particular HLC release, it treats those as illustrative of the architecture rather than as independently audited figures; the analytical claims of the thesis do not depend on any single unverified cell value.

Table of Contents

Abstract.....	2
Declaration and Note on Evidentiary Basis.....	4
Table of Contents.....	5
Acknowledgements.....	7
List of Figures.....	8
List of Tables.....	8
1 Introduction.....	11
1.1 The translation problem in geographic reference.....	11
1.2 Object of study.....	12
1.3 Research questions.....	12
1.4 Interdisciplinary stance.....	13
1.5 Contributions.....	13
1.6 Structure of the thesis.....	13
2 The Landscape of Geographic Reference Standards.....	15
2.1 Geopolitical enumeration: ISO 3166, UN M49 and GENC.....	15
2.2 Hydrographic delimitation: IHO S-23 and its unratified successors.....	16
2.3 Biogeographic classification of the living surface.....	17
2.4 Reference-data and interoperability literatures.....	17
3 Theoretical Framework.....	19
3.1 What a classification commits us to.....	19
3.2 Mereology: the logic of part and whole.....	19
3.3 Namespaces, identifiers and ownership.....	20
3.4 An epistemology of preserved conflict.....	20
4 Research Design and Methods of Analysis.....	24
4.1 Research paradigm and rationale.....	24
4.2 Object, unit and scope of analysis.....	24
4.3 Data sources and provenance.....	25
4.4 Extraction and verification procedure.....	25
4.5 Analytical lenses.....	25
4.6 Evaluation criteria.....	26
4.7 Threats to validity and mitigations.....	26
4.8 Ethical and reflexive considerations.....	27
5 Architecture and Notation of the Framework.....	30
5.1 File topology and separation of concerns.....	30
5.2 The shared notation as a controlled micro-language.....	31
5.3 Inline attribute tags as qualified references.....	31
5.4 The master integrator: a partitioned numeric code space.....	32
5.5 Coordinates as orientation, not geometry.....	32
6 Construction Methodology and Editorial Method.....	33
6.1 Sources, precedence and the reuse-before-mint rule.....	33
6.2 The treatment of contested and historical entities.....	33
6.3 Reconciling competing editions: the hydrographic case.....	33
6.4 Distribution as derivative, open artefacts.....	33
7 Empirical Characterisation of the Corpus.....	35
7.1 Overall scale and file composition.....	35
7.2 The type structure of the land hierarchy.....	36
7.3 The partitioned master code space in practice.....	37
7.4 Cross-reference density: the integrated schemes.....	39
7.5 The ecoregional sheets and their source scales.....	40
7.6 The hydrographic multi-edition crosswalk.....	41
8 The Land Subsystem (codehl).....	45
8.1 Scope and role as master integrator.....	45

8.2 The parallel type taxonomy.....	45
8.3 Crosswalk density and the multi-attribute entry.....	45
8.4 Assessment of the land subsystem.....	46
9 The Water Subsystem (codehy).....	47
9.1 Scope.....	47
9.2 Multiple hydrographic standards in parallel.....	47
9.3 Assessment of the water subsystem.....	47
10 The Ecoregional Subsystem (codehw).....	49
10.1 Scope and the deliberate absence of a master.....	49
10.2 Common structure across heterogeneous sources.....	49
10.3 Assessment of the ecoregional subsystem.....	49
11 Namespace Governance (codehu) and the Shared Front Matter (codehx)....	51
11.1 The namespace layer as the guarantor of identity.....	51
11.2 The front matter as the framework's semantic contract.....	51
11.3 Declared limitations as an epistemic virtue.....	51
12 Evaluation.....	55
12.1 Against the FAIR principles.....	55
12.2 Against crosswalk-quality criteria.....	55
12.3 Fitness for declared purpose.....	56
12.4 Comparative positioning.....	56
13 Applied Perspectives and Use Scenarios.....	57
13.1 Scenario: harmonising cross-border epidemiological data.....	57
13.2 Scenario: integrating marine observations across hydrographic authorities.....	57
13.3 Scenario: linking biodiversity records to a coherent ecoregional frame	58
13.4 Scenario: a historical time-series across changing states.....	58
13.5 Comparison with alternative reference resources.....	59
13.6 Anti-patterns: where HLC should not be used.....	60
13.7 Synthesis: the niche the scenarios reveal.....	60
14 Limitations: Contingent and Constitutive.....	61
14.1 Constitutive limitations.....	61
14.2 Contingent limitations.....	61
14.3 The central risk restated.....	62
15 Conclusions and Future Work.....	63
15.1 Answers to the research questions.....	63
15.2 Contribution restated.....	63
15.3 Future work.....	64
15.4 Closing statement.....	64
References.....	65
Appendix A Notation Reference (Analytical Summary).....	66
Appendix B Glossary.....	66
Appendix C Register of Integrated Schemes.....	68
Appendix D A Worked Crosswalk.....	70
D.1 The entry as stored.....	70
D.2 Reading the entry.....	70
D.3 What the worked example demonstrates.....	70

Acknowledgements

This monograph analyses a body of work built, over many years and without institutional support, by a single scholar. Whatever the assessment that follows concludes about the framework's limits, the achievement of assembling and reconciling some forty reference schemes across three domains, alone and in the open, deserves to be acknowledged at the outset. The study also stands on the labour of the standards bodies and research teams whose classifications the framework integrates; their published work is the ground on which both the framework and this analysis rest.

List of Figures

Figure 5.1 The layered file architecture and dependency direction.

Figure 7.1 Relative scale of the four data files.

Figure 7.2 Entity-type distribution across the 4,146 land entries.

Figure 7.3 Depth profile of the land hierarchy.

Figure 7.4 Partitioning of the master (HLC-M) numeric code space.

Figure 7.5 The twelve ecoregional sheets of codehw by size.

List of Tables

Table 2.1 Principal geopolitical enumeration standards.

Table 2.2 Biogeographic classifications and their native scales.

Table 4.1 Analytical lenses and their evidence.

Table 5.1 File responsibilities and dependencies.

Table 7.1 Scale and composition of the five files.

Table 7.2 Entity-type distribution of the land file.

Table 7.3 Partitioning of the master code space.

Table 7.4 Ecoregional sheets and source scales.

Table 7.5 Excerpt of the IHO multi-edition reconciliation.

Table 12.1 FAIR assessment summary.

Table 13.1 Comparison of HLC with alternative reference resources.

Table A.1 Notation operators and their functions.

PART I

Orientation

1 Introduction

1.1 The translation problem in geographic reference

Every serious effort to describe the world quantitatively begins with an act of naming and enumeration. Before a public-health analyst can compare mortality across states, before an oceanographer can pool observations from two cruises, and before a conservation planner can aggregate species records into a region, each must commit to a scheme that partitions the Earth into named, coded units. These schemes are not neutral. They encode purposes: ISO 3166 exists to give machine-stable identifiers to countries and their subdivisions for postal, banking and computing use; UN M49 exists to arrange countries into statistical groupings without prejudice to their political status; the biogeographic ecoregion systems exist to direct scarce conservation resources toward biologically coherent units. Because their purposes differ, their partitions differ, and because their partitions differ, data expressed in one scheme cannot be moved into another without loss, ambiguity or contestation.

The difficulty is not merely that codes differ superficially. It is that the schemes disagree at the level of ontology — about what the units of the world are. UN M49 assigns a numeric code to a statistical grouping such as “Sub-Saharan Africa” that has no counterpart in ISO 3166. The IHO's S-23 recognises a set of named seas whose boundaries were fixed for the convenience of national hydrographic offices and “have no political significance whatsoever,” yet which nonetheless collide with claims of sovereignty, as the unresolved dispute over the naming of the sea between the Korean peninsula and the Japanese archipelago has demonstrated for two decades. The marine ecoregion system nests 232 ecoregions within 62 provinces within 12 realms; the terrestrial system nests 867 ecoregions within 14 biomes within 8 realms; these numbers are incommensurable by construction. To move between registers is therefore to translate between worldviews, and translation between worldviews cannot be lossless.

The conventional engineering response to such heterogeneity is **canonicalisation**: designate one scheme as authoritative, map everything else onto it, and discard or flatten the residue. This is the logic of the master data warehouse and of the single golden record. It is efficient, and for many operational purposes it is correct. But it purchases efficiency at the price of information: the disagreements between sources — which are frequently the most scientifically and politically interesting facts about a place — are precisely what canonicalisation destroys. The HLC Operational Framework is an extended argument against reflexive canonicalisation in the domain of geographic reference. Its central design decision is to treat conflict between authoritative sources as data to be preserved rather than noise to be removed.

1.2 Object of study

The object of this thesis is the HLC Operational Framework in its version 7.8, a body of structured reference data authored and maintained by a single independent scholar and disseminated as openly readable derivative files. The framework consists of five coordinated components:

1. **codehx** — a common front matter that fixes the shared notation, editorial conventions, coordinate policy and declared limitations for the entire system, so that these need not be restated in each data file;
2. **codehu** — a namespace-governance layer that declares, for the space of two- and three-character base codes, which external or internal authority owns each code, thereby preventing collisions between the many coding systems the framework integrates;
3. **codehl** — the land-based master file, which carries the geopolitical hierarchy from continents to sub-national units and which also hosts the master integrator, the partitioned numeric code space that anchors the whole system;
4. **codehy** — the water-based file, which organises oceans, seas, rivers, lakes, currents, straits and undersea features and reconciles competing hydrographic standards;
5. **codehw** — the ecoregional file, which assembles the major terrestrial, freshwater, marine and pelagic biogeographic classifications as parallel, sibling views of the living surface of the Earth.

The working files are maintained as spreadsheets and a word-processing document; these are not published. What is published is a set of derived artefacts in portable, non-proprietary formats — a human-readable rendering and a tabular rendering — under a stable public address. The tabular renderings use tab separation for the narrower files and semicolon separation for the wider ones, the latter chosen because the semicolon does not occur inside the source cells and can therefore serve as an unambiguous delimiter.

1.3 Research questions

The thesis is organised around four questions.

1. **RQ1 (Positioning).** What problem in the landscape of geographic, hydrographic and ecological reference does HLC address, and how does its response differ from the canonicalising strategies that dominate reference-data practice?
2. **RQ2 (Architecture).** What are the formal mechanisms — notation, mereology, namespace discipline, code-space partitioning and inline attribute tagging — by which the framework binds heterogeneous classifications while preserving their disagreements?

3. **RQ3 (Evaluation).** Measured against recognised criteria for interoperability, provenance and reusability, what does the framework achieve, and where does it fall short?
4. **RQ4 (Sustainability).** Which of the framework's limitations are contingent features of its current governance, and which are constitutive of its design and therefore irreducible?

1.4 Interdisciplinary stance

HLC cannot be adequately assessed from within any single discipline, because it deliberately spans several. Its geopolitical layer is an artefact of political geography and of standardisation studies; its code architecture is an artefact of information science and of the theory of controlled vocabularies; its hydrographic layer belongs to the history of cartography and to marine governance; its ecoregional layer belongs to biogeography and conservation science; and its overall stance toward conflicting sources is an epistemological position with roots in the philosophy of classification. This thesis is therefore explicitly interdisciplinary. It borrows the vocabulary of formal ontology and mereology to describe the framework's part-whole logic; it borrows the FAIR principles and the crosswalk literature from data science to evaluate it; and it borrows from science-and-technology studies the recognition that classifications are never innocent, that every ordering of the world advantages some purposes and some actors over others.

1.5 Contributions

- A formal reconstruction of the HLC architecture as a conflict-preserving crosswalk, expressed in the shared vocabulary of ontology engineering and information science rather than in the framework's own idiom.
- A comparative situating of HLC against the principal source standards it integrates, using their canonical documentation, that clarifies exactly which incompatibilities the framework absorbs and how.
- An evaluation against interoperability and reusability criteria that separates the framework's genuine achievements from its aspirational claims.
- A sustainability analysis that distinguishes contingent from constitutive limitations and that frames the framework's principal risk as one of governance rather than of design.

1.6 Structure of the thesis

The monograph is organised in five parts. **Part I** orients the reader: Chapter 1 states the problem and the research questions, Chapter 2 surveys the source standards, and Chapter 3 develops the theoretical apparatus. **Part II** sets out the research design and the methods of analysis in Chapter 4. **Part III** anatomises the framework: Chapter 5 reconstructs its architecture and notation, Chapter 6

its construction method, and Chapter 7 characterises the corpus empirically with figures drawn from the version 7.8.2 release. **Part IV** analyses the four subsystems in turn — land (Chapter 8), water (Chapter 9), ecoregions (Chapter 10) and namespace governance with the front matter (Chapter 11). **Part V** assesses and concludes: Chapter 12 evaluates the framework, Chapter 13 develops applied use scenarios, Chapter 14 states its limitations, and Chapter 15 concludes and identifies future work.

2 The Landscape of Geographic Reference Standards

This chapter surveys the standards that HLC integrates and the scholarly literatures that bear on their integration. The purpose is not encyclopaedic completeness but the identification of the specific incompatibilities that the framework is designed to absorb.

2.1 Geopolitical enumeration: ISO 3166, UN M49 and GENC

The three dominant enumerations of countries and territories were built for different institutions and different ends, and they diverge accordingly. ISO 3166-1 provides three parallel country identifiers — a two-letter code intended as the general-purpose identifier, a three-letter code that associates more legibly with the country name, and a three-digit numeric code that is script-independent and identical to the numeric code maintained by the United Nations. ISO 3166 currently enumerates 249 countries, of which 193 are UN member states; it further provides, in Part 2, codes for principal subdivisions, and in Part 3, codes for names withdrawn since the standard's first publication in 1974. Crucially, ISO does not itself decide country names; it draws them from United Nations sources.

UN M49, the Standard Country or Area Codes for Statistical Use, assigns a three-digit numeric code to each country or area and, above the level of the country, to a hierarchy of statistical groupings — world, region, sub-region and intermediate region. The numeric codes are deliberately conservative: unlike the ISO alphabetic codes, they generally do not change when a country's name changes, only when its territorial extent changes significantly. The UN is explicit that the assignment of countries to groupings is a matter of statistical convenience that implies no judgement about political status. This disclaimer is not a formality; it is the mechanism by which a body composed of contending states can publish a shared classification at all.

The United States GENC standard re-implements the geopolitical entity space for US federal use, tracking ISO 3166 closely but diverging where US policy recognises entities, boundaries or names differently, and adding identifiers in numeric ranges reserved for entities that ISO does not code. The existence of GENC alongside ISO is itself an instance of the core problem: two authorities, each legitimate within its own remit, producing overlapping but non-identical enumerations of the same world.

HLC absorbs these three enumerations not by choosing among them but by carrying them side by side. A single geopolitical entity in the framework may simultaneously bear its ISO alphabetic and numeric codes, its M49 statistical code, its GENC identifier and the framework's own master code, with each held in a distinct, labelled slot. The disagreements — an entity coded by GENC but not by ISO, a grouping present in M49 but absent from ISO — survive as recorded facts.

Standard	Maintainer	Primary identifier(s)	Governing purpose	Update logic
ISO 3166-1	ISO (Geneva)	alpha-2, alpha-3, numeric-3	General-purpose machine identification	Alpha codes may change with names
UN M49	UN Statistics Division	numeric-3 + grouping codes	Statistical aggregation	Codes change with territory, not name
GENC	US Government (NSG)	re-implemented entity codes	US federal interoperability	Tracks ISO; diverges on US policy
DIAM 65-18	US Defense Intelligence Agency	geopolitical & hydrographic codes	Defence data standardisation	Manual revision

2.2 Hydrographic delimitation: IHO S-23 and its unratified successors

The delimitation of the oceans and seas rests, remarkably, on a document from 1953. The IHO's special publication S-23, *Limits of Oceans and Seas*, appeared first in 1928, was revised in 1937, and reached its third and still-current edition in 1953. Its preface is candid about its own modest ambition: the limits were drawn “solely for the convenience of National Hydrographic Offices” so that publications headed with the name of a sea would refer to the same area, and they “have no political significance whatsoever.” Straits joining two seas were allotted to one of them by a convention adopted in 1919. A fourth edition was drafted in 1986 and nearly approved; work resumed in 1998 under a new title and with some sixty renamed seas; it was resubmitted in 2002 and then suspended, pending a resolution of the dispute over the name of the Japan Sea that has never come. In 2020 the member states adopted a compromise: numeric identifiers for seas, the retirement of efforts to revise the analog S-23, and the development of a digital successor dataset, S-130, planned for release around 2026.

For a crosswalk framework this history is not an antiquarian aside; it is the central fact of the hydrographic domain. There is no single, current, ratified delimitation of the seas. Instead there are at least four partially incompatible authorities in active reference use — the in-force third edition of 1953, the never-ratified fourth-edition draft of 1986, the older aeronautical chart delimitation, and the Defense Intelligence Agency's manual — and their correspondences are genuinely many-to-many. The IHO's own appendices document the problem directly, tabulating, sea by sea, that the same body of water bears different codes and sometimes different extents across the third edition, the fourth-edition draft, the aeronautical delimitation and the DIA manual, and warning that “even the number of principal water bodies varies from organization to organization.” HLC's water file is, in effect, a machine-actionable version of exactly this cross-reference, extended to rivers, lakes and undersea features.

2.3 Biogeographic classification of the living surface

The partition of the Earth's biota into nested regions is a distinct intellectual tradition, developed largely for conservation planning and only incidentally for interoperability. The terrestrial system of Olson and colleagues (2001) mapped 867 ecoregions nested within 14 biomes and 8 biogeographic realms, a roughly fourfold increase in resolution over the earlier scheme of Udvardy, and it rapidly became the de facto framework for global terrestrial conservation. Its marine counterpart, the Marine Ecoregions of the World of Spalding and colleagues (2007), applied the same nested logic to coastal and continental-shelf waters, describing 232 ecoregions within 62 provinces within 12 realms, extending from the intertidal zone to the 200-metre depth contour. The surface pelagic waters were classified separately by Spalding and colleagues (2012) into 37 pelagic provinces within four realms and seven ecologically defined biomes. Freshwater systems were treated by Abell and colleagues (2008). Later syntheses — the RESOLVE bioregions of Dinerstein and colleagues (2017) and the European EUNIS habitat classification — reorganised and extended these units again.

Two features of this tradition matter for HLC. First, these systems are explicitly nested hierarchies of the same formal shape — realm, province/biome, ecoregion — which makes them structurally amenable to a common tabular treatment even though their contents differ. Second, they are avowedly non-authoritative in the political sense: their boundaries are hypotheses about biological coherence, revised as knowledge improves, and they are cross-referenced to, rather than derived from, one another. HLC's ecoregional file assembles these as sibling views with no master among them, which faithfully reflects the fact that biogeography itself recognises no single correct partition of life.

Classification	Reference	Top tier	Middle tier	Base units
Terrestrial (TEOW)	Olson et al. 2001	8 realms	14 biomes	867 ecoregions
Marine coastal (MEOW)	Spalding et al. 2007	12 realms	62 provinces	232 ecoregions
Pelagic (PPOW)	Spalding et al. 2012	4 realms	7 biomes	37 provinces
Freshwater (FEOW)	Abell et al. 2008	8 realms	habitat types	426 ecoregions
Bioregions (RESOLVE)	Dinerstein et al. 2017	8 realms	14 biomes	846 ecoregions

2.4 Reference-data and interoperability literatures

The problem HLC addresses is a recognised one in information science, where it is discussed under the headings of crosswalks, concordances, semantic interoperability and reference-data management. A crosswalk is a mapping

between the elements of two schemes; its difficulty grows with the degree to which the schemes disagree about their units, and the literature distinguishes clean one-to-one mappings from the far more common one-to-many, many-to-one and many-to-many correspondences that arise when schemes carve the world differently. The FAIR principles — that data should be Findable, Accessible, Interoperable and Reusable — supply an evaluative frame that this thesis will apply in Chapter 12, particularly the interoperability requirement that data use shared vocabularies and qualified references to other data, and the reusability requirement of rich, accurate provenance.

Science-and-technology studies contributes a complementary and more critical perspective. In the analysis of classification associated with Bowker and Star, every working classification embeds decisions about what counts, imposes a residual category on what does not fit, and renders some perspectives visible while silencing others. From this vantage HLC's conflict-preserving stance can be read as an attempt to keep the silenced perspectives on the page — to refuse the closure that ordinary classification enforces. Whether that refusal is fully achievable, or whether the framework merely relocates the act of judgement from the assignment of units to the curation of alternatives, is a question the thesis returns to in Chapters 11 and 12.

3 Theoretical Framework

To analyse HLC rigorously requires a vocabulary adequate to what it actually does. This chapter assembles that vocabulary from four sources: the theory of classification, formal mereology, the discipline of namespaces and identifiers, and an epistemology of preserved conflict.

3.1 What a classification commits us to

A classification is not merely a list; it is a set of ontological commitments about what entities exist and how they relate. When UN M49 posits “Sub-Saharan Africa” as a coded unit, it commits to the existence of that grouping as an object about which statements can be made. When ISO 3166 declines to code such groupings, it declines that commitment. A framework that carries both must be able to host incompatible ontological commitments simultaneously without collapsing them. This is the first theoretical demand HLC places on itself, and it distinguishes the framework from an ordinary database, whose schema typically enforces a single ontology.

The classical criteria for a good classification — that its categories be jointly exhaustive and mutually exclusive — are precisely what a multi-source reference system cannot honour globally, because its sources are exclusive and exhaustive only relative to their own purposes. HLC therefore does not aim at a single exhaustive-and-exclusive partition. It aims instead at a coherent superposition of several such partitions, each locally well-formed, bound together by explicit cross-reference. The coherence it seeks is not the coherence of a single tree but the coherence of a well-documented atlas of overlapping trees.

3.2 Mereology: the logic of part and whole

The dominant relation in every layer of HLC is the part-whole relation: a subdivision is part of a country, a country part of a region, a sea part of an ocean, an ecoregion part of a province. Formal mereology, the theory of parthood, supplies the appropriate discipline. Parthood is transitive — if a district is part of a state and the state part of a country, the district is part of the country — and this transitivity is what makes hierarchical roll-up meaningful. But the framework confronts a complication that pure mereology does not: an entity may be part of different wholes under different classifications. A coastal ecoregion is part of a marine province under MEOW and, simultaneously, adjacent to a terrestrial ecoregion under TEOW; a strait is allotted to one sea under the 1919 convention but borders two. HLC handles this through what it designates as legitimate alternative hierarchies — additional, non-redundant parenthoods marked explicitly so that they are never mistaken for duplicated data.

This is the framework's most theoretically interesting move. It replaces the single-parent tree, which is the default of nearly all administrative and file-system hierarchies, with a controlled polyhierarchy in which multiple parenthoods

coexist under declared conditions. To prevent the contradictions that polyhierarchy invites, the framework imposes a discipline of attribute locality: the descriptive attributes of an entity are stored only at its original occurrence, and alternative-hierarchy occurrences carry a pointer rather than a copy. Because the attributes exist in exactly one place, two occurrences of the same entity can never disagree about its properties — a design that converts a potential source of inconsistency into a structural guarantee of consistency.

3.3 Namespaces, identifiers and ownership

A system that ingests codes from ISO, the UN, the US government, a defence manual, several hydrographic authorities and a dozen ecological classifications faces an immediate hazard: the same short string means different things to different authorities, and two authorities may issue colliding codes. The information-science response is the namespace — a declared scope within which identifiers are unique and an owner is responsible for their assignment. HLC elevates namespace management to a dedicated layer whose sole function is to declare, for each region of the code space, which authority owns it. A two-letter code is adjudicated by one rule; a three-letter code by another; numeric ranges are partitioned among the standards and the framework's own extensions. The effect is that every identifier in the system can be traced to an owning authority, which is a precondition both for provenance and for controlled extension.

The distinction between an identifier and a locator is relevant here. HLC's codes are identifiers — they denote entities and are intended to remain stable even as descriptions change — and the framework's commitment to redirects rather than deletions, so that a withdrawn identifier points to its successor rather than vanishing, mirrors the stability logic of ISO 3166-3 and of persistent-identifier schemes generally. Stability of identifier under change of description is exactly the property that makes longitudinal analysis possible, and it is a property HLC shares with the more conservative of its sources, notably UN M49.

3.4 An epistemology of preserved conflict

The framework's defining commitment — that conflicting classifications are preserved as information rather than resolved — deserves to be stated as an epistemological thesis rather than an engineering preference. The thesis is that, in a domain where legitimate authorities disagree, the disagreement is itself knowledge, and a reference system that erases it is less truthful than one that records it. This position has a respectable pedigree. In the treatment of contested territory, recording a claim without adjudicating it is the neutral scholarly stance; in the treatment of the seas, the IHO's own cross-reference tables preserve rather than resolve the divergences among authorities; in biogeography, the coexistence of competing regionalisations is understood as normal science rather than as error to be corrected.

The counter-argument is equally clear and must be granted. Preservation of conflict shifts the burden of resolution onto the user, who must still decide, for any given purpose, which alternative to adopt; and the selection of which alternatives are legitimate enough to record is itself an editorial judgement that reintroduces, at one remove, the very subjectivity that preservation was meant to avoid. The framework does not escape judgement; it relocates it from the assignment of a single answer to the curation of a set of answers. Whether that relocation is an improvement depends on whether the curation is transparent and principled — a question this thesis treats as empirical and returns to in the evaluation.

PART II

Research Design

4 Research Design and Methods of Analysis

This chapter states how the framework was studied. Because HLC is at once an artefact of information engineering, a scholarly claim about classification, and a body of empirical data, no single method suffices; the study therefore combines document analysis, formal reconstruction, quantitative corpus characterisation and criterion-based evaluation. The chapter sets out the research paradigm, the object and unit of analysis, the data sources and their provenance, the extraction and verification procedure, the analytical lenses applied, and the threats to validity together with the measures taken against them.

4.1 Research paradigm and rationale

The study is interpretive and analytical rather than experimental. Its aim is not to test a hypothesis about the world but to understand, reconstruct and appraise a designed artefact and the epistemological commitments it embodies. This places the work within the tradition of design-oriented and artefact-centred research, in which the object of study is a human-made system whose rationale, structure and fitness are to be made explicit and judged. The choice of paradigm follows from the object: a reference framework cannot be evaluated by controlled experiment, because its value lies in coherence, coverage and fidelity to sources rather than in any measurable behavioural outcome. It can, however, be evaluated rigorously — against its own stated goals, against the standards it integrates, and against established criteria for reference-data quality — and it can be characterised quantitatively from its own contents. The study therefore triangulates three complementary modes of evidence: what the framework says of itself, what its data reveal, and what independent standards and literatures establish.

4.2 Object, unit and scope of analysis

The object of analysis is the HLC Operational Framework, version 7.8.2, dated 18 August 2026, in the five files that constitute it. The primary unit of analysis is the entry — a single coded entity together with its hierarchical position, its type, its master code and its inline attribute tags — of which the land file alone contains 4,146. Secondary units are the worksheet, at which coverage and structure are characterised, and the file, at which architecture and dependency are analysed. The scope is deliberately bounded to the framework as released: the study analyses the published derivative logic and the working spreadsheets as supplied, but does not attempt to audit the correctness of every external code against its issuing authority, an undertaking that would require the cooperation of some forty standards bodies and lies beyond a single study. Where correctness is discussed, it is at the level of architecture and method, not of the individual cell.

4.3 Data sources and provenance

Two classes of source underpin the study. The first is the framework itself: the front-matter document and the four workbooks, read in full, from which every quantitative figure in Chapter 7 is computed directly. The second is the canonical documentation of the standards the framework integrates — the published texts and datasets of ISO 3166, UN M49, GENC, the Defense Intelligence Agency manual, the IHO publication on the limits of oceans and seas, and the terrestrial, freshwater, marine, pelagic and deep-sea ecoregion classifications — which supply the independent yardstick against which the framework's treatment of each domain is assessed. The provenance of every figure is thus traceable to one of these two classes, and the study is careful throughout to attribute each claim to the class from which it derives, so that a reader can distinguish a fact about the HLC cells from a fact about an external standard.

4.4 Extraction and verification procedure

Quantitative characterisation proceeded in four steps. First, each workbook was parsed programmatically and every worksheet inventoried for its populated row count and maximum populated column, yielding the scale figures of Chapter 7. Second, the land file's entries were tabulated by their one-letter type to produce the type distribution, and by the separator depth of their hierarchy path to produce the depth profile. Third, the master-code definition sheet and the namespace file were read to reconstruct the partitioning of the numeric code space. Fourth, targeted entries — a representative country, a representative sea across the several hydrographic editions, and a sample from each ecoregional sheet — were extracted in full to verify that the architecture described in Chapters 5 and 6 is realised as claimed in the data. Each derived figure was recomputed independently to guard against parsing error, and figures that inform the argument were cross-checked against the framework's own front-matter statements of scale, which they corroborate.

This procedure yields a form of internal triangulation: the same structural claim is checked against the front matter's prose, against the definition sheets and against the raw data. Where these agree, as they did throughout, the claim is reported as established; the study does not rely on any figure that could not be reproduced from the released files.

4.5 Analytical lenses

Four lenses, introduced in Chapter 3, are applied systematically to the reconstructed framework. The **mereological lens** examines the part-whole logic of each subsystem and the controlled polyhierarchy that permits multiple parenthoods. The **namespace lens** examines identity and ownership: how codes are attributed to authorities and how collisions are prevented. The **crosswalk lens** examines the inline attribute tags as qualified references and assesses the

breadth and fidelity of the correspondences they encode. The **epistemic lens** examines the framework's treatment of conflict — whether disagreement between sources is genuinely preserved, and at what cost. Each substantive chapter foregrounds the lens most pertinent to its subsystem while drawing on the others, and the evaluation of Chapter 12 integrates all four against external criteria.

Lens	Question it answers	Primary evidence	Chapter s
Mereological	How are part and whole, and multiple parenthood, structured?	Hierarchy paths; alt-hierarchy markers	3, 7, 8–10
Namespace	Who owns each identifier and how are collisions avoided?	codehu ownership table; code ranges	3, 7, 11
Crosswalk	How wide and faithful are the cross-references?	Inline attribute tags; INFO scheme list	6, 7, 12, 13
Epistemic	Is conflict preserved, and at what cost?	IHO multi-edition sheet; alternatives notation	3, 9, 12, 14

4.6 Evaluation criteria

The appraisal in Chapter 12 is criterion-based rather than impressionistic. Three established frames are applied. The **FAIR principles** supply the reusability frame, assessing findability, accessibility, interoperability and reusability. **Crosswalk-quality criteria** from the concordance literature supply the integration frame, assessing coverage, fidelity to many-to-many correspondence, and the handling of genuine disagreement. **Fitness-for-purpose** supplies the pragmatic frame, judging the framework against its own declared envelope of use and against the purposes it explicitly disclaims. Using declared, external criteria in place of the analyst's unaided judgement is itself a methodological safeguard: it makes the evaluation contestable, because a reader who rejects a rating can point to the criterion and the evidence on which it rests.

4.7 Threats to validity and mitigations

Four threats deserve explicit treatment. The first is **construct validity**: the risk that the study measures something other than what it claims. This is mitigated by defining the unit of analysis precisely and by tying every quantitative claim to a reproducible computation over the released files. The second is **source-fidelity risk**: the possibility that the framework misrepresents an external standard, which the study cannot fully exclude without a cell-by-cell audit; it is mitigated by grounding all statements about external standards in their canonical documentation and by confining correctness claims to the architectural level. The

third is **single-artefact generality**: the study concerns one framework, so its conclusions about conflict-preserving design are illustrative rather than universally established; this is acknowledged and reflected in the cautious framing of the theoretical contribution. The fourth is **investigator distance**: because the framework is the work of a single author who is also the party most knowledgeable about it, the study relies on the released artefacts and public standards rather than on privileged testimony, so that its findings rest on inspectable evidence alone.

4.8 Ethical and reflexive considerations

The framework touches politically sensitive matter — contested territories, disputed sea names, the enumeration of peoples and states — and the study treats these as the framework itself does: descriptively and without adjudication. Recording a claim is not endorsing it, and the analysis is careful to present contested cases as the data present them, noting competing claims without ranking them. Reflexively, the study recognises that any appraisal of a classification is itself performed from a position, and it therefore makes its criteria explicit so that its own judgements are as open to inspection as the artefact it examines. The aim throughout is an assessment that the framework's author, its potential users and its potential critics could each recognise as fair.

PART III

Anatomy of the Framework

5 Architecture and Notation of the Framework

This chapter reconstructs the framework's formal machinery: its file topology, its shared notation, its inline attribute tags, and the partitioned master code space that anchors the whole.

5.1 File topology and separation of concerns

The five-file decomposition is an application of the software-engineering principle of separation of concerns to reference data. Each file has a single, well-defined responsibility, and the responsibilities are layered so that the more stable and general sits beneath the more specific and volatile. The front matter fixes conventions; the namespace layer governs identity; the land file anchors the master code space and carries the geopolitical hierarchy; the water and ecoregion files extend the system into their domains within reserved code ranges. The files are coordinated but never merged, which yields two benefits. First, each file can adopt the representational logic appropriate to its domain — the geopolitical hierarchy is a near-tree with controlled exceptions, whereas the ecoregional file is an explicit collection of sibling hierarchies with no master. Second, the shared conventions live in exactly one place, so that a change to notation or policy need not be propagated across five documents.

Layered file architecture and dependency direction

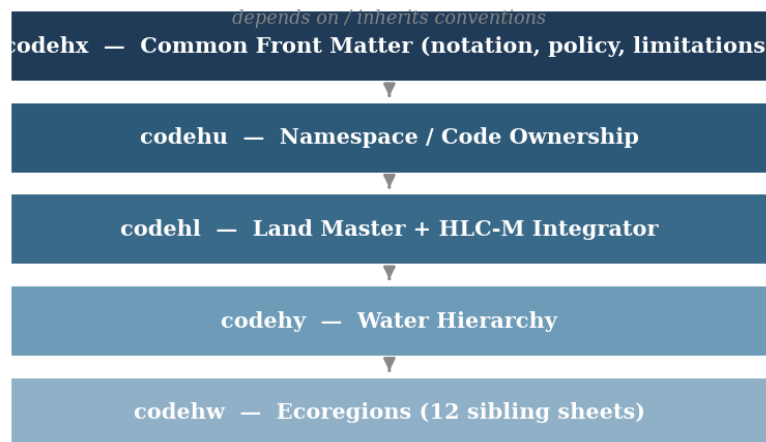


Figure 5.1 The layered file architecture. Each file inherits the conventions of the layer above; the land file hosts the master integrator on which the water and ecoregion files draw.

File	Responsibility	Dominant structure	Depends on
codehx	Shared conventions, notation, limitations	Prose front matter	—
codehu	Namespace / code ownership	Ownership table over code space	codehx
codehl	Land hierarchy +	Controlled	codehx,

	master integrator	polyhierarchy + code space	codehu
codehy	Water hierarchy	Controlled polyhierarchy	codehx, codehu, codehl
codehw	Ecoregions	Parallel sibling hierarchies	codehx, codehu, codehl, codehy

5.2 The shared notation as a controlled micro-language

The framework's notation is best understood as a small controlled language whose operators carry precise, documented meanings. A vertical bar marks alternatives that are equally valid, refusing to privilege one over another. A right-pointing marker denotes a redirect to a canonical or successor entry, encoding the stability-under-change policy directly in the data. An exclamation marker flags a code whose authoritative definition lives outside the present sheet, making cross-file dependency explicit. An asterisk designates a legitimate alternative hierarchy — a genuine second parenthood rather than accidental duplication — and is the syntactic carrier of the controlled polyhierarchy described in Chapter 3. A small set of arithmetic-like markers records values that are inherited, encroached upon, or extended relative to a source. Each operator is a deliberate act of compression: it lets a single cell express a relationship that would otherwise require prose, and it does so uniformly across all five files.

The virtue of a controlled notation is that it is simultaneously human-readable and machine-actionable: a person can read the alternatives and redirects directly, while a parser can act on them deterministically. The cost is a learning curve. The notation is dense, and it is not self-explanatory; without the front matter a reader cannot fully decode it. This is the price of the framework's expressiveness, and it positions HLC as an expert instrument rather than a casual lookup table — a characterisation the thesis develops in the evaluation.

5.3 Inline attribute tags as qualified references

The mechanism by which HLC binds an entity to its representations in external schemes is the inline attribute tag: a compact, labelled key-value construct of the form scheme-then-value that attaches, to a single entry, its identifier under a named external authority. A geopolitical entity may thus carry, in one place, its UN numeric code, its several national and institutional codes, and its cross-references to hydrographic and ecological schemes, each in a distinct tagged slot. In the vocabulary of the FAIR principles, these tags are qualified references to other data: they do not merely assert that a correspondence exists but name the scheme under which it holds, which is exactly the qualification that machine interoperability requires. The tag system is what turns a set of five hierarchies

into a crosswalk, because it is through the tags that a unit in one register is tied to its counterparts in every other.

5.4 The master integrator: a partitioned numeric code space

The anchor of the whole system is a numeric master code space, hosted in the land file, that assigns to every entity a code drawn from deliberately reserved ranges. The partitioning is semantic: broad numeric bands are allotted to the continents and their subordinate regions, further bands to supranational organisations and to historical entities, and specific ranges are reserved for hand-off to the water and ecoregion files so that those domains can mint master codes without colliding with the land hierarchy. This is the same design idea that underlies the allocation of numeric ranges in long-lived technical standards: by fixing the meaning of ranges in advance, the system permits distributed, non-colliding extension over time. The master code space is what allows the five files to remain physically separate yet logically integrated: a code minted in the ecoregion file is guaranteed, by the partitioning, to be unique across the entire framework and to be recognisable, from its range, as ecoregional in origin.

When an entity already possesses a code in the land or water files, the ecoregion file reuses it rather than minting a new one; only where no global code exists is a fresh master code derived from a local shortcut. This reuse-before-mint discipline is the operational expression of the framework's integration: identity is assigned once, at the most authoritative available layer, and inherited everywhere else.

5.5 Coordinates as orientation, not geometry

The framework attaches geographic coordinates to entities but is emphatic that these are orientation anchors only — indicative points, not boundaries, polygons or inputs to a geographic information system — and that no spatial geometry is planned. This is a consequential scoping decision, and it is defensible on its own terms: the framework's purpose is reference and crosswalk, not spatial analysis, and a point anchor is sufficient to orient a human reader or to seed a geocoding lookup while imposing none of the maintenance burden or false precision of polygon boundaries. The decision nonetheless imposes a hard ceiling on applicability, discussed in Chapter 14, because containment queries, mapping and spatial overlay — the operations for which the ecoregion sources were originally built as GIS layers — lie permanently outside the framework's reach.

6 Construction Methodology and Editorial Method

6.1 Sources, precedence and the reuse-before-mint rule

The framework is assembled by ingesting recognised standards and adding proprietary extensions only where those standards leave gaps. A precedence discipline governs the ingestion: where an authoritative external code exists, it is adopted and attributed to its owner; where several authorities disagree, all are recorded through the notation and the attribute tags; and only where no external authority reaches does the framework mint its own identifier, in a reserved range declared by the namespace layer. This precedence — external before internal, reuse before mint — is what keeps the framework anchored to recognised standards rather than drifting into a private ontology, and it is the methodological counterpart to the architectural code-space partitioning of Chapter 5.

6.2 The treatment of contested and historical entities

The framework's editorial method is most clearly revealed in hard cases. Contested territories are recorded neutrally, with claims noted and no sovereign parent asserted, which is the scholarly convention for disputed status and which the framework operationalises through its alternatives notation. Historical entities are retained rather than deleted, carried in dedicated ranges of the master code space, so that time-series data referring to former states or former names remains resolvable. Withdrawn identifiers are redirected to their successors rather than removed. Taken together these practices implement a policy of non-destructive maintenance: the reference record grows and is annotated but does not lose information, which is the property that makes it usable for longitudinal and historical work.

6.3 Reconciling competing editions: the hydrographic case

The water file's reconciliation of the several IHO editions and the defence and aeronautical delimitations is the framework's paradigmatic act of conflict preservation, and it directly mirrors the cross-reference tables that the standards bodies themselves publish. Rather than declaring one edition current and discarding the rest — which no authority has been able to do, given the two-decade impasse over the successor to the 1953 edition — the framework tabulates the correspondences, retaining each edition's code for each body of water alongside the others and marking where extents, not merely codes, diverge. This is not a workaround for missing consensus; it is a faithful representation of a domain in which consensus does not exist, and it is arguably more honest than any system that would manufacture a false single answer.

6.4 Distribution as derivative, open artefacts

The working files — spreadsheets and a word-processing document — are not published; the published artefacts are derived renderings in portable, non-

proprietary formats: a human-readable document form and a delimited tabular form. The tabular renderings use tab separation for the narrower files and semicolon separation for the wider ones, the semicolon chosen precisely because it does not occur within the source cells and can therefore serve as an unambiguous field delimiter. This separation of working format from distribution format is good data-publishing practice: it decouples the maintainer's tools from the consumer's, guarantees that the public artefacts carry no proprietary dependency, and fixes a stable public address at which the derivatives are served. It also has a governance implication, developed in Chapter 14: because only derivatives are published, the reproducibility of the framework depends on the maintainer's continued rendering of the private working files.

7 Empirical Characterisation of the Corpus

This chapter characterises the framework quantitatively from the version 7.8.2 release dated 18 August 2026. Its purpose is to ground the architectural argument of the preceding chapters in the actual shape and scale of the data, and to make the framework's coverage, its type structure, its code-space partitioning and its cross-reference density concrete. All figures in this chapter are taken directly from the release files; where a figure derives from the source standards rather than from the HLC cells, this is stated.

7.1 Overall scale and file composition

The five files comprise, in aggregate, on the order of eleven thousand populated data rows distributed across twenty-eight worksheets, of which five are informational front matter, four are code-space or type definitions, and the remainder carry data. The land file is by a wide margin the largest single data table; the ecoregion file is the most internally heterogeneous, holding twelve distinct classification sheets. The distribution of substance across the files is summarised below.

File	Data sheets	Principal data table	Populated rows	Max columns
codehx	— (prose front matter)	—	7 numbered sections	—
codehu	DATA (HU), COPY (HD)	DATA (HU)	761 (+248 DIAM copy)	6
codehl	DATA (HL) + 3 control sheets	DATA (HL)	4,146	18
codehy	DATA (HY), DIAM (HD), IHO (HI)	DATA (HY)	1,456	14
codehw	12 classification sheets	TEOW / BEOW largest	~4,600 across sheets	7

Two observations follow directly from these figures. First, the framework's centre of gravity is the land file, which alone holds more populated rows than the water and namespace files combined and which carries the widest rows at eighteen columns; this is the physical manifestation of its role as master integrator. Second, the ecoregion file's substance is spread across many sheets of similar structure rather than concentrated in one, which is the physical manifestation of its deliberate refusal of a master sheet.

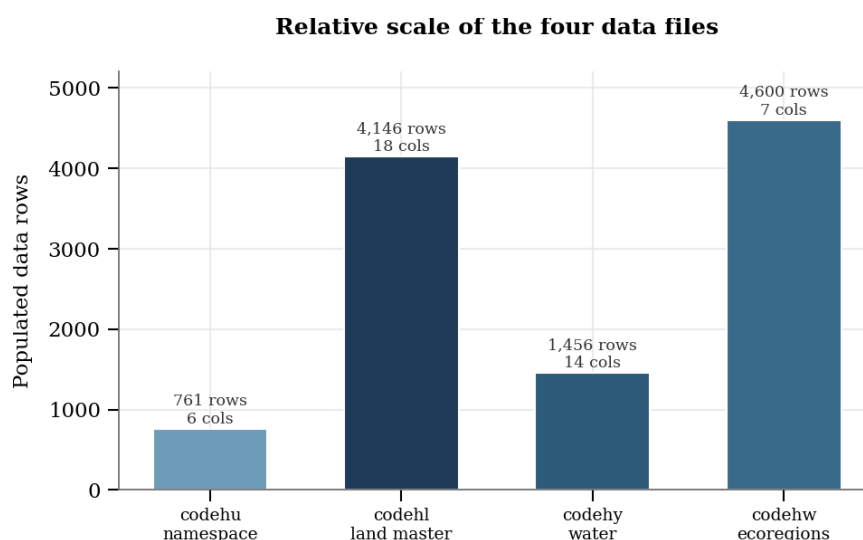


Figure 7.1 Relative scale of the four data files (populated rows and column widths). *codehl* dominates as the master integrator; *codehw* is spread across twelve sibling sheets.

7.2 The type structure of the land hierarchy

Every entry in the land file carries a one-letter type drawn from the parallel Earth/Water taxonomy. The empirical distribution of these types across the 4,146 land rows is informative, because it reveals what the file actually is: not primarily a list of countries but a far larger structure of sub-national divisions, historical entities and pseudo-regions within which the countries are a minority.

Type	Meaning (per ## Types sheet)	Count
D	Division (level-4 administrative)	886
V	Subdivision (level-4 sub-administrative)	750
H	Historical entity	663
U	Unit / lower structure	528
S	Subregion	315
C	Country (level-3)	252
X	Unspecified / technical	221
M	Multiples (cross-state grouping)	181
E	Economic region	94
R	Region	77
W	Water-typed entry	77
T	Tectonic	36
N / G / P / I / L	Null / Global / Province / Division- I / Logical	21 / 15 / 14 / 9 / 6

The single most consequential figure here is that countries proper account for 252 of the 4,146 entries — about six per cent. The overwhelming majority of the file describes structures above the country (regions, subregions, multiples,

economic and geographical groupings) and below it (886 divisions and 750 subdivisions), together with a substantial historical layer of 663 retained former entities. This confirms empirically the claim of Chapter 8 that the land file is a deep hierarchy anchored on, but far exceeding, the roster of present-day states, and it demonstrates that the framework's value lies precisely in the connective tissue around the countries rather than in the country list itself.

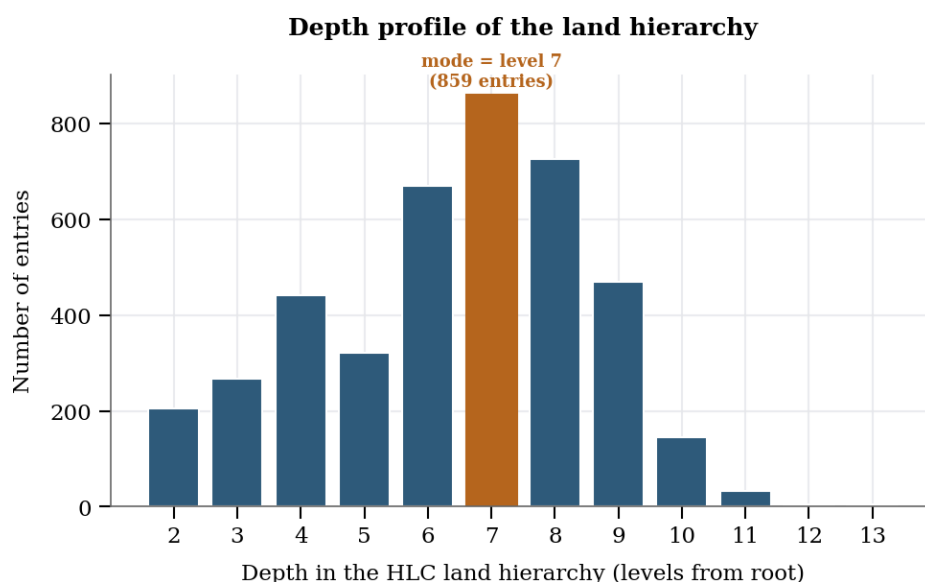


Figure 7.3 Depth profile of the land hierarchy (levels from the root of the HLC tree). The distribution peaks at level 7 and reaches thirteen levels, confirming a deep, uneven mereological structure.

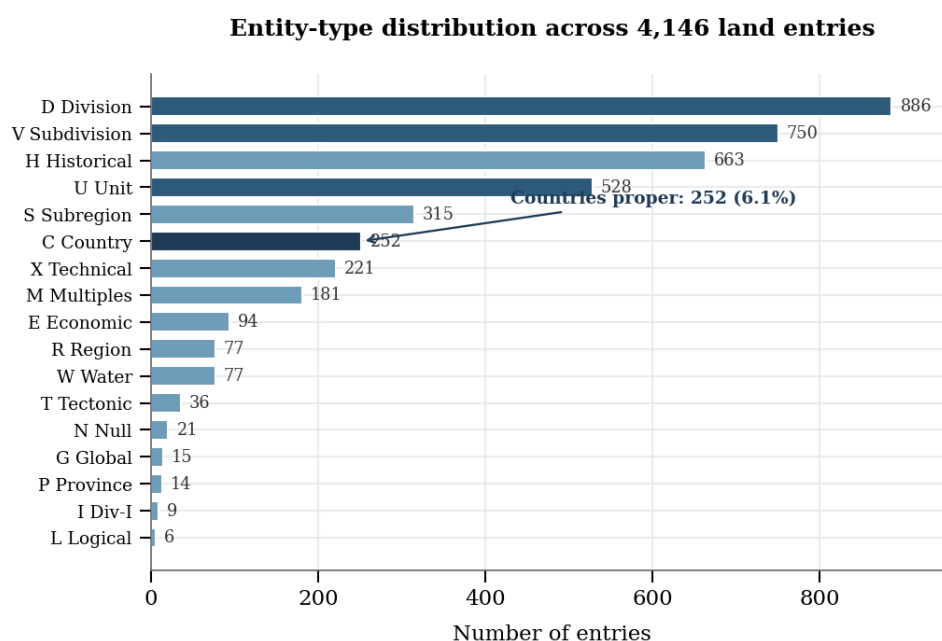


Figure 7.2 Entity-type distribution across the 4,146 land entries. Countries proper (type C) are only 6.1% of the file; divisions, subdivisions, units and historical entities dominate.

7.3 The partitioned master code space in practice

The master integrator's numeric code space is partitioned by leading digit and by reserved sub-ranges, with each range assigned to an owning file (HL for land, HY for water, HW for ecoregions, or an external-source escape range). The

empirical partitioning, read from the HLC-M definition sheet, is reproduced in condensed form below; it shows how a single flat numeric space is carved so that the physically separate files can mint non-colliding, self-identifying codes.

Range (leading digit / band)	Owner	Content
0nn	HL	Special, organizational, historical, world islands
1nn / 2nn / 3nn	HL	Countries — Asia / Africa / Europe
4nn(n)	HL	Northern America (countries & mainland waters)
53-55n	HL	Southern America
56-59nn	HW	Terrestrial sub-ecoregions & sub-realms
60nn / 610-688	HY / HL	Mainland waters / Australasia, Oceania, Antarctica
70-79	HY	Oceans, seas and derived CWWN+ codes
800-810 / 897-899	HW	Realms, ecosystems, marine ecoregions (tree level)
811-86n	HL	Sub-structures by continent
87-88 / 890-896	HL	World and regional organizations
90-99	HL	Waters (sort range), main structures & epochs

The design permits controlled external extension: the specification reserves a numeric-plus-letter escape convention (substituting the penultimate digit with a letter P-Y) so that codes ingested from external gazetteers can be admitted without colliding with the framework's own ranges. This is the code-space embodiment of the reuse-before-mint and namespace-ownership disciplines described in Chapters 5 and 6: the partition is not decorative but load-bearing, since it is what allows a code minted in the ecoregion file to be globally unique and recognisable as ecoregional from its range alone.

Partitioning of the master (HLC-M) numeric code space

0nn	HL	Special / historical / world islands
1nn	HL	Asia (countries)
2nn	HL	Africa (countries)
3nn	HL	Europe (countries)
4nnn	HL	Northern America + waters
53-55n	HL	Southern America
56-59nn	HW	Terrestrial sub-ecoregions
60nn	HY	Mainland waters
610-688	HL	Australasia / Oceania / Antarctica
70-79	HY	Oceans, seas, CWWN+
800-810	HW	Realms & ecosystems
811-86n	HL	Sub-structures by continent
87-88	HL	World & regional organizations
890-896	HL	Regional organizations
897-899	HW	Marine ecoregions
90-99	HL	Waters sort / structures / epochs

HL structural / org / historical
 HY hydrographic (reserved)

HL geopolitical (countries/regions)
 HW ecoregional (reserved)

Figure 7.4 Partitioning of the master (HLC-M) numeric code space by leading band and owning file. Reserved ranges let the physically separate files mint globally unique, self-identifying codes.

7.4 Cross-reference density: the integrated schemes

The land file's columns are, in effect, a register of the external classifications the framework harmonises. Beyond the ISO, UN M49 and GENC identifiers, the documented inline-attribute-tag schemes in the land file alone include the Correlates of War project, the World Geographical Scheme for Recording Plant Distributions, the terrestrial, biological and freshwater ecoregion adaptations, the World Waterbody Names scheme, the Treasury International Capital and World Area Codes classifications, the ITU and NATO code sets, the EU Geonomenclature, and a series of national statistical classifications — Australian, New Zealand, German, Italian, British, Argentine, South African, Japanese, Swiss, French, Norwegian and Canadian — together with the Unicode CLDR exceptions and several historical-information sources. More than thirty distinct external schemes are thus cross-referenced from a single file, and the water and ecoregion files add the fisheries, hydrographic and oceanographic schemes on top.

The density is best shown by a worked example. The entry for Germany occupies one row yet resolves simultaneously into more than a dozen external registers, each held in its own tagged slot, exactly as the architecture of Chapter 5 predicts:

Tree 930-961-939-839-390 · HLC-M 390 · type C · ISO DE / DEU · {UN:276} {LO:024} {OA:ED} · {CW:260,GFR} {PD:GER} · {TI:1-100-2} {WC:429-2} · {R1:D,49} · {EU:(-11-)004} {NT:GE,DEU} ·

{AU:2304} · {IT:216} · {AR:438} {ZA:417} {JP:213} · {CH:(13)207}
{FR:142} {NO:144} · (Deutschland).

A single reference lookup on this row therefore yields the entity's UN statistical code, its labour-organization and civil-aviation prefixes, its Correlates-of-War and plant-distribution codes, its US capital-flow and transport codes, its telecommunication region, its EU and NATO codes, and its representation in seven national statistical systems — the crosswalk value proposition made fully concrete. Because of the attribute-locality rule, all of these attributes live only here, at the canonical entry, so no alternative-hierarchy occurrence of Germany elsewhere in the framework can contradict them.

7.5 The ecoregional sheets and their source scales

The ecoregion file assembles twelve classification sheets, each importing a recognised system at its own native resolution and re-expressing it under the shared column pattern of hierarchy path, master code, technical and shortcut codes, source reference code and description. The source scales, as documented in the file's own front matter and confirmed by the sheet sizes, are as follows.

Sheet	System (source)	Native scale	HLC rows
TEOW (TW)	Terrestrial Ecoregions (Olson et al. 2001)	8 realms, 14 biomes, 867 ecoregions	1,150
BEOW (BW)	Bioregions (RESOLVE, Dinerstein et al. 2017)	8 realms, 14 habitats, 846 ecoregions	1,172
FEOW (FW)	Freshwater Ecoregions (Abell et al. 2008)	8 realms, 426 ecoregions	554
MEOW (MW)	Marine Ecoregions (Spalding et al. 2007)	12 realms, 62 provinces, 232 ecoregions	367
PPOW (PW)	Pelagic Provinces (Spalding et al. 2012)	4 realms, 7 biomes, 37 provinces	101
GOODS (OW)	Open Oceans & Deep Seabed (Watling 2013)	benthic & pelagic deep-sea units	162
EUNIS (SW)	European Nature Information System (2021)	marine + terrestrial habitat hierarchy	680
GIWA (GW)	Global Int'l Waters Assessment (1999)	9 megaregions, 66 regions	137
GEZ (ZW)	FAO Global Ecological Zones (2000)	5 zones, 20 ecoregions	38
LMES (LW)	Large Marine Ecosystems (NOAA)	5 realms, 66 ecosystems	90
PAME (EW)	Arctic Marine Environment (2018)	2 realms, 18 ecosystems	117

The WWF Global 200 priority set (238 ecoregions — 142 terrestrial, 53 freshwater, 43 marine) and the Longhurst biogeochemical provinces are woven through these sheets as cross-cutting grouping items rather than as separate hierarchies, and the front matter records candidly where the framework has added its own grouping items to compensate for gaps in a source's regulation — for instance the harmonising items added to align the RESOLVE bioregions with the Olson terrestrial ecoregions. This is the empirical face of conflict preservation: the sources are neither merged nor silently patched but carried in parallel with their divergences and the maintainer's bridging items explicitly marked.

The twelve ecoregional sheets by size (codehw)

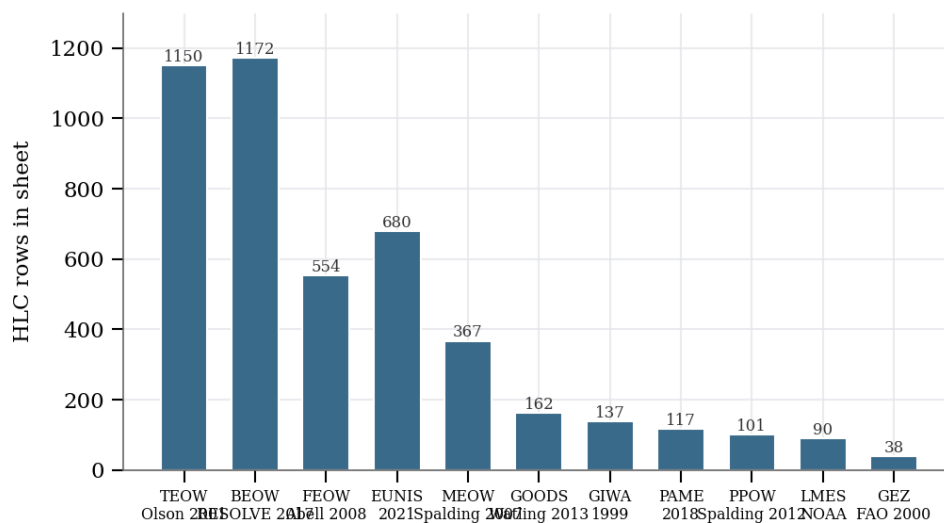


Figure 7.5 The twelve ecoregional sheets of codehw by size, from TEOW (1,150 rows) to GEZ (38). Each imports a recognised classification at its native resolution under a shared column pattern.

7.6 The hydrographic multi-edition crosswalk

The clearest empirical demonstration of the framework's conflict-preserving stance is the dedicated IHO reconciliation sheet in the water file, which tabulates each named water body against its identifier in five distinct authorities side by side: the 2002 revision, the never-ratified fourth-edition draft, the in-force third edition of 1953, the older aeronautical-chart delimitation, and the defence manual. A single row of that sheet illustrates the many-to-many reality directly.

Water body	HLC-M	IHO 2002	IHO 4th	IHO 3rd	ACIC M49	DIAM+
North Atlantic Ocean	733	1	1	23	B	2L(1A) OK
Skagerrak	7435	1.1	1.1	3	B24	1S
North Sea	7419	1.2	1.2	4	B23	1N
St. George's Channel	7455	—	1.4	19	B22	1Q
Irish Sea	7452	1.4	1.4	19	B22	1R

The table makes visible what prose can only assert: that the same body of water carries different identifiers under each authority, that some bodies are absent from one edition and present in another (St. George's Channel has no 2002 identifier), and that the numbering schemes are structurally incompatible — the third edition's flat sequence bears no relation to the hierarchical decimal of the later drafts. A framework that adopted any one column and discarded the rest would misrepresent a domain that its own governing body has been unable to unify for more than three decades. The HLC water file instead preserves every column, which is the honest representation and the framework's central thesis realised in a single sheet.

PART IV

The Subsystems

8 The Land Subsystem (codehl)

8.1 Scope and role as master integrator

The land file is the largest and the most central of the five, and it plays a dual role: it carries the geopolitical hierarchy from continents through regions and sub-regions to countries, dependencies and sub-national units, and it hosts the master integrator — the partitioned numeric code space that anchors identity for the entire framework. Because the master code space lives here, the land file is not merely one subsystem among equals but the layer on which the water and ecoregion files depend for their code allocations. This centrality reflects a substantive claim about the domain: the geopolitical partition of land is the most stable, the most widely standardised and the most densely cross-referenced of the three domains, and is therefore the natural anchor for a system that must bind all three.

8.2 The parallel type taxonomy

Alongside the entity hierarchy the land file maintains a taxonomy of entity types organised by level — from the global tier down through region, sub-region and country to the smallest units — with typed variants that mark, for example, historical, economic and botanical readings of an otherwise geographic unit. This typing is what allows the framework to carry, within one file, several logically distinct partitions of the same territory: the ordinary administrative hierarchy, the historical succession of states, and the economic or biogeographic groupings that cut across administrative lines. The type taxonomy is the mechanism by which the controlled polyhierarchy of Chapter 3 is made legible: each alternative parenthood is tagged with the type of hierarchy it belongs to, so that a reader or a parser can select a single consistent view — the purely administrative view, say, or the historical view — by filtering on type.

8.3 Crosswalk density and the multi-attribute entry

A mature entry in the land file is a dense object. Beyond its master code and its position in the hierarchy, it may carry its ISO alphabetic and numeric codes, its UN statistical code, its GENC identifier and a substantial set of inline attribute tags binding it to national and institutional classifications and, where relevant, to the hydrographic and ecoregional schemes. This density is the framework's value proposition made concrete: it is precisely because a single entry gathers an entity's identifiers across many authorities that the framework can serve as a crosswalk. The cost of the density is the width of the file, which is what motivates the semicolon-delimited distribution format for the wider files described in Chapter 6.

8.4 Assessment of the land subsystem

The land subsystem's strengths are its anchoring role, its stable and partitioned code space, and its crosswalk density; its principal vulnerability is that the very centrality that makes it valuable also concentrates risk, since an error or inconsistency in the master integrator propagates to every dependent file. The attribute-locality discipline of Chapter 3 mitigates one class of such errors by guaranteeing that an entity's attributes cannot be internally contradicted, but it cannot guard against an erroneous assignment at the single authoritative location, which by construction propagates everywhere. The subsystem is, in short, as reliable as the editorial judgement embodied at its anchor points — a dependency the thesis examines under governance in Chapter 14.

9 The Water Subsystem (codehy)

9.1 Scope

The water file organises the hydrographic domain: oceans, seas, gulfs and straits at the surface; rivers and lakes on the continents; and currents, gyres and undersea features. Its structural logic follows the land file's controlled polyhierarchy — a strait allotted to one sea under the 1919 convention yet bordering two is represented through the alternatives notation — but its content is governed by a different and more fractured set of authorities, because, as Chapter 2 established, the hydrographic domain has no single ratified delimitation.

9.2 Multiple hydrographic standards in parallel

The file carries several coding systems side by side. A defence-derived system provides ocean and sea codes; a world-waterbody naming system, extended by the framework, supplies a broader inventory; and a reordered area system aligns with the delimitations used in fisheries and marine management. Above these, inline attribute tags bind each water body to the regional and thematic classifications — the fisheries management areas, the marine-management commissions, and the oceanographic province and ecoregion systems — under which it is also known. Two dedicated attribute sheets carry the definitional detail for oceans and seas and reconcile the several editions of the IHO delimitation against one another and against the aeronautical delimitation, so that the same body of water can be resolved across the in-force 1953 edition, the never-ratified 1986 draft, and the older aeronautical chart, each retaining its own code and, where they differ, its own extent.

This is the framework's conflict-preserving stance at its most consequential. The IHO itself has been unable, across more than thirty years, to ratify a successor to its 1953 edition, and has ultimately adopted numeric identifiers precisely to sidestep an unresolvable naming dispute. A crosswalk that recorded only one edition would misrepresent the domain; the water file's parallel retention of all of them is the accurate representation of a genuinely unsettled field.

9.3 Assessment of the water subsystem

The water subsystem is where the framework's central thesis is best vindicated, because the hydrographic domain is objectively one in which no single authority can be adopted without loss. Its strength is the faithful, machine-actionable reconciliation of incompatible editions; its limitation is inherited from its sources — the absence of ratified consensus means that any downstream user must still choose an edition for any given purpose, and the framework, by design, declines to make that choice for them. As in the land file, the burden of final

selection is preserved and passed to the user, which is the honest but demanding consequence of conflict preservation.

10 The Ecoregional Subsystem (codehw)

10.1 Scope and the deliberate absence of a master

The ecoregion file assembles the major biogeographic classifications of the living surface — terrestrial, freshwater, marine coastal and benthic, and surface pelagic, together with regional and thematic systems — as parallel sibling views. Its defining structural decision is that, unlike the land file, it contains no master sheet: no one classification is privileged as the anchor of the others. This faithfully reflects the state of biogeography, in which competing regionalisations coexist as legitimate alternatives rather than as approximations to a single correct partition. The file's role is not to adjudicate among these systems but to make them mutually navigable, so that a datum expressed in one can be located relative to the others.

10.2 Common structure across heterogeneous sources

Although the source classifications differ in content and in the number of their tiers, they share the same formal shape — a nested hierarchy of realm, an intermediate tier of province or biome, and a base tier of ecoregion — and the file exploits this shared shape by imposing a common column pattern across all of them: a hierarchy path, the framework's master and local codes, the source's own reference code, and a description. This uniform treatment is what makes the otherwise incommensurable systems comparable within one file. Where a unit already possesses a master code from the land or water files — a coastal ecoregion adjacent to a coded sea, for instance — that code is reused; where it does not, a fresh master code is derived from the local shortcut within the ecoregional range reserved in the master code space. Priority-ecoregion groupings that cut across the base classifications are woven in as cross-cutting tags rather than as a separate hierarchy, which again reflects the reuse-before-mint and preserve-alternatives disciplines that run through the whole framework.

Represented system	Domain	Formal tiers
Terrestrial ecoregions	Land biota	realm / biome / ecoregion
Freshwater ecoregions	Inland waters	realm / habitat / ecoregion
Marine coastal ecoregions	Shelf seas	realm / province / ecoregion
Pelagic provinces	Surface open ocean	realm / biome / province
Deep-sea benthic/pelagic	Open ocean & seabed	broad benthic / pelagic units
Ecological zones & bioregions	Cross-cutting syntheses	zone / biome / ecoregion

10.3 Assessment of the ecoregional subsystem

The ecoregional subsystem is the clearest illustration of the framework's polyhierarchical philosophy, because here the refusal of a single master is not a

concession but a correct description of the science. Its strength is that it brings systems designed in isolation — and licensed and maintained by different bodies — into a single navigable structure without distorting any of them. Its principal limitation is one of currency and geometry: the source classifications are periodically revised, and, being GIS layers in origin, they carry spatial boundaries that the framework, by its coordinate policy, does not reproduce. A user needing the actual ecoregion polygons must return to the source datasets; the framework offers the crosswalk and the identifiers, not the geometry.

11 Namespace Governance (codehu) and the Shared Front Matter (codehx)

11.1 The namespace layer as the guarantor of identity

The namespace file is the least conspicuous and most structurally important of the data files, because it is what prevents the framework's ambition from collapsing into collision. Its function is to declare, for each region of the code space, which authority owns it: which rule adjudicates two-character codes, which adjudicates three-character codes, and how numeric ranges are partitioned among the external standards and the framework's own extensions. It also records the distinction between land-based and water-based entries within the shared code space, so that a code's domain can be read from its governance entry. Without this layer, the ingestion of codes from many authorities would produce ambiguity the moment two authorities issued the same string; with it, every identifier resolves to an owning authority, which is the precondition for both provenance and safe extension.

In the vocabulary of Chapter 3, the namespace file is where the framework's identifier discipline is made explicit and auditable. It is also, in effect, the framework's constitution: the document that fixes who may assign what, and therefore the document that makes distributed, non-colliding growth of the reference record possible over time.

11.2 The front matter as the framework's semantic contract

The front matter is a short prose document, but it carries disproportionate weight, because it is the single source of truth for everything the data files leave unstated: the meaning of every notation operator, the editorial conventions, the coordinate policy, and — importantly — the framework's declared limitations. By factoring these into one shared preface, the framework avoids restating them across five files and, more subtly, guarantees that all five files speak the same language. The front matter is thus the semantic contract between the maintainer and every consumer of the derivatives: it is what makes the dense notation of Chapter 5 decodable and what makes the framework's scope-claims — reference and crosswalk, not authority, not geometry, not legal or spatial use — explicit rather than implied.

11.3 Declared limitations as an epistemic virtue

It is unusual, and to the framework's credit, that the front matter states plainly what the system is not. It disclaims the status of an authority; it disclaims spatial geometry; it declares itself a snapshot rather than a live service; it acknowledges its dependence on a single maintainer's editorial judgement and on uneven granularity inherited from its sources. In the terms of research-data ethics this candour is itself a form of quality: a reference resource that documents its own

boundaries enables its users to use it correctly and protects them from misapplication. The declared limitations are not a disclaimer appended for caution; they are, read carefully, an accurate specification of the framework's intended envelope of use.

PART V

Assessment and Outlook

12 Evaluation

This chapter evaluates the framework against three recognised frames: the FAIR principles for research data, the criteria for crosswalk quality, and the framework's fitness for its declared purposes. The evaluation is deliberately even-handed, separating genuine achievement from aspiration.

12.1 Against the FAIR principles

On findability, the framework scores well: entities carry globally unique, stable identifiers in a partitioned code space, and the derivatives are served at a fixed public address. On accessibility, it scores well: the public artefacts are open, non-proprietary and retrievable without special tooling. On interoperability, it scores highly by design, since the entire purpose of the inline attribute tags is to provide qualified references binding each entity to its representations under named external schemes — this is the FAIR interoperability requirement realised almost literally. On reusability, the picture is mixed: provenance is strong at the level of scheme ownership, because every code resolves to an owning authority, but weaker at the level of individual editorial decisions, since the rationale for a particular alternative-hierarchy assignment lives in the maintainer's judgement rather than in an auditable record. The framework is, in FAIR terms, strongly interoperable and findable, adequately accessible, and reusable subject to the caveat that some provenance is tacit.

FAIR dimension	Mechanism in HLC	Assessment
Findable	Stable unique codes; fixed public address	Strong
Accessible	Open, non-proprietary derivative formats	Strong
Interoperable	Inline attribute tags as qualified cross-references	Strong (by design)
Reusable	Scheme-level provenance; declared limitations	Adequate; decision-level provenance tacit

12.2 Against crosswalk-quality criteria

A crosswalk is judged by its coverage, by the fidelity with which it represents many-to-many correspondences, and by its handling of the cases where schemes genuinely disagree. On coverage, the framework is unusually broad, spanning geopolitical, hydrographic and ecological domains that are almost never brought into a single concordance. On fidelity to many-to-many correspondence, it is strong, because its notation and attribute tags are expressly built to record alternatives and multiple parenthoods rather than forcing one-to-one maps. On the handling of disagreement, it is exemplary in the specific sense that it neither

hides nor resolves conflict but records it — the reconciliation of the several hydrographic editions being the clearest case. The framework's crosswalk weakness is not representational but evaluative: it tells the user that alternatives exist and what they are, but it offers limited machine-readable guidance on which alternative to prefer for a given purpose, because supplying such guidance would require exactly the canonicalisation the framework declines.

12.3 Fitness for declared purpose

Judged against its own declared envelope — a specialist reference and crosswalk instrument, explicitly not an authority, not a GIS and not a legal source — the framework is fit for purpose to a high degree. It does what it says it does: it gathers, aligns and cross-references heterogeneous classifications while preserving their disagreements, and it publishes the result in open, stable, machine-readable form. Judged against purposes it explicitly disclaims — cartographic rendering, spatial containment analysis, or use as an authoritative legal register — it is, as its own front matter warns, unsuitable. The correct evaluative conclusion is therefore conditional: within its declared scope the framework is a strong and unusual instrument; outside that scope it is, by its own honest account, inapplicable.

12.4 Comparative positioning

No directly comparable artefact exists that spans all three domains with this stance. Institutional gazetteers and authority files typically cover one domain and canonicalise within it; the standards bodies themselves publish domain-specific cross-reference tables — the IHO's own appendices being a notable example — but do not integrate across the geopolitical, hydrographic and ecological registers, and do not present their crosswalks as a single coordinated, openly distributed system. HLC's distinctive contribution is precisely this cross-domain integration under a uniform notation and a shared code space. Its distinctiveness is also its vulnerability: an artefact with no peer has no external community to validate, correct or continue it, which returns the analysis to the question of governance.

13 Applied Perspectives and Use Scenarios

The preceding chapters established what the framework is and how well it is built. This chapter asks what it is for, by working through concrete scenarios in which a conflict-preserving crosswalk earns its keep, by comparing HLC with the reference resources a practitioner might otherwise reach for, and by identifying the uses for which it is unsuited. The scenarios are deliberately drawn from different disciplines, because the framework's distinguishing claim is cross-domain integration, and its value is best judged where a single task must cross the boundaries between the political, hydrographic and ecological registers.

13.1 Scenario: harmonising cross-border epidemiological data

Consider an analyst assembling a multi-decade panel of health indicators from national statistical offices, the World Health Organization and academic sources. The records arrive keyed variously by ISO alpha-2, by UN M49 numeric code, by national statistical codes and, for older data, by the names of states that no longer exist. The recurring difficulty is not any single translation but their accumulation: a former state must be resolved to its successors for the years after its dissolution and retained for the years before; a territory reported under a national scheme must be matched to its ISO identity; a UN regional aggregate must be decomposed into countries. HLC addresses precisely this pattern. Because each entry gathers its ISO, M49 and many national codes in one row, a single join resolves most of the incoming keys; because historical entities are retained with redirects to their successors, the time dimension is handled without dropping the pre-dissolution years; and because the master code is stable under name change, the panel's key remains constant even where the underlying name does not. The scenario exercises the crosswalk and namespace lenses together, and it is one in which the framework's breadth of national-scheme coverage — Australian, German, Italian, Japanese, Swiss, French, Norwegian, Canadian and others — is directly useful.

13.2 Scenario: integrating marine observations across hydrographic authorities

An oceanographer pooling observations from several national cruises confronts the hydrographic domain's central problem: the same body of water is named and bounded differently by the authorities under which each cruise reported. One dataset delimits the seas by the in-force 1953 edition; another follows a fisheries-management scheme; a third uses a defence delimitation. There is no ratified single answer to adopt, as the two-decade impasse over the successor to the 1953 edition attests. HLC's water file is built for exactly this situation. Its dedicated reconciliation sheet tabulates each sea against its identifier in the 2002 revision, the fourth-edition draft, the third edition, the aeronautical delimitation and the defence manual, so that an observation reported under any one can be located under all the others. The framework does

not tell the oceanographer which delimitation to prefer — that remains a scientific judgement — but it makes the alternatives explicit and mutually resolvable, which is the necessary precondition for pooling the data at all. This scenario is the epistemic lens made operational: conflict preservation is not an abstraction here but the difference between being able to merge the datasets and not.

13.3 Scenario: linking biodiversity records to a coherent ecoregional frame

A conservation scientist aggregating species-occurrence records wishes to summarise them by ecoregion, but the records are tagged inconsistently — some to terrestrial ecoregions, some to freshwater units, some only to a country or a marine province. The ecoregional classifications were designed in isolation and do not natively cross-reference one another. HLC's ecoregion file assembles the terrestrial, freshwater, marine, pelagic and deep-sea systems under a shared column pattern and a common master code, with the priority-ecoregion groupings woven across them, so that a record tagged in any one system can be related to its neighbours in the others and rolled up to a realm consistently. Where a coastal ecoregion already carries a master code from the land or water files, that code is reused, so the ecological and geopolitical summaries share a key. The scenario exercises the mereological and crosswalk lenses, and it exposes the framework's honest limit: because HLC holds no polygons, the scientist must still return to the source GIS layers to assign a raw coordinate to an ecoregion. HLC harmonises the identifiers, not the geometry — a division of labour the framework states plainly and the scenario respects.

13.4 Scenario: a historical time-series across changing states

A political scientist constructing a long-run series of state attributes needs identifiers that survive the appearance, disappearance, merger and renaming of states across a century. Most reference schemes are poorly suited to this, because they optimise for the present and treat the past as withdrawn or absent. HLC's retention of 663 historical entities in dedicated ranges of the master code space, together with its redirect-rather-than-delete policy, makes it unusually apt for the task: a datum from a former state resolves to a stable historical code, and the redirect links it to the present-day successor for analyses that require continuity. The framework's cross-reference to the Correlates of War project — a standard source in the discipline — means the series can be tied to the field's established identifiers as well. The scenario demonstrates that conflict preservation extends across time as well as across authorities: the past is neither erased nor silently mapped onto the present but retained and explicitly connected.

13.5 Comparison with alternative reference resources

A practitioner weighing HLC will naturally compare it with the resources already in common use. The comparison is best made on the dimensions that matter for cross-domain harmonisation: the breadth of domains covered, the treatment of conflicting authorities, the presence of spatial geometry, the governance model, and the handling of historical entities. The table summarises the comparison in neutral terms; none of these resources is a strict substitute for another, because they were built for different purposes.

Resource	Domain breadth	Conflict handling	Geometry	Governance	History
HLC	Geopolitical + hydrographic + ecological	Preserves multiple authorities explicitly	None (point anchors)	Single maintainer	Retained + redirects
ISO 3166	Geopolitical only	Canonical single answer	None	Institutional (ISO)	Withdrawn codes (Part 3)
UN M49	Geopolitical + statistical groupings	Single statistical view	None	Institutional (UN)	Limited
GeoNames	Places (geopolitical + features)	Single merged view	Point coordinates	Community + owner	Limited
Wikidata	Universal (incl. places)	Multiple statements, sourced	Coordinates + shapes (via links)	Open community	Strong (historical items)
GADM / Natural Earth	Administrative boundaries	Single authored view	Full polygons	Project teams	Limited

Two observations follow. First, no listed resource matches HLC's simultaneous span of the geopolitical, hydrographic and ecological registers under one notation and one code space; the boundary datasets and the single-domain standards are each deeper in their own territory but narrower across territories. Second, the resource closest to HLC in spirit is Wikidata, which likewise records multiple sourced statements rather than a single answer; but Wikidata is a general knowledge base with community governance and linked geometry, whereas HLC is a specialised, internally consistent, single-author crosswalk with a designed code space and no geometry. The two are complementary rather than competing: HLC's disciplined cross-domain identifiers could enrich a Wikidata-based workflow, and Wikidata's geometry and community scale could supply what HLC by design does not.

13.6 Anti-patterns: where HLC should not be used

Responsible advocacy for a tool includes stating where it must not be applied. HLC is unsuited to any task requiring spatial computation — mapping, containment, adjacency or overlay — because it holds no boundaries; a user needing these must use a geometry dataset and may, at most, use HLC's identifiers as a join key to one. It is unsuited to service as a legal or authoritative register of sovereignty or borders, a role its own front matter disclaims, because it records claims without adjudicating them and is not issued by any competent authority. It is unsuited to applications demanding guaranteed real-time currency, because it is a periodically released snapshot rather than a live service. And it is a poor fit for a casual, single-domain lookup that a mainstream standard already serves cleanly, since its expressive notation imposes a learning cost that such a task does not repay. Recognising these anti-patterns is not a concession but a condition of using the framework well: its value is real precisely where these constraints do not bind.

13.7 Synthesis: the niche the scenarios reveal

Taken together, the scenarios delineate a coherent niche. HLC is most valuable when a task must cross domain boundaries, must reconcile several legitimate authorities without erasing their differences, and must remain stable across changes of name and across time — and when the user can supply spatial geometry separately if it is needed at all. This is not a mass-market niche, but it is a real and under-served one, occupied by researchers and data engineers who work at the seams between the political, hydrographic and ecological orderings of the Earth and who have, until now, had to stitch those seams by hand. The framework's applied contribution is to have done that stitching once, systematically, and to have published the result in a form that others can reuse.

14 Limitations: Contingent and Constitutive

RQ4 asked which of the framework's limitations are contingent — remediable without altering the design — and which are constitutive — irreducible consequences of what the framework is. This chapter answers that question directly, because the distinction determines what could be improved and what must simply be accepted.

14.1 Constitutive limitations

- **Absence of spatial geometry.** The coordinate-as-anchor policy is a defining scope decision, not an omission. Because the framework holds no boundaries, it cannot support containment, overlay or mapping, and no incremental effort would change this without turning the framework into a different kind of artefact. This limitation is constitutive.
- **Residual editorial judgement.** Conflict preservation relocates judgement from the assignment of a single answer to the curation of legitimate alternatives; it does not eliminate judgement. Which alternatives are legitimate enough to record is an irreducibly interpretive decision. This limitation is constitutive of any conflict-preserving design.
- **Burden of final selection on the user.** Because the framework declines to canonicalise, a user must still choose an alternative for any concrete purpose. This is the necessary complement of preservation and cannot be removed without abandoning the framework's central commitment.
- **Uneven granularity.** Detail is inherited from the sources, which are themselves uneven; the framework cannot be more uniform than the standards it integrates. This limitation is constitutive so long as the framework remains faithful to its sources.

14.2 Contingent limitations

- **Single-maintainer governance.** The framework is authored and maintained by one person without an editorial board or formal review. This is the most serious limitation, but it is contingent: it could be remedied by institutionalising governance, recruiting reviewers, or transferring stewardship, none of which would alter the design. Until then the framework carries a substantial bus-factor risk and a dependence on one individual's continuity and consistency.
- **Tacit decision-level provenance.** The rationale for individual assignments lives in the maintainer's judgement rather than in an auditable log. This is contingent: an explicit decision record or change log would raise reusability without changing the architecture.
- **Snapshot currency.** The derivatives capture the sources at a moment and drift as the sources evolve. This is contingent on release cadence and

could be mitigated by more frequent, versioned releases and by explicit source-version stamping.

- **Reproducibility dependence on private working files.** Because only derivatives are published, external reproduction of the framework depends on the maintainer's rendering of unpublished spreadsheets. This is contingent: publishing a schema, or the transformation from working files to derivatives, would close the gap.
- **Steep learning curve.** The dense notation demands study of the front matter before use. This is contingent on documentation and tooling — worked examples, a parser, or an interactive lookup would lower the barrier without simplifying the underlying model.

14.3 The central risk restated

The analysis supports a clear conclusion: the framework's design limitations are largely constitutive and therefore honestly unavoidable, whereas its most serious practical limitations are contingent and centre on governance. The principal risk to HLC is not that it is wrongly designed but that it is thinly resourced — that a valuable, unusual and internally coherent artefact depends for its continuity, correctness and currency on a single unpaid individual. This reframes the improvement agenda away from redesign and toward institutionalisation: the most valuable interventions would strengthen governance, provenance and cadence, not the model itself.

15 Conclusions and Future Work

15.1 Answers to the research questions

RQ1 (Positioning). HLC addresses the translation problem among geopolitical, hydrographic and ecological reference systems, and it does so by rejecting the canonicalisation that dominates reference-data practice in favour of conflict preservation: it records the disagreements between authoritative sources as first-class data rather than resolving them away.

RQ2 (Architecture). The framework binds heterogeneous classifications through four coordinated mechanisms — a controlled notation that expresses alternatives, redirects and multiple parenthoods; a controlled polyhierarchy disciplined by attribute locality; a namespace layer that assigns ownership across the code space; and a partitioned master code space that guarantees unique, domain-recognisable identity across five physically separate files — with inline attribute tags serving as the qualified cross-references that turn five hierarchies into a crosswalk.

RQ3 (Evaluation). Measured against the FAIR principles and crosswalk-quality criteria, the framework is strongly interoperable and findable, adequately accessible, and reusable subject to tacit decision-level provenance; it is exemplary in representing and preserving disagreement and correspondingly limited in offering machine-readable guidance on which alternative to prefer. Within its declared scope it is fit for purpose; outside that scope it is, by its own account, inapplicable.

RQ4 (Sustainability). The framework's constitutive limitations — no geometry, residual judgement, user-borne selection, inherited granularity — are irreducible consequences of its design and stance; its most serious practical limitations — single-maintainer governance, tacit provenance, snapshot currency, reproducibility dependence and a steep learning curve — are contingent and remediable. The central risk is one of resourcing and governance, not of design.

15.2 Contribution restated

The thesis has reconstructed HLC as a conflict-preserving crosswalk architecture, situated it against the standards it integrates, evaluated it against recognised criteria, and separated its contingent from its constitutive limitations. The broader theoretical contribution is an articulated defence of conflict preservation as an epistemological stance in reference-data design: in a domain where legitimate authorities disagree, a system that records the disagreement is more truthful than one that erases it, provided the curation of alternatives is transparent. HLC is a substantial, if imperfect, instantiation of that stance.

15.3 Future work

1. **Institutionalise governance.** Establish an editorial board, a review process and a succession plan to convert the single-maintainer artefact into a sustainable community resource.
2. **Publish a formal schema and transformation.** Release the schema of the derivatives and the rendering from working files to public artefacts, closing the reproducibility gap and enabling third-party validation.
3. **Add decision-level provenance.** Maintain an auditable change log and per-assignment rationale to raise reusability without altering the model.
4. **Version-stamp sources and increase cadence.** Record the version of each ingested standard and release more frequently to bound snapshot drift.
5. **Build tooling.** Provide a reference parser, worked examples and an interactive lookup to lower the learning curve and to demonstrate the notation's machine-actionability.
6. **Optional geometry bridge.** Without importing polygons, publish stable joins from HLC identifiers to the source GIS datasets, so that users may combine HLC's crosswalk with external geometry while preserving the framework's scope.

15.4 Closing statement

The HLC Operational Framework is best understood exactly as it presents itself: a specialist, single-author reference and crosswalk system that harmonises the geopolitical, hydrographic and ecological orderings of the Earth without forcing them into a false unity. Its intellectual significance lies in the discipline with which it preserves contested classification, and its practical fragility lies not in that design but in the slender governance that currently sustains it. Strengthen the governance and the framework becomes a durable public good; leave it unresourced and a genuinely original contribution remains hostage to the continuity of one person. The design has answered its own question well; the open question is institutional, and it is answerable.

References

- Abell, R., Thieme, M. L., Revenga, C., et al. (2008). Freshwater Ecoregions of the World: A New Map of Biogeographic Units for Freshwater Biodiversity Conservation. *BioScience*, 58(5), 403–414.
- Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*. MIT Press.
- Defense Intelligence Agency (1994). *Geopolitical Data Elements and Related Features*, Data Standard No. 4 (DIAM 65-18). Washington, DC: Defense Intelligence Agency.
- Dinerstein, E., Olson, D., Joshi, A., et al. (2017). An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *BioScience*, 67(6), 534–545.
- International Hydrographic Organization (1953). *Limits of Oceans and Seas*, Special Publication No. 23, 3rd edition. Monaco: International Hydrographic Bureau.
- International Hydrographic Organization (1986). *Limits of Oceans and Seas*, Special Publication No. 23, Draft 4th edition. Monaco: International Hydrographic Bureau.
- International Organization for Standardization (2020). ISO 3166-1: Codes for the representation of names of countries and their subdivisions — Part 1: Country codes. Geneva: ISO.
- Olson, D. M., Dinerstein, E., Wikramanayake, E. D., et al. (2001). Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience*, 51(11), 933–938.
- Spalding, M. D., Fox, H. E., Allen, G. R., et al. (2007). Marine Ecoregions of the World: A Bioregionalization of Coastal and Shelf Areas. *BioScience*, 57(7), 573–583.
- Spalding, M. D., Agostini, V. N., Rice, J., & Grant, S. M. (2012). Pelagic Provinces of the World: A Biogeographic Classification of the World's Surface Pelagic Waters. *Ocean & Coastal Management*, 60, 19–30.
- United Nations Statistics Division (1999, and online editions). *Standard Country or Area Codes for Statistical Use* (Series M, No. 49). New York: United Nations.
- US National System for Geospatial Intelligence (2020). *Geopolitical Entities, Names, and Codes* (GENC) Standard. Springfield, VA: NGA.
- Weynans, J. (2026). HLC Operational Framework, version 7.8 (codehx, codehu, codehl, codehy, codehw). Derivative distribution at <https://weynans.info/hlc>.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.

Appendix A Notation Reference (Analytical Summary)

The following summarises the analytical reading of the framework's controlled notation developed in Chapter 5. It is a scholarly gloss on the operators, not a substitute for the framework's own front matter, which remains the authoritative source.

Operator (role)	Analytical meaning	Theoretical function
Alternatives	Equally valid variants held side by side	Refuses privileging; preserves conflict
Redirect	Points to canonical or successor entry	Stability of identifier under change
External-definition flag	Authoritative definition lies in another file	Makes cross-file dependency explicit
Alternative-hierarchy mark	Legitimate second parenthood	Carrier of controlled polyhierarchy
Inherited / encroached / extended	Value relative to a source	Records derivation and boundary relations
Attribute tag (scheme:value)	Qualified reference to an external scheme	Turns hierarchies into a crosswalk

Appendix B Glossary

Canonicalisation — The reduction of multiple representations to a single privileged one; the strategy HLC deliberately declines.

Controlled polyhierarchy — A hierarchy permitting multiple declared parenthoods, disciplined so as to avoid contradiction.

Crosswalk — A mapping between the elements of two or more classification schemes.

Attribute locality — The rule that an entity's attributes are stored at exactly one occurrence, guaranteeing internal consistency.

Conflict preservation — The recording of disagreement between authoritative sources as first-class data rather than resolving it.

Master integrator — The partitioned numeric code space, hosted in the land file, that anchors identity across the framework.

Namespace — A declared scope within which identifiers are unique and an owning authority is responsible for assignment.

Qualified reference — A cross-reference that names the scheme under which a correspondence holds.

Appendix C Register of Integrated Schemes

The following register lists the external classification and coding schemes cross-referenced from the framework, as documented in the information sheets of the land and water files. It is presented to make concrete the breadth of the harmonisation; the tag is the label under which each scheme's values appear in the inline attribute tags.

Tag	Scheme	Domain
UN:	UN M49 statistical codes	Geopolitical
LO:	ILO country/territory index	Geopolitical
OA:	ICAO airport prefix codes	Geopolitical / aviation
CW:	Correlates of War (Gleditsch & Ward)	Political science
PD:	World Geographical Scheme for Plant Distributions	Botanical
TW:	Terrestrial Ecoregions (Olson et al. 2001)	Ecological (land)
BW:	Bioregions (RESOLVE, Dinerstein et al. 2017)	Ecological (land)
FW:	Freshwater Ecoregions (Abell et al. 2008)	Ecological (freshwater)
OW:	Global Open Oceans & Deep Seabed (Watling 2013)	Ecological (marine)
WN:	CWWN+ World Waterbody Names (NGA)	Hydrographic
FA:	HLC-F / FAO area codes	Fisheries
TI:	US Treasury International Capital classification	Economic
WC:	World Area Codes (US DOT / BTS)	Transport
ID:	World Inequality Database	Economic
R*:	ITU regions and dial codes	Telecommunication
ED:	Earth System Science Data groups	Earth science
NT:	NATO country codes	Institutional
EU:	EU Geonomenclature (EUSS)	Statistical (EU)
AU: / NZ:	Australian / New Zealand country classifications	National statistics
DE: / IT: / UK:	German / Italian / UK codes	National statistics
AR: / ZA: / JP:	Argentina / South African / Japan codes	National statistics
CH: / FR: / NO:	Swiss / French / Norwegian codes	National statistics

CA: / RU:	Canadian / Russian classifications	National statistics
DR:	Unicode CLDR exception codes	Computing / locale
WG:	WGS84 orientation centroids	Geodetic (anchors only)

The register is not exhaustive of every historical-information source the framework draws upon, but it captures the principal schemes and demonstrates the point argued throughout: that the framework's contribution lies in binding more than thirty distinct authorities into a single navigable structure.

Appendix D A Worked Crosswalk

This appendix traces a single entity through the framework to illustrate, end to end, how the mechanisms described in Chapters 5 through 7 combine in practice. The entity is Germany; the reading is drawn directly from its land-file entry.

D.1 The entry as stored

*Tree 930-961-939-839-390 · HLC-M 390 · type C · ISO DE / DEU ·
 {UN:276} {LO:024} {OA:ED} {CW:260,GFR} {PD:GER} {TI:1-100-
 2} {WC:429-2} {R1:D,49} {EU:(-11-)004} {NT:GE,DEU} {AU:2304}
 {IT:216} {AR:438} {ZA:417} {JP:213} {CH:(13)207} {FR:142}
 {NO:144} · (Deutschland).*

D.2 Reading the entry

The hierarchy path 930-961-939-839-390 locates Germany within the master code space: the leading 9 places it in the structural range, and the terminal 390 is its own master code, which the leading 3 marks as European. The type C identifies it as a country. The ISO slot carries both the alpha-2 and alpha-3 identifiers. Each subsequent tag binds the entity to one external authority: its UN statistical code is 276; its ILO index is 024; it carries Correlates-of-War identifiers including the historical GFR label; its plant-distribution, capital-flow, transport, telecommunication, EU and NATO codes follow; and its representation in seven national statistical systems is recorded, each under its own tag. The parenthetical endonym closes the entry.

D.3 What the worked example demonstrates

Three principles are visible in this single row. The **crosswalk** principle: one lookup yields the entity's identity in more than a dozen registers, which is the framework's reason for being. The **attribute-locality** principle: because every attribute lives only here, at the canonical entry, no alternative-hierarchy occurrence of Germany elsewhere in the framework can contradict it. The **stability** principle: the master code 390 is invariant under any future change to the country's name, so a time series keyed on it survives such change. The worked example is thus the whole thesis in miniature — harmonisation across many authorities, held consistent by design, and stable across time.