

DOCTORAL THESIS

Anchored Plural Harmonization

Evolving the Main Principle of the HLC Operational Framework
and Its Implementation across Land, Water and Ecoregions

*An interdisciplinary study in knowledge organisation, geography,
political science, hydrography and ecology*

Object of study: HLC Operational Framework v8.0.0
(codehx · codehu · codehl · codehy · codehw), maintained by Jens Weynans

Candidate: _____ Supervisor(s): _____

Faculty / Institution: _____ Date of submission: _____

Contents

Abstract.....	4
Abbreviations.....	5
1 Introduction.....	6
1.1 Problem statement.....	6
1.2 Research gap.....	6
1.3 Aim and research questions.....	6
1.4 Methodology.....	7
1.5 Contributions.....	7
1.6 Structure.....	7
2 Literature Review and Interdisciplinary Foundations.....	8
2.1 Knowledge organisation and classification theory.....	8
2.2 Interoperability of knowledge organisation systems.....	8
2.3 Geographic ontology and geospatial semantics.....	9
2.4 Gazetteers: contemporary and historical.....	9
2.5 Political geography, statehood and disputes.....	10
2.6 Official statistics and country code standards.....	10
2.7 Hydrography and marine management.....	11
2.8 Biogeographic and ecological regionalisation.....	11
2.9 Identity, reference and persistent identifiers.....	12
2.10 Synthesis: requirements and research gap.....	12
3 Evolving the Main Principle.....	14
3.1 Point of departure.....	14
3.2 Stage S1 — Non-replacement.....	14
3.3 Stage S2 — Anchored harmonization.....	14
3.4 Stage S3 — Explicit plurality.....	15
3.5 Stage S4 — Governed plurality.....	15
3.6 Stage S5 — Conceptual placement.....	15
3.7 Stage S6 — Temporal persistence and derivability.....	15
3.8 The evolved principle.....	15
3.9 Formal model.....	16
3.10 APH compared with alternative integration principles.....	17
4 Architecture of the HLC Framework.....	18
4.1 A five-part artefact with separated concerns.....	18
4.2 Code layers.....	19
4.3 Entity type system.....	19
4.4 Number-space allocation.....	20
4.5 Notation as a relation language.....	20
4.6 Inline attribute tags.....	21
5 Implementation of the Principle, File by File.....	22
5.1 codehx — the constitution.....	22
5.2 codehu — namespace ownership (A6) and persistence (A8a).....	22
5.3 codehl — the anchor (A2, A3), plurality (A4), locality (A5), conceptual placement (A7).....	23
5.4 codehy — fidelity (A1) and parallel hydrographic standards.....	24
5.5 codehw — sibling views (A4) and derivability (A8b).....	25
5.6 Synthesis: where each axiom lives.....	26
6 Empirical Evaluation: Structural Audit of v8.0.0.....	27
6.1 Design of the audit.....	27
6.2 Results.....	27
6.3 Analysis of deviations.....	28

6.4 Summary of the evaluation.....	29
7 Discussion.....	30
7.1 Fulfilment matrix.....	30
7.2 Interdisciplinary value.....	30
7.3 Tensions within the principle.....	31
7.4 Threats to validity.....	31
8 Further Evolution of Principle and Implementation.....	33
8.1 Refinements of APH.....	33
8.2 A normalised serialisation.....	33
8.3 Automated validation.....	34
8.4 Governance.....	34
8.5 Transferability of APH.....	34
9 Conclusion.....	35
References.....	36
Appendix A Audit Procedure.....	40
Appendix B Deviation Register (v8.0.0).....	41

Abstract

Spatial reference knowledge is fragmented across dozens of authoritative but mutually incompatible classification systems: country codes (ISO 3166, UN M49, GENC), statistical geonomenclatures, historical state registers, hydrographic limits (IHO S-23, DIAM 65-18, CWWN), fishery areas (FAO, NAFO, ICES), and more than a dozen ecological regionalisations (TEOW, FEOW, MEOW, GOODS, GET and others). Existing integration strategies either impose one canonical hierarchy, thereby overwriting disagreement, or maintain flat pairwise correspondence tables, which scale quadratically and lose hierarchical context.

This thesis develops a single integration principle from the founding maxim of the Hierarchical Location Codes (HLC) Operational Framework — *harmonization, not replacement* — and demonstrates how HLC version 8.0.0 implements it. Through six successive refinements (non-replacement, anchoring, explicit plurality, governance, conceptual placement, temporal persistence with derivability) the maxim is evolved into the **Principle of Anchored Plural Harmonization (APH)**, formalised as a set-theoretic model with eight axioms. The principle is then traced through the five-part artefact (codehx, codehu, codehl, codehy, codehw) and tested by a structural audit of 10,471 hierarchy rows, 836 namespace entries and roughly 9,700 inline attribute tags.

The audit shows that the axioms of non-replacement, anchoring, namespace ownership and temporal persistence are implemented with high fidelity (e.g. complete ownership of all 676 alpha-2 code pairs; identical DIAM+ copies across files; 99.95 % referential integrity after cross-file resolution). Explicit plurality is realised through a compact relation notation whose semantics are sound but only partly machine-readable. Derivability, introduced in v8.0 through the HLC-D column, is in transition: 63.4 % of ecological rows have converged to their derived code. The thesis concludes that APH is a coherent and transferable principle for pluralistic reference systems, and that HLC is a valid proof of concept whose main remaining gaps concern machine-actionable serialisation, validation automation and governance continuity.

Keywords: knowledge organisation; geographic classification; crosswalk; gazetteer; namespace governance; ecoregions; hydrography; political geography; persistent identifiers; pluralistic ontologies.

Abbreviations

Table 0. Abbreviations used in this thesis.

Abbreviation	Meaning
APH	Principle of Anchored Plural Harmonization (developed in this thesis)
HLC	Hierarchical Location Codes
HLC-2 / HLC-3	HLC 2-gram and 3-gram codes (ownership governed by codehu)
HLC-M	HLC main code (3–4 characters, leading digit); anchor identifier
HLC-X / HLC-S	HLC technical code / HLC alphanumeric shortcut for cross-references
HLC-D	HLC derived code (proposed HLC-M computed from HLC-S or source code)
IAT	Inline Attribute Tag of the form {scheme:value}
CWWN(+)	Codes for World Waterbody Names (NGA) (+ HLC extensions)
DIAM(+)	Defense Intelligence Agency Manual 65-18 codes (+ HLC extensions)
FAO(+)	FAO Major Fishing Areas (+ HLC re-ordered variant)
GENC	Geopolitical Entities, Names, and Codes (US government)
IHO S-23	Limits of Oceans and Seas, International Hydrographic Organization
SDMX(+)	Statistical Data and Metadata eXchange code lists (+ HLC extensions)
TEOW, BEOW, FEOW, MEOW, PPOW, BPOW	Terrestrial, Biological (RESOLVE), Freshwater, Marine, Pelagic and Benthic ecoregion/province schemes
GOODS, GET, GEZ, BYES	Global Open Oceans and Deep Seabed; IUCN Global Ecosystem Typology; FAO Global Ecological Zones; Bailey's Ecoregions
GIWA, LMEs, PAME, AMAs, EUNIS	Global International Waters Assessment; Large Marine Ecosystems; Arctic Council PAME areas; CBMP Arctic Marine Areas; European Nature Information System
SKOS	Simple Knowledge Organization System (W3C)

1 Introduction

1.1 Problem statement

Every discipline that deals with places maintains its own partition of the Earth. Statisticians partition by reporting territory, political scientists by sovereign or historical polity, hydrographers by named water bodies with legally or cartographically fixed limits, fishery scientists by management areas, and ecologists by biotic and abiotic homogeneity. Each partition is internally coherent and institutionally authoritative. None is reducible to another: the Mediterranean is a sea for the IHO, a fishing area (37) for the FAO, a group of ecoregions for MEOW, a set of provinces for PPOW, and a political neighbourhood for the Union for the Mediterranean.

Data integration across these domains therefore faces a structural dilemma. A **canonical** approach selects one hierarchy and re-expresses all others in its terms; it is simple to use but destroys the information carried by disagreement. A **pairwise** approach preserves each source but multiplies correspondence tables and loses the hierarchical context in which a code is meaningful. Neither addresses entities that change over time (state succession, renamed seas, re-numbered ecoregions) or entities whose status is contested.

1.2 Research gap

The knowledge-organisation literature has long recognised that classifications are social and plural artefacts (Bowker & Star, 1999), and web standards such as SKOS provide mapping relations between concept schemes. Gazetteer research has addressed place identity and footprints (Hill, 2006; Goodchild & Hill, 2008). What is missing is an **operational principle** that states, at the level of design rules, how a single maintainable reference system can (a) host many authoritative schemes without overriding them, (b) provide stable identifiers for longitudinal use, (c) represent disagreement as data, and (d) do so across land, water and ecological domains simultaneously. The HLC Framework is, to the author's knowledge, one of very few artefacts that attempts all four; it has not previously been analysed as the implementation of a principle.

1.3 Aim and research questions

The aim is twofold: to evolve the framework's main principle into an explicit, formal and transferable form, and to show how the HLC Framework implements it. Four research questions follow:

1. **RQ1** — What is the main principle of the HLC Framework, and how can it be evolved from a maxim into a formally specified design principle?
2. **RQ2** — Through which architectural, notational and coding mechanisms does HLC v8.0.0 implement each component of that principle?
3. **RQ3** — To what degree is the implementation consistent, as measured by reproducible structural indicators?
4. **RQ4** — What are the principle's limits, and which evolutions would close the gap between principle and implementation?

1.4 Methodology

The study combines three methods. **Conceptual analysis** reconstructs the principle from the framework's own normative text (codehx) and positions it against classification theory, geographic ontology and knowledge-organisation standards. **Design-science reasoning** (Hevner et al., 2004) treats HLC as an artefact whose design rules can be abstracted, formalised and evaluated against their intended properties. **Structural audit** operationalises each axiom into measurable indicators computed directly on the v8.0.0 workbooks (Chapter 6). All figures reported were computed programmatically on the files codehx80.docx, codehu80.xlsx, codehl80.xlsx, codehy80.xlsx and codehw80.xlsx; no external data were modified.

The analysis is interdisciplinary by necessity: the principle is abstract (information science), but its test cases are geographic (continents, climate zones, tectonic plates), political (sovereignty, disputes, succession), statistical (national country code lists), hydrographic (IHO, CWWN, FAO areas) and ecological (fourteen regionalisations). Each chapter draws on the discipline whose phenomena it discusses.

1.5 Contributions

- A six-stage derivation of the HLC maxim into the **Principle of Anchored Plural Harmonization (APH)** with eight axioms and three derived propositions (Chapter 3).
- A mapping of every APH axiom to concrete HLC mechanisms — files, code layers, notation symbols, number ranges and derivation grammars (Chapters 4–5).
- A reproducible set of structural indicators and their values for HLC v8.0.0 (Chapter 6).
- A fulfilment matrix and a roadmap for machine-actionable serialisation, validation and governance (Chapters 7–8).

1.6 Structure

Chapter 2 reviews the literature across the disciplines involved and derives the requirements and research gap. Chapter 3 evolves and formalises the principle. Chapter 4 describes the architecture of HLC as a whole; Chapter 5 analyses the implementation file by file. Chapter 6 presents the audit, Chapter 7 discusses fulfilment and tensions, Chapter 8 proposes further evolution, and Chapter 9 answers the research questions.

2 Literature Review and Interdisciplinary Foundations

This chapter reviews the literature on which the principle developed in Chapter 3 rests. It is organised by discipline rather than chronology, because HLC combines domains that are rarely discussed together. Each section ends with the implication for a pluralistic reference system; Section 2.10 condenses these implications into requirements and locates the research gap.

2.1 Knowledge organisation and classification theory

2.1.1 Classification as situated practice

Classical classification assumes mutually exclusive, jointly exhaustive classes defined by necessary and sufficient conditions. Prototype theory (Rosch, 1978) showed that human categories often have graded membership around central cases — a finding with obvious geographic counterparts in the "Middle East" or the "Arctic". Bowker and Star (1999) demonstrated empirically that classifications embody institutional choices, carry residual categories and coexist with rival schemes; their earlier concept of the **boundary object** (Star & Griesemer, 1989) — an entity robust enough to keep its identity across communities yet plastic enough to be read differently by each — describes what a country, a sea or an ecoregion is in multi-source data.

Knowledge organisation (KO) as a field has drawn the epistemological consequence. Hjørland (2008) argues that the different approaches to KO correspond to different theories of knowledge, so that the question of what a classification should be cannot be answered independently of whose knowledge it organises. Mazzocchi (2018), reviewing the typologies of knowledge organisation systems (KOS), examines the notion of *classificatory perspectivism* and asks whether pluralism in KO can take a pragmatically integrative form rather than ending in fragmentation.

2.1.2 Fragmentation and reinvention

Soergel (1999) observed that classifications — increasingly called ontologies — are developed independently in many communities with little mutual learning, which results in costly reinvention and suboptimal products. The geographic domain illustrates this vividly: statistical offices, military cartographers, fishery bodies and ecologists each maintain their own partition of the Earth, frequently with overlapping code shapes.

Implication. A reference system serving several communities cannot presuppose one partition. The identity of an entity must be separable from its placement in any single classification, and pluralism must be integrative, not merely additive.

2.2 Interoperability of knowledge organisation systems

Zeng and Chan (2004) provide the standard review of KOS interoperability methods: derivation or modelling, translation and adaptation, satellite and leaf-node linking, direct mapping, co-occurrence mapping, **switching** through an intermediate scheme, linking through a temporary union list, and linking through a thesaurus-server protocol. Two families are relevant here. *Direct mapping* produces pairwise correspondence tables; *switching* routes all mappings through a hub vocabulary. The review

notes that link storage and maintenance are recurring practical problems — a point that becomes acute when source schemes are revised.

The W3C Simple Knowledge Organization System (SKOS; W3C, 2009) standardised the representation of concept schemes: hierarchical relations (broader, narrower), polyhierarchy, multiple labels and inter-scheme mapping relations (exactMatch, closeMatch, broadMatch, narrowMatch, relatedMatch). SKOS supports plurality at the level of *representation*, but it prescribes neither a hub nor rules for namespace ownership, attribute placement or deprecation.

Ontology matching extends the problem to formal ontologies. Euzenat and Shvaiko (2013) point out that using ontologies does not by itself reduce heterogeneity but moves it to a higher level, and they catalogue correspondence types beyond equivalence — subsumption, disjointness and others. Their treatment assumes that matching is performed between independent ontologies; it does not address a curated system in which the correspondences themselves are the primary artefact.

Implication. The hub (switching) pattern is the scalable option, but a hub must be stable, governed and capable of expressing non-equivalence relations. These are design questions left open by the interoperability literature.

2.3 Geographic ontology and geospatial semantics

Smith and Mark (2001) distinguish *bona fide* boundaries (physical discontinuities such as coastlines) from *fiat* boundaries (created by human decision, such as national borders or fishery area limits). Most boundaries relevant to HLC are fiat, and many are fiat boundaries imposed on fuzzy bona fide phenomena — currents, biomes, climate zones.

Fonseca, Egenhofer, Agouris and Câmara (2002) proposed an *ontology-driven GIS* in which ontologies act as a system integrator, and in which multiple ontologies at different levels of generality allow information to be extracted at different stages of classification. Kuhn (2005) located the need for geospatial semantics in the componentisation of GIS into interoperating services and proposed *semantic reference systems* as an analogue of spatial reference systems, stressing the grounding of semantics in physical processes and measurements.

Both lines of work argue for explicit, multi-level semantic structures alongside geometric data. Neither treats the case in which the semantic structure *replaces* geometry as the primary integrator, which is the situation of a reference system without polygons.

Implication. A framework may legitimately separate conceptual placement from geometric containment, provided it declares the semantics of its placements and of any coordinates it carries.

2.4 Gazetteers: contemporary and historical

Gazetteer research models a place as a triple of name(s), footprint(s) and feature type (Hill, 2006); Goodchild and Hill (2008) framed digital gazetteers as an independent research agenda. Janowicz and Keßler (2008) argue that the typing schemes of gazetteers benefit from an ontological formalisation,

and they document the difficulties of converting feature-type thesauri into ontologies — evidence that the *type* dimension of a place record is as problematic as its footprint.

Historical gazetteers raise further demands. Southall, Mostern and Berman (2011) argue that historians need places whose location is uncertain, alternative names and their evolution, and the contexts in which names were used; they propose gazetteers in which dates serve mainly to reference particular instances of names rather than to fix precise creation and deletion dates. Mostern and Johnson (2008) go further and model historical gazetteers as records of *events* in the histories of places, linked by chronological, mereological and causal relations.

Implication. Names, types and time must be first-class. A reference system should keep historical entities within the same identity space as current ones, record successor relations explicitly, and tolerate imprecise temporal and spatial extents.

2.5 Political geography, statehood and disputes

Statehood is not a stable attribute. Crawford (2006) treats the creation, continuity and extinction of states as matters governed by international law, discussing contested cases such as Taiwan and Palestine and the effects of the dissolution of Yugoslavia and the USSR. Fazal (2007) observes that about half of the states on an 1816 world map no longer exist, and her study includes a revision of the Correlates of War list of state-system members — showing that even the canonical political-science state register is itself subject to scholarly revision.

Boundaries are equally unsettled. Prescott and Triggs (2008) survey an interdisciplinary literature spanning economics, geography, history, law and political science on the evolution of frontiers into boundaries and on boundary disputes. At sea, Prescott and Schofield (2005) document the maritime political boundaries of the world, and Schofield (2013) describes the South China Sea as one of the most geographically and geopolitically complex ocean spaces, with multiple competing territorial and jurisdictional claims.

Implication. A neutral reference system must represent succession, dissolution, renaming and contested sovereignty without adjudicating them — which requires a claim relation distinct from hierarchical parenthood.

2.6 Official statistics and country code standards

ISO 3166 has three parts: country codes (alpha-2, alpha-3, numeric), subdivision codes, and codes for formerly used country names (ISO, n.d.). ISO does not itself define country names, which are taken from United Nations sources (ISO, n.d.). The numeric codes originate in the UN M49 standard of the United Nations Statistics Division and were included in ISO 3166 from its 1981 edition; because merged or split countries receive new numeric codes, numeric identifiers are useful for datasets with historical country information (UNSD, M49).

National statistical offices and international bodies nonetheless publish country lists with deliberate deviations from ISO: separate codes for territories ISO does not distinguish, different aggregates, or

retained historical codes. SDMX (ISO 17369) provides the exchange format for statistical code lists but not a reconciliation of their contents.

Implication. Statistical interoperability requires holding the deviating national and international lists in parallel, attached to a common identity, rather than normalising them to ISO.

2.7 Hydrography and marine management

The IHO publication S-23 *Limits of Oceans and Seas* exists as the 3rd edition (1953) and as unratified later drafts, which differ in naming and delimitation; the ACIC chart of limits is a further military-cartographic variant. The NGA-derived CWWN codes and the DIAM 65-18 codes provide alphanumeric identifiers for water bodies. Fishery management uses another partition: the FAO Major Fishing Areas, refined regionally by NAFO, ICES and GFCM. Regional conventions (OSPAR, HELCOM) and national hydrologic units (USGS HUC, Water Survey of Canada, CCM2) complete a landscape in which the same water body carries several concurrent identifiers, each authoritative in its own mandate. Maritime boundary scholarship (Prescott & Schofield, 2005) adds the legal layer of jurisdictional zones that cut across all of these.

Implication. Plurality in hydrography includes plurality of *editions* of one authority. A reference system must be able to hold versions side by side without selecting one.

2.8 Biogeographic and ecological regionalisation

2.8.1 A proliferation of schemes

Biogeographic regionalisation has multiplied since Udvardy's provinces (1975). Terrestrial systems include TEOW (Olson et al., 2001), its RESOLVE revision (Dinerstein et al., 2017), Bailey's ecoregions (Bailey, 2014) and FAO's Global Ecological Zones (FAO, 2001); freshwater has FEOW (Abell et al., 2008); marine has MEOW (Spalding et al., 2007), PPOW (Spalding et al., 2012), the deep-sea biogeography of GOODS (Watling et al., 2013), Longhurst's biogeochemical provinces (Longhurst, 2007) and Large Marine Ecosystems (Sherman & Alexander, 1989). The IUCN Global Ecosystem Typology (Keith et al., 2022) adds a function-based rather than geographic hierarchy. Regionalisations are used operationally, for instance to identify biomes and ecoregions in which habitat conversion greatly exceeds protection (Hoekstra et al., 2005).

2.8.2 Expert versus data-driven delimitation

Kreft and Jetz (2010) note that regionalisations were traditionally established by experts on qualitative evidence, and that the lack of transparency limited their utility; their quantitative framework produced regions that both resemble and differ from the classic divisions. Holt et al. (2013), combining distributions and phylogenies of 21,037 vertebrate species, identified 20 zoogeographic regions grouped into 11 realms and found no support for several previously defined regions. Vilhena and Antonelli (2015) state that there is no consensus on how best to identify and delimit biogeographical regions, and propose a network method designed to handle transition zones. For the oceans, Costello et al. (2017) distinguished 30 marine realms — 18 on continental shelves and 12 offshore and deep-sea — from the distributions of about 65,000 species.

2.8.3 Regionalisation as a hierarchical but contested system

Morrone (2018) defines a biogeographical regionalisation as a hierarchical system categorising areas by their biotas and proposes a seven-step protocol ending with the definition of boundaries and transition zones and with area nomenclature. He also recalls critics who doubted that any single regionalisation could hold for different taxa. The ecological literature therefore supplies both the case for hierarchy and the case for plurality: each scheme is hierarchical, but the schemes are siblings, partially derived from one another, and none is final.

Implication. Ecological schemes must be held as parallel views with their own hierarchies; documented derivations between schemes should be recorded, and identifiers for new schemes should be generated reproducibly.

2.9 Identity, reference and persistent identifiers

Kent (1978) argued that data models must separate real-world entities from their representations. Hayes and Halpin (2008) distinguish *access* from *reference*: identifiers for things that cannot be retrieved refer by description, and reference by description is inherently ambiguous — additional ontological distinctions often increase rather than reduce ambiguity. Presutti and Gangemi (2008) address the identity of resources and entities on the Web from an ontological perspective.

On the practical side, the FAIR principles (Wilkinson et al., 2016) require globally unique, persistent, resolvable identifiers with rich metadata. McMurry et al. (2017) derive ten lessons on the design, provision and reuse of identifiers for large-scale data integration, emphasising persistence and resolvability as the most common failure points.

Implication. Anchors must be persistent and never reused; deprecated identifiers must resolve; and the system should accept that reference to geographic entities is description-based, so that plurality of descriptions is a feature to be managed rather than eliminated.

2.10 Synthesis: requirements and research gap

Table 2.1 maps the literature strands to the requirements they impose; Table 2.2 states what each strand leaves open.

Table 2.1. Requirements derived from the literature.

#	Requirement	Principal sources
R1	Preserve every source scheme without redefining its codes or boundaries	ISO 3166; national statistics; IHO S-23 editions; Soergel (1999)
R2	Provide a stable identity independent of any single source	Kent (1978); Wilkinson et al. (2016); McMurry et al. (2017)
R3	Represent polyhierarchy, alternatives, succession and dispute as data	Bowker & Star (1999); Crawford (2006); Fazal (2007); Southall et al. (2011)
R4	Prevent code collisions between schemes that reuse code shapes	Zeng & Chan (2004); SDMX/DIAM practice
R5	Declare the spatial semantics (conceptual vs. geometric)	Smith & Mark (2001); Kuhn (2005); Hill (2006)

#	Requirement	Principal sources
R6	Keep deprecated entities resolvable across time	ISO 3166-3; McMurry et al. (2017); Mostern & Johnson (2008)
R7	Generate identifiers reproducibly for new schemes	Kreft & Jetz (2010); Morrone (2018); proliferation of regionalisations

Table 2.2. Literature strands, contributions and open questions.

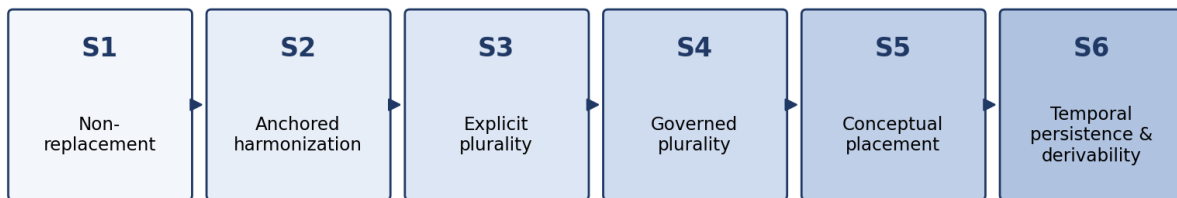
Strand	Key contribution	What remains open
KO theory	Classification is situated and plural; perspectivism may be integrative	Operational rules for integrating perspectives in one system
KOS interoperability	Switching (hub) and mapping methods; SKOS relations	Governance of the hub, namespace ownership, deprecation
Ontology matching	Correspondence types beyond equivalence	Curated correspondences as the primary artefact
Geospatial semantics	Ontologies as integrators; semantic reference systems	Integration without geometry
Gazetteers	Name-footprint-type model; historical and event-based gazetteers	Multi-domain scope (land, water, ecology) with shared identity
Political geography	Statehood, succession, disputes, boundary evolution	Neutral machine-readable representation of claims
Official statistics	ISO/M49 coding; national deviations	Parallel retention of deviating lists
Hydrography	Multiple authoritative partitions and editions	Version-parallel storage and cross-linking to land and ecology
Biogeography	Many hierarchical, partly derived, contested regionalisations	Sibling storage with reproducible identifiers
Identifiers	Persistence, resolvability; ambiguity of reference	Persistence rules for curated, plural hierarchies

The gap is therefore not the absence of any single component — polyhierarchy, hub mapping, persistent identifiers and neutral dispute records all exist in the literature — but the absence of an **operational principle** that combines them into design rules for one maintainable reference system across land, water and ecological domains. Chapter 3 develops such a principle from the HLC maxim; Chapters 4–6 test it against the HLC artefact.

3 Evolving the Main Principle

3.1 Point of departure

The HLC Common Front Matter states that the framework's purpose is "harmonization, not replacement". It integrates recognised standards, adds hierarchical HLC extensions "where existing standards are insufficient", preserves "multiple coding schemes, versions, and hierarchical interpretations in parallel", treats classificatory disagreement "as information, not error", and maintains "stable internal identifiers for longitudinal consistency" (codehx, §§1, 5, 6). These statements are normative but informal. They do not specify what must hold for a given row, file or release to comply. This chapter evolves the maxim in six stages; each stage answers a weakness of the previous one and adds one constraint. Figure 3.1 summarises the derivation.



Principle of Anchored Plural Harmonization (APH)

each stage adds one constraint; the conjunction of all six defines APH

Figure 3.1. Six-stage evolution of the HLC maxim into the Principle of Anchored Plural Harmonization.

3.2 Stage S1 — Non-replacement

In its minimal form the maxim is a negative rule: *do not replace source standards*. It responds to requirement R1. Taken alone, S1 is compatible with a mere library of unconnected code lists; it guarantees fidelity but not integration. It also leaves open how HLC may modify a source when the source is incomplete or internally inconsistent. The refinement adopted by HLC is **marked modification**: an HLC deviation from a source is admissible only if the original value remains visible (original codes in brackets, HLC additions marked by "+"). Non-replacement thus becomes *non-destructive* rather than *non-interventionist*.

3.3 Stage S2 — Anchored harmonization

Integration requires a pivot. S2 introduces a single internal identifier per entity to which all source codes attach (R2). The pivot converts a network of pairwise correspondences into a hub-and-spoke structure. For n schemes, full pairwise coverage requires

$$M_{\text{pair}}(n) = n(n - 1) / 2 \quad \text{versus} \quad M_{\text{hub}}(n) = n \quad (3.1)$$

mapping sets. With the roughly 40 inline tag keys used in codehl, this is 780 versus 40. The pivot must be stable across releases, since otherwise every downstream join breaks at each version. S2 thus

implies an **identity-before-classification** stance: the anchor denotes the entity, not its position in a hierarchy.

3.4 Stage S3 — Explicit plurality

An anchor alone still invites a single canonical placement. S3 requires that disagreement between sources — about names, codes, parents, status or extent — be stored as data (R3). Four kinds of plurality are distinguished: *nominal* (alternative names or codes for the same entity), *hierarchical* (the same entity under several legitimate parents), *temporal* (predecessor and successor entities) and *political* (competing claims). Each needs a distinct representation, because merging them would conflate, for example, a renaming with a partition.

3.5 Stage S4 — Governed plurality

Unconstrained plurality produces two failure modes: *code collision*, when different schemes use the same code shape (DIAM+ and SDMX+ both use `nx` two-character codes; ISO alpha-2 and HLC pseudo-regions share `aa`), and *attribute drift*, when duplicated entries accumulate contradictory properties. S4 adds two governance rules (R4): **namespace ownership** — every code string in a given code layer has exactly one owning scheme — and **attribute locality** — descriptive attributes live only at the original entry, while duplicates carry placement only.

3.6 Stage S5 — Conceptual placement

Spatial reference systems are routinely misread as geometric authorities. S5 declares the spatial semantics (R5): placement expresses **best conceptual fit**, not containment. Coordinates are **orientation anchors** — centroid-like reference points — and are not footprints, boundaries or inputs to spatial analysis. This makes it legitimate to place a strait under two oceans or a climate zone across continents without implying overlap computations, and it protects the framework from being used for delimitation.

3.7 Stage S6 — Temporal persistence and derivability

Stages S1–S5 describe a single release. S6 extends the principle through time and growth (R6, R7). **Persistence**: identifiers are never deleted or reused; deprecated entities remain with a redirect to their successor. **Derivability**: for newly integrated schemes, anchors are generated by a deterministic rule from the source code, so that integration is reproducible and does not depend on ad-hoc allocation. Where an anchor already exists elsewhere in the system, it takes precedence over the derived value.

3.8 The evolved principle

Principle of Anchored Plural Harmonization (APH). A pluralistic reference system shall integrate authoritative classification schemes by (i) anchoring each referenced entity to one stable internal identifier, (ii) recording the codes and placements of every source beside that anchor without destructive modification, (iii) expressing every disagreement as a typed relation, (iv) under governed namespaces and with attributes held at a single locus, (v)

placing entities by conceptual fit rather than geometric containment, and (vi) preserving and deterministically deriving identifiers over time.

APH differs from the original maxim in being *checkable*: each clause can be translated into a condition on data. The next section provides this translation.

3.9 Formal model

3.9.1 Primitives

- E — the set of entities (referents: countries, seas, ecoregions, organisations, epochs, plates, zones).
- $S = \text{Sext} \cup \{\text{HLC}\}$ — the set of schemes; Sext are external sources (ISO-2, M49, CWWN, TEOW, ...).
- $\kappa_s : E \rightarrow C_s$ — the partial code function of scheme s into its code set C_s .
- $\alpha : E \rightarrow A$ — the anchor function into the set of HLC main codes A (HLC-M).
- V — a set of hierarchical views; each placement is a pair (e, τ) where τ is a root-to-node path in view $v \in V$.
- $\rho : E_{\text{dep}} \rightarrow E$ — the redirect function from deprecated to successor entities.
- $\omega : E \rightarrow \mathbb{R}^2$ — orientation coordinates (WGS84 reference point).
- $\text{own} : K \rightarrow S$ — the ownership function on code strings K of a given code layer.

3.9.2 Axioms

Table 3.1. Axioms of APH.

Axiom	Formal statement	Stage
A1 Fidelity	For all $s \in S_{\text{ext}}$ and e : the recorded value of $\kappa_s(e)$ equals the source value, or the source value is retained as annotation next to a marked HLC value.	S1
A2 Anchoring	α is total on E , and for consecutive releases $r, r+1$: $\alpha_r(e) = \alpha_{r+1}(e)$ for every e present in both.	S2
A3 Hub mapping	Every correspondence between schemes s and t is obtained by composition $\kappa_t \circ \alpha^{-1} \circ \alpha \circ \kappa_s^{-1}$ via the anchor; no direct s - t table is primary.	S2
A4 Typed plurality	Every disagreement is expressed by exactly one relation type from $R = \{\text{alt, redirect, logical, duplicate, inherit, range, claim}\}$; no assertion is deleted to resolve a conflict.	S3
A5 Attribute locality	For each e there is exactly one primary placement $p_o(e)$; all other placements of e carry no descriptive attributes.	S4
A6 Namespace ownership	For each code layer L , own_L is a function: every string $k \in K_L$ has exactly one owning scheme.	S4
A7 Conceptual placement	Parent assignment is a curatorial relation; $\omega(e)$ is a reference point and no containment relation is inferred from ω .	S5
A8 Persistence & derivability	(a) No anchor is removed or reassigned; deprecated e has $\rho(e)$ defined and ρ is acyclic. (b) For entities of a newly integrated scheme without prior anchor, $\alpha(e) = \delta(\sigma(e))$ for a deterministic grammar δ over the shortcut $\sigma(e)$.	S6

3.9.3 Derived propositions

Proposition 1 (non-contradiction of duplicates). Under A5, no two placements of the same entity can assert contradictory attributes. *Proof sketch:* attributes are defined only at $p_o(e)$; any other placement contributes the empty attribute set, and the union of a set with the empty set cannot contain a contradiction not already present in the set itself. ■

Proposition 2 (linear integration cost). Under A2 and A3, adding a scheme s_{n+1} requires one mapping set $\kappa_{(n+1)}$ against A , and yields correspondences to all n existing schemes by composition. The marginal cost is constant in n , in contrast to n for pairwise designs. ■

Proposition 3 (longitudinal resolvability). Under A8(a), any anchor ever issued resolves, in a finite number of redirect steps, to either a current entity or a retained historical entity. *Proof sketch:* anchors are never removed; p is acyclic on a finite set, so iteration terminates. ■

The propositions state what APH *buys*: consistency despite duplication, scalability despite plurality, and longevity despite change.

3.10 APH compared with alternative integration principles

Table 3.2. Integration principles against requirements R1–R7 (+ satisfied, o partial, – not addressed).

Principle	R1	R2	R3	R4	R5	R6	R7	Typical weakness
Canonical hierarchy (one master scheme)	–	+	–	+	o	o	o	erases disagreement
Pairwise crosswalk tables	+	–	o	–	o	–	–	quadratic growth; no context
Federated linked data (independent URIs + mappings)	+	o	+	+	o	o	–	no common anchor; mapping sparsity
Formal ontology alignment (OWL)	+	o	o	+	+	o	o	high modelling cost; brittle with fiat categories
APH (this thesis)	+	+	+	+	+	+	+	requires curation and governance

APH does not claim novelty in each component; polyhierarchy, persistent identifiers and mapping relations are established. Its contribution is the **conjunction** of the eight axioms into one principle, and — as the following chapters show — an existing artefact that implements the conjunction across three domains.

4 Architecture of the HLC Framework

4.1 A five-part artefact with separated concerns

HLC v8.0.0 consists of one normative document and four data workbooks (Figure 4.1). The separation is itself an implementation decision: each APH clause is assigned a primary locus, so that governance (who owns a code), anchoring (where an entity lives) and domain plurality (which sources describe it) are not entangled in one table.

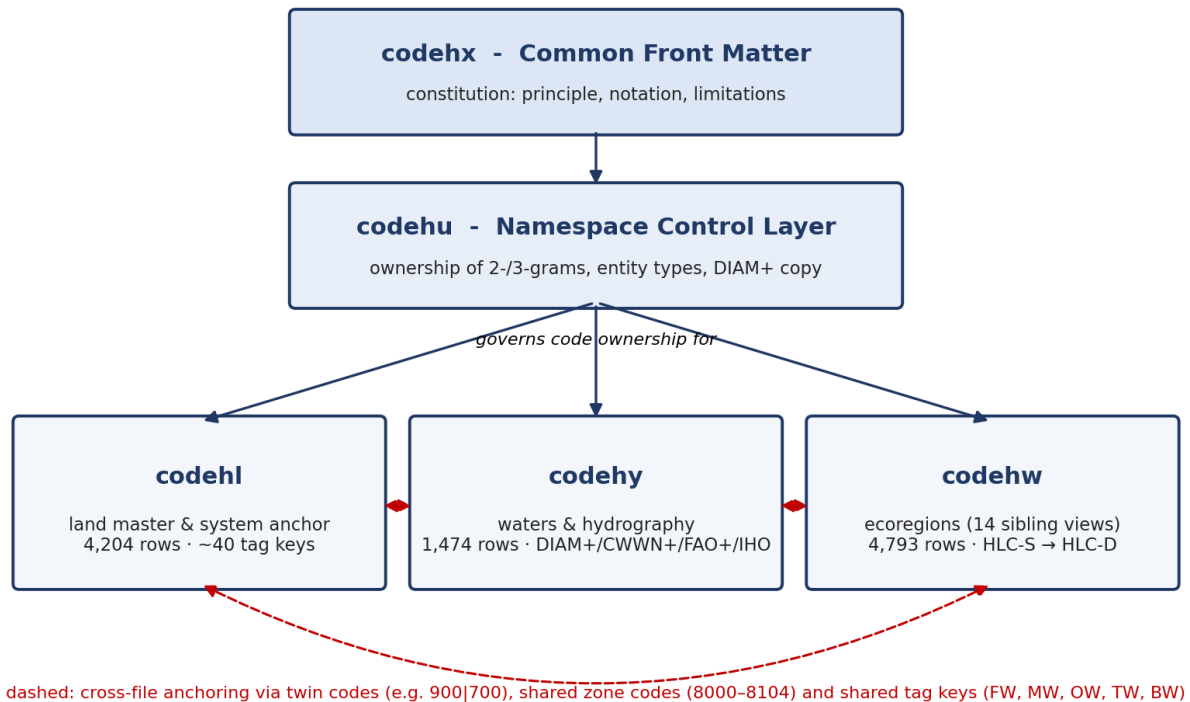


Figure 4.1. Architecture of HLC v8.0.0 and cross-file anchoring. Row counts are data rows with a non-empty tree path.

Table 4.1. File roles and their principal APH clauses.

File	Primary role	Sheets (v8.0.0)	Primary APH clauses
codehx	Common front matter: purpose, notation, conflict handling, spatial semantics, limitations, source list, change history	— (document)	Constitution of all clauses; A7 declared
codehu	Namespace control: ownership of 2-grams and 3-grams across ISO, HLC, SDMX+, DIAM+, GENC, CWWN+; entity type system	INFO, TYPES, DATA (836 rows), COPY (DIAM+, 247 rows)	A6, A8(a), A4 (redirect, logical)
codehl	Land-based master hierarchy and system anchor; number-space allocation; ~40 crosswalk tag keys	INFO, HLC-M, DATA (4,204 rows), CHART	A2, A3, A4, A5, A7
codehy	Water and hydrographic hierarchy; DIAM+, CWWN+, FAO+, IHO/ACIC in parallel	INFO, DATA (1,474), CWWN+, DIAM+, FAO+, IHO+	A1, A4, A5
codehw	Ecoregions and ecosystems as 14 sibling views; derivation grammar	INFO, HLC-M, 14 scheme sheets (4,793), CHART	A8(b), A4 (sibling views)

The historical development confirms that the architecture was evolved rather than designed up front (Figure 4.2): the land file preceded the others by roughly a year, the namespace layer and the water file appeared in 2023, the ecological file in late 2024, and the normative text was only separated into its own document in v7.8 (08/2026). The separation of the constitution from the data is thus the most recent architectural step — consistent with a maturing principle being made explicit after the fact.

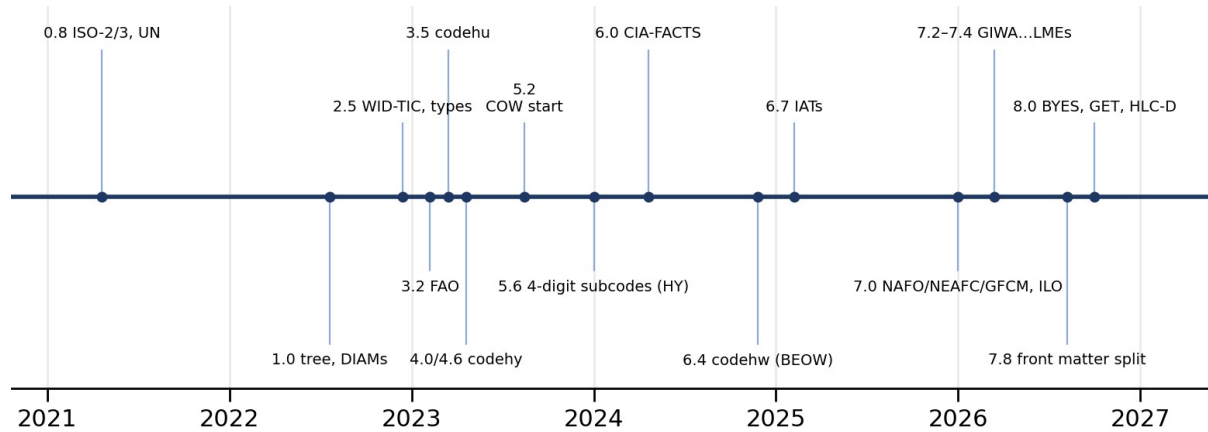


Figure 4.2. Selected milestones from the codehx change history (version and main integration).

4.2 Code layers

HLC separates identifiers by function into layers. Each layer has its own shape grammar (a = alpha, n = numeric, x/y/z = alphanumeric) and its own ownership rule.

Table 4.2. HLC code layers.

Layer	Shape	Function	Ownership
HLC-2	aa, an, axxx, ax:	Short key: ISO alpha-2, HLC pseudo-areas/regions (aa), pseudo-subregions (an), dependency numbering (axxx), attribute and organisation codes (ax:)	codehu for all files, except DIAM+ nx codes
HLC-3	aaa, axx, Unx, Xnx, Xaa	ISO alpha-3; HLC additions; SDMX+-derived organisations (U); DIAM+-derived water regions (X+digit); GENC entities (X+alpha)	codehu
HLC-M	n + 2-3 x, ranges (!-), inheritance (~ + -)	Main anchor code; carries the position in the number space	codehl HLC-M sheet (allocation), each file (assignment)
## type	nn / a	Entity type and level (see 4.3)	codehu TYPES
HLC-X HLC-S	3-5 characters	Technical code and alphanumeric shortcut for cross-reference; carries M49 (nnn), inherited M49 (~ nnn), GENC (: 9nn), CWWN+ (# naa)	file-specific
HLC-D	as HLC-M	Derived proposal for HLC-M computed from HLC-S / source code (v8.0)	codehw derivation grammar
IAT	{key:value}	Inline crosswalk to external schemes	key registry per file

4.3 Entity type system

The TYPES sheet (codehu) defines a two-dimensional typology: a *level* axis (0 global, 1 region, 2 subregion, intermediate multiples, 3 country, 4 division, 5 province/unit) and a *qualifier* axis (single, multiple, geographical, economic, historical, botanical, administrative). Parallel columns define

terrestrial ("HLC-Earth") and non-terrestrial ("Others") readings — e.g. level 3 is *Country* on land and *Organisation* otherwise; level 1 is *Region* or *Eco-region*. Additional non-hierarchical types cover tectonic (T), functional (F), logical (L), NUL (N), waters (W), international (I) and unspecified (X). The type system is what allows a single tree to hold climate zones, polities, organisations and plates without category mistakes: the type makes explicit *what kind of claim* a placement makes.

4.4 Number-space allocation

The HLC-M sheet in codehl allocates the leading digits of the anchor code among the three data files and, within codehl, by continent and structural level (Figure 4.3). Examples: 1nn–5nn for historical entities, divisions and small islands by continent; 600–610 and 700–797 for waters (codehy); 56–59, 689–699 and 798–810 for ecological branches and trees (codehw); 87n world organisations; 91n–96n main structures by continent; 977–978 tectonic plates; 99n historical world epochs. A dedicated range permits external numeric codes (e.g. GeoNames) with a character substitution in the penultimate position to keep value areas disjoint.

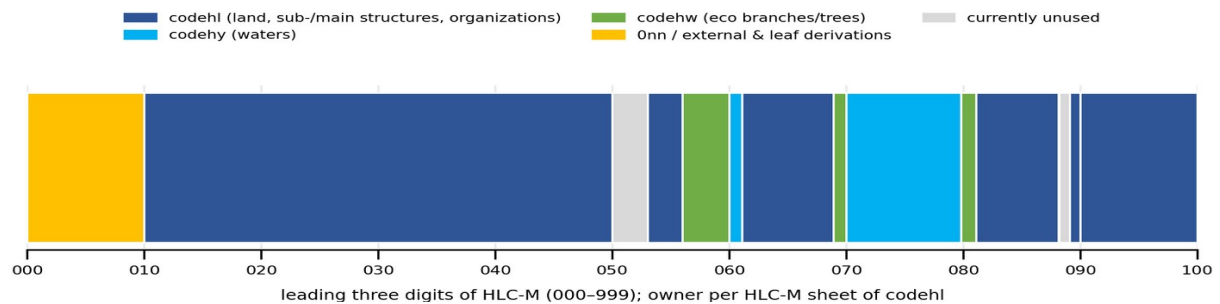


Figure 4.3. Allocation of the HLC-M number space by owning file (simplified from the HLC-M sheet of codehl).

The allocation is the mechanism that makes a single anchor namespace possible across independently edited files: each file can mint anchors in its own ranges without coordination, and a code's leading digits reveal its owning file.

4.5 Notation as a relation language

The shared notation of codehx §3 is, in effect, a compact relation language embedded in cell values. Table 4.3 reads each symbol as a typed relation and relates it to the axioms.

Table 4.3. HLC notation interpreted as typed relations.

Symbol	Reading as relation	Example (v8.0.0)	Axiom
	alt(e, c_1, c_2): equally valid codes or names	AF A1 (ISO alpha-2 and HLC-X); 800 689 (twin anchors)	A4, A3
>	redirect(e_{old}, e_{new})	DD>DE German Democratic Republic; AB>AQ British Antarctic Territory	A8(a)
:	logical/organisational entry	AA:30 ASEAN; DE: Destatis country system	A4, A6
*	duplicate placement in a legitimate alternative hierarchy	722* South China Sea under Pacific and under Asia	A4, A5
~ + -	inherit / encroach number ranges from neighbours	153~ Brunei; 241+- Ethiopia	A6 (range sharing)
!-	range extension	390!-99 Germany; 189!-94 China	A6

Symbol	Reading as relation	Example (v8.0.0)	Axiom
()	optional extension; original source value	OK(5A) XK: DIAM+ code with original in brackets	A1
{k:v}	mapping to external scheme k	{UN:608} {CW:840,PHI} {PD:PHI}	A3
"claimed by X, Y"	claim(e, {X, Y}) without sovereign parent	Abu Musa and Tunb Islands (claimed by IR, AE)	A4
(#)	intentionally empty attribute (placement only)	secondary placements of climate zones under GEZ	A5

4.6 Inline attribute tags

IATs store external codes as {key:value} inside designated columns (UNO, TERRA, AQUA, M04–M10, OTHERS in codehl; *FAO & related* and *OTHERS* in codehy). The design trades normalisation for editability: a single row carries the full crosswalk for an entity, while the column grouping (M04–M10) clusters sources thematically (e.g. national statistics offices). A row may carry several tags per cell; the Philippines row, for example, holds {UN:608} {IL:084} {OA:RP}, {CW:840,PHI} {PD:PHI}, {TC:4-480-6} {WC:766}, {R3:PHL,63}, {NT:RP,PHL} {EU:708}, {AU:5204}, {DE:462} {IT:323}, {AR:312} {ZA:336} {JP:117} and {CH:(47)534} {FR:220} {NO:428} — a hub-and-spoke crosswalk (A3) to over twenty schemes in one record.

5 Implementation of the Principle, File by File

5.1 codehx — the constitution

codehx formulates each APH clause in natural language and assigns it framework-wide scope. Its structure maps onto APH as follows: §1 (purpose) states S1–S2; §3 (notation, attribute-locality rule, IATs) states S3–S4; §4 (coordinates) states S5; §5 (conflict handling) specifies the relation types of A4 including neutral recording of disputes; §6 (limitations) states persistence ("deprecated entities are retained with redirects rather than deleted") and acknowledges the curatorial character of hierarchy choices. The change history functions as the release log that A2 presupposes. The separation of this text in v7.8 is a governance act: one normative source now binds four data files, removing the risk of diverging per-file definitions.

5.2 codehu — namespace ownership (A6) and persistence (A8a)

5.2.1 Ownership of the 2-gram and 3-gram space

The DATA (HU) sheet contains 836 entries keyed by HLC-2. All 676 alpha-2 letter pairs (AA–ZZ) are either assigned or explicitly reserved; no HLC-2 key occurs twice. Ownership classes are encoded by shape: ISO alpha-2 for recognised countries (251 type-C rows), HLC pseudo-codes for regions (85), subregions (74), waters (80), multiples (41) and special entities (45 type X), and organisation keys of the form aa:nx (133 rows containing ":"), which bind an SDMX+-derived organisation code (Unx) to a two-letter host key — e.g. AA:3O → ASEAN, CB:4F → ECB, CW:3W → Correlates of War. Logical entries such as DE:, CH:, AU:, CA:(E) and CA:(W) register national classification systems themselves as entities, with (E)/(W) distinguishing terrestrial from aquatic registers.

Collision control is explicit where two schemes share a shape: DIAM+ water codes (nx) are owned by DIAM+ and replicated in the COPY sheet rather than re-minted. A programmatic comparison shows that the DIAM+ copy in codehu and the DIAM+ sheet in codehy agree on the HLC-M of all 245 shared DIAM+ codes (after normalising bracket annotations) — replication without drift.

5.2.2 Reservation and redirects

Sixteen HLC-2 keys are reserved or free, including one prospective reservation (BU — Bougainville, annotated as a possible 2027 case). Nine entries carry redirects: DD>DE (German Democratic Republic), CS>(HH) (Czechoslovakia), AN>(HH) (Netherlands Antilles → AW, BQ, CW), VD>(VN) (Republic of Viet-Nam), CE>(KI), TA>(SH), AB>AQ, EL>(GR) and R9>LM. The notation distinguishes redirection to a live code from redirection to a historical bucket (HH), thereby encoding the difference between succession and dissolution. This is A8(a) at the namespace level: codes are never freed for reuse.

5.2.3 The type registry

Moving the TYPES sheet from codehl to codehu (v7.9) relocated the type vocabulary to the governance layer. The consequence is that types, like codes, have a single owner and are shared by reference.

5.3 codehl — the anchor (A2, A3), plurality (A4), locality (A5), conceptual placement (A7)

5.3.1 Scope and structure

codehl holds 4,204 placements in a single tree whose paths reach up to twelve levels. The type distribution (Figure 5.1) shows that the file is far more than a country list: divisions (885), subdivisions (750), historical entities (663) and units (528) outnumber countries (252) by a wide margin, and the tree also hosts eco-regions (94), waters (77), international organisations (73), tectonic plates (36) and logical registers (23).

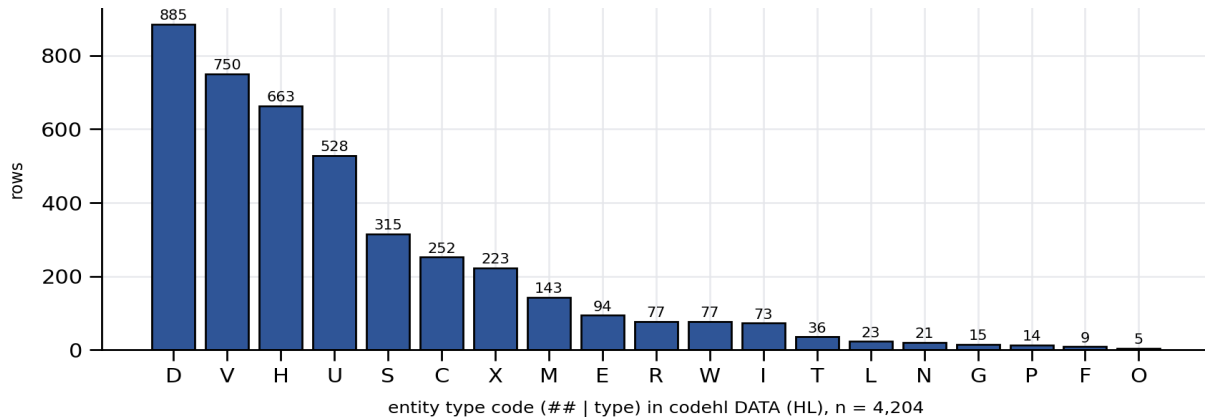


Figure 5.1. Entity types in codehl (DATA (HL)).

Top-level branches reflect the number-space allocation: 800 geographical/climate zones, 810 Global Ecological Zones, 898/899 biogeographic realms and marine ecosystems, 900–909 oceans and international waters, 910–96x continents, 98x organisations and 99x historical world epochs. The land file therefore anchors not only land entities but the *roots* of every other domain, which is what the codehx designation "master integrator" means in operational terms.

5.3.2 Anchoring and hub mapping

Every row has an HLC-M; 4,157 distinct anchor values occur. Crosswalks are attached to the anchor by IATs: codehl contains 9,089 tags drawn from about 40 keys. For the 252 country rows, coverage is near-complete for the core schemes — UN M49 252, Eurostat 251, SACC 248, US TIC 247, Correlates of War 240, WGS-RPD 237, NATO 236 — with an average of 18.7 tags per country (maximum 23). National statistical systems of Germany, Switzerland, France, Italy, Norway, Argentina, South Africa and Japan are held in parallel columns. This is the hub of Proposition 2: any two of these schemes can be joined through the anchor without a dedicated correspondence table.

5.3.3 Twin anchors across files

Where a domain root belongs to another file, codehl uses **twin anchors** in |-notation: 900|700 (High Seas), 901|701 (Arctic Ocean), 906|776* (Indian Ocean), 800|689 (Geographical Zones), 810|690 (GEZ), 898|698, 899|699. The first value sorts the entry within the land tree; the second is the owning file's anchor. Twin anchors realise A3 across file boundaries: codehl can place the Arctic Ocean conceptually without minting a competing identity. Empirically, 209 anchor values are shared between

codehl and codehy, 536 between codehl and codehw, 310 between codehy and codehw, and 63 occur in all three.

5.3.4 Hierarchical, temporal and political plurality

Hierarchical plurality is expressed by *-duplicates: the South China Sea (722*) appears both under the Pacific (West-Central) and under South-Eastern Asia; the Atlantic (Central) under the Atlantic and under Africa; the Indian Ocean (West) under the Indian Ocean and under Eastern Africa. *Cross-classification* is expressed by re-placing climate zones (8000–8104) under GEZ ecological zones (8105–8109) with type X and all attribute columns set to (#): the zone is present in the GEZ view but carries no attributes there — a textbook application of A5.

Temporal plurality is carried by 663 type-H rows spanning from ancient polities (Elamites 3200–539 BC; Early Dynastic Period 2900–2350 BC; Sabaean Kingdom) through colonial units (Aden Protectorate 1872–1963; French protectorate in Morocco 1912–56) to recent entities (North and South Yemen; neutral zones between Kuwait, Iraq and Saudi Arabia). History is not a separate file but a type within the same tree, so that historical and current entities share anchors, hierarchy and crosswalks.

Political plurality is recorded in 44 rows annotated "claimed by ...". Their placement pattern is instructive: 24 are typed X (unspecified), and typical cases are placed in neutral regional containers rather than under a claimant — Aksai Chin (claimed by TW, CN, IN), Siachen Glacier (IN, PK), Paracel Islands (CN, TW, VN), Abu Musa and Tunb Islands (IR, AE) — frequently with their GENC code (e.g. {US:XAA,935}). Where de-facto administration is placed under a state, the claim remains in the record (e.g. Kuril Islands, claimed by JP; Liancourt Rocks). The mechanism implements A4's *claim* relation and codehx §5 without the framework adjudicating sovereignty.

5.3.5 Conceptual placement

Coordinates appear as {WG:...} tags in 208 rows, often as latitude bands rather than points (e.g. Arctic Climate Zone: North Pole ↔ Arctic Circle {WG:(90N)(663339N/60N)}). Placements are made by fit: climate zones by latitude definition, realms by source (Udvardy; Heilprin/Newton), tectonic plates (36 rows, major/middle/minor) as their own branch — none of which could be derived from national boundaries. A7 is thus not merely a disclaimer; it is what enables the co-existence of categorically different partitions within one tree.

5.4 codehy — fidelity (A1) and parallel hydrographic standards

codehy holds 1,474 placements from global zones and biogeographic realms through oceans, seas, straits, currents, gyres, polynyas, water masses, inland waters and vertical depth zones. Its main data sheet carries, side by side for each water body: DIAM+, CWWN+ (with original codes), FAO+, WGS84 reference, FAO-related tags (NAFO, ICES, GFCM), other tags (USGS HUC, Water Survey of Canada, CCM2, HELCOM, OSPAR, Longhurst, GIWA, GOODS/MEOW, FEOW, ISTAT) and four IHO/ACIC columns.

5.4.1 Non-destructive modification

Fidelity (A1) is visible in the code cells. DIAM+ values retain the original code in brackets: 0K(5A)|XK (Arctic), 0L(9A)|XL (Atlantic), 0O(6A)|XO (Indian Ocean). CWWN+ uses every predefined CWWN

code and confines HLC additions to declared extension ranges (0AA–0ZZ, 1AR–1AZ, 1RA–1ZZ, ..., 9AA–9YZ), so that an HLC code can never be mistaken for an NGA code. FAO+ is explicitly labelled as an *HLC-style* re-ordering with original FAO numbers retained in the related column. The version history documents the restoration of mostly original CWWN codes in v7.1 — a correction towards fidelity.

5.4.2 Four editions of one standard

The IHO columns hold S-23 in the 2002 draft, the 1986 4th-edition draft, the 1953 3rd edition, and the ACIC chart side by side. The Lincoln Sea, for example, is 9.13 (2002), 9.13 (4th), 17 A (3rd) and A13 (ACIC). The framework does not choose an edition; the dedicated IHO+ sheet serves as a selection and sorting aid. This is plurality of *versions* of a single authority — a case that neither canonical nor pairwise designs handle naturally.

5.4.3 Hierarchical plurality in water

80 rows in codehy carry *-duplicates, reflecting that marginal seas, straits and archipelagic waters legitimately belong to several parent basins depending on the standard consulted. Oceanographic features that no national standard covers — currents (North Equatorial Current, NECC, SEC), gyres (Beaufort Gyre), drifts (Transpolar Drift), water masses (Arctic Bottom Water) and polynyas (North Water) — receive HLC anchors in the CWWN+ extension ranges. The vertical dimension (epipelagic to hadopelagic, 8990–8999) is linked to GOODS via {OW:OD9x} tags. Of all rows, 1,018 (69 %) carry an orientation reference.

5.5 codehw — sibling views (A4) and derivability (A8b)

5.5.1 Sibling, not subset

codehw holds fourteen regionalisations on separate sheets with 4,793 rows in total: GIWA (136), GET (145), BYES (58), GEZ (41), TEOW (1,149), BEOW (1,171), FEOW (555), EUNIS (679), BPOW (86), GOODS (102), MEOW (366), LMEs (89), PAME incl. AMAs (116) and PPOW (100). The file declares that none of the sheets is a master or a subset of another. Documented inter-scheme dependencies are integrated rather than hidden: WWF Global 200 groupings as sub-levels in TEOW and additional items in FEOW and MEOW; Longhurst provinces in PPOW; AMAs in PAME; the identity of PPOW with the pelagic part of GOODS is noted with highlighting. Sort-only levels (-00-, -0000-) are declared non-hierarchical.

5.5.2 The derivation grammar

The HLC-M (HW) sheet specifies, per scheme, a rewrite grammar δ from the scheme's shortcut (HLC-S) or reference code to an HLC-M code. Examples:

Table 5.1. Selected rules of the codehw derivation grammar δ .

Scheme	Source pattern (HLC-S / REF)	Derived HLC-M (HLC-D)
GIWA	'GI'<H J K S V X Y><xy:01–66,89–94>	0'G'<xy>
GET	'E'<4 5 6 7 9><yz:00–99>	0<I J K L N><yz>
GEZ	'GZ'<xy:00–99>	0'Z'<xy>
TEOW	'PA'<71 74 ... 83><xy:00–99>	3<P Q ... Y><xy>
FEOW	'F'<x:1–7><yz:00–99>	<1–8><B ... H><yz>

Scheme	Source pattern (HLC-S / REF)	Derived HLC-M (HLC-D)
GOODS	'VT'<xy:00-09>	9'Q'<xy>
LMEs	'LM'<xy:01-99>	9'L'<xy>
PPOW	'PP'<xy:01-99>	9'P'<xy>

The grammar is injective within each scheme by construction (distinct source codes map to distinct derived codes within the scheme's target range), and the target ranges of different schemes are disjoint by their leading characters (0G, 0Z, 9Q, 9L, 9P, ...). Derivation therefore satisfies A8(b) and, together with the number-space allocation, A6.

5.5.3 Precedence of existing anchors

v8.0 introduced the HLC-D column in all codehw sheets. The rule is: HLC-M equals HLC-D *unless* an anchor for the same entity already exists in codehl or codehy. The GIWA sheet illustrates the precedence: the Kara Sea has HLC-D 01A1 but HLC-M 7071, the anchor minted in codehy; the Yellow Sea has HLC-D 0A34 but HLC-M 7124; the Amur River Basin has HLC-D 0B30 but HLC-M 6002. Precedence prevents a second identity for an entity already anchored — the operational meaning of *anchored* in APH. The codehw info sheet also records the maintainer's intention to replace locally derived values progressively with global ones; derivation is thus positioned as a reproducible fallback, not as a competing identity scheme.

5.6 Synthesis: where each axiom lives

Table 5.2. Localisation of APH axioms in the artefact.

Axiom	Principal mechanism(s) in HLC v8.0.0	Locus
A1 Fidelity	Original codes in brackets; declared extension ranges; "+"-marked additions; editions in parallel	codehy, codehu COPY
A2 Anchoring	HLC-M on every row; stable-identifier policy	codehl (primary), all
A3 Hub mapping	IATs; twin anchors (x y); shared anchor values across files	codehl, codehy
A4 Typed plurality	> : * ~ + - !- notation; "claimed by"; sibling sheets; type registry	all; codehx \$3, \$5
A5 Attribute locality	Attributes only at original entry; (#) at secondary placements	codehl, codehy
A6 Namespace ownership	2-gram/3-gram ownership table; HLC-M number-space allocation; disjoint derivation ranges	codehu, codehl HLC-M, codehw HLC-M
A7 Conceptual placement	Coordinates as reference points/bands; no geometry; fit-based parents	codehx \$4; all
A8 Persistence & derivability	Redirects (>, (HH)); historical type H; reservation; HLC-D with anchor precedence	codehu, codehl, codehw

6 Empirical Evaluation: Structural Audit of v8.0.0

6.1 Design of the audit

Each axiom that makes a claim about data was operationalised as one or more indicators computable from the workbooks without external reference data. The audit therefore tests *internal* conformance to APH, not the correctness of individual codes against their source standards (see 7.4). Computation used cell values as stored; tree paths were normalised by removing duplicate markers (*), trailing hyphens, bracketed level comments and the declared sort-only levels (-00-, -0000-). A parent was counted as *resolvable* if its code exists as a node in the same sheet or, for intended cross-file anchoring, in any file of the framework.

6.2 Results

Table 6.1. Audit indicators and results.

#	Indicator	Axiom	Result (v8.0.0)
I1	Coverage of the alpha-2 space (AA–ZZ) by assigned or reserved HLC-2 keys	A6	676 / 676 (100 %)
I2	Uniqueness of HLC-2 keys in codehu	A6	836 keys, 0 duplicates
I3	Uniqueness of HLC-3 codes in codehu	A6	1 duplicate (MES)
I4	Country rows in codehu present in codehl	A2	251 / 251; 28 differ only in key notation (e.g. AS US60 vs AS (US-60))
I5	HLC-3 agreement codehu ↔ codehl (directly keyed countries)	A2	223 / 223 (100 %)
I6	HLC-M agreement codehu ↔ codehl (directly keyed countries)	A2	222 / 223 (99.6 %); deviation: BV 677~ vs 677
I7	DIAM+ replication codehu COPY ↔ codehy DIAM+ (HLC-M)	A1, A6	245 / 245 shared codes identical
I8	Referential integrity of tree paths (all files)	A2	10,471 rows; 51 local orphans, 46 resolve cross-file; 5 unresolvable → 99.95 %
I9	Path–anchor congruence in codehl (last path segment = HLC-M, allowing padding and range markers)	A2	4,179 / 4,204 (99.4 %); 25 off-by-one rows
I10	Attribute locality for repeated anchors in codehl	A5	47 groups; 34 compliant; 13 with attributes at > 1 placement
I11	Cross-file anchor sharing (distinct HLC-M values)	A3	HLnHY 209; HLnHW 536; HYnHW 310; all three 63
I12	Crosswalk density: tags in codehl / codehy	A3	9,089 / 645; countries 18.7 tags on average
I13	Coverage of core schemes for 252 countries	A3	UN 252; EU 251; SACC 248; TIC 247; COW 240; WGS-RPD 237; NATO 236
I14	Orientation references	A7	codehl 208 rows; codehy 1,018 / 1,474 (69 %); no geometry present
I15	Persistence markers	A8a	9 redirects and 16 reservations in codehu; 663 historical rows in codehl
I16	Recorded claims	A4	44 rows "claimed by ..."; 24 of them typed X (no sovereign type)
I17	Derivation convergence in codehw (HLC-M = HLC-D)	A8b	3,038 / 4,793 (63.4 %)

6.3 Analysis of deviations

6.3.1 Referential integrity and path–anchor congruence (I8, I9)

Of 51 locally orphaned paths, 46 are intended cross-file references — e.g. GEZ placements under Bailey domains (7996–7998) or GET groupings under zones owned by codehl — and resolve once the framework is treated as one namespace. The five unresolvable rows all sit under one node (...-976-7881-..., Antarctic sub-structures), i.e. a single missing parent. The 25 off-by-one rows cluster in runs (Russian Far East 1320–1323; Madhya Pradesh historical units 1273–1276; Mecklenburg 3988/3989 vs 3938/3939), which indicates renumbering that was propagated to one column but not to the other rather than independent errors.

6.3.2 Attribute locality (I10)

The 13 non-compliant groups fall into three classes, each touching a different axiom:

- **Same entity, attributes duplicated** (4 groups: South China Sea 722, Atlantic Central 751, Indian Ocean West 777, Antarctic Ocean 907). These are genuine A5 deviations: the secondary copy repeats tags. They are harmless while copies are identical but reintroduce the drift risk that A5 excludes.
- **View-specific reuse of an anchor** (4 groups: 6974 large-river habitats in two realms; 7320, 7760, 7820 hydrothermal vs. hadal provinces of the same ocean). The anchor denotes the *ocean* or *habitat type*, but the rows denote distinct provinces. This is an identity question (A2): the entity of the placement is not the entity of the anchor.
- **Distinct entities, colliding anchors** (5 groups: 8037, 6978, 1666, 1324, 1272). Each coincides with an off-by-one path from 6.3.1; the collision is a consequence of incomplete renumbering, not of design.

6.3.3 Derivation convergence (I17)

Figure 6.1 shows convergence per scheme. It is high where the ecological sheet is the first and only locus of the entities (GET 95 %, EUNIS 97 %, BYES 83 %) and low where entities were anchored earlier in the water file (MEOW < 1 %, LMEs 10 %, GIWA 18 %). Low values are therefore not errors: they measure the working of anchor *precedence* (5.5.3). A convergence indicator that excludes rows whose HLC-M is a codehl/codehy anchor would be the appropriate quality metric; on the present data it cannot be computed exactly without an explicit precedence flag.

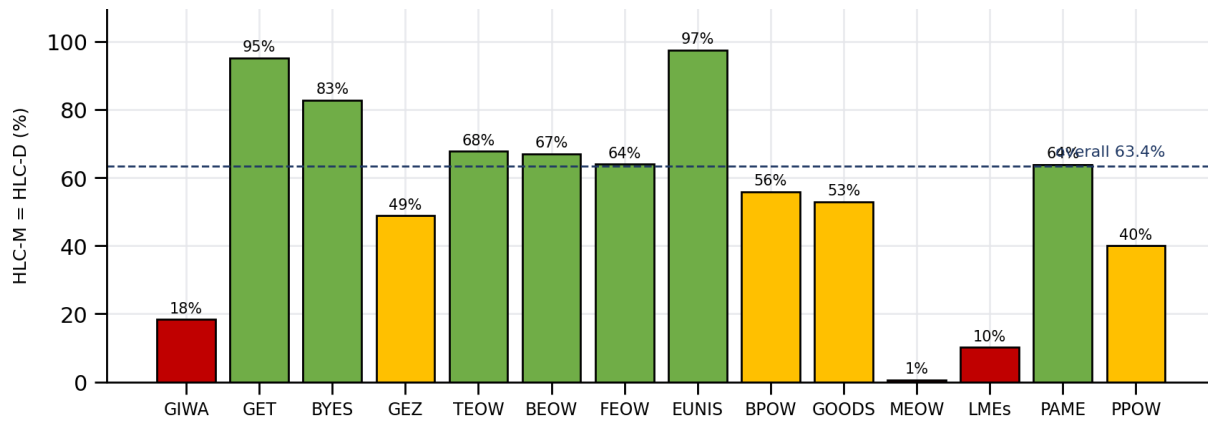


Figure 6.1. Share of rows with HLC-M = HLC-D per codehw scheme.

6.3.4 Notational and documentary deviations

Minor deviations concern consistency of the relation language rather than data: dependency keys written as AS|US60 in codehu and AS (US-60) in codehl; one organisation key (CM:0C) with HLC-3 CCM|0C instead of the U-prefixed pattern; sheet references in the codehy INFO sheet ("CWWN+ (HC)", "FAO+ (HD)") that differ from the sheet names ("CWWN+ (-C)", "FAO+ (-F)"); an incomplete line in the codehu HLC-2 definition; the spelling "HCL" for "HLC" in several info texts; type codes 34/35 listed out of numeric order; and per-file version labels of 7.9 within an 8.0.0 release.

6.4 Summary of the evaluation

The artefact conforms closely to APH on the structural level: namespace ownership, replication, referential integrity and inter-file country consistency are all at or above 99.5 %. Deviations concentrate in two places — attribute duplication for multi-parent water bodies, and renumbering residues — and the most informative finding is conceptual rather than numerical: anchors are occasionally used for *types of places* (ocean, habitat) where the placement denotes an *instance* (a province). This distinction is not yet expressed in the relation language.

7 Discussion

7.1 Fulfilment matrix

Table 7.1. APH fulfilment by HLC v8.0.0.

Axiom	Implementation	Evidence	Status
A1 Fidelity	Brackets, extension ranges, parallel editions	I7; 5.4.1–5.4.2	fulfilled
A2 Anchoring	HLC-M on every row; stable-identifier policy	I4–I6, I8, I9	largely fulfilled (renumbering residues)
A3 Hub mapping	IATs, twin anchors, shared anchors	I11–I13	fulfilled
A4 Typed plurality	Relation notation; claims; sibling sheets	I16; 5.3.4; 5.5.1	fulfilled semantically; partly machine-readable
A5 Attribute locality	(#) at secondary placements	I10	largely fulfilled (4 groups duplicated)
A6 Namespace ownership	Ownership table; number space; disjoint derivation	I1–I3	fulfilled
A7 Conceptual placement	Reference points/bands only; declared in codehx	I14; 5.3.5	fulfilled
A8a Persistence	Redirects, (HH), reservations, type H	I15	fulfilled; formal release-to-release check not yet possible
A8b Derivability	δ grammar, HLC-D, anchor precedence	I17; 5.5.2–5.5.3	implemented, in transition

7.2 Interdisciplinary value

Official statistics. HLC provides what national country code systems lack individually: a single join key across Destatis, BFS, INSEE, ISTAT, SSB, INDEC, Stats SA, ABS/SACC, Eurostat, US TIC and others, with deviations from ISO retained instead of corrected. For longitudinal statistics the historical type and redirect mechanism allow series to be re-aggregated across state succession.

Political science and history. The combination of Correlates of War codes, historical entities back to the Bronze Age and neutral claim records makes HLC a reference layer for datasets that must be comparable across regime changes without adopting a sovereignty position. The treatment of disputed areas — recorded, typed, not subordinated — is a methodological position in itself and one compatible with the neutrality norms of international organisations.

Hydrography and marine management. Parallel IHO editions, DIAM+, CWWN+ and FAO/NAFO/ICES/GFCM areas in one record address a practical problem of marine data integration: the same observation is reported against legally, cartographically and managerially different partitions. HLC anchors oceanographic features (currents, gyres, water masses) that no official gazetteer covers.

Ecology and biogeography. Holding fourteen regionalisations as siblings, with their documented derivations (Global 200, Longhurst, PPOW/GOODS), respects the epistemic status of each as a model with its own purpose. The derivation grammar offers a reproducible way to cite ecoregions across schemes with one code family.

Information science. HLC is an existence proof that polyhierarchy, persistent identity and typed mapping can be maintained in a widely available medium by a single curator at a scale of about 11,000 rows — and it shows where that medium reaches its limits.

7.3 Tensions within the principle

7.3.1 Plurality versus usability

Every relation type added to satisfy A4 increases the cognitive load on users. The HLC notation compresses nine relation types into symbols inside cell values; this is efficient for the curator but opaque to newcomers and to software. The tension is inherent in APH: the more disagreement is kept, the more interpretation is required downstream. It can be mitigated by serialising relations explicitly (Chapter 8) rather than by reducing plurality.

7.3.2 Stability versus correctability

A2 demands stable anchors; data quality demands correction. The off-by-one runs (6.3.1) show the cost of renumbering in a system without separate placement identifiers: when a code is corrected in the tree path but not in the anchor column, identity and position diverge. APH implies that corrections should change *placements*, never *anchors*; the implementation would benefit from a rule that makes this explicit.

7.3.3 Identity of instances versus types

The reuse of ocean or habitat anchors for distinct provinces (6.3.2) reveals an under-specified point of the principle itself. APH as formulated in Chapter 3 assumes that each anchor denotes one entity. In ecological and marine schemes, however, many codes denote *intersections* (province = ocean × depth zone; ecoregion = realm × biome). A refined A2 should either mint anchors for intersections or introduce a composite anchor (e.g. 7320@8996) that is distinct from both components.

7.3.4 Curatorial judgment versus reproducibility

codehx openly states that the legitimacy of alternative hierarchies is a curatorial decision. This is consistent with Bowker and Star's view that classification is always situated, but it means that parts of HLC cannot be reproduced from its sources alone. A8(b) addresses reproducibility for identifiers; a comparable provenance record for *placements* (source or rationale per *-duplicate) would extend it to hierarchy.

7.3.5 Medium versus model

The spreadsheet medium enables rapid curation and human review, and the planned TSV/SSV exports keep the barrier to use low. It does not enforce keys, types or relation integrity. Most deviations found in Chapter 6 are of a kind that a relational or graph store would reject at insertion time.

7.4 Threats to validity

- **Internal audit only.** Indicators test conformance to APH, not the correctness of codes against ISO, UN, IHO, NGA or ecological sources.

- **Normalisation choices.** Path normalisation (padding, range markers, sort levels) affects I8–I10; alternative rules would shift counts slightly but not the pattern of findings.
- **Heuristic key matching.** Inter-file country matching used HLC-2 keys; 28 entries required notation-tolerant matching.
- **Single release.** Persistence (A8a) can be fully tested only by comparing consecutive releases, which were outside the material analysed.
- **Researcher interpretation.** The reconstruction of APH is an interpretation of the framework's normative text; the maintainer's intentions may differ in emphasis.

8 Further Evolution of Principle and Implementation

8.1 Refinements of APH

The discussion motivates three refinements of the axioms. They tighten the principle without altering its character.

A2' (identity/placement separation). Each placement (e, τ, v) has its own placement identifier; corrections change placements, never anchors. The anchor denotes the entity; the tree path denotes one position of it in one view.

A2'' (composite entities). An entity defined as the intersection of entities of different views (e.g. ocean $\times$ depth zone, realm $\times$ biome) receives its own anchor or a composite anchor $e_1@e_2$ distinct from both components.

A4' (provenance of plurality). Every alternative placement (*), claim and redirect carries a source reference or a curatorial rationale.

8.2 A normalised serialisation

The cell-embedded relation language can be serialised losslessly into five tables, which could be generated from the workbooks by a parser and published alongside the existing TSV/SSV files:

Table 8.1. Proposed normalised serialisation of HLC.

Table	Key	Content	Source in HLC
entity	anchor	HLC-M, type, preferred label, status (current / historical / reserved)	HLC-M, ## type, description
placement	placement_id	anchor, view, parent placement, primary flag, sort order	tree path, * marker
code	(anchor, scheme, value)	scheme key, value, role (preferred / alternative / original), source version	HLC-2, HLC-3, HLC-X S, IATs, bracketed originals
relation	(type, from, to)	redirect, claim, inherit/encroach, range, composite	>, (HH), "claimed by", ~ + -, !-
namespace	(layer, pattern)	owning scheme, reserved flag	codehu DATA, HLC-M sheets

A linked-data rendering follows directly (Table 8.2). It would make the framework FAIR-compliant in the sense of Wilkinson et al. (2016) — resolvable identifiers, machine-readable relations — without changing its content.

Table 8.2. Mapping HLC constructs to SKOS and related vocabularies.

HLC construct	SKOS / Dublin Core rendering
HLC-M anchor	skos:Concept with skos:notation
tree path in view v	skos:broader within skos:ConceptScheme v; polyhierarchy via additional skos:broader for * placements
alternative code or name	skos:notation (typed per scheme) / skos:altLabel
{k:v} IAT	skos:exactMatch or skos:closeMatch to the concept of scheme k
> redirect	dct:isReplacedBy; owl:deprecated true

HLC construct	SKOS / Dublin Core rendering
"claimed by X, Y"	custom property hlc:claimedBy (no skos:broader to X or Y)
type H, date ranges	dct:temporal
WGS84 reference	geo:lat / geo:long, explicitly typed as reference point

8.3 Automated validation

The indicators of Chapter 6 can be turned into release gates. A minimal rule set:

1. **V1** Every HLC-2 and HLC-3 value has exactly one owner in codehu (I1–I3).
2. **V2** Every tree parent resolves within the framework namespace (I8).
3. **V3** The last path segment equals the HLC-M modulo declared padding and range markers (I9).
4. **V4** For each anchor, exactly one placement carries attributes (I10).
5. **V5** Replicated tables (DIAM+ COPY, types) are identical across files (I7).
6. **V6** Every anchor of release *r* exists in release *r*+1, either current or with a redirect (A8a).
7. **V7** Redirect chains are acyclic and terminate (Proposition 3).
8. **V8** Every derived HLC-D lies in the target range of its scheme and is unique within it (A8b).
9. **V9** Every IAT key is registered in the file's key legend.
10. **V10** Key notation for dependencies is uniform across files.

Rules V1–V5 and V8–V10 are executable on a single release; V6–V7 require the preceding release. All deviations reported in 6.3 would have been flagged by V2–V4 and V10.

8.4 Governance

The single-maintainer model, stated openly in codehx, is both a strength (consistency of judgment, absence of external influence) and the principal continuity risk. APH itself is maintainer-independent; what is maintainer-dependent is curatorial judgment (7.3.4). Three measures would decouple them: an explicit licence and reuse statement covering both HLC content and the integrated source codes; citable, versioned releases with persistent identifiers for each release; and a public change-request channel through which placement rationales (A4') can be proposed and recorded.

8.5 Transferability of APH

Nothing in APH is specific to geography. The same conjunction of anchoring, fidelity, typed plurality, governed namespaces and persistence applies wherever authoritative classifications disagree and must be joined longitudinally — occupational and industrial classifications, species taxonomies with competing checklists, or disease classifications across revisions. Clause (v), conceptual placement, generalises to *declared semantics of membership*: a reference system must state whether membership is definitional, empirical or curatorial.

9 Conclusion

RQ1. The main principle of HLC is *harmonization, not replacement*. Evolved through six stages, it becomes the Principle of Anchored Plural Harmonization: authoritative schemes are integrated by anchoring each entity to one stable identifier, recording every source unchanged beside it, expressing every disagreement as a typed relation under governed namespaces with attributes at a single locus, placing entities by conceptual fit, and preserving and deterministically deriving identifiers over time. Formalised in eight axioms, APH yields three properties: non-contradiction of duplicates, linear integration cost and longitudinal resolvability.

RQ2. HLC v8.0.0 implements APH through a separation of concerns across five files: a constitution (codehx), a namespace layer governing all 2- and 3-grams (codehu), a land-based anchor tree that also roots every other domain (codehl), a water file holding DIAM+, CWWN+, FAO+ and four IHO/ACIC editions in parallel (codehy), and an ecological file of fourteen sibling schemes with a derivation grammar (codehw). The mechanisms are a layered code system, an allocated number space, twin anchors, inline attribute tags and a compact relation notation.

RQ3. The structural audit shows high conformance: complete ownership of the alpha-2 space, identical replicated tables, 99.95 % referential integrity, 99.4 % path-anchor congruence and 99.6 % inter-file anchor agreement for countries. Deviations concentrate in attribute duplication for four multi-parent water bodies and in residues of renumbering; derivation convergence (63.4 %) reflects the intended precedence of existing anchors.

RQ4. The limits lie in machine-actionability of the relation language, in the identity of composite entities, in provenance for curatorial placements, and in governance continuity. Refinements A2', A2'' and A4', a normalised serialisation with a SKOS mapping, ten validation rules and three governance measures would close these gaps without compromising the principle.

HLC thus stands as a coherent proof of concept for a general principle: that a reference system can be *integrative without being imperial* — joining land, water and life, past and present, agreement and dispute, under one set of identifiers while leaving every source as it is.

References

Primary sources

Weynans, J. (2026). *HLC Operational Framework — Common Front Matter*, v8.0.0 (codehx80.docx), 2026-10-01.

Weynans, J. (2026). *HLC namespace control layer* (codehu80.xlsx); *Hierarchical Location Codes — land master* (codehl80.xlsx); *waters* (codehy80.xlsx); *ecoregions* (codehw80.xlsx), v8.0.0.

Secondary literature and standards

Abell, R., et al. (2008). Freshwater ecoregions of the world: A new map of biogeographic units for freshwater biodiversity conservation. *BioScience*, 58(5), 403–414.

Bailey, R. G. (2014). *Ecoregions: The Ecosystem Geography of the Oceans and Continents* (2nd ed.). Springer.

Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out: Classification and Its Consequences*. MIT Press.

Brummitt, R. K. (2001). *World Geographical Scheme for Recording Plant Distributions* (2nd ed.). Hunt Institute for Botanical Documentation / TDWG.

Costello, M. J., Tsai, P., Wong, P. S., Cheung, A. K. L., Basher, Z., & Chaudhary, C. (2017). Marine biogeographic realms and species endemicity. *Nature Communications*, 8, 1057. <https://doi.org/10.1038/s41467-017-01121-2>

Crawford, J. R. (2006). *The Creation of States in International Law* (2nd ed.). Oxford University Press.

Dinerstein, E., et al. (2017). An ecoregion-based approach to protecting half the terrestrial realm. *BioScience*, 67(6), 534–545.

Euzenat, J., & Shvaiko, P. (2013). *Ontology Matching* (2nd ed.). Springer. <https://doi.org/10.1007/978-3-642-38721-0>

FAO (2001). *Global Ecological Zoning for the Global Forest Resources Assessment 2000*. FRA Working Paper 56. Rome: FAO.

Fazal, T. M. (2007). *State Death: The Politics and Geography of Conquest, Occupation, and Annexation*. Princeton University Press.

Fonseca, F. T., Egenhofer, M. J., Agouris, P., & Câmara, G. (2002). Using ontologies for integrated geographic information systems. *Transactions in GIS*, 6(3), 231–257. <https://doi.org/10.1111/1467-9671.00109>

Gleditsch, K. S., & Ward, M. D. (1999). A revised list of independent states since the Congress of Vienna. *International Interactions*, 25(4), 393–413.

Goodchild, M. F., & Hill, L. L. (2008). Introduction to digital gazetteer research. *International Journal of Geographical Information Science*, 22(10), 1039–1044.

Hayes, P. J., & Halpin, H. (2008). In defense of ambiguity. *International Journal on Semantic Web and Information Systems*, 4(2), 1–18. <https://doi.org/10.4018/jswis.2008040101>

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.

Hill, L. L. (2006). *Georeferencing: The Geographic Associations of Information*. MIT Press.

- Hjørland, B. (2008). What is knowledge organization (KO)? *Knowledge Organization*, 35(2/3), 86–101.
- Hoekstra, J. M., Boucher, T. M., Ricketts, T. H., & Roberts, C. (2005). Confronting a biome crisis: Global disparities of habitat loss and protection. *Ecology Letters*, 8(1), 23–29. <https://doi.org/10.1111/j.1461-0248.2004.00686.x>
- Holt, B. G., et al. (2013). An update of Wallace's zoogeographic regions of the world. *Science*, 339(6115), 74–78. <https://doi.org/10.1126/science.1228282>
- International Hydrographic Organization (1953). *Limits of Oceans and Seas* (Special Publication S-23, 3rd ed.). Monaco: IHO.
- ISO (n.d.). *ISO 3166 — Country Codes*. International Organization for Standardization. <https://www.iso.org/iso-3166-country-codes.html>
- ISO 3166-1:2020. *Codes for the representation of names of countries and their subdivisions — Part 1: Country code* (4th ed.). Geneva: ISO.
- ISO 17369. *Statistical data and metadata exchange (SDMX)*. Geneva: International Organization for Standardization.
- Janowicz, K., & Keßler, C. (2008). The role of ontology in improving gazetteer interaction. *International Journal of Geographical Information Science*, 22(10), 1129–1157. <https://doi.org/10.1080/13658810701851461>
- Keith, D. A., et al. (2022). A function-based typology for Earth's ecosystems. *Nature*, 610, 513–518.
- Kent, W. (1978). *Data and Reality*. North-Holland.
- Kreft, H., & Jetz, W. (2010). A framework for delineating biogeographical regions based on species distributions. *Journal of Biogeography*, 37(11), 2029–2053. <https://doi.org/10.1111/j.1365-2699.2010.02375.x>
- Kuhn, W. (2005). Geospatial semantics: Why, of what, and how? In S. Spaccapietra & E. Zimányi (Eds.), *Journal on Data Semantics III* (LNCS 3534, pp. 1–24). Springer. https://doi.org/10.1007/11496168_1
- Longhurst, A. R. (2007). *Ecological Geography of the Sea* (2nd ed.). Academic Press.
- Mazzocchi, F. (2018). Knowledge organization system (KOS): An introductory critical account. *Knowledge Organization*, 45(1), 54–78. <https://doi.org/10.5771/0943-7444-2018-1-54>
- McMurry, J. A., et al. (2017). Identifiers for the 21st century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data. *PLoS Biology*, 15(6), e2001414. <https://doi.org/10.1371/journal.pbio.2001414>
- Morrone, J. J. (2018). The spectre of biogeographical regionalization. *Journal of Biogeography*, 45(2), 282–288. <https://doi.org/10.1111/jbi.13135>
- Mostern, R., & Johnson, I. (2008). From named place to naming event: Creating gazetteers for history. *International Journal of Geographical Information Science*, 22(10), 1091–1108. <https://doi.org/10.1080/13658810701851438>
- Olson, D. M., et al. (2001). Terrestrial ecoregions of the world: A new map of life on Earth. *BioScience*, 51(11), 933–938.

- Olson, D. M., & Dinerstein, E. (2002). The Global 200: Priority ecoregions for global conservation. *Annals of the Missouri Botanical Garden*, 89(2), 199–224.
- Prescott, V., & Schofield, C. (2005). *The Maritime Political Boundaries of the World* (2nd ed.). Martinus Nijhoff.
- Prescott, V., & Triggs, G. D. (2008). *International Frontiers and Boundaries: Law, Politics and Geography*. Martinus Nijhoff.
- Presutti, V., & Gangemi, A. (2008). Identity of resources and entities on the Web. *International Journal on Semantic Web and Information Systems*, 4(2), 49–72.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27–48). Erlbaum.
- Schofield, C. (2013). What's at stake in the South China Sea? Geographical and geopolitical considerations. In R. Beckman, I. Townsend-Gault, C. Schofield, T. Davenport & L. Bernard (Eds.), *Beyond Territorial Disputes in the South China Sea* (pp. 11–46). Edward Elgar.
- Sherman, K., & Alexander, L. M. (Eds.) (1989). *Biomass Yields and Geography of Large Marine Ecosystems*. Westview Press.
- Smith, B., & Mark, D. M. (2001). Geographical categories: An ontological investigation. *International Journal of Geographical Information Science*, 15(7), 591–612.
- Soergel, D. (1999). The rise of ontologies or the reinvention of classification. *Journal of the American Society for Information Science*, 50(12), 1119–1120.
- Southall, H., Mostern, R., & Berman, M. L. (2011). On historical gazetteers. *International Journal of Humanities and Arts Computing*, 5(2), 127–145. <https://doi.org/10.3366/ijhac.2011.0028>
- Spalding, M. D., 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.
- Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, "translations" and boundary objects. *Social Studies of Science*, 19(3), 387–420.
- Udvardy, M. D. F. (1975). *A Classification of the Biogeographical Provinces of the World* (IUCN Occasional Paper 18). Morges: IUCN.
- United Nations Statistics Division (UNSD). *Standard Country or Area Codes for Statistical Use (M49)*. New York: United Nations.
- Vilhena, D. A., & Antonelli, A. (2015). A network approach for identifying and delimiting biogeographical regions. *Nature Communications*, 6, 6848. <https://doi.org/10.1038/ncomms7848>
- W3C (2009). *SKOS Simple Knowledge Organization System Reference*. W3C Recommendation.
- Watling, L., Guinotte, J., Clark, M. R., & Smith, C. R. (2013). A proposed biogeography of the deep ocean floor. *Progress in Oceanography*, 111, 91–112.
- Wilkinson, M. D., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018.

Zeng, M. L., & Chan, L. M. (2004). Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology*, 55(5), 377–395. <https://doi.org/10.1002/asi.10387>

Further sources integrated in HLC (e.g. GENC, DIAM 65-18, CWWN, NAFO, ICES, GFCM, OSPAR, HELCOM, USGS HUC, CCM2, GIWA, GET, BPOW, EUNIS, PAME/AMAs, national statistical classifications) are cited in the form given in the respective HLC files.

Appendix A Audit Procedure

Table A.1. Operational definitions used in Chapter 6.

Step	Procedure
A.1 Extraction	All sheets read with cell values as stored; rows counted if the tree-path (or key) column is non-empty.
A.2 Path normalisation	Remove "**", trailing "-", bracketed level comments " (...)", and sort-only levels "00" and "0000".
A.3 Parent resolution	Local: parent path exists in the same sheet. Cross-file: last code of parent path exists as a node code anywhere in codehl, codehy or codehw.
A.4 Path-anchor congruence	Compare last path segment with HLC-M after stripping , !-, *, ~, +, -; accept 3-digit anchors padded with a trailing 0 in the path; exclude root rows ending in "-".
A.5 Attribute locality	Group codehl rows by base HLC-M; a row is attributed if any of TERRA, AQUA, M04–M10 holds a value other than empty or "(#)"; a group is compliant if at most one row is attributed.
A.6 Key matching	HLC-2 key = text before the first " ", ">", ":", or space, brackets removed; codehu type C rows matched to codehl rows by this key.
A.7 DIAM+ replication	Key = DIAM+ code up to the first "(", " ", ">", "**" or "+"; HLC-M compared for all common keys.
A.8 Derivation convergence	Per codehw scheme sheet: share of rows with HLC-M string equal to HLC-D string.

Appendix B Deviation Register (v8.0.0)

Table B.1. Deviations identified by the audit, with the validation rule that would detect them.

#	File	Location / item	Deviation	Rule
1	codehl	...-976-7881-6741/42/43/46/47	parent 7881 not present	V2
2	codehl	1320–1323, 1273–1276, 3988/3989, 8056, 6977, 1665 and others (25 rows)	path segment ≠ HLC-M (off-by-one)	V3
3	codehl	722, 751, 777, 907	attributes at more than one placement	V4
4	codehl	6974, 7320, 7760, 7820	one anchor for distinct provinces/habitats	A2"
5	codehu/codehl	BV	HLC-M 677~ vs 677	V5
6	codehu/codehl	28 dependency keys (AS, AW, HK, TW, ...)	key notation differs	V10
7	codehu	HLC-3 "MES"	duplicate HLC-3 value	V1
8	codehu	CM:OC	HLC-3 CCM OC lacks U prefix	V1
9	codehy	INFO sheet	sheet references differ from sheet names	doc.
10	codehu	INFO sheet	incomplete HLC-2 definition line; "HCL" spelling	doc.
11	codehu	TYPES	codes 34/35 out of order	doc.
12	codehx	change history	per-file 7.9 labels within 8.0.0 release	doc.