Coordinated Plurality in Hierarchical Location Coding

The Principle of Harmonization without Replacement, its Theoretical Evolution, and its Implementation in the HLC Operational Framework (Version 8.1.2)

Doctoral Thesis

Interdisciplinary study in Information Science, Geography, Philosophy of Classification, Biogeography and Data Engineering

submitted by

Jens Weynans

Primary corpus: HLC Operational Framework, Common Front Matter v8.1.2 (25 October 2026) and data files codehu, codehl, codehy, codehw

October 2026

Abstract

Geographic reference data are produced by many institutions, each with its own purpose, granularity, update cycle and political stance. ISO 3166, UN M49, national statistical classifications, the hydrographic limits of the International Hydrographic Organization, the FAO fishing areas and more than a dozen ecoregion schemes describe overlapping parts of the same world, yet none of them can be reduced to another without loss. The common engineering reaction is either to select one scheme as master and map everything onto it, or to define a new, supposedly better scheme. Both reactions destroy information: the first silences the non-master schemes, the second adds one more competitor.

This thesis examines a third strategy, expressed in the HLC Operational Framework as the maxim *harmonization, not replacement*. It develops this maxim from a programmatic slogan into an explicit principle, termed **Coordinated Plurality**, and decomposes it into ten constituent sub-principles: non-replacement, conflict-as-information, legitimate polyhierarchy, attribute locality, namespace sovereignty, identity persistence, conceptual (non-spatial) assignment, domain-appropriate logics, bounded plurality and declared epistemic limits. Each sub-principle is grounded in an established body of thought from philosophy of classification, information and library science, political and physical geography, biogeography, official statistics and computer science, and is given a formal expression.

Methodologically, the study is a single embedded case study with a theory-building aim, combining qualitative document analysis, formal reconstruction, directed content analysis with a codebook derived from the sub-principles, scripted extraction of corpus indicators and cross-domain case tracing. The corpus comprises the five framework artefacts (the common front matter codehv and the data files codehu, codehl, codehy and codehw), which together contain 837 registered namespace strings, 4,213 land entries, 1,490 water entries and 5,207 rows across 15 ecological scheme sheets.

The analysis identifies the concrete mechanisms of implementation: a federated file architecture, a namespace control layer for two- and three-character codes, a partitioned identifier space (HLC-M), a compact notation grammar, inline attribute tags as an embedded crosswalk (9,081 tag instances from 43 registered scheme keys in the land file alone), starred duplicate placements, redirect chains for historical entities, a deterministic code-derivation pipeline for ecoregions, and explicit admission rules for alternative hierarchies. Cross-domain case studies (the South China Sea and its disputed islands, transcontinental states, Antarctica, three zonal logics, and the Greek EL/GR code divergence) show the mechanisms operating together.

Seven of the ten sub-principles are fully implemented and three partially; the partial ratings fall in the integrity and governance layers. Several residual tensions are shown to be inherent to the principle: stability versus consolidation, expressiveness versus usability, neutrality versus naming, and plurality versus single-curator judgement. The thesis concludes with a generalised reference model, a mapping of HLC constructs to SKOS/RDF, checkable validation rules and a research agenda.

Keywords: classification pluralism; crosswalk; polyhierarchy; geographic coding; ISO 3166; UN M49; ecoregions; hydrographic limits; namespace governance; identifier persistence; knowledge organisation systems; case study; HLC

Note on Scope and Sources

The primary corpus consists of the HLC Operational Framework Common Front Matter, version 8.1.2 dated 25 October 2026, and the working files codehu, codehl, codehy and codehw in version 8.1. All statements about HLC refer to these artefacts; where a statement is an interpretation rather than a description, it is marked as such. Examples are quoted with their original codes so that every claim can be verified against the files. Quantitative indicators were extracted by script from the same files (Chapter 3).

The author of this thesis is also the compiler and maintainer of the HLC framework. Section 3.2 sets out how this insider position is handled methodologically.

In accordance with the brief, this thesis does not catalogue clerical errors, typographical slips or isolated data inconsistencies. Such findings belong to quality assurance, not to the analysis of the principle. Where a structural pattern (for example, two different placement modes for disputed territories) bears on the principle itself, it is discussed as a design question.

All figures were produced by the author. Figures 3, 4, 10 and 14 are generated by script directly from the corpus files; Figure 9 is drawn from the HLC-M allocation table in codehl; the remaining figures are schematic diagrams of structures documented in the corpus.

Table of Contents

Abstract 2

Note on Scope and Sources 3

Abbreviations 8

1 Introduction 9
- 1.1 The Problem of Many Correct Answers 9
- 1.2 Three Strategies of Integration 9
- 1.3 Research Questions 10
- 1.4 Thesis Statement 10
- 1.5 Approach in Brief 10
- 1.6 Structure of the Thesis 11

2 Interdisciplinary Foundations 12
- 2.1 Philosophy of Classification: Pluralism without Relativism 12
- 2.2 Classification as Infrastructure 12
- 2.3 Knowledge Organisation: Polyhierarchy, Facets and Mapping 12
- 2.4 Geography: Fiat Boundaries, Contested Space and Naming 13
- 2.5 Biogeography and Ecology: Partitions as Models 13
- 2.6 Official Statistics and Standards Governance 14
- 2.7 Computer Science: Identifiers, Namespaces and Integration 14
- 2.8 History: Temporal Entities and Succession 14
- 2.9 Synthesis: Requirements Derived from the Disciplines 15

3 Research Design and Methodology 16
- 3.1 Paradigm and Design Logic 16
- 3.2 Positionality: Insider Research 17
- 3.3 Corpus Definition and Profile 17
- 3.4 Data Extraction Procedure 19
- 3.5 Analytical Procedures 19
- 3.6 Case Selection 21
- 3.7 Evaluation Procedure 21
- 3.8 Quality Criteria 21
- 3.9 Methodological Limitations 22

4 Evolution of the Principle: From Maxim to Coordinated Plurality 23
- 4.1 Genealogy: The Principle Emerged from Practice 23
- 4.2 Basic Concepts 24
- 4.3 Formal Model 24
- 4.4 The Ten Constituent Sub-Principles 25
- 4.5 Interdependencies and Layering 28
- 4.6 Neutrality Reconsidered: Attributed Neutrality and Placement Modes 29
- 4.7 Distinction from Neighbouring Approaches 30
- 4.8 The Evolved Principle 31

5 The Architecture of HLC as Implementation of the Principle 32
- 5.1 Federated File Architecture 32
- 5.2 The Namespace Control Layer (codehu) 33
- 5.3 The Entry-Type Matrix 33
- 5.4 The Identifier Space HLC-M 34
- 5.5 Tree Paths 36
- 5.6 The Notation as an Assertion Grammar 36
- 5.7 Inline Attribute Tags as an Embedded Crosswalk 36
- 5.8 Governance Artefacts 37

6 Implementation in the Land Domain (codehl) 39
- 6.1 Column Architecture 39
- 6.2 Layering of Partitions before Politics 39
- 6.3 Statistical and National Plurality 39
- 6.4 Contested Territories and Placement Modes 40

6.5 Transcontinental States ... 41
6.6 Historical Entities and Succession ... 42
6.7 Organisations and Logical Entries ... 42
6.8 Points, Installations and Below-Country Granularity ... 43

7 Implementation in the Hydrographic Domain (codehy) ... 44
7.1 Four Parallel Code Systems ... 44
7.2 Temporal Plurality: Four Editions of the Limits of Oceans and Seas ... 44
7.3 Polyhierarchy of Seas ... 45
7.4 Fisheries and Regulatory Areas as Attributed Assertions ... 46
7.5 Water Columns and Depth Zones ... 46
7.6 Inland Waters and Cross-Domain Links ... 46
7.7 Coordinates as Anchors ... 47

8 Implementation in the Ecological Domain (codehw) ... 48
8.1 Sibling Views without Master ... 48
8.2 The Identifier Pipeline: from Source Code to HLC-M ... 49
8.3 The Replacement Grammar ... 50
8.4 Integration of Overlay Schemes ... 50
8.5 Explicit Cross-Scheme Correspondences ... 51
8.6 Umbrella Typologies ... 51
8.7 Links from codehw to the Other Files ... 51

9 Cross-Domain Case Studies ... 52
9.1 The South China Sea and its Islands ... 52
9.2 Egypt, Sinai and the Suez Canal ... 52
9.3 Antarctica ... 53
9.4 Three Zonal Logics ... 54
9.5 Greece: EL and GR ... 54
9.6 Synthesis of the Cases ... 54

10 Evaluation ... 55
10.1 Rating Scheme ... 55
10.2 Conformance Matrix ... 55
10.3 Inherent Tensions ... 56
10.4 The Spreadsheet Substrate ... 57
10.5 What the Implementation Adds to the Principle ... 58

11 Generalisation and Transfer ... 59
11.1 A Reference Model for Coordinated Plurality ... 59
11.2 Mapping HLC to SKOS and RDF ... 60
11.3 Validation Rules Derived from the Principle ... 60
11.4 Transfer to Other Domains ... 61
11.5 Research Agenda ... 61

12 Conclusion ... 62

Appendix A: Formal Grammar of the HLC Notation ... 63

Appendix B: HLC-M Identifier-Range Allocation ... 64

Appendix C: Inline Attribute Tag Keys (Selection) ... 65

Appendix D: Glossary ... 66

Bibliography ... 67
Primary Sources ... 67
Methodological Literature ... 67
Theoretical Literature ... 67
Integrated Schemes and Standards (as referenced by HLC) ... 68

List of Tables

Table 1: Abbreviations used in this thesis 8
Table 2: Requirements for a harmonizing location framework derived from the disciplines..... 15
Table 3: Alignment of research questions, methods, data and outputs..... 17
Table 4: Corpus profile (extracted by script from the v8.1 files) 18
Table 5: Codebook for the directed content analysis (abridged) 20
Table 6: Case selection matrix..... 21
Table 7: Quality criteria and the measures taken..... 21
Table 8: Phases in the development of HLC as documented in the change history 23
Table 9: Dependencies between the sub-principles..... 29
Table 10: Placement modes for contested referents 30
Table 11: Coordinated plurality compared with neighbouring approaches..... 30
Table 12: Distribution of responsibilities across the HLC files..... 32
Table 13: Code classes registered in codehu for the two-character layer (HLC-2) 33
Table 14: Structure of the HLC entry-type matrix (codehu, TYPES) 34
Table 15: The HLC notation as an assertion grammar 36
Table 16: Observed placement modes for contested entries in codehl 40
Table 17: Representation of transcontinental states..... 41
Table 18: The four parallel code systems in codehy..... 44
Table 19: Multiple placements of the South China Sea branches in codehy 45
Table 20: Ecological schemes held as siblings in codehw..... 48
Table 21: Selected rules of the codehw replacement grammar 50
Table 22: The South China Sea across the HLC files..... 52
Table 23: Antarctic claims and units in codehu..... 53
Table 24: Three zonal logics in HLC..... 54
Table 25: Sub-principles observed in the case studies..... 54
Table 26: Conformance of HLC v8.1 with the sub-principles of Coordinated Plurality..... 55
Table 27: Reference model for Coordinated Plurality and its instantiation in HLC..... 59
Table 28: Mapping of HLC constructs to SKOS, Dublin Core and PROV..... 60
Table 29: Semantics of the grammar productions 63
Table 30: HLC-M range allocation by owner..... 64
Table 31: Selected inline attribute tag keys 65
Table 32: Glossary of key terms 66

List of Figures

Figure 1: Three strategies for integrating plural classification schemes 10
Figure 2: Research design: five phases with iterative refinement and cross-cutting quality measures . 16
Figure 3: Corpus profile: (a) coded entries per data file; (b) rows per ecological scheme sheet in codehw 18
Figure 4: Distribution of inline-attribute-tag keys in codehl (top 15 keys) 19
Figure 5: Development phases and selected milestones of HLC, 2021–2026 (from the codehv change history) 23
Figure 6: The ten sub-principles in four layers, with "requires" dependencies 29
Figure 7: Four placement modes for contested referents, with examples from codehl 30
Figure 8: Federated architecture of HLC: policy, namespace and content layers (entry counts from v8.1) 32
Figure 9: Ownership of the numeric HLC-M space by first two digits (codehl, HLC-M sheet) 35
Figure 10: The embedded crosswalk: nineteen attributed assertions on the codehl entry for Iran 37
Figure 11: Whole-and-parts representation of Egypt and the cross-domain identity of the Suez Canal41
Figure 12: Polyhierarchy of the South China Sea branches: primary and starred secondary placements45
Figure 13: The codehw identifier pipeline, traced through one TEOW row 49
Figure 14: Conformance profile of HLC v8.1 by sub-principle and layer, with selected corpus indicators 56
Figure 15: Substrate-agnostic reference model for Coordinated Plurality 59

Abbreviations

Table 1: Abbreviations used in this thesis

Abbreviation	Meaning
ACIC	Aeronautical Chart and Information Center (Chart of Limits of Seas and Oceans, M49-1)
BEOW	Biological Ecoregions of the World (RESOLVE, Dinerstein et al. 2017)
BPOW	Benthic Provinces of the World (Maureaud et al. 2023)
BYES	Bailey's Ecoregions of the World (map 1989; book 1998, 2nd ed. 2014)
COW	Correlates of War (state system data; Gleditsch & Ward variant)
CWWN / CWWN+	Codes for World Waterbody Names / with HLC extensions
DIAM / DIAM+	Defense Intelligence Agency Manual 65-18 codes / with HLC extensions
EOW / SEOW	Ecosystems of the World (Goodall, ed.) / Subterranean Ecosystems
EUNIS	European Nature Information System habitat classification
FAO / FAO+	FAO Major Fishing Areas / HLC-restructured FAO-style areas
FEOW	Freshwater Ecoregions of the World (Abell et al. 2008)
GENC	US Geopolitical Entities, Names, and Codes
GET	IUCN Global Ecosystem Typology v2.1
GEZ	FAO Global Ecological Zones (FRA 2000)
GIWA	Global International Waters Assessment (UNEP)
GOODS	Global Open Oceans and Deep Seabed (Watling et al. 2013)
HLC	Hierarchical Location Codes
HLC-2 / HLC-3	Two- and three-character code layers controlled by codehu
HLC-M	HLC main numeric/alphanumeric identifier (3–4 characters)
HLC-X / -S / -D	Technical code / shortcut code / derived code
IAT	Inline Attribute Tag, syntax {scheme:value}
IHO S-23	Limits of Oceans and Seas (IHO), editions 1953, 1986 draft, 2002 draft
ISO 3166-1/-2/-3	Country codes / subdivision codes / formerly used codes
LME	Large Marine Ecosystem
M49	UN Standard Country or Area Codes for Statistical Use
MEOW	Marine Ecoregions of the World (Spalding et al. 2007)
PAME / AMA	Protection of the Arctic Marine Environment / Arctic Marine Areas
PPOW	Pelagic Provinces of the World (Spalding et al. 2012)
SDMX	Statistical Data and Metadata eXchange (ISO 17369)
SKOS	W3C Simple Knowledge Organization System
TEOW	Terrestrial Ecoregions of the World (Olson et al. 2001)
WGS-RPD	World Geographical Scheme for Recording Plant Distributions
WGS84	World Geodetic System 1984

1 Introduction

1.1 The Problem of Many Correct Answers

Asking which region Egypt belongs to has several correct answers. In UN M49, Egypt is a country of Northern Africa. In the World Geographical Scheme for Recording Plant Distributions, the Sinai Peninsula belongs to a different botanical region than the Nile valley. Several national statistical classifications place Egypt with the Middle East for trade purposes. In a hydrographic perspective, the Suez Canal is a waterway connecting the Mediterranean with the Red Sea, and is thus an object of the water domain rather than of the land domain. None of these answers is wrong. Each is correct relative to a purpose, a community of practice and a moment in time.

Information systems, however, tend to require a single answer. A database column holds one region code; a foreign key points to one parent; a dropdown list offers one hierarchy. The pressure towards one answer is a technical pressure, not an epistemic one, but it shapes what users come to believe about the world. When a statistical office adopts one regional grouping, the alternatives disappear from view, even though they continue to govern other datasets with which the first will eventually have to be combined.

Location coding is a particularly acute case of this problem for four reasons. First, the objects are politically charged: sovereignty, recognition and naming are contested. Second, the objects change: states merge, split, rename and disappear, and their codes are retired or reused. Third, the objects span ontological domains: land, water and living systems interpenetrate, so that one place may simultaneously be a country, a coastal sea, a fishing area and an ecoregion. Fourth, the classifications are produced by actors with incompatible mandates: a standards body, a statistical service, an intelligence agency, a fisheries organisation and a conservation NGO each partition the earth for their own reasons.

1.2 Three Strategies of Integration

Faced with a plurality of schemes, an integrator can follow one of three strategies (Figure 1).

1. **Replacement (unification).** A new scheme is designed that claims to supersede the existing ones. This is the classical strategy of standardisation. Its strength is simplicity for downstream users; its weakness is that it rarely succeeds in practice. Existing schemes persist because they are embedded in law, budgets and legacy data, so the new scheme becomes one more competitor.
2. **Reduction (master-mapping).** One existing scheme is designated as master, and all others are mapped onto it. This is the dominant strategy of data warehousing. Its weakness is asymmetric loss: whatever the master cannot express (a different parent, a finer subdivision, a historical predecessor, a disputed status) is either discarded or forced into an approximate match.
3. **Harmonization (coordinated plurality).** All schemes are retained as authoritative within their own domains. A coordinating layer provides stable identifiers, makes correspondences explicit, records divergences instead of resolving them, and controls the shared namespace so that the coordinating layer itself does not become a source of ambiguity.

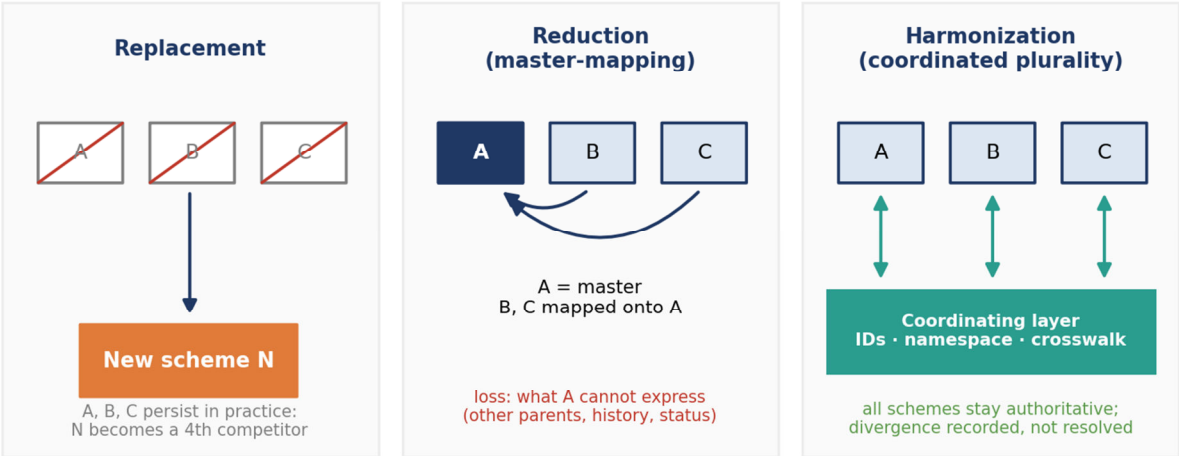

Figure 1: Three strategies for integrating plural classification schemes

The HLC Operational Framework positions itself explicitly in the third category. Its common front matter states that its purpose is "harmonization, not replacement" and that classificatory disagreement is treated "as information, not error". These formulations are the starting point of this thesis.

1.3 Research Questions

The thesis addresses two linked questions, corresponding to the two aims set for it.

- **RQ1 (evolution of the principle).** What exactly does "harmonization, not replacement" mean when it is developed from a maxim into a principle that can guide design decisions, be tested against an implementation and be transferred to other domains? Which constituent sub-principles does it entail, how are they grounded across disciplines, and how do they relate to one another?
- **RQ2 (implementation).** Through which concrete mechanisms (architecture, identifiers, notation, data structures, governance rules) does the HLC framework implement this principle in its land, water and ecological domains, to what degree is each sub-principle realised, and which tensions arise from the implementation?

1.4 Thesis Statement

Central claim
Harmonization without replacement can be stated as a coherent design principle, **Coordinated Plurality**, consisting of ten interdependent sub-principles. The HLC framework is a working implementation of this principle. It realises the principle not through a single feature but through the interplay of a federated architecture, a controlled namespace, a compact assertion notation and an embedded crosswalk. Its residual tensions are properties of the principle itself and indicate where any implementation of coordinated plurality must make explicit trade-offs.

1.5 Approach in Brief

The study is designed as a single embedded case study of the HLC framework with a theory-building aim. It combines conceptual analysis, formal reconstruction, directed content analysis of the artefacts, scripted extraction of corpus indicators and cross-domain case tracing. Chapter 3 sets out the research design, the corpus, the analytical procedures, the evaluation scheme and the quality criteria in detail.

1.6 Structure of the Thesis

Chapter 2 assembles the interdisciplinary foundations. Chapter 3 presents the research design and methodology. Chapter 4 develops the principle: definitions, a formal model and the ten sub-principles. Chapter 5 describes the architecture through which HLC implements the principle across all files. Chapters 6 to 8 analyse the implementation in the land, water and ecological domains. Chapter 9 presents cross-domain case studies. Chapter 10 evaluates the implementation and discusses inherent tensions. Chapter 11 generalises the findings into a reference model, maps HLC to semantic-web standards and outlines further research. Chapter 12 concludes. The appendices formalise the notation, reproduce the identifier-range allocation and the principal tag keys, and provide a glossary.

2 Interdisciplinary Foundations

The principle examined here belongs to no single discipline. Its premises come from philosophy (what classifications are), information science (how classifications are represented and linked), geography (what spatial units are and who defines them), biogeography and ecology (why nature admits several partitions), official statistics (how classifications are governed) and computer science (how identifiers and namespaces behave). This chapter extracts from each field the specific insight on which the principle relies.

2.1 Philosophy of Classification: Pluralism without Relativism

Classical essentialism holds that the world is divided into natural kinds and that a correct classification follows these divisions. Under this view, competing classifications are either equivalent or partly wrong, and the task of an integrator is to find the right one. The essentialist view has been challenged from several directions. Dupré's *promiscuous realism* argues that there are many equally real ways of dividing the world, each grounded in real features but selected by different interests. Goodman's *ways of worldmaking* emphasises that classificatory schemes are constructed under criteria of rightness that are plural but not arbitrary. Foucault's discussion, at the opening of *The Order of Things*, of a fictional Chinese encyclopaedia quoted by Borges illustrates how alien an unfamiliar classificatory principle can appear while being internally coherent.

Three consequences matter for location coding. First, the existence of several partitions of the same territory is not a defect to be repaired but a normal condition. Second, pluralism is not relativism: a partition can still be wrong relative to its own criteria, so each scheme must be represented faithfully on its own terms. Third, the integrator needs criteria for *which* partitions to admit, otherwise the integration dissolves into an unbounded list. The HLC admission rules (Section 4.4.9) answer this third consequence.

2.2 Classification as Infrastructure

Bowker and Star's *Sorting Things Out* treats classifications as infrastructure: largely invisible, embedded in practice, and consequential for those who are classified. Two concepts are directly relevant. The first is the *boundary object* introduced by Star and Griesemer, an artefact that is robust enough to maintain identity across communities while being flexible enough to serve each community's needs. A shared code for a place is a boundary object: the statistician, the fisheries manager and the ecologist each read it differently, yet must be able to agree that they mean the same thing. The second is the *residual category* ("other", "not elsewhere classified"), through which a classification hides what it cannot handle. A harmonizing framework must make residuals explicit rather than inheriting them silently.

From this perspective, the decisive design question is not "which classification is right" but "how can several classifications coexist as infrastructure without one of them invisibly dominating". HLC's answer is to make each source scheme visible at the level of every single entry, through inline attribute tags.

2.3 Knowledge Organisation: Polyhierarchy, Facets and Mapping

Library and information science has long distinguished between monohierarchies, in which every concept has exactly one broader concept, and polyhierarchies, in which a concept may have several.

Thesaurus standards (ISO 25964) admit polyhierarchy, and the W3C SKOS recommendation provides `broader` and `narrower` relations without imposing a tree. Ranganathan's faceted classification takes a different route: instead of multiplying parents, it decomposes subjects into independent facets (personality, matter, energy, space, time), each with its own scheme.

Mapping between independently maintained vocabularies is a further concern. SKOS distinguishes `exactMatch`, `closeMatch`, `broadMatch`, `narrowMatch` and `relatedMatch`. The distinction matters because correspondences between classifications are rarely one-to-one. A UN M49 region and a WGS-RPD botanical region overlap but do not coincide; a FAO fishing division and an IHO sea area have different limits drawn for different purposes.

HLC combines both traditions. Its tree is polyhierarchical (via starred duplicates), and its tag columns behave like facets: TERRA, AQUA, telecommunications, statistics, trade and national classifications are kept in separate columns rather than merged into one hierarchy. Its mapping semantics are expressed in a lightweight notation (parentheses for inherited, old or approximate values) rather than in an explicit relation vocabulary, which is analysed in Chapter 11.

2.4 Geography: Fiat Boundaries, Contested Space and Naming

Smith and Varzi distinguish *bona fide* boundaries, which correspond to physical discontinuities such as coastlines, from *fiat* boundaries, which exist by human decision, such as most political borders and all statistical regions. Most units in location coding are fiat objects. Their limits depend on an authority's decision, and different authorities decide differently. Even apparently physical objects such as seas become fiat objects when they are delimited, as the successive editions of the IHO *Limits of Oceans and Seas* demonstrate.

Political geography adds that many fiat objects are contested. Disputed islands, divided territories, occupied zones and unclaimed land (*terra nullius*) are a structural part of the world map. Critical cartography, following Harley, has shown that maps are never neutral: the choice of a name, a colour or a parent asserts a position. A framework that wants to be neutral therefore cannot simply avoid positions. It must instead record *whose* position each assertion represents.

Toponymy contributes the observation that names are themselves plural and political: exonyms and endonyms, historical and current names, and names in competing languages all refer to the same object. A coding framework must separate identity (the code) from naming (the labels).

2.5 Biogeography and Ecology: Partitions as Models

Ecoregionalisation is a modelling exercise. Terrestrial schemes (Olson et al. 2001; Dinerstein et al. 2017), freshwater schemes (Abell et al. 2008), marine coastal schemes (Spalding et al. 2007), pelagic schemes (Spalding et al. 2012; Longhurst), deep-sea schemes (Watling et al. 2013; Maureaud et al. 2023) and functional typologies (Keith et al. 2022) partition the biosphere according to different criteria: species assemblages, physiography, climate, biogeochemistry, depth or ecosystem function. They are explicitly designed for different purposes, from conservation priority-setting to ocean biogeochemistry.

Two lessons follow. First, ecological partitions are not competing approximations of one true partition. They are complementary models whose disagreement reflects different explanatory targets. Second, ecological schemes have their own internal hierarchies (realm, biome, province, ecoregion) that do not

nest into political hierarchies. Any attempt to merge them into one tree would produce artefacts. This is why HLC keeps ecoregion schemes as *sibling views* without a master (Chapter 8).

2.6 Official Statistics and Standards Governance

Official statistics depend on classifications that are stable enough for time series and flexible enough to reflect change. ISO 3166 is maintained by a maintenance agency that adds, changes and withdraws codes, and ISO 3166-3 records formerly used codes with four-letter identifiers. UN M49 is explicitly a statistical classification: its groupings imply no political judgement about the status of countries or territories. National statistical offices (for example in Germany, Switzerland, Norway, Italy, France, Australia, New Zealand, South Africa, Argentina and Japan) maintain their own country and territory classifications, which differ from ISO in the treatment of dependencies, disputed areas and historical units. SDMX (ISO 17369) provides exchange formats for such code lists.

The governance lesson is that code lists are authoritative only within their mandate. An integrator that does not respect this would implicitly claim a mandate it does not have. HLC's explicit disclaimer ("not an authority") is therefore a precondition of the principle, not a mere legal caveat.

2.7 Computer Science: Identifiers, Namespaces and Integration

Data integration research distinguishes *global-as-view* integration, in which a global schema is defined in terms of the sources, from *local-as-view* integration, in which each source is described as a view over a global schema (Lenzerini 2002). Coordinated plurality is closer to a third, federated model: sources keep their schemas, and a mediating layer stores correspondences and identities without claiming to subsume the sources.

Identifier theory contributes two insights. First, identifiers should be persistent and should not change when names, attributes or classifications change ("cool URIs don't change", in Berners-Lee's formulation). Second, identifiers must be unambiguous within a namespace, so that the same string never refers to two different things in the same context. Code lists that reuse short strings (two letters, two digits) across purposes are a well-known source of integration errors. A coordinating framework must therefore control its namespace even more carefully than the sources do, because it hosts all of their strings at once.

Ontology engineering, from Gruber's definition of an ontology as an explicit specification of a conceptualisation onwards, shows that alignments between vocabularies are themselves data, with provenance and confidence. An alignment between a political region and an ecoregion is a claim by someone, not a fact about the world.

2.8 History: Temporal Entities and Succession

Historical geography deals with entities that existed for a period and then disappeared, merged or split: empires, protectorates, princely states, sultanates, occupation zones. Their relations to present entities are relations of succession, partition and absorption, not of containment. The Correlates of War project and the Gleditsch-Ward list of independent states are examples of datasets that encode such entities with explicit time spans for quantitative research. A location framework that wants to support longitudinal analysis must represent such entities and their succession relations without confusing them with present-day containment.

2.9 Synthesis: Requirements Derived from the Disciplines

Table 2: Requirements for a harmonizing location framework derived from the disciplines

Discipline	Insight	Requirement for the framework
Philosophy of classification	Several partitions can be equally real; pluralism is not relativism	Retain multiple partitions; represent each faithfully; define admission criteria
Infrastructure studies	Classifications are boundary objects with invisible residuals	Shared identifiers across communities; make residual and special categories explicit
Knowledge organisation	Polyhierarchy, facets, typed mapping relations	Allow multiple parents; keep facets separate; distinguish exact from approximate correspondence
Geography	Most units are fiat; boundaries and names are contested	Separate identity from name and boundary; record whose assertion a placement is
Biogeography	Ecological partitions are complementary models	Hold schemes as siblings; avoid forced nesting into political trees
Official statistics	Code lists are authoritative only within their mandate	Never redefine external codes; disclaim authority
Computer science	Persistent, unambiguous identifiers; alignments as data	Stable IDs; controlled namespace; correspondences with provenance
History	Entities have life spans and succession relations	Retain historical entities; express succession by redirect, not containment

3 Research Design and Methodology

This chapter states how the two research questions are answered. It sets out the research paradigm and design logic, the handling of the author's insider position, the corpus and its profile, the extraction procedure, the analytical procedures, the selection of cases, the evaluation scheme and the quality criteria. Figure 2 summarises the design.

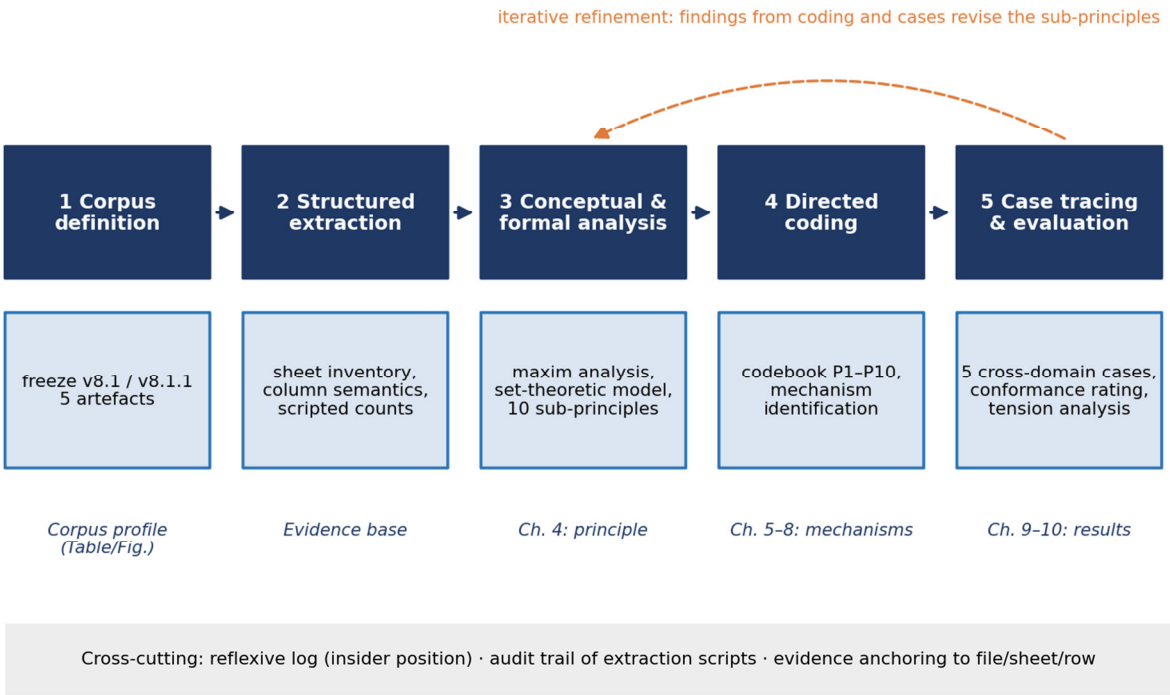

Figure 2: Research design: five phases with iterative refinement and cross-cutting quality measures

3.1 Paradigm and Design Logic

The study answers two questions of different kinds. RQ1 asks what a principle *means*: it is conceptual and requires interpretation, explication and formalisation. RQ2 asks how an artefact *realises* that principle: it is empirical and requires systematic examination of the artefact. The study therefore adopts an interpretive stance for RQ1 and a structured, evidence-anchored stance for RQ2, linked by a common formal model.

The overall design is a **single embedded case study** in the sense of Yin (2018): the case is the HLC framework in version 8.1/8.1.2, and the embedded units of analysis are the five artefacts, their sheets, their mechanisms and five selected referents traced across all files. A single case is appropriate for three reasons. First, the case is *revelatory*: HLC is, to the author's knowledge, the only documented framework that integrates political, statistical, hydrographic and ecological schemes under an explicit non-replacement maxim, so the phenomenon cannot be studied comparatively at present. Second, the case is *critical* in Flyvbjerg's (2006) sense: if the principle can be implemented coherently with deliberately simple tools by a single curator, it can be expected to be implementable under more favourable conditions. Third, the aim is *theory-building* (Eisenhardt 1989): the principle is developed in iteration with the evidence rather than tested as a fixed hypothesis.

Because the study evaluates a designed artefact against design principles, it also borrows from design-science research (Hevner et al. 2004). The ten sub-principles function as design requirements; the

HLC files are the artefact; the conformance matrix in Chapter 10 is the evaluation; the reference model, the RDF mapping and the validation rules in Chapter 11 are the generalised design knowledge. The study does not build the artefact as part of the research; it evaluates and generalises an existing one.

Table 3: Alignment of research questions, methods, data and outputs

Question	Method	Data	Output (chapter)
RQ1 What does the principle mean?	Conceptual analysis; interdisciplinary synthesis; formal reconstruction	codehv statements; literature (Chapter 2)	Ten sub-principles, layering, formal model, evolved principle (Chapter 4)
RQ1 How did it arise?	Genealogical reading of the change history	codehv change history 2021–2026	Four development phases (Section 4.1)
RQ2 Which mechanisms implement it?	Directed content analysis with codebook; scripted extraction	All sheets of codehu, codehl, codehy, codehw	Mechanism inventory (Chapters 5–8)
RQ2 Do mechanisms interact consistently?	Cross-domain case tracing	Five referents across all files	Case analyses (Chapter 9)
RQ2 To what degree is each sub-principle realised?	Rule-based rating; tension analysis	Coded evidence and indicators	Conformance matrix, tensions (Chapter 10)

3.2 Positionality: Insider Research

The author is the compiler and maintainer of HLC. This is insider research in the sense of Brannick and Coghlan (2007): research by a full member of the system studied. Insider status brings *pre-understanding* (detailed knowledge of design intent and of decisions not visible in the files) and *access* (complete, unrestricted access to all versions). It also brings two risks: a tendency to read intended properties into the artefact, and reluctance to identify weaknesses.

Five measures address these risks.

1. **Evidence anchoring.** Every claim about HLC is tied to a verifiable location (file, sheet, column, code or row). Design intent that is not documented in the artefacts is not used as evidence.
2. **Criteria before evaluation.** The sub-principles and the rating rules (Section 3.7) were fixed before the conformance assessment, so that the evaluation cannot be adjusted to the result.
3. **Scripted indicators.** Quantitative statements are produced by extraction scripts rather than by recollection (Section 3.4).
4. **Explicit search for disconfirming evidence.** For each sub-principle, the coding included a dedicated category for counter-instances (Section 3.5.3), and the evaluation reports limiting factors with the same prominence as supporting evidence.
5. **Separation of description and judgement.** Chapters 5 to 9 describe mechanisms; judgement is concentrated in Chapter 10, where it can be checked against the descriptive chapters.

The insider position is also an asset that the design uses deliberately: the genealogical analysis (Section 4.1) and the identification of implicit rules of practice would be difficult for an outsider to perform with the same completeness.

3.3 Corpus Definition and Profile

The corpus is defined by version, not by date of access. It comprises the Common Front Matter in version 8.1.2 (23 October 2026) and the four data files in version 8.1, frozen for the duration of the

analysis. Earlier versions are used only through the change history recorded in the front matter. Published PDF, TXT and TSV/SSV renditions are not analysed separately because they are derived from the same workbook sources. External standards are not part of the corpus; they are consulted only to the extent needed to interpret HLC's representation of them.

Table 4: Corpus profile (extracted by script from the v8.1 files)

Artefact	Sheets	Principal data sheet	Coded entries	Selected indicators
codehv (DOCX)	–	Front matter text	–	Purpose, notation, policies, limits, admission rules, licence, change history 2021–2026
codehu	4	DATA (HU)	837	134 colon (logical/organisational) entries; 327 entries with | alternatives; 9 redirects >; 251 type C, 86 R, 80 W, 76 S, 68 L, 63 I, 45 X entries
codehl	4	DATA (HL)	4,213	Tree depth up to 13 levels; 9,081 inline tags from 43 keys; 44 rows with claim annotations; 208 rows with WGS84 tags; 663 historical (H) entries
codehy	6	DATA (HY)	1,490	102 starred placements; 656 inline tags from 11 keys; 954 WGS84 anchors; 152 rows with IHO 2002 and 107 with IHO 1953 identifiers
codehw	18	15 scheme sheets	5,207	Largest sheets BEOW 1,171, TEOW 1,149, EUNIS 740, FEOW 575; smallest GEZ 26

Figure 3 shows the size of the data files and the distribution of rows across the ecological scheme sheets; Figure 4 shows the most frequent tag keys in the land file. Both figures are generated from the corpus with the extraction procedure described in Section 3.4. The tag-key distribution already indicates the breadth of the embedded crosswalk: the most frequent key belongs to the Correlates of War data (historical states), followed by the botanical WGS-RPD scheme and the Australian SACC classification, while UN M49 ranks only fourth. No single external scheme dominates the crosswalk.

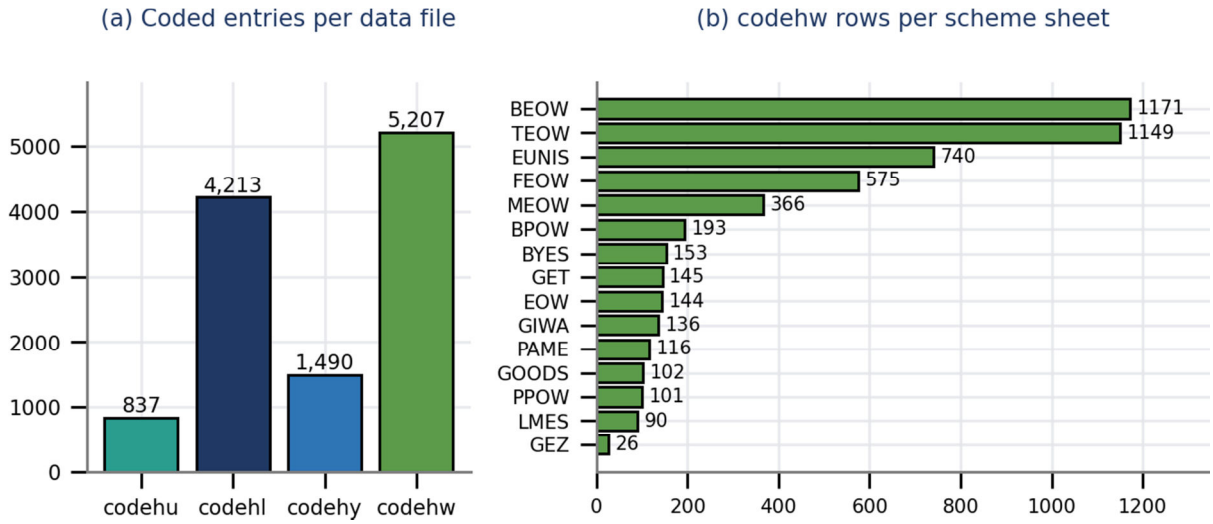

Figure 3: Corpus profile: (a) coded entries per data file; (b) rows per ecological scheme sheet in codehw

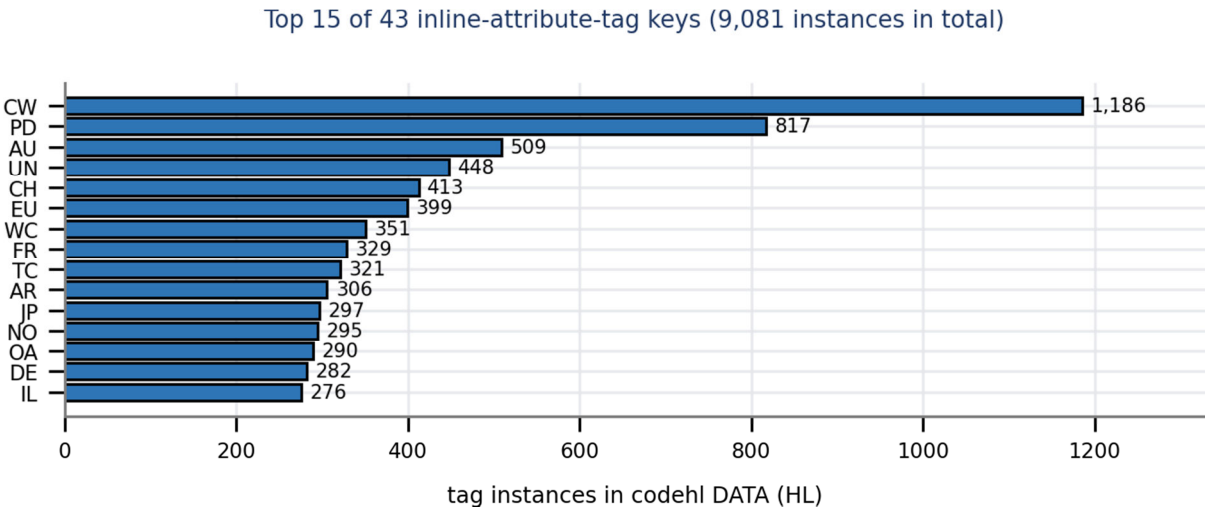

Figure 4: Distribution of inline-attribute-tag keys in codehl (top 15 keys)

3.4 Data Extraction Procedure

Extraction followed a fixed protocol so that every count in the thesis can be reproduced.

1. **Sheet inventory.** All workbooks were opened read-only with cell values (not formulas). For each sheet, the header row was recorded and each column was assigned a semantic role (identifier, type, code layer, description, tag column, coordinate, edition column).
2. **Entry definition.** An *entry* is a row whose identifier column is non-empty: HLC-2 in codehu, HLC-M in codehl and codehy, HLC-M in each codehw scheme sheet. Blank separator rows and header rows are excluded.
3. **Pattern extraction.** Inline tags were extracted with the regular expression `\{([A-Z0-9]{2}):` across all columns of a row; starred placements by the presence of `*` in the identifier column; redirects by `>` in the HLC-2 column; logical entries by `:`; claim annotations by the string "claimed"; coordinate anchors by WGS84 tags or the coordinate column.
4. **Structural measures.** Tree depth was computed from the number of separators in the tree path; type distributions from the first character of the type column.
5. **Qualitative excerpting.** For each mechanism, representative rows were excerpted with their full identifiers. Every example quoted in Chapters 5 to 9 is such an excerpt.

The counts are descriptive indicators, not measures of quality. They show the extent to which a mechanism is used; whether its use conforms to a sub-principle is a qualitative judgement made under the rules in Section 3.7.

3.5 Analytical Procedures

3.5.1 Conceptual Analysis of the Maxim

The programmatic statements of the front matter ("harmonization, not replacement", "disagreement as information, not error", "coordinated but not merged", "not an authority") were analysed for their presuppositions and implications. Each statement was asked three questions: what must be true of a framework for the statement to hold; which failure would the statement prevent; and which discipline provides the reason for the statement. The answers were synthesised with the requirements in Section 2.9 into candidate sub-principles.

3.5.2 Formal Reconstruction

The candidate sub-principles were restated as constraints on a set-theoretic model of a coordinating framework (Section 4.3). Formalisation served two methodological purposes: it exposed ambiguities in the informal statements (for example, whether neutrality means the absence of assertion or the attribution of assertion), and it produced operational definitions against which the artefacts could be coded.

3.5.3 Directed Content Analysis

The artefacts were coded with a directed approach in the sense of Hsieh and Shannon (2005): the initial codes were derived from the sub-principles rather than from the data, and new codes were added where the data contained mechanisms that the initial codebook did not anticipate. Document analysis followed Bowen (2009): documents were treated as social artefacts produced for a purpose, and their statements were triangulated with the data structures they describe.

Table 5: Codebook for the directed content analysis (abridged)

Code	Operational definition	Indicators in the artefacts	Counter-instance code
P1 Non-Replacement	External code appears unaltered and attributed to its scheme	IATs `{k:v}`; original codes in brackets; separate HLC-owned schemes	External code altered or unattributed
P2 Conflict as Information	Divergent external positions both recoverable	Claim annotations; edition columns; parallel national tags	Single position adopted silently
P3 Polyhierarchy	Entry reachable from several parents with marked secondary placement	Starred identifiers; whole-and-parts entries	Duplicate without marker
P4 Attribute Locality	Attributes only at primary placement	Secondary rows without tags/coordinates	Attributes on a starred row
P5 Namespace Sovereignty	One owner and class per string / range	codehu register; HLC-M range table; replacement grammar	String with two owners or classes
P6 Identity Persistence	Stable identifiers; succession by redirect	>; historical entries; deterministic derivation	Reassigned or deleted identifier
P7 Conceptual Assignment	Membership not based on geometry	Coordinate policy; anchors as points/ranges	Geometric claim in hierarchy
P8 Domain Logics	Domain-specific hierarchies linked by identifiers	File federation; type matrix; sibling sheets	Entry forced into foreign type system
P9 Bounded Plurality	Alternatives admitted by stated criteria	Admission rules; registered scheme keys	Unexplained alternative hierarchy
P10 Declared Limits	Limits of authority, precision, currency stated	Limitations section; licence scope	Implicit claim of authority

Two codes were added inductively during analysis because the initial codebook did not foresee them: *whole-and-parts decomposition* (transcontinental states, Section 6.5) as an alternative realisation of P3, and *edition parallelism* (four IHO/ACIC editions, Section 7.2) as a temporal realisation of P2. Both were subsequently integrated into the evolved principle (Section 4.8), which is the iterative refinement loop shown in Figure 2.

3.5.4 Cross-Domain Case Tracing

For each selected referent, all occurrences in all five artefacts were located by identifier, code and name; each occurrence was recorded with its placement, type and tags; and the relations between occurrences were classified as containment, secondary placement, cross-domain identity, succession or attributed assertion. The resulting traces are reported in Chapter 9.

3.6 Case Selection

Cases were selected purposively for maximum variation on four dimensions that stress different sub-principles: number of domains involved, degree of contestation, temporal depth and scale of the unit. Random selection would not serve the purpose, because the aim is to observe mechanisms under strain, not to estimate frequencies.

Table 6: Case selection matrix

Case	Domains	Contestation	Temporal depth	Scale	Primary sub-principles stressed
South China Sea and islands	Water, land, ecology	High (six claimants)	Four IHO/ACIC editions	Sea region	P2, P3, P4, P8
Egypt, Sinai, Suez	Land, water	Low (historical zones)	COW-coded demilitarised zones	State and sub-units	P3, P6, P8
Antarctica	Land, water, ecology	Frozen claims	Claims since early 20th c.	Continent	P2, P8, P10
Three zonal logics	All (zones)	None (scientific plurality)	Low	Global	P2, P3, P9
Greece EL/GR	Namespace only	Institutional divergence	Low	Single string	P1, P5, P6

3.7 Evaluation Procedure

Each sub-principle was rated on the scale *full*, *partial* or *implicit* using fixed decision rules.

- **Full** requires (a) an explicit statement of the rule in codehv or an INFO sheet, (b) a mechanism realising it in every file where it is relevant, and (c) no counter-instance that reflects a structural gap rather than an isolated slip.
- **Partial** applies when (a) holds but (b) or (c) does not: the mechanism is absent from some relevant file, depends on inference by the reader, or has a structural counter-instance.
- **Implicit** applies when (a) does not hold but the rule can be reconstructed from consistent practice in the data.

A rating required evidence from at least two independent locations in the corpus. Isolated clerical inconsistencies were not treated as counter-instances, in line with the scope of the study; structural patterns were. Tensions (Section 10.3) were identified where improving conformance with one sub-principle would necessarily reduce conformance with another.

3.8 Quality Criteria

Table 7: Quality criteria and the measures taken

Criterion	Meaning (Lincoln & Guba 1985; Yin 2018)	Measures in this study
Credibility / construct validity	Findings represent the phenomenon	Operational codebook; multiple sources (front matter, INFO sheets, data rows); quantitative indicators triangulated with excerpts
Transferability / external validity	Findings can inform other contexts	Analytical generalisation to a substrate-agnostic reference model; explicit transfer domains (Section 11.4)
Dependability / reliability	Procedure is consistent and repeatable	Frozen corpus version; documented extraction protocol; scripted counts; fixed rating rules

Criterion	Meaning (Lincoln & Guba 1985; Yin 2018)	Measures in this study
Confirmability	Findings are grounded in data, not in the researcher	Evidence anchoring to file/sheet/row; counter-instance codes; separation of description (Ch. 5–9) and judgement (Ch. 10)
Internal validity (pattern matching)	Inferences are warranted	Predicted patterns from the sub-principles matched against observed mechanisms; rival explanation considered (mechanism as convenience vs principle)

One rival explanation deserves explicit mention: that the mechanisms of HLC are pragmatic conveniences of spreadsheet maintenance rather than implementations of a principle. The genealogy (Section 4.1) shows that several mechanisms arose before the principle was stated, which supports the rival partially. The study accommodates this by treating the principle as inductively derived from practice; the claim is not that the mechanisms were designed from the principle, but that they jointly satisfy it and that the principle explains why they are needed together.

3.9 Methodological Limitations

- **Single case.** Generalisation is analytical, not statistical. The reference model in Chapter 11 is a proposition for further testing, not an established finding.
- **Insider analyst.** Measures against bias are described in Section 3.2; they reduce but cannot eliminate the risk. Independent replication of the rating by a second analyst would strengthen the results.
- **No user study.** The study examines the artefact, not its use. Claims about usability (Section 10.3.2) are analytical.
- **Snapshot.** The corpus is frozen at version 8.1/8.1.2; later versions may change the ratings, particularly for the integrity and governance layers.
- **Indicator scope.** Counts depend on the pattern definitions in Section 3.4; for example, a claim annotation phrased without the word "claimed" is not counted. Indicators are therefore lower bounds of mechanism use.

4 Evolution of the Principle: From Maxim to Coordinated Plurality

This chapter answers RQ1. It first traces how the principle emerged historically within HLC, then defines the concepts needed to state it precisely, builds a formal model, decomposes the principle into ten sub-principles, analyses their interdependencies and finally states the evolved principle in a single formulation.

4.1 Genealogy: The Principle Emerged from Practice

The change history recorded in the front matter shows that HLC did not begin with a principle. It began with an inventory of codes and arrived at an explicit principle only after five years of accumulation. Four phases can be distinguished (Figure 5).

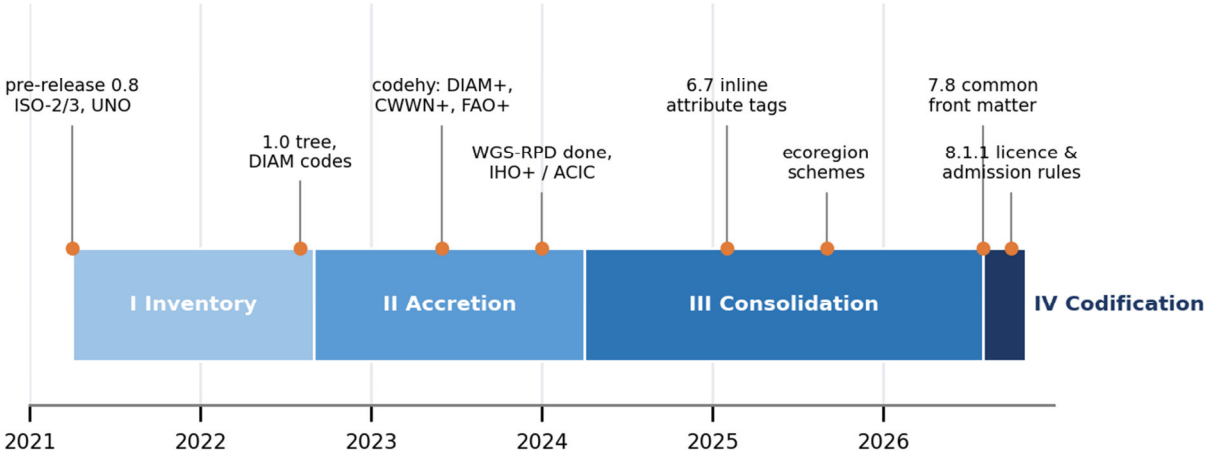

Figure 5: Development phases and selected milestones of HLC, 2021–2026 (from the codehv change history)

Table 8: Phases in the development of HLC as documented in the change history

Phase	Period	Documented milestones	Interpretation
I Inventory	04/2021 – 08/2022	Pre-release 0.8 with ISO-2, ISO-3 and UNO; 1.0 with a hierarchical tree and first DIAM codes; national classifications CH, CA, EU, FR, AR added	A crosswalk of country codes; a single tree is introduced as organising device
II Accretion	09/2022 – 03/2024	AU, ITU+, ESSD, WID-TIC, type table, FAO, classification chart, NZ, ZA, JP, NO; four-digit subcodes (5.0); WGS-RPD completed; COW integration started; codehy released with DIAM+, CWWN+, FAO+; IHO+ and ACIC combined; WGS84 anchors	Sources are added one by one; each addition forces new structural devices (subcodes, types, water file)
III Consolidation	04/2024 – 07/2026	CIA-FACTS; history codes; COW completed; inline attribute tags (6.7, 02/2025); IT+, ISTAT, MIN, ILO; codehu released (3.5 in 2023, extended 7.x); NAFO, NEAFC, GFCM; ecoregion schemes BEOW, GIWA, MEOW, FEOW, TEOW, PAME, LMEs, PPOW, GOODS, GEZ, EUNIS; tectonic plates; WAC	The growing number of sources requires compaction (IATs), namespace control (codehu) and a separate ecological domain (codehw)
IV Codification	08/2026 – 10/2026	Split-off of the common front matter (7.8); SDMX+ and types sheet moved to codehu (7.9); BYES, EOW/SEOW, GET, HLC-D derivation; IAT documentation (8.1); licence and admission rules	The implicit rules of practice are written down as philosophy, notation and governance

Two observations follow. First, the principle is *inductive*: it formulates what had proved necessary while integrating real sources. This gives it a practical credibility that a principle designed in advance would lack, but it also means that some of its parts are stated only implicitly in the data. Second, each phase transition was triggered by scale. A crosswalk with three sources can live in a single table; one with forty sources requires compaction, namespace governance and federation. The principle is therefore best understood as the set of conditions under which a crosswalk can grow without collapsing.

4.2 Basic Concepts

The following definitions fix the vocabulary used in the rest of the thesis. They are formulated so that they apply to HLC but are not specific to it.

- **Referent.** A portion of the world that someone wishes to identify: a territory, a water body, an ecosystem type, an organisation's membership area, a historical polity.
- **Scheme.** An independently maintained classification or code list with its own mandate, for example ISO 3166-1, UN M49, IHO S-23 (1953), FEOW or the Swiss list of states and territories.
- **Code.** A string assigned by a scheme to a referent. A code is meaningful only together with its scheme.
- **Entry.** A record in the coordinating layer that represents one referent and carries the coordinating layer's own identifier.
- **Placement.** The position of an entry in a hierarchy of the coordinating layer, expressed as a path of identifiers.
- **Assertion.** A statement that a scheme assigns a certain code, parent, name or status to a referent. Placements and crosswalk values are assertions.
- **Authority.** The party whose mandate gives an assertion its validity within a domain.
- **Coordinating layer.** The set of identifiers, placements, correspondences and rules that connects schemes without replacing them.

The central conceptual move is the reinterpretation of every placement and every code as an *assertion* attributable to an authority. Once this move is made, conflict between schemes is no longer contradiction within the framework. Two assertions with different authorities can both be true, in the sense of being correctly recorded, even if they cannot both be adopted in one application.

4.3 Formal Model

The model below is a reconstruction. It uses the minimum of formalism needed to state the sub-principles in a testable way.

> *Let S be a finite set of schemes. Each scheme $s \in S$ has a code set $C(s)$, a (possibly partial) broader relation $B(s) \subseteq C(s) \times C(s)$, and a mandate $\mu(s)$ describing the domain in which it is authoritative.*
>
> *Let E be the set of entries of the coordinating layer. Each entry e has an identifier $m(e)$ from an identifier space M, a type $\tau(e)$ from a type system T, and a non-empty set of labels $L(e)$, one of which is preferred.*

> *A hierarchy H is a set of paths over identifiers. The placement function π assigns to each entry a non-empty set of paths π(e) □ H. Exactly one path $\pi_0(e)$ is primary; all others are marked as secondary.*
>
> *The crosswalk is a relation X □ E × S × C × Q, where (e, s, c, q) states that scheme s assigns code c to the referent of e with qualifier q (for example: direct, inherited, former, approximate).*
>
> *The namespace is a function ν that assigns to each string w in a code layer ℓ exactly one owner ν(w, ℓ) ∈ S ∪ {coordinating layer} and one code class.*
>
> *The attribute function α assigns descriptive values (names, notes, coordinates, crosswalk entries) to entries. The redirect function ρ maps a deprecated entry to the set of entries that succeed it.*
>
> *An admission predicate Adm decides whether an alternative hierarchy h may be added to H.*

With these elements, a coordinating framework is the tuple F = (S, E, M, T, H, π, X, ν, α, ρ, Adm). The sub-principles are constraints on this tuple.

4.4 The Ten Constituent Sub-Principles

Each sub-principle is presented with a statement, its rationale and disciplinary grounding, a formal expression, and the failure mode that it prevents.

4.4.1 P1 Non-Replacement

P1 Non-Replacement
External schemes remain authoritative within their mandates. The coordinating layer never alters, reinterprets or reuses an external code, and introduces its own codes only where no external scheme provides one.

Rationale. Replacement fails because legacy schemes are embedded in law and data (Section 1.2), and because the integrator lacks the mandate of the source authorities (Section 2.6). Non-replacement is the precondition for the coordinating layer to be accepted by the communities whose schemes it hosts.

> *For all s ≠ coordinating layer and all (e, s, c, q) □ X: c □ C(s) as published by s. An own code is minted for e in layer ℓ only if no external scheme designated for ℓ provides a code for e.*

Failure mode. A *shadow standard*: the coordinating layer silently "corrects" external codes or introduces look-alike codes, so that users can no longer tell which code is authoritative.

4.4.2 P2 Conflict as Information

P2 Conflict as Information
Where schemes disagree about parentage, extent, status, name or code, the disagreement is recorded, not resolved. Every recorded position remains attributable to its source.

Rationale. Under pluralism (Section 2.1), disagreement reflects different criteria rather than error. Resolving it would require a meta-criterion that the coordinating layer cannot justify. Moreover, the

pattern of disagreement is itself valuable knowledge, for example for statisticians reconciling two regional aggregates.

> *If s and t assign different parents to the referent of e, then either both parent relations appear in π(e), or both are recoverable from X. No rule of the framework selects one of them as true.*

Failure mode. Silent arbitration: the integrator chooses a winner without documenting the losing position.

4.4.3 P3 Legitimate Polyhierarchy

P3 Legitimate Polyhierarchy
An entry may appear in several hierarchies of the coordinating layer. Secondary appearances are marked as such and represent legitimate alternative hierarchies, not data redundancy.

Rationale. Many referents genuinely belong to several wholes: a sea to two oceans, a country to two continents, an island group to a state and to a biogeographic province. Knowledge organisation standards admit polyhierarchy for this reason (Section 2.3). Marking secondary appearances distinguishes intentional polyhierarchy from accidental duplication.

> $|\pi(e)| \geq 1$; *for every* $p \square \pi(e) \setminus \{\pi_0(e)\}$ *the path p is flagged as secondary.*

Failure mode. Either *forced monohierarchy*, which loses valid relations, or *unmarked duplication*, which makes counting, aggregation and deduplication unreliable.

4.4.4 P4 Attribute Locality

P4 Attribute Locality
Descriptive attributes of an entry reside only at its primary placement. Secondary placements carry position, not description.

Rationale. Polyhierarchy introduces the risk that copies of an entry drift apart. Locality guarantees a single source of truth for each entry, which is the integrity counterpart of plurality: the framework is plural in its relations but singular in its descriptions.

> $\alpha(e)$ *is defined with respect to* $\pi_0(e)$ *only. For* $p \neq \pi_0(e)$*, the record at p contains m(e), the marker and the path, and nothing that could contradict* $\alpha(e)$*.*

Failure mode. Contradictory duplicates: two copies of the same entry state different names, codes or coordinates.

4.4.5 P5 Namespace Sovereignty

P5 Namespace Sovereignty
Every string in every shared code layer has exactly one owner and one code class. Ownership is registered centrally before a string is used, and the coordinating layer's own codes are confined to ranges that do not collide with external codes.

Rationale. A framework that hosts many schemes hosts all their strings at once. Two-letter and two-digit codes are scarce, and the same string often means different things in different schemes. Without

central ownership control, the coordinating layer would itself become a source of ambiguity (Section 2.7).

> *ν is a function: for each (w, ℓ) there is exactly one owner. Ranges of M allocated to the coordinating layer are disjoint from ranges allocated to imported codes.*

Failure mode. Code collision: one string denotes two referents in the same layer, or an own code later collides with a newly published external code.

4.4.6 P6 Identity Persistence

P6 Identity Persistence
Identifiers of the coordinating layer remain stable when names, attributes, external codes or placements change. Deprecated or historical entities are retained and linked to their successors by redirects, not deleted.

Rationale. Longitudinal analysis requires that an identifier mean the same thing over time (Sections 2.6, 2.7). Succession relations (partition, merger, renaming) are not containment, and must be expressed by a separate relation.

> *$m(e)$ does not change when $\alpha(e)$, X or $\pi(e)$ change. For each deprecated entry e, $\rho(e)$ is defined and non-empty, or e is explicitly marked as having no successor.*

Failure mode. Identifier drift: historical data become uninterpretable because their codes were reassigned or deleted.

4.4.7 P7 Conceptual Assignment

P7 Conceptual Assignment
Membership of an entry in a parent is a conceptual assignment by best fit, not a result of geometric containment. Coordinates, where present, are orientation anchors, not boundaries.

Rationale. Most units are fiat objects whose boundaries differ between authorities (Section 2.4). Encoding geometry would require choosing one authority's boundary, which violates P1 and P2. Separating conceptual membership from geometry keeps the framework neutral with respect to boundary disputes, and keeps it maintainable by a small curatorial team.

> *For a placement e under parent f, no claim is made that the extent of e lies within the extent of f in any particular boundary dataset. Coordinates in $\alpha(e)$ carry no topological semantics.*

Failure mode. Spurious precision: users treat a hierarchy as a spatial containment model and compute areas or point-in-polygon results that the framework cannot support.

4.4.8 P8 Domain-Appropriate Logics (Federation)

P8 Domain-Appropriate Logics
Different domains (land, water, ecosystems) are organised according to the classification logic that is appropriate to each. Domains are coordinated through shared identifiers and tags, not merged into a single hierarchy.

Rationale. Political hierarchies, hydrographic hierarchies and ecological hierarchies follow different principles of division (Sections 2.4, 2.5). A single tree would force one logic onto the others. Federation preserves each logic while allowing navigation across domains.

> *H is partitioned into domain hierarchies H(1) … H(k) with domain-specific type systems. Cross-domain links are expressed by shared identifiers m(e) or by entries in X, never by forcing an entry of H(i) into the type system of H(j).*

Failure mode. Category mistakes: ecoregions as subdivisions of countries, or seas as parts of continents.

4.4.9 P9 Bounded Plurality

P9 Bounded Plurality
An alternative hierarchy or classification is admitted only if it is defined in an identifiable published source, used by an identifiable community, organisation or discipline, and not derivable from a hierarchy already included. Each admission is recorded with a reason.

Rationale. Pluralism without bounds becomes relativism and, practically, an unmaintainable list (Section 2.1). The three criteria correspond to three tests: *existence* (is it published?), *relevance* (does anyone use it?), and *parsimony* (does it add information?).

> $Adm(h) \Leftrightarrow published(h) \wedge community(h) \wedge \neg derivable(h, H)$. *Every admitted h carries a recorded reason.*

Failure mode. Unbounded proliferation or, conversely, *curatorial arbitrariness* if admission is decided without stated criteria.

4.4.10 P10 Declared Epistemic Limits

P10 Declared Epistemic Limits
The framework states openly what it is not: not an authority, not a boundary source, not real-time, uneven in granularity, and dependent on curatorial judgement. These declarations are part of the principle, because they define the conditions under which its outputs may be used.

Rationale. A coordinating layer that does not disclaim authority implicitly claims it, which would contradict P1. Declaring temporal status (snapshot), granularity and judgement allows users to calibrate trust, which is the standard expectation for any reference dataset.

> *F is accompanied by a statement of scope that includes at least: authority status, spatial semantics, temporal status, granularity policy, curatorial model, and licence.*

Failure mode. Over-trust: the framework is used for legal, cadastral or delimitation purposes for which it is unsuitable.

4.5 Interdependencies and Layering

The ten sub-principles are not independent rules. Several presuppose others, and some exist only to neutralise risks created by others. The following table records the dependencies; Figure 6 arranges them in layers.

Table 9: Dependencies between the sub-principles

Sub-principle	Requires	Neutralises risk created by	Layer
P1 Non-Replacement	P5, P10	–	Foundational
P2 Conflict as Information	P3 or X (crosswalk), P1	–	Representational
P3 Legitimate Polyhierarchy	P4, P5	P2 (where conflicts concern parentage)	Representational
P4 Attribute Locality	P3 (presupposes copies)	P3	Integrity
P5 Namespace Sovereignty	–	P1 (hosting many schemes)	Integrity
P6 Identity Persistence	P5	P2 (temporal conflicts)	Integrity
P7 Conceptual Assignment	P10	P1, P2 (boundary disputes)	Representational
P8 Domain-Appropriate Logics	P5 (shared identifiers)	P3 (cross-domain parentage)	Representational
P9 Bounded Plurality	P10 (judgement declared)	P2, P3 (proliferation)	Governance
P10 Declared Epistemic Limits	–	P1, P7 (over-trust)	Foundational

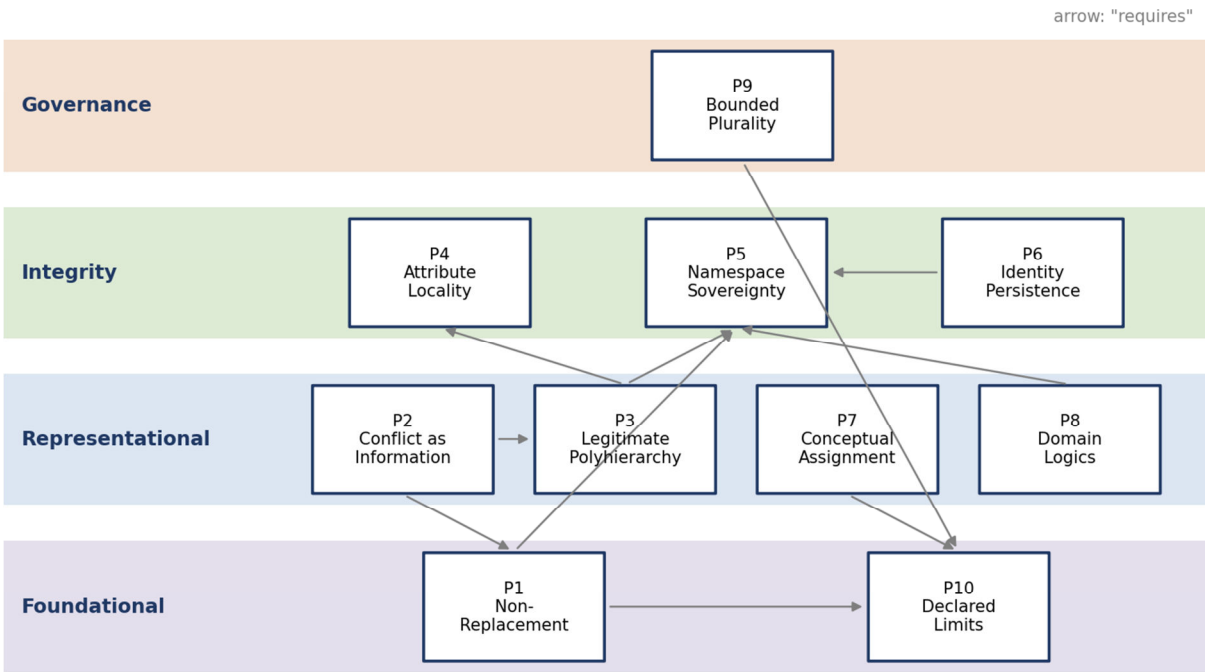

Figure 6: The ten sub-principles in four layers, with "requires" dependencies

Four layers result. The **foundational layer** (P1, P10) defines the stance of the coordinating layer towards authorities and users. The **representational layer** (P2, P3, P7, P8) determines what can be expressed. The **integrity layer** (P4, P5, P6) ensures that what is expressed stays consistent over copies, strings and time. The **governance layer** (P9) limits growth. A framework that implements only the representational layer would be expressive but unstable; one that implements only the integrity layer would be stable but would not be plural. P5 is the most depended-upon sub-principle: four others require it, which explains why namespace control became a separate file (codehu) once the framework grew.

4.6 Neutrality Reconsidered: Attributed Neutrality and Placement Modes

HLC states that territorial disputes are "recorded neutrally" with "no sovereign parent assigned". Developing this statement exposes a conceptual difficulty that any implementation of the principle

must address. A hierarchy *always* asserts something: placing an island under a state asserts a relation; placing it elsewhere asserts another. Neutrality cannot mean the absence of assertion. It can only mean that the framework does not adopt an assertion as its own.

The thesis therefore proposes the concept of **attributed neutrality**: a framework is neutral when every contested placement is either attributed to the authority that makes it, or made under a placement mode that is explicitly non-sovereign. Four placement modes can be distinguished (Figure 7), and all four occur in HLC (Section 6.4).

Table 10: Placement modes for contested referents

Mode	Definition	Assertion made by the placement
De jure	Under the state that holds the legal title recognised by a reference authority	Recognised title, not control
De facto (administering)	Under the state that currently administers the area	Control, not title
Neutral container	Under a non-sovereign parent such as a sea, a region or a special group for contested areas	Location only
Multi-placement	Under several parents, with one primary and the others marked	Plurality of claims

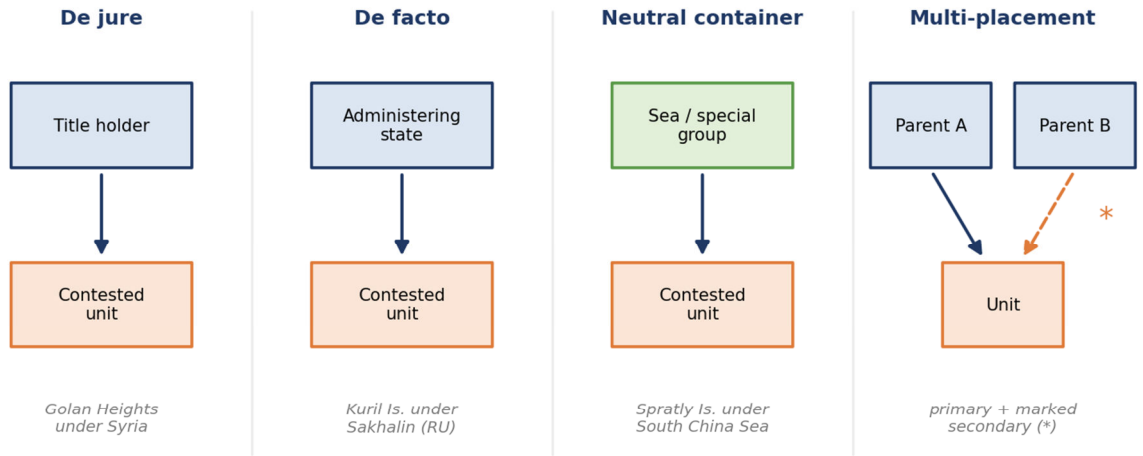

Figure 7: Four placement modes for contested referents, with examples from codehl

The evolved principle requires that the placement mode be *declared* for each contested entry. This is a refinement of P2 and P10 rather than a new sub-principle: conflict is information only if the reader can tell which position a placement represents.

4.7 Distinction from Neighbouring Approaches

Table 11: Coordinated plurality compared with neighbouring approaches

Approach	Treatment of plurality	Identifier policy	Difference from coordinated plurality
Master data management	Reduces sources to a golden record	Golden-record IDs	Resolves conflicts; HLC records them (P2)
Gazetteers (e.g. GeoNames)	One feature hierarchy, alternate names	Feature IDs	Single administrative hierarchy; limited multi-scheme parentage
Single reference ontology	Sources mapped to one conceptual model	Ontology IRIs	Imposes one logic across domains; HLC federates (P8)
SKOS concept-scheme mapping	Schemes kept separate; typed matches	One IRI per scheme concept	Close in spirit; lacks an own coordinating identifier and namespace control (P5, P6)

Approach	Treatment of plurality	Identifier policy	Difference from coordinated plurality
Faceted classification	Independent facets	Facet notations	Facets but no polyhierarchy within a facet; HLC combines both

4.8 The Evolved Principle

Principle of Coordinated Plurality
A coordinating framework for classifications shall (1) leave every source scheme authoritative and unaltered within its mandate; (2) record disagreements between schemes as attributed assertions, including the placement mode of every contested entry; (3) allow entries to occupy several marked positions while keeping their description in one place; (4) control every shared string and identifier through a single register of ownership; (5) keep its own identifiers persistent and express succession by redirect; (6) assign membership conceptually rather than geometrically; (7) organise each domain by its own logic and connect domains through shared identifiers; (8) admit alternatives only by published, used and non-derivable criteria; and (9) declare openly the limits of its authority, precision and currency.

Compared with the original maxim, the evolved principle (a) separates the stance towards sources from the integrity mechanisms that make the stance sustainable, (b) adds the requirement to declare placement modes, (c) admits decomposition into whole and parts and parallel editions as legitimate realisations of plurality (both found inductively, Section 3.5.3), and (d) makes the bounding of plurality and the declaration of limits part of the principle rather than appendices to it.

5 The Architecture of HLC as Implementation of the Principle

This chapter describes the framework-wide structures through which HLC implements the principle. Domain-specific mechanisms follow in Chapters 6 to 8.

5.1 Federated File Architecture

HLC consists of one shared document and four data files (Figure 8). The front matter states that "the files are coordinated but not merged" and that each "uses classification logic appropriate to its domain". This is a direct implementation of P8.

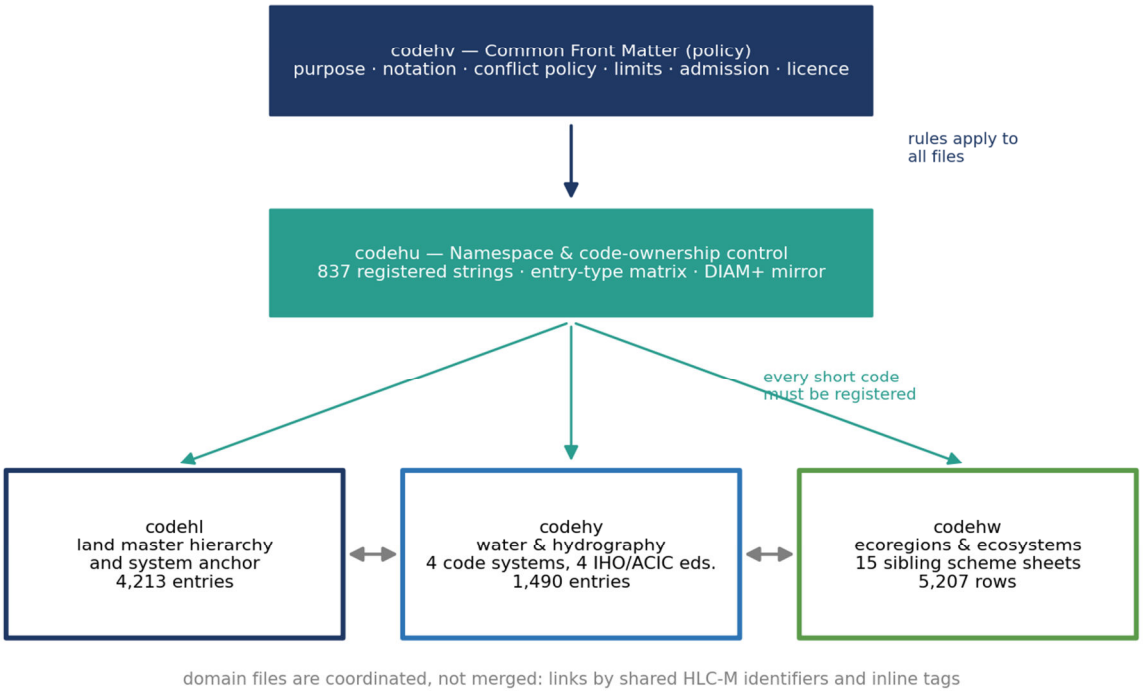

Figure 8: Federated architecture of HLC: policy, namespace and content layers (entry counts from v8.1)

Table 12: Distribution of responsibilities across the HLC files

File	Primary sub-principles served	Mechanism
codehv	P1, P2, P9, P10	States purpose, notation, conflict policy, coordinate policy, limitations, admission rules and licence once for all files
codehu	P5, P1	Registers ownership of every two-character code and of three-character HLC codes; defines entry types; mirrors DIAM+ ownership
codehl	P3, P6, P2, P8 (anchor)	Holds the land master hierarchy, historical entities, the HLC-M range allocation and the broadest crosswalk
codehy	P2, P3, P8	Holds water hierarchies with four parallel code systems and four IHO/ACIC limit editions
codehw	P8, P1, P5	Holds ecological schemes as siblings and derives HLC-M codes deterministically from source codes

The architecture separates *policy* (codehv), *namespace* (codehu) and *content* (codehl, codehy, codehw). This separation reflects the layering of Section 4.5: foundational and governance rules are stated once, integrity rules are centralised, representational content is federated. The decision taken in

version 7.8 to split the front matter out of the data files, and in 7.9 to move the type table to codehu, completed this separation.

codehl is designated as "master integrator that anchors the full system". This does not contradict federation. The anchor role is navigational: codehl contains the top-level entries of all domains (geographical zones, biogeographic realms, marine ecosystems, oceans, continents) so that every domain hierarchy can be reached from one root, while the detailed logic of each domain remains in its own file. In the formal model, codehl provides the shared top of H without imposing its type system on the domain hierarchies.

5.2 The Namespace Control Layer (codehu)

codehu implements P5 by assigning an owner to every two-character string from AA to ZZ and to every three-character HLC string; its DATA sheet holds 837 registered entries. Its INFO sheet lists the admissible code classes for each layer.

Table 13: Code classes registered in codehu for the two-character layer (HLC-2)

Pattern	Owner	Meaning	Example from codehu
aa	ISO 3166-1	Active country code (no ISO reservations)	DE Germany; JP Japan
aa	HLC	Pseudo area, region or country	XU Europe; YJ Western Asia; DC Dead Countries
an	HLC	Pseudo (sub-)region	Z0–Z9 climate and ecological zones; R1–R8 biogeographic realms
aa-aa	ISO 3166-2	Subdivision, separated by hyphen	AE-DU Dubai; MY-12 Sabah
axxx	HLC	Dependency of another country with internal number	AS|US60; AW|NLAW
ax:	HLC	Logical / organisational / source-scheme entry	EU:4Z European Union; MW:0M MEOW; DE: Destatis
nx	DIAM+ (mirrored)	Water regions	7B Baltic Sea; 6U Suez Canal

Three design decisions stand out. First, ISO codes are admitted only for *active* countries, so that ISO reservations and transitional codes cannot be confused with HLC pseudo-codes. Second, the colon marks a separate namespace for logical and organisational entries (134 entries): `EU:4Z` (the European Union as organisation) is distinct from any territorial use of EU. Third, DIAM+ water codes of the form digit-plus-character are not owned by HLC but mirrored on a separate COPY sheet, which states that DIAM+ "bears the codes ownership". The framework thus records not only its own namespace but also which parts of the namespace it does *not* own.

The three-character layer (HLC-3) follows the same logic: ISO alpha-3 for active countries, HLC three-character codes depending on HLC-2, `U`+nx codes for organisations derived from SDMX+, `X`+nx codes for water regions derived from DIAM+, and GENC codes of the form `X`+aa for geopolitical entities defined by the US government.

5.3 The Entry-Type Matrix

The TYPES sheet in codehu defines a two-digit type system that combines a hierarchical level with a semantic variant. It implements P8 at the level of individual entries: each entry declares which kind of

unit it is, so that a geographical region, an economic region and a botanical region at the same level are not confused.

Table 14: Structure of the HLC entry-type matrix (codehu, TYPES)

Code band	Earth-related type	Alternative type	Level	Variants
00	X unspecified / contested / special	N NUL entry	–	–
08 / 09	T tectonic; F functional	L logical	–	–
10	W waters	–	–	–
20–22	G global	I international (organisation)	0	single / multiple oceans or continents
30–36	R region / area	E eco-region / system	1	single, multiple, geographical, historical, economic, botanical
40–46	S subregion	–	2	as level 1
50–57	M multiples	–	intermediate	political, geographical, historical, economic, botanical, administrative areas of several states
60–64	C country	O organisation	3	single, historical
70–76	D division	V subdivision	4	as level 1, plus free zones
80–89	P province	U unit	5	single, multiple, historical, city district, community
90	H historical	–	–	–

Two features matter for the principle. The type `X` explicitly hosts "unspecified/contested/special" entries, which gives contested areas a type of their own instead of forcing them into the country type; codehl contains 222 such entries. This is the type-level counterpart of the neutral-container placement mode (Section 4.6). The *Multiples* band (50–57) represents areas composed of several states, such as the Gulf States, Benelux or the Antarctic territorial claims (144 M entries in codehl). It recognises that many referents are groupings rather than subdivisions, which a strict containment hierarchy cannot represent.

5.4 The Identifier Space HLC-M

HLC-M is the main identifier of every entry: an alphanumeric string of three or four characters beginning with a digit. The HLC-M sheet in codehl partitions this space among the files and among continents. Figure 9 visualises the numeric part of the allocation; Appendix B reproduces it in full. Its structure shows how P5 is applied to the coordinating layer's own identifiers.

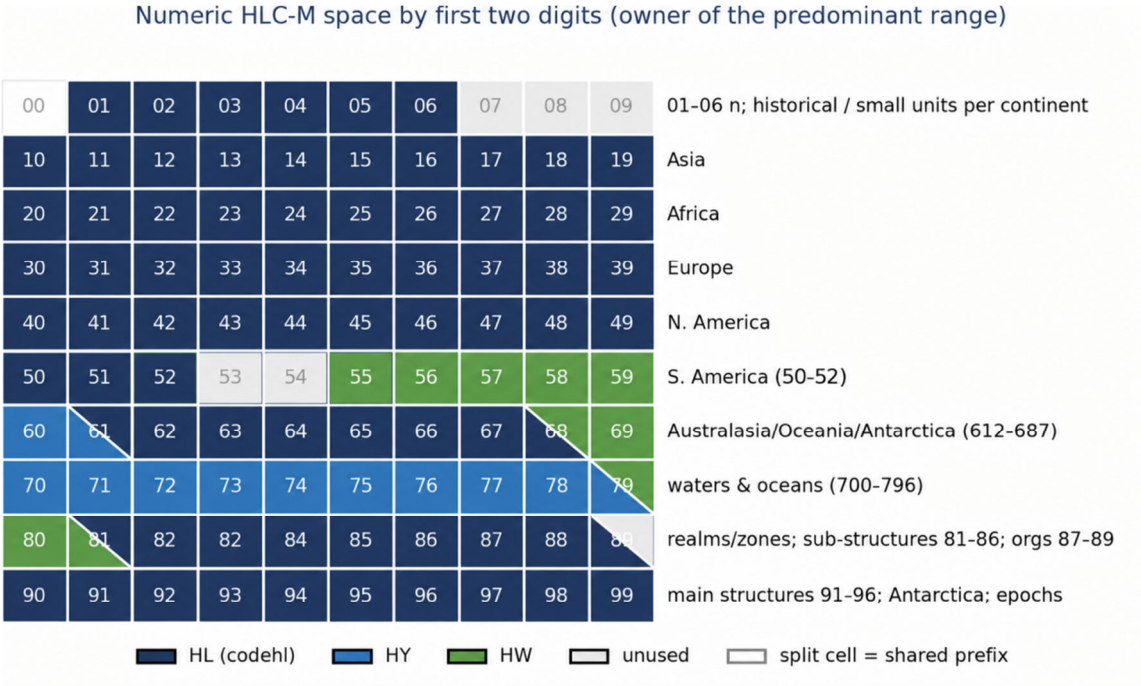

Figure 9: Ownership of the numeric HLC-M space by first two digits (codehl, HLC-M sheet)

- **Ownership by file.** Every range is assigned to HL, HY or HW. For example, `700–796` (numeric) belongs to HY (waters and oceans), `700–796` followed by a letter to HW (water-based ecosystems), and `811–819` to HL (sub-structures of Asia).
- **Ownership by continent.** Land ranges are allocated by continent: `1nn` Asia, `2nn` Africa, `3nn` Europe, `4nn` Northern America, `50–52`n Southern America, `612–687` Australasia, Oceania and Antarctica. The same continental digits recur in the sub-structure ranges `81–86` and the main-structure ranges `91–96`, so that the identifier carries a mnemonic of its continent.
- **Ownership by character pattern.** Within the same leading digit, the character class of the following positions separates owners: ‘0’–‘9’anx is reserved for HW leaf codes, ‘0’–‘9’aa for CWWN+-derived water codes in HY, and numeric patterns for HL.
- **Reserved and external ranges.** Ranges `07–09`n, `53–54`n and `891–896` are declared "currently unused". A dedicated rule admits external numeric codes (for example from GeoNames) by replacing the penultimate digit with a letter from P to Y, "to keep the value areas separate and clearly distinguish".

The last rule is a precise implementation of P1 and P5 together: external codes are admitted unchanged in substance, but transformed in a reversible way so that they cannot collide with HLC-owned codes.

Two further notational devices refine the identifier space. The range marker `!-` extends an identifier over a block (for example `110!-27` India, `189!-94` China, `747!-50` Baltic Seas), reserving sub-identifiers for the entry’s internal structure. The markers `~`, `+` and `-` record that a number is not exclusively used by its holder but shared with the preceding or following entry (`153~` Brunei, `182+` South Korea, `184-` North Korea). The optional `(0)` records that a trailing zero is omissible (`807(0)` Oceanic Realm). These devices make the allocation itself transparent: a reader can see where an identifier has neighbours that compete for the same numeric space.

5.5 Tree Paths

Every data row begins with a tree path that concatenates HLC-M identifiers from the root to the entry, for example `910-811-911-918-814-169-165-1652` for Dubai (Asia – Temperate Asia – West and Central Asia – Western Asia – Arabian Peninsula – Gulf States – United Arab Emirates – Dubai). In codehl, paths reach a depth of up to 13 levels. The path is redundant with the parent relation but has two functions. It makes every placement self-describing, so that a starred duplicate can be read without consulting its original. It also makes the sort order of the spreadsheet identical to a depth-first traversal of the tree, which is a pragmatic way to keep a hierarchy maintainable in a tabular medium.

In codehy and codehw, a second hierarchical number appears in parentheses after the path (for example `(0.3.14)` for the Subantarctic Climate Zone in GIWA, or the FAO+ number `29.1 (23.13)` for a duplicated South China Sea branch). The parenthesised number records the original position of a duplicate, which links each secondary placement back to its primary placement.

5.6 The Notation as an Assertion Grammar

The shared notation of the front matter is compact, but it is not a set of typographical conventions. It is a small grammar for assertions. The following table relates each symbol to the assertion it encodes and to the sub-principle it serves. Appendix A gives a formal grammar.

Table 15: The HLC notation as an assertion grammar

Symbol	Assertion encoded	Sub-principle	Example
|	Equivalence of alternative codes or names for the same referent	P2, P1	`AF|A1`; `GBR|826`
>	Deprecated entry; the successor is the target	P6	`DD>DE` German Democratic Republic; `EL>(GR)`
:	Logical, institutional or organisational entry (separate namespace)	P5, P8	`EU:4Z`; `DE:`
*	Secondary placement in a legitimate alternative hierarchy	P3, P4	`722*` South China Sea under Maritime South-East Asia
~ + -	Shared numeric space with predecessor or successor entry	P5	`153~`; `241+-`
!-	Number-range extension reserved for the entry	P5, P6	`420!-28` USA
( )	Optional extension, annotation, or original/former value	P1, P2	`807(0)`; `2TP(3FA)`
!	After a country code in a claim: administering party	P2	`claimed by EG!,SD`
{k:v}	Inline attribute tag: scheme k assigns value v	P1, P2	`{UN:364}`; `{IE:IIIa}`

Read together, the symbols allow a single cell to express equivalence (`|`), succession (>), namespace (`:`), alternative placement (`*`), allocation (`~ + - !-`), provenance of former values (`( )`), contested administration (`!`) and external assertion (`{k:v}`). This density is what allows a spreadsheet to carry the relations that a graph database would represent as typed edges.

5.7 Inline Attribute Tags as an Embedded Crosswalk

Inline attribute tags (IATs), introduced in version 6.7 (02/2025), store external assertions in the form `{scheme:value}` inside thematically grouped columns. The INFO sheet of codehl documents the

key registry; codehy and codehw document their own keys. IATs are the most direct implementation of P1 and P2 in the framework: each tag is an assertion attributed to a named scheme, stored next to the entry it describes and kept in the scheme's own code. codehl alone contains 9,081 tag instances from 43 keys, and codehy a further 656 from 11 keys (Figure 4).

The tag columns of codehl are organised by theme: UNO (M01: M49, ILO, ICAO), TERRA (M02: Correlates of War, WGS-RPD, SDMX+, BEOW, TEOW), AQUA (M03: FAO+, FEOW, MEOW, GOODS, CWWN+), M04 (US TIC, WAC, World Inequality Database), M05 (ITU, ESSD), M06 (NATO, EUSS), M07 (SACC, NZSCC), M08 (Destatis, ISTAT, UK), M09 (Argentina, South Africa, Japan), M10 (Switzerland, France, Norway) and OTHERS (CLDR, GENC, OKSM, opendatasoft, historical codes, WIPO, ICAO, WGS84 anchors). This is a faceted arrangement in Ranganathan's sense: each column groups schemes with a common kind of mandate.

The row for Iran illustrates the density of the crosswalk (Figure 10). In one row, codehl records nineteen external assertions, each in its own notation, attached to one HLC entry (HLC-M `166`, HLC-2 `IR`, HLC-3 `IRN`). Note that the telecommunications tag is keyed `R3`, encoding that Iran belongs to ITU Region 3, whereas its neighbour Iraq carries an `R1` key: the tag key itself carries information.

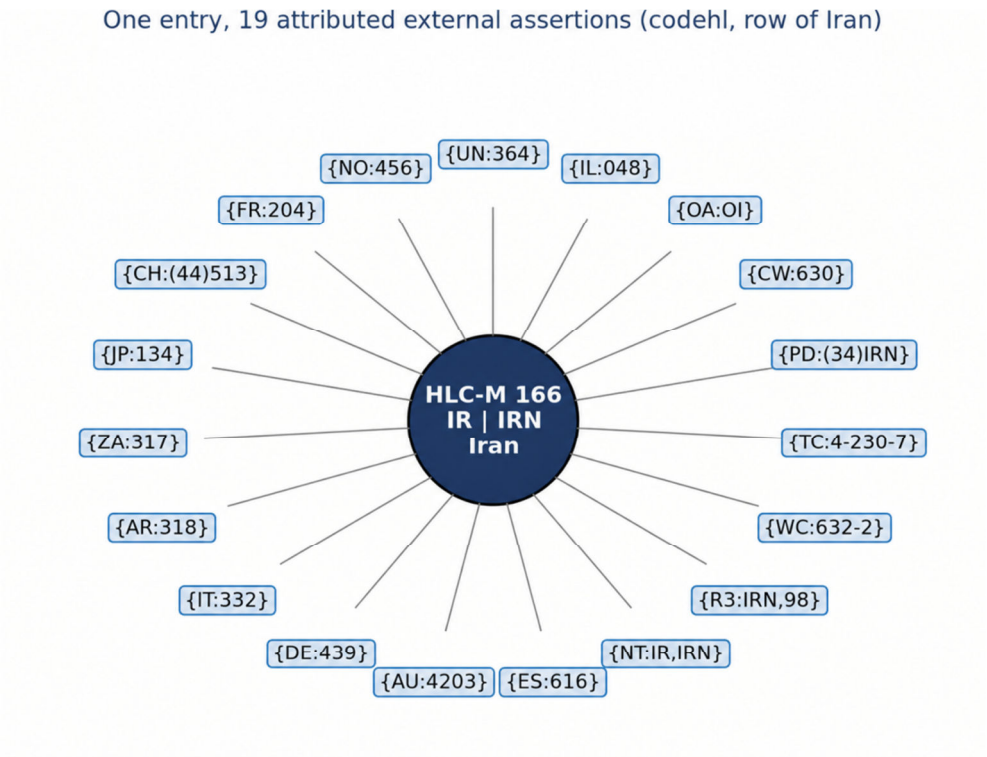

Figure 10: The embedded crosswalk: nineteen attributed assertions on the codehl entry for Iran

Parentheses inside tag values carry a qualifier. `{UN:(145)}` on the Sinai Peninsula records an inherited M49 region (Western Asia) for a unit that M49 itself does not code; `{UN:(536)}` on the Iraq–Saudi neutral zone records a former M49 code; `{CH:(44)513}` records the regional group of the Swiss classification in parentheses before the country number. In the terms of the formal model, the parentheses implement the qualifier q of the crosswalk relation X.

5.8 Governance Artefacts

Version 8.1.2 adds explicit governance: a licence (CC BY 4.0 for the published PDF and TXT renditions, covering only HLC's own contributions, while external codes remain subject to their

owners' terms), a feedback channel requiring file, code and proposed change, and the admission rule for alternative hierarchies. The licence clause separating HLC's contributions from external codes is itself an implementation of P1 at the legal level: the framework does not claim rights over what it does not own.

6 Implementation in the Land Domain (codehl)

codehl is the largest hierarchical file (4,213 entries) and the one in which political, historical and statistical plurality is most visible. This chapter analyses how its data structure implements the sub-principles in the land domain.

6.1 Column Architecture

The DATA sheet of codehl has a fixed sequence of eighteen columns: tree path, HLC-M, type, HLC-2, HLC-3, UNO|M01, HLC-X|S, description, TERRA|M02, AQUA|M03, M04 to M10, and OTHERS. The order is significant. Identity (tree, HLC-M), kind (type) and the controlled code layers (HLC-2, HLC-3) come first; external statistical identity (M49, ILO, ICAO) follows; the human-readable description is in the middle; external classifications follow as tag columns. The layout thus separates *who the entry is* from *what others say about it*, which is the tabular expression of the distinction between entries and assertions (Section 4.2).

6.2 Layering of Partitions before Politics

The tree of codehl does not start with continents. It starts with partitions that are independent of political geography: Geographical Zones (`800`, with the climate zones Z0 to Z4 and their subzones), Global Ecological Zones (`810`, Z5 to Z9), Biogeographic Realms and Habitats (`898`, realms R1 to R8 with TEOW, BEOW and FEOW subdivisions as tags), and Marine Ecosystems (`899`, depth zones from epipelagic to hadal with GOODS, PPOW, hydrothermal and hadal provinces). Only then follow the oceans (`900` to `908`), offshore installations in international waters (`909`) and the continents (`910` onwards).

This ordering implements P8 within the anchor file. The physical and ecological partitions are given the same rank as the political ones, rather than being subordinated to them. It also implements P3: the same latitudinal zone appears under Geographical Zones and again under Global Ecological Zones, because the two zonal systems group the same bands differently (the FAO polar zone, for example, contains both the Arctic and the Antarctic climate zones).

6.3 Statistical and National Plurality

For countries, codehl holds up to twenty external assertions per row (Section 5.7). The schemes differ systematically in three respects, all of which the tag columns preserve.

- **Continental assignment.** Several national classifications prefix the country number with a continental group (for example `{CH:(44)513}` for Iran, `{CH:(21)304}` for Algeria). These groups do not coincide with M49 regions, and the tags keep the national grouping visible.
- **Treatment of dependencies.** Australian and New Zealand classifications (SACC, NZSCC) and EU Geonomenclature code dependencies and special territories differently from ISO. Exceptions are recorded in OTHERS, for example `{AU:UJ,UJJ,922}` alongside the Asian continent entry.
- **Historical codes.** Former codes are kept with their scheme: `{OD:Y1,YE1,886}` for North Yemen and `{OD:Y2,YE2,720}` for South Yemen from opendatasoft, `{HI:BU,BUR}` for Burma, `{HI:TP,TMP,Z21,Z38}` for Timor-Leste.

No column is privileged. M49 appears first among external identifiers because it is the statistical reference with the widest adoption, but it is not used to override any other scheme; in frequency it ranks only fourth among tag keys (Figure 4). The column UNO|M01 combines M49 with ILO and ICAO codes, which shows that the first column is organised by *institutional family* (United Nations system) rather than by priority.

6.4 Contested Territories and Placement Modes

Contested territories are the hardest test of P2. HLC uses three mechanisms: a textual claim annotation (`claimed by X,Y`, found in 44 rows of codehl), the administering-party marker `!` (`claimed by EG!,SD` for the Halaib Triangle), and the entry type X for "unspecified/contested/special" units. The placement modes identified in Section 4.6 can all be observed in the data.

Table 16: Observed placement modes for contested entries in codehl

Entry (HLC-M, HLC-2)	Annotation in HLC	Parent in codehl tree	Placement mode
Golan Heights (1759, GO)	occupied by IL, claimed by SY	Syria (South)	De jure
Kuril Islands (3301, KX)	claimed by JP	Sakhalin (RU)	De facto (administering)
Senkaku Islands (1809, DY)	claimed by CN,TW	Kyushu (JP)	De facto (administering)
Liancourt Rocks (1823, JA)	claimed by JP and KR	North Kyongsang (KR)	De facto (administering)
Paracel Islands (1053, RP)	claimed by CN,TW,VN	South China Sea (722*)	Neutral container (water parent)
Spratly Islands (1054, YX)	BN,CN,MY,PH,TW,VN	South China Sea (722*)	Neutral container (water parent)
Aksai Chin (1001, HQ)	claimed by TW,CN,IN	Asian Regions & Territories (100)	Neutral container (special group)
Halaib Triangle (2033, HL)	claimed by EG!,SD	North-East African Regions (203)	Neutral container, administering party marked
Bir Tawil (2036, TB)	unclaimed territory	North-East African Regions (203)	Neutral container (terra nullius)
Abyei Area (236, RN)	claimed by SD,SS	(United) Sudan, historical (229)	Historical common predecessor

Two implementation strategies are combined. Where a dispute concerns a small unit inside a state that administers it, HLC places the unit under the administering state and attributes the counter-claim in the annotation. Where several states claim an area that none uncontestedly administers, HLC uses a non-sovereign parent: a sea for island groups, or a special "Regions & Territories" group (type X) for land areas. The Abyei case shows a third, historically grounded variant: placement under the common predecessor state (the united Sudan, typed H), which avoids choosing between its successors.

The analysis shows that HLC implements attributed neutrality in substance: every contested placement carries its counter-assertion. What the data do not yet contain is an explicit field that declares the placement mode. It has to be inferred from the parent. This is the point at which the evolved principle (Section 4.8) goes beyond the current implementation.

6.5 Transcontinental States

States that span continents expose the limits of monohierarchy. HLC resolves them by a combination of P3 (polyhierarchy) and P6 (stable identity) that is distinctive enough to describe in detail. Figure 11 shows the pattern for Egypt.

Table 17: Representation of transcontinental states

State	Whole-state entry (codehu)	Continental part entries (codehl)	M49 assertion
Egypt	`EG|G0`, EGY, HLC-M `885`, type C, "Asia+Africa (150,250)"	Egypt (Asia) `150`, type X, HLC-X `G0A`, under Sinai Peninsula; Egypt (Africa) `250`, type X, HLC-X `G0F`, under North-Eastern Africa	`{UN:(818)}` on both parts
Türkiye	`TR|V6`, TUR, HLC-M `886`, "(Asia+Europe) (160,360)"	Türkiye (Asia) `160`, HLC-X `V6A`, under Western Asia; European part in the 3nn range	`{UN:(792)}`
Kazakhstan	`KZ|L6`, KAZ, HLC-M `884`, "(Asia+Europe) (140,340)"	Kazakhstan (Asia) `140`, HLC-X `L6A`, under Central Asia	`({UN:398})`
Russian Federation	`RU|S3`, RUS, HLC-M `883`, "(Asia+Europe) (130–132,330–332)"	Russian Federation (Asia) `130!-32`, HLC-X `S3A`, under Northern Asia; European part in the 3nn range	`{UN:(643)}`

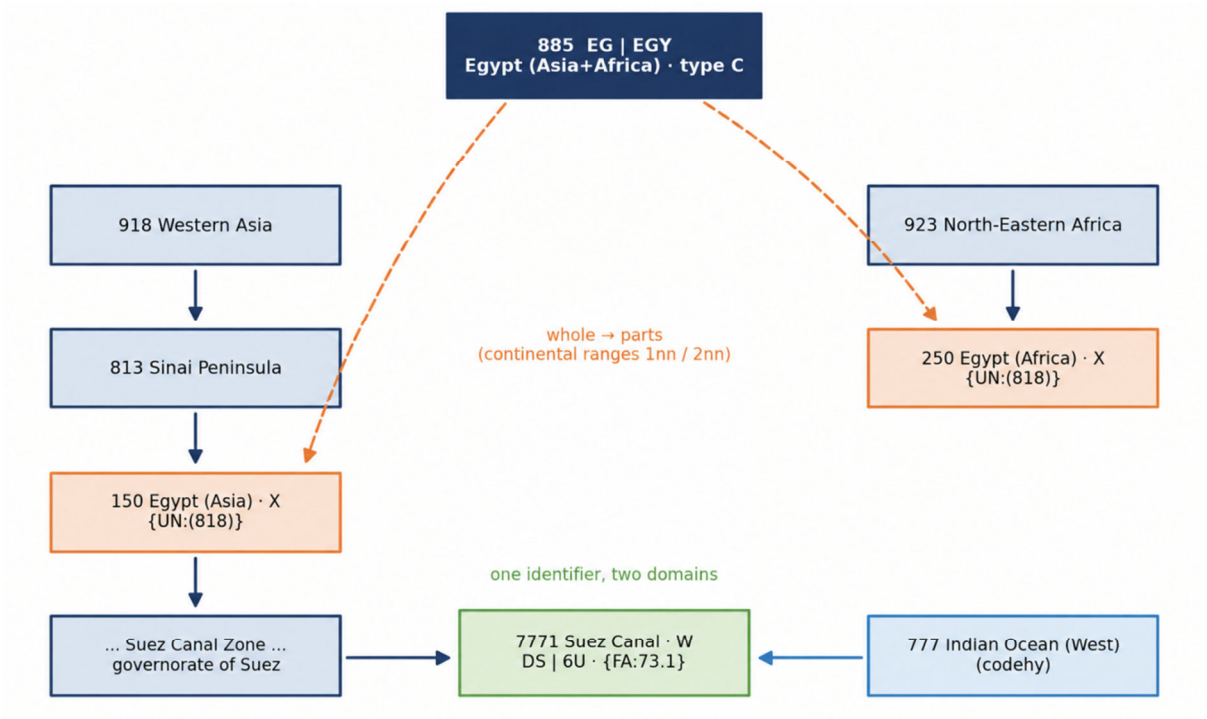

Figure 11: Whole-and-parts representation of Egypt and the cross-domain identity of the Suez Canal

The pattern is consistent. The whole state receives an identifier in the `88n` range (Unions – Cross Continent) and is placed in codehl under the continental spanning structures (`983`); its continental parts receive identifiers in the respective continental ranges (`1nn` Asia, `2nn` Africa, `3nn` Europe). The parts are typed X (special), not C (country), because they are not countries; their M49 codes are given in parentheses, marking them as inherited from the whole. Statistical tags such as `{ES:(-21-)(220)}` or `{DE:287}` are attached to the parts in the same way.

This design satisfies P3 without starred duplicates: instead of placing one entry twice, HLC creates one whole and several parts, each with its own stable identifier. It satisfies P6 because the whole-state identifier remains stable regardless of how the parts are grouped. It satisfies P8 because continents

remain clean partitions of land while the state remains one entity. It satisfies P2 because analyses that treat Egypt as African, as Middle Eastern, or as both, can all be reconstructed from the data.

6.6 Historical Entities and Succession

codehl contains 663 historical entities (type H), ranging from ancient polities (Elamites, Sabaean Kingdom, Himyarite Kingdom) through colonial and protectorate units (Aden Protectorate with its sheikhdoms and sultanates, Federated and Unfederated Malay States, the Shan States, the princely states of India) to twentieth-century entities (Republic of Viet-Nam, the Neutral Zones between Iraq, Kuwait and Saudi Arabia, occupation zones). The Correlates of War key `CW` is the most frequent tag key in the file (1,186 instances), which links these entities to quantitative datasets on interstate relations.

Three mechanisms implement P6 for these entities.

1. **Retention under present-day parents.** Historical entities are placed under the present-day unit that covers their core area, as children rather than predecessors. For example, the Aden Protectorate (`1615`) and its constituent states sit under South Yemen (`1618`), which in turn sits under Yemen (`161!-64`). This placement is conceptual (P7): it locates the historical entity in space without claiming continuity of statehood.
2. **Redirects in the namespace.** Deprecated two-letter codes point to successors: `DD>DE` (German Democratic Republic to Germany), `VD>(VN)` (Republic of Viet-Nam), `CS>(HH)` (Czechoslovakia) and `AN>(HH)` (Netherlands Antilles to AW, BQ, CW). `HH` is itself a registered HLC-2 code for "countries divided into several parts" in the ISO 3166-3 sense, so a redirect to `(HH)` asserts that the successor is a plurality.
3. **Time-qualified labels.** Descriptions carry periods where relevant: "Nagorno-Karabakh (1991–2023)", "Spanish protectorate in Morocco (1912–56)", installations with "(1999–2019)". The time span is part of the referent's identity, not an attribute that could be overwritten.

Historical epochs are also represented as independent top-level structures: `99n` "Historical World Epochs" (codehl) and the HLC-2 pseudo-codes XB, XE, XG, XH, XI, XJ, XR, XV (Ancient Greece and Byzantium, Celtic World, Gauls-Germanic-Goths, Ancient and Medieval Far East, Islamic World and Ottoman Empire, Ancient Near East, Ancient Rome, Ancient India) together with the epochs JB to JZ per continent (codehu). These entries implement P8 temporally: history is given its own hierarchy rather than being forced into the present-day political tree.

6.7 Organisations and Logical Entries

International organisations and source schemes appear in codehu as colon entries (`EU:4Z`, `NT:6B` NATO, `AE:2P` APEC, `UN:` M49 itself as `UNO|XL3`), with HLC-3 codes in the `U+nx` range derived from SDMX+ and HLC-M codes in the `87n` (world organisations) and `89n` (regional organisations) ranges. Organisational membership areas are therefore entries in their own right, typed I (international) or O (organisation) or L (logical), and never confused with territorial units. This implements P8 between the territorial and the institutional domain and P5 through the separate colon namespace.

A notable design choice is that the source schemes themselves are registered as logical entries: `MW:0M` (MEOW), `TW:0T` (TEOW), `FW:0F` (FEOW), `PD:0D` (WGS-RPD), `CW:3W` (Correlates of War), `HD:` (HLC-D derived codes), `HM:` (HLC-M). In this way, the keys used in inline attribute tags are

themselves controlled codes in the namespace. Every assertion in the crosswalk thus points to a registered authority, which closes the loop between P1 (attribution to an authority) and P5 (controlled strings).

6.8 Points, Installations and Below-Country Granularity

codehl extends below the country level to divisions (D, 885 entries), subdivisions (V, 750), provinces (P) and units (U, 527), and includes point-like entities such as offshore oil and gas terminals with codes of the form `VZ-xxx` and WGS84 anchors (for example the Norwegian fields under the Norwegian Sea and the North Sea, or terminals in the Java Sea). These entries are placed under the water body in which they lie, not under the state that licenses them. This is a consistent application of P7 and P8: the installation is conceptually located in a sea, and its national relation is given in the label, for example "Heidrun (NO)".

The granularity is explicitly uneven. Some countries are subdivided into historical states and modern provinces, others into a handful of regional units. The front matter declares this unevenness as a reflection of the source standards (P10), and the type system makes it explicit at each level.

7 Implementation in the Hydrographic Domain (codehy)

Water bodies are a test case for coordinated plurality because they are physically continuous and only become discrete objects by delimitation. Every sea is a fiat object, and different authorities delimit and group seas differently. codehy (1,490 entries) is accordingly the file with the most systematic parallelism.

7.1 Four Parallel Code Systems

Every water entry in codehy carries four identifiers side by side: the HLC-M identifier, a DIAM+ code, a CWWN+ code and an FAO+ number. Each of these has a different owner and a different purpose.

Table 18: The four parallel code systems in codehy

System	Owner and origin	Format	HLC treatment	Example (Bismarck Sea)
HLC-M	HLC	3–4 characters, `7nn(n)`, `7nAA`	Primary identifier, stable	`7205`
DIAM+	Defense Intelligence Agency Manual 65-18 with HLC changes	`nx`	Original codes kept in brackets; HLC additions marked +, specials !	`4K`
CWWN+	Codes for World Waterbody Names with HLC extensions	`naa`	All predefined codes used; original codes in brackets where shifted; HLC ranges declared	`2TP(3FA)`
FAO+	HLC FAO-style area codes	Dotted hierarchy	Restructured order and ranges; original FAO area in the "FAO & related" column	`26.15`

This structure is a textbook implementation of P1 combined with P5. Where HLC extends an external system (DIAM+, CWWN+), it keeps the original code visible (in brackets) and confines its own additions to declared ranges. The INFO sheet of codehy lists the CWWN+ extension ranges explicitly (`0AA–0ZZ`, `1AR–1AZ`, `1RA–1ZZ`, and analogous blocks for each ocean, plus `9AA–9YZ`). A reader can therefore always tell whether a given CWWN+ code is original or an HLC extension. The note that "SDMX+ (organizations etc.) partly uses also the same codes (nx), so be aware of the item type" shows that the framework is aware of the residual namespace overlap between DIAM+ waters and organisation codes, and resolves it by the type field rather than by changing either code system.

The FAO+ numbering is the only case where HLC restructures an external system. It does so by creating a new, HLC-owned scheme ("FAO style areas codes, but with a different structure/order and number ranges") while the original FAO codes are kept in a separate column, for example `23` for the Northwest Pacific with the original FAO area `(61)`. This respects P1: the external FAO codes are not changed; a parallel HLC scheme is added and clearly named as such.

7.2 Temporal Plurality: Four Editions of the Limits of Oceans and Seas

For each sea, codehy records an HLC selection code (IHO+) followed by the sea's identifier in the IHO S-23 draft of 2002, the draft fourth edition of 1986, the third edition of 1953, and the ACIC chart of limits (M49-1). For the North Sea, these five columns read `(I0C)`, `1.2`, `1.2`, `4`, `B23`; for the Kattegat `(I1L)`, `2.9`, `2.7`, `2`, `B25`. In total, 152 rows carry a 2002 identifier and 107 a 1953 identifier.

This is the clearest instance of P2 applied along the time axis. The successive IHO editions are not equivalent: seas are added, split or renumbered between editions, and the 2002 draft was never formally adopted. A system that kept only the current edition would lose the ability to interpret older publications and charts; a system that kept only the 1953 edition, which remains the last formally approved one, would ignore later refinements. HLC keeps all of them in parallel and lets the user decide which edition governs a given dataset.

7.3 Polyhierarchy of Seas

Seas belong to several larger wholes. HLC implements this with starred duplicates (P3; 102 starred placements in codehy) and the attribute-locality rule (P4). The South China Sea region illustrates the mechanism (Figure 12).

Table 19: Multiple placements of the South China Sea branches in codehy

Placement path (HLC-M)	FAO+ position	Status	Parent context
712-7133 South China Seas (NW Pacific)	23.13	Primary	Pacific Ocean (Northwest)
722-7133*	29.1 (23.13)	Secondary	South China Sea (IHO & SeaVoX Area)
781-7133*	79.11 (23.13)	Secondary	Australasian Mediterranean Sea
722-7220 South China Seas (West-Central Pacific)	29.2	Primary	South China Sea (IHO & SeaVoX Area)
781-7220*	79.12 (29.2)	Secondary	Australasian Mediterranean Sea
780-7803 South China Seas (East Indian Ocean)	79.3	Primary	Indian Ocean (East)
722-7803, *781-7803*	29.3 (79.3), 79.14 (79.3)	Secondary	South China Sea; Australasian Mediterranean Sea

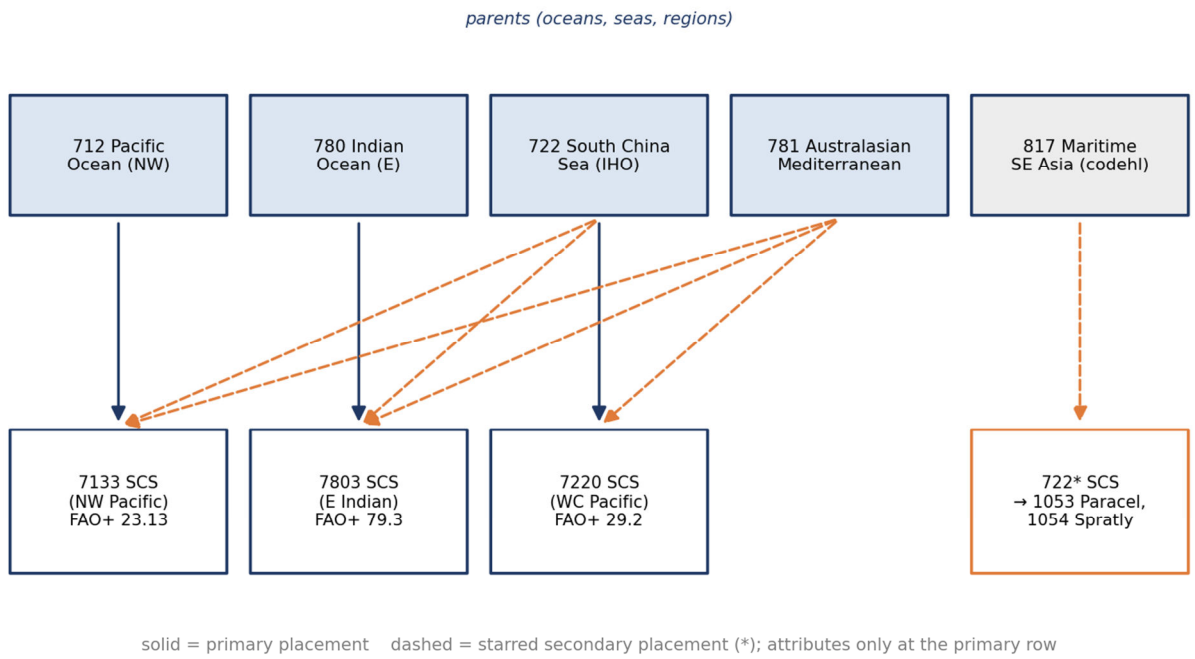

Figure 12: Polyhierarchy of the South China Sea branches: primary and starred secondary placements

The South China Sea region thus appears simultaneously as part of the Pacific, as part of the Indian Ocean, as an IHO sea area in its own right, and as part of the Australasian Mediterranean Sea, an

oceanographic concept that groups the seas between Asia and Australia. Each perspective is legitimate and used by an identifiable community (P9). The secondary rows carry only the path, the starred identifier, the parenthesised original position and the description; coordinates, tags and IHO identifiers are present only at the primary rows (P4).

The same mechanism handles biogeographic water provinces that span oceans. Temperate Australasia (`7272`) groups the Southwest Pacific (`7280*`) and the Southeast Indian Ocean (`7768*`); Temperate South America (`7273`) groups the Southeast Pacific (`7290*`) and the Southwest Atlantic (`7600*`); the Northern European parts of the Arctic (Norwegian and Barents seas) appear again under the Northeast Atlantic (`741-7081*`, `741-7082*`). These cross-ocean groupings correspond to MEOW realms and are marked as such by tags (`{MW:MY00}`, `{MW:MW00}`); MEOW is the second most frequent tag key in codehy (190 instances, after the USGS hydrologic units with 247).

7.4 Fisheries and Regulatory Areas as Attributed Assertions

The "FAO & related" column holds the original FAO area numbers and the subdivisions of regional fisheries bodies as tags: `{NF:...}` for NAFO divisions (for example `{NF:1A}` for Melville Bay and Disko Island), `{IE:...}` for ICES areas (for example `{IE:IIIa}` for Skagerrak and Kattegat, `{IE:IVb}` for the German Bight). The "OTHERS" column adds GFCM subareas (`{FM:...}`), HELCOM subbasins (`{HC:A01}` Kattegat to `{HC:A17}` Bothnian Bay), OSPAR regions (`{OP:I}` to `{OP:V}`), the Catchment Characterisation Model (`{CM:...}`), Italian ISTAT/MIN sea codes (`{IT:...}`), Longhurst provinces, MEOW ecoregions and GOODS provinces.

These regulatory partitions are drawn for management purposes and do not coincide with IHO limits. HLC does not attempt to reconcile them. It attaches each regulatory code to the HLC entry whose conceptual extent best matches it (P7), and records the regulatory code in its own notation (P1). Where a regulatory unit has no corresponding sea, HLC creates an entry for it, as with the NAFO divisions along West Greenland (`7029` "Canadian Eastern Arctic – West Greenland Shelf" with `{NF:0}`) or the ICES subdivisions of the Northeast Atlantic (`7415` Iceland Grounds with `{IE:Va}`).

7.5 Water Columns and Depth Zones

codehy opens with two non-geographical partitions: the climate zones (`689`, mirroring `800` in codehl) and the Marine Eco-Systems by depth (`699`): epipelagic, neritic and oceanic epipelagic with littoral, sublittoral, infralittoral and circalittoral subdivisions, mesopelagic, bathypelagic, bathyal provinces, hydrothermal vent provinces, abyssopelagic, abyssal and hadal provinces, each tagged with the GOODS code (`{OW:OD90}` to `{OW:OD99}`). This vertical partition is orthogonal to the horizontal partition of seas. Keeping it as a separate hierarchy implements P8 within the water domain: depth zones and sea areas are different logics and are not nested into each other.

7.6 Inland Waters and Cross-Domain Links

Inland waters are grouped by continent (`787` Waters of Asia to `796` Waters of the former USSR) and, below that, under the regions of codehl (for example `787-9180` Western Asia with the Caspian Sea, Lake Van and the Euphrates and Tigris). For the United States, codehy embeds the USGS Hydrologic Unit Code structure: the regions `{US:01}` (New England) to `{US:19}` (Alaska) with their subregions (`{US:0101}` and following) placed under the USA entry (`790-4200`), and the regions for Hawaii

(`{US:20}`), the Caribbean (`{US:21}`) and the Pacific islands (`{US:22}`) attached to the corresponding island entries.

This is a cross-domain link by shared identifier (P8): the land entries `912`, `914`, `918`, `420` and others are reused as parents in the water file without duplicating their descriptions. Conversely, codehl contains water entities in the land tree where they are conceptually part of a land unit, such as the Suez Canal (`7771`) under the Suez governorate (Figure 11). The `Waters (Former USSR)` hierarchy (`796`) is entirely composed of starred duplicates, which reconstructs a historical political grouping of water bodies without altering their primary placement: an application of P3 and P6 to historical regions.

7.7 Coordinates as Anchors

codehy gives WGS84 centroids in a compact degrees-and-minutes notation for 954 entries (`{7635N,7357W}` North Water Polynya; `{00N,00E}` Gulf of Guinea, noted as Null Island). Several features show that the coordinates are deliberately non-geometric (P7). Some are ranges rather than points (`{10-20N,126E}` Mindanao Current; `{5-20N}` North Equatorial Current). Some are omitted where a point would be misleading (gyres, water masses, very large regions). The Common Front Matter states that there is "no plan to implement GIS shapelines". The anchors serve orientation and plausibility checks, not spatial computation.

Abstract water entities without location are included as well: water masses (Arctic Bottom Water, North Atlantic Deep Water, Antarctic Bottom Water), gyres and currents. They are conceptual referents that matter to oceanography and fisheries, and their inclusion shows that the water domain is organised by conceptual relevance rather than by mappability.

8 Implementation in the Ecological Domain (codehw)

Ecological classifications are where pluralism is most openly acknowledged by the source disciplines themselves (Section 2.5). codehw is accordingly built on the most radical form of federation in the framework: "None of the worksheets is a master sheet or a subset of another, sometimes, however, siblings."

8.1 Sibling Views without Master

codehw holds fifteen schemes in separate sheets with 5,207 rows in total (Figure 3b); a sixteenth scheme, the draft World Geographical Scheme for Marine Biota Distributions (WGS-RMB, 07/2025), is documented in the INFO sheet with its High Seas level-4 units highlighted, but has no sheet of its own in version 8.1. Each sheet preserves the native hierarchy of its scheme: realms, biomes and ecoregions for TEOW; realms, bioregions and ecoregions for BEOW; realms, habitat types and ecoregions for FEOW; domains, divisions and provinces for Bailey; realms, biomes and functional groups for GET; volumes and chapters for EOW; and so on.

Table 20: Ecological schemes held as siblings in codehw

Sheet	Scheme	Native structure (as documented in HLC)	Logic of division	Rows
GIWA (IW)	Global International Waters Assessment, UNEP 1999	9 megaregions, 66 regions	Transboundary water assessment units	136
GET (GW)	IUCN Global Ecosystem Typology v2.1 (Keith et al. 2022)	10 realms, 25 biomes, 110 functional groups	Ecosystem function across realms	145
BYES (YW)	Bailey's Ecoregions (map 1989; book 1998/2014)	Oceanic 3 domains, 13 divisions; continental 4 domains, 30 divisions	Macroclimate	153
EOW (EW)	Ecosystems of the World (Goodall, ed.); Subterranean Ecosystems	Volumes and chapters	Monographic synthesis	144
GEZ (ZW)	FAO Global Ecological Zones, FRA 2000 (FAO 2001)	5 domains, 20 ecological zones	Climate and vegetation physiognomy	26
TEOW (TW)	Olson et al. 2001	8 realms, 14 biomes, 867 ecoregions	Terrestrial biodiversity	1,149
BEOW (BW)	Dinerstein et al. 2017	8 realms, 14 habitat types, 846 ecoregions	Revised terrestrial ecoregions	1,171
FEOW (FW)	Abell et al. 2008	8 realms, 12 habitat types, 426 ecoregions	Freshwater fish distribution	575
EUNIS (UW)	European Nature Information System, 2021/22/26	567 level-3 habitats	Habitat types	740
BPOW (WW)	Maureaud et al. 2023	Benthic provinces by bathymetric type	Deep-sea floor biogeography	193
GOODS (OW)	Watling et al. 2013	Pelagic and benthic provinces	Open ocean and deep sea	102
MEOW (MW)	Spalding et al. 2007	12 realms, 62 provinces, 232 ecoregions	Coastal and shelf biogeography	366
LMES (LW)	Sherman & Alexander 1989	Ecosystems 001–066	Large marine ecosystem management	90
PAME (AW)	Arctic Council PAME (LME map 2013), with AMAs (Gill et al. 2011)	18 Arctic LMEs, AMA additions	Arctic marine management	116

Sheet	Scheme	Native structure (as documented in HLC)	Logic of division	Rows
PPOW (PW)	Spalding et al. 2012	4 realms, 7 biomes, 37 provinces	Pelagic biogeography	101

Row counts include grouping and structural rows and therefore exceed the number of ecoregions or provinces in the native scheme. The decision not to designate a master is a direct consequence of P2 and P8. There is no ecological meta-criterion that would make TEOW superior to BEOW, or MEOW superior to PPOW: they model different things. Choosing a master would turn a scientific plurality into an administrative hierarchy.

8.2 The Identifier Pipeline: from Source Code to HLC-M

Every row in codehw carries five identifier columns: HLC-M, HLC-D, HLC-X, HLC-S and REF-CODE. They form a derivation pipeline (Figure 13), which can be followed in a single row of the TEOW sheet.

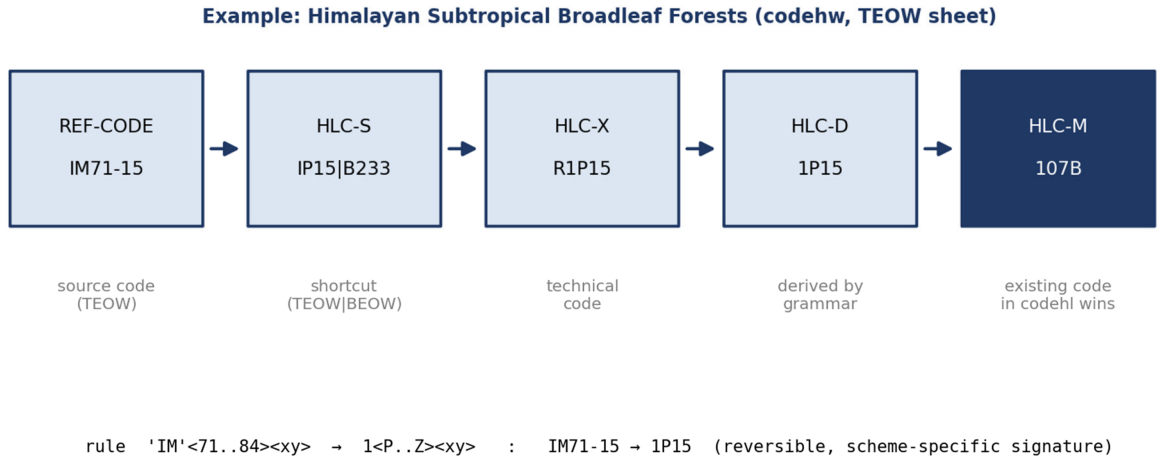

Figure 13: The codehw identifier pipeline, traced through one TEOW row

1. **REF-CODE** is the code as given by the source scheme (`IM71-15` for the TEOW ecoregion Himalayan Subtropical Broadleaf Forests).
2. **HLC-S** is an HLC shortcut for cross-referencing (3–5 characters), here combining the TEOW shortcut and the BEOW number (`IP15|B233`).
3. **HLC-X** is the technical code, here a realm-prefixed form of HLC-D (`R1P15`).
4. **HLC-D** is the derived code, computed from REF-CODE by the replacement grammar in the HLC-M sheet (`1P15`).
5. **HLC-M** is the final identifier. It equals HLC-D unless the same referent already has an HLC-M code in codehl or codehy, in which case that existing code is used; here `107B`, because the ecoregion is also an entry in the land tree under the Himalaya.

This pipeline implements four sub-principles at once. The source code is kept unchanged (P1). The derivation is deterministic and documented, so that every HLC-D code can be recomputed from its source (P6). Existing identifiers take precedence over derived ones, so that a referent present in several files has one HLC-M identity (P5, P8). The INFO sheet adds that HLC-D "is the proposed value for HLC-M" and that users who want "to remain independent of other parts of the framework" may use it directly: the ecological file can be used standalone without losing identifier stability.

8.3 The Replacement Grammar

The HLC-M sheet of codehw specifies, for each scheme, a rewriting rule from source-code patterns to HLC-M patterns. Examples:

Table 21: Selected rules of the codehw replacement grammar

Scheme	Source pattern (HLC-S / REF-CODE)	HLC-M pattern
GIWA	‘G’<G|H|I|J|K|R|V|X|Y><xy:01–66,89–94>	0‘G’<xy>
GET	<‘GT’|‘GM’><10–99>	0<‘I’|‘J’><10–99>
BYES	<‘YM’|‘Y’><x:1–4><10–99>	<x:1–4><‘M’|‘N’><10–99>
GEZ	‘GZ’<xy:00–99>	0‘Z’<xy>
TEOW	‘IM’<71|72|…|84><xy:00–99>	1<P|Q|R|…|Z><xy>
TEOW (G200 groups)	‘G’<xy:00–14,24><z>	<1–8><‘P’–‘Z’><xy:‘A’–‘O’,‘Y’><z>
FEOW	‘F’<1–7><00–99>	<1–8><B|C|D|E|F|G|H><00–99>
MEOW	‘M’<K|L|…|Z><xy:01–99>	9‘M’<xy>
LMES	‘LM’<xy:01–99>	9‘L’<xy>
PPOW	‘PO’<xy:01–99>	9‘P’<xy>

Three properties of the grammar are relevant to the principle. First, it is *designed to be injective*: distinct source codes map to distinct HLC-M codes, because each scheme is assigned its own character signature (for instance `9M`, `9L`, `9P` for MEOW, LMEs and PPOW; `0G`, `0Z` for GIWA and GEZ; a leading realm digit followed by a biome letter for TEOW). Second, it is *range-respecting*: the produced codes fall into the HLC-M ranges that codehl assigns to HW (‘0’–‘9’anx for leaf codes, `90AA–99ZZ` for marine leaf codes). Third, it is *reversible*: the source code can be reconstructed from the HLC-M code and the scheme, so that no information is lost by the transformation. These properties make the grammar the most formally rigorous implementation of P5 in the framework.

8.4 Integration of Overlay Schemes

Several ecological schemes are overlays on others: WWF Global 200 (Olson & Dinerstein 2002) selects and groups TEOW, FEOW and MEOW ecoregions; Longhurst provinces partly correspond to PPOW provinces; AMAs refine PAME areas. HLC integrates overlays without merging them, using four techniques documented in the INFO sheet.

- **Grouping levels.** Global 200 groups G001 to G142 are inserted as sublevels in TEOW (for example `5686` "G001 – Guinean moist forests" grouping three TEOW ecoregions). Where Global 200 lacks groups for certain mangroves and Antarctic units, HLC adds grouping items G241–G244 and T611–T617 "due to regulation lacks".
- **Reserved number blocks.** Global 200 freshwater units G143–G195 are integrated into FEOW as numbers 843–895; Global 200 marine units G196–G238 into MEOW as 296–338; Longhurst provinces into PPOW as 41–56; AMA additions into PAME as AW72–AW77; additional BEOW items 980–999 harmonise BEOW with TEOW.
- **Highlighting of completeness items.** In BPOW, items included "only for completeness" are distinguished from those "for common use" by colour, and in PPOW the two biomes and provinces that belong to GOODS but not to PPOW are highlighted.

- **Sort-only levels.** Where a source hierarchy is fluid, the sublevel `-0000-` is inserted only "to ensure the right sort order and must be omitted from the hierarchical analysis". This keeps tabular order and hierarchical meaning separate.

Each technique keeps the overlay scheme recognisable as a distinct assertion while placing its units where they belong in the base scheme. The reserved number blocks are a further application of P5: overlay units are given their own ranges within the base scheme's numbering, so that they cannot be mistaken for base units.

8.5 Explicit Cross-Scheme Correspondences

The "Details" column of several sheets records correspondences between schemes as text. Bailey's oceanic divisions list the PPOW provinces they contain (for example "510 – Inner Polar Division (PO01, PO34, PO35)") and note where a division is known only to one author ("Dietrich only", "Elliott (1954) – unknown to Dietrich"). Bailey's continental divisions cite GEZ equivalents ("130 – Subarctic Division (GZ61, GZ62)"). EOW volumes cite the TEOW, BEOW, FEOW, GET, GEZ and EUNIS units they cover ("EOW06-01 boreal forest/taiga – TEOW/BEOW: PA/NA 76"). GET functional groups cite the TEOW biome numbers, EUNIS habitats, GOODS depth zones and PAME areas that instantiate them.

These correspondences are *qualified* in the text: an asterisk marks partial membership (`PO15*`), "(partial)" or "(sliver)" marks partial overlap with EUNIS, "nested" marks a group contained in another. They implement P2 at the level of inter-scheme alignment: correspondences are recorded as statements with their limitations, not as equivalences. In SKOS terms, the text distinguishes close, broad and narrow matches, although it does not yet use a controlled vocabulary for them (Section 11.2).

8.6 Umbrella Typologies

Two schemes in codehw are explicitly "umbrella" systems: GET, which covers terrestrial, freshwater, marine, subterranean and transitional realms in one hierarchy, and EOW, the monographic Elsevier series extended by the Subterranean Ecosystems volumes (EOW30 to EOW35). Umbrella systems are the closest ecological analogue to a master scheme. HLC nevertheless keeps them as siblings. Their role is to *index* other schemes through correspondences, not to subordinate them. This is the strongest evidence that HLC's non-master policy is principled: even where a candidate master exists and claims comprehensiveness, it is not promoted.

8.7 Links from codehw to the Other Files

Ecological entries appear in the other files in two ways. In codehl, the biogeographic realm tree (`898`) carries TEOW, BEOW and FEOW codes as tags down to the biome level, and individual ecoregions appear as type E entries (99 in codehl) under the land units they characterise (for example the Sri Lanka ecoregions `102D`, `102E`, `102L`, `102M`, `102W` under Sri Lanka, or the Himalayan ecoregions `107B` to `107Z` under the Himalaya). In codehy, MEOW, GOODS and FEOW codes are attached as tags to seas and inland waters. GIWA, finally, is organised under the climate-zone tree, linking ecological assessment units to the zonal partition of codehl. The ecological domain is thus woven into the others by identifiers and tags while retaining its own hierarchies (P8).

9 Cross-Domain Case Studies

The preceding chapters analysed mechanisms one file at a time. The principle, however, shows its value where mechanisms interact. Five cases, selected by the criteria in Section 3.6, trace selected referents through all files.

9.1 The South China Sea and its Islands

The South China Sea combines every difficulty: it is a sea claimed in part by several states, it is crossed by oceanographic and ecological boundaries, it contains disputed island groups and it is subdivided differently by IHO, FAO, MEOW and GIWA.

Table 22: The South China Sea across the HLC files

File	Representation	Mechanism and sub-principle
codehu	`VS|3U`, SCS, X3U, HLC-M `722`, type W, "South China Sea (IHO & SeaVoX Area)"; `RX|5U` "South China Sea & East Indian Archipelago" `721`; `YX` Spratly, `RP` Paracel (type X)	Namespace: HLC-2 pseudo-code and DIAM+ code side by side (P5); islands typed as contested (P2)
codehy	Primary `722` under `721` under `719` (West-Central Pacific); branches `7133`, `7220`, `7803` with starred copies under `722` and under `781` Australasian Mediterranean Sea; IHO columns `(I5B)`, `6.1`, `6.1`, `49`, `D18`; `{MW:MQ25}`; FAO `(71.1)`	Polyhierarchy (P3) with locality (P4); four IHO/ACIC editions (P2); MEOW and FAO as attributed tags (P1)
codehl	Primary `722` under the West-Central Pacific (`903-719-721-722`) and `722*` under Maritime South-East Asia (`817`), with children Paracel Islands (`1053`) and Spratly Islands (`1054`), each annotated with claimants and GENC codes `{US:XP,XPR,928}`, `{US:XS,XSP,929}`	Neutral-container placement for contested islands (Section 4.6); GENC as attributed assertion (P1)
codehw	GIWA unit "South China Sea (LM36)" `7220` under Southwest Pacific (`7280`); TEOW/BEOW ecoregion "South China Sea Islands" (`1P48`, `IP48|B267`) with the claimant ISO codes CHN, MYS, PHL, TWN, VNM	Ecological assessment and terrestrial ecoregions as sibling views (P8); shared identifier `7220` across files

The case shows how the same referent is reached from four directions without any file overriding another. The island groups never receive a sovereign parent: in the land file they hang under the sea, in the ecological file they belong to an ecoregion whose description lists all claimants. The sea itself has three oceanic contexts, of which one is primary and two are marked (Figure 12). The four IHO/ACIC columns allow readers of charts from 1953 to 2002 to locate the sea in their edition.

9.2 Egypt, Sinai and the Suez Canal

Egypt shows how transcontinental identity, sub-national units, water entities and historical demilitarised zones are combined (Figure 11).

- **Whole and parts.** The whole state `EG|G0` (HLC-M `885`) is typed C. Its parts Egypt (Asia) `150` and Egypt (Africa) `250` are typed X in codehl and sit under the Sinai Peninsula (`813`, Western Asia) and North-Eastern Africa (`923`). Both carry `{UN:(818)}` (Section 6.5).
- **Sub-national structure.** Under Egypt (Asia), the Sinai units are arranged as Northeast Egypt (`1503`), Areas of Sinai (`1504`, `{CW:6512}`), Sinai demilitarised zones (`1501`, `{CW:6513}`), North Sinai governorate (`1502`, ISO `EG-SIN`) with the Gidi and Mitla passes, and the Suez Canal Zone (`1505`, `{CW:6514}`) with the governorates of Port Said and Suez.

- **Water inside land.** The Suez Canal (`7771`, `DS|6U`, `{FA:73.1}`, `{WN:7RB}`) appears in codehl at the end of the path `…-813-150-1503-1505-1509-7771` and in codehy under the Western Indian Ocean (`777`). The same HLC-M identifier links the two placements across domains (P8).
- **Ecological and botanical context.** The Sinai Peninsula carries `{PD:(34)SIN}` (WGS-RPD) and `{FW:F435}` (FEOW), recording that botanical and freshwater schemes treat Sinai as a unit of its own.

Three different kinds of relation are visible here and are kept distinct: containment of governorates in a state part (land hierarchy), conceptual location of a canal in a governorate and in an ocean (cross-domain polyhierarchy), and historical overlay of demilitarised zones as entities with COW codes (temporal plurality).

9.3 Antarctica

Antarctica is governed by the Antarctic Treaty system, which freezes territorial claims. Seven states maintain claims, some overlapping; one sector is unclaimed. A framework that placed the continent under any claimant would violate the treaty's logic; one that ignored the claims would lose information used by national statistics and mapping agencies.

Table 23: Antarctic claims and units in codehu

HLC-2	HLC-M	Type	Description in HLC
AQ|A7	670!-80	C	Antarctica (ISO country code)
AB>AQ	671	M	British Antarctic Territory (GB)
CQ	672(0)	M	Chilean Antarctic Territory (CL)
TQ	673(0)	M	Argentine Antarctica (AR)
NQ	674	M	Dependencies of Norway (NO)
UQ	675(0)	M	Australian Antarctic Territory (AU)
RQ	676	M	Ross Dependency (NZ)
FQ	6794	H	Adélie Land (FR)
DR	7859	D	Dronning Maud Land (NO)
DQ	6749	D	Peter I Island (NO)
MB	6705	D	Marie Byrd Land (except Peter I Island, unclaimed territory)
WQ / EQ / TX	974 / 976 / 975	S	West Antarctica, East Antarctica, Transantarctic Mountains

The claims are typed M (*Multiples*: area of several states) rather than D (division), which avoids asserting that they are subdivisions of the claimant state. The British claim is recorded with a redirect to AQ (`AB>AQ`), stating that the code is subordinate to the continent rather than to the United Kingdom. Physical subdivisions (West and East Antarctica, the Transantarctic Mountains) form a second, non-political partition. In codehy, the Southern Ocean (`782`) and its Pacific, Atlantic and Indian sectors carry the surrounding seas, and the "Waters of Antarctica" (`794`) hold subglacial lakes and polar freshwaters with FEOW tags (`{FW:QA70}`, `{FW:QA64}`). In codehw, TEOW adds the grouping items G241 (Continental Antarctica) to G244 (Southern Indian Ocean Islands) under QA81 Tundra, with leaf units that reuse the land identifiers (`6705` Marie Byrd Land, `6746` Dronning Maud Land Tundra, `9750` Transantarctic Mountains). The continent is therefore represented in a political-claims view, a physical view, a hydrographic view and an ecological view, all sharing identifiers and none subordinated to another.

9.4 Three Zonal Logics

Latitudinal zonation looks like the simplest partition of the earth. HLC shows that there are at least three.

Table 24: Three zonal logics in HLC

Logic	Codes	Criterion	Location in HLC
Astronomical climate zones	Z0–Z4 with subzones ZK, ZC, ZN, ZS, ZL, ZQ (`8000–8006`, `8101–8104`)	Arctic and Antarctic circles, tropics (e.g. 66°33′39″, 23°26′22″)	codehl `800`, codehy `689`, GIWA in codehw
FAO Global Ecological Zones	Z5–Z9 (`8105–8109`), GEZ subzones GZ51–GZ96	Number of months above 10 °C; frost; vegetation physiognomy	codehl `810`, codehw GEZ sheet
Bailey domains	Polar, Temperate, Dry, Humid Tropical (oceanic and continental)	Macroclimate after Dietrich (oceans) and Köppen–Trewartha (continents)	codehw BYES sheet

The three logics are related but not equivalent. The GEZ Polar zone (`8105`) contains both astronomical frigid zones (`8000` and `8004`) as children; the GEZ Boreal zone contains the Subarctic and Subantarctic zones; the GEZ Tropical zone contains the northern and southern Equator zones. Bailey's domains are cross-referenced to GEZ zones in the details column. Meta-groupings such as Polar Domain (`7996`, "Z0+ZK+Z4"), Temperate Domain (`7997`, "Z1+Z3"), Polar Boreal Zones (`7998`) and Subtropics Zones (`7999`) are typed M and attributed ("by Bailey/Dietrich", "Global Ecological Zones (GEZ, FRA 2000)"). Instead of choosing one zonation, the framework encodes all three and the relations between them, including their attribution.

9.5 Greece: EL and GR

The smallest case shows the principle at the level of a single string. ISO 3166-1 codes Greece as GR. The European Union uses EL in its statistics and institutions. A system that adopted one would misinterpret data coded with the other. HLC registers both in codehu: `GR|I4`, GRC, HLC-M `371!-73`, type C ("Greece (SEE) – EL (Ellas, Ellada)"), and `EL>(GR)`, ELL, the same HLC-M range, type X ("Greece (Ellas, Ellada) –> GR (used by the EU)"). The EU code is kept, typed as special, attributed to the EU and redirected to the ISO entry, which remains primary. This single pair of rows implements P1 (both codes unaltered), P2 (the divergence is recorded and attributed), P5 (both strings owned and classified) and P6 (one identity behind two codes).

9.6 Synthesis of the Cases

Table 25: Sub-principles observed in the case studies

Case	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
South China Sea	●	●	●	●	●	○	●	●	○	○
Egypt / Sinai / Suez	●	●	●	○	●	●	●	●	○	○
Antarctica	●	●	●	○	●	●	●	●	○	●
Three zonal logics	●	●	●	●	●	○	●	●	●	○
Greece EL/GR	●	●	○	○	●	●	○	○	○	○

● = sub-principle clearly operative in the case; ○ = not specifically exercised by the case. Every sub-principle is exercised by at least one case, and P1, P2 and P5 by all five; P9 and P10 are exercised least, which is consistent with their character as framework-level rather than entry-level rules.

10 Evaluation

This chapter applies the evaluation procedure of Section 3.7. It assesses to what degree HLC implements each sub-principle, identifies the structural tensions that the implementation reveals, and discusses the significance of the substrate (spreadsheets) and of the curatorial model.

10.1 Rating Scheme

Each sub-principle is rated **full**, **partial** or **implicit** according to the decision rules in Section 3.7: full requires an explicit rule, a mechanism in every relevant file and no structural counter-instance; partial applies when the rule is explicit but the mechanism is not uniform, depends on inference or has a structural counter-instance; implicit applies when the rule is only reconstructable from practice. Each rating rests on evidence from at least two independent locations.

10.2 Conformance Matrix

Table 26: Conformance of HLC v8.1 with the sub-principles of Coordinated Plurality

Sub-principle	Rating	Principal evidence (with corpus indicators)	Limiting factor
P1 Non-Replacement	Full	IATs in source notation (9,081 in codehl, 43 keys; 656 in codehy, 11 keys); original DIAM+ and CWWN+ codes in brackets; FAO+ declared as separate HLC scheme; licence excludes external codes	HLC pseudo-codes share the visual format of ISO two-letter codes; distinction relies on the codehu register
P2 Conflict as Information	Full	44 claim annotations in codehl; administering-party marker `!`; four IHO/ACIC editions (152 rows with 2002, 107 with 1953 identifiers); parallel national classifications; qualified correspondences in codehw	Placement mode of contested entries must be inferred from the parent (Section 6.4)
P3 Legitimate Polyhierarchy	Full	102 starred placements in codehy; starred entries in codehl (e.g. `722*`, `751*`); whole-and-parts pattern for four transcontinental states	Starring is applied less uniformly to zone mirrors in codehl than to sea duplicates in codehy
P4 Attribute Locality	Full	Secondary rows carry path, identifier and description only; tags, coordinates and IHO codes at primary rows	Rule enforced by curation, not by structure
P5 Namespace Sovereignty	Partial	837 registered strings in codehu; 134 colon entries; HLC-M range table; replacement grammar in codehw; reserved CWWN+ extension ranges	Residual reuse of `nx` strings by DIAM+ waters and SDMX+ organisations, disambiguated by type rather than by string
P6 Identity Persistence	Partial	9 redirects > in codehu; 663 historical entries retained in codehl; deterministic HLC-D derivation	Front matter admits that consolidation of ranges is sometimes unavoidable; no published identifier-change log at code level
P7 Conceptual Assignment	Full	Explicit coordinate policy; 954 anchors in codehy and 208 in codehl as points or ranges; no geometry; installations under seas	–
P8 Domain-Appropriate Logics	Full	Five-file federation; type matrix; 15 sibling sheets in codehw; depth zones separate from sea areas; history as separate epoch hierarchy	–
P9 Bounded Plurality	Partial	Three admission criteria stated in v8.1.2; scheme keys registered as logical entries	Criteria are recent; recorded reasons for earlier admissions are not yet visible in the data files

Sub-principle	Rating	Principal evidence (with corpus indicators)	Limiting factor
P10 Declared Epistemic Limits	Full	Limitations section: single maintainer, not an authority, no GIS geometry, not for legal use, snapshot, uneven granularity, editorial judgement	–

Figure 14: Conformance profile of HLC v8.1 by sub-principle and layer, with selected corpus indicators

Seven of ten sub-principles are fully implemented, three partially. No sub-principle is absent. The partial ratings concern the integrity and governance layers (P5, P6, P9), whereas the foundational and representational layers are fully realised (Figure 14). This pattern matches the genealogy in Section 4.1: representational devices were introduced early in response to concrete sources, while integrity and governance were codified only in the most recent phase.

10.3 Inherent Tensions

Some limiting factors in the matrix are not deficiencies of HLC but tensions within the principle itself. Any implementation of coordinated plurality will meet them, and must decide them explicitly.

10.3.1 Stability versus Consolidation

P6 demands that identifiers never change. Yet a framework that grows by accretion will discover that early allocations were too narrow (a continent's range fills up) or too wide (a block remains empty). HLC's front matter acknowledges this: identifiers are intended to be stable, but "sometimes changes or consolidation of the number ranges are unavoidable". The tension cannot be removed, only managed. The standard resolution in identifier practice is to never reuse a retired identifier and to publish a machine-readable change log mapping old to new identifiers. HLC already uses the redirect operator for entities; extending it to identifier changes would align P6 with the reality of growth.

10.3.2 Expressiveness versus Usability

The compact notation (Section 5.6) allows a single cell to carry several assertions, which is what makes a crosswalk of more than forty schemes manageable in spreadsheets. The same density raises the entry barrier: a reader must know that `{UN:(145)}` denotes an inherited value, that `153~` denotes shared numeric space, and that `!` means different things after a country code and after a DIAM+ code. The more plural the framework, the more it needs a precise grammar; the more precise the grammar, the more it requires training. The appropriate resolution is to separate the authoring notation from the

publication format, for example by publishing a normalised machine-readable rendition alongside the human-readable files. HLC already publishes TSV and SSV renditions; a fully normalised export is the natural next step.

10.3.3 Neutrality versus Naming

Labels must use words, and words carry positions: "occupied", "claimed by", "administered", "unclaimed". HLC uses short descriptive phrases that state who claims what. The analysis in Section 4.6 shows that neutrality is achievable only as attributed neutrality: each loaded term should be traceable to a source that uses it, and the placement mode should be explicit. Where a label uses a term that is itself contested, attributing the term to the body that uses it would bring the label into line with P2.

10.3.4 Plurality versus Single-Curator Judgement

P9 bounds plurality by criteria, but applying the criteria requires judgement, and HLC states that this judgement is exercised by one maintainer. A single curator guarantees consistency of style and decisions; it also concentrates the risk of bias and of discontinuity. The front matter handles this tension in the only way available to a one-person project: by declaring it (P10), by publishing the criteria (P9), and by inviting corrections with a defined format. The tension would be reduced, not removed, by an advisory review of admission decisions and by recording the reason for each admission alongside the admitted scheme's key in codehu.

10.3.5 Locality versus Readability

P4 places all attributes at the primary placement. A reader browsing a secondary placement therefore sees an entry without its tags, coordinates or codes. This is the price of integrity. The tree path and the parenthesised original position mitigate the cost by pointing to the primary row. In a database rendition the cost disappears, because secondary placements become edges pointing to a single node.

10.4 The Spreadsheet Substrate

HLC is authored in spreadsheets. This choice has consequences for each layer of the principle.

- **Representational layer.** Spreadsheets are monohierarchical by nature: one row, one position. HLC overcomes this with tree paths and starred rows, which is an effective workaround but encodes graph structure in strings.
- **Integrity layer.** Spreadsheets do not enforce uniqueness, referential integrity or controlled vocabularies. P4 and P5 are therefore enforced by curation. The namespace register in codehu and the range table in codehl are the manual equivalent of database constraints.
- **Governance layer.** Version history is maintained at file level in the front matter rather than at row level.

The substrate also has advantages that are often underestimated: it is transparent, editable without tooling, readable by domain experts, and durable. For a single-maintainer reference framework, these properties support P10 directly, because every user can inspect every assertion. The extraction procedure of Section 3.4 demonstrates the other side: because the substrate is regular, mechanism use can be counted by simple scripts. The appropriate conclusion is not that the substrate should be replaced, but that it should be complemented by a validation layer and a normalised export (Chapter 11).

10.5 What the Implementation Adds to the Principle

The relation between principle and implementation runs in both directions. HLC does not only realise the principle; its practice contributes elements that the original maxim did not contain and that the evolved principle in Section 4.8 has adopted.

- **Whole-and-parts entities** for referents that straddle partitions (transcontinental states) are an alternative to starred duplicates. They show that polyhierarchy can be realised by decomposition as well as by multiple placement.
- **Registered authorities.** Making each tag key a controlled logical entry turns provenance into part of the namespace.
- **Deterministic derivation** of identifiers from source codes (codehw) shows that P1 and P6 can be reconciled for very large external schemes: the coordinating identifier is a reversible function of the source code.
- **Edition columns** for a single scheme (IHO 1953, 1986, 2002, ACIC) show that P2 applies along the time axis within one authority, not only between authorities.
- **Sort-only levels** separate the tabular order of the medium from the hierarchical semantics of the content.

11 Generalisation and Transfer

If Coordinated Plurality is a principle and not only a description of HLC, it must be expressible independently of HLC's files and transferable to other domains. This chapter performs the analytical generalisation foreseen in the research design (Section 3.1): it proposes a reference model, maps HLC constructs to semantic-web standards and outlines further research.

11.1 A Reference Model for Coordinated Plurality

The reference model consists of six components (Figure 15), each corresponding to one or more sub-principles and each instantiated in HLC.

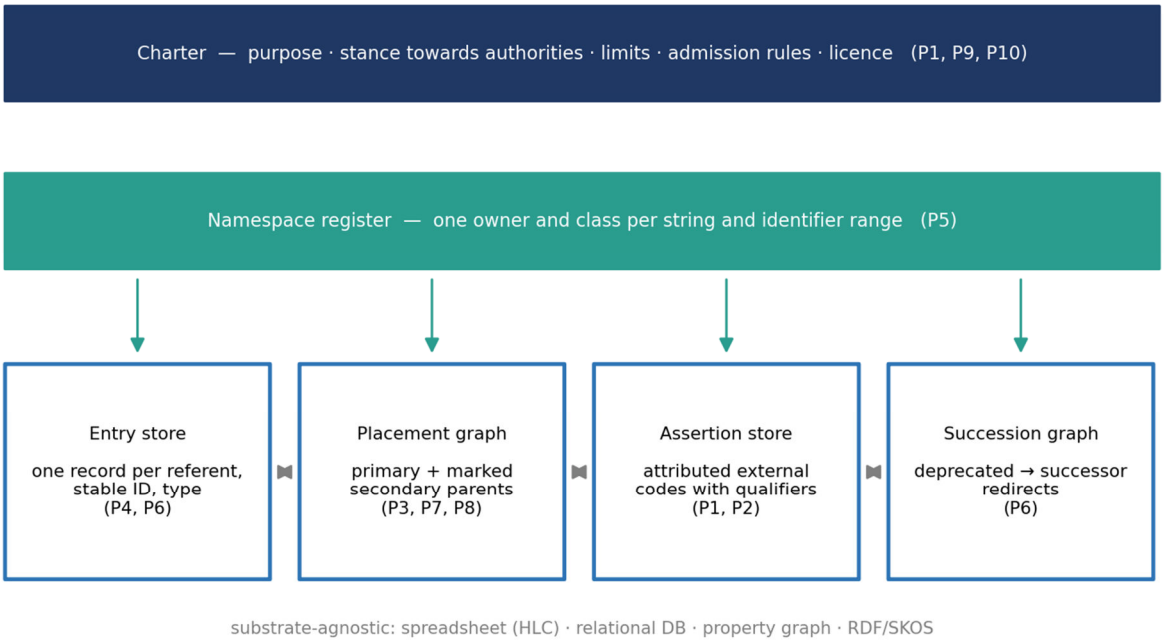

Figure 15: Substrate-agnostic reference model for Coordinated Plurality

Table 27: Reference model for Coordinated Plurality and its instantiation in HLC

Component	Function	Sub-principles	HLC instantiation
Charter	States purpose, stance towards authorities, limits, admission rules, licence	P1, P9, P10	codehv (Common Front Matter)
Namespace register	Assigns one owner and class to every shared string and identifier range	P5	codehu (HLC-2, HLC-3, types); HLC-M range table in codehl
Entry store	Holds one record per referent with stable identifier, type and description	P4, P6	Primary rows in codehl, codehy, codehw
Placement graph	Relates entries by primary and secondary parent relations per domain	P3, P7, P8	Tree paths and starred rows; domain files
Assertion store (crosswalk)	Holds attributed external assertions with qualifiers	P1, P2	Inline attribute tags; edition columns; details columns
Succession graph	Relates deprecated entries to successors	P6	Redirect operator >; historical entities with time spans

The model is deliberately agnostic about the substrate. It can be realised in spreadsheets (as in HLC), in a relational database (entries, placements and assertions as tables), in a property graph (entries as nodes, placements and successions as typed edges, assertions as node properties with provenance) or in RDF.

11.2 Mapping HLC to SKOS and RDF

Publishing HLC as linked data would make its assertions machine-actionable without changing its principle. The following mapping shows that every HLC construct has a natural counterpart in existing W3C vocabularies, which is further evidence that the principle is general.

Table 28: Mapping of HLC constructs to SKOS, Dublin Core and PROV

HLC construct	Proposed RDF representation
Entry with HLC-M identifier	skos:Concept with IRI derived from HLC-M; skos:notation for HLC-M, HLC-2, HLC-3
Domain file / sibling sheet	skos:ConceptScheme per domain and per source scheme
Primary placement	skos:broader to the parent concept
Starred secondary placement	Additional skos:broader, typed by a sub-property (e.g. hlc:alternativeBroader) with the source hierarchy as provenance
| alternative code or name	Additional skos:notation with datatype per scheme; skos:altLabel
> redirect	dct:isReplacedBy (and dct:replaces in reverse); owl:deprecated true
: logical / organisational entry	Separate concept scheme for organisations; membership via a dedicated property
Inline attribute tag {k:v}	skos:notation typed with the datatype of scheme k, or skos:exactMatch/closeMatch to the concept v in scheme k
Parenthesised tag value (k:(v))	skos:broadMatch or a qualified relation (inherited / former / approximate) with PROV attribution
Claim annotation and !	Reified assertion: claimant, administering party, placement mode, source
Type code	rdf:type to a class in an HLC type ontology derived from the codehu matrix
WGS84 anchor	geo:lat / geo:long with an explicit note that the point is an anchor, not a geometry
Edition columns (IHO 1953/1986/2002, ACIC)	One concept scheme per edition; skos:exactMatch or closeMatch between edition concepts and the HLC entry

Two HLC features go beyond plain SKOS and require small extensions: the qualifiers of correspondences (inherited, former, approximate, partial) and the placement mode of contested entries. Both can be expressed with reified statements or RDF-star, attributing each assertion to a source with PROV-O. This is precisely the requirement stated by P2: conflict is information only if each position remains attributable.

11.3 Validation Rules Derived from the Principle

The formal model of Section 4.3 yields checkable constraints. They do not change HLC's content; they make its integrity layer verifiable. The extraction procedure of Section 3.4 already implements the pattern-matching that most of these rules require.

1. Every HLC-2 and HLC-3 string used in any file is registered in codehu with exactly one owner and class (P5).
2. Every HLC-M identifier falls into a range assigned to the file in which it is primary (P5).

3. Every identifier occurs at most once without a secondary marker across all files, except where a deliberate cross-file shared identifier is declared (P3, P8).
4. Every secondary row carries no attribute values beyond path, identifier, marker, original position and description (P4).
5. Every redirect target exists or is a registered plurality code such as `HH` (P6).
6. Every tag key used in an IAT is registered as a logical entry in codehu (P1, P5).
7. Every entry of type X with a claim annotation declares a placement mode (P2, Section 4.6).
8. Every HLC-D code in codehw is recomputable from its REF-CODE by the replacement grammar (P6).

Such rules can be implemented as a short script over the published TSV/SSV renditions and run before each release, which would turn the curatorial enforcement noted in Section 10.4 into automated assurance without altering the authoring workflow.

11.4 Transfer to Other Domains

The principle applies wherever several authoritative classifications describe overlapping referents. Three examples indicate its range.

- **Languages.** ISO 639 parts, Glottolog and Ethnologue classify languages and dialects with different criteria and hierarchies. Coordinated plurality would keep each scheme authoritative, record genealogical and areal groupings as alternative hierarchies, and treat disputed language/dialect status as attributed assertion.
- **Occupations and industries.** ISCO, national occupation codes, NACE, NAICS and ISIC are linked by official correspondence tables that are many-to-many and revised across versions. Edition columns and qualified correspondences, as in codehy, directly address this.
- **Health classifications.** ICD-10, ICD-11, national modifications and clinical terminologies coexist in health information systems. The redirect and succession mechanisms of HLC correspond to the needs of version transitions, and attribute locality to the needs of a single clinical concept with several classification placements.

11.5 Research Agenda

- **Independent replication.** A second analyst applying the codebook and rating rules of Chapter 3 to the same corpus would test the reliability of the conformance ratings.
- **Empirical usability.** How do users with different backgrounds (statisticians, ecologists, historians) read and use the compact notation, and which publication format best preserves the assertions for each group?
- **Placement-mode ontology.** A controlled vocabulary for placement modes and claim types, tested against the full set of contested entries in HLC and in other gazetteers.
- **Identifier evolution.** Quantitative analysis of identifier stability across HLC releases, to calibrate the tension between P6 and growth.
- **Alignment quality.** Comparison of HLC's qualified correspondences with geometric overlap measures where boundary data exist, to assess how well conceptual assignment (P7) approximates spatial relations.
- **Governance models.** Comparison of single-curator, editorial-board and community models for admission decisions under P9.

12 Conclusion

This thesis set out to evolve the maxim "harmonization, not replacement" into an explicit principle and to show how the HLC framework implements it.

Methodologically, the study was designed as a single embedded case study with a theory-building aim. It combined conceptual analysis, formal reconstruction, directed content analysis with a codebook derived from the principle, scripted extraction of corpus indicators and cross-domain case tracing, under explicit measures for the author's insider position and fixed evaluation rules.

For the first aim (RQ1), the maxim has been developed into the **Principle of Coordinated Plurality**. It rests on interdisciplinary foundations: the pluralist philosophy of classification, the view of classifications as infrastructure, polyhierarchy and mapping in knowledge organisation, the fiat and contested nature of geographic units, the model character of ecological partitions, the mandate-bound authority of statistical code lists, the theory of persistent identifiers and namespaces, and the treatment of temporal entities in historical geography. The principle has been decomposed into ten sub-principles arranged in four layers (foundational, representational, integrity, governance), formalised on a common model and related through explicit dependencies. Developing the principle also produced a conceptual result beyond the original text: neutrality in a hierarchical framework can only be *attributed* neutrality, which requires that the placement mode of contested entries be declared.

For the second aim (RQ2), the analysis has shown that HLC implements the principle through a coherent set of mechanisms: a federated five-file architecture separating policy, namespace and content; a namespace register of 837 strings; a partitioned identifier space with explicit allocation and transparent sharing; a compact assertion grammar; inline attribute tags that act as an embedded, faceted crosswalk (more than 9,700 tag instances in the land and water files); starred placements governed by attribute locality; whole-and-parts entities for transcontinental states; redirects and time-qualified historical entities; four parallel water code systems and four parallel editions of sea limits; fifteen sibling ecological schemes with a deterministic, reversible identifier derivation; and published admission rules and limitations. The case studies demonstrated that these mechanisms interact consistently across domains. Seven of the ten sub-principles are fully implemented and three partially, with all partial ratings in the integrity and governance layers that were codified most recently.

The remaining tensions (stability versus consolidation, expressiveness versus usability, neutrality versus naming, plurality versus single-curator judgement, locality versus readability) belong to the principle rather than to its implementation. They mark the points at which every framework that harmonises without replacing must make explicit, documented choices. HLC already makes most of them visible; the reference model, the RDF mapping and the validation rules proposed in Chapter 11 indicate how they can be made verifiable.

The broader contribution is a demonstration that pluralism in classification is not merely tolerable but can be engineered. A single curator, using deliberately simple tools, has built a framework in which forty and more authorities coexist without one silencing another, because the framework adds identity, ownership and attribution to them rather than replacing them. That is the meaning of harmonization without replacement, and it is transferable well beyond geography.

Appendix A: Formal Grammar of the HLC Notation

The grammar below reconstructs the notation used in codes and tags across the HLC files in Extended Backus–Naur Form. It is descriptive: it captures the forms observed in version 8.1 and the meaning assigned to them by the front matter and the INFO sheets.

alpha = "A" | ... | "Z" ; digit = "0" | ... | "9" ; alnum = alpha | digit ;

hlc2 = alpha alpha | alpha digit | digit alnum | iso2 "-" alnum { alnum } ;

logical = [alnum alnum] ":" [alnum alnum] ;

hlcm = digit alnum alnum [alnum] [range | share | optzero] [""] ;*

range = "!-" digit digit ; share = "~" | "+" | "-" | "+-" ; optzero = "(0)" ;

alternatives = code { "|" code } ;

redirect = code ">" (code | "(" code ")") ;

tag = "{" key ":" value "}" ; key = alnum alnum ; value = ["(" text ")"] text ;

claim = "claimed by " iso2 ["!"] { "," iso2 ["!"] } ;

path = hlcm { "-" hlcm } [" (" levelnumber ")"] ;

Table 29: Semantics of the grammar productions

Production	Semantics	Sub-principle
alternatives	All members denote the same referent; the first is preferred	P1, P2
redirect	Left side is deprecated; right side is its successor; parentheses denote a successor that is itself a code of another layer or a plurality	P6
logical	Entry in the logical/organisational namespace; (E)/(W) split an authority into terrestrial and aquatic roles	P5, P8
hlcm with range	Identifier reserving a block for internal structure	P5
hlcm with share	Identifier sharing numeric space with neighbouring entries	P5
hlcm with *	Secondary placement; attributes at primary placement only	P3, P4
tag	Assertion by scheme key; parenthesised value marks inherited, former or approximate	P1, P2
claim	Contested status; ! marks the administering party	P2
path	Placement as sequence of identifiers; parenthesised level number gives original position of a duplicate	P3, P7

Appendix B: HLC-M Identifier-Range Allocation

Reproduced and condensed from the HLC-M sheet of codehl (v8.1); Figure 9 visualises the numeric part. Patterns follow the front-matter notation: a = alpha, n = numeric, x = alphanumeric; quoted characters are literal.

Table 30: HLC-M range allocation by owner

Range / pattern	Owner	Content
‘0’–‘9’ anx	HW	Ecoregions and ecosystems, leaf codes derived from other schemes
‘0’–‘9’ aax	HW	Ecoregions and ecosystems, branch codes
‘0’–‘9’ aa	HY	Waters, oceans, seas (CWWN+ codes incl. HLC extensions)
‘0’–‘9’ … an	External	External numeric codes (e.g. GeoNames) with penultimate digit replaced by P–Y
07–09 n; 53–54 n(x); 891–896	–	Currently unused
01n + 1nn(x)	HL	Historical, divisions, small islands, countries: Asia
02n + 2nn(x)	HL	Historical, divisions, small islands, countries: Africa
03n + 3nn(x)	HL	Historical, divisions, small islands, countries: Europe
04n + 4nn(n)	HL	Historical, divisions, small islands, countries and mainland waters: Northern America
40AA–49ZZ	HY	Waters of Northern America derived from CWWN+
05n + 50–52 n(x)	HL	Historical, divisions, small islands, countries: Southern America
55–59 nn; 56AA–59ZZ	HW	Terrestrial and biogeographic sub-ecoregions and sub-realms (branch level)
600–611 n	HY	Mainland waters (excluding Northern America)
06n + 612–687 (x)	HL	Countries etc.: Australasia, Oceania, Antarctica
688–699 n	HW	Realms, ecosystems, ecoregions, zones (branch level)
69AA–69ZZ	HY	Mainland waters derived from CWWN+
700–796 (n); 70AA–78ZZ	HY	Waters, oceans, seas (incl. CWWN+-derived)
700–796 a	HW	Water-based realms, ecosystems, ecoregions, zones
797–810 (x)	HW	Realms, ecosystems, ecoregions, zones (tree level)
811–819; 82n; 83n; 84n; 85n; 86n	HL	Sub-structures: Asia; Africa; Europe; Northern America; Southern America; Australasia, Oceania, Antarctica
87n	HL	World organisations
880–890	HL	unions, cross-continent (incl. transcontinental states)
897–899	HL	Marine ecosystems and ecoregions (tree level)
90AA–99ZZ	HW	Marine ecoregions and ecosystems (leaf level)
90n	HL	Waters and oceans (world, secondary range for sorting)
91n–96n	HL	Main structures and epochs: Asia, Africa, Europe, Northern America, Southern America, Australasia and Oceania
970–976; 977–978; 979	HL	Antarctica and adjacent territories; tectonic plates; Outlying Oceania
98n; 99n	HL	Continental spanning structures; historical world epochs

Appendix C: Inline Attribute Tag Keys (Selection)

Selected tag keys as documented in the INFO sheets of codehl and codehy. The selection covers the keys used in the examples of this thesis; Figure 4 shows their frequency in codehl.

Table 31: Selected inline attribute tag keys

Key	Scheme	Column / file
UN	UN M49 (incl. old, invalid, inherited; GENC extensions > 900)	UNO|M01 (codehl)
IL	ILO index of countries, territories and areas	UNO|M01
OA	ICAO airport prefix code	UNO|M01
CW	Correlates of War (Gleditsch & Ward variant)	TERRA|M02
PD	WGS-RPD botanical units	TERRA|M02
SX	SDMX+ incl. IMF codes	TERRA|M02
BW / TW	BEOW / TEOW (HLC scheme adaptation)	TERRA|M02
FA	FAO+ (original FAO in brackets)	AQUA|M03
FW / MW / OW	FEOW / MEOW / GOODS	AQUA|M03; OTHERS (codehy)
WN	CWWN+	AQUA|M03
TC / WC / ID	US TIC / World Area Codes / World Inequality Database	M04
R0–R3 / ED	ITU region and dial codes / ESSD groups	M05
NT / EU	NATO country codes / EU Geonomenclature	M06
AU / NZ	SACC / NZSCC	M07
DE / IT / UK	Destatis / ISTAT and MIN / UK regions	M08
AR / ZA / JP	INDEC / Stats SA / Trade Statistics of Japan	M09
CH / FR / NO	Swiss FSO / INSEE / Statistics Norway	M10
CA, DR, HI, AO, OD, RU, US, WG, WI	SCCAI; CLDR; historical codes; ICAO; opendatasoft; OKSM; GENC; WGS84 anchor; WIPO	OTHERS
NF / IE	NAFO divisions / ICES areas	FAO & related (codehy)
CM, FM, IW, HC, LH, OP, US	CCM2; GFCM; GIWA; HELCOM; Longhurst; OSPAR; USGS HUC	OTHERS (codehy)

Appendix D: Glossary

Table 32: Glossary of key terms

Term	Definition in this thesis
Assertion	A statement attributable to an authority that assigns a code, parent, name or status to a referent
Attributed neutrality	Neutrality achieved by attributing every contested placement to its source or by using a declared non-sovereign placement mode
Attribute locality	Rule that descriptive attributes reside only at an entry's primary placement
Bounded plurality	Admission of alternative hierarchies only under published criteria of existence, relevance and non-derivability
Coordinated Plurality	The evolved principle of harmonization without replacement (Section 4.8)
Coordinating layer	Identifiers, placements, correspondences and rules that connect schemes without replacing them
Embedded case study	A case study with several units of analysis nested within the single case (Yin 2018)
Entry	A record in the coordinating layer representing one referent
Fiat object	A unit whose boundaries exist by human decision
Placement mode	The kind of claim a placement represents: de jure, de facto, neutral container, multi-placement
Polyhierarchy	A hierarchy in which an entry may have several parents
Scheme	An independently maintained classification or code list with its own mandate
Sibling view	A scheme held in parallel with others without master or subset relation

Bibliography

Primary Sources

Weynans, J. (2026a). *HLC Operational Framework – Common Front Matter*, version 8.1.2, 23 October 2026 (codehv).

Weynans, J. (2026b). *Hierarchical Location Codes: Namespace and Code-Ownership Control* (codehu), version 8.1.

Weynans, J. (2026c). *Hierarchical Location Codes: Land-Based Master Hierarchy* (codehl), version 8.1.

Weynans, J. (2026d). *Hierarchical Location Codes: Water and Hydrographic Hierarchy* (codehy), version 8.1.

Weynans, J. (2026e). *Hierarchical Location Codes: Ecoregions and Ecosystems* (codehw), version 8.1.

Methodological Literature

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27–40. https://doi.org/10.3316/QRJ0902027

Brannick, T., & Coghlan, D. (2007). In defense of being "native": The case for insider academic research. *Organizational Research Methods*, 10(1), 59–74. https://doi.org/10.1177/1094428106289253

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532–550. https://doi.org/10.2307/258557

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2), 219–245. https://doi.org/10.1177/1077800405284363

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105. https://doi.org/10.2307/25148625

Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. https://doi.org/10.1177/1049732305276687

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. Beverly Hills, CA: Sage.

Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods* (6th ed.). Thousand Oaks, CA: Sage.

Theoretical Literature

Berners-Lee, T. (1998). *Cool URIs don't change*. W3C Style Guide. https://www.w3.org/Provider/Style/URI

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

Dupré, J. (1993). *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.

Foucault, M. (1970). *The Order of Things: An Archaeology of the Human Sciences*. London: Tavistock (original work *Les mots et les choses*, Paris: Gallimard, 1966).

Goodman, N. (1978). *Ways of Worldmaking*. Indianapolis: Hackett.

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2), 199–220. https://doi.org/10.1006/knac.1993.1008

Harley, J. B. (1989). Deconstructing the map. *Cartographica*, 26(2), 1–20. https://doi.org/10.3138/E635-7827-1757-9T53

ISO 25964-1:2011. *Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval*. Geneva: ISO.

Lenzerini, M. (2002). Data integration: A theoretical perspective. In *Proceedings of the 21st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS 2002)* (pp. 233–246). https://doi.org/10.1145/543613.543644

Ranganathan, S. R. (1933). *Colon Classification*. Madras: Madras Library Association.

Ranganathan, S. R. (1967). *Prolegomena to Library Classification* (3rd ed.). Bombay: Asia Publishing House.

Smith, B., & Varzi, A. C. (2000). Fiat and bona fide boundaries. *Philosophy and Phenomenological Research*, 60(2), 401–420. https://doi.org/10.2307/2653492

Star, S. L., & Griesemer, J. R. (1989). Institutional ecology, 'translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907–39. *Social Studies of Science*, 19(3), 387–420. https://doi.org/10.1177/030631289019003001

W3C (2009). *SKOS Simple Knowledge Organization System Reference*. W3C Recommendation, 18 August 2009.

W3C (2013). *PROV-O: The PROV Ontology*. W3C Recommendation, 30 April 2013.

Integrated Schemes and Standards (as referenced by HLC)

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. https://doi.org/10.1641/B580507

Bailey, R. G. (1989). Explanatory supplement to ecoregions map of the continents. *Environmental Conservation*, 16(4), 307–309. https://doi.org/10.1017/S0376892900009711

Bailey, R. G. (1998). *Ecoregions: The Ecosystem Geography of the Oceans and Continents*. New York: Springer. 2nd ed. 2014, https://doi.org/10.1007/978-1-4939-0524-9

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

Dinerstein, E., Olson, D., Joshi, A., et al. (2017). An ecoregion-based approach to protecting half the terrestrial realm. *BioScience*, 67(6), 534–545. https://doi.org/10.1093/biosci/bix014

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

Gill, M. J., Crane, K., Hindrum, R., et al. (2011). *Arctic Marine Biodiversity Monitoring Plan (CBMP-Marine Plan)*. CAFF Monitoring Series Report No. 3. Akureyri: CAFF International Secretariat.

Gleditsch, K. S., & Ward, M. D. (1999). A revised list of independent states since the Congress of Vienna. *International Interactions*, 25(4), 393–413. https://doi.org/10.1080/03050629908434958

Goodall, D. W. (Ed.-in-chief) (1977–). *Ecosystems of the World* (multi-volume series). Amsterdam: Elsevier.

International Hydrographic Organization (1953). *Limits of Oceans and Seas* (Special Publication S-23, 3rd ed.). Monaco: IHB. Draft 4th ed. 1986; draft 2002.

ISO 3166-1, ISO 3166-2, ISO 3166-3. *Codes for the representation of names of countries and their subdivisions*. Geneva: ISO.

ISO 17369:2013. *Statistical data and metadata exchange (SDMX)*. Geneva: ISO.

Keith, D. A., Ferrer-Paris, J. R., Nicholson, E., et al. (2022). A function-based typology for Earth's ecosystems. *Nature*, 610(7932), 513–518. https://doi.org/10.1038/s41586-022-05318-4

Longhurst, A. R. (2007). *Ecological Geography of the Sea* (2nd ed.). Amsterdam: Academic Press.

Maureaud, A. A., Reygondeau, G., Ingenloff, K., Vigneron, J. G., Watling, L., Winner, K., & Jetz, W. (2023). A global biogeographic regionalization of the benthic ocean. *OSF Preprints*. https://doi.org/10.31219/osf.io/nkjvf

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. https://doi.org/10.1641/0006-3568(2001)051[0933:TEOTWA]2.0.CO;2

Olson, D. M., & Dinerstein, E. (2002). The Global 200: Priority ecoregions for global conservation. *Annals of the Missouri Botanical Garden*, 89(2), 199–224.

PAME (2013). *Large Marine Ecosystems (LMEs) of the Arctic Area: Revision of the Arctic LME Map* (2nd ed.). Akureyri: Arctic Council PAME. Factsheet series 2018.

Sherman, K., & Alexander, L. M. (Eds.) (1989). *Biomass Yields and Geography of Large Marine Ecosystems*. AAAS Selected Symposium 111. Boulder: Westview Press.

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. https://doi.org/10.1641/B570707

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. https://doi.org/10.1016/j.ocecoaman.2011.12.016

United Nations Statistics Division. *Standard Country or Area Codes for Statistical Use (M49)*. New York: United Nations.

UNEP. *Global International Waters Assessment (GIWA)*. Kalmar/Nairobi: UNEP, from 1999.

US Defense Intelligence Agency. *DIAM 65-18: Geopolitical Data Elements and Related Features*. Washington, DC.

US National Geospatial-Intelligence Agency. *Geopolitical Entities, Names, and Codes (GENC) Standard*. Washington, DC.

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.