

AP-011 — Laboratorios Alexandria

# The Epistemic Desert Map: Identifying Unexplored Cross-Domain Research Frontiers

*Laboratorios Alexandria*

contact@laboratoriosalexandria.com

June 2026

## Abstract

Maps are valued not only for what they show but for what they leave blank. We present a systematic identification of epistemic deserts—regions of the cross-domain scientific landscape where verified connections between disciplines are absent or near-absent—derived from a knowledge base of over 360,000 scientific documents organized across a 302-subfield taxonomy. Using a corpus of 5,351 cross-domain correlations distributed across 64 domain pairs, we identify 14 pairs with a single correlation and classify them using a three-type taxonomy: verified deserts, unexplored deserts, and structurally distant deserts. We present the empirical desert map alongside criteria for prioritizing exploration of specific desert regions, and argue that the desert map is a research planning instrument that tells us where the unexplored frontiers are, which of them are most likely to yield discoveries, and which are sterile boundaries not worth crossing.

## 1 Introduction

The companion paper in this series (AP-006) demonstrated that cross-domain fertility is non-uniformly distributed: certain domain pairs produce orders of magnitude more epistemic correlations than others. That paper looked upward—at the peaks of the fertility landscape. This paper looks downward—at the valleys, the gaps, the regions where no connections have been found.

We call these regions epistemic deserts. The term is chosen deliberately. A desert is not necessarily lifeless—it is unexplored or inhospitable. An epistemic desert between two scientific domains may indicate that the domains have nothing to teach each other (a genuine boundary), or it may indicate that no one has looked (an unexplored frontier). Distinguishing between these two interpretations is the central methodological challenge of this paper.

The practical stakes are significant. If a desert represents a genuine boundary, then research programs targeting that domain pair will fail. If it represents an unexplored frontier, then it is precisely the site where the next cross-domain breakthrough is most likely—because it is the territory that existing research programs have systematically missed.

## 2 Identifying Epistemic Deserts

### 2.1 The Taxonomy

The knowledge base is organized using a hierarchical taxonomy of 4 domains, 26 fields, and 302 subfields, aligned with the OpenAlex classification system. This taxonomy provides the coordinate system for mapping cross-domain connections and identifying gaps. At the field level (26 fields), there are 325 unique pairwise combinations. At the subfield level (302 subfields), there are 45,451 unique combinations. The vast majority of these combinations have zero verified correlations—not because they have been evaluated and found empty, but because the system has not yet explored them.

### 2.2 Desert Classification

We classify epistemic deserts into three categories based on available evidence:

**Type I: Verified deserts.** Domain pairs that have been systematically explored and have produced zero or near-zero correlations despite adequate representation of both domains in the knowledge base. These are candidates for genuine disciplinary boundaries.

**Type II: Unexplored deserts.** Domain pairs with zero or near-zero correlations where at least one domain has limited representation in the knowledge base. These are not verified absences—they are gaps in coverage. The absence of correlations reflects the absence of data, not the absence of connections.

**Type III: Structurally distant deserts.** Domain pairs where the epistemic distance between fields is so great that correlations are unlikely to be detected by current methods. These pairs may contain genuine connections that require theoretical frameworks not yet available—or they may be genuinely independent. The desert itself cannot tell us which.

## 3 The Empirical Desert Map

The following table presents the 14 domain pairs at the tail of the fertility distribution—each with a single verified correlation—along with a preliminary desert type classification:

Domain Pair	Cor r.	Type	Notes
Materials Sci. × Overfitting Registry	1	Type I	Operational × scientific
Life Sci. × Market Regime	1	Type II	Under-sampled intersection
Human Behavior × AI Security/Privacy	1	Type II	Emerging field pair
Ubiquitous Computing × Human-Computer Int.	1	Type III	Sub-field overlap expected
Computation/AI × Materials Sci.	1	Type I*	Asymmetric duplicate (222 inverse)
Computation/AI × Operational	1	Type I	Operational × scientific

Daily			
Computation/AI × Public Health	1	Type II	Surprising given COVID-era work
Energy × Validity of Conclusions	1	Type III	Meta-epistemic × applied
Space/Astrophys. × Knowledge Frontiers	1	Type III	Meta-epistemic × applied
Space/Astrophys. × Validity of Conclusions	1	Type III	Meta-epistemic × applied
Knowledge Frontiers × Market Regime	1	Type II	Under-sampled intersection
Human-Computer Int. × Human Behavior	1	Type I*	Asymmetric duplicate (2 inverse)
Operational Daily × Overfitting Registry	1	Type I	Operational × operational
AI Ethics × Human Behavior	1	Type I*	Asymmetric duplicate (5 inverse)

*Table 1. The 14 singleton domain pairs in the correlation corpus. Type I\* indicates pairs whose low count is an artifact of detection order asymmetry (the inverse pair has substantially more correlations).*

Several patterns emerge from the empirical desert map. First, three of the fourteen singletons (marked Type I\*) are artifacts of detection order asymmetry—the system detected correlations in one direction (e.g., Materials Science → Computation/AI, 222 correlations) but only once in the reverse direction (Computation/AI → Materials Science, 1 correlation). These are not genuine deserts; they are measurement artifacts that would disappear under symmetric pair consolidation.

Second, several singletons involve intersections between scientific domains and operational or meta-epistemic categories (Overfitting Registry, Operational Daily Activity, Validity of Conclusions). These reflect the boundaries between knowledge about the world and knowledge about the system that studies the world—a distinction that is genuinely structural rather than artifactual.

Third, the Type II deserts are the most actionable for research planning. Computation/AI × Public Health has only one correlation despite the massive growth of computational epidemiology. This is likely an under-sampling problem: the knowledge base may not yet contain sufficient computational public health literature. Similarly, Life Sciences × Market Regime and Human Behavior × AI Security/Privacy are intersections where active research exists but has not yet been captured by the detection system. These Type II deserts are priority candidates for knowledge base expansion.

## 4 The Scale of the Unexplored

The 14 singleton pairs represent the visible deserts—pairs that have at least been touched by the detection system. The far larger population of deserts is invisible: subfield pairs with zero correlations that have never been explored at all.

At the subfield level, 45,451 unique pairwise combinations exist. The system has detected correlations in a small fraction of these. The remainder—tens of thousands of subfield pairs—are epistemically unexplored. Even at the field level, the 5,351 correlations are concentrated in a small number of pairs: the top three pairs account for 54.5% of all correlations (AP-006). The long tail contains dozens of field pairs with minimal or no verified connections.

The practical implication is that the current map of cross-domain knowledge is heavily biased toward well-explored territory. The Life Sciences  $\times$  Computation axis and the Materials  $\times$  Life nexus are well-mapped because they are naturally fertile and because institutional research programs have invested heavily in them. The deserts remain terra incognita.

## 5 Criteria for Desert Exploration Priority

Not all deserts are equally worth exploring. We propose four criteria for prioritizing exploration of specific epistemic deserts:

Criterion 1: Adjacent fertility. If a desert exists between two domains that are each highly fertile with a common third domain, the desert may represent an unexplored bridge. For example, both Computation/AI (1,456 correlations with Life Sciences) and Energy/Propulsion (145 correlations with Life Sciences) are independently fertile with Life Sciences. The Computation/AI  $\times$  Energy/Propulsion pair has 203 correlations, suggesting that the adjacent fertility criterion successfully predicts moderate-to-high connectivity—but the prediction should be tested on less-explored pairs.

Criterion 2: Methodological compatibility. Domains that share experimental methods, mathematical formalisms, or data types are more likely to have undiscovered connections than domains with incommensurable methodologies. A desert between two domains that both use network analysis, for instance, is more promising than a desert between a domain that uses wet-lab experimentation and one that uses pure mathematical proof.

Criterion 3: Temporal convergence. Domains that are independently trending toward similar questions or methods may be approaching a convergence point. A desert between two converging domains is a high-priority exploration target because the convergence suggests that connections may be forming that have not yet been detected.

Criterion 4: Practical demand. Some deserts have practical significance regardless of theoretical promise. If a policy problem requires integrating knowledge from two domains that have no verified connections, the desert represents a practical research priority that justifies exploratory investment even if the theoretical probability of finding connections is low.

## 6 Deserts in the Context of the Companion Papers

The epistemic desert map connects directly to the findings of three companion papers in this series, forming a coherent analytical framework for cross-domain discovery.

AP-005 (Verified Negative Isomorphisms) argues that rigorous demonstrations of the absence of cross-domain correspondence are first-order epistemic results. Some deserts, upon exploration, will turn out to contain verified negatives—confirmations that no structural

correspondence exists between the domains. The desert map identifies where to look for these negatives; AP-005 provides the framework for evaluating them.

AP-006 (The Materials-Life Nexus) demonstrated that cross-domain fertility is non-uniformly distributed. The desert map is the complement of that fertility distribution: the peaks in AP-006 are the well-watered regions; the deserts are the valleys. Together, they provide a complete topographic map of the cross-domain landscape. Critically, AP-006's finding that fertility does not uniformly predict quality adds nuance to desert prioritization: a desert adjacent to a high-fertility, high-quality pair (Materials Science × Life Sciences, with 25.8% Grade A+B rate) is more promising than one adjacent to a high-fertility, low-quality pair (Life Sciences × Markets, with 75% Grade D).

AP-010 (The Maturity Gradient) demonstrated that cross-domain discoveries follow a quantifiable attrition funnel (2.1% survival from initial evaluation to validated status). This attrition rate provides a baseline expectation for desert exploration: if a desert exploration campaign detects 100 candidate correlations, approximately 2 will survive to validated status. This expectation calibrates the investment required for productive desert exploration.

## 7 Sources of Uncertainty and Limitations

The desert map is limited by the knowledge base from which it is derived. Domains with fewer documents will have more Type II (unexplored) deserts, regardless of their genuine connectivity. Expanding the knowledge base will reclassify some deserts as fertile territory.

The 302-subfield taxonomy, while comprehensive, imposes a particular structure on the organization of knowledge. Different taxonomies would produce different desert maps. The map is relative to the coordinate system, not absolute.

The criteria for exploration priority (Section 5) are proposed but not empirically validated. Testing these criteria would require systematically exploring deserts and measuring the yield—an experiment that has not yet been conducted at sufficient scale.

The distinction between Type I (verified) and Type II (unexplored) deserts requires judgment about what constitutes adequate exploration. We do not provide a formal threshold for this distinction, acknowledging that the boundary between “we looked and found nothing” and “we haven't looked enough” is inherently fuzzy.

The detection order asymmetry (three singletons that are artifacts of pair direction) is a technical limitation that should be addressed in future work by consolidating symmetric pairs. This consolidation would reduce the 64 observed pairs to fewer unique combinations and would eliminate some artificial deserts.

## 8 Conclusion

The map of cross-domain scientific knowledge has vast blank regions—epistemic deserts where connections between disciplines are absent or unexplored. Of the 64 domain pairs in the current corpus, 14 contain a single correlation, and tens of thousands of subfield pairs remain entirely

unexplored. These deserts are not merely gaps in our knowledge; they are the frontier of cross-domain discovery.

By identifying, classifying, and prioritizing these deserts, we transform the absence of knowledge into a research planning instrument. The empirical desert map (Table 1) distinguishes genuine boundaries from measurement artifacts and unexplored frontiers, providing actionable guidance for knowledge base expansion and targeted exploration. Some deserts will yield discoveries. Others will yield verified negatives that delimit the boundaries of useful cross-domain research. Both outcomes are valuable.

The worst outcome is the one that currently prevails: deserts that remain unexplored because the research community's attention is concentrated on already-fertile territory. The epistemic desert map is an invitation to look where no one has looked, with the understanding that what we find—or confirm that we cannot find—will be informative in either case.

*Intellectual property for all systems, methods, and protocols described in this paper is proprietary to Laboratorios Alexandria. All rights reserved.*

## References

- [1] Uzzi, B., Mukherjee, S., Stringer, M. & Jones, B. (2013). Atypical Combinations and Scientific Impact. *Science*, 342(6157), 468–472.
- [2] Porter, A. L. & Rafols, I. (2009). Is Science Becoming More Interdisciplinary? *Scientometrics*, 81(3), 719–745.
- [3] Boyack, K. W. & Klavans, R. (2010). Co-Citation Analysis, Bibliographic Coupling, and Direct Citation. *JASIST*, 61(12), 2389–2404.
- [4] Chen, C. (2006). CiteSpace II: Detecting and Visualizing Emerging Trends. *JASIST*, 57(3), 359–377.
- [5] Leydesdorff, L. & Rafols, I. (2011). Indicators of the Interdisciplinarity of Journals. *Scientometrics*, 87(2), 267–292.
- [6] Merton, R. (1973). *The Sociology of Science*. University of Chicago Press.
- [7] Kuhn, T. (1962). *The Structure of Scientific Revolutions*. University of Chicago Press.
- [8] Popper, K. (1959). *The Logic of Scientific Discovery*. Routledge.