SolvxAI
Data to Decision Powered by AI

SOLVXAI WHITE PAPER SERIES | NO. 05 | MEMORY & ONTOLOGY

The Memory That Knows What a Well Is

Why general-purpose memory cannot hold engineering knowledge, and what has to replace it

Published by the SolvxAI Research Team
July 2026 · Edition 1.0
Contact: sd@nexaconsultinc.com

IN BRIEF

A single physical well answers to half a dozen names, sits inside a hierarchy, and changes what it means as it is re-interpreted. This paper assembles the evidence on all three problems and reaches a conclusion that is less flattering to structured approaches than we expected. The industry's own purpose-built well identifier is described, by the body that maintains it, as one of the primary causes of misintegrated well data. Meanwhile the published comparisons show that adding structure to a retrieval system is not reliably an improvement — on one evaluation, ontology structure without the underlying text performed far worse than plain vector search. We argue that domain memory is necessary and that structure alone is not sufficient, and we set out what an operator should actually require.

Contents

1. Executive summary

The hard problem in engineering memory is not storage or recall. It is deciding that two records describe the same well.

Three findings frame this paper, and the third is the one that stops it being a sales argument for knowledge graphs:

- **Well identity is a documented, unsolved industry problem.** The association that maintains the North American well-numbering standard reports that the identifier is simultaneously the chief means of linking well data and "one of the primary causes of missing or incorrectly integrated well data." [1] Its own survey documents cores and logs attached to the wrong wellbore because different operators and vendors created different identifiers for the same hole. [1]
- **Retrieval by similarity cannot answer the questions engineers ask.** Aggregation over text requires finding *all* the evidence rather than some of it, and published work states plainly that both database-query and retrieval-augmented approaches fail to achieve that completeness. [5] Multi-hop questions fail even when the supporting evidence is supplied. [6]
- **But structure is not automatically better, and we will not claim it is.** In a systematic comparison, graph-based and conventional retrieval were *complementary, with no consistent winner*, and graph construction cost roughly forty times more. [7] On a mathematics textbook, conventional retrieval beat the graph approach outright. [9] And in one evaluation, an ontology graph **without** the underlying source text scored 15% against 60% for plain vector search. [8]

The conclusion we reach is therefore narrower than the one we set out to reach. Domain memory is necessary because identity, relationship and time are real problems that general memory does not address. It is not sufficient, and an ontology that floats free of the evidence underneath it is worse than no ontology at all.

> **A CLAIM WE WITHDREW**
>
> We began this paper intending to argue that the petroleum data standards — the well-numbering standard, the exchange formats, the open data platform — exist because generic data models were tried and found inadequate. We could not substantiate it. Every stated rationale we were able to verify describes **heterogeneity between vendors, applications and regulators**, not the inadequacy of general models. The argument may still be true; it is not something we can cite, so we do not assert it.

2. The identity problem

Start with the problem that has to be solved before any of the others matter.

2.1 One well, many names

A well is referred to by the name the geologist uses, the name the production engineer uses, a regulatory identifier, a shortened form on a plot axis, and whatever a vendor's export decided to call it. These are not errors. They are legitimate names for one physical hole.

The body that maintains the North American well identification standard investigated this directly, interviewing thirteen regulators, fifteen operators, three service companies and three vendors. Its finding on why stakeholders carry more than one identifier is that there is more than one *context*: a drilling permit against a production report, one corporate system against another, in-house against external communication, geoscience analysis against management reporting. [1]

2.2 The identifier does not settle it

It is tempting to assume the regulatory number resolves this. The same report documents why it does not. Among stakeholders using the standard number, some digits are applied inconsistently, with regulators omitting or zero-padding them and a few using the county segment differently; vendors and operators consequently create look-alike identifiers to harmonise across states. Some operators modify an identifier from one system for use in another — stripping a twelve-digit number to ten for a financial system — requiring special rules to reverse. [1]

The consequences are stated concretely: cores and logs cannot be attached to the correct wellbore without considerable manual effort, or are attached to the wrong wellbore, explicitly because different operators or vendors created different identifiers for the same wellbore. Regulators' own ranked list of issues includes duplicate identifiers assigned to different wells. [1]

THE SAME CONCEPT, SIX JURISDICTIONS, SIX INCOMPATIBLE STRUCTURES

Colorado	**05-123-23881**	
Texas	**42-001-32559**	— sidetrack not identified
Alberta	**100/15-35-067-11W6/00**	— well not identified
Norway	**6407/9-A-4**	
United Kingdom	**211/26a-P4**	
Brazil	**3-ESSO-4-SPS**	

Reported by the PPDM Well Identification Project, 2012 [1].

Figure 1. The same concept expressed in six jurisdictions. Two of these structures do not identify a sidetrack or the parent well at all, so the information required to link a record to the right hole is absent from the identifier itself.

Canada's experience is documented separately and independently. The 1965 unique well identifier was found to have three specific limitations: it is **not permanent**, it identifies well *events* rather than the parent well, and it does not distinguish drilling events from completion events. [2] Amendment frequency grew roughly **300% over fifteen years**, with a review-and-amendment lag of several weeks after drilling

finished — named in the source as a data-quality problem in its own right. [2]

> **The industry's own purpose-built identifier, mandated and decades old, still does not reliably identify a well. A general-purpose memory system is not going to do better by accident.**

2.3 This is a named problem with a literature

Deciding whether two records refer to the same real-world thing is a long-studied problem — record linkage, or entity resolution. The authoritative review states the reason exact-match keys are insufficient in terms that transfer directly: exact matching links records only when they agree on all common attributes, which degrades as soon as textual variables are involved, and a designated unique key **can itself be duplicated**. [3] The founding 1959 case had no unique identifier at all; the insight was that no single field was reliable but that together they supported probabilistic linkage. [3]

The review also names an assumption that matters here: filtering candidate matches on fields assumed to agree amounts to "a deterministic a priori judgment that these fields are error-free." [3] For well data that judgement is demonstrably wrong.

Modern methods do not dissolve the problem. In a peer-reviewed evaluation, matchers fine-tuned on one product domain lost between 22% and 61% of their score when transferred to unseen entities. [4] Entity resolution generalises poorly, which is precisely the regime an operator is in when a new asset arrives.

3. The relationship problem

The second failure is structural rather than about identity.

Ask what is known about a reservoir and the answer must be assembled from knowledge attached at several levels — some at the reservoir, most at the wells within it, some at the completions within those. Ask which offsets might be interfering and the answer depends on a spatial relationship that no document states in those words. Both are traversals of structure wearing the costume of a question.

The published work is direct about the limit. Aggregation over unstructured text requires exhaustive evidence collection — a system must "find all," not merely "find one" — and the authors state that both text-to-query and retrieval-augmented approaches fail to achieve that completeness. [5] On multi-hop questions, embedding models are reported inadequate at retrieving the evidence, and frontier models reason poorly over such questions *even when the supporting evidence is supplied* — so the failure is not purely retrieval. [6]

An engineering memory that can only return documents resembling the question will answer "what do we know about this reservoir" with whatever text most resembles the phrase, which is not the same thing and is not checkable.

4. The time problem

The third failure is that engineering knowledge is revised, and a store of current belief cannot represent having been wrong.

A volumetric estimate from an early study is superseded by a later re-interpretation. A store holding only the current value silently rewrites history. A store holding everything without temporal structure cannot say which value was live when a decision was signed. Neither supports an audit, which asks what was believed *then*.

This is recognised in the benchmarks rather than being our own framing. A peer-reviewed long-term memory benchmark treats **knowledge updates** as one of five first-class abilities, alongside multi-session and temporal reasoning — superseded facts are a category, not an edge case. [10] Its headline result is a **30% accuracy drop** for commercial assistants and long-context models on retaining information across sustained interaction. [10]

A second peer-reviewed benchmark, built because prior work had evaluated no more than five conversation sessions, reports that models struggle with long-range temporal and causal dynamics, and that retrieval augmentation improves matters but still lags human performance substantially. [11]

> **HOW WE GRADE THE MEMORY LITERATURE**
>
> Almost every search on agent memory returns vendor material first — companies selling memory products, writing about the need for memory products. We excluded all of it. The two anchors above are peer-reviewed conference papers; where we cite anything weaker in this paper, we say so at the point of citation.

5. Where the argument for structure gets uncomfortable

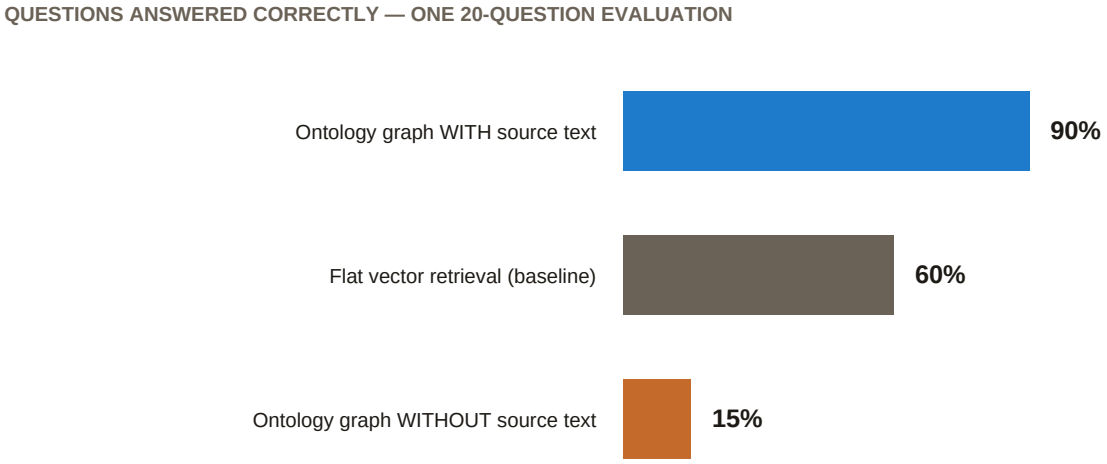
If structured memory were simply better, this section would not exist.

The honest state of the evidence is that adding structure to retrieval is not reliably an improvement, and we would rather set that out than have a reader discover it.

A systematic comparison of conventional and graph-based retrieval concluded that the two are **complementary, with no consistent winner**. Conventional retrieval won on single-hop questions; the graph approach won on multi-hop. On one dataset, 13.6% of questions were answered only by the graph system and 11.6% only by the conventional one — neither subsumes the other. Graph construction took roughly **forty times longer**. [7]

On a mathematics textbook, conventional embedding retrieval beat the graph approach on both retrieval accuracy and overall score, with the authors attributing the gap to the graph retrieving excessive and sometimes irrelevant content because of its entity-based structure. [9]

And the sharpest result is this one:



n = 20 questions. Far too small to be a rate — reported because the ORDERING is the finding, not the magnitudes. [8]

Figure 2. Ontology structure helps only when it stays attached to the source text. Detached from it, the same structure scored far below plain vector retrieval. The evaluation is small — twenty questions — so the ordering is the finding and the magnitudes should not be quoted as rates. [8]

An ontology-derived knowledge graph combined with the underlying text chunks answered 18 of 20 questions. Plain vector retrieval answered 12. The same ontology graph **without** the text answered 3. [8] Symbolic structure alone was substantially worse than no structure at all.

We report the sample size prominently because twenty questions cannot establish a rate. What it does illustrate is a mechanism that matches the other two results: structure that summarises away its evidence loses more than it gains.

THE CRITICISM WE THINK IS FAIR

The standing academic critique of ontology-driven systems is not that they do not work. It is that there is **no single correct way to model a domain** and no global truth — different stakeholders hold different semantics, and the approach is brittle to differing views — and that the cost of publishing and maintaining structured content is prohibitive. [12] Anyone who has watched an industry argue about what counts as one well will recognise the first objection immediately.

5.1 What we could not find

We looked specifically for a peer-reviewed, adequately powered study showing that a domain ontology measurably improves retrieval or reasoning in an engineering or subsurface setting. **We did not find one.** The strongest positive evidence available is preprint-grade, small-sample, and from general domains.

That is a genuine gap in the record, and it means the case for domain memory in upstream currently rests on the identity, relationship and time arguments in sections 2 to 4 — which are well documented — rather than on a demonstrated retrieval benchmark in this field, which does not yet exist.

6. What follows

Putting the two halves together gives a more specific requirement than “use a knowledge graph.”

- **Identity must be resolved, and the resolution must be inspectable.** Since no identifier is sufficient [1][2] and exact matching provably fails on textual attributes [3], a system has to decide identity as a judgement and then be able to show its working — which names it merged, and why. A false merge silently corrupts the knowledge of two wells at once.
- **Structure must stay attached to evidence.** This is the clearest practical lesson in the whole register [8][9]. A digest that cannot be followed back to the artefact that produced it is the configuration that underperformed plain search.
- **Relationships must be represented as relationships.** Containment and adjacency are properties of the physical world, not of the prose, and no amount of similarity produces them [5][6].
- **Time must be first-class, including disagreement.** Superseded findings must be retained rather than overwritten, and two sources that disagree must be representable as disagreeing — because a store that can only hold settled belief will resolve conflicts by recency, and recency is not evidence. [10]
- **The cost has to be justified per question, not in principle.** Given the roughly forty-fold construction cost [7] and the absence of a consistent winner, structure earns its place on the questions that need traversal and aggregation — and a system should be honest that on simple lookups it may be buying nothing.

7. Questions worth asking

- **1.** When the same well appears under three different names in my data, what does the system do — and can it show me why it decided they were the same?
- **2.** Can every stored claim be followed back to the document that produced it?
- **3.** When a finding is superseded, what happens to the old one? Can the system tell me what it believed on the date a decision was taken?
- **4.** When two studies disagree, does the system represent the disagreement or pick one?
- **5.** Ask for something requiring completeness — every well in this unit meeting a criterion. Does the answer come with a count of what it examined?
- **6.** What does the memory cost to maintain per study, and what is paid on every query whether or not it is used?
- **7.** Is any of this demonstrable on my data, rather than described?

8. Limits of this paper

Stated in the same spirit as the rest of the series.

The identity evidence [1][2] is strong, specific and industry-authored, but one of the two documents has a co-author employed by a commercial well-data vendor, and both are association working papers rather than peer-reviewed research. The reference addresses printed in one of them are now dead links, which is itself a small illustration of the durability problem.

Most of the retrieval and memory evidence in sections 3 to 5 is preprint grade. The exceptions — the two memory benchmarks [10][11], the entity matching evaluation [4], and the Semantic Web critique [12] — are peer-reviewed and are the load-bearing citations where we rely on them.

Figure 2 rests on a twenty-question evaluation. We report it because the ordering is consistent with two independent results, not because twenty questions establish anything on their own.

And, as stated in section 1, we withdrew our original framing about the petroleum data standards when we could not verify it.

9. Conclusion

A memory that does not know what a well is cannot hold engineering knowledge. It can only hold text that mentions wells.

The three problems are real and documented. A physical well answers to many names, and the industry's own identifier does not settle which. The questions engineers ask are traversals and aggregations that similarity search cannot perform. Knowledge is revised, and a decision has to be defensible against what was known at the time.

What does not follow is that structure alone fixes any of it. The published comparisons are mixed, the construction cost is real, and detaching structure from its evidence made things measurably worse. The requirement is narrower: identity resolved and inspectable, structure anchored to provenance, relationships represented as relationships, and time carried honestly.

That is buildable, and it is specific enough to be tested. The right question to put to any vendor, including us, is the seventh one in section 7.

How SolvxAI has already resolved this

The requirement list in section 6 is not hypothetical. It is a description of SolvxAI's Reservoir Decision Memory, built for exactly the three problems this paper documents.

Knowledge is held on a domain ontology of the things an asset actually consists of — field, reservoir, formation, lease, well, wellbore, completion, operator — not on documents that mention them. Identity is resolved as a judgement made against everything already known, with every observed alias retained, so the decision that two records describe one well is inspectable rather than assumed from an identifier this paper shows cannot carry that weight. Every stored fact stays linked to the artefact that produced it — the configuration that failed in Figure 2 is the one we did not build. And the record is honest about time: superseded findings are retained rather than overwritten, and two sources that disagree are held as a disagreement, not resolved by recency.

Question 7 in section 7 has a short answer: yes — on your data, not ours.

ABOUT SOLVXAI

SolvxAI is the agentic operating system for upstream oil and gas. To see this working against your own asset, or to request the technical papers in this series, contact sd@nexaconsultinc.com.

References

Sources are listed with publication year; preliminary, self-reported, or vendor-authored work is flagged as such.

[1] **Smith, B. (IHS Inc.) and Fisher, D. (PPDM Association).** *The PPDM Well Identification Project: A Global Framework for Well Identification and its Application to Well Numbering in the United States*, preliminary report ahead of PNEC, May 2012. Industry-association working report; stated methodology of structured interviews with 13 regulators, 15 operators, 3 service companies and 3 vendors. **One co-author is affiliated with a commercial well-data vendor.** https://dl.ppdm.org/dl/513

[2] **Huber, A. and Fisher, D. (PPDM Association).** “The Canadian Well Identification System.” GeoConvention 2015. Conference paper. https://geoconvention.com/wp-content/uploads/abstracts/2015/040_GC2015_The_Canadian_Well_Identification_System.pdf

[3] **Binette, O. and Steorts, R. C.** “(Almost) All of Entity Resolution.” arXiv:2008.04443v3, 17 January 2022. Authoritative review; second author is also a principal mathematical statistician at the US Census Bureau. A journal version is commonly cited; we verified the arXiv version. https://arxiv.org/pdf/2008.04443

[4] **Peeters, R., Steiner, A., Bizer, C.** “Entity Matching using Large Language Models.” EDBT 2025. **Peer-reviewed conference.** https://openproceedings.org/2025/conf/edbt/paper-81.pdf

[5] **Zhu, H., et al.** “Aggregation Queries over Unstructured Text: Benchmark and Agentic Method.” arXiv:2602.01355v2, 4 February 2026. **Preprint**, not peer-reviewed. https://arxiv.org/pdf/2602.01355

[6] **Tang, Y. and Yang, Y.** “MultiHop-RAG: Benchmarking Retrieval-Augmented Generation for Multi-Hop Queries.” arXiv:2401.15391, 27 January 2024. **Preprint**, widely cited. https://arxiv.org/abs/2401.15391

[7] **Han, H., et al.** “RAG vs. GraphRAG: A Systematic Evaluation and Key Insights.” arXiv:2502.11371v3, 4 March 2026. **Preprint.** Cited here for evidence that cuts against structured retrieval. https://arxiv.org/html/2502.11371v3

[8] **da Cruz, T., Tavares, B., Belo, F.** “Ontology Learning and Knowledge Graph Construction: A Comparison of Approaches and Their Impact on RAG Performance.” arXiv:2511.05991, 8 November 2025. **Preprint**; **n = 20 questions** — illustrative only, not a rate. https://arxiv.org/html/2511.05991v1

[9] **Chen, E., et al. (Carnegie Mellon University).** “Comparing RAG and GraphRAG for Page-Level Retrieval Question Answering on a Math Textbook.” arXiv:2509.16780v2, 26 September 2025. **Preprint.** Cited for a result where the graph approach lost outright. https://arxiv.org/pdf/2509.16780

[10] **Wu, D., et al.** “LongMemEval: Benchmarking Chat Assistants on Long-Term Interactive Memory.” ICLR 2025. **Peer-reviewed conference.** The 30% figure is from the authors’ own abstract. https://arxiv.org/abs/2410.10813

[11] **Maharana, A., Lee, D.-H., Tulyakov, S., Bansal, M., Barbieri, F., Fang, Y.** “Evaluating Very Long-Term Conversational Memory of LLM Agents.” ACL 2024. **Peer-reviewed.** We verified the abstract; figures circulating from this paper's body were not independently confirmed and are not quoted here. https://aclanthology.org/2024.acl-long.747/

[12] **Hogan, A. (IMFD and University of Chile).** “The Semantic Web: Two Decades On.” *Semantic Web Journal*, IOS Press. **Peer-reviewed.** Argues both for and against each critique; the author is a Semantic Web researcher, and we represent the paper as the balanced assessment it is. https://www.semantic-web-journal.net/system/files/swj2229.pdf