A REFERENCE FRAMEWORK FOR AI-NATIVE EDISCOVERY

The Living Discovery Model

The EDRM* has served the industry since 2006. This is an alternative framework for discovery inside a corporation where a governed LLM can query the company's own systems in place, conversationally, without anything leaving where it lives.

CHARISMA STARR

livingdiscoverymodel.org

The EDRM* was created to represent solutions that required electronic documents to move before anyone could look at them. Identify them, preserve them, collect them, process them, review them, then finally produce them. Nine stages, largely sequential, because the underlying technology forced sequence on you. Data had to be copied out of a live system and into a repository before a human, or a keyword search, could touch it.

That constraint is gone. A properly governed LLM can be given permission to a corporation's live systems and asked, in plain language, to find, evaluate, and explain relevant information without anything ever leaving where it lives.

When the constraint that built the old model disappears, the model built on top of it should not survive unchanged.

EDRM 2.0 opened for public comment this summer, and it extends analysis across the workflow and grounds discovery in a wider governance lifecycle. It is a real improvement. But it keeps the phase structure. It treats AI as a capability you add inside each stage rather than a reason some of those stages no longer need to exist as separate events.

Here is the alternative: six shifts instead of nine phases, one of which is not a phase at all.

I call it "The Living Discovery Model" because it allows you and your data to live and breathe.

01. Connected Corpus

REPLACES IDENTIFICATION AND COLLECTION

Instead of collecting data into a repository, the system is given permissioned, federated access to where it already lives: email, chat, file shares, and systems. Nothing is copied out. Nothing sits somewhere new waiting to be searched. Identification and collection used to be two separate events precisely because you had to find data before you could move it. When nothing moves, both collapse into one act of connecting.

02. Continuous Indexing

REPLACES PROCESSING

Chunking, embedding, entity resolution, and deduplication run continuously in the background rather than as a single batch job that has to finish before review can start. Processing used to be a gate. Now it is a standing condition of the corpus, always current.

03. Retrieval-Augmented Discovery

REPLACES IDENTIFICATION AND ANALYSIS

Search-term negotiation gives way to conversational, iterative querying. Finding a document and evaluating whether it matters happen in the same turn, not as sequential gates separated by weeks of linear review. This is where the old model's biggest assumption breaks down: that you must finish finding before you start understanding.

Law firms and eDiscovery providers maintain their role as experts, guiding and consulting, without their clients exporting the data and handing it over. Company data remains safe, inside company-governed systems, thus reducing additional third-party architecture analysis, security review, and risk.

04. AI-First Review with Validation

REPLACES REVIEW

The system makes the first pass on responsiveness and privilege. Humans validate through statistical sampling, the same discipline that already underlies continuous active learning and TAR 2.0, just chatbot-mediated instead of relevance-model-mediated. The skill shifts from reading every document to auditing a process.

05. Explainable Production

REPLACES PRODUCTION AND EMPOWERS PRESENTATION

Every document produced carries a reasoning trail: the query or inference that surfaced it, and why it was deemed responsive. Production and presentation used to be separate because you prepared a static export first and explained it to the courtroom later. When eDiscovery lives inside the client's own connected environment, those steps align instead: exhibits, Bates-stamped versions, and presentations are produced and revealed as needed, online or through export, rather than assembled after the fact.

Case strategy and privileged work product can live in that same corporate environment, properly labeled as privileged and walled off from indexing entirely. Storage and retrieval are separate questions: sitting in company infrastructure doesn't mean sitting in the corpus that the AI queries.

The result is a cleaner chain of custody than the old export-and-explain model allowed. Client data and attorney-client work product never leave the corporate environment.

Preservation

THE SPINE, NOT A PHASE

This is the piece that has to be rebuilt from scratch, not relabeled. In the old model, preservation was accomplished by the act of collection itself: once you copied the data out and locked it down, it was safe from deletion or alteration because it no longer lived inside the systems that could change it.

In this model, the data never leaves. It stays in place, inside the corporation's own governed systems, the same systems where it originated and where people still use it every day. That is a real advantage: no separate copy to maintain, no question about whether an export still matches its source, no third-party repository that becomes its own point of failure.

But it means preservation has to do different work, in two parts. First, the live system itself needs enforceable hold controls: deletion paused, versions locked, the same protection collection used to provide, now applied directly in the system of origination. Second, because a chatbot queries that live data conversationally rather than through a one-time export, there has to be an immutable, auditable log of exactly what the system accessed, when, and what it did. That log is the new defensibility anchor. It is not something that happens once. It is what makes everything else defensible.

The Open Question

The hard, unsolved problem is defensibility. Courts will ask how you know you found everything, and what you didn't find and why. Today's recall and precision statistics were built for a search that runs once against a static index. They do not map cleanly onto a conversational, iterative, non-deterministic retrieval process. Those who elegantly solve validation for AI-native discovery will set the standard the rest of the industry gets measured against. That is the piece, and presenting it well is the road less traveled. I wish you luck in your journey and hope that this article may be of some benefit to the years ahead.

* EDRM.net

Charisma Starr has more than 25 years of experience in legal technology and eDiscovery, including serving as Head of Legal Technology and eDiscovery at Exelon, a founding member of Reveal, and head of Kirkland & Ellis's Litigation Support department in Chicago. She co-led the development of EDRM 2.0 before stepping away to pursue a more fundamental reimagining of the model. Drawing on feedback from her EDRM 2.0 discussions, she hopes this article supports software companies and thought leaders shaping the future of eDiscovery. As her views continue to evolve, she wholeheartedly welcomes feedback and ideas. You can reach her on LinkedIn.