

Софт не умеет смотреть ваши видео — он находит, рекомендует и ищет тайтлы через метаданные. Discovery не лучше, чем метаданные под ним. Соберите четыре вида, примите стандарты идентичности и обмена, смешайте источники разметки и постройте пайплайн до запуска. Версии стандартов в 2026 меняются — сверяйте вживую.

1 · ЧЕТЫРЕ ВИДА, КОТОРЫЕ НУЖНО СОБРАТЬ

- ☐ **Descriptive** — название, синопсис, жанр, актёры, настроение, тема, теги. Топливо для рекомендаций и поиска; дальше «жанра».
- ☐ **Structural** — сериал → сезон → эпизод, главы, трейлер ↔ фильм. Даёт группировку, resume, «следующий эпизод».
- ☐ **Technical** — разрешение, кодек, битрейт, дорожки аудио/субтитров, DRM. В основном фильтры; пишет энкодер.
- ☐ **Administrative** — владелец, территория, окна, модель, рейтинг. Решает, может ли тайтл вообще появиться.

2 · ИДЕНТИЧНОСТЬ И ОБМЕН (стандарты, не свой код)

- ☐ **EIDR** — примите универсальный content ID (на базе DOI; сериал→сезон→эпизод); требуйте его у лицензируемого контента.
- ☐ **Канонический ID** — присвойте внутренние ID всему без EIDR, чтобы «это тот же тайтл?» никогда не было догадкой.
- ☐ **MovieLabs MEC** — принимайте каталоги в схеме Common Metadata / MEC, чтобы поля попадали в поля, а не в свой парсер.
- ☐ **Schema.org** — публикуйте VideoObject / Movie / TVSeries в JSON-LD, чтобы Google и AI-ассистенты читали ваши тайтлы.

3 · РАЗМЕТКА И ОБОГАЩЕНИЕ (смешайте три, всегда QC)

- ☐ **Ручная разметка** — тонкие теги (настроение, тема); богатейшее топливо, но медленно, дорого и непоследовательно без словаря.
- ☐ **Провайдеры** — купите готовые метаданные для массового каталога; слабы на нишевом/оригинальном контенте, на котором вы выделяетесь.
- ☐ **AI-обогащение** — авто-теги сцен, объектов, речи, настроения в масштабе каталога; нужен QC (одна ошибка портит рекомендации).
- ☐ **Фингерпринтинг (ACR)** — опознать контент без меток и поймать дубли; связать каждый ассет с канонической записью.

4 · ПАЙПЛАЙН (постройте до запуска)

- ☐ **Ingest** — принимайте множество диалектов и форматов поставщиков без своего импортёра на каждую сделку.
- ☐ **Нормализация** — сведите каждое поле к одному словарю и одному ID; уберите дубли.
- ☐ **Обогащение** — закройте пробелы данными человек + AI + провайдеры; добавьте глубину descriptive, нужную discovery.
- ☐ **Governance** — владейте истиной: точность, свежесть и кто чинит неверный тег.

ПОРЯДОК ДЕЙСТВИЙ — ИДЕНТИЧНОСТЬ, ОБМЕН, ОБОГАЩЕНИЕ, GOVERNANCE; ГЛУБИНА DESCRIPTIVE ДО ЗАПУСКА

Метаданные — это топливо, которое сжигает discovery, и слой discovery никогда не будет лучше, чем метаданные под ним: мусор на входе — мусор на выходе. По порядку. Первое: соберите все четыре вида и не принимайте каталог только с названием и жанром, потому что descriptive и structural метаданные — это то, что читают рекомендации, поиск, ряды главного экрана и путь cold start для новых тайтлов. Второе: решайте идентичность и обмен стандартами, а не своим кодом — дайте каждому тайтлу один универсальный ID (EIDR), чтобы все системы соглашались, какой это тайтл, и принимайте каталоги в общей схеме (MovieLabs Common Metadata / MEC), чтобы поля поставщика попадали в ваши поля без парсера на каждую сделку; публикуйте структурированные данные Schema.org, чтобы Google и AI-ассистенты тоже читали ваш каталог. Третье: обогащайте, смешивая ручную разметку, сторонних провайдеров и AI плюс фингерпринтинг — и контролируйте качество машинного вывода, потому что одна ошибка метки портит все рекомендации ниже по течению. Четвёртое: постройте пайплайн ingest-нормализация-обогащение-governance до запуска, а не как пункт бэклога — бедные метаданные делают хороший каталог невидимым и превращают лицензионные платежи в списания. Версии стандартов (MEC, обязательные поля структурированных данных Google) меняются — сверяйте перед планированием.