EIDR: INITIAL PROOFS OF CONCEPT-MATCHING AND API INTEGRATION

(REV 1.1)

Getting started with an EIDR POC: Initial steps and high-level overview of Matching and API Integration.

1 Introduction

Now that you have become an EIDR Member, we suggest you begin with two, possibly parallel, Proofs of Concept (POCs): a Matching POC and an API (Application Programming Interface) Integration POC, which will help you become acclimated to the EIDR system. This document describes the framework for you to get started on these EIDR POCs. The detail of this framework can be used to create a project schedule for the effort.

2 Initial Proofs of Concept

The intention of this section is to provide a high-level, step-by-step overview of the general tasks and high-level boilerplate durations for the standard initial EIDR POC.

All POCs start with data review and preparation:

- Data Review and Preparation
  - Record and Workflow Definition
  - Data Model Mapping

Then, depending on the workflow and other factors, such as the type and quantity of records, and the likelihood of overlap with existing records, one or both of the following paths may be pursued:

- Matching
  - Record Matching Process
  - Party (Principal Agent) Matching
  - Manual or Bulk Title Registration
  - Modification of Records
• Utilizing the API
  o API or SDK (Software Development Kit) Integration

The last stage in a POC is to document the results and lessons learned so that the EIDR community can benefit from the experience gained.

• Documenting and Sharing

2.1 Data Review and Preparation

2.1.1 Record and Workflow Definition

The first task in any POC is to define the types of records to be registered or matched and the workflow.

EIDR has a rich data model supporting movies and episodic content as well as edits, products, and digital assets. Records can be registered in a hierarchy that also includes other relationships, such as elements in a composite or a digital asset as a packaging of a product. This document is primarily focused on POCs involving title-level movie records. Workflows for matching and registering episodic content are still under development.

Once the types of records have been determined, the intended workflow needs to be defined. Matching of a large catalog requires a semi-manual workflow to ensure quality results. Registration of green field titles, edits, and assets can proceed along a more automated path, using direct API integration. If API integration is used, then the workflow also needs to cover handling of asynchronous API responses for registrations that end up in manual review within EIDR.

2.1.2 Data Model Mapping

This section describes the process by which the Member’s data model is mapped to the EIDR Data Model to (1) define and map all correlative data fields, and (2) identify any gaps and determine if those gaps will cause any issues with the creation of records.

1. Member reviews their data model and maps it to the EIDR data model, paying particular attention to the EIDR data fields needed to register a record.
2. If needed, the Member can request a review of the data model mapping of EIDR, to help validate that the Member’s data model is aligned with EIDR.
3. Member supplies sample set (for example, 10 titles) of data for the fields identified to EIDR in a pre-defined Excel template.
4. EIDR reviews this initial sample to confirm mapping and identify any gaps.
5. If there are missing data fields or other anomalous issues, the Member iterates with the EIDR team to close these gaps.

Output: Mapping of Member’s data model and practices to EIDR's.

An estimated duration of **5 to 7 business days** should be allocated for this task.
2.2 Matching

2.2.1 Record Matching Process

The next step in the process is for the Member to match its data against the EIDR data. We've defined five mechanisms for matching records.

1. Use of Member’s internal matching tools or a third-party service to match against an EIDR-provided dump of the EIDR database.
2. Use of ID mappings known by third parties, e.g. AMG IDs, which Rovi can map to EIDRs.
3. Use of ID mapping with IDs that are registered as Alternate IDs in EIDR, such as an ISAN or IMDB ID.
4. Use of the EIDR matching tools, which are currently under development.

Generally, none of the processes can be fully automated. Even the results of using ID mappings should receive some manual review. All of the processes benefit from the use of spreadsheets to evaluate the quality of the matches and review questionable records. A typical matching process may progress as follows:

1. EIDR can provide a data file (flat file) of the EIDR registry to the Member for matching.
2. Member reviews and matches titles and develops a list of titles that are not matched on the EIDR database file (gap records).
3. At this pint the member should add the matched IEDR’s into their internal systems
4. Member inputs the gap records it wishes to register on an EIDR provided template.
5. Using the provided data, the EIDR team performs a spot check of the gap records to confirm there are no matches in EIDR.
6. Member then follows the Party Matching steps, see below.

Output: Gap titles identified and verified for Bulk Ingest process.

The duration of this group of tasks will vary based on the number of records, but for approximately 200 title level records, an estimated duration of **10 to 15 business days** should be scheduled for automated and manual title matching and verification. The time needed for larger title sets will depend on the available tools and data quality.

2.2.2 Party (Principal Agent) Matching

Once a member has aligned its data model with EIDR, and all needed data fields for a registration are available, the next step is to confirm all Production and Distribution parties are present in EIDR.

The goal in Party Matching is to ensure that all the parties that a Member may want when registering titles are present in the EIDR system before registration begins.
1. Member builds a list of Parties (Production and Distribution companies) they will need as part of the registration process.

2. EIDR provides a list of the current EIDR Parties in an Excel template for the Member to match against.

3. Member compares EIDR Party IDs with their own generated list.

4. Member fills in their names for the matched records and lists any unmatched parties from their list.

5. If there are missing parties, Member submits a help desk ticket to add the missing parties to the EIDR system.

6. EIDR Support Team adds parties as appropriate, based on help desk request, and returns a list of EIDR Party IDs.

7. Once all parties are added to the EIDR system and the Member has recorded all the associated EIDR Party IDs, registration can start.

Output: Mapping of Member’s producer/distributor database to EIDR’s.

The duration of this group of tasks will vary based on the number of production/distribution companies, the quality of the data, and the number of missing parties, but an estimated duration of **5 to 10 business days** should be scheduled for all party matching efforts.

### 2.2.3 Manual or Bulk Title Registration

After Data Model Alignment, Record Matching, and Principal Agent Matching have been completed, the registration process can start.

1. The Member provides a full set of data fields required for the gap titles identified in the output of the Matching Process in the template provided.

2. EIDR spot checks records:
   a. Identifies records with missing data fields:
      i. For records with missing data, those records are reviewed and missing data manually added.
      ii. In some cases, classes of records may be deferred for later registration.
   b. Spot check data quality for those records with full data fields:
      i. Of the records spot checked, errors are manually addressed.
      ii. If there is a high percentage of records in error during the spot check, a more thorough review will be needed, and a manual addition of missing data will need to be performed by the Member.

3. Once a final Excel spreadsheet is provided to EIDR, it is converted utilizing an EIDR provided tool into XML.

4. The XML file is ingested into the EIDR sandbox.

5. After ingest, the Member and EIDR spot check the data and fields:
   a. Errors corrected in input files if possible; otherwise, record made for manual correction.

6. Once quality is confirmed, modified ingest files are sent to the Registry:
   a. EIDR performs any manual fixes noted in 5a.
   b. EIDR registers the records.
c. EIDR spot checks the Registry.
   d. EIDR provides the Member with a list of the newly assigned IDs.

**Output:** Gap data added to EIDR database.

*NOTE:* The member is responsible for adding the newly create EIDR IDs to its own data systems.

The duration of this group of tasks will vary based on the number of records, but for approximately 200 title level records, an estimated duration of **5 to 10 business days** should be scheduled for registration into the EIDR system. If the number of gap titles is very small, the gap registration could alternatively be done through the Web UI.

### 2.2.4 Modifications and Record Update

If the POC involves adding or correcting metadata in the matched EIDR records, e.g. adding Alternate IDs or replacing the release year with a full date, this can be handled in one of three ways:

- Either EIDR or the Member can extract the alternate IDs and metadata corrections from the spreadsheets and update the records using existing tools, or
- The Member can do so by integrating these updates into an API-based workflow, or
- EIDR Support can provide assistance for smaller numbers of modifications.

1. After Record Matching, if the Member has additional fields such as alternate IDs or more accurate metadata, the Member supplies a CSV file with the EIDR IDs and additional metadata fields via a Help Desk Ticket.
2. The EIDR Support Team spot checks metadata for accuracy and alternate IDs for proper granularity.
3. The EIDR Support team updates the records.

### 2.3 Utilizing the API

The API path is one that can be done in parallel with the steps defined above in the Matching Path. The API integration is the process by which the Member uses the EIDR API or SDKs to integrate EIDR Record Creation and EIDR Search into the Member's relevant systems.

#### 2.3.1 API Integration (High-Level Overview)

The duration of this group of tasks will vary widely based on based on the team assigned, the number of integration points (search, register, modify, etc.), and the current state of existing systems within the Member. An estimated duration of between **60 to 90 business days** should be scheduled for API integration and validation.

1. Establish Member development team for integration
2. Review all EIDR API and SDK documentation
3. Decide whether to use Java SDK, .NET SDK, or raw HTTP (REST API).
a. We strongly recommend using Java or .NET, with use of the REST API reserved for simple applications such as a Web UI.

4. Set up training session with EIDR technical team

5. Define and document (or verify existing) data field mappings and hierarchies

6. Define and document workflows, handling of existing matches, asynchronous responses, ACL additions, initial and ongoing record updates

7. Register Member and associated users with EIDR registry accounts (Sandbox)

8. API/SDK integration
   a. Perform integration for one or more data types (e.g. Movie title, Episode version, etc.) and workflows (register, update, etc.) using the Sandbox
   b. Validate the data results in the Sandbox, possibly starting from a clean Sandbox refreshed from production. This includes validating the data quality and any hierarchies.
   c. Validate that the workflows are performing adequately for collisions and asynchronous responses. This may also include EIDR administrative functions, such as granting ACL permissions on existing records.

9. Iterate #8 for other data types and workflows required for the POC

10. Member and EIDR agree when POC is ready to move to production

11. Register Member and associated users with EIDR registry accounts (Production)

12. Move to Production with a limited set of records

13. Validate the data results in Production

14. Place into normal operation

2.4 Documenting and Sharing

• Member documents practices and lessons learned to share with other members
• Present material at a POC TWG (Technical Working Group) or other POC forum