EIDR
Entertainment Identifier Registry

EIDR: INTERIM BEST PRACTICE – REQUIRED BASE OBJECT DATA

This document details the interim best practice for indicating that certain required Base Object Data cannot be determined.

Contents

1 Introduction 2
2 Substituting Required Data 2
2.1 Country of Origin 2
2.2 Original Language 3
2.3 Release Date 4
2.4 Approximate Length 5
3 Compensating Data 6
3.1 Minimum Root Record Data Requirements 7
3.2 Episode Deduplication 7
3.3 Identifying Foreign Territory Releases 8
3.4 Complete Credits Flag 8
3.5 Metadata Authority 9
4 Substituted vs. Provisional Data 9
4.1 Title 10
4.2 Country of Origin 10
4.3 Original Language 10
4.4 Release Date 10
4.5 Event Date 11
4.6 Approximate Length 11
4.7 Credits 11
5 Impact on Deduplication 11

EIDR Entertainment Identifier Registry

1 Introduction

EIDR Content records must contain sufficient descriptive information to uniquely identify the records and distinguish them from other, confusingly similar records so that the accuracy of the EIDR registry is not compromised.

All registrants should always strive to provide complete and accurate metadata for every EIDR registry record, expending reasonable efforts to obtain any missing data and to verify all data provided to EIDR. This not only helps ensure that their own registrations will be as accurate as possible, but that subsequent registrations will correctly deduplicate against prior registrations of the same work.

However, practical considerations often make it prohibitively expensive to research and verify certain pieces of information for certain Content records. In these situations, accommodations must be made, so long as the integrity of the registry is not placed in jeopardy. In no circumstances will "spoofed," or knowingly fabricated, data be allowed in the EIDR registry.

To this end, mechanisms have been developed to allow records to be recorded that do not have known values for the following Base Object Data elements:

- Country of Origin
- Original Language
- Release Date (for Edit records only)
- Approximate Length

The Base Object Data changes will need to be incorporated into the EIDR development schedule, so in the meantime, temporary practices have been identified that will allow the immediate registration of records where one or more of these fields does not contain valid information.

2 Substituting Required Data

2.1 Country of Origin

The Country of Origin field contains one or more ISO 3166-1 2-character Country Codes that identify the home country(ies) for the company(ies) exerting primary creative control over the creation of the work. (Generally, the home country(ies) of the Associated Org(s) with a producer role.) This field is used to help:

- Identify when a title record has been registered with a foreign territory release title in the Resource Name instead of the original release title. (Potentially an Edit registered as a title record.)
- Identify duplicate works submitted with non-aligned metadata. (Particularly when the titles are in different languages.)

- Distinguish between remakes. (A new version of a foreign film produced for the local market.)

NOTE: The order of the Country of Origin entries does not affect discovery or deduplication, but is assumed to reflect the decreasing order of significance.

Country of Origin is of relatively little value in automated deduplication, but is quite important in manual reviews of match candidates and in automated bulk data scans, such as registry data quality reviews and catalog matching exercises.

Because there are many different definitions of country of origin used throughout the ecosystem, and different definitions may be applied to the same work for different purposes at different points in time, many distributors, broadcasters, and data aggregators do not track this information in their master title systems. (This is not an issue for producers, since they are themselves the primary creative controllers identified by the Country of Origin field.)

2.1.1 Interim Practice

Provide a single Country of Origin set to "AQ" for Antarctica.

This is a valid ISO 3166-1 country code and will be accepted by the EIDR registry's validation rules. However, since Antarctica has no local government (it is an internationally neutral continent), it cannot have locally incorporated businesses. This means it can be a filming location, but it can never be a valid Country of Origin by EIDR's definition. So, any time it is encountered, it will be interpreted to mean that the Country of Origin is "undetermined."

2.1.2 Long-Term Solution

- Add the code "XX" for "Undetermined" (an ISO 3166-1 user-assigned value) to the EIDR validation rules.
- Convert all existing "AQ" codes to "XX".
- Add a registry validation rule that rejects any records that contains more than one Country of Origin when at least one of them is "XX". (XX can only appear by itself.)
- Set the deduplication rules so that a mismatch involving an "XX" Country of Origin does not lower the overall match score.

2.2 Original Language

The Original Language field contains one or more RFC 5656 Language Codes that identify the significant language(s) present in the original version of the work. This field is used to help:

- Identify when a title record has been registered with a foreign territory release title in the Resource Name instead of the original release title. (Potentially an Edit or Manifestation registered as a title record.)

- Identify duplicate works submitted with non-aligned metadata.
- Distinguish between remakes.

NOTE: The order of the Original Language entries does not affect discovery or deduplication, but is assumed to reflect the decreasing order of significance.

Original Language is of relatively little value in automated deduplication, but is quite important in manual reviews of match candidates and in automated bulk data scans, such as registry data quality reviews and catalog matching exercises.

Because an acquired work may not retain its original language at the point when it is acquired, a down-stream distributor, broadcaster, or data aggregator may not know what languages were present to a significant degree in the original. (This is not an issue for producers, since they created the original work.)

NOTE: Original and Version Language replaced Primary and Secondary Language in the EIRD 2.0 release.

2.2.1 Interim Practice

Provide a single Original Language set to "und" for "Undetermined" with a Language Mode of "audio".

This is a valid ISO 639-2 (RFC 5656 compliant) language code and will be accepted by the EIDR registry's validation rules.

2.2.2 Long-Term Solution

- Add a registry validation rule that rejects any records that contains more than one Original Language when at least one of them is "und". (Undetermined can only appear by itself.)
- Set the deduplication rules so that a mismatch involving an "und" Original Language does not lower the overall match score.

2.3 Release Date

The Release Date is a year or full date (with a strong preference for full date, particularly for Episodes) that records the original public release of each work, derived version, encoding, etc.[1] EIDR does not record subsequent release dates in new avail windows, distribution channels, markets, etc. This field is used to help:

- Identify duplicate works submitted with non-aligned metadata.
- Distinguish between remakes, sequels, etc.

[1] Excluding festivals, preview or press screenings, gala premiers, etc., but including awards qualification runs, limited releases, etc.

- Correctly sequence episodic works.

The original Release Date is readily available, and shall be provided, for all title records. However, it is quite often not available for derived Edits. The common practice is to simply copy the Release Date from the parent record. However, this does not draw attention to the fact that the Release Date is, in fact, unknown and gives a false sense of confidence in the data.

2.3.1 Interim Practice

In Edit records only, provide a Release Date of "3210-01-23" and set the Provisional data flag to true. (See *EIDR: Interim Best Practice – Provisional Data*.)

This is a valid calendar date, so it will be accepted by the EIDR registry's validation rules, but it is not a date that is likely to be entered by accident – particularly when it is accompanied by a Provisional data flag.

2.3.2 Long-Term Solution

- Change the registry validation rules so that an Edit record can be submitted without a Release Date. Instead, inherit the Release Date from the parent record as needed.[2]
- Remove the Release Date and Provisional flag from all Edit records with a "3210-01-23" Release Date.
- Set the deduplication rules so that a missing Edit Release Date is treated as an inherited value.

2.4 Approximate Length

The Approximate Length is an XML standard duration (xs:duration) that records the elapsed time from first to last frame of picture. It is assumed to be approximate at the title level and gain increasing specificity moving down the hierarchical registration tree through Edits to Clips and Manifestations. Made for TV works often use scheduled timeslot in place of length of picture. This field is used to help:

- Distinguish between work types (shorts from features, sitcoms from dramas, etc.)
- Distinguish between different versions & manifestations.

The duration of a significant percentage of catalog works is not known. For example, many works may simply be designated "short" or "feature" with no indication of their actual length. This affects title records and extends down to Edits, where the differences in duration for different edits is not readily available.

[2] If a content record does not define a value in a field that is eligible for inheritance, then the parent's value will be substituted in a full metadata query.

2.4.1 Interim Practice

Provide an Approximate Length of "PT0.001S" (one millisecond) and set the Provisional data flag to true. (See *EIDR: Interim Best Practice – Provisional Data*.)

This is a valid duration, so it will be accepted by the EIDR registry's validation rules, but it is not a duration that is likely to be entered by accident – particularly when it is accompanied by a Provisional data flag.

NOTE: Do not use this for Series or Season records where there is no standard duration for the child Episodes. Instead, continue to provide "PT0H" for the Approximate Length.

2.4.2 Long-Term Solution

- Change the registry validation rules so that Edit records can be submitted without an Approximate Length. Instead, inherit the Approximate Length from the parent record, as needed.
- Remove the Approximate Length and Provisional flag from all Edit records with a "PT0.001S" Approximate Length.
- Set the deduplication rules so that a mismatch involving a "PT0.001S" Approximate Length does not lower the overall match score and so that a missing Edit Approximate Length is treated as an inherited value.

  NOTE: It could be possible for an Edit to inherit an Approximate Length of "PT0.001S" from its parent.

3 Compensating Data

To help ensure that relaxing the required data rules for Base Object Data does not have a significant negative impact on the accuracy of the EIDR registry' deduplication process, when one or more of the previously required fields is substituted with the appropriate "undetermined" value in a root record, additional participant metadata must be provided to compensate. In certain cases, additional participants may not exist, so a special provision will be needed to identify these works and exempt them from the additional compensating data requirements.[3] For Episodes, additional participant information would likely not be useful, so special deduplication rules will be applied. Finally, to ensure that the "undetermined" values are not abused, only Metadata Authorities will be allowed to use them.

[3] The interim solution proposed will not actually relax the requirements, just flag records with complete credits blocks that contain less than the maximum possible values. Only the long-term solution will be able to override the data validation rules.

3.1 Minimum Root Record Data Requirements

- If all required fields are provided (none are substituted with an "undetermined" value), then the current minimum participant data requirements apply:
  - At least one Associated Org (with any role)

    OR
  - At least one Director

    OR
  - Four Actors
- If at least one of the Base Object Data fields in a root record[4] contains an "undetermined" value, then the minimum participant requirements are increased to:
  - At least one Associated Org identified as Producer OR one Director

    AND
  - At least two Actors.

Registrants are always encouraged to provide more than the minimum number of participants, preferably at least one production company, one director, and four actors (in first billed order).

NOTE: If it is not possible to meet the enhanced compensating data minimums because a work simply does not have that many participants, then the compensating data minimums will be waived if the registrant asserts that all participants that exist have been provided by setting the Complete Credits flag to "true".

3.2 Episode Deduplication

When considering a pair of Episode records for deduplication, the following conditions will force automatic manual review:

- One or both of the records has a system-generated title.

  AND
- At least one of the following applies:
  - One or both of the records has a release year instead of a full release date.

    OR
  - The two records do not have at least one identifying number from the same domain.

[4] This would not include Seasons, Episodes, or Edits.

3.3 Identifying Foreign Territory Releases

To retain the ability to identify foreign territory releases that are more properly registered as Edits or Manifestations, rather than title records, every record must contain at least one valid entry for Country of Origin or Original Language. That is, both of these fields cannot contain "undetermined" values at the same time.

3.3.1 Interim Practice

In the near-term, this will be enforced by policy and periodic registry data reviews.

3.3.2 Long-Term Solution

Reject any record where both Country of Origin and Original Language have been set to "undetermined" values ("XX" and "und", respectively).

3.4 Complete Credits Flag

There are works that have fewer participants than may be required for registration under the compensating data rules. For example, an animated work with no dialogue or a nature film with no narration will have no actors; many works are produced by a single production company, so more than one cannot be provided.

3.4.1 Interim Practice

If all of the participants for a work (production companies, directors, and cast members) have been identified in a record, but the minimum compensating data requirements have not been met, add "CR:Complete;" to the Registrant Extra field.

3.4.2 Long-Term Solution

- Add an optional "Complete" flag to the Credits block that can be set to "true" when the registrant has provided all the participant information that applies to the identified work.[5] (Missing or "false" Complete flags are ignored.)

  NOTE: This may result in a Credits block that has a Complete = true attribute, but no sub-tags or other data payload.

- Reject any record with an empty Credits block unless Complete = true.
- Ignore the compensating data participant minimums when Complete = true and accept however many participants have been identified. (The standard minimum metadata rules will still apply.)[6]

[5] Works from the early Silent Era may not have identified directors. In these cases, the producer generally performed the functions now attributed to the director and can be identified as the director in the EIDR record. Actors can include anyone who appears in the visual record or who is heard in the soundtrack, including individuals appearing as themselves in interviews, press conferences, political debates, etc.

- Convert all existing "CR:Complete;" entries in the Registrant Extra field to a Complete = true attribute in the Credits block.
- Force a manual deduplication review of any record being submitted for registration that has the Complete flag set true.

3.5 Metadata Authority

Only a declared Metadata Authority shall be allowed to substitute one of the "undetermined" values for a required Base Object Data field. The Metadata Authority thereby asserts that they are providing the best and most complete metadata available and commits to updating the EIDR registry record should more or more accurate data become available in the future.

NOTE: Being a record's Metadata Authority does imply any ownership interest, intellectual property rights, or involvement in the creation or distribution of the work being described. If a work has more than one Metadata Authority, their order does not imply precedence – they are all treated equally.

3.5.1 Interim Practice

By policy, any registrant wishing to use one of the "undetermined" values in a record must list their Party as a Metadata Authority for the record

3.5.2 Long-Term Solution

Add a new registry validation rule that rejects any record registration that includes an "undertermined" value unless the record contains a Metadata Authority.

- If the registrant is the Superparty, the Metadata Authority can be any Party with the Metadata Authority role.
- Any other registrant must list itself as the Metadata Authority.

NOTE: So long as the record includes "undetermined" values, it must have at least one valid Metadata Authority.

4 Substituted vs. Provisional Data

Certain required values may be identified as being "undetermined" (they simply do not apply to the record in question or cannot be obtained through reasonable effort), while others values may be flagged as being "provisional" (subject to change in the near term, but eventually available in a reliable form). A select few support both options.

---

[6] This means that with the Complete flag set true, the content record will still require at least one Associated Org OR one director OR four actors. The Complete flag only relaxes the increased minimum data requirements applied under the compensating data rules.

Data Element	Undetermined	Provisional
Title	No	Yes
Country of Origin	Yes	No
Original Language	Yes	No
Release Date	Yes	Yes
Event Date	No	Yes
Approximate Length	Yes	Yes
Credits	Yes	No

4.1 Title

A work always has a name, even if it is an internal or working title, so there is no need to allow for an undetermined value.

4.2 Country of Origin

A work's country of origin does not normally change during development or production (short of exiting turn-around owned by a company from a different country, which is a rare enough occurrence that it can be handled as a manual exception), so there is no provisional option. However, it may be un-knowable at the time of initial registration, and so supports an undetermined value.

4.3 Original Language

A work's original language is not a matter of uncertainty for its producers (one does not start out to make a Spanish language telenovela and somehow end up with a German language program), so there is no provisional option. However, it may not be obvious to down-stream distributors who only possess a dubbed version, and so supports an undetermined value.

4.4 Release Date

Prior to its original release, a work's release could change due to any number of circumstances, so release date necessarily has a provisional option. At some fixed point in time, a work is either released or abandoned, and so for EIDR's purposes, the release date is eventually known for all works. As a result, undetermined release dates are not allowed on title records. However, the specific release date for a particular version of a work may not be knowable, so Edits may inherit their release date from their parent rather than directly asserting a date that is unverifiable. Since you can only apply a

provisional flag to a release date that exists, you cannot have a record with a release date that is both provisional and undetermined at the same time – it must be one or the other.

4.5 Event Date

Event Date is an optional field, valid only for Live Events. It support a provisional flag, since the planned dates for live events often change, but it does not have an undetermined option since it is simply not provided if it is not known.

4.6 Approximate Length

Approximate length is a special case. To begin, it is approximate.[7] During development, production, and a good portion of post-production, the target length could vary and so may need to be flagged as provisional, in addition to being approximate. Older catalog works and works that are both acquired and exploited in foreign territories are often recorded in centralized title systems without durations. Tracking down and measuring the duration of such works is often cost-prohibitive, if even possible. Therefore, approximate length also supports an undetermined option. However, since duration can be determined by inspection (unlike original language, which may not be apparent upon viewing the work), any work with an undetermined duration must also be flagged as being provisional – holding out hope against that day when a party with sufficient proprietary interest in the work will arise and provide an approximate length.

4.7 Credits

Credits, if not known, are not provided, and so do not have a provisional flag. (During development when a work's cast may change, the content record can simply be updated to match the current state of the attached talent. As the work moves through production, the final cast eventually becomes known.) By the same logic, they do not support an explicit undetermined option. They do, however, support an option to indicate that they are complete as provided. This avoids the need to provide "filler" values just to satisfy minimum data requirements when they simply do not apply to a particular work.

5 Impact on Deduplication

The interim practices recommended above will lower a work's deduplication score, resulting in fewer works that receive a high-confidence match (and therefore require manual review). There will also likely be works that fall out of manual review and are incorrectly identified as gap records when one party has provided the required values while another party has not. As a result, the interim practices should be used with

[7] The assumption being that a work in the abstract does not have a specific length, but rather one that approximates the modal length of the derived versions. Moving down the inheritance tree to Edits, Clips, and Manifestations, the length becomes more certain and more specific.

caution and only when no other reasonable alternative is at hand (including manual search prior to registration).

The long-term solutions include necessary adjustments to deduplication to account for the undetermined data. This will result in more works entering manual review than would otherwise, but will avoid false high-confidence or gap record scores.