

# **Abstract Metadata in Public Broadcasting**

## **Part 1: Introduction**

Version 2.1  
May 2017

# Table of Contents

- [1 Introduction to This Document](#)
  - [Document Organization](#)
  - [External Resources](#)
- [2 Broadcasting Content on the Interconnection System](#)
  - [The Broadcasting Lifecycle Phases](#)
    - [Figure 1: Content Lifecycle](#)
    - [Acquisition](#)
    - [Production](#)
    - [Distribution](#)
    - [Delivery](#)
    - [Archive](#)
    - [Transition](#)
  - [Figure 2: Content Moving Through Lifecycle Stages](#)
  - [The Interconnection System](#)
    - [v5 Interconnection System Overview](#)
    - [Enhancement Overview](#)
- [3 The Metadata System](#)
  - [A Non-Broadcasting Abstract Metadata System](#)
  - [Abstract Models in Broadcasting](#)
  - [Types of Broadcasting Content Metadata](#)
  - [Identifiers](#)
  - [The Impact of an Abstract Metadata Hierarchy in Broadcasting](#)
- [4 Metadata in the Broadcasting Content Lifecycle](#)

# 1 Introduction to This Document

Public Broadcasting Service (PBS) acts as the caretaker of the Interconnection System (IXS) and the main distributor of content. For next evolution of the IXS, PBS has implemented a hierarchical abstract metadata system using the Entertainment Identifier Registry (EIDR). This system will issue unique identifiers to newly registered content based on basic metadata, like titles, genre, etc., making new content more organized and easily discoverable by stations.

This is the first part of a series of documents concerning the metadata system. The other documents in this series are as follows:

- **Part 2: Core Models**  
This part describes the core abstract metadata models that compose the content library.
- **Part 3: Time-based Descriptive Metadata**  
This part explains the metadata that provides detailed information about media content.

This section introduces the next evolution of the IXS and the need for an hierarchical abstract metadata system.

## Document Organization

This document is organized as follows:

1. “Broadcasting Content on the Interconnection System” provides the historical context behind the abstract metadata system implemented for the Interconnection System.
2. “The Metadata System” explains abstract metadata hierarchies in general, applies the general idea to broadcasting, and lists some use cases where an abstract metadata hierarchy in broadcasting would benefit station workers.
3. “Metadata in the Broadcasting Content Lifecycle” unites the content lifecycle described in Section 1 with some of the metadata concepts described in Section 2.

## External Resources

The following resources provide extra information about the broadcasting lifecycle and IXS:

- [PBS-TOP1] PBS Technical Operating Procedures, Part 1, Program Submission, November 2016. <http://mypbs.org/workarea/DownloadAsset.aspx?id=60130968219>

- [PBS-TOP2] PBS Technical Operating Procedures, Part 2, Program Distribution from PBS, 2015. <http://mypbs.org/workarea/DownloadAsset.aspx?id=60130968217>

The following resources discuss metadata in the movie industry and EIDR standards:

- [CM] [TR-META-CM](http://www.movielabs.com/md/md) MovieLabs Common Metadata, v2.4, <http://www.movielabs.com/md/md>
- [EIDR-TO] *EIDR Technical Overview*, November 2010. <http://eidr.org/technology/#docs>
- [PBS-EIDR] *Recommendation for adoption of EIDR for v6 Interconnection System*, January 8. Internal Document.

For metadata in general, see the following:

- [NISO-UM] *Understanding Metadata*, 2004. <http://www.niso.org/publications/press/UnderstandingMetadata.pdf>
- [NISO-MD] *Metadata Demystified*, 2003. [http://www.niso.org/standards/resources/Metadata\\_Demystified.pdf](http://www.niso.org/standards/resources/Metadata_Demystified.pdf)

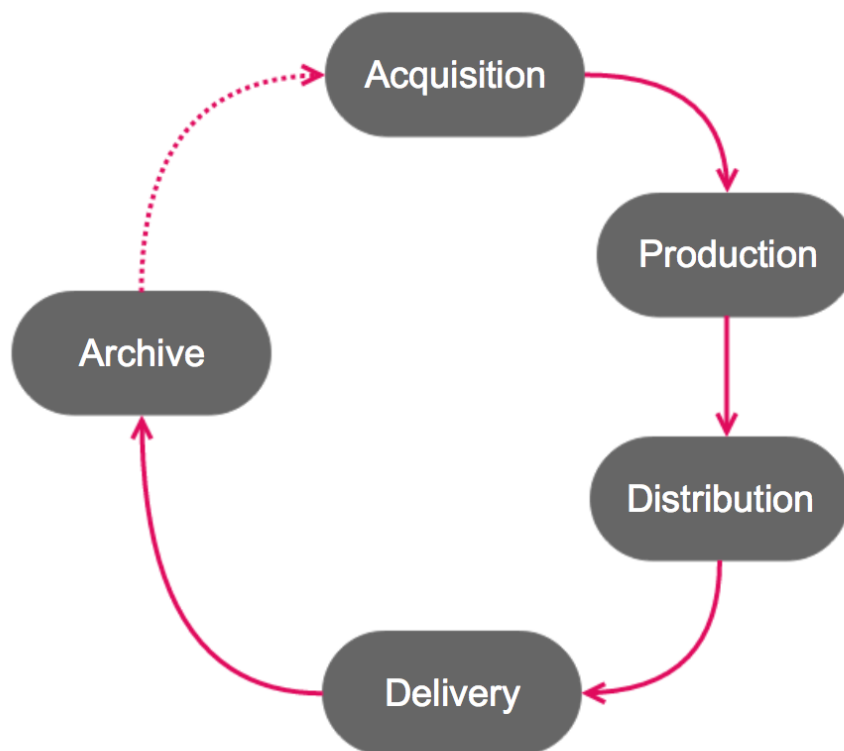
## 2 Broadcasting Content on the Interconnection System

The public broadcasting content undergoes a typical lifecycle, and that content is described by the metadata system explained in this document. This section explains the importance and reasoning behind the metadata system by providing contextual information about the lifecycle of broadcasting content and the IXS that delivers content to broadcasting stations.

### The Broadcasting Lifecycle Phases

While there are many exceptions that can happen for individual pieces of content, the standard content lifecycle can be represented in [Figure 1](#).

Figure 1: Content Lifecycle



The typical first-time flow for most content will follow this sequence:

1. [Acquisition](#)
2. [Production](#)
3. [Distribution](#)
4. [Delivery](#)
5. [Transition](#)
6. [Archive](#)

## Acquisition

This stage represents the business and legal tasks that are required to establish a relationship or environment that facilitates the content creation.

## Production

This stage represents the creation of content essence, including footage creation (Source Media) and editing (Stories, Releases).

## Distribution

This stage represents the preparation required to deliver content to a wider audience, such as when PBS sends information over the [Interconnection System to PBS stations](#).

## Delivery

This stage represents content delivery from stations to viewers, such as television broadcast or digital or online availability.

## Archive

If Archived, content is no longer actively produced, distributed, or broadcast. Archiving is temporary storage for content until additional lifecycles -- such as re-runs or syndication -- are triggered.

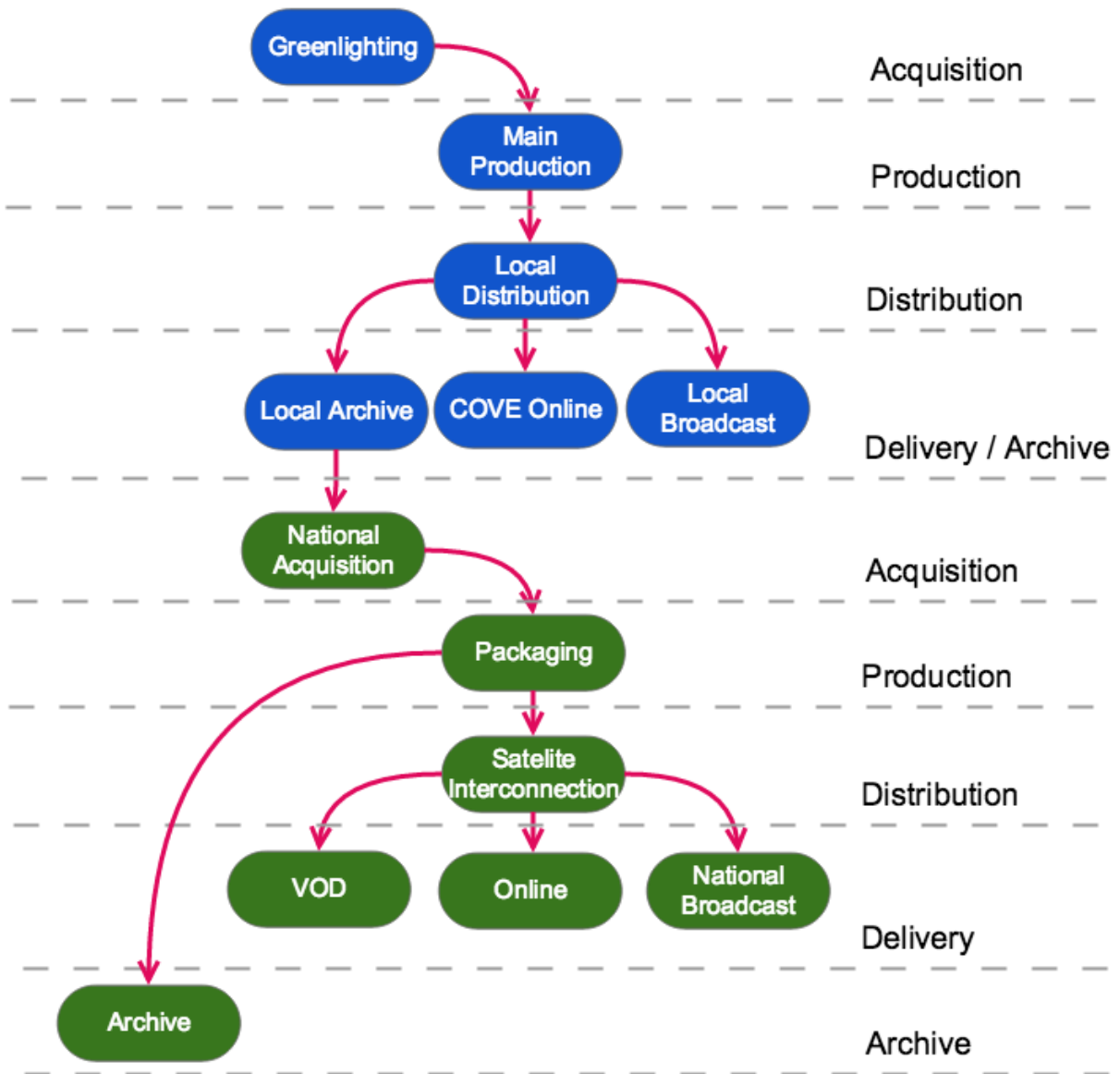
## Transition

After content has completed the entire cycle at least once, the following transitions could occur:

- Content could be included in online archives and available for streaming (rights permitting), like the American Archive of Public Broadcasting or the Louisiana Digital Media Archive.
- Content could be pulled from Archive to run through the cycle again from the beginning.
- Content could be jumped to Acquisition and then to Distribution and Broadcast.

[Figure 2](#) show how content could repeat the content lifecycle.

Figure 2: Content Moving Through Lifecycle Stages



## The Interconnection System

At the Distribution phase during the content lifecycle phase, PBS distributes content to the Interconnection System.

The Interconnection System is the telecommunications system facilitating content exchange between a network of Public Media entities, including PBS, APT, NETA, and various PBS and non-PBS public broadcasting stations. Since its inception, the Interconnection System has gone through five iterations (v1 - v5) and ongoing enhancements are continually being applied. This metadata specification is intended to be implemented in a future enhancement.

This section provides an overview of past versions of the IXS. This is not intended to provide technical operating specifications. See [\[PBS-TOP1\]](#) and [\[PBS-TOP2\]](#) for detailed specifications.

## **v5 Interconnection System Overview**

The v5 Interconnection System delivered content solely through digital satellite transmissions. Satellite distribution depends on a small number of uplink sites across the country and on participating stations with specialized equipment that had been distributed when v5 was deployed. Uplink sites are few because of prohibitive setup costs, and both the uplink sites and station receivers were prone to experience “rain fades” or snow outages caused by local weather phenomena degrading satellite signals to or from the satellite transponders in space.

Occasional quality degradation or technological glitches aside, content delivery in v5 has been reliable due to redundant content feeds.

- There are 7 linear feeds that provided content in various formats and time zone variants. See [\[PBS-TOP1\]](#) for the format specifications.
- The v5 system also provided a Non-Real Time (NRT) feed, which delivered pre-recorded files sequentially and unidirectionally from PBS to recipient stations using IP-over-satellite technology.

Scheduled content is obtained from various sources: some is sent to PBS on video tape, some is pre-recorded in a digital video file, and some is delivered via another linear feed over satellite or direct fiber link. Once the content is all aggregated at PBS, PBS distributes the content on schedule over the IXS, and the individual stations chose which feed to use for downloading or streaming.

PBS sets the national schedule for the linear feeds. PBS, APT, and NETA set the NRT feed schedule. Occasionally, inconsistent metadata input from all the various content sources may erroneously indicate that a feed is present in a schedule when it is not.

The sequential nature of both the linear and NRT feed means that a station may miss the time window to access or download a particular file or stream because of technical issues or weather problems. A station may choose to cache the entire daily output of an NRT stream locally to sort through later.

## **Enhancement Overview**

Metadata is information that describes an asset, formatted in a way to be recognized by both computers and humans. A significant enhancement to the IXS will be to overlay a hierarchical system that uses metadata within the Interconnection System to enable non-linear file delivery. This means that the metadata system overlay will allow station traffic systems to request specific video asset files from the centralized storage on the IXS at any time as long as the rights window for that content is still valid.



The ability to demand specific video content whenever schedulers want — as opposed to waiting for content to be delivered during a certain time window via linear or NRT stream — means that searching for video content on the Cloud should be streamlined and optimized.

The new metadata system is designed as a centralized database with near-real-time updates and an exposed API that station traffic systems can use to consume data directly or create interfaces as is needed. The API allows any system to query for specific metadata to find content.

Because of metadata's key role in content discovery, standards about what metadata is tagged and how it is tagged need to be established rigorously. This is especially important given the various organizations and system processes that deliver video files to the IXS. A set of standards for all new video content, whether station-submitted or distributor-submitted, will reduce duplicate submissions and improve the ability to uniquely identify content.

The metadata system borrows and relies heavily on precedents set by the Entertainment Identifier Registry (EIDR), which is a global registry that provides a unique ID for media content based on select metadata entries. This model's strength is its capability to group tangible assets under abstract concepts and tag both the abstract concepts and the tangible assets with searchable and editable metadata.

The IXS metadata system is detailed further in the following section.

# 3 The Metadata System

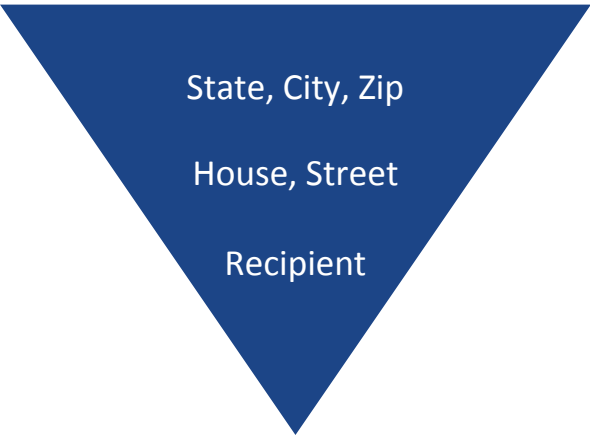
The IXS metadata system depends on a hierarchy of abstract concepts.

A hierarchical abstract metadata model is a system that provides searchable order to particular items. Items can be defined by specific metadata descriptors, and a hierarchy can be built out of logical relationships between the items. A user can then use the hierarchy to pinpoint an exact item they want.

## A Non-Broadcasting Abstract Metadata System

The U.S. Postal System wants to find Fred Rogers in order to give him some letters. They could begin their search with his name, but that starting point could turn up a bevy of Fred Rogers. Then the U.S. Postal System would have to turn to other information that they'll have on hand, like City, State, or Street Address, to narrow down the right Fred Rogers.

A more efficient way to find Fred Rogers is by using the following abstract hierarchy:



Model	Precision	Example
City, State, Zip Code	General	Pittsburgh, PA 15650
House, Street	Less General	1 Mister Rogers Street
Recipient	Exact	Fred Rogers

In this hierarchy, City, State, and Zip Code are abstract models. The order in which they are defined helps to streamline the search. So for the abstract model “City” to be of optimal use, the State should be specified beforehand. Following the hierarchical order reduces confusion, like trying to find the Neighborhood of Makebelieve in Charleston, WV instead of Charleston, SC.

Each abstract model can be loaded with metadata descriptors, which are identifying details understandable by both humans and machines.

Depending on the search’s purpose, some metadata descriptors may only be helpful in certain situations. For example, the “Recipient” model could contain metadata descriptors like Fred Rogers’ Height and Weight. If the mailman had to identify Fred Rogers by physical appearance, these metadata descriptors would be invaluable. However, for the U.S. Postal Service’s purpose of dropping mail off at the right house, the most useful metadata descriptor attached to the Recipient is his Name.

In the case of broadcasting, the goal is to make content easily discoverable. This includes both content distributed on the Interconnection System and content stored within a single broadcasting entity.

For an abstract metadata hierarchy to work in Public Broadcasting, the first step is identifying the most useful broadcasting abstract models, their relationship with each other, and the most useful metadata descriptors needed to accomplish common search goals. The second step is to build a system to ensure that the metadata descriptors for each broadcasting abstract model are specified in a timely way.

## Abstract Models in Broadcasting

In the non-broadcasting example, the U.S. Postal Service relied on abstract models like Area Code, State, Zip Code, Street Address, Recipient Name.

In public broadcasting, main abstract models are as follows:

- Franchise (optional)
- Series
- Season
- Sub Series (optional)
- Sub Season (optional)
- Episode

This list also reflects the expected hierarchy. In this hierarchy, everything ultimately drills down to the identity of an exact Episode. From there, a station may want to find specific tangible assets like Versions and Manifestations:

- Episode
  - Versions
  - Manifestations

Each of these abstract models and tangible assets need to be further defined by metadata descriptors.

## Types of Broadcasting Content Metadata

PBS is a large entity with many moving parts, and one department's search goal is different from another department's search goal. As a result, the broadcasting industry needs a lot of metadata descriptors in order to enable all types of searches.

At the same time, lumping all the possible metadata into a single category would be terribly unhelpful. Logical categories to better organize metadata are as follows:

1. **Technical/Structural Metadata** describes the technical aspects or renderings of a piece of content.  
Examples: HD Level, Aspect Ratio, Audio Type
2. **Relational Metadata** describes the relationships between metadata concepts.  
Example: Franchise contains a Series; Episodes are associated with One Time Onlys.
3. **Descriptive Metadata** is metadata that describes the content of the show. These fields become the basic identifying features of a show.  
Examples: Title, Primary Genre, Primary Language, Primary Genre, etc.
4. **Business Metadata** describes how and when content can be used, as well as who can use the content and what organization and people should be credited for roles or funding.  
Examples: Deals describing intellectual property rights and Usage Window information.
5. **Schedule Metadata** covers data generated by schedules and actual broadcast times.  
Examples: Channels, Feeds, and Schedule information.  
This metadata type will be addressed in a later phase.
6. **Preservation Metadata** contains information needed to archive and preserve content.  
This metadata type will be addressed in a later phase.

Each of the [Abstract Models in Broadcasting](#) can have any number of metadata fields that various searches can use to pinpoint a particular concept's unique identity. For example, a Series could have the following metadata values:

- **Title**="Downton Abbey"

- **Synopsis**=“Life in the Edwardian country house of Downton Abbey has its ups and downs for the Crawley family and their cadre of servants.”
- **Genre**=“Historical Drama”

Having many metadata descriptors is crucial to truly identify a piece of content. But imagine if a person insisted that he be addressed as “Male, 5 feet and 9¾ inches tall, 29 years old and 4 months, with a predilection for mince pies and lemonade...” Finding a particular person using all those details can definitely work, but so would a simple identifier generated specifically for a unique set of metadata descriptors, like a full name or social security number.

## Identifiers

While one could assume that using the Series Title “Downton Abbey” is a unique enough to track all associated Seasons and Episodes associated with this Series, relying on an identifier formatted like *SeriesTitle+SeasonNumber+EpisodeName* can get rather cumbersome. One station may then want to truncate the Title to something shorter, while another station may choose to represent the Season as S1 or S01, thus creating multiple, unnecessary identifiers for the same asset.

In short: there are many possible schemas, formats, and policies for assigning an ID, which can cause duplicate entries under different IDs on different systems or can lose content altogether due to miscategorization or missing metadata.

Each piece of content created for and or entered into the IXS needs to be registered in a standardized way to prevent the quirks of individual station from inhibiting content discovery.

For the metadata system, PBS uses the following four identifiers:

1. EIDR
2. UID
3. NOLA identifiers
4. Package Numbers

**EIDR** IDs are new to the Interconnection System, and they are the key in making a large portion of content metadata interoperable with external platforms and entities. The Entertainment Identifier Registry provides this global identifier for media content upon registration, and this ID is recognized by any organization or platform that uses EIDR — including Nielsen, Rovi, SMPTE, Netflix, CableLabs, Comcast, and others. Registering all PBS content in EIDR reduces misidentification, duplicated entries, or poor search visibility outside of PBS.

The strength of EIDR’s uniqueness and interoperability allows for increased automation; for example, correcting any metadata in EIDR disseminates this correction automatically to all platforms recognizing the EIDR ID.

The EIDR ID is still not the only ID used in PBS because EIDR is a public registry, so any piece of content will become visible to all third-party competitors once registration is complete -- even if production is not finished.

**UIDs** are IXS identifiers assigned to a piece of content before EIDR IDs are obtained. This ID is used until the content is ready for distribution and should never be exposed externally. Upon publication, the EIDR ID should be used instead.

**NOLA** code identifiers are PBS in-house identifiers assigned to specific pieces of content scheduled for the linear feeds. It has two parts: the Root code and the Episode code. The Root code is created for a Series, and the Episode code is appended to the Root code to create a Series-Episode relationship. NOLA codes get exposed externally upon publication. NOLA codes are not necessarily unique for content, but rather codes for association (i.e. an Episode may have multiple NOLA codes). Because PBS issues NOLA codes, content from other contributors to the IXS will not have a NOLA code assigned.

**Package Numbers** are PBS in-house identifiers assigned to Releases and Manifestations. These IDs are used to track specific versions being distributed on the IXS. This ID gets exposed externally upon distribution.

All of these IDs work to track content (and all its metadata) as it moves through the broadcasting lifecycle. Ultimately, all content should have an EIDR ID upon distribution to continue tracking versions, uses, and corrections.

## The Impact of an Abstract Metadata Hierarchy in Broadcasting

Once metadata descriptors are specified for a piece of content, the values are visible and accessible by all workers at all stages in the content lifecycle. The metadata system can then use those metadata descriptor values to funnel search queries down to a precise data point, like the recipient's name in the mailing example.

Thus, required metadata fields need to have values specified as soon as possible so that decision-makers at crucial points in the content lifecycle can be informed.

Some specific examples include the following:

- A programmer would like to create a programming schedule. If expected description, genre, usage rights, and duration are documented in the Acquisition stage and properly updated throughout the Production stage, then searching for possible fits will be faster.

- A network system analyzes bandwidth capacity and assesses the impact of traffic on content options. Expected technical metadata should be made available during the Acquisition phase and updated throughout the Production phase.
- A traffic coordinator at the Distribution and Delivery stages can adjust the schedule using Episode and Release metadata to determine if the content is appropriate and up-to-date.
- A content producer can specify Deal metadata at the Acquisition phase to authorize access and enforce access levels throughout the Production stage.
- Content producers can use metadata from the Delivery phase to determine how many organizations are consuming their content to prioritize future projects.

## 4 Metadata in the Broadcasting Content Lifecycle

The following table gives an overview of what [Metadata Types](#) should be defined at what [Lifecycle Phase](#). The third column lists examples of required metadata fields.

The specific abstract metadata fields that are required to describe broadcasting content are listed and detailed in Part 2: Specific Fields and Values.

Lifecycle Phase	Metadata Types	Example
<a href="#">Acquisition</a>	Business/Legal Descriptive	<ul style="list-style-type: none"><li>• Organization (if new)</li><li>• Usage Windows</li><li>• Deals</li><li>• Series</li><li>• Season</li><li>• Some Episode information (e.g. title)</li><li>• Franchise and Sub-Series, if applicable</li></ul>
<a href="#">Production</a>	Descriptive Technical Business/Legal	<ul style="list-style-type: none"><li>• Episode (episode length)</li><li>• Credits (cast/crew)</li><li>• Story</li><li>• Source Material and Segment</li></ul>
<a href="#">Distribution</a>	Descriptive Technical Business/Legal	<ul style="list-style-type: none"><li>• Episode</li><li>• Story</li></ul>



<a href="#">Delivery</a>	Technical Schedule	<ul style="list-style-type: none"> <li>• Releases (with appropriate In</li> <li>• Manifestations</li> </ul>
<a href="#">Temporary/Archive</a>	[TBD]	[TBD]