Media and entertainment metadata governance

Richard W. Kroon

Director of Technical Operations, Entertainment Identifier Registry Association, USA

Richard W. Kroon is the Director of Technical Operations for the Entertainment Identifier Registry Association. He has previously held senior positions with the Motion Picture Association of America, Technicolor, MovieLabs and Southbay Motion Picture Technologies. In addition to being an award-winning videographer, Mr Kroon is the author of numerous technical papers and several books, including 'A/V A to Z' and '3D A-to-Z'. Mr Kroon has a BS in computer science from the University of Southern California, an MBA from Auburn University, an MA in media and communications psychology from Touro University Worldwide and post-graduate certificates in film and television from UCLA Extension's Entertainment Studies.

Entertainment Identifier Registry Association, 1416 N 13th St., Boise, ID 83702, USA
Tel: +1 818 207 3037; E-mail: rkroon@eidr.org

Cameron Crosby

Partner, WaveSeven, USA

Cameron Crosby is the cofounder of WaveSeven, a consulting firm servicing the media and entertainment industry. He has over 25 years of professional service experience delivering world-class solutions, leading and growing professional service organisations, and recruiting and mentoring great talent. Mr Crosby has worked with all of the major studios as well as other top entertainment companies involved in film, television, music, streaming, publishing, videogame and entertainment events/ticketing. He has worked across the industry spectrum with special emphasis on intellectual property management, rights management, digital distribution, metadata governance and financial systems.

WaveSeven, 5306 La Forest Dr., La Canada, CA 91011, USA
Tel: +1 626 705 0705; E-mail: cameron.crosby@waveseven.net

Abstract   This paper delves into the pivotal role of data-driven strategies in media and entertainment, unravelling the symbiotic relationship between content and data in shaping industry success. The paper explores how effective data governance is essential for navigating the complexities of content creation, distribution, usage and adaptation, and addresses the challenges and nuances unique to the industry. The cited research highlights the multifaceted nature of data governance, emphasising the need for strategies that align with a complex and evolving industry. The insights in this paper help pave the way for media and entertainment entities to harness data governance as an enabler of innovation and excellence, rather than a burden to bear.

KEYWORDS: data, data governance, data-driven, standards, content strategy

DATA DRIVES THE CONTENT BUSINESS

Bill Gates famously said, in his 1996 essay titled 'Content is king', that 'Content is where I expect much of the real money will be made on the internet'.[1] Gates was building on the earlier assertion that Content is King from Sumner Redstone, who turned National Amusements, a chain of movie theatres in the northeast USA,

first into Viacom and then into what is now Paramount Global. Redstone's point was that 'People don't watch CBS. They watch what's on it'.[2] The ultimate foundation of Redstone's belief was provided by J.W. Click and Russell N. Baird, writing in 1974 for *Magazine Editing and Production*, who said 'Content is king. It is the meaning that counts. Form and technical considerations, though important, cannot substitute for content'.[3] They are all correct. Today's multinational media conglomerates are producers, distributors, exhibitors, broadcasters and streamers. They often focus on technology, platforms, brands and other very important issues, but at their core, they are content companies, and content is the foundation of their supply chains.

Everything a media business does should be in service of maximising the value derived from their content and measured against its contribution to their key content strategy goals:[4]

- maximise content exploitation and measure performance (if you can't measure it, did it actually happen?);
- align content to consumers (never rely on consumer lock-in or catalogue volume and just hope for the best);
- expedite direct-to-consumer penetration in a fully digital workflow (reducing distribution delays and costs — which, for long-tail content, should approach zero);
- drive efficiency across the portfolio in a frictionless supply chain through fully automated distribution;
- integrate into the broader media industry ecosystem (we are all partners in the supply chain and need to work together for our mutual benefit); and
- allow for data-driven decision making (quantify what you have measured and then act on those insights).

Content is undeniably king, but the content 'kingdom' requires strong governance and excellent data: the 'coin of the realm'.

Achieving one's content strategy and profitability goals requires quality, available and reliable data, all of which are wholly dependent on a fit-for-purpose programme of data governance. Just as content organisations have traditionally fostered a creative culture, they must also foster a strong data culture. According to a recent study:

> 'building a data culture has a direct impact on a company's ability to get ahead or stay ahead of their competition … [and] having a strong data culture is linked to achieving revenue goals, with those in the top-tier more likely to have exceeded their revenue goals in the past 12 months'.[5]

DATA AND DATA GOVERNANCE

Data, in its essence, is *raw* information; information is refined, curated data; knowledge is useful and, most importantly, *used* information.[6] At the foundation of it all, however, is quality data. If the data is wrong, incomplete or unavailable (ie does not exist; exists but cannot be found; or is findable but unusable in its present form), then all subsequent uses risk fundamental flaws. Bad decisions may be made. Process cost and friction can increase while process speed may decrease. Sales opportunities may go unnoticed, leading to reduced profits and potential harm to the enterprise. The best hedge against all of this is quality data, which is to say, curated data, secured and made available via a proper data governance strategy.

Data governance enables and operationalises a set of internal standards that define how an organisation creates, collects, stores, processes and uses its data. At its core, data governance orchestrates an organisation's people, processes, technology and data within a structured governance framework with clearly defined roles and responsibilities that operate under a set of common policies and procedures, as noted in Figure 1. It is not, as
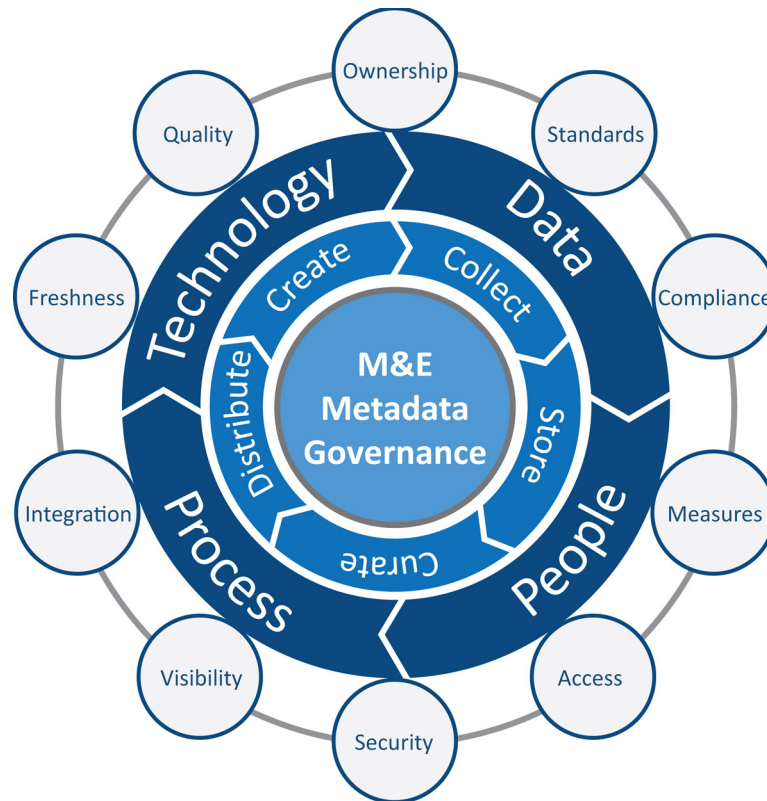
Figure 1: WaveSeven's M&E Metadata Governance Framework

some see, a 'bureaucratic evil' or a roadblock to creativity or productivity. In the past, data governance focused almost exclusively on compliance and enforcement,[7] and earned a deservedly bad reputation. Such defensive approaches protected data but did not clearly align with the overall organisation's mission to use and positively and profitably exploit the content. Effective data governance should be neither defensive or offensive, but progressive in anticipating and adjusting to the needs of both the industry and the organisation.

The Gartner consulting group would have us avoid any accrued ill-will by talking about 'information governance' and its focus on 'the effective and efficient use of information in enabling an organisation to achieve its goals'.[8] Whatever we call it, a successful data governance strategy must be tailored to an organisation across five primary dimensions:

- *Who* — who owns the data, who is responsible for it, who are the stewards of it, and who can use it?
- *How* — what processes, procedures and systems are employed in the creation, collection, storage, curation and dissemination of the data?
- *What* — what is it, what does it mean (how does the business define its data), what are its key parameters (how fresh must it be, how accurate must it be, how complete must it be, etc to be useful to the organisation)?
- *Where* — where is the data stored, where can it be found, where can it be used (and where can it not be used)?
- *When* — when in the life cycle is the data available and ready for use?

A well-thought-out data governance plan will also recognise several key data

characteristics. Data may be (in various combinations):

- *objective* — it is a fact;
- *subjective* — it is open to interpretation;
- *contextual* — what the data is/how it is interpreted depends on the situation or timing;
- *durable* — it does not change over time; and
- *ephemeral* — it is specific to a point in time or tailored to a very specific use.

Durable and objective data is generally the most reliable because there is a single source of truth — it is a fact and always will be — it does not change over time and does not depend on the current situation or one's perspective. However, there is a role to be played by the subjective (genre, synopsis, ratings, etc) and ephemeral data. Contextual data is a particular problem and can break an overly rigid governance plan, when what was otherwise durable and objective data must be altered or re-interpreted to account for special situations (eg previously blacklisted screenwriters may be shown as originally credited or under their actual names after the blacklist has ended).

'IF IT WERE EASY, EVERYBODY WOULD DO IT'[9]

Data governance is not without its challenges. Implementing data governance in an ongoing operation is akin to building an aeroplane while flying it, but there is no other viable option. There certainly is risk in implementing a practice of enterprise data governance, but what is the risk of doing nothing? You must accept the risk of moving forward and begin the work or you can wait for the inevitable crash, because a soft landing is not likely if data practices are undisciplined.

You may have deeply held beliefs regarding your organisation's data and how it should be managed, but it is better to find common agreement across the supply chain than to survive on the warmth of your conviction that 'at least you're right and they're all wrong'. This agreement begins within an organisation, where multiple interests may operate multiple systems, each tailored to a parochial need. Without effective governance across the enterprise, the time, effort and expense spent repeatedly reconciling these issues and forcing movement through the supply chain will consume a significant portion of the tight margins available. Without effective coordination (regarding standards) across the supply chain, processes will no longer flow quickly and smoothly. They will grind to a halt as humans are brought in to finesse the transition between partners. Without shared identification and commonly understood descriptive data, the cost of integrating an acquired catalogue and the delay getting it ready for sale will diminish the derived value significantly.

Every industry has data governance challenges. What is more, many of these challenges are the same for every organisation:[10]

- understanding the business value of data governance;
- thinking IT owns the data;
- providing limited or misallocated resources;
- housing siloed data;
- maintaining poor-quality data leading to lack of trust in data;
- having poor data context; and
- lacking data controls.

Despite this, a recent survey[11] showed that only 10 per cent of companies have adopted data governance across the whole organisation, while 98 per cent cited a need for additional investment in data and analytics to get ahead or stay ahead of the nearest competition.

Media and entertainment organisations face all these data governance challenges, to be sure, but there are also many that are unique to the industry and uniquely difficult to deal with, including:

- content evolution;
- 'experience' vs 'search' goods;
- release versions;
- multiple identifiers; and
- 'creative' workarounds.

Taken together, these lists capture the primary challenges that a successful media and entertainment industry data governance initiative needs to address, even if just to acknowledge them and note that they do not apply or are excluded from the current plan.

Content evolution

If we learned nothing else from Nick Chapman's (Kevin Bacon) experience in 'The Big Picture',[12] it is that, in Hollywood, nothing stays the same. In that case, his first feature film went from art house to commercial pablum in the death by a thousand cuts that can be the Hollywood development project. Even films that do not go through such fundamental change will often go through multiple titles, writers, directors, cast members and scripts on their way from initial pitch to final release. A project may start at one studio, be cancelled, go into turnaround and later be picked up for production by a different studio. A work's initial release format can change, transforming a concept initially envisioned as a theatrical movie into a multi-part television mini-series and vice versa. As the content evolves, so must the descriptive metadata and the system(s) supporting its creation, distribution and delivery.

Take the case of a busted pilot — a pilot produced for a series that was never sold. There are sales systems in current use where a work's type cannot be changed once it has been recorded. So, pilots produced in anticipation of a series sale cannot be entered into the sales system. All associated sales efforts must be recorded and managed by other means until the series is sold or not. If the series is sold, then the series can be created in the sales system and the pilot can finally be entered as the first episode. If the series is not sold, then the pilot can finally be entered as a stand-alone work and sold for distribution on its own. This flaw in some legacy systems is because, when these sales systems were built, their creators failed to account for the fact that the fundamental nature of an audiovisual work is not immutable. The little plastic bit on the end of a shoelace is always a little plastic bit on the end of a shoelace, but a pilot can become a movie, and a movie can become a pilot — and may even be both at the same time in different territories and distribution channels. The mercurial nature of audiovisual works extends into their descriptive metadata from the ever-changing multiplicity of titles that may be used over a work's life down to the duration that can change with every version of each release. Nothing but change is certain, and inflexible systems and supporting practices will increasingly limit an organisation's potential.

'Experience' vs 'search' goods

The manufacturing and marketing of mass-produced goods is reasonably straightforward. There are specific model numbers to reference when ordering, detailed specifications and tolerances when manufacturing, etc, so the associated data is almost entirely objective (and durable) in nature. There are few grey areas. But entertainment products are 'experience goods',[13] where a consumer cannot be sure if they will be satisfied by the product (price, quality, features, etc) until after they have consumed it. This means that the data used to describe, market and measure audiovisual

works is necessarily full of grey areas of interpretation and preference.

Experience goods are in contrast to highly standardised 'search goods'[14] (eg consumer products). Search goods are distinguished from experience goods by having specific qualities that can be determined by the consumer before purchase. For example, if you have liked McDonald's hamburgers in the past, then you will like them in the future. If each Big Mac were the result of the unique creative vision of the person manning the griddle with access to a nearly unlimited variety of ingredients, customers would have to carefully consider their purchasing decisions with each visit and McDonald's would have to completely revamp its marketing approach.

Because movies and television are experience goods, significant effort is put into describing these entertainment products in a way that will highlight their unique nature, differentiating them in the marketplace and attracting the attention of those who are most likely to value the product — want to pay to see it (with their time and/or money); will be glad that they have seen it; will recommend others see it; and will be favourably disposed to associated products (ie those featured in commercials, product placements, etc) and interested in buying them. This naturally leads to a profusion of subjective data to describe these highly subjective products. Even otherwise objective factors may be subject to contextual adjustments to help maximise audience attraction.

The two most cited examples of subjective or contextual entertainment data used to attract consumers are genre and synopsis. There is no one or right answer when it comes to specifying these values (eg the sci-fi drama 'The Martian'[15] won a Golden Globe award for Best Picture — Musical/Comedy,[16] presumably because there was less competition in the comedy category that year[17]). There are also contextual data points, where subjectivity

is layered on top of objective data. Billing order for a cast list is objective, but the order in which to list the cast depends on the context. (For example, Cantinflas, who played the supporting character Passepartout in 'Around the World in 80 Days',[18] was given top billing above David Niven, as the lead character Phileas Fogg, in Latin American markets because Cantinflas, a popular Mexican comedian, was locally more famous.[19]) All of these variations are examples of version management issues.

Release versions

Content continues to evolve even after creation. There was a time when the 'locked cut', 'answer print' and 'final mix' signalled an end to the creative process. Digital production and distribution have made creating different cuts easier which has led to an explosion of versions.

In 2010, Carolyn Giardina, Tech Editor for *The Hollywood Reporter*, wrote about how the release of James Cameron's 'Avatar'[20] 'changed the rules forever' by requiring the creation of 110 different versions to support its initial theatrical release.[21] This included 18 different versions for the US domestic market and 92 for international release, all spanning three aspect ratios, Scope (2:39:1), flat (1:85:1) and Imax (1:43:1); two visual geometries (2D and 3D); and various combinations of subtitled and dubbed versions in 47 languages beyond the original English (and Na'vi). Then, in 2023, Giardina wrote again about how the theatrical release of 'Avatar: The Way of Water'[22] required 'a whopping 1,065 unique delivery versions [that] included combinations of 2D, 3D, HDR, 4K, varying light levels, aspect ratios, [multiple frame rates, including] a high frame rate of 48 frames per second, a range of audio formats, 51 languages supported with subtitles and 28 languages supported by dubbing'.[23] That is almost a full order of magnitude (10✕) increase in the number of unique versions in just 13 years.

Granted, these 'Avatar' releases are outliers, but they are bellwethers of the industry-wide trend toward version specialisation designed to optimise the viewer experience and eke out the maximum possible financial return for each picture. In 2010, the original Avatar release (110 versions) was an extreme outlier. At the time, Ted Gagliano, President of Feature Post-Production at 20th Century Fox, said, 'No studio has ever faced what we faced on ['Avatar']'.[24] Today, the typical Marvel tentpole picture requires around 500 unique versions to fulfil all of the various delivery requirements for different presentation formats, content localisation, etc.[25]

Version management is not just an issue for the producer or distributor at the head of the theatrical supply chain. A single multiplex theatre may require multiple versions of the same work to accommodate the unique characteristics of each auditorium; a single television broadcaster may require multiple versions, each tailored to terrestrial broadcast or catch-up television. A streaming service provider must distribute multiple versions to adhere to diverse content restriction protocols that differ across territories, in addition to managing numerous language variations and more. The proliferation of versions continues well after initial release. A theatrical film may have multiple creative cuts in addition to versions produced to comply with local censorship standards, runtime restrictions, etc. The movie 'Blade Runner'[26] has had seven different creative cuts (workprint prototype, San Diego sneak preview, US theatrical release, international theatrical release, US broadcast version, 'The Director's Cut' and 'The Final Cut').[27] David Lynch felt that the extended version of 'Dune'[28] deviated so far from his vision for the film that he had his name removed, substituting Alan Smithee as the director (standard Directors Guild of America practice at the time for films disowned by their directors) and Judas Booth as the screenwriter (a combination of Judas of Iscariot, the betrayer of Christ, and John Wilkes Booth, Abraham Lincoln's assassin, as the Writers Guild of America did not have a standard pseudonym for such cases).[29] In the USA, 'Demolition Man'[30] presented a future where all restaurants were Taco Bell, while in some international markets, all restaurants were Pizza Hut, thanks to product placement deals. In 'Inside Out',[31] Riley played either ice hockey or soccer to help increase the local cultural appeal in different markets.

Version management is also not just a content issue. Each version of the content has data that defines it (title, cast, synopsis, genre, duration, parental advisory, etc.). There is near-universal agreement on the authoritative nature of the metadata published by a work's creator to describe the work's initial, domestic (or home territory) release. But then things begin to diverge. There are natural variations caused by language localisation — the English synopsis will not be the same as the Spanish synopsis, but it may also not be just a direct translation as that would not take into account cultural and target market differences. Land wars have been fought in Europe over genre classifications: everyone has their own list (which is naturally the best), their own hierarchical decomposition into more precise genres (which is naturally the most logical) and their own way of generating compound/hybrid genres (which is naturally the most descriptive). This plethora of incompatible genres further complicates issues because traditionally, genre is driven by marketing and then used for analytics, where the fact that no two genre lists map cleanly from one to the other makes meaningful competitive analysis something of a black art. Even the definition of what is or is not a version is not settled science. For each version of a work, there are also multiple versions of the descriptive data, usually produced by different parties for different purposes but also including iterations of the metadata from the same source, which introduces a metadata version control problem on top of everything else.

The complexities of content and data version management continue throughout the content supply chain across all territories and distribution channels and on into the 'return loop' where data is reported back up the supply chain in the form of financial performance, audience measurement, revenue, royalties and residuals. The data management challenges continue, now multiplied, as all the downstream distributors and sub-licensees report back common information in a myriad of forms and formats.

Multiple identifiers

Unique identifiers are sadly not unique. According to Andrew W. Tanenbaum, 'The wonderful thing about standards is that there are so many of them to choose from'.[32] The same could be said of identifiers. In a recent case study,[33] a long-running television series with 402 episodes over 23 seasons was found to have 24,275 different identifiers (an average of 60 per episode), each assigned for a specific purpose and linked to potentially different — and sometimes outright contradictory — descriptive metadata (even to the extent that the same programme had different episode and season numbers according to different data sources). Each identifier serves a purpose and is useful within its context and user base, so you cannot just get rid of all of them. Accounting systems generate internal identifiers that are often used in a wider context as project identifiers. Media asset management (MAM) systems apply their own identifiers to track all the content recorded therein. Licensing systems will assign an ID at one end of a transaction, while the receiving licensing system at the other end of the transaction may assign it a different ID. One particular Entertainment Identifier Registry (EIDR) record contains 85 alternative IDs for the same feature film,[34] including 13 different identifiers issued for the same film from the same source.

'Creative' workarounds

Because of all the content and data variations inherent in the media and entertainment industry, in addition to all the common data governance challenges faced by any business unit, it is generally impossible to impose a single, common view across an entire enterprise. In this context, data creators and consumers are a bit like teenagers: you can enforce a strict curfew, but unless you make allowances for special situations, they will just sneak out a window and come home with a tattoo. If data governance is purely defensive — focused on rigid, proscriptive controls with penalties for non-conformance — people who are just trying to get their work done will see data governance as a hindrance to their business rather than a critical aid. Instead of finding ways to work within the data governance framework and help improve it over time, they will just work around it in increasingly creative and ultimately unsustainable ways.

WHERE DO WE GO FROM HERE?

The structure of the traditional three-act play is often described as 'Act One, get your hero up a tree; Act Two, throw rocks at him; Act Three; get him down'.[35] Our organisation data hero is definitely up a tree. Media and entertainment companies drive their content pipelines on data, but that data is often uncontrolled and seemingly uncontrollable. In fact, it is frequently controlled by multiple people using separate systems, which do not talk to one another (the people or the systems). When forced to integrate, systems are often left alone and a translation layer is added on top so that the least amount of friction is felt by the users. This only encourages the creation of more disconnected systems with divergent data practices: a.k.a. data silos. It is more effective, efficient and economical in the long run to unify behind a single vocabulary and identification of the authoritative source(s) per value/field.

In addition to the traditional challenges of data governance faced across all industries, content companies must also overcome a series of unique challenges stemming from the highly variant nature of creative content as they try to manage their data and guide its evolution into actionable information and organisational knowledge. While specific solutions may not be one-size-fits-all, is there at least an approach to a solution that can guide the entire industry to a successful conclusion that gets our hero out of the tree?

A DATA GOVERNANCE PLAYBOOK

Data governance allows an organisation to leverage automation and limit manual intervention throughout the creation, curation and dissemination of its content; it helps eliminate redundant efforts and reduces rework; and it allows an organisation to focus on its content strategy and objectively measure its performance against its goals.

The process begins with an organisational data inventory, including an objective and critical self-assessment of an organisation's data management maturity, including:

- *Data* — what is it; who creates it; how do they find it; and how do they use it? Is the source authoritative; how is that determined; and is that source available to others? How does it drive the business? What are the associated risks (of bad data or data leaks)?
- *Standards* — what internal data standards and standardised processes are in place and how are they applied; what standards are being used within critical supply chains; what standards are employed by key business partners; etc?
- *Governance* — what controls are in place to help ensure that the correct data is collected, stored, protected and presented where, when and how needed to those

with a legitimate need for and right to the data?
- *Systems* — what technologies are employed in support of the above?

People, process and technology must all work together towards a common goal within data governance, each informing, but not dominating, the others. Data governance efforts often begin by selecting the technologies that will be employed, or just assuming that which is already in place will be suitable for the tasks ahead, but that is a bit like building a house starting with the plumbing before the foundation, or assuming that the existing structure will function as a boutique hotel when it was originally constructed as a service station. The technologies are critical, but their selection and implementation should be driven by the data, standards and governance employed, rather than driving how those will be defined. Once the over-emphasis on technology has been addressed, process and controls often come to the fore. But, again, those responsible for data governance cannot be allowed to dominate. If the right people are not leading the effort, if the staff employed are not properly trained and motivated, no amount of process or advanced technology will deliver the desired goals. Data governance must be a synthesis of co-equal partners.

There may be an urge to implement a data governance plan, declare victory and stop there. That will eventually lead to abandoning data governance, writing off the investment as a bad idea because the plan failed to keep pace with the evolving enterprise, or because no one reviewed the existing data to determine how well it conforms to the new requirements and then took the necessary steps to resolve any issues. As noted in the data maturity scale shown in Figure 2, governance is not a single, static item. It is a continuum that evolves in sophistication and reach as the organisation's data management matures.
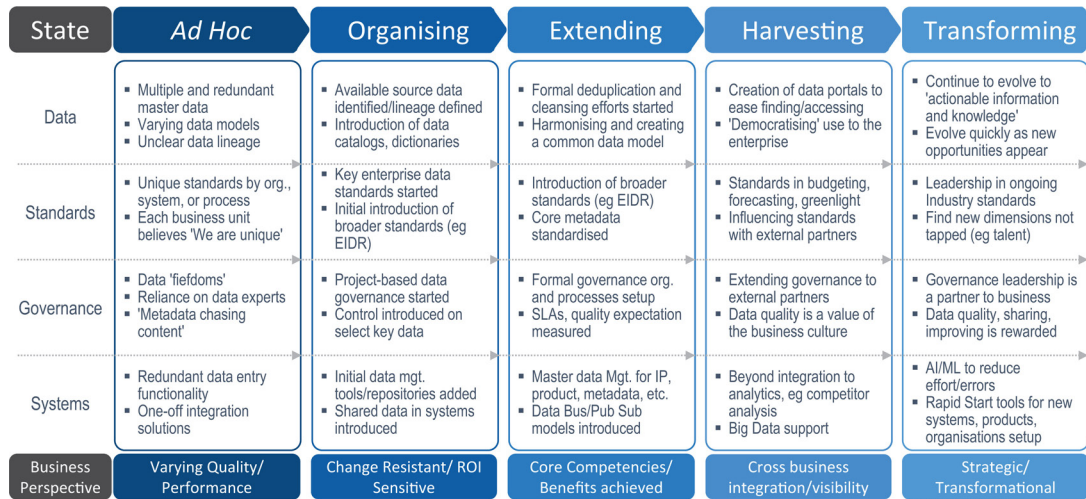
State	Ad Hoc	Organising	Extending	Harvesting	Transforming
Data	▪ Multiple and redundant master data ▪ Varying data models ▪ Unclear data lineage	▪ Available source data identified/lineage defined ▪ Introduction of data catalogs, dictionaries	▪ Formal deduplication and cleansing efforts started ▪ Harmonising and creating a common data model	▪ Creation of data portals to ease finding/accessing ▪ 'Democratising' use to the enterprise	▪ Continue to evolve to 'actionable information and knowledge' ▪ Evolve quickly as new opportunities appear
Standards	▪ Unique standards by org., system, or process ▪ Each business unit believes 'We are unique'	▪ Key enterprise data standards started ▪ Initial introduction of broader standards (eg EIDR)	▪ Introduction of broader standards (eg EIDR) ▪ Core metadata standardised	▪ Standards in budgeting, forecasting, greenlight ▪ Influencing standards with external partners	▪ Leadership in ongoing Industry standards ▪ Find new dimensions not tapped (eg talent)
Governance	▪ Data 'fiefdoms' ▪ Reliance on data experts ▪ 'Metadata chasing content'	▪ Project-based data governance started ▪ Control introduced on select key data	▪ Formal governance org. and processes setup ▪ SLAs, quality expectation measured	▪ Extending governance to external partners ▪ Data quality is a value of the business culture	▪ Governance leadership is a partner to business ▪ Data quality, sharing, improving is rewarded
Systems	▪ Redundant data entry functionality ▪ One-off integration solutions	▪ Initial data mgt. tools/repositories added ▪ Shared data in systems introduced	▪ Master data Mgt. for IP, product, metadata, etc. ▪ Data Bus/Pub Sub models introduced	▪ Beyond integration to analytics, eg competitor analysis ▪ Big Data support	▪ AI/ML to reduce effort/errors ▪ Rapid Start tools for new systems, products, organisations setup
Business Perspective	Varying Quality/ Performance	Change Resistant/ ROI Sensitive	Core Competencies/ Benefits achieved	Cross business integration/visibility	Strategic/ Transformational

Figure 2: WaveSeven's Metadata Maturity Scale

When implementing a data governance strategy in a content-focused organisation, there are four opportunities for significant data management improvement:

- implementing common nomenclature;
- implementing common identification;
- providing flexibility within bounds; and
- creating a golden record.

Common nomenclature

Globally, there is regular disagreement among media and entertainment taxonomies. For example, the episodic content structure of Series-Season-Episode in the USA is Brand-Series-Episode in the UK. There is common agreement on 'Episode', but sharp disagreement on 'Series'. To make matters worse, Japanese broadcasters do not use the concept of Series or Brand — they just have a number of Episodes with very similar titles. Within an organisation, it is possible (and necessary) to come to agreement on a core set of terms — or, failing that, at least agreement on how one maps to the other and the contexts within which each will be used. Between organisations, it may not be possible to come to complete agreement (viz., the US/UK 'series' disconnect), but

you can at least work with your key supply chain partners to agree on the conceptual definitions and data models, then use a glossary to translate between partners who must maintain separate nomenclature (recognising that some specificity necessarily will be lost in the translation).

The earlier discussion of 'versions' lumped two distinct types of versions — creative and technical — into one bucket, but it is important to distinguish between the two when formulating an overarching data management strategy. In 2010, Sony Pictures Entertainment studied the problem of categorising an audiovisual work including all of its derived versions. After analysing several different options (dubbed *alpha*, *beta*, etc), they settled on the alpha structure[36] (now known as the Sony Alpha) that breaks every audiovisual work into three categories:[37]

- *Title:* the work in general — all costs, work products, etc roll up to the Title;
- *Alpha:* a standardised description of picture, audio and/or content differences — a.k.a. creative cuts,[38] often identified by a unique editorial timeline; and
- *Asset:* the derived technical versions — objects that can be viewed.

According to the Sony model, an Alpha exists when a change is made to a programme (inclusive of everything from the opening distribution logo through to the last of the end credits or final logo/brand message) due to a story change (a change to the creative content, but not including changes focused on the viewing experience), for legal/rights reasons (including swapping out licensed footage, changing the music cues, etc), and/or as required by local censorship or broadcast standards and practices.

An Asset is then a recording of the Alpha version, which could be rendered to film, tape or disk in an analogue or digital format. Different Assets may be generated for technical reasons, including new/different timed text tracks (subtitles), audio tracks (dubbing), aspect ratios, colour space/dynamic range, bitrate encoding, etc. These are all grouped under the general heading of 'technical versions'. The individual components may be grouped together in a single deliverable asset or separated into component modules for separate delivery, as needed.

When EIDR was being designed, it adopted the Sony Alpha model for its work hierarchy, ultimately changing the names to:[39]

- abstraction;
- edit;
- manifestation.

While retaining essentially the same definitions. (EIDR adds a '0' layer above the original 3 for collections: franchise, distribution bundle, series, season, etc).

So, when one talks about the 1,065 different versions for 'Avatar: The Way of Water', a relatively small number of those are creative cuts or Alphas (25 such Edits are identified in the EIDR registry[40]), while the vast majority are technical versions (ie Sony Assets a.k.a. EIDR Manifestations).

Common identification

Successful content data management requires unambiguous content identification (if you can't identify it, you can't sell it and you certainly can't automate it[41]) along with accurate descriptive metadata. Thus, it is necessary to reconcile all the IDs in play so that you can begin to automate and measure and adjust.[42] EIDR expresses the core tenants of identifier management using two Rules. First, you must accept that 'The Highlander Rule' ('There can only be one') does not apply to identifiers.[43] Rather than eliminate all other identifiers, you must find a way to manage them using a shared identifier that can be used 'across organisations and in multiple applications, connecting as needed to and from other identifiers coined for other purposes. One that can capture the relationships among various types of resources that use different content standards and metadata structures'.[44] Cross-referencing third-party identifiers is a core function of the EIDR Content ID registry. Next, to be useful, a globally unique, shared identifier must satisfy 'The Tolkien Rule':[45]

> One ID to rule them all,
> One ID to find them,
> One ID to bring them all and in the global ecosystem bind them.
> The utility of EIDR Content IDs is extended by their assurance that each ID will be:[46]

- *globally unique* (each ID references a specific thing);
- *permanent* (once issued, they never go way);
- *consistent* (they always return the same result);
- *open* (not proprietary);
- *large* (enough IDs for everyone);
- *resolvable* (via simple web resolution).

So, while an organisation may have multiple internal IDs from multiple, special-purpose systems and its supply chain partners will have a suite of IDs of their own, they can all be brought together using a common EIDR ID.

Flexibility within bounds

Having no structure and allowing unlimited flexibility is not the answer — that way lies madness. Data governance is instituted to create order from such chaos, but the opposite extreme is equally damaging. A governance practice that is too rigid, too proscriptive or too difficult to change will break. If data governance is perceived as a data jail or a mechanism to punish transgressors, then it has already failed. People who are just trying to get their work done despite a restrictive governance system will constantly strain against the guardrails, causing the system to break even sooner than it might otherwise. Worse yet, users may ignore the governance system entirely and choose to work around it rather than within it, leading to hidden pools of deviating data and non-compliant practices that may go unnoticed until a catastrophic breach occurs. So, instead, create a participative plan that actively seeks regular contributions from the bottom up, that encourages people to be creative within prescribed bounds and identifies blockers early so satisfactory solutions can be found. Such a data governance system will grow and evolve with the enterprise while working to keep the data safe and the gears turning smoothly.

The golden record

At the centre of the data governance plan must be a single, authoritative record for each described object — each creative work in the abstract, their derived creative cuts (edits), their derived technical versions (encodings), unique collections, etc. From that can be derived custom views or augmentations as needed, but always tying back to one central source. Each data element in the enterprise must have a defined source within the enterprise and a data steward, responsible for its collection, creation and curation. As soon as there are two independent sources, there will be uncertainty as to which source is correct. This follows from the adage that 'A man with one watch always knows what time it is, but a man with two watches is never sure'. Maintaining a single source with replication also eliminates redundant collection and curation efforts and potential sources of error.

This is not to say that the central object record is static. Different data will be required at different levels of precision at different points in a described asset's life cycle. For example, a release date cannot become a fact until after the work has been released. Up to that point, it is, at best, a plan. Early in the life cycle, the object is an aspiration. The nature of the available data also changes as you move down the association tree from abstraction to version to encoding. For example, the duration may be expressed in minutes at the abstraction level or rounded to a broadcast timeslot of 30 or 60 minutes, but an edit can have a more specific duration as it represents a particular editorial timeline, which can be measured down to the frame. An edit may then be described in minutes and seconds. An encoding, which represents a particular viewable asset, may have its duration specified down to the frame. It would make no sense to provide (or even allow) this level of precision at the abstraction level as the abstraction can represent a multitude of edits, each with a slightly different duration.

The desire to curate and disseminate authoritative descriptive metadata that can be exchanged across the enterprise and among supply chain partners has led to the creation of the International Broadcasting Convention (IBC) Accelerator Media Innovation Programme for the Authenticated Data Structure (ADS), now under development by a consortium of international media and entertainment industry partners.[47] ADS will allow metadata authorities to publish certified data sets (where the source is verified and the data is guaranteed to be unchanged) where those who need the descriptive metadata (downstream distributors, sub-licensees,

broadcasters, streamers, etc) can discover it (via a search service by identified work, language, territory, data publisher, etc) and locate it (download it directly from the ADS service or follow the provided pointer to obtain it from a commercial data provider). This will give content creators and other data authorities greater control over how their works are described in the marketplace and consumers greater assurance that the descriptions are complete, accurate and tailored to their interest.

SUMMARY

The intricate web of media and entertainment thrives on content, but quality and reliable data fuels its success. The fusion of creative excellence and strategic data governance defines the industry's future, enabling seamless content delivery, targeting and satisfying diverse audience preferences, and unlocking additional value throughout the supply chain for the individual participants. Rather than viewing data governance as a barrier or a cudgel, embrace it as a catalyst for innovation. With this approach, media and entertainment companies can transcend challenges and carve a path toward sustained excellence and profitability in a dynamic and highly competitive landscape.

References

(1) Karbasfrooshan, A., (2021), 'The origins and paradox of "content is king"', 1st April, available at: https://contextisking.com/2020/12/06/the-origins-and-paradox-of-content-is-king (accessed 12th October, 2023).

(2) Archive of American Television, (2016), 'Sumner Redstone on the phrase "content is king"', available at: https://www.youtube.com/watch?v=kums8SGXHoo (accessed 12th October, 2023).

(3) Karbasfrooshan, ref. 1 above.

(4) Haralson, R., (2023), 'Leveraging EIDR for Title & Version Strategy', presentation at the EIDR Annual Participant Meeting, Culver Theater, Culver City, CA, 30th May.

(5) Wakefield Research, (2022), 'The Alation State of Data Culture Report: Q1 2022 Report', available at:

https://www.alation.com/resource-center/reports/state-data-culture-report-2022-q1 (accessed 12th October, 2023).

(6) Saber, M., (2021), 'The differences between data, information, and knowledge, and why you never find it when it's needed!', 15th November, available at: https://www.linkedin.com/pulse/differences-between-data-information-knowledge-why-mohamed-1e (accessed 12th October, 2023).

(7) Alation, (2023), 'Data governance is valuable: moving to an offensive strategy', available at: https://www.alation.com/resource-center/whitepapers-ebooks/data-governance-is-valuable-moving-to-an-offensive-strategy (accessed 12th October, 2023).

(8) Gartner, (2023), 'Information governance', available at: https://www.gartner.com/en/information-technology/glossary/information-governance (accessed 12th October, 2023).

(9) Marshall, P., (1992), 'A League of Their Own', see: https://doi.org/10.5240/58B6-6CEB-E482-F6F1-BA31-A.

(10) Sandwell, D., (2023), 'The top 7 data governance challenges organizations face and how to address them', Quest, 16th January, available at: https://blog.quest.com/the-top-7-data-governance-challenges-organizations-face-and-how-to-address-them (accessed 12th October, 2023).

(11) Wakefield Research, ref. 5 above.

(12) Guest, C., (1989), 'The Big Picture', see: https://doi.org/10.5240/A0B1-295F-3955-F8F7-B72B-I.

(13) Nelson, P., (1970), 'Information and consumer behavior', *Journal of Political Economy*, Vol. 78, No. 2, pp. 311–329.

(14) *Ibid*.

(15) Scott, R., (2015), 'The Martian', see: https://doi.org/10.5240/7DBC-D245-E700-BF93-138C-P.

(16) Hollywood Foreign Press Association, (2016), 'The Martian', Golden Globe Awards, available at: https://www.goldenglobes.com/film/martian (accessed 12th October, 2023).

(17) Bishop, B., (2016), 'The Martian won the Golden Globe for Best Comedy because everything is a lie', *The Verge*, 10th January, available at: https://www.theverge.com/2016/1/10/10732126/golden-globes-2016-best-comedy-musical-the-martian-nonsense (accessed 12th October, 2023).

(18) Anderson, M., Farrow, J., (1956), 'Around the World in 80 Days', see: https://doi.org/10.5240/0AEC-059D-559D-4376-A63B-Y.

(19) IMDb, (1993–2023), '*Around the World in 80 Days*: Trivia', available at: https://www.imdb.com/title/tt0048960/trivia/?ref_=tt_trv_trv (accessed 12th October, 2023).

(20) Cameron, J., (2009), 'Avatar', see: https://doi.org/10.5240/C840-E543-A58F-5C59-1B1C-T.

(21) Giardina, C., (2010), 'How "Avatar" changed the rules of deliverables', *Hollywood Reporter*, 24th March, available at: https://www.hollywoodreporter.com/news/general-news/how-avatar-changed-rules-deliverables-22027 (accessed 12th October, 2023).

(22) Cameron, J., (2022), 'Avatar: The Way of Water', see: https://doi.org/10.5240/5494-6F7D-368F-0AF7-0A3D-C.

(23) Giardina, C., (2023), '"Avatar: The Way of Water" team raced to create 1,065 delivery versions for theaters', *Hollywood Reporter*, 22nd February, available at: https://www.hollywoodreporter.com/movies/movie-news/avatar-the-way-of-water-team-1235331142 (accessed 12th October, 2023).

(24) Giardina, ref. 21 above.

(25) Giardina, ref. 23 above.

(26) Scott, R., (1982), 'Blade Runner', see: https://doi.org/10.5240/EA73-79D7-1B2B-B378-3A73-M.

(27) Wikipedia, (2023), 'Versions of Blade Runner', available at: https://en.wikipedia.org/wiki/Versions_of_Blade_Runner#:~:text=Seven%20different%20versions%20of%20Ridley,Cut%2C%20and%20the%20Final%20Cut (accessed 12th October, 2023).

(28) Lynch, D., (1984), 'Dune', see: https://doi.org/10.5240/7ECC-15E0-7585-62DA-DA8C-V.

(29) IMDb, (1990–2023), 'Dune: Trivia', available at: https://www.imdb.com/title/tt0087182/trivia/?ref_=tt_trv_trv#director-trademark (accessed 12th October, 2023).

(30) Brambilla, M., (1993), 'Demolition Man', see: https://doi.org/10.5240/2031-528E-D451-B973-9F6C-0.

(31) Docter, P., Del Carmen, R., (2015), 'Inside Out', see: https://doi.org/10.5240/05C4-4304-CE7C-9153-63F1-S.

(32) Tanenbaum, A. S., (2011), 'Computer Networks', 5th edn, Prentice Hall, New York, NY.

(33) Haralson, ref. 4 above.

(34) Zemeckis, R., (2004), 'The Polar Express', see: https://doi.org/10.5240/0E8C-C250-E484-9794-A9F2-0.

(35) *Bridgeport Herald* (1897) 'At the theatres', 21st November, *Bridgeport Herald*, p. 9, col. 3, available at: https://news.google.com/newspapers?id=-NUyAAAAIBAJ&sjid=7wAGAAAAIBAJ&pg=2275,4401184&dq=up-a-tree+throw+get-him-down&hl=en (accessed 12th October, 2023).

(36) Ward, P., Yasgur, J. R., (2015), 'US Patent No. 8990174: System and Method for Identifying Media Assets', available at: https://image-ppubs.uspto.gov/dirsearch-public/print/downloadPdf/8990174 (accessed 12th October, 2023).

(37) Sony Pictures Entertainment, (2013), 'Rights Management and MediaCentre Overview', available at: https://slideplayer.com/slide/8734666 (accessed 12th October, 2023).

(38) Yasgur, J. R., (2012), 'Universal Media Identifiers in Action: Sony Pictures Case Study', 27th February, available at: https://eidr.org/documents/Sony_Lessons_20120229.pdf (accessed 12th October, 2023).

(39) Kroon, R. W., (2020), 'Introduction to the EIDR Data Model', available at: https://www.eidr.org/documents/Introduction%20to%20the%20EIDR%20Data%20Model.pdf (accessed 12th October, 2023).

(40) Cameron, ref. 22 above.

(41) Kroon, R. W., (2022), 'EIDR for Catalogs', EIDR, available at: https://www.eidr.org/assets/EIDR-for-Catalogs.pdf (accessed 12th October, 2023).

(42) Kroon, R. W., (2016), 'The secret to automating multi-party asset workflows', *Journal of Digital Media Management*, Vol. 4, No. 2, pp. 105–115.

(43) Seidel, C., Ripping, S., Edelman, M., (2020), 'Crawl, walk, run — making your digital supply chain meet the needs of your growing business one step at a time', presentation at OTT.X Summit, 1st September.

(44) *Ibid*.

(45) Kroon, R. W., Drewry, R., Leigh, A., McConnachie, S., (2019), 'The power and promise of identification', EIDR, available at: https://www.eidr.org/assets/The-Power-and-Promise-of-Identification.pdf (accessed 12th October, 2023).

(46) Kroon, R. W., (2016), 'The secret to automating multi-party asset workflows', *Journal of Digital Media Management*, Vol. 4, No. 2, pp. 105–115.

(47) International Broadcasting Convention, (2023), 'Authenticated Data Standard', available at: https://show.ibc.org/accelerator-media-innovation-programme/accelerator-project-authenticated-data-standard (accessed 12th October, 2023).