

M+E

JOURNAL

Charting the Metaverse

The metaverse will prove transformative for M&E. But only when the entire industry gets on board.

22.01

THE COST OF BAD DATA

By Hollie Choi, Executive Director, EIDR



Hollie Choi is the executive director of the Entertainment ID Registry (EIDR). During her 15-plus years in media and entertainment she's become an expert in supply chain automation, metadata governance and workflows, and digital and media asset management systems. hchoi@eidr.org @eidrdotorg

Data Management Review puts the financial impact of bad data within an organization at an average of 15 percent of net revenue annually. Let's do a tiny bit of math. Variety reported that in 2019 (pre-pandemic), the media and entertainment industry was said to have made \$101 billion across theatrical and home entertainment. Fifteen percent of that is approximately \$15.2 billion. Bad data causes this much damage to your bottom line? It's not hard to believe.

Everyone has bad data lurking somewhere in their organization. Many content creators started putting out movies in the early 1900s, others started only a few years ago. Both business models produce the opportunity for bad data. It is understandable. It's not as if anyone at the start said, "Hey, let's make movies and television shows, but first, let's get a taxonomist in here and build a governance team." Data is a byproduct of content production, and therefore, understandably, a secondary concern.

It can be difficult to quantify losses resulting from bad data, or to understand the consequences, because some causes are not always immediately clear. I'm going to suggest that bad data hurts your brand. Don't believe me? Let's say you settle in with your family and your popcorn to stream your favorite show. You turn on your device and navigate to the streaming service of your choice, find the title, and hit play. But the description you read does not match the episode that plays. What I think when I see this is, "Wow, their metadata is lacking proper controls." Most consumers don't think that, though. What they think is, "What is this? This isn't what I want. Wow, this studio is messed up."

In the olden days (you know, five years ago), most people could not identify which studio made their favorite show. Now the major studios have streaming platforms with some variation of their name. The consumer is now drawing a direct connection from the data and content they see to the creator and their brand. If I see bad data, or the wrong content shows up, it causes a ding in our relationship. Enough dings and I might just call the service a piece of junk.

What about the more obvious causes? Operational and system inefficiencies create a huge loss

*THAT'S DATA GOVERNANCE:
finding the authoritative
source of data and then
managing the quality.*

of productivity in that they enable duplication of effort, manual work, broken business processes, and/or poorly designed line of business applications? These are possible to quantify, but it can be challenging. If we were to look at the operations of a single business unit in a single company, I'll bet we could find at least one process that is caused by a lack of automation, which is only used to mitigate data issues from upstream sources. Let's follow that stream metaphor for a minute.

Imagine you're standing next to a stream. Suddenly, you hear a person yelling. They are in the water, holding on for dear life, in danger of drowning. So, you get in the water, safely of course, and you save them. Just then, you hear another person calling out for help, and another, then another. Soon, all you're doing is pulling people out of the water. Would it not make more sense to send someone upstream to find out how those people keep getting in the river in the first place, and put up some sort of fence, or a warning sign at least?

That's what data governance is: finding the authoritative source of data and then managing the quality. Without doing that initial scan of the upstream processes and systems, automation does not make sense. Why would you want to proliferate bad data? Excellent automation cannot prevent bad data from showing up in front of the consumer if there is no fence or warning sign upstream!

What we need is a source of truth. One that is curated by the best data people in the industry. It should have a single, persistent, unique identifier and a defined set of core metadata so that we can identify all our content. Then we could use that to automate the digital supply chain of the entire industry. Then we need to make the registry available to anyone and any system that needs it.

That exists, and it's called EIDR. ■