Why the Collectibles Market Needs Standardization: Sports Cards

Intro

If you've spent any time trying to make sense of the sports card market, whether you're a serious collector or emerging hobbyist, you've probably run into the same issue across all marketplaces: **nothing is standardized**. Varying descriptions, inconsistent data and missing or misattributed card qualities (such as labels for "insert" and "parallel") all contribute to an opaque marketplace that makes buying and selling cards more difficult than it needs to be.

For sellers, free text listings are manual and error prone. For buyers, a lack of standardization makes it difficult to search, compare and verify marketplace listings. It's a major roadblock in bringing the card industry into the 21st century.

So how can this be improved? By building a robust data standardization pipeline that takes messy, human-generated inputs (like ebay listings) and converts them into a consistent format using a defined taxonomy. With a unified taxonomy, it becomes significantly easier to understand, search for and trade collectibles.

Read on to see how prevalent the problem is, and how it can be fixed. Collectors should be collecting, not wasting time battling messy, inconsistent data.
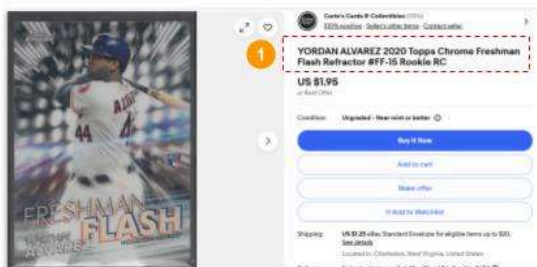

The Problem: No Common Language

The fundamental issue stems from the fact that the card marketplace, unlike those in other industries, lacks a unified system for organizing listings. Collectors are forced to cross-reference marketplace data with checklists and images just to participate confidently. And even after extensive research, the available resources often fall short of providing a clear, holistic view of the market for a given player, set or card.
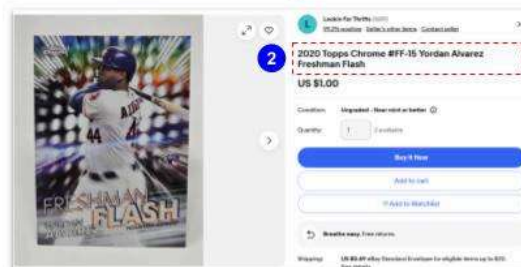
To illustrate how this plays out in practice, let's look at a few examples.

Below are three different ebay listings for the same card, a **2020 Topps Chrome Freshman Flash Yordan Alvarez.**
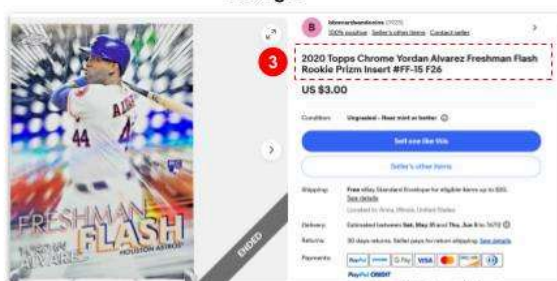

While categories like year, manufacturer and product are listed accurately, key pieces of information are either missing or incorrectly included across all three cards.

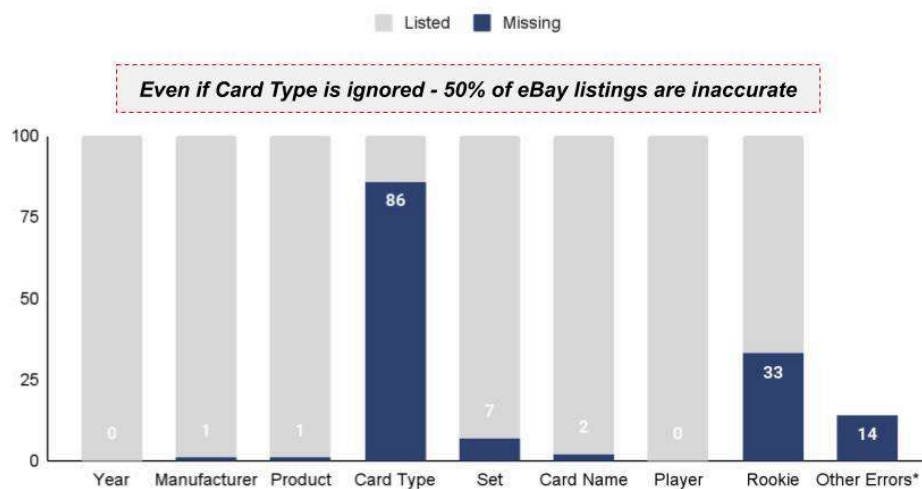|	Listing 1	Listing 2	Listing 3	Complete Listing
Year	2020	2020	2020	2020
Manufacturer	Topps	Topps	Topps	Topps
Product	Chrome	Chrome	Chrome	Chrome
Card Type	—	Insert	—	Insert
Set	Freshman Flash	Freshman Flash	Freshman Flash	Freshman Flash
Card Name	#FF-15	#FF-15	#FF-15	#FF-15
Player	Yordan Alvarez	Yordan Alvarez	Yordan Alvarez	Yordan Alvarez
Rookie	Y	Y	N	Y
Additional Incorrect Information	Refractor	Prizm, F26	na	na

- **Listing 1** miscategorizes the card as a "**Refractor**" and does not list the card as an **insert.** For this set, if the card was a refractor, the refractor type would be specified and it would be numbered, neither of which it is.

- **Listing 2** incorrectly labels the card as a "**Prizm**" insert and includes the text "**F26**" at the end. Prizm is a separate manufacturer to Topps entirely, and F26 does not describe any characteristic of the card itself. It could reflect any number of things, such as the seller's inventory code, card show or auction reference or even something else entirely.

- **Listing 3**, similar to listing 1, excludes the fact that the card is an **insert** and also does not mention that this is a **rookie** card for the player.

While some of these omissions or mischaracterizations may seem minor, *these are identical cards and should be listed with consistent information*. Such inconsistencies place an unnecessary burden on the buyer. For collectors trying to identify cards from a set or theme, or compare listings and prices, issues like this can be frustrating, erode confidence and increase the risk of making misinformed purchases.

But this is just three cards, a more comprehensive analysis of the data may tell a different story.

Let's look at **100 additional listings** we have identified for this card on eBay.

Applying the same analysis to a larger dataset, this chart shows how often information is missing across the 8 key fields identified above. Notably, only **11 out of 100** listings include all the required details.

*Inaccurate information included in the listing of the card, such as mislabels or extraneous text

As shown, the most frequently omitted category is "Insert." But if it's missing so often, does it really matter?

It does! And for several important reasons.

- It's one of the five core Card Types that every baseball card falls into, and it's especially important to collectors, many of whom aim to complete sets within or across types like base cards and inserts.

- Prices can range from multiples of 10x (sometimes more) depending on the card type

- For this specific example, it provides context to the Rookie designation, which is often valued differently depending on the type of card the Rookie is (base vs. insert vs. auto)

So how prevalent is this issue?

Based on a random data set of **~150k** cards, we estitime that **80%** of listings on ebay do not correctly identify the listing as one of the five core card types

1. **Base**
2. **Auto**
3. **Parallel**
4. **Insert**
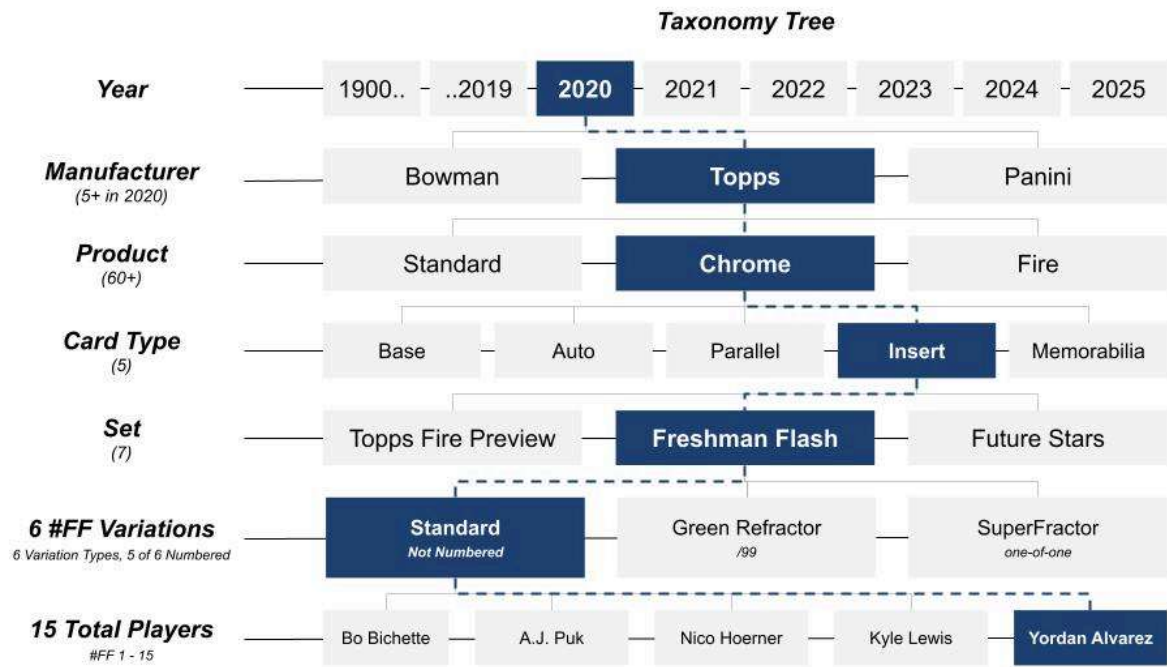5. **Memorabilia (Relic, Game-used etc.)**

This matters because Card Type isn't just a label, it's fundamental to how collectors assess value, rarity and relevance. But even if you overlook the inaccuracies in Card Type, errors and omissions across the other seven key fields can result in a combined 50% error rate.

Together, these issues present a serious challenge for the market.

The Solution: Create a Taxonomy

To fix this, we need to build a **taxonomy**, a structured hierarchy applied to a standardized set of data that can be used to characterize what kinds of cards exist and how they relate to one another. Think of it like a biological classification system, but for cards: *Card Types* like *Base, Parallel, Insert, Autograph* and *Memorabilia*, would exist below groupings like *Year, Manufacturer, Product*, and above groupings like *Set, Card Name and Player*.

To make this clearer, let's look at an example using our **Yordan Alvarez #FF-15** case study

Taxonomy Tree

Then outside of this there could be additional components, such as

- *Is the card a rookie*
- *Is the card numbered (such as the Refractor variations above)*
- *Has it been graded (if so what is the cert number)*

So what was once many variations of the same listing

- **YORDAN ALVAREZ 2020 Topps Chrome Freshman Flash Refractor #FF-15 Rookie RC**
- **2020 Topps Chrome Yordan Alvarez Freshman Flash Rookie Prizm Insert #FF-15 F26**
- **2020 Topps Chrome #FF-15 Yordan Alvarez Freshman Flash**

We now have one standardized listing

- **2020 - Topps - Chrome - Insert - Freshman Flash - Yordan Alvarez - Houston Astros - Rookie**

With the right taxonomy, the card market will gain clarity and consistency through accurate labeling. In the near term, this will improve search, discoverability, and pricing at scale. In the

long term, it will lay the foundation for model training, sophisticated analytical tools and enhanced collector research capabilities.

And this is just the beginning, with a clean taxonomy, a plethora of opportunities in the market is opened. Better pricing tools, more accurate collection management systems, and various AI technologies that currently are not feasible will become commonplace and beneficial to collectors. An agreed upon taxonomy for standardization is foundational infrastructure, and something the card world has lacked for too long.

If you're working in this space or just passionate about making the hobby smarter, I'd love to hear from you. Let's build the language the card market needs.

mparsons@onecovet.com