STAR WARS DATABANK
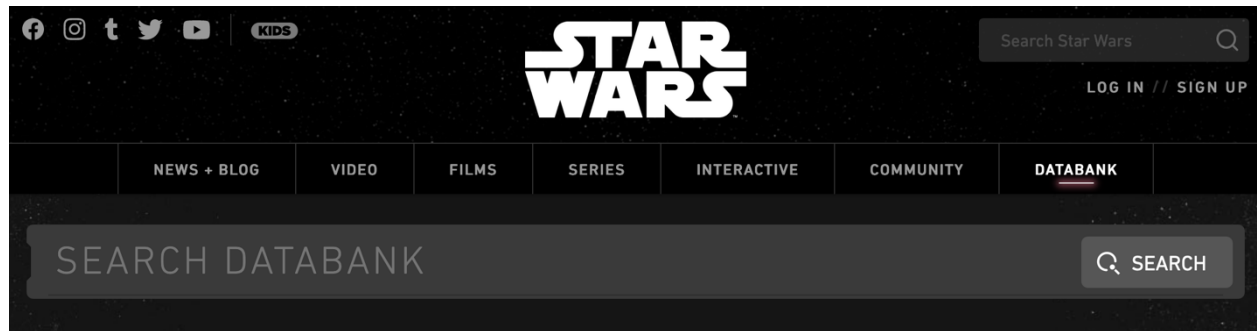## WHEN A GALAXY OF POTENTIAL GETS FORCE CHOKED

An Information System Analysis
By
Kate Polley
IST 616, FALL 2020

INTRODUCTION
� ᚖ᠎ᚖ᠎73 ᚖ᠎ᚖ᠎73

The *Star Wars* Databank is the official canonical database that exists as a part of StarWars.com (starwars.com/databank). The current version is the third incarnation of the *Star Wars* online encyclopedia. The Databank only covers items found in the movies (the trilogies and anthology films) and the television series (animated and live action); items found exclusively in the new canon books or comics are not included, with the exception of characters that will be part of the upcoming multimedia project *Star Wars: The High Republic*. Unlike the original version of the *Star Wars* encyclopedia, Expanded Universe/Legends and behind-the-scenes content are also outside of the scope of the Databank. Databank items include characters, locations, vehicles, weapons, organizations, species, and more.

STRUCTURE
ᚖ᠎ᚖ᠎73 ᚖ᠎ᚖ᠎73 VI

At the minimum, each Databank entry includes a brief description of the item and a list of the works where that item appears. More significant entities may include a number of other **metadata attributes**, as well as a written history of the subject, related items from the Databank, and links to other areas of the website like the gallery and video collections. Below is the top of Ahsoka Tano's Databank page as an example.

AHSOKA TANO

Ahsoka Tano, a Togruta female, was the Padawan learner to Anakin Skywalker and a hero of the Clone Wars. Alongside Anakin, she grew from headstrong student into a mature leader. But her destiny laid along a different path than the Jedi.

APPEARANCES	AFFILIATIONS	GENDER	DIMENSIONS	SPECIES	VEHICLES	WEAPONS
Star Wars Rebels, Star Wars: The Clone Wars	Jedi Order	Female	Height: 1.7m	Togruta	Alderaan Cruiser, Jedi Starfighter	Lightsaber, Ahsoka's lightsabers

The metadata provided for each entry can exclusively be categorized as **descriptive metadata** as described by Joudrey and Taylor (2018, p. 190). Attributes include appearances in Star Wars media, organizational affiliations, gender, dimensions, species, locations, vehicles, and weapons. Much of this metadata, including vehicles, locations, and weapons captures **relationship data** (Joudrey & Taylor, 2018, p. 190); referring to the example of Ahsoka Tano above, the metadata available shows, for instance, that the entity "Ahsoka Tano" is related to the entity "Lightsaber" in that the lightsaber is a weapon that Ahsoka uses. Other attributes, including gender, dimensions, and species may be more accurately described as **identifying data** (Joudrey & Taylor, 2018, p. 190) as they are used to describe identifying characteristics of the given entity.


It is important to note that the **values** of the metadata elements given are not always complete, even considering the limitation that only the films and TV shows are covered by the Databank. Most elements can have multiple values, but relevant values are frequently missing. For example, in the entry for Anakin Skywalker, the only vehicle listed is the Delta-7 Jedi Starfighter, leaving out both the Eta-2 Jedi Starfighter (shown above), a vehicle made iconic in the Battle of Coruscant, the opening scene of *Star Wars Episode III: Revenge of the Sith,* and the *Twilight,* a freighter that Skywalker owned during the Clone Wars and that was seen on screen just as frequently as his Delta-7.

Another kind of metadata that is not present in the Databank but that would be useful to users is **linked data**. Linked data models resources as a graph or network of triple (**subject-predicate-object**) statements (Riley, 2017, p.10) where subjects and objects are the nodes and predicates are the edges. When describing a vast and heavily interconnected universe such as that of *Star Wars,* this kind of data representation would be very useful. The ability to query linked data would allow users to answer such questions as "Who are all of the characters that have possessed the Darksaber?" or "What planets has Obi-Wan Kenobi visited?" which currently cannot be answered by the database. This kind of information could be of interest to fans and even of use to the vast number of creators working on the current canon in order to maintain continuity. If modeled after linked data systems like *Wikidata,* the sources of information as well as qualifying information about the statement predicates could be included in statements. Examples of this, taken from the *Wikidata* page for Anakin Skywalker can be seen to the right. Relating back to the previous examples of Skywalker's starfighters, a linked data system could express that Anakin flew the Delta-7 at the beginning of the Clone Wars and the Eta-2 at the end of the Clone Wars, even providing specific years to further clarify this data to fans.
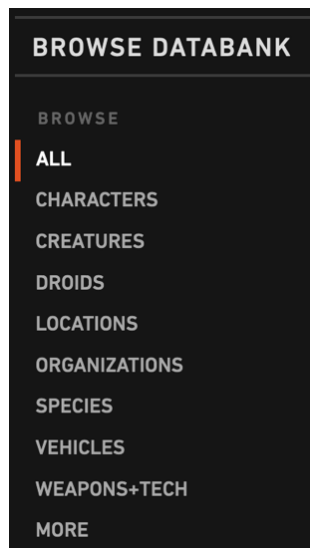
ORGANIZATION

Beyond the scope of individual entities, the organization of the database if fairly minimal. On the landing page, users can find recently added items related to the latest *Star Wars* media. Databank items organized by appearance can also be found by navigating through the "Films" and "Series" sub-sections of StarWars.com, but this requires leaving the Databank. The section that allows users to browse the Databank breaks down the collection of entities into the distinct classes seen in the image on the left. The **classification** of the Databank resources is incredibly **broad**, as it only uses very high-level class divisions without any subdivisions. Despite the underutilization of **hierarchy** in the Databank's current classification scheme, it is clear that this can be considered classification instead of the more general **categorization** because of the presence of certain properties like "**mutual exclusivity**" and "systematic and predictable rules for association and distinction" that Kwasnik (1999, p. 26) describe as necessary for classification hierarchies. For example, consider the two classes "characters" and "droids." There are certainly characters in *Star Wars* that are droids, but the Databank makes clear the rules that are used to sort entities; for example, C-3PO is listed exclusively under "characters," but the broader category of "Protocol Droid" is listed under "droids."

The example of C-3PO begins to illustrate some of the potential missed by not having a hierarchical system with multiple levels, as well as some of the possible problems that could arise with such a system. For example, in a new hierarchy, C-3PO could be considered an instance of "protocol droid," which could then be a subclass of "droid." However, this loses the information that C-3PO is a character, unless "droid" were to be made a subclass of "character." This then begs the question of whether or not all droids can be considered characters, due to the rules of **inheritance** and **transitivity,** which state that a subclass inherits the properties of every superclass above it in the hierarchy (Kwasnik, 1999, p. 25). On the other hand, it could be useful to be able to divide "character" into subclasses like "humans," "droids," and other species to make it easier for users searching for a specific type of character. Another possible area for hierarchy is the class "locations," which could easily be broken into subclasses such as "settlements" versus "planets," and "planets" could then have the subclasses of "core worlds," "mid rim," and "outer rim" for example.

As it stands, the current classification system does not make the searching a very positive experience for users. With only nine classes, eight of which each contain dozens of entries organized only by alphabetical order, general **browsing** is supported, but it can be incredibly difficult for one to find particular entities without knowing a specific name.

501ST LEGION
DATABANK

ALLIANCE HIGH COMMAND
DATABANK

AT-AT PILOTS
DATABANK
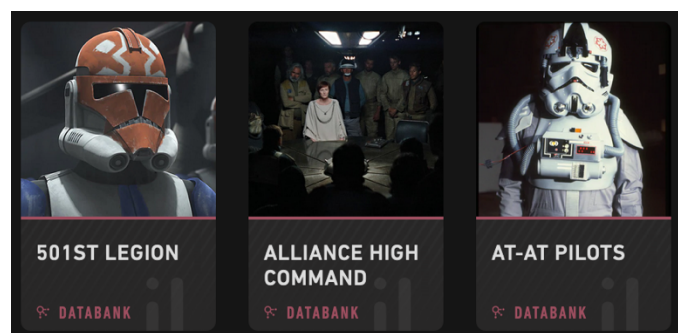
Beyond the classification of entries, searching the Databank is further made difficult by the variety of formats that entity labels, particularly those of characters, can take. To clarify, characters may be listed in any of a variety of formats, including "FirstName LastName," "FirstName 'Nickname' LastName," and more. This variety would seem to make sense based on the fact that, for example, some characters only have a single name, others have formal titles, and still others may have names that appear in different written forms. However, similar cases are not always handled in a similar fashion, which may cause confusion. Consider the following examples that reveal inconsistency in Databank naming.

AMILYN HOLDO   **GENERAL HUX**   **C-3PO (SEE-THREEPIO)**   **R2-D2**
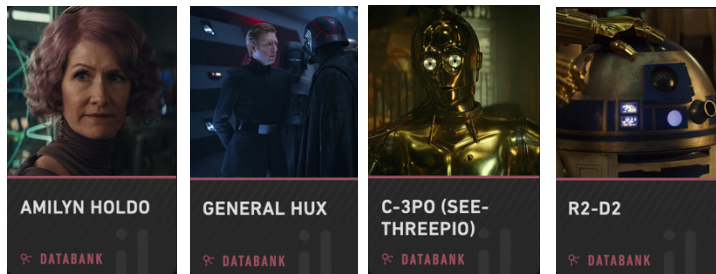
Let us begin by looking at the entries for Holdo and Hux, two characters who appeared exclusively in the Sequel Trilogy. Both were known in the films primarily by their respective titles, Vice Admiral and General. In addition, they both have canon first names given to them, those being Amilyn and Armitage respectively. However, while Holdo is listed under her first and last name, Hux is listed by title and last name. In another example, C-3PO (often written as "Threepio" in text) is listed with both number and phonetic name. However, fellow leading droid R2-D2 (commonly written as "Artoo") is listed only by model number. These are only two of many examples of odd or unclear naming conventions in the Databank.

These cases illustrate how similar characters may be treated differently in the Databank, and that can cause problems for users browsing a system organized alphabetically by first letter (is Hux listed under G for General or A for Armitage?) and for users performing queries on the Databank. For example, a query for "Vice Admiral Holdo" returns several characters and vehicles, but the entry for Amilyn Holdo is not found, resulting in a search with 0% **recall**, because the relevant entry from the Databank was not retrieved, and 0% **precision,** because none of the retrieved objects were relevant to the search. In addition, a search for "Armitage Hux" returns no results at all. This is a very poor performance when searching for what should be easily recognized names and forms of name within the *Star Wars* universe.

As another example, consider Leia Organa. If a user were to search for her common title of "Princess Leia," the entry for Leia Organa is the fourth entry listed in the results, following such items as the WED Treadwell droid (a droid Luke was with when he saw the space battle involving Leia's blockade runner over Tatooine), which calls into question how result **relevance** is calculated by the search engine (Joudrey & Taylor, 2018, p. 130). This is important because there is only one way to search the databank: a general search bar that offers no advanced search fields; in fact, the option on the results page to "refine your search" only offers the ability to search the other parts of the website outside the Databank.

LEIA ORGANA

When using the search bar, entries with names containing your search query will pop up below the bar, but if you elect to click "search" rather than selecting one of the given options, the system appears to run a **keyword search** that looks for query terms within entry titles and descriptions. The search function also does not appear to use any of the metadata elements described at the beginning of this evaluation. For example, a search for "Delta-7," referencing the starfighter discussed in a previous case, returns only the Delta-7 and Eta-2 (because the Eta-2 is compared to the Delta-7) Jedi Starfighters, but not, for example, the entry for Anakin Skywalker, as someone who has the Delta-7 listed as one of his "vehicles."

While the elements being used in a query would need to be addressed in the search algorithm, the poor query performance in regard to inconsistent naming suggests a need for **authority-controlled access points.** Because of the different forms that names can take (i.e., Leia Organa versus Princess Leia versus General Organa), it would make sense to have not only the **authorized access point** specified as the title or label of each Databank entry, but also **variant access points** so that "any search on an alternate form of name will cause a reference to be displayed to the searcher" (Joudrey & Taylor, 2018, p. 374). It would be incredibly beneficial for users to be able to search for various forms of names or labels with greater success than is currently possible.

CONCLUSION

The *Star Wars* Databank is, in theory, a great tool for fans of the franchise to look up resources and learn information about a wide variety of people, places, and things that existed a long time ago in a galaxy far, far away. However, there are quite a few issues with the current version of the Databank that prevent the system from reaching its full potential. These issues and potential upgrades, from updating metadata to expanding classification and introducing new access points, as described in this analysis, are just a few areas that should be explored in order to improve the Databank and better serve its users.

REFERENCES

Joudrey, D. N., & Taylor, A. G. (2018). *The organization of informatio*n (4th ed.). Libraries Unlimited.

Kwasnik, B. (1999). *The role of classification in knowledge representation and discovery*. Library Trends, 48(1): 22-47.

Riley, J. (2004). *Understanding Metadata*. National Information Standards Organization.