

Arctos Agents Roadmap

Disclaimer

This is an attempt to functionally summarize various issues. This is not *my* proposal. Corrections and redirections are exceptionally welcome.

Overview

- <https://github.com/ArctosDB/arctos/issues/6813>

Concerns

1. How do we communicate this? The verbatim agent change involved a LOT of apparently-failed attempts at communicating, can we do better?
2. “Curated by...”: <https://github.com/ArctosDB/arctos/issues/7235>
3. Can we somehow make it harder to use high-quality agents? Maybe existing reports are sufficient.
4. As of 20240110 meeting it looks like we’re proposing with the below, will work with individual collections to migrate any verbatims, etc. (Start with Derek.)

Proposed Structural Changes

1. Remove “requirements” for additional data with Agents.
2. Remove unique key on preferred_agent_name.
3. Add unique “key” on ORCID, Wikidata, Library of Congress, Github, or collectionID
4. Maintain a history of changes.
5. Make all agent information attributes that allow for dates and determiners see
 - [- Arctos Agents - Functional Requirements
 - a. Agent Experiment - [+ Agent Attribute Experiment
6. Delete nothing. Mark attributes as ‘hidden in some UI by some agent’ (or similar). Find some way to maintain a history of whatever remains in table Agent. “Deleted” (including through “update” - which would be flag-and-insert) Agents will be entirely flag+UI; they (along with everything else) will continue to exist. ‘Version’ everything, forever.

Implications of Changes

This is largely driven by <https://github.com/ArctosDB/arctos/issues/7135>, a request to replace verbatim agents with agents. This will make verbatim agents entirely avoidable. Verbatim agents will continue to be a part of agent mergers, and will remain available to anyone who wants to use them. (I will continue to recommend them to incoming collections, for example.)

The lack of controls will, barring redirection here, mean that any string may be a preferred agent name, and an agent may consist of only a preferred agent name. “A”, “Å”, and “👤”, each with no additional information, will be acceptable agents.

The lack of a guaranteed unique “data key” means that there may be any number of agents whose entire data consists of preferred name ‘A.’ This necessarily means that string-based tools (eg, record bulkloader) will not be able to successfully disambiguate them. The agent picker will be expanded to include some form(s) of the “Arctos key” such that eg, [“https://arctos.database.museum/agent/10014199”](https://arctos.database.museum/agent/10014199) will select exactly one Agent.

Considerations

This will provide a mechanism through which proper attribution will be difficult to provide. For example, agents

- Michelle S. Koo
- Michelle S. Koo
- Michelle Koo
- Michelle S Koo
- Michelle Koo
- Machelle Koo
- Michelle K0o
- Michelle Koo
- Koo, Michelle
- MK
- K,M

might all exist. Some might be created with our Michelle in mind, some might have been intended to refer to someone else. Some might be used for one person, or several people, or not at all. (The change history - whatever form it may take - will be intended to help convey create+edit intent, but it will remain possible for a collection entering records to just provide a string and let whatever happens happen.) <https://arctos.database.museum/agent/10014199> is unlikely to be able to receive credit for her activities in that environment. It has been suggested that the Agent Committee may upon request attempt to clean up such data; it is not clear how realistic that may be.

External identifiers (such as ORCID [\[link\]](#) and Wikidata [\[link\]](#)) will continue to serve as “quality indicator flags” (in addition to their inherent functionality). Collections who wish to provide proper attribution to collectors and other donors should be strongly encouraged to create, maintain, and use agents which carry these data.

User Guidance

I hope that the Agent Committee will provide guidance for efficiently operating in this system, but it will also probably not be possible to prevent some collection from creating a million “MK” agents (that is, large masses of maybe-not-so-great data from within which higher-quality data may be difficult to find).

DLM's take on 20240110 meeting: the Committee WILL NOT control what Agents are created nor how they are used. The Committee may find some way to do something clever with data quality flags. Data quality reports will continue to run (eg there will be a notification when an agent does something before their birth date).

Expenses

Maintaining a history should be expected to require significant data. However, storage is relatively cheap and I don't foresee this causing short-term problems. Efficiently accessing such data may require more and more expensive resources, but I expect that to be at least a few years out.