MEMORANDUM

BUREAU OF LAND MANAGEMENT

DATE: **September 21, 2022**

FROM: **Patrick Alexander**

Ecologist, National Operations Center

SUBJECT: **Plant names in Terrestrial AIM data**

Background— The Bureau of Land Management (BLM) collects data about the distribution and abundance of different kinds of plants as part of broader efforts to understand the condition of natural resources on the nation's public lands. The Assessment, Inventory, and Monitoring (AIM; www.blm.gov/aim) Strategy is a standardized national framework for ecological monitoring on BLM lands. Terrestrial AIM, which focuses on upland habitats rather than streams or wetlands, is the largest BLM program collecting plant biodiversity data. The scientific names of plants play a central role in the collection of accurate data and our ability to use that data effectively.

Researchers studying plant diversity separate the task into two components: nomenclature and taxonomy. Consider a set of three personal names: "John Doe", "Doe, John H.", "David Smith". The first two are different forms of one name. The third is a separate name. This list of available names & data about which are variants of each other is the nomenclature. The nomenclature of plants is governed by a set of rules, the International Code of Nomenclature for algae, fungi, and plants (ICNafp). There is a single correct nomenclature.

We don't yet know if there is a single taxon (the two names belong to one person, whose name has changed) or two taxa (two different people). This gives us two possible taxonomies: John Doe = David Smith; John Doe ≠ David Smith. In the context of people, our intuition is that they simply are or are not the same person; find out which and get everyone on the same page. In the context of biodiversity data, however, we should think of different taxonomies as equally valid alternatives. While there is consensus on many aspects of plant taxonomy, this is a field of research rather than a body of static knowledge—there *is not* a single correct taxonomy.

Understanding plant names has both context-independent (nomenclature) and context-dependent (taxonomy) aspects. Standardization and centralization are desirable in nomenclature. We want to use the same plant names clearly and unambiguously across contexts. Documentation and translation are the key concepts for taxonomy. We want to know how plant names map onto taxa in different data collection and analytical contexts, as we work with data across these contexts.

Current practice in Terrestrial AIM— Our current data structures and processes are built around a set of state species lists whose structure and content is largely inherited from the USDA's PLANTS (Plant List of Accepted Nomenclature, Taxonomy, and Symbols; plants.usda.gov) database. The role of the state species lists has not been well-defined. They serve in part to fill a need for attribute data associated with each taxon, e.g.: duration, growth habit, sage-grouse preferred forbs. PLANTS includes most of this information, but often provides a set of possible values where we need a single value. The state species lists also serve in part to provide flexibility to address limitations of PLANTS.

The state species lists and PLANTS use an accepted names list model, enumerating the taxa of a standardized, prescriptive taxonomy. This model assumes uniform taxonomic understanding across data collectors and data users, and across time. Consequently, it does not provide us with the basic data and structures for working with multiple taxonomies. Nomenclatural and taxonomic relationships are not distinguished. This is roughly analogous to knowing there is a relationship between the names "John Doe" and "David Smith", but not what it is—they could be the same person, or they could be friends, relatives, co-workers.

Existing nomenclatural data— The International Plant Names Index (IPNI; ipni.org) and Tropicos (tropicos.org) are the two primary databases for plant nomenclature. They don't fulfill all of our needs for nomenclatural data (e.g., they do not provide an equivalent to the plant codes of PLANTS). PLANTS provides nomenclatural data for plants of the United States to some extent, although the nomenclatural data first needs to be separated from taxonomic data.

Existing taxonomic data— There are several national-level taxonomies available, including PLANTS, the Integrated Taxonomic Information System (ITIS; itis.gov), the Biota of North America Program (BONAP; bonap.org), and Plants of the World Online (POWO; powo.science.kew.org). Many state or regional taxonomies are also available, though many of these are available only in print. No single resource provides the associated attributes needed in Terrestrial AIM, at least not in a form that can be used as-is.

Proposed handling of nomenclature— I have compiled a draft national nomenclature data set based on: consistency with PLANTS (all relationships between plant codes and plant names are maintained as-is); consistency with existing Terrestrial AIM data (all plant codes in Terrestrial AIM are in the nomenclatural data); using data from IPNI and Tropicos to fill in gaps in PLANTS and separate the nomenclatural content in PLANTS from its taxonomic content. This data set is provisionally called TANN (Terrestrial AIM National Nomenclature). The data structure used by TANN is is intended to provide the nomenclatural data that is likely to be relevant in biodiversity data management as compactly as possible (see the attached document, "Nomenclatural Concepts"). I propose that Terrestrial AIM strictly apply TANN—any plant code used in Terrestrial AIM must be present in TANN and be used to mean the same plant name as in TANN. Ideally, over time TANN and PLANTS will converge. If the two converged in all other respects, TANN might ultimately become a documentation of historical usage in Terrestrial AIM, rather than having a role in ongoing data collection. However, there might always be some uses cases in Terrestrial AIM that are not covered by PLANTS.

Proposed handling of taxonomy— Rather than proposing a particular taxonomy, I propose a data structure and conceptual framework for working with multiple taxonomies (described in the attachment, "Nomenclatural Concepts"). If we have multiple taxonomies using a common data structure, we can work across them for analytical and reporting tasks. The basic information provided by a taxonomy is the set of names included within each taxon, and which of those names should be used. A taxonomy may also include associated attribute data about that taxon (duration, growth habit, etc.). A taxonomy can fulfill two basic roles: input or output.

An input taxonomy is used in data collection. It documents how plant names are used by field crews (i.e., what taxon a name refers to). The ideal input taxonomy perfectly matches how plant names are used by a particular field crew. This ideal is intrinsically difficult to achieve, so it is better viewed as the scale on which we measure improvement: The better a taxonomy matches how plant names are used by crews, the better it is at fulfilling the input taxonomy role. The threshold for improvement is low, even if the ideal end state is not attainable.

An output taxonomy is used in data analysis and reporting. Its role is to provide internal consistency in data products and help data users correctly interpret plant names. The ideal output taxonomy varies with the audience. Some audiences are best served by a modern taxonomy that incorporates recent research, some by a conservative taxonomy that prioritizes consistency with historical usage. BONAP and POWO are more modern, PLANTS is more conservative, ITIS is somewhere in between. For users of a particular state flora, that taxonomy might be the best.

At present there remain many details, especially related developing and maintaining input taxonomies, that are beyond my current scope. However, the state species lists are obvious candidates for conversion to input taxonomies. For output taxonomies, it may be best for us to convert the major national-level taxonomies to our common data structure (or create a repeatable conversion workflow; at present conversion is trivial for ITIS and PLANTS) but generally leave work on state or regional taxonomies to user groups who would find those taxonomies useful.

Discussion— Integrating nomenclature and taxonomy into the management of biodiversity data is a difficult problem that has not been solved. My hope is to establish the building blocks we'll need to do it well. One of the challenges is that simpler approaches that work well within a limited context generally do not scale well. I think we are in a good position in this regard, as we're operating at a large enough geographic and temporal scale to encounter these limitations, but not at such a large scale that our practices are deeply entrenched and inflexible.

Attachments—

1. Nomenclatural concepts and AIM data (nomenclatural-concepts_PJA13Sep22.pdf; 24 pages)