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)

Nomenclatural concepts and AIM data

Patrick J. Alexander

Initial draft 20 Apr 2022, current draft 2 Sep 2022.

This document provides an overview of botanical nomenclature, a proposed data format and process for handling botanical nomenclature in AIM data, and a discussion of the limitations of other options.

Section 0. Background: How did I get here? What is the problem?

I started managing biodiversity data in late 2004, as a graduate student working at the New Mexico State University Herbarium (NMC). We stored specimen data for the herbarium in a Microsoft Access database called "Maii'tsoh", developed by Chris Frazier at the University of New Mexico. I worked on the herbarium database throughout graduate school. After graduation I had a post-doctoral position with the herbarium, during which my time was focused on database QA / QC and preparations for migrating to a more capable and well-supported database (Specify). In 2010 I started collecting biodiversity data analogous to the AIM species richness protocol (with a smaller sample area), as well. As this dataset grew, I realized that it needed to be linked to plant nomenclature & taxonomy data. Over a decade or so of collecting biodiversity data and learning how to work with these linked data sets, I improved and periodically restructured the nomenclature / taxonomy data, eventually arriving at the data structure and conceptual framework presented here.

I've become convinced that most data structures and processes for handling nomenclature & taxonomy in the context of biodiversity data are based on a fundamental misunderstanding, the One True Taxonomy concept. There is a single correct taxonomy and our goal, both as taxonomists and as biodiversity data managers, is to find it, update all of our data to follow it, and get everyone else to do the same. For a taxonomist focusing on a particular group of plants, this is reasonable enough. Their research is focused on creating the most accurate understanding of diversity and relationships within that group of plants as possible. The One True Taxonomy viewpoint has obvious appeal to authors of floras, as well, given that (especially in print) a flora must adopt a single taxonomy and surely we'd prefer it to be correct. Non-taxonomists often adopt a weaker version, where we might replace "True" with "Standard". The emphasis is less on the correctness of the taxonomy and more on trying to get everyone to use the same names. One True Taxonomy breaks down as your scope widens to encompass multiple taxonomic works on a particular group of plants, or geographic areas covered by different floras. The correct taxonomy also has an annoying habit of changing over time, since this is a field of ongoing research. Anyone working with biodiversity data beyond a narrow geographic and temporal scale will have to work with multiple taxonomies. For One True Taxonomists, this means there are errors that must be rooted out. Our response should focus on correctness and standardization. However, we will still have to work with multiple taxonomies.

I've arrived at the opposite viewpoint. Relationships between taxonomies are fundamental to understanding biodiversity. Translation between taxonomies is the core function of nomenclature & taxonomy when working with biodiversity data. Translation should be automated, transparent, and reversible. Although there are benefits to standardization, focusing on working with multiple taxonomies is usually more rewarding than trying to convert people to a standard.

I think the situation is partially analogous to GIS coordinate systems. Some degree of standardization is helpful, as working with data in a hodgepodge of different coordinate systems can cause problems. However, no single coordinate system is ideal across all contexts. Organizations and individuals also have idiosyncratic preferences that may or may not make much sense. One True Coordinate System would be a terrible philosophy for GIS software design, focused on getting everyone to agree on a single coordinate system instead of building coordinate transformation into the software. We can't standardize *instead of* being able to transform coordinates. Coordinate transformation is just basic GIS functionality, it's not optional.

Coordinate transformation in GIS software is also helpful for thinking about what an implementation of nomenclatural & taxonomic functionality might look like. Someone writing the coordinate transformation code needs to understand how to do the math for transforming a pair of coordinates from NAD83 UTM Zone 13 to WGS84 Latitude / Longitude. Someone using GIS software needs to be familiar with some basic concepts related to coordinate systems, and to be able to interact with a few user interface elements that specify what coordinate system is used in map display or data import / export. This document is focused on the more detailed level of nomenclatural & taxonomic functionality. We could also compare this level to calculating % cover from tblLPIDetail and tblLPIHeader. Data collectors and most data users should be not be working at this level.[1]

There are also differences between GIS coordinate systems and nomenclature & taxonomy. I think standardizing on a single coordinate system would be comparatively straightforward. Also, One True Coordinate System has not (so far as I know) been a dominant paradigm in the geographic community, while One True Taxonomy is common among taxonomists as well as data specialists working in biodiversity and bioinformatics. Also, coordinate transformation is a mathematical problem that is very well understood and very well implemented in software, while taxonomic translation is poorly understood and poorly handled by most biodiversity data platforms[2]. This means that there are far fewer existing resources for us to build on.

Before I continue, it's worth briefly defining "taxonomic translation" and illustrating some of the problems one might encounter. Generally speaking, any two plant identification resources you encounter will use at least slightly different taxonomies. The taxonomies of USDA PLANTS, iNaturalist, the Integrated Taxonomic Information System (ITIS), Plants of the World Online (POWO), Ackerfield's *Flora of Colorado*, and Allred's *Flora Neomexicana III* are frequently in agreement, but no two are identical for the plants that overlap between them. While individual botanists may try to follow a particular taxonomic resource, it is inevitable that a particular person's "head taxonomy" is not identical with that resource. Whenever we're taking plant names from one taxonomy and recording or exporting them in another taxonomy, this is an act of taxonomic translation. When the source and destination taxonomy are in agreement, this is trivial and hardly worth calling "translation". The *Gutierrezia sarothrae* of PLANTS is the same as the *Gutierrezia sarothrae* of iNaturalist. It is not always obvious whether or not the source and destination taxonomy are in agreement. Comparing *Flora Neomexicana III* to PLANTS: Is *Eriocoma hymenoides* the same as *Achnatherum hymenoides*?[3] Is *Boechera fendleri* the same as *Arabis fendleri*?[4] Is *Cylindropuntia imbricata* the same as *Cylindropuntia imbricata*?[5] If the source and destination taxonomies are well-defined and in an interoperable data structure[6], translations can be scripted. We can design processes to do the translation well, rather than hoping field crews or data users will get it right on their own.

---

[1] I don't think I've made this point well in prior discussions, causing concern that this is way too complicated for field crews. The terradactyl scripts are complicated, too. The data collection UI is important, and separate.

[2] The Symbiota platform for herbarium data (e.g., https://swbiodiversity.org/seinet/) is an outlier. Although not very prominent in the interface, it has a nice taxonomic translation function. If you search the collections and then switch to the "Species List" tab, the "Taxonomic Filter" pull-down menu allows you to select between alternate taxonomies.

[3] Yes! This species was moved to a new genus, with no change in the set of plants that belong to it.

[4] No! This species was moved between genera *and* the set of plants belonging to it changed. *Boechera fendleri* of *Flora Neomexicana III* is much more narrowly defined than *Arabis fendleri* of PLANTS.

[5] No! *Cylindropuntia imbricata* of *Flora Neomexicana III* includes both *Cylindropuntia imbricata* and *Cylindropuntia spinosior* of PLANTS.

[6] These two are.

This is a challenging topic. Nomenclature & taxonomy define the relationship between our data and organisms in the field, but few people collecting or working with biodiversity data have taxonomic backgrounds taxonomy. Since I do, part of my goal is to help interested non-taxonomists get up to speed. I know this document is long and dense. It's as close to an "easy button" as I can provide, but that doesn't mean it's easy!

SECTION 1. NOMENCLATURE & TAXONOMY.

STUDYING BIODIVERSITY—The classification of biodiversity, often called systematics, is a field of research in which patterns of variation among organisms are used to classify them into a set of hierarchically related, named groups. Each group is called a taxon. Each level of the hierarchy is a rank. Various sources of information can be used to understand variation among organisms, including morphology, genetics, ecology, geography, phenology, and cytology. Nomenclature is the subfield focused on applying names to taxa. Taxonomy is the subfield focused primarily on understanding how many taxa there are and the relationships between them. While taxonomy is open-ended inquiry into patterns of variation, nomenclature is relatively rule-bound.[7] The present discussion is limited to plants and focuses primarily on nomenclature at the ranks of species, subspecies, and variety. Taxonomy is discussed as it relates to nomenclature, without consideration of how organisms are grouped into taxa.

CIRCUMSCRIPTION—The circumscription of a taxon is the specification of which organisms fall within it. Circumscriptions usually focus primarily on morphology and nomenclature, but also include at least brief descriptions of ecology, phenology, and other forms of data. Circumscriptions are a combination of intensional definition (characteristics used to assign an individual to that taxon, e.g., a written description of the morphology of the species) and extensional definition (examples of the taxon, e.g., a list of herbarium specimens belonging to the taxon). When we talk about "a taxonomy" rather than about the field as a whole, we mean one particular grouping of organisms into taxa, among various other possible groupings. In other words, a particular set of circumscriptions, usually with a particular geographic or taxonomic scope.

NAMES—We need to apply names to taxa. Common names are often sufficient, but since common names are established by use—whatever people call a taxon is its common name, or one of them—they are inherently ambiguous. So we have scientific names, which must follow the rules of the International Code of Nomenclature for algae, fungi, and plants (ICNafp: https://www.iapt-taxon.org/nomen/main.php). The goal of the ICNafp is to leave taxonomists with freedom to circumscribe taxa as they see fit, while providing a framework to apply unambiguous and well-defined names to those taxa. Once we have a circumscription for a given taxon, the ICNafp rules determine which existing name is given to that taxon, or if a new name or changed name is needed. There isn't necessarily a wrong answer to a question about how many species exist within a particular genus, but there are definitely wrong answers when we ask what names to use for those species.

For the purposes of this discussion, I will use "name" to mean any plant name that follows the ICNafp rules, or any plant name that was intended to follow or is presented as if it follows the ICNafp rules. I use the word "sound" to refer to any name that follows the ICNafp rules, and "unsound" to refer to any name that does not.[8]

---

[7] Sometimes, the fields of biodiversity research as a whole is called taxonomy and the subfield focused on differentiating taxa is called systematics. I prefer to minimize use of the term "systematics" since it is ambiguous.

[8] I think it is useful to have a term for names that meet all the requirements of the ICNafp and a term for names that do not, but the ICNafp does not provide them—which I think is an odd oversight. Using the ICNafp terminology, instead of "sound" we would say a name was effectively published, validly published, and legitimate. In previous drafts I used the words "valid" and "invalid", but these can cause confusion with the conceptually different ICNafp term "validly published". "Sound" and "unsound" are not ideal, but they're the least worst options I have come up with.

There are a few cases like this where I have found that the concepts most useful to me in understanding and working with plant names do not quite correspond with terms defined in the ICNafp and established in use among taxonomists. My use of

The name of a species is a binomial: the genus name plus a specific epithet (ICNafp chapter III §4; e.g., *Ericameria nauseosa*). Additional ranks below the species rank can also be used for infraspecific taxa. The name of a subspecies is a trinomial: the name of a species plus a subspecific epithet (ICNafp chapter III §5). The name of a variety is also a trinomial: the name of a species plus a varietal epithet. A rank marker is usually inserted before a subspecific or varietal epithet (*Ericameria nauseosa* subsp. *consimilis*; *Ericameria nauseosa* var. *oreophila*).

Additional ranks below species are encountered occasionally. Form ("fo." or "f.") is the most common in recent decades. Subvarieties ("subvar."), subforms ("subf."), and unranked infraspecific taxa are rare. Use of multiple infraspecific ranks can result in tetranomials (*Ericameria nauseosa* subsp. *consimilis* var. *oreophila*) or even pentanomials, hexanomials, or heptanomials[9]. Most botanists in recent decades do not use multiple ranks within a species. When multiple ranks are used, their order from most to least inclusive is: subspecies, variety, subvariety, form, subform. Various distinctions between subspecies and varieties have been proposed. In practice there is no agreed distinction between them, except that subspecies is the higher rank when both are used.

A species never has a lone subspecies. It has two or more subspecies, or none. When subspecies are recognized within a species, one of these always has a subspecific epithet identical to the specific epithet. This is an autonym, sometimes also called the nominate subspecies (e.g., *Ericameria nauseosa* subsp. *nauseosa*). This autonym exists automatically once any other subspecies is named. It does not need to be published as a new or changed name. The same applies to varieties, and to subvarieties, forms, & subforms when those ranks are used.

Each name has an author, or multiple authors. The authors are usually given in a standardized form established in the International Plant Names Index (IPNI: https://ipni.org/), e.g. "L." for Linnaeus. The authors are formatted differently depending on whether it is a new name or a name that has been modified from its original form. With the exception of autonyms, infraspecific names have authorship independent from the species they belong to. Autonyms have no separate authorship.

When multiple authors have published the same name, only one of these (the first one published, with some exceptions) is sound. For instance, Linnaeus published the name *Andropogon hirtus* L. in 1753, and Lojacono published the name *Andropogon hirtus* Degen ex Lojac. in 1909. Though identical in spelling of the name itself, these are different names and might refer to different taxa. Linnaeus's name has priority, Lojacono's is unsound. Including authors of names is generally unnecessary if we can be confident that we are not using unsound names, or when we are using a clearly defined reference taxonomy that includes authors. In some cases, though, unsound names have become frequently used and it may not be possible to tell which taxon is being referred to without including the author.

NEW NAMES (ICNAFP CHAPTER V §2)—To create a new name, there are four requirements. 1) It must be effectively published, either by making printed material available or by electronic PDF in a journal or book that has an ISSN or ISBN. 2) The name must follow rules regarding spelling & typography and avoid duplicating an existing name. 3) It must have a written description (covering the overall morphology of the plant) or a diagnosis (a statement of what features distinguish it from a similar or closely related taxon) in English or Latin. 4) A type specimen must be specified. The author of a new name is usually the author of the paper or book in which the name is published, but occasionally a different authorship is given. The author follows the name (*Chrysocoma*

---

nonstandard terminology will likely make some taxonomists queasy, but I feel that strict adherence to ICNafp terminology sometimes forces my writing, or the data structure and scripts written to interact with it, to become unnecessarily convoluted. My intent is to prioritize clarity.

[9] The ICNafp defines the name of an infraspecific taxon as a trinomial, regardless. Additional ranks between the specific epithet and the final epithet can be used to provide a more complete classification, but are not technically part of and do not affect the nomenclatural status of the infraspecific taxon denoted by the final epithet.

nauseosa Pursh). In some cases, the author of the name may indicate that their work is based on the unpublished work of another botanist. Although these cases can get murky, the author of the unpublished work goes first, then "ex", then the publishing author of the name (*Chrysocoma nauseosa* Pall. ex Pursh).

CHANGED NAMES (ICNAFP CHAPTER V §3)—We can also change the rank of an existing name or move it to a different genus. For subspecies and varieties, we could also change which species they belong to. The existing name in its originally published form is the basionym (*Chrysocoma nauseosa*). The changed name may be a new combination (the specific epithet from the basionym, with a different genus than that of the basionym, e.g., *Ericameria nauseosa*), or a new status (from the basionym *Chrysothamnus nauseosus* subsp. *graveolens*, we might create the name at new status *Chrysothamnus nauseosus* var. *graveolens*), or a new status & combination (from the basionym *Chrysothamnus oreophilus*, we might create *Ericameria nauseosa* var. *oreophila*).

When a new combination, new status, or new status & combination is published, the author of the basionym is given in parentheses and the author who changed the basionym follows (*Chrysocoma nauseosa* Pall. ex Pursh becomes *Ericameria nauseosa* (Pall. ex Pursh) G.L.Nesom & G.I.Baird). For subspecies and varieties, we may indicate only the authorship of the subspecies or variety (from the basionym *Chrysothamnus oreophilus* A.Nelson, *Ericameria nauseosa* var. *oreophila* (A.Nelson) G.L.Nesom & G.I.Baird) or the authorship of the variety and of the species to which it belongs (*Ericameria nauseosa* (Pall. ex Pursh) G.L.Nesom & G.I.Baird var. *oreophila* (A.Nelson) G.L.Nesom & G.I.Baird). Autonyms are a special case, since it has no independent authorship. The authorship of the species is given before the rank term, or authors are omitted: *Ericameria nauseosa* (Pall. ex Pursh) G.L.Nesom & G.I.Baird var. *nauseosa* or *Ericamerica nauseosa* var. *nauseosa.*

These changed names have the same type specimen as the basionym. In strict usage, the term "basionym" means only "the originally published name on which a new combination (or new status, &c.) is based". In the following I use "original name" to mean "any name in its originally published form". This provides a consistent term to use regardless of whether changed names have been created based on a particular original name. So, for example, *Chrysothamnus oreophila* A.Nelson is the basionym of *Ericameria nauseosa* (Pall. ex Pursh) G.L.Nesom & G.I.Baird, but *Cirsium funkiae* Ackerfield is not a basionym because no changed names are based on it. Both *Chrysothamnus oreophila* A.Nelson and *Cirsium funkiae* Ackerfield, however, are original names.

WHICH NAME?—A name is applied to a particular taxon when the type of that name is a member of that taxon. If there are multiple names that apply to a particular taxon, the oldest name at that taxon's rank has priority. The oldest name, or a new combination based on it, is the single correct name for that taxon.

Suppose we've found a potentially new species of *Ericameria*. What do we call it? First, we check for any original names whose types are members of our "new" species. If there are none, we need to publish a new name for it. If there is an existing original name whose type is a member of our new species, we check to see if it, or any name based on it, is at species rank. If yes, and the existing name is in the genus *Ericameria*, that is the correct name for our new species. If yes, but the existing name is not in the genus *Ericameria*, we need to make a new combination to move the name to *Ericameria*. If neither the basionym nor any changed name name based on it is at species rank, we get to choose: we can either publish a new name, or create a name at species rank based on the existing original name.

If there are multiple original names whose types are members of our "new" species, we'll need to find the set of all applicable names--the original names and all changed names derived from them. The oldest name at species rank in the pile has priority. As above, if that name is in *Ericameria*, it is the correct name for our new species. If not, we'll need to publish a new combination. If none of the names in the pile is at species rank, we get to choose between publishing a new name or creating a name at species rank based on one of the existing basionyms. We could choose any of the existing basionyms. Priority applies within a rank, not across ranks.

SYNONYMS—Given a particular taxonomy and its circumscription of a particular taxon, all names whose types belong to that taxon are synonymous. Only one of these synonyms will be the correct name for that taxon in that taxonomy. A different one of these synonyms might be the correct name for that taxon in a different taxonomy. Synonyms come in two flavors: nomenclatural synonyms and taxonomic synonyms. A set of nomenclatural synonyms includes a single original name and all changed names derived from it[10]. They are also called homotypic synonyms, because the changed names have the same type as the original name. As a result, the set of nomenclatural synonyms applies to whichever taxon that type belongs to. Nomenclatural synonyms are synonymous in all possible taxonomies. Taxonomic synonyms are, or are derived from, different original names. They are also called heterotypic synonyms, since the different original names have different types. These types may or may not belong to the same taxon. When the types belong to the same taxon, the names are taxonomic synonyms. This means they are synonymous within the context of a particular taxonomy. In a different taxonomy, they may or may not be synonyms.

Returning to the circumscription, we can add nomenclature as an aspect of a taxon's circumscription along with morphology, ecology, geography, genetics, &c. The nomenclatural circumscription of a taxon is the set of names included within it, i.e. the set of names considered synonymous. Ideally, the nomenclatural circumscription lists both the original names and all changed names included in a taxon. Including at least one member of each set of nomenclatural synonyms is adequate, as the others apply by definition to the same taxon.

A simple example in *Geranium* is given in Tables 1 & 2, showing two alternate taxonomies. The original names are in the left column. The set of sound names is in the central column, including both the original names and any changed names derived from them. The taxon names are in the right column. The relationship between the names in the left two columns and the right column shows the nomenclatural circumscription of a taxon. Since the authors are given in the table, I will omit them here. In Table 1, *Geranium caespitosum* includes the original names *Geranium caespitosum*, *Geranium fremontii*, and *Geranium fremontii* var. *parrryi*, and all names derived from them. *Geranium eremophilum* and *Geranium caespitosum* var. *eremophilum* are nomenclatural synonyms. A taxonomy that applied these two names to different taxa would be incorrect. That is not one of the allowed options under the ICNafp. Similarly, there is no choice about using the name *Geranium atropurpureum* for the species that includes the types of *Geranium atropurpureum* (published in 1898) and *Geranium eremophilum* (published in 1913). Using the younger name would be incorrect. *Geranium eremophilum* and *Geranium caespitosum* are heterotypic synonyms in the taxonomy of Table 2, but they are not synonymous in the taxonomy of Table 1. *Geranium caespitosum* var. *caespitosum* and *Geranium atropurpureum* var. *atropurpureum* are autonyms of *Geranium caespitosum* and *Geranium atropurpureum*. If we included geography along with nomenclature in these circumscriptions, we would see that under the taxonomy of Table 1, usage of the name *Geranium caespitosum* for plants in most of New Mexico would be misapplication of that name to plants correctly called *Geranium atropurpureum*. However, under the taxonomy of Table 2, the name *Geranium caespitosum* would be correctly applied to plants throughout New Mexico.

MISAPPLIED NAMES—Whenever a taxon is called by a name that is neither its correct name nor a synonym of it, that name is misapplied. While synonymy is a feature of relationships between names and relationships of names to a taxonomy, misapplication is more context-dependent. A name that is misapplied to one taxon in one context

[10] With the exception of autonyms; see the end of Section 3. There are other exceptions to the relatively simple view presented here, but apart from autonyms these are infrequent and omitted for the present discussion.

may be correctly applied to a different taxon in a different context, even in the same taxonomy. Misapplied names are similar to misidentifications, although when a name is said to be misapplied this implies a common practice within a community of naturalists—not an occasional error.

SPELLING OF NAMES (ICNAFP CHAPTER VIII)—The originally published spelling of a name is the correct spelling of that name, except for correction of errors as defined by the ICNafp. Latin has grammatical gender, and because scientific names are based on Latin they do as well. When new combinations are made, the gender of the specific epithet often needs to be changed to match the gender of the new genus. Spellings other than the correct spelling are orthographic variants. As with misapplied names, this usually implies that some group of people believed the orthographic variant to be the correct spelling of the name, rather than it being a single person's error. Unlike misapplied names, the taxon referred to by an orthographic variant is still unambiguous rather than context-dependent. The names *Phacelia caerulea* Greene and *Phacelia coerulea* Greene both refer to the same plant in all possible taxonomies, although only one of them is the correct name.[11]

A multiplication sign is sometimes used to mark names of hybrids (e.g., *Centaurea* × *moncktonii* C.E.Britton). The ICNafp (Article H.3) indicates that the multiplication sign is optional, not actually part of the name, and that for nomenclatural purposes the name is identical with and without the multiplication sign. "X" or "x" is often, incorrectly, substituted for "×". Multiplication signs are also used in hybrid formulae (e.g., *Boechera carrizozoensis* × *perennans*). In this case, the two components of the hybrid formula (*Boechera carrizozoensis* P.J.Alexander and *Boechera perennans* (S.Watson) W.A.Weber) are ICNafp names, but the hybrid formula is not.

UNSOUND NAMES—As mentioned above, I call names that violate ICNafp rules in some way "unsound". The code itself does not use the word "unsound" and does not provide a corresponding term, instead defining various different kinds of violations. An isonym is a name that is identical to and derived from the same basionym as an existing name (e.g., if published the new status & combination *Ericameria nauseosa* var. *oreophila* (A.Nelson) P.J.Alexander, this would be an isonym of *Ericameria nauseosa* var. *oreophila* (A.Nelson) G.L.Nesom & G.I.Baird). A superfluous name is a violation of the rule of priority, if someone publishes a new name while indicating that an older name at the same rank is a synonym—the older name should have been used, the new name is superfluous. A name is marked "pro synonymo" if the author to whom the name is attributed cited the name only as a synonym, rather than proposing it as the correct name of a taxon. A nomen nudum (naked name) is a name that does not have a description or diagnosis. A suppressed name (nomen rejiciendum) is a name that was published following the rules of ICNafp, but which the botanists at an International Botanical Congress decided should be rejected for other reasons. A junior homonym is a name that is identical to but not derived from the same basionym as an existing name (e.g., if I were to publish the new name *Ericameria nauseosa* var. *oreophila* P.J.Alexander). Unpublished names are often marked "ined.", for the Latin "ineditus".

Sometimes we can infer the correct taxon of an unsound name, sometimes not. Isonyms and superfluous names are similar to nomenclatural synonyms: though unsound, the taxon they apply to is unambiguous so long as the context (the application of the related sound name) is clear. Nomina nuda, suppressed names, junior homonyms, and unpublished names, though, may or may not have a clear reference to a taxon.

SUMMARY—To understand how plant names are related, what name is correct for a plant in a particular taxonomy, and how to translate names accurately between taxonomies, we need a data structure that:

1) Indicates the status of any particular name, including its rank, whether it is the correct spelling of the name or an orthographic variant, and whether or not it violates any rules of the ICNafp.

---

[11] *Phacelia coerulea* Greene is the correct name.

2) Allows us to track three kinds of relationships between names: i) the relationship between the set of original names and the set of all names; ii) the taxonomic relationship between a set of names and the correct name of a taxon under a particular taxonomy; iii) context-specific misapplication of names.

This information will let us recognize and avoid unsound names, understand what plant is being referred to in cases of misspelling or misapplication, and understand how different taxonomies are related to each other. For translating between taxonomies, it is especially important that we be able to distinguish cases in which a set of names will refer to the same taxon regardless of the taxonomy (nomenclatural synonyms, orthographic variants, &c.) and cases in which names are synonymous under one taxonomy but not another.

Table 1. Relationship between original names, names, and taxa in a subset of the genus *Geranium*.

original name	name	taxon
Geranium caespitosum E.James	*Geranium caespitosum* E.James	*Geranium caespitosum* E.James
	Geranium caespitosum var. *caespitosum*	
Geranium fremontii Torr. ex A.Gray	*Geranium caespitosum* var. *fremontii* (Torr. ex A.Gray) Dorn	
	Geranium fremontii Torr. ex A.Gray	
Geranium fremontii var. *parryi* Engelm.	*Geranium caespitosum* var. *parryi* (Engelm.) W.A.Weber	
	Geranium fremontii var. *parryi* Engelm.	
	Geranium parryi (Engelm.) A.Heller	
Geranium atropurpureum A.Heller	*Geranium atropurpureum* A.Heller	*Geranium atropurpureum* A.Heller
	Geranium atropurpureum var. *atropurpureum*	
	Geranium caespitosum subsp. *atropurpureum* (A.Heller) W.A.Weber	
Geranium eremophilum Wooton & Standl.	*Geranium caespitosum* var. *eremophilum* (Wooton & Standl.) W.C.Martin & C.R.Hutchins	
	Geranium eremophilum Wooton & Standl.	

Table 2. A different taxonomy for the same names as Table 1.

original name	name	taxon
Geranium caespitosum E.James	*Geranium caespitosum* E.James	*Geranium caespitosum* E.James
	Geranium caespitosum var. *caespitosum*	
Geranium fremontii Torr. ex A.Gray	*Geranium caespitosum* var. *fremontii* (Torr. ex A.Gray) Dorn	
	Geranium fremontii Torr. ex A.Gray	
Geranium fremontii var. *parryi* Engelm.	*Geranium caespitosum* var. *parryi* (Engelm.) W.A.Weber	
	Geranium fremontii var. *parryi* Engelm.	
	Geranium parryi (Engelm.) A.Heller	
Geranium atropurpureum A.Heller	*Geranium atropurpureum* A.Heller	
	Geranium atropurpureum var. *atropurpureum*	
	Geranium caespitosum subsp. *atropurpureum* (A.Heller) W.A.Weber	
Geranium eremophilum Wooton & Standl.	*Geranium caespitosum* var. *eremophilum* (Wooton & Standl.) W.C.Martin & C.R.Hutchins	
	Geranium eremophilum Wooton & Standl.	

Section 2. Proposed handling of nomenclatural & taxonomic data.

An expansion of the basic structure of Tables 1 and 2 provides taxonomic flexibility while maintaining nomenclatural clarity. "Original name" and "name" are nomenclatural, relatively rule-bound & inflexible, and should be nationally consistent. These columns provide the sound names that can be applied to plants, as well as unsound names that should not be used, without prescribing how these names are grouped into taxa. In order to serve this purpose, these columns would need to be relatively complete and continuously updated, so that absent names will not be needed. Additional nomenclatural fields would provide consistent plant codes, indicate ranks of names, mark unsound names & orthographic variants, and so on. Informal names that are pragmatically useful are included in the nomenclatural data, provided they are explicitly marked. The "taxon" column groups names into taxa. It is flexible and should be allowed to vary locally. Any "taxon" entry is a sound name or is defined as a set of sound names; appears in the "name" data; contains its own name; contains all or none of each set of nomenclatural synonyms. The minimal taxon definition is a single sound name that is the name of the taxon. The taxonomy data also includes attributes used in calculating AIM indicators or likely to be generally useful in analyses (habit & duration, native / exotic status, plant family, &c.). In the simplest implementation, there would be a national nomenclatural dataset and the state species lists would be converted appropriately to serve as taxonomies. The next step upwards in complexity would be to maintain national taxonomy data as well.

Two tables: Nomina and Taxa—The proposed data structure is presented as two tables, one for nomenclature (Table 3, "Nomina"), one for taxonomy (Table 5, "Taxa"). The fields of these tables are defined in tables 4 & 6. The "namCode" field serves as a key joining them.[12] In practice, it may generally be preferable to join Nomina & Taxa and work with them as a unit. Conceptually and in terms of data management, however, they are separate. Nomina should be a single, centrally-managed resource. It provides the set of names that can be used for many separate taxonomies. This common foundation can then be used to translate between taxonomies. Each taxonomy table has a particular geographic scope: the continental US for the national taxonomy; state political or BLM administrative boundaries for state taxonomies. Taxonomies can vary regionally, while the set of sound ICNafp plant names does not. The fields included in Taxa should also be expected to vary regionally, as additional attributes are useful to meet local needs. The fields in Table 6 are a floor, not a ceiling.

The Nomina table is in essence a two-level hierarchy. Original names are the more inclusive (parent) level. Names are the less inclusive (child) level. The attributes of names are given primarily at the level of names, with some of this information duplicated for original names to make the data more legible. The Taxa table is an

[12] To assign all nomenclatural synonyms to the same taxon when some are absent from a particular taxonomy (as will presumably be the norm), some additional work is needed after joining Nomina and Taxa by namCode. Let's call that joined table nominaTaxa. We should next add oriCode to Taxa based on matching namCode. Then we join Nomina and Taxa by oriCode and remove all records where the namCode is present in Taxa. Last, we can append this output to nominaTaxa. Using R and the 'tidyverse' package, this code could be used:

```
    taxa <- separate_rows(taxa,namCode,sep="\\|")
    nominaTaxa <- left_join(taxa,nomina,by="namCode")
    homotypSyns <- taxa %>%
        mutate(oriCode = nomina$oriCode[match(namCode, nomina$namCode)]) %>%
        filter(!oriCode == "") %>%
        select(-namCode) %>%
        distinct(oriCode,.keep_all = TRUE)
    homotypSyns <- left_join(homotypSyns,nomina,by="oriCode") %>%
        filter(!namCode %in% nominaTaxa$namCode)
nominaTaxa <- rbind(nominaTaxa,homotypSyns)
```

abbreviated four-level hierarchy. From most to least inclusive: family; genus; species; infraspecies[13]. This hierarchy is represented as two levels: family; taxon. Attributes are associated exclusively with taxa, not families. Provided the names are properly formatted and assigned the correct ranks, these two levels can be expanded to the full four levels in an automated fashion. The "missing" names are: for species, the parent genus is not explicitly represented; for infraspecies, the parent species and parent genus are not explicitly represented. Parent genera can be created by taking the first word of each name. Parent species can be created by taking the first two words of each name.

In this data structure, taxa and names are clearly distinguished. The full range of name statuses and relationships between names can be represented. Nomenclatural synonymy is represented by the relationship between the originalName + oriCode and the name + namCode. Taxonomic / heterotypic synonymy is represented by the relationship between the name + namCode and the taxon + taxCode. Misapplication is represented by the relationship between the name + namCode and the taxCode in the "misapplied" field. The namKind field distinguishes between ICNafp names (value starts with "1", "2", or "3") and non-ICNafp names (value starts with "0"). The namStat field marks orthographic variants and unsound names. The legitimate ICNafp name that should be used instead is not explicitly given, but can usually be inferred from context.[14]

I've attempted to reduce the set of fields to the minimum needed to capture relationships between names and attributes that will be useful in an AIM context. However, in some cases superfluous fields make the data more human-readable and usable. This leads, for instance, to duplicate author and rank fields. I assume most database managers would say that this data should be split into a greater number of related tables. I find related tables to be difficult to handle outside of dedicated database software that most users will find difficult to use without a custom UI. I think two tables that will often be united into one is probably the minimal level of structural complexity needed, and is much more portable and legible across software, platforms, and user skill levels.

Implementation—The basic implementation process would be:

1) Create the national nomenclature dataset. I have a preliminary version including: all names in the AIM & LMF SpeciesIndicators tables up to the 2020 field season; all non-native plants known in the continental US; all plants known in New Mexico, Montana, & Wyoming; all names in PLANTS that are easily matched in IPNI or Tropicos. The data for most of the nomenclatural fields of Table 3 are complete, except IPNI & Tropicos IDs.[15] The list has ±96,000 names and is probably complete enough for use. For comparison, PLANTS has ±86,000 names, ±70,000 of which are in my nomenclature data. A reasonably complete dataset for the continental U.S. would probably have 120,000–130,000 names. So long as the nomenclatural data includes all names in the AIM data, that level of completeness isn't critical or time-sensitive. The set of state species lists currently has ±2800 names not in my nomenclature data. Adding these would be the near term priority. In general, a set of names that is well-formatted, correctly spelled, includes relationships between original & changed names, and has correct authors in the IPNI standard forms can be added to the nomenclature data in a mostly-automated fashion. Most of the effort arises from missing basionyms, incorrect or inconsistently formatted authors, and orthographic variants.

2) Ensure unknown codes are clearly marked in AIM data. We need to be able to separate known plant codes from unknown plant codes so that we can treat them differently—known plant codes need to match the nomenclatural data & can be translated, unknown plant codes do not. I have a list of current unknown codes in the AIM data,

---

[13] We could recognize additional levels here, e.g. for multiple infraspecific ranks. However, it is better not done.

[14] Developing good processes to use in this context is on my to-do list. I've wished to avoid having an additional field for this purpose, but that may ultimately turn out to be the best course of action.

[15] Population of these fields can be mostly automated so long as the name and author data are in good shape. My focus so far has been getting to that point rather than attempting to add this information in the interim.

many anomalously formatted. If new unknown codes can all be expected to use the standard format, this list + a simple script to recognize the standard unknown code formatting should be adequate. Over the next field season or two we can find out if the formatting is now consistent enough, or if a better solution is needed.

3) Have a firm rule that no plant codes or names enter the AIM data except those in the national nomenclature dataset and unknown codes.[16] For all names + codes marked as informal names in the national nomenclature data, there must be a definition in either the national taxonomy data or the applicable state taxonomy.

4) Convert the current state species lists to the Taxa format proposed here. This should be relatively straightforward once all of the codes are present in the nomenclatural data. Whether or not we wish to systematically populate some of these fields (e.g., native / non-native status at the state level) for all taxa on a state list is an open question that could have an impact on the level of effort needed at this stage.

5) Flesh out the national taxonomy dataset to the point where the taxa and attributes cover at least the plants recorded in the AIM / LMF data. At present, ±40,000 of ±96,000 names in the nomenclatural data are assigned to taxa. Some of the attribute fields remain very sparsely populated. Most of the attribute data can be pulled from the state species lists and PLANTS, with resolution of discrepancies as needed. For the taxonomy, I prefer to avoid both the state species lists and PLANTS. Taxonomic data from the ITIS, Tropicos, POWO, and some other resources can be imported in a mostly-automated fashion. There is probably no single resource that it would be appropriate for us to follow completely. Completeness of the taxonomy across the entire nomenclature dataset is unlikely in the foreseeable future. Luckily, this completeness is neither critical nor time-sensitive.

6) Update the national nomenclatural dataset and national + state taxonomy datasets continuously. Additions to the national names list must meet one of the following criteria: a) accompanied by the name's IPNI or Tropicos ID; b) accompanied by a link to the work in which the name is published; c) submitted with a definition that is a set of plant codes or names that are in the national nomenclature dataset.[17]

MISCELLANEA—There are several formatting decisions for plant names & authors that are not dictated by the ICNafp or the IPNI standardized author list. These are minor issues where a consistent format is desirable, but which alternative is chosen is not very consequential. My suggestions are as follows:

1) Use of "×". This is optional in ICNafp names, difficult to type, and causes formatting headaches. I prefer to omit it from all ICNafp names. In hybrid formulae, however, it serves a useful purpose and can not be discarded.

2) Rank abbreviations. I prefer: subsp., var., subvar., fo., subfo. The first three are standard among most plant taxonomists. "Fo." and "subfo." are used by Tropicos but otherwise uncommon. I prefer them because "fo." is less easily confused with the "f." of some authors (e.g., Hook.f.). Once one has "fo.", "subfo." follows. I omit rare & unusual ranks (e.g., "proles"), marking these as unranked infraspecies.

3) Spaces in authors. With rare exceptions, IPNI omits spaces from authors (e.g., "M.Bieb."). Other resources, including Tropicos & PLANTS, use spaces before surnames ("M. Bieb."). I prefer to omit them. Text is easier to parse when each author is an unbroken group of letters & punctuation.

4) Diaeresis. According to ICNafp Article 60.7, diacritical marks are omitted from plant names except: "The diaeresis, indicating that a vowel is to be pronounced separately from the preceding vowel (as in *Cephaëlis*, *Isoëtes*),

[16] We may want to further require that only sound names / codes or informal names / codes enter the AIM data. This is less necessary, though, and may have disadvantages that have not occurred to me.

[17] An advantage of a dataset that is fairly narrowly focused on plant names is that it reduces the work involved in updates. I think this has been a difficulty for similar efforts. The more accessory information (maps, pictures, &c.) is tied to the plant name data, the more difficult it is to update the plant name data without potentially requiring a cascade of additional data updates. Allowing / expecting incompleteness in the taxonomy data should also make updates more efficient.

is a phonetic device that is not considered to alter the spelling; as such, its use is optional." I prefer to include diaereses as an aid to pronunciation. Also, ornamented letters are fun.

When a particular state wishes to strictly follow the taxonomy of PLANTS or some external resource in their species list, the species list is probably best formatted as a list of included taxa plus any needed attributes (habit, duration, *Centrocercus* functional groups, &c.). It would be easier to populate the other data fields in an automated fashion from a copy of the PLANTS taxonomy than to separately maintain this data in each state species list and then resolve errors that would inevitably creep in.

Table 3. Nomina. Proposed format for nomenclatural data, with a subset of *Geranium*. The "namCode" field serves as a key linking to the taxonomy data.

originalName	oriAuth	oriCode	oriKind	namRel	name	namAuth	namCode	isPLANTScode	namStat	namKind	unsoundType	IPNIid	TropicosID
				informal name	Geranium cf. eremophilum		GERcfERE	FALSE	informal name	0s			
				NA	Geranium		GERAN	TRUE	legitimate	1g		327764-2	40013761
Geranium atropurpureum	A.Heller	GEAT2	2s	new name	Geranium atropurpureum	A.Heller	GEAT2	TRUE	legitimate	2s		109072-2	13900935
Geranium atropurpureum	A.Heller	GEAT2	2s	autonym	Geranium atropurpureum var. atropurpureum		GEATA	TRUE	legitimate	3vn			50248135
Geranium atropurpureum	A.Heller	GEAT2	2s	new status & combination	Geranium caespitosum subsp. atropurpureum	(A.Heller) W.A.Weber	GECAA3	TRUE	legitimate	3s		109089-2	50087044
Geranium eremophilum	Wooton & Standl.	GEER	2s	new status & combination	Geranium caespitosum var. eremophilum	(Wooton & Standl.) W.C.Martin & C.R.Hutchins	GECAE	TRUE	legitimate	3v		881088-1	100364071
Geranium eremophilum	Wooton & Standl.	GEER	2s	new name	Geranium eremophilum	Wooton & Standl.	GEER	TRUE	legitimate	2s		109139-2	50109347
Geranium fremontii	Torr. ex A.Gray	GEFR2	2s	new name	Geranium fremontii	Torr. ex A.Gray	GEFR2	TRUE	legitimate	2s		277412-2	50111143
Geranium caespitosum	E.James	GECA3	2s	new name	Geranium caespitosum	E.James	GECA3	TRUE	legitimate	2s		277409-2	13900205
Geranium caespitosum	E.James	GECA3	2s	autonym	Geranium caespitosum var. caespitosum		GECAC3	TRUE	legitimate	3vn			100364074
Geranium fremontii	Torr. ex A.Gray	GEFR2	2s	new status & combination	Geranium caespitosum var. fremontii	(Torr. ex A.Gray) Dorn	GECAF	TRUE	legitimate	3v		277410-2	50187843
Geranium fremontii var. parryi	Engelm.	GEFRP	3v	new combination	Geranium caespitosum var. parryi	(Engelm.) W.A.Weber	GECAP2	TRUE	legitimate	3v		109090-2	100342361
Geranium fremontii var. parryi	Engelm.	GEFRP	3v	new name	Geranium fremontii var. parryi	Engelm.	GEFRP	TRUE	legitimate	3v		109150-2	50111144
Geranium fremontii var. parryi	Engelm.	GEFRP	3v	new status & combination	Geranium parryi	(Engelm.) A.Heller	GEPA2	TRUE	legitimate	2s		109241-2	13900158

Table 4. Definitions of fields used in Table 3.

	field name	field type	possible values, if factor	description
1	originalName	character		The original published form of a plant name, similar to the ICNafp definition of "basionym" except that a name is a basionym only in relation to a changed name based upon it, while here I include the original form of any name. This and the other original names fields are not populated for genera, sections or subgenera, informal names, hybrid formulae, and so on.
2	oriAuth	character		The author(s) of the original name, using the standardized IPNI forms of author names.
3	oriCode	character		The code of the originalName. USDA PLANTS codes are used when available. For more details, see the description for "namCode".
4	oriKind	factor	2s, 3s, 3v, 3sv, 3f, 3sf, 3u	This field records the rank of each originalName. 2s = species; 3s = subspecies; 3v = variety; 3sv = subvariety; 3f = form; 3sf = subform; 3u = infraspecies with rank not specified.
5	namRel	factor	NA, informal name, new name, new combination, new status & combination, new status, replacement name, superfluous	The relationship between the name and the originalName. NA = no original name, or original name not tracked here (genera, hybrid formulae, &c.); autonym = infraspecific name that repeats the specific epithet (nominate subspecies, nominate variety, etc.); informal name = not an ICNafp name; new name = the name and originalName are identical; new combination = the name is a new combination at the same rank as the originalName; new status & combination = the name is a new combination at a different rank than the originalName; new status = a change in rank that does not introduce any new epithet; replacement name = a name that is based on an unsound name and replaces that name as the basionym for subsequent changed names; superfluous = a name that was published in violation of priority, i.e. when a prior original name, or a changed name based on it, should have been used (namStat should also be "superfluous"). All names linked to the same originalName are nomenclatural synonyms.
6	name	character		A plant name. Most are published scientific names following (or intended to follow) the ICNafp. Informal designations are also included to account for groups of species that are frequently indistinguishable in the field, specification of a perennial member of a genus, or occasionally to include unpublished taxa or other anomalous cases. While the inclusion of these informal names in the same field as scientific names of plants is not ideal, the fields "namRel", "namStat", and "namKind" allow these informal names to be easily identified as such and pragmatically it is too convenient to have all possible names living in a single field to warrant splitting them into separate fields.
7	namAuth	character		The author(s) of the name, using the standardized IPNI forms. Authors are omitted for autonyms.

8	namCode	character		The code of the name. USDA PLANTS codes are used when available, when not available 6-7 letter codes are generated: first three letters of the genus, first three letters of the species, first letter of the infraspecies when applicable, followed by a number when multiple entries in the list have the same letter code. Codes for informal groups of species are generally formed by inserting "cf" between the first three letters of the genus and the first three letters of one of the included species. Codes for sections or subgenera are formed by inserting "s" between the first three letters of the genus and the first three letters of the section or subgenus name. When it is useful to designate annual or perennial species of a genus, an "af", "pf", &c., is appended to the genus code. Every row has a unique namCode.
9	isPLANTScode	logical		"TRUE" when the entry in namCode is a USDA PLANTS code, "FALSE" when it is not.
10	namStat	factor	, legitimate, unsound, ineditus, isonym, junior homonym, misattribution, nomen nudum, orthographic variant, pro synonymo, superfluous, suppressed, tautonym, informal name, misapplication, data anomaly	The status of the name under ICNafp. Legitimate = legitimate ICNafp name; unsound = not a legitimate ICNafp name, but not further categorized; isonym = identical in both name and type to a prior name; junior homonym = a name with the same spelling as a prior name, but having a different type or author than the previously published name; misattribution = name given incorrect authorship; nomen nudum = a name not validly published because it is missing a description / diagnosis; orthographic variant = an incorrect spelling of a name that has entered use alongside the correct spelling; pro synonymo = a name published only as a synonym, not proposed by the author as a new name; superfluous = a name that was published in violation of priority, i.e. when a prior original name, or a changed name based on it, should have been used; suppressed = a name that is rejected under ICNafp although otherwise legitimate; tautonym = a name in which the genus and specific epithet are identical; informal name = deliberately non-ICNafp names used as designations for informal groups of species & the like; misapplication = a "name" that results from a misrepresentation of misapplication, a relationship between names, as if it were itself a name; data anomaly = various other data anomalies, primarily to accommodate plant codes that have been in use but duplicate the nomenclatural meaning of other codes. Empty cells imply legitimate, ICNafp-compliant names. Isonyms, misattributions, orthographic variants, and superfluous names can be treated like nomenclatural synonyms. Ineditus names, nomina nuda, and junior homonyms can not.
11	namKind	factor	0g, 0s, 0u, 0x, 1g, 2s, 3s, 3sn, 3v, 3vn, 3sv, 3f, 3fn, 3sf, 3u, 3un	The rank or kind of each name. 0g = groups of species for convenience (e.g., genus + habit code; a group of similar species often difficult to distinguish in the field); 0s = section or subgenus; 0u = undescribed species-level taxon; 0x = hybrid formula; 1g = genus; 2s = species; 3s = subspecies; 3sn = nominate subspecies; 3v = variety; 3vn = nominate variety; 3sv = subvariety; 3f = form; 3fn = nominate form; 3sf = subform; 3u = infraspecies without rank; 3un = unranked infraspecific autonym.
12	unsoundType	character		Additional explanatory text for unsound names (marked in "namStat").
13	IPNIid	character		The International Plant Names Index name ID. IPNI does not include autonyms or mosses. Otherwise, ICNafp names should have IPNI IDs, although (like any resource) IPNI is not complete.
14	TropicosID	character		The Tropicos name ID. Tropicos maintains records for autonyms and mosses, so all ICNafp names should have Tropicos IDs.

Table 5. Taxa. Proposed format for taxonomy data, with a subset of *Geranium*. The "namCode" field serves as a key linking to the nomenclatural data. New Mexico is the geographic scope for this example.

namCode	taxon	taxCode	taxKind	taxDef	family	comments	common	exotic	misapplied	invasive	noxious	encroacher	GrowthHabit	GrowthHabitSub	Duration	SG_Group
GERcfERE	Geranium cf. eremophilum	GERcfERE	0s	GEER | GECAE	Geraniaceae		purple cluster geranium	FALSE					Non-Woody	Forb	Perennial	
GERAN	Geranium	GERAN	1g		Geraniaceae		geranium	FALSE					Non-Woody	Forb	Perennial	
GEAT2 | GEATA | GECAA3 | GECAE | GEER	Geranium atropurpureum	GEAT2	2s		Geraniaceae		western purple geranium	FALSE					Non-Woody	Forb	Perennial	
GEFR3 | GECA3 | GECAC3 | GECAF | GECAP2 | GEFRP | GEPA2	Geranium caespitosum	GECA3	2s		Geraniaceae		pineywoods geranium	FALSE	GEAT2				Non-Woody	Forb	Perennial	

Table 6. Definitions of fields used in Table 5.

	field name	field type	possible values	description
1	namCode	character		One or more namCodes, matching namCodes in Nomina. Formatted as a |-separated list.
2	taxon	character		The name of the taxon. This must match a name included in the taxon.
3	taxCode			The code of a taxon, following the same format as "namCode". The taxCode must be one of the namCodes included in the taxon.
4	taxKind	factor	0g, 0s, 0u, 0x, 1g, 2s, 3s, 3sn, 3v, 3vn, 3sv, 3f, 3fn, 3sf, 3u, 3un	The rank or kind of a taxon, following the same format as "namKind".
5	taxDef	character		A |-separated list of namCodes, defining an informal taxon. Definitions of informal names should generally remain consistent between taxonomies, although it may be appropriate to, e.g., define a code meaning "perennial *Astragalus*" to include all perennial *Astragalus* in the national taxonomy, but only those perennial *Astragalus* that occur in Wyoming for that state's taxonomy.
6	family	character		The family to which a taxon belongs.
7	comments	character		An area for taxon comments that do not fit elsewhere.
8	common	character		Common names for taxa, as used by USDA PLANTS or perhaps an alternate local name source.
9	exotic	factor	, TRUE, FALSE, ABSENT, UNKNOWN	Relative to the area covered by the taxonomy, TRUE = present, not native; FALSE = present, native; ABSENT = absent; UNKNOWN = unknown. Empty implies "FALSE".
10	misapplied	character		The taxCode of the taxon to which a name / namCode has been misapplied in the area covered by the taxonomy.
11	invasive	factor	, TRUE, FALSE	Relative to the area covered by the taxonomy, TRUE = invasive (not native and causing / likely to cause ecological or economic harm, or harm to human health); FALSE = not invasive. Empty cells imply "FALSE".
12	noxious	character		The noxious weed status, if any. For the national taxonomy, a |-separated list of national & lower-level noxious statuses, given as [state abbreviation]:[noxious weed category].
13	encroacher	character		For state taxonomies: TRUE = a native species that is the target of management actions to reduce its abundance. For the national taxonomy: a |-separated list of the states in which a given taxon is considered an encroacher.
14	GrowthHabit	factor	Woody, Non-Woody	Records whether or not a taxon is woody. Trees, shrubs, subshrubs, and most succulents are woody.
15	GrowthHabitSub	factor	Forb, Graminoid, NonVascular, Sedge, Shrub, SubShrub, Succulent, Tree	Records a taxon's growth habit category.
16	Duration	factor	Annual, Perennial	Records the longevity of plants belonging to a taxon. Biennials are recorded as "Annual".
15	SG_Group	factor	TallStaturePerennialGrass, PreferredForb, Sagebrush, ShortStaturePerennialGrass, NonSagebrushShrub	Records the sagegrouse-related functional group to which a plant belongs.

Section 3. Translating between taxonomies.

Fully developing processes for taxonomic translation will be an ongoing process, as there are many details to sort out. There are simple and straightforward options for translation that will introduce some errors, and more complicated options that allow lossless translation but require more contextual information and the ability to handle identifications as sets of names rather than only single names.

The simplest option is to look up the source names in "name" or "namCode" and then translate them to the matching values in "taxon" or "taxCode". This is often the only option if there is little contextual information, for instance if the source data simply has a list of names of unknown provenance and there is little information in the destination taxonomy regarding potential misapplication of names. However, whenever a taxon in the source taxonomy is split into multiple taxa in the destination taxonomy, this simple translation will introduce errors. An identification that was correct in reference to the source taxonomy may become incorrect in reference to the destination taxonomy. Using *Geranium* as an example, if the taxonomy shown in Tables 1, 3, and 5 is the destination taxonomy and the taxonomy shown in Table 2 is the source taxonomy, plants called *Geranium caespitosum* in a source data set using the source taxonomy could be either *Geranium caespitosum* or *Geranium atropurpureum*. A simple name to taxon lookup would call them all *Geranium caespitosum*, which would sometimes be incorrect relative to the destination taxonomy.

When there is data in the destination taxonomy about misapplication of names, we can do a two-step translation. First, look up whether the name used in the source data has an entry in the regionally appropriate "misapplied" field, and translate to that taxon if so; then proceed with a name to taxon lookup as before. This will reduce translation errors, but not eliminate them. In the *Geranium* example, almost all usage of the name *Geranium caespitosum* is misapplication correctable to *Geranium atropurpureum*. However, *Geranium caespitosum* does, rarely, occur in north-central New Mexico. Correcting all *Geranium atropurpureum* to *Geranium caespitosum* will introduce errors in a very small percentage of cases. This expected error rate is an important consideration in populating data in "misapplied" fields. When a single taxon in the source taxonomy maps to multiple taxa in the destination taxonomy, and these taxa all occur in the relevant region for the "misapplied" field, we will get the same translation errors as when doing a simple name to taxon lookup.

When we have reasonably complete data for both the source and destination taxonomies, we can trace the taxon used in the source data back to the set of original names included within that taxon in the source taxonomy. Given the name *Geranium caespitosum* and the source taxonomy in Table 2, we would get the following list: *Geranium caespitosum*, *Geranium fremontii*, *Geranium fremontii* var. *parryi*, *Geranium atropurpureum*, *Geranium eremophilum*. Then we look up what taxon or taxa matches that set of original names in the destination taxonomy: *Geranium atropurpureum*, *Geranium caespitosum*. So we translate "*Geranium caespitosum*" to "*Geranium atropurpureum* or *Geranium caespitosum*". Or we could use a two-step process as in the paragraph above, translating based on the "misapplied" field first and then proceeding with the source taxon → original names → destination taxon lookup. In this case, that would translate *Geranium caespitosum* to *Geranium atropurpureum*. When a single taxon in the source taxonomy maps to multiple taxa in the destination taxonomy and this is not resolved by the "misapplied" field, we will be stuck with translating a single source taxon to multiple destination taxa. This is inconvenient, but unfortunately not avoidable. There is no information that allows us to arrive at a single destination taxon in such a case, and representing this as "taxon A or taxon B" is the only way of representing the destination taxon without either losing information or introducing spurious (likely incorrect) information.[18] How best to handle these cases is unclear. The simplest approach is to create informal names in the destination taxonomy, e.g. adding a code "GERcfCAE" defined as "GEAT2 or GECA3".

However, since the nomenclature data is hard-coded, this would be a relatively inflexible and fragile approach—unless the code already exists or someone manually adds it, the translation wouldn't work.

In even the best case scenario, some level of translation error will be unavoidable. The two-step taxon → misapplied then source taxon → original names → destination approach described above, combined with good handling of cases with multiple destination taxa would ensure accurate translation from one nomenclatural circumscription to another. However, there are cases when two taxa may have the same nomenclatural circumscription but, for instance, different morphological circumscriptions. These cases are probably rare. In practice, there is probably not any feasible way to address them, so I think they simply create some level of inherent background error in the process.

TAXONOMIC COMPLETENESS—Our ability to accurately translate between two taxonomies increases as each taxonomy is more completely specified. "Completeness" in this context is the proportion of the original names in the nomenclatural data that are matched to taxa. So long as the relationship between nomenclatural synonyms is captured in the nomenclatural data, it does not matter which nomenclatural synonym is used to match an original name to a taxon. This is one of the main benefits of having a data structure that explicitly tracks nomenclatural synonymy.

In practice we should assume that taxonomies are rarely, if ever, going to be complete. This is a problem to the extent that names absent from a taxonomy are present in the data being translated. A good translation process will need to include handling of these names, e.g. by leaving them untranslated but with a marker of some kind indicating that they are untranslated.[19] Pragmatically, I think we should strive toward taxonomic completeness while keeping in mind that this is unlikely to be attainable and that an imperfect translation between taxonomies is better than none.

AUTONYMS—Autonyms can be problematic, as this is the only context in which names derived from the same original name refer to different taxa. Table 7 provides an example. *Lotus procumbens* could be considered a synonym of *Acmispon procumbens* or *Acmispon procumbens* var. *procumbens*. I think the first option is preferable as a default, but the second option is often used in existing taxonomy databases. Both options will cause errors using naïve translation processes. This can be seen by considering translations from the alternative taxonomies shown in tables 8, 9, and 10 to the taxonomy shown in Table 7. The taxonomies of Table 7 & Table 8 recognize the same taxa but in different genera. If we fill in the "?" with the autonym, a plant IDed as *Lotus procumbens* becomes *Acmispon procumbens* var. *procumbens*. This creates a more precise identification in the output than existed in the original data. However, when translating from Table 9 to Table 7, the ranks of the taxa change as well as the genus to which they are assigned. In this context, we should translate *Lotus procumbens* to *Acmispon procumbens* var. *procumbens*. If we fill in the "?" with the species, translating to *Acmispon procumbens* will produce a coarser ID in the output than in the original.

This can be solved by a translation rule: When the output taxonomy recognizes an autonymic infraspecific taxon and the source taxonomy does not, translate to the autonymic taxon unless the source taxon includes original names excluded by the autonymic taxon. When both source and output taxonomies recognize an

[18] As a result, if we have two alternative taxonomies, one recognizing a single taxon and the other recognizing multiple taxa, so long as the taxa in the second taxonomy can be identified by the field crew it is better to use this taxonomy.

[19] The Latin phrase "incertae sedis" is often used in this context, meaning "of uncertain placement".

autonymic infraspecific taxon, translate species to species and autonyms to autonyms.[20] Translating *Lotus procumbens* with Table 7 as the output taxonomy, the source taxonomy of Table 8 yields *Acmispon procumbens*, that of Table 9 yields *Acmispon procumbens* var. *procumbens*, and that of Table 10 yields *Acmispon procumbens*.

Tables 7 & 8. Two alternative taxonomies for a subset of the genus *Acmispon*.

<table>
<tr><th>original name</th><th>name</th><th>taxon</th></tr>
<tr><td rowspan="5">Hosackia procumbens Greene</td><td>Acmispon procumbens (Greene) Brouillet</td><td>Acmispon procumbens (Greene) Brouillet</td></tr>
<tr><td>Acmispon procumbens var. procumbens</td><td rowspan="2">Acmispon procumbens var. procumbens</td></tr>
<tr><td>Lotus procumbens var. procumbens</td></tr>
<tr><td>Lotus procumbens (Greene) Greene</td><td rowspan="2">?</td></tr>
<tr><td>Hosackia procumbens Greene</td></tr>
<tr><td rowspan="3">Lotus leucophyllus var. jepsonii Ottley</td><td>Lotus leucophyllus var. jepsonii Ottley</td><td rowspan="3">Acmispon procumbens var. jepsonii (Ottley) Brouillet</td></tr>
<tr><td>Acmispon procumbens var. jepsonii (Ottley) Brouillet</td></tr>
<tr><td>Lotus procumbens var. jepsonii (Ottley) Ottley</td></tr>
</table>

<table>
<tr><th>original name</th><th>name</th><th>taxon</th></tr>
<tr><td rowspan="5">Hosackia procumbens Greene</td><td>Lotus procumbens (Greene) Greene</td><td>Lotus procumbens (Greene) Greene</td></tr>
<tr><td>Acmispon procumbens var. procumbens</td><td rowspan="2">Lotus procumbens var. procumbens</td></tr>
<tr><td>Lotus procumbens var. procumbens</td></tr>
<tr><td>Acmispon procumbens (Greene) Brouillet</td><td rowspan="2">?</td></tr>
<tr><td>Hosackia procumbens Greene</td></tr>
<tr><td rowspan="3">Lotus leucophyllus var. jepsonii Ottley</td><td>Lotus leucophyllus var. jepsonii Ottley</td><td rowspan="3">Lotus procumbens var. jepsonii (Ottley) Ottley</td></tr>
<tr><td>Acmispon procumbens var. jepsonii (Ottley) Brouillet</td></tr>
<tr><td>Lotus procumbens var. jepsonii (Ottley) Ottley</td></tr>
</table>

[20] This sentence may seem to state the obvious, but it contradicts taxonomies that fill in the "?" with an autonymic infraspecific taxon. If we fill in the "?" with a species, the taxonomy is also telling us to do the obvious.

Tables 9 & 10. Two more alternate taxonomies for the names of tables 7 & 8.

original name	name	taxon
Hosackia procumbens Greene	*Lotus procumbens* (Greene) Greene	*Lotus procumbens* (Greene) Greene
	Acmispon procumbens var. *procumbens*	
	Lotus procumbens var. procumbens	
	Acmispon procumbens (Greene) Brouillet	?
	Hosackia procumbens Greene	
Lotus leucophyllus var. *jepsonii* Ottley	*Lotus leucophyllus* var. *jepsonii* Ottley	*Lotus jepsonii* (Ottley) ineditus
	Acmispon procumbens var. *jepsonii* (Ottley) Brouillet	
	Lotus procumbens var. *jepsonii* (Ottley) Ottley	

original name	name	taxon
Hosackia procumbens Greene	*Lotus procumbens* (Greene) Greene	*Lotus procumbens* (Greene) Greene
	Acmispon procumbens var. *procumbens*	
	Lotus procumbens var. procumbens	
	Acmispon procumbens (Greene) Brouillet	
	Hosackia procumbens Greene	
Lotus leucophyllus var. *jepsonii* Ottley	*Lotus leucophyllus* var. *jepsonii* Ottley	
	Acmispon procumbens var. *jepsonii* (Ottley) Brouillet	
	Lotus procumbens var. *jepsonii* (Ottley) Ottley	

SECTION 4. EXISTING / ALTERNATIVE STRUCTURES FOR NOMENCLATURAL DATA.

USDA PLANTS—The USDA's Plant List of Accepted Nomenclature, Taxonomy, & Symbols (PLANTS) database has two main advantages. It provides a consistent set of names, along with images & maps, for plants throughout the United States. It provides a consistent, national list of plant codes for efficient communication and data entry. However, the current, public-facing PLANTS data structure is not well-designed for botanical nomenclature. It does not distinguish clearly between names and taxa. It has a single kind of relationship, between "Accepted Symbol" and "Synonym Symbol". As a result, it is difficult to identify nomenclatural synonyms or distinguish between true synonymy and misapplication of names. PLANTS is also poor at recording & communicating the status of names. This information is provided by text accompanying the authors. Orthographic variants are designated by "orth. var." (e.g., SCLI16, *Scirpus littoralis* Schrad., orth. var.), misapplication of a name is designated using "sensu" or "non" (SEPL5, *Selaginella plagiochila* sensu Krug & Urb., non Baker; CAAT13, *Carex atrata* auct. non L. p.p[21]), unsound names are designated using varying text (SEBI4, *Senecio bicolor* auct. non (Willd.) Todaro, nom. illeg.), and so on. This approach is based on traditional academic publications on botanical nomenclature. It has the benefit of being human-readable for people with a background in the field. The meaning of these terms is probably not clear to most users, though, and data in this format is difficult to work with in a scripted or automated fashion.

The PLANTS database also has issues related to data quality and maintenance. It generally represents a taxonomic viewpoint of the 1980s and 1990s, with sporadic subsequent updates. Most names published in the last

[21] Compare with CAAT5, *Carex atrata* L. Using CAAT13 would imply a conscious decision to misapply the name. This is like trying to catch a data entry error by an option in a pull-down menu that says, "I'm entering the data incorrectly right now."

decade or two are absent. Many basionyms are absent. The basically static nature of the PLANTS database is sometimes seen as a benefit in terms of stability. The names and codes on PLANTS generally stay the same, leading to greater continuity in data over time and reducing the need for staff to learn new names. I think concerns related to stability are entirely reasonable. However, this is better addressed by having good systems for reliably and transparently translating between taxonomies than by trying to adhere to a static taxonomy that is increasingly out of date and incompatible with recent floras and other identification resources. Messy translation problems aren't avoidable. A more static and less locally-responsive set of taxonomic data does not reduce the need for messy translation, but places more of the translation process prior to data entry, as field personnel try to figure out which PLANTS codes correspond with names used in online resources and floras. We should expect, then, that stability in PLANTS hides rather than prevents inconsistency in the use of names. Less of the process is visible and within the reach of QA / QC.

More generally, relying on an external reference limits our ability to ensure that the data are appropriate for our usage. Unless a resource like PLANTS can accommodate all of our use cases, to some extent we would be forced to adopt multiple nomenclatural data sets rather than one. I think this is what most users of PLANTS have done—usually without good documentation of deviations from PLANTS.

State species lists—The current AIM state species lists do a good job of addressing some of the limitations of PLANTS, by allowing local flexibility and updates to nomenclature & taxonomy. Done well, a state species list can give field crews within a state a consistent and well-defined set of names to use that will be both up to date and more consistent with the identification resources that crews will be using. However, the data structure doesn't resolve the limitations of the PLANTS database and in some ways exacerbates them. Instead of one kind of relationship there are two: between "SpeciesCode" & "SynonymOf" and between "SpeciesCode" & "UpdatedSpeciesCode". Unfortunately, neither relationship is well-defined and may be used to mark taxonomic synonyms, misapplications, and so on. There are also a couple of additional, murkier relationships between names that have crept into the data: sometimes "SpeciesCode" is a PLANTS code but "ScientificName" is a name that does not match that code; sometimes multiple names are entered in "ScientificName". One benefit of the simpler PLANTS data structure is that it is clear which code and name is accepted in the PLANT S taxonomy, while in the AIM state species lists this is sometimes ambiguous.

The AIM state species lists also usually contain some unknown plant codes or anomalous plant codes which, while presumably useful within their intended context, are not defined so that their meaning or purpose can be understood. To some extent, I think this stems from the ambiguous purpose of the state species lists. Is a state species list intended to provide documentation of which plants occur in the state? To provide the list of allowed plant codes for field crews? To document how plant codes & names are used in a state's AIM data? To serve as an error-tracking form for past mistakes in plant names & codes? To assign habit & duration attributes to AIM data? Two or more of these purposes often conflict, but they are generally combined in state species lists.

To some extent, we can view the PLANTS database and state species lists as being similar conceptually but at opposite ends of a continuum from rigid to flexible. The PLANTS database is rigid in its approach to both taxonomy & nomenclature. The state species lists treat both taxonomy & nomenclature flexibly. It is better to place flexible taxonomy on top of rigid nomenclature. This is my core principle in developing an alternative.

Taxon concepts—An alternative discussed previously in relation to AIM species lists is the "taxon concept" approach advocated by Cam Webb at the University of Alaska Museum of the North (there is an introduction here: https://alaskaflora.org/pages/blog8.html). My understanding, which may be incorrect, is that taxon concepts are motivated by a desire to incorporate more aspects of a taxon's circumscription than just the nomenclatural circumscription. This would be particularly relevant for taxonomic translation in a scenario mentioned in the prior section—source & destination taxa with identical nomenclatural circumscriptions but different geographic ranges or morphological circumscriptions. While I understand the motivation, I do not think it is pragmatically feasible in a meaningful way. Returning to *Geranium*, as it happens I have location data (Figure 1) providing geographic circumscriptions of *Geranium atropurpureum* and *Geranium caespitosum* corresponding to the nomenclatural circumscriptions of Table 1. This is certainly a useful addition to the nomenclatural table. I don't think it could be applied at scale, across multiple states & multiple floras for each, without substantially greater resources than are available within either the BLM or the taxonomic research community. Morphology would probably be more difficult. Lacking these kinds of databases, it's not clear what we would do with knowledge that two authors have different morphological circumscriptions for a taxon. How do they differ? How does that relate to the identifications in our observational data?

Taxon concepts as currently envisioned rely, instead, on reference to published works. This is a mixed blessing. A reference to a published checklist or flora is easily achievable but not, on its own, very useful. Several taxonomic resources have implemented something like this limited taxon concept approach. The NatureServe entry for *Ericameria parryi* is typical:

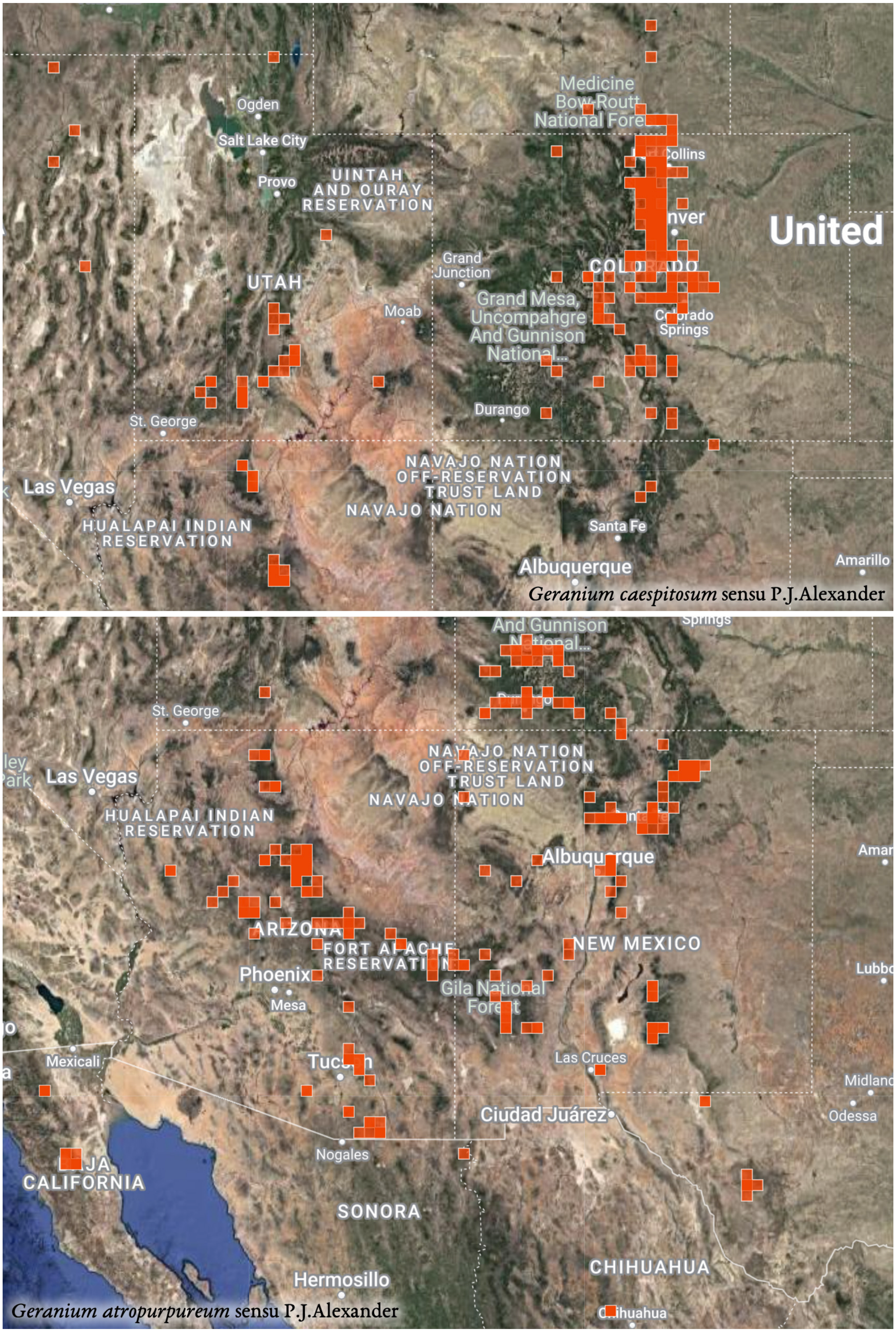
> "Concept Reference: Kartesz, J.T. 1994. A synonymized checklist of the vascular flora of the United States, Canada, and Greenland. 2nd edition. 2 vols. Timber Press, Portland, OR.
> Name Used in Concept Reference: *Chrysothamnus parryi*"

The 1994 BONAP checklist is the basis for PLANTS taxonomy of that time period, and consequently the two are very similar.[22] Rather than providing added value relative to a list of syonyms, this refers to an external list of synonyms. Providing the synonyms directly would be more useful. It's also not clear how we could meaningfully associate taxon concepts with field IDs from AIM crews. Do AIM crew members know if their concept of *Ericameria parryi* is that of Ackerfield or the Flora of North America? Do we, as data users, know if these concepts are different? If we have the synonymies of Ackerfield and the FNA, we could easily check if the nomenclatural circumscriptions differ, and this could be automated. Beyond that, I don't think an additional level of granularity or precision in recording identifications is realistically achievable at present.

Ultimately, if we want to know something about the morphology of a plant recorded in our field data, we need to go look at the plant.

[22] However, I have not seen this document and do not know if it is available digitally.

Figure 1. Geographic distributions of *Geranium caespitosum* and *Geranium atropurpureum*.



Comments on piñon / juniper ecology in relation to Grand Staircase–Escalante National Monument.

Patrick J. Alexander | Ecologist | BLM National Operations Center | 1 Aug 2022

Context—My personal field experience with plant ecology is primarily in New Mexico, with some experience in Arizona & Colorado but relatively little in Utah. My understanding of plant ecology in GSENM at this point is based on published research, remote sensing products, aerial imagery, and AIM plot data, not field experience in GSENM. This document should be interpreted accordingly, as a provisional understanding from someone who is reasonably well-equipped to understand the ecological situation *within the limits* imposed by working at a distance, but is absolutely not able to ground-truth his understanding of the landscape from personal experience. Neither you nor I should be at all surprised to find my understanding inadequate!

Margolis (2014) provides a description of the relationship between fire, grazing, and piñon / juniper ecosystems for Glorieta Mesa, in north-central New Mexico. In short, the plant community was historically a savanna with relatively sparse piñon and juniper and a well-developed herbaceous layer dominated by perennial grasses. The historic fire regime was characterized by frequent (±8 years, at Glorieta Mesa), low-intensity surface fires that caused mortality on young piñon & juniper, while rarely causing appreciable mortality on more mature trees. The introduction of grazing led to a reduction of fine surface fuels to below a level that could carry low-intensity surface fire. Fire became rare around 1900 and fire frequency was "0" from 1920 to 2000 in the Glorieta Mesa data. Without fire, piñon & juniper seedlings had low mortality. Tree density increased to ±6 times that of the pre-grazing plant community. In southern New Mexico, juniper establishment appears to be particularly associated with years of very high winter / spring precipitation and good monsoonal precipitation (my unpublished assessment of precipitation records & aerial imagery). Prior research suggests juniper seedlings are susceptible to drought and that competition with perennial grasses (in addition to their role in fueling low-intensity surface fires) may reduce their ability to establish (Johnsen 1962; Smith et al. 1974). I believe this basic picture holds for most piñon / juniper and juniper savanna ecosystems in New Mexico, where *Juniperus monosperma* is usually the dominant juniper, often with *Juniperus deppeana*, *Juniperus scopulorum*, *Juniperus arizonica*, or *Juniperus pinchotii*. It may or may not hold for the northwestern corner of the state where *Juniperus osteosperma* is found. At the higher and lower elevation margins of piñon / juniper and juniper savanna ecosystems in New Mexico the dynamics probably become somewhat different. In any case, in approaching GSENM, my initial expectation was that this basic picture is the norm in piñon / juniper ecosystems, and that GSENM is likely to be similar. That didn't turn out to be a good expectation.

Remote sensing data for GSENM—The remote sensing data from the Rangeland Analysis Platform uses data from monitoring plots, including AIM plots, to develop ground-truthed estimates of fractional foliar cover from Landsat imagery. RAP appears to do a very good job of capturing variation in tree cover in GSENM, based on comparison of the RAP tree cover products to what I can see on aerial imagery going to back to 1993. Where I see trees on aerial imagery, RAP also sees trees; where I see shrublands, bare ground, or vegetation treatments that removed trees on aerial imagery, RAP agrees that trees are sparse or absent. Using this kind of subjective assessment, I can't evaluate the relationship between the % tree cover reported by RAP and the true % tree cover, but I can verify that RAP is capturing the same relative densities of tree cover in different areas that I can see visually. However, I noticed some variation between years that seemed more likely to reflect differences in the quality of Landsat imagery, so decided to focus on a pair of rasters averaging the oldest (1986–1995) and newest (2012–2021) decades of available RAP data.

The RAP data shows relatively small increases (fractional tree foliar cover ±2–8% higher) from 1986–1995 to 2012–2021 across most of the piñon / juniper ecosystems of GSENM. So far as I can tell from aerial imagery, these result primarily from increases in the size of trees that were apparently mature or nearly mature in the earliest aerial imagery (±1993). These increases in tree cover are not clearly evident when looking at the aerial imagery. Some areas show larger increases (±10–20%) from 1986–1995 to 2012–2021, and these appear to be associated either with trees recolonizing older vegetation treatments (pre-1986; for many of these the dates are not known) or growth of trees that apparently colonized pockets of shrubland within the broader matrix of piñon / juniper woodlands at some point in the early to middle 20th century. These larger changes are visually apparent in aerial imagery. The RAP data also shows widespread but small (±1–4%) declines in shrub cover in the piñon / juniper ecosystems of GSENM.

Unlike the southern New Mexico sites I'm familiar with, where large and visually conspicuous pulses of juniper establishment occurred around 20–30 years ago in areas where junipers had previously been sparse, I did not see any evidence of big establishment events while looking at aerial imagery in GSENM. Small junipers are harder to see, though, forcing me to rely more on zooming in to small areas scattered around the landscape than on more systematically assessing larger parts of the landscape. So, my understanding of any recent juniper establishment events is probably not that good, and I would be especially unlikely to spot establishment events in the last decade or so. Given the resolution of Landsat imagery (30 m pixels), RAP & similar Landsat-derived products are not able to distinguish young from mature junipers.

AIM data for GSENM—The time horizon of the AIM and LMF data doesn't support inference of changes over time in GSENM. The primary role of these data in the context of this document, then, is calibration of the RAP Landsat estimates.

Published literature in GSENM—Harris et al. (2003) used a combination of field methods and remote sensing to compare the vegetation of Deer Spring Point (grazed) to No Man's Mesa (very little livestock grazing in the early 20th century, none known since; native ungulates apparently also having little impact on the vegetation). They found that total foliar cover was significantly higher and bare soil significantly lower at the grazed Deer Spring Point compared to the relatively ungrazed No Man's Mesa. Shrubs had significantly higher fractional cover[1] at Deer Spring Point, cryptobiotic crusts, grasses[2], and cacti significantly higher fractional cover at No Man's Mesa[3]. Tree cover was not significantly different between the two sites.

Guenther et al. (2004) made a similar comparison between these two sites using field methods without remote sensing data. Guenther et al. report their results as absolute foliar cover values, which are difficult to directly compare with the fractional cover values of Harris et al. However, the two studies reach similar conclusions. Contrary to Harris et al., Guenther et al. find that bare ground is significantly lower on No Man's Mesa than at Deer Spring point. Consistent with Harris et al., Guenther et al. find that on the relatively ungrazed No Man's Mesa cryptobiotic crust cover is significantly higher, total foliar cover and shrub cover significantly lower than on the grazed Deer Spring Point. Guenther et al. find no significant difference in tree

[1] This is a somewhat unusual way of reporting the data, as the fraction of total foliar cover rather than as foliar cover. Given that the total foliar cover also differs between the two sites, it is possible for the two sites to have significantly different fractional cover values even if the absolute foliar cover values do not differ.

[2] For grasses, the significant difference apparently holds for fractional cover but not absolute foliar cover.

[3] There's an error in Table 2 of Harris et al. that can cause some confusion here. I believe the first row should be read as "grasses", the second row—which is not labelled—as "forbs".

cover between the two sites. Grasses a minor component (±3% foliar cover) of both sites, with no significant difference between them. Annual plants are sparse (<2% cover) on both sites, but significantly higher at the grazed Deer Spring Point. Non-native plants were few at Deer Spring Point, absent from No Man's Mesa.

Barger et al. (2009) followed with research on tree demography and growth at these two sites. They found no significant differences in tree density, diameter, basal area, growth rate, or age of the oldest trees. Trees were significantly younger at the relatively ungrazed No Man's Mesa (average age 124 years) than at the grazed Deer Spring Point (average age 157 years). If grazing were causing an increase in tree establishment, we would expect the opposite, for trees to be younger at Deer Spring Point. Barger et al. (2009) highlighted the synchrony in establishment and growth rates between the two sites as evidence that climatic variation was the primary factor controlling these processes, not the difference in grazing history. The main limitation of these two studies is that they look at a single grazed / ungrazed comparison. The ecological dynamics may be different in other piñon / juniper ecosystems in GSENM.

Harris & Asner (2003) used a combination of field methods and remote sensing data to evaluate how vegetation changes with distance from livestock water sources. They distinguished photosynthetic vegetation, non-photosynthetic vegetation, and bare ground, but did not distinguish between plant functional groups (perennial grasses, trees, etc.). To address differences in livestock travel distance over different types of terrain, they categorized slopes into three categories (1–5°, 5–15°, >15°). They found significantly less vegetation (of both categories) and significantly more bare ground near livestock water sources. On the steepest terrain category, this effect extends ± 0.3 km from water sources, while on shallower slopes it extends 0.5–0.7 km from water.

Witt & Shaw (2010) summarized plot data from piñon / juniper woodlands in GSENM, collected from 1981 to 2005 as part of the U.S. Forest Service's Forest Inventory & Analysis program. They focus on increased mortality, especially of piñon but also of juniper, during drought from 2003 to 2005.

Published literature for other Colorado Plateau piñon / juniper ecosystems—Floyd et al. (2000, 2004, 2015) studied piñon / juniper ecosystems on Mesa Verde. They found that the fire-return interval in the natural fire regime was ±400 years and that low-intensity surface fires did not play an appreciable role in this ecosystem in the time period covered by their data (±1700 to 2004). The natural fire regime is apparently characterized by infrequent but high-intensity, stand-replacing fires. They also found that the outlines of fires that took place 200–300 years ago were still visible on aerial imagery, and that charred snags from fires of this age could still be found when the sites were visited in person. Where historical fires were more frequent on Mesa Verde, plant communities are dominated by resprouting shrubs (*Quercus gambelii, Amelanchier utahensis, Cerocarpus montanus*, &c.) rather than piñon / juniper woodlands. Recent changes in the piñon / juniper woodlands include a period of unusually large, high-intensity fires from 1996 to 2003, apparently the result of drought following two decades of unusually wet weather that led to increased plant biomass. Mortality of piñon (and juniper, to a lesser extent) due to the combined effects of hotter temperatures, drought, and engraver beetles (*Ips confusus*) has continued to 2015, having the greatest effect on mature piñons. Floyd et al. (2015) highlight the possibility that the effects of recent climatic shifts could lead to overall decline or loss of the current, old-growth or persistent, piñon / juniper woodlands at Mesa Verde.

Floyd et al. (2008) studied stand structure on Navajo Point, at the south end of Kaiparowits Plateau. They estimated the fire return interval at ≥600 years, and found that there had not been fundamental change in stand structure associated with human activity during the 20th century. Similar to work on Mesa Verde, they could identify the outlines of fires 100–300 years old on aerial imagery, and could find charred snags from fires of this age during field work. Kennard & Moore (2013) studied stand structure and fire history in persistent piñon / juniper woodlands of Colorado National Monument, west of Grand Junction. They found no evidence of large (>100 ha) stand-replacing fires and infer that such fires have been absent from this ecosystem for a perhaps a millennium. Drought and small (< 10 ha) fires, rather than large fires, have probably been the primary causes of tree mortality. Shinneman & Baker (2009a) studied several dozen sites around the margins of the Uncompahgre Plateau , south of Grand Junction and west of Montrose. They found high-intensity, stand-replacing fires occurring at >400–600 year intervals and no evidence of low-severity surface fires. They identified long-term patterns of tree mortality and drought driven by climate. Using sites within Colorado National Monument as a an ungrazed comparison, Shinneman & Baker (2009a) found significantly higher seedling / sapling densities at grazed sites for piñon but not for juniper. Both Floyd et al. (2008) and Kennard & Moore (2013) highlight the risk of cheatgrass (*Bromus tectorum*) reducing the resistance of persistent piñon / juniper woodlands to fire & resilience of woodlands following fire. Shinneman & Baker (2009b) also discuss the role of cheatgrass in this context. They did not find a significant difference in post-fire cheatgrass cover between seeded and unseeded sites. However, regression models found that higher native plant species richness and higher cover of biological soil crusts were associated with lower post-fire cheatgrass cover. Shinneman & Baker (2009b) suggest that grazing is likely to play a role in altering fire ecology due to the effect of grazing on ecosystem attributes related to post-fire resilience.

Vankat (2017) studied change from 1935 to 2011 in piñon / juniper woodlands on the south rim of the Grand Canyon, in Grand Canyon National Park. They divided piñon / juniper woodlands into three categories: persistent piñon / juniper woodland, seral piñon / juniper woodland (affected by stand-replacing fires prior to 1935), and transitional piñon / juniper woodland (at the higher elevations of piñon / juniper ecosystems, grading into ponderosa woodlands). Within persistent piñon / juniper woodland, they found that drought-induced mortality on more mature piñon and juniper was followed by establishment events, primarily of juniper. Density and basal area of tree increased substantially in seral piñon / juniper woodland, apparently representing post-fire recovery towards the tree cover typical of persistent piñon / juniper woodland. Transitional piñon / juniper woodland was stable. Vankat attributes change primarily to climatic variation over time. He echoes Floyd et al. (2000, 2004, 2008), Shinneman & Baker (2009), and Romme et al. (2009) in describing the natural fire regime as having infrequent (recurring on time scales of multiple centuries) but high-intensity, stand-replacing fires that play a smaller role within the context of climate-driven mortality and establishment cycles.

Romme et al. (2009) provided a synopsis of research on piñon / juniper ecosystems and fire ecology. They divide piñon / juniper ecosystems into three general categories: persistent piñon / juniper woodlands; piñon / juniper savannas; wooded shrublands. The authors suggest that, although these broad categories are useful, there is substantial variation within each category as well. They identify persistent piñon / juniper woodlands as especially prevalent on the Colorado Plateau[4], and as associated especially with shallower or rockier soils and climates with relatively bimodal (rather than primarily winter / spring or primarily monsoonal) precipitation. In a prior paper, Romme et al. (2007) provide a key allowing sites to be assigned to one of these three categories. They characterize persistent piñon / juniper woodlands as being primarily areas where total tree foliar cover is >10% and at least 10% of the trees are over 150 years old, or with lower tree foliar cover at particularly rocky, unproductive sites or following severe disturbance (fire, woodcutting, etc.). In both papers, Romme et al. characterize persistent piñon / juniper woodlands as burning infrequently, with fires typically small and stand-replacing; having relatively low and discontinuous surface fine fuel loads under most natural conditions; being stable over centuries in the absence of fire and not increasing in tree cover due to recent changes in fire regime; having patterns of tree mortality and establishment driven primarily by climate

[4] So far, the papers I've found on piñon / juniper ecosystems of the Colorado Plateau either call them persistent woodlands or present data showing that they match the Romme et al. definition of persistent woodlands.

rather than fire; having recent periods of more numerous or larger fires driven primarily by climate, secondarily by invasion of cheatgrass. They also state that there have been 20th century increases in tree cover in many or most persistent piñon / juniper woodlands (Romme et al. 2007) but with little landscape-scale effect on total tree cover (Romme et al. 2009). A potential role of grazing in changing tree cover, tree demography, or fire regime in persistent piñon / juniper woodlands is discussed in both papers, but they state (in Romme et al. 2009) that the empirical evidence is "sparse and mixed".

Conclusions—My current understanding of the fire ecology and tree dynamics in piñon / juniper ecosystems of GSENM is that: These are primarily persistent piñon / juniper woodlands and are probably not significantly affected by grazing or fire suppression with regard to fire frequency or severity, or the abundance or age distribution of trees. The natural fire regime is probably characterized by relatively small (100–1000 ha?), infrequent (fire return interval of 300–600 years?), but high intensity, stand-replacing fires. Natural climatic variation on decadal time scales and, in recent decades, unnatural increases in temperature and changes in the timing or amount of precipitation are probably the major drivers of change in tree abundance / age composition. Climate change is probably the primary anthropogenic force affecting piñon / juniper dynamics in GSENM, and is not directly amenable to management at the scale of GSENM.

This basic picture seems to be inconsistent with the fire regime / vegetation condition classes in LANDFIRE products. I do not have a good enough understanding of how this LANDFIRE product is developed to have a good guess why this might be. I view remote-sensing products like this primarily as a way of extending our understanding over much larger parts of the landscape than we can directly sample. As a result, I approach them by first trying to if they are consistent with an understanding based other data sources. When this is not the case, I consider other data sources—especially direct field measurements—to have priority and the remote-sensing product to be in doubt[5]. Or, at least, while it may tell me something useful, I find myself uncertain what that something might be. I've found several times that a remote-sensing product is accurately reporting something about the landscape, but not what I expected it to report.[6]

There may well be multiple piñon / juniper types at GSENM, which differ in one or more of the characteristics described above. Other piñon / juniper types are more likely to be found on deeper and less rocky soils relative to the average of piñon / juniper ecosystems at GSENM.

A focus on a single site, No Man's Mesa, in studies of ungrazed ecosystems in GSENM is a major constraint on our ability to interpret the available research as characterizing piñon / juniper ecosystems in the monument as a whole.

The largest published effects of grazing on the ecology of piñon / juniper ecosystems in GSENM are decreases in cryptobiotic crust cover & increases in shrub cover in grazed areas, and decreases in total foliar cover near water sources / in areas more readily accessible to livestock. The few papers I've run across so far that directly address native & introduced plant diversity and annual plant cover suggest that grazing causes declines

---

[5] A related observation on this point: I often use southern New Mexico as my test case, since this is where I can most easily & accurately check a remote-sensing product against what's happening on the landscape. There are reasons to suspect that southern New Mexico is not a very good test case for this particular LANDFIRE product. Nonetheless, I can't really match up the patterns of variation depicted in the LANDFIRE fire regime / vegetation condition class product with patterns of variation in plant ecology on the landscape.

[6] In one case, I found that a product intended to depict areas more susceptible to juniper encroachment instead showed me where junipers were historically—presumably, naturally—more abundant. Certainly useful, but with management implications close to the opposite of what was intended.

in native plant diversity, higher introduced plant diversity, higher annual cover, and greater susceptibility to invasion by cheatgrass, especially after fire. Given the severity of cheatgrass invasion throughout much of the western U.S., a better understanding of this topic would be desirable. The papers I've come across so far are more suggestive than definitive. However, I haven't made any effort to review the literature on cheatgrass except insofar as it overlaps with studies of tree dynamics & fire ecology. I don't know how well other research fills this gap.

Cited works—

N.N. Barger, H.D. Adams, C. Woodhouse, J.C. Neff, & G.P. Asner, 2009. Influence of livestock grazing and climate on pinyon pine (*Pinus edulis*) dynamics. *Rangeland Ecology and Management* 62: 531–539.

M.L. Floyd, D.D. Hanna, & W.H. Romme, 2000. Fire history and vegetation pattern in Mesa Verde National Park, Colorado, USA. *Ecological Applications* 10: 1666–1680.

M.L. Floyd, D.D. Hanna, & W.H. Romme, 2004. Historical and recent fire regimes in piñon–juniper woodlands on Mesa Verde, USA. *Forest Ecology and Management* 198: 269–289.

M.L. Floyd, D.D. Hanna, J. Spence, W.H. Romme, M. Winterowd, & D.P. Hanna, 2008. Fire history of piñon–juniper woodlands on Navajo Point, Glen Canyon National Recreation Area. *Natural Areas Journal* 28: 26–36.

M.L. Floyd, W.H. Romme, M.E. Rocca, D.P. Hanna, & D.D. Hanna, 2015. Structural and regenerative changes in old-growth piñon-juniper woodlands following drought-induced mortality. *Forest Ecology and Management* 341: 18–29.

D. Guenther, T.J. Stohlgren, & P. Evangelista, 2004. A comparison of a near-relict site and a grazed site in a pinyon-juniper community in the Grand Staircase–Escalante National Monument, Utah. *Pp. 153–162 in:* C. Van Riper & K.L. Cole [EDS.], *The Colorado Plateau: cultural, biological and physical research*. Tucson, AZ, USA: University of Arizona Press.

A.T. Harris & G.P. Asner, 2003. Grazing gradient detection with airborne imaging spectroscopy on a semi-arid rangeland. *Journal of Arid Environments* 55: 391–404.

A.T. Harris, G.P. Asner, & M.E. Miller, 2003. Changes in vegetation structure after long-term grazing in pinyon-juniper ecosystems: integrating imaging spectroscopy and field studies. *Ecosystems* 6: 368–383.

T.N. Johnsen, 1962. One-seed juniper invasion of northern Arizona grasslands. *Ecological Monographs* 32: 187–207

D. Kennard & A. Moore, 2013. Fire history, spatial structure, and mortality in a piñon–juniper woodland in the Colorado National Monument. *Natural Areas Journal* 33: 296–306.

E.Q. Margolis, 2014. Fire regime shift linked to increased forest density in a piñon–juniper savanna landscape. *International Journal of Wildland Fire* 23: 234–245.

W.H. Romme, C.D. Allen, J.D. Bailey, W.L. Baker, B.T. Bestelmeyer, P.M. Brown, K. Eisenhart, M.L. Floyd, D.W. Huffman, B.F. Jacobs, R.F. Miller, E.H. Muldavin, T.W. Swetnam, R.J. Tausch, & P.J. Weisberg, 2009. Historical and modern disturbance regimes, stand structures, and landscape dynamics in piñon–juniper vegetation of the Western United States. *Rangeland Ecology and Management* 62: 203–222.

W.H. Romme, C.D. Allen, J.D. Bailey, W.L. Baker, B.T. Bestelmeyer, P.M. Brown, K. Eisenhart, M.L. Floyd-Hanna, D.W. Huffman, B.F. Jacobs, R.F. Miller, E.H. Muldavin, T.W. Swetnam, R.J. Tausch, & P.

Weisberg, 2007. Historical and modern disturbance regimes of piñon–juniper vegetation in the western US. Nature Conservancy, Fire Learning Network, Tallahassee, FL, USA.

D.J. Shinneman & W.L. Baker, 2009a. Historical fire and multidecadal drought as context for piñon–juniper woodland restoration in western Colorado. *Ecological Applications* 19: 1231–1245.

D.J. Shinneman & W.L. Baker, 2009b. Environmental and climatic variables as potential drivers of post-fire cover of cheatgrass (*Bromus tectorum*) in seeded and unseeded semiarid ecosystems. *International Journal of Wildland Fire* 18: 191–202.

J.L. Vankat, 2017. Post-1935 changes in pinyon-juniper persistent woodland on the South Rim of Grand Canyon National Park, Arizona, USA. *Forest Ecology and Management* 394: 73–85.

C. Witt & J.D. Shaw, 2010. Characteristics of pinyon-juniper woodlands in Grand Staircase–Escalante National Monument: changes since monument establishment and prospects for future monitoring. *Pp. 70–78 in: Proceedings: Learning from the Land, Grand Staircase–Escalante National Monument Science Symposium, 2006*. U.S. Department of the Interior, Bureau of Land Management, Escalante, UT.

Using AIM Data for Land Use Planning

[DRAFT Land Use Planning Handbook]

I. Principles for data analysis.

1. Analyses should be as simple as possible, but not simpler. At minimum, any analysis should have a straightforward descriptive sentence identifying what data were used, how the data were analyzed, and why the analysis matters. When more complicated descriptions of analytical methods are appropriate, provide this simple description first, then proceed to additional details.

2. All models are wrong, some models are useful. Ecology is complicated, our data are always incomplete, and our analyses require simplifying assumptions. Our understanding of the world will always be inaccurate in various ways. We should approach data analysis by trying to understand and describe these limits and inaccuracies. Then we can evaluate whether or not the analysis is useful.

3. When we don't know, write "We don't know." We often do not have the appropriate data, or perhaps the expertise, time, or analytical tools, to provide a useful answer to an important resource question. We should assume that this is common knowledge among both BLM staff and the public. Our role in NEPA, then, is to identify the gaps in our knowledge and state them clearly. For resource questions that are particularly important or occur frequently, "We don't know" should be followed by "and here's how we're going to find out."

4. Data is not decoration. When we present an analysis in a NEPA document, the reader should not wonder why it is there. The analysis might measure the likely effects of alternatives, explain how a proposed action or mitigation measure was developed, provide the context a reader will need to understand a subsequent effects analysis, and so on. When the data is intended to play a decisive role, we should demonstrate to the reader that different analytical results lead to different management decisions.

5. If we can't fail, we can't succeed. Whenever we will use data to measure the outcome of a resource use or management action against our objectives, there is a temptation to provide some wiggle room so that we don't force our future selves to admit failure. This is normal. However, when an objective isn't clear enough for us—and the public—to tell when we have failed, is isn't clear enough for us to tell when we have succeeded. We also learn from failure. If we want to make better decisions in the future, failure is part of the process.

6. Match the analysis to the question. In land use plans, the analyses we are likely to use can be divided into four categories: condition; effects; sustainability; trend. Each is appropriate in some contexts and inappropriate in other contexts. The next section describes these categories as well as the data and context appropriate for each.

II. A taxonomy of analyses.

- Condition analyses report on measured indicator values, or other ecosystem attributes calculated from measured indicator values, either directly or in relation to benchmark values or reference conditions.

- Trend analyses report change in indicator values over time. As with condition analyses, these may be in relation to benchmarks or reference conditions.

- Effects analyses measure the change in ecosystem attributes caused by an action, whether a resource use, management action, form of mitigation, &c.

- Sustainability analyses measure the effect of current resource use on future resource availability. This may be limited to a single resource use / resource, or may evaluate the effects of one resource use on a different resource. Sustainability analysis is a specialized form of effects analysis.

Further detail on each analysis type is provided below. Condition and effect analyses are divided into several subtypes.

Condition analysis.

- Descriptive condition analysis. In the simplest case, ecosystem attributes are presented as-is, without interpretation that relates the values to management objectives or resource uses. This is generally appropriate when an analysis is used to provide context. A good descriptive condition analysis requires data that are representative of the analysis area: plots that are randomly sampled from throughout the analysis area, or plots from a stratified random sample are used in a weighted analysis to correct for different sampling probabilities on different parts of the landscape. When the ecosystem attributes we're interested in vary seasonally, we may also need to check whether or not the sampling times are representative. Limitations in time or analytical staff may make it appropriate to use an unweighted plot-count analysis with stratified random sample designs, although this reduces our ability to infer the condition of the analysis area.

    - Examples: % cover of sagebrush; average soil stability; number of species per plot; mapping of plant community types.

- Benchmark analysis. We can use a defined threshold, or benchmark, to separate data into two categories, typically representing desired vs. unwanted outcomes. Sometimes we may use three or more categories. Then we can report the number of plots or proportion of the analysis area in each category. This is generally appropriate whenever management decisions or downstream analytical results will be based on categorical management objectives. These management objectives may be established by external regulations, agency policy, or prior NEPA. In developing a land use plan we will often find ourselves both defining benchmarks and evaluating current conditions relative to them in the same document. When we are defining benchmarks we should provide a clear rationale—perhaps with its own supporting analysis—explaining how we arrived at the benchmark value. As for a descriptive condition analysis, representativeness is the primary criterion for evaluating whether data is suitable to use in a benchmark analysis.

    - Examples: meeting / not meeting land health standards; suitable / marginal / unsuitable Habitat Assessment Framenwork (HAF) categories; evaluating plots against a canopy gap size threshold for wind erosion.

- Reference condition analysis. Reference conditions are those we believe would exist on the landscape in the absence of any adverse impacts from resource uses or management actions. A measure of similarity to / departure from reference conditions can provide a useful interpretive framework for our data. Other measures, for instance habitat suitability for a focal species, can also place data on a numerical scale of more suitable / less suitable or more desirable / less desirable conditions. In some cases, a single scale may be applied to an entire analysis area. In other cases, we divide the landscape into parts with different reference conditions. Ecological sites are often used for this purpose. When we partition the landscape, we need to consider whether or not our data are representative relative to each ecological site. Reference condition analyses are also sensitive to how the data are associated with ecological sites. The same data may indicate either reference or highly departed conditions, depending on which ecological site it is evaluated against.

- Examples: range condition scores and related similarity indices; categorization of plots into ecological states that correspond to different levels of departure from reference.

Trend analysis. Trend analyses are similar to condition analyses, except reported over time rather than for a single time period, and can be broken into the same subcategories: descriptive trend; benchmark trend; reference trend. Trend analyses are most informative when change over time is inferred from repeated measurements at the same monitoring plots. Trend can also be inferred from multiple representative samples using different monitoring plots, but in this case a portion of the change between time points will be attributable to the change in sampling. Interpretation of trend should take condition into account. Increasing, decreasing, or stable trends have different meanings depending on where the starting and ending values are relative to the potential variation in indicator values. Also, decreasing trends may not be possible if initial values are near the low end, and increasing trends may not be possible if initial values are near the high end, of the potential variation.

- Examples: change in % cover of sagebrush over time; change in % of plots meeting land health standards over time; change in range condition scores over time.

Effects analysis.

- Categorical effects analysis. When an action is either present or absent, an effects analysis measures the difference between areas where the action has occurred and areas unaffected by it. The ideal sample design uses randomized treatment / control pairs of monitoring plots. This standard is often difficult to meet in practice and requires sampling designed specifically for this purpose. When we conduct an effects analysis using existing data from other sample designs, we need to address what other differences between the affected and unaffected plots, apart from the action being evaluated, are likely to affect our understanding of the action's effects.
    - Examples: vegetation treatment effectiveness; grazing effects measured using grazing exclosures; measuring fire effects at burned vs. unburned sites.
- Continuous effects analysis. When an action occurs on an intensity gradient rather than being present or absent, an effects analysis may measure the correlation between that intensity gradient and ecosystem attributes. When evaluating the suitability of data for this kind of analysis, we need to consider whether or not the entire range of the intensity gradient is represented and what other factors may vary along with intensity in our sampling. A sample designed specifically for this purpose may be needed, in which case systematic sampling (e.g., monitoring plots at fixed intervals along the intensity gradient) may be an alternative to random sample designs.

- Examples: indirect effects, e.g. spread of invasive plants, radiating outward from disturbance; effects of different grazing intensities across an allotment; effects of varying levels of recreational use; effects of different levels of fuel removal.

Sustainability analysis.

This is the most complicated of the analysis types. Sustainability analyses can be broken into two parts. First, we need to understand the relationship between current resource use and future resource availability. We can visualize this as a response curve, with resource use on the *x* axis and future resource availability on the *y* axis (Figure 1). Depending on context, resource use might be expressed as an absolute amount of use or as a percent use of the currently available resources. Three general patterns of response are shown. A: resource use has no effect on future resource availability. B: resource use has no effect on future resource availability below a threshold value (dashed line), and above that threshold increased resource use causes declines in future resource availability. C: resource use increases future resource availability up to a threshold value (dashed line), and above that threshold increased resource use causes declines in future resource availability. "A" indicates that all levels of resource use are sustainable. "B" indicates that values below the threshold are sustainable. "C" indicates that sustainability is highest at an intermediate level of resource use. Research to create a response curve for a particular resource use in a given ecosystem is generally difficult and time-consuming. Where published data establishing response curves is not available, a sustainability analysis may not be possible and collaboration with external research partners might be needed to resolve the issue. However, a given response curve may have broad applicability in an ecoregion and support many sustainability analyses.

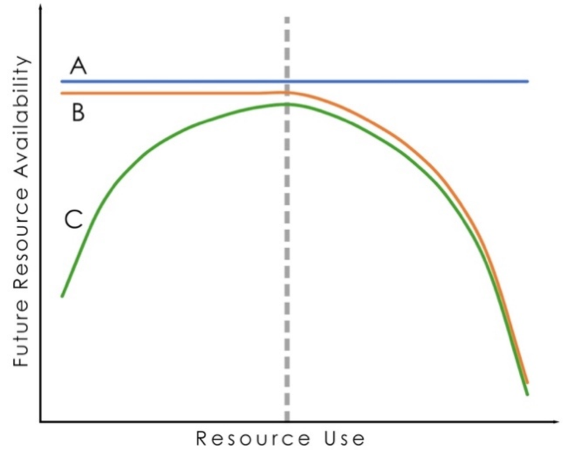


Figure 1. Three generalized response curves for relationships between resource use and future resource availability.

Second, once we understand the relationship between resource use and future resource availability, we can use data about levels of use and levels of current resource availability to understand where we are on the response curve. Current levels of use give us insight into the sustainability of current use. Because current resource availability depends on past resource use, this can give us insight into the sustainability of past resource use. For instance, suppose we are working with a response curve of type "B" and both current use and current resource availability are low (relative to the threshold and relative to the maximum resource availability, respectively). This would suggest that current resource use is sustainable, but resource availability is reduced by unsustainable use in the past. If current use and current resource availability are both high, on the other hand, that would suggest past resource use has been sustainable but current use is unsustainable and likely to cause a decline in future resource availability.

- Examples: grazing sustainability for a particular species of grass or a particular plant community; forest product sustainability in a particular woodland type.

III. Role of AIM data in different stages of the land use planning process.

Preparation Plan.

- Use the AIM data portal to summarize the AIM and LMF data available in the planning area. Also consult with the local and state AIM project leads to understand what data is being collected and is likely to become available in the immediate future.

- In addition to direct use of AIM data, consider remote-sensing products, many of which use AIM data:

    - LANDFIRE

    - Landscape Cover Analysis and Reporting Tools (LandCART)

    - Rangeland Analysis Platform (RAP)

- Keep AIM data in mind when developing the preliminary planning criteria. The planning criteria describe what kinds of analysis needs and methods are expected in the LUP process. Can we meet the analysis needs using existing AIM data? Are there data gaps that the AIM program could fill within the expected timeline of LUP development?

Analysis of the Management Situation (AMS).

- Providing context for the LUP is the primary purpose of the AMS.

- Condition analyses are generally a good fit. Descriptive analyses are sufficient in some cases. Benchmark analyses are appropriate when benchmarks are established above the LUP level but may be predecisional in contexts where the LUP will play a role in developing benchmarks. Trend analyses are also a good fit for the AMS, depending on the time scale of the available data and the appropriate time scale for relevant ecosystem processes.

- Existing effect or sustainability analyses should be included when they provide useful context (e.g., describing the effects of a resource use in the ecoregion), but developing new effect or sustainability analyses is generally not a priority at the AMS stage.

- The planning criteria of the Preparation Plan are more fully developed in the AMS:

    - What kinds of questions will analyses need to answer?

    - What kinds of analyses will we need?

        - Land health standards apply in most LUPs. Benchmark analyses are usually appropriate for land health standards.

        - The use of HAF in LUPs that address sage-grouse also implies a benchmark analysis, following established HAF protocols.

        - Benchmark analyses are likely to apply to other habitat suitability questions, as well.

- Direction for managing National Conservation Lands usually includes avoiding harm to certain resources. This implies effects analyses, evaluating the effects of resource uses or planned management actions on the protected resources.

- Endangered Species Act compliance is based primarily on effects analyses.

- Vegetation treatments are usually best evaluated by categorical effects analyses.

- For resource uses that include substantial surface disturbance (e.g., renewable and non-renewable energy development, OHV use), both categorical analyses of direct effects & reclamation or mitigation measures and continuous effects analyses of indirect effects may be needed.

- The multiple use / sustained yield mission established by FLPMA implies sustainability analyses for resource uses.

NOI and Scoping

- Can communicate AIM as the BLM natural resource monitoring approach which will inform the RMP/EIS

- Can highlight some of the current conditions and changes that you are addressing with the planning effort

DRAFT Land Use Plan

- Affected Environment

    - Draws heavily on the AMS

- Environmental Consequences/Alternatives analysis based on AMS

    - AIM provides a common set of indicators that can connect to field office monitoring, the literature, and remote sensing datasets to help structure the effects of planned actions on natural resources

    - RMP objectives (part of each alternative)

        - Specifying RMP objectives

        - Benchmarks where feasible and defensible (e.g., invasive plants, state water quality standards)

    - Look to analyses of past AIM data in the office – what worked and didn't? including land use plan effectiveness reports, treatment effectiveness reports, land health assessments, etc.

    - Look to research that uses AIM methods or data to link condition to land uses (e.g., SageSTEP treatment effectiveness)

- Look to predictions from modeling which may use AIM data (e.g., John Bradford climate predictions, Sofaer)
-

Proposed RMP

- Monitoring plan -- how will you monitor whether the plan is effective
    - Including implementation level monitoring
    - Provide standard language to use AIM for effectiveness monitoring, e.g. (from Carlsbad RMP):

Assessment Inventory Monitoring (AIM): At the implementation phase, after the record of decision for the RMP has been signed, AIM monitoring objectives will be developed. Monitoring objectives will include monitoring indicators, the condition determination method, the condition benchmarks, a time objective for achieving the desired results, and the proportion required to meet the benchmark. The data will be used to determine landscape condition and inform management decisions.

Chapter 2:

Assessment Inventory Monitoring (AIM): At the implementation phase, after the record of decision for the RMP has been signed, AIM monitoring objectives will be developed. Monitoring objectives will include monitoring indicators, the condition determination method, the condition benchmarks, a time objective for achieving the desired results, and the proportion required to meet the benchmark. The data will be used to determine landscape condition and inform management decisions. AIM monitoring objectives will be developed for soils, wildlife habitat condition, special status species habitat condition, NLCS units, treatment effectiveness, rangeland health and permit renewals, fuels treatment effectiveness, and reclamation effectiveness. However, AIM monitoring objectives are not limited to these resources and land use activities. Additional resources or land use activities, for which monitoring objectives can be developed, may be identified at the implementation phase as needed