

## Ten Practical Questions to Improve Data Quality

Authors: McCord, Sarah E., Welty, Justin L., Courtwright, Jennifer, Dillon, Catherine, Traynor, Alex, et al.

Source: Rangelands, 44(1) : 17-28

Published By: Society for Range Management

URL: <https://doi.org/10.1016/j.rala.2021.07.006>

---

The BioOne Digital Library (<https://bioone.org/>) provides worldwide distribution for more than 580 journals and eBooks from BioOne's community of over 150 nonprofit societies, research institutions, and university presses in the biological, ecological, and environmental sciences. The BioOne Digital Library encompasses the flagship aggregation BioOne Complete (<https://bioone.org/subscribe>), the BioOne Complete Archive (<https://bioone.org/archive>), and the BioOne eBooks program offerings ESA eBook Collection (<https://bioone.org/esa-ebooks>) and CSIRO Publishing BioSelect Collection (<https://bioone.org/csiro-ebooks>).

Your use of this PDF, the BioOne Digital Library, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at [www.bioone.org/terms-of-use](http://www.bioone.org/terms-of-use).

Usage of BioOne Digital Library content is strictly limited to personal, educational, and non-commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

---

BioOne is an innovative nonprofit that sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.



# Ten practical questions to improve data quality

By Sarah E. McCord, Justin L. Welty, Jennifer Courtwright, Catherine Dillon, Alex Traynor, Sarah H. Burnett, Ericha M. Courtright, Gene Fults, Jason W. Karl, Justin W. Van Zee, Nicholas P. Webb, and Craig Tweedie

## On the Ground

- High-quality rangeland data are critical to supporting adaptive management. However, concrete, cost-saving steps to ensure data quality are often poorly defined and understood.
- Data quality is more than data management. Ensuring data quality requires 1) clear communication among team members; 2) appropriate sample design; 3) training of data collectors, data managers, and data users; 4) observer and sensor calibration; and 5) active data management. Quality assurance and quality control are ongoing processes to help rangeland managers and scientists identify, prevent, and correct errors in past, current, and future monitoring data.
- We present 10 guiding data quality questions to help managers and scientists identify appropriate workflows to improve data quality by 1) describing the data ecosystem, 2) creating a data quality plan, 3) identifying roles and responsibilities, 4) building data collection and data management workflows, 5) training and calibrating data collectors, 6) detecting and correcting errors, and 7) describing sources of variability.
- Iteratively improving rangeland data quality is a key part of adaptive monitoring and rangeland data collection. All members of the rangeland community are invited to participate in ensuring rangeland data quality.

**Keywords:** monitoring, data quality, quality assurance, quality control.

*Rangelands* 44(1):17–28

doi 10.1016/j.rala.2021.07.006

Published by Elsevier Inc. on behalf of The Society for Range Management.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

economic, and ecological impacts. Data collected on rangelands are diverse, collected by observers, sensors, and remote sensing through inventories, monitoring, assessments, and experimental studies. Rangeland data are used and re-used in a variety of management and research contexts. Rangeland data applications include, but are not limited to, adjusting stocking rates<sup>1</sup>; evaluating conservation practices<sup>2</sup>; assessing land health at local, regional scales, and national scales<sup>3–5</sup>; determining restoration effectiveness<sup>6,7</sup>; developing or improving models<sup>8,9</sup>; and advancing our understanding of rangeland ecosystems responses to management decisions<sup>10</sup> and natural disturbances.<sup>11</sup> To evaluate progress toward meeting management objectives, managers often use a combination of datasets.<sup>12</sup> Use-based monitoring, such as forage utilization, enables managers to adapt management in response to short-term thresholds.<sup>1</sup> Site-scale monitoring data collected using probabilistic sample designs are often used to infer condition and trend across spatial and temporal scales,<sup>13</sup> such as the Natural Resources Conservation Service (NRCS) National Resources Inventory (NRI) and Bureau of Land Management (BLM) Assessment Inventory and Monitoring (AIM) programs. In all uses of rangeland data, confidence in data-supported decision-making is boosted by high-quality data and eroded by errors and data issues. These issues also relate to rangeland research, where inference from research studies, experimental monitoring, treatments, and practices are also used to support management decisions.<sup>6</sup> For example, the National Wind Erosion Research Network (NWERN) uses a small number of research sites to calibrate dust emission models that can then be run on monitoring datasets such as AIM and NRI to provide managers and conservation planners with dust estimates.<sup>8</sup> If the data from NWERN were found to be faulty, all subsequent dust estimates across multiple study sites would also be faulty. Therefore, any discussion of rangeland data must be paired to a discussion of data quality among land and natural resource managers, conservation planners, and researchers.

Ensuring data quality involves more than maintaining and managing data. This distinction is often overlooked in rangeland research and management,<sup>14</sup> despite the widely recognized need for quality data to support effective decision-making. Data quality describes the degree to which data are useful for a given purpose due to their accuracy, pre-

## Introduction

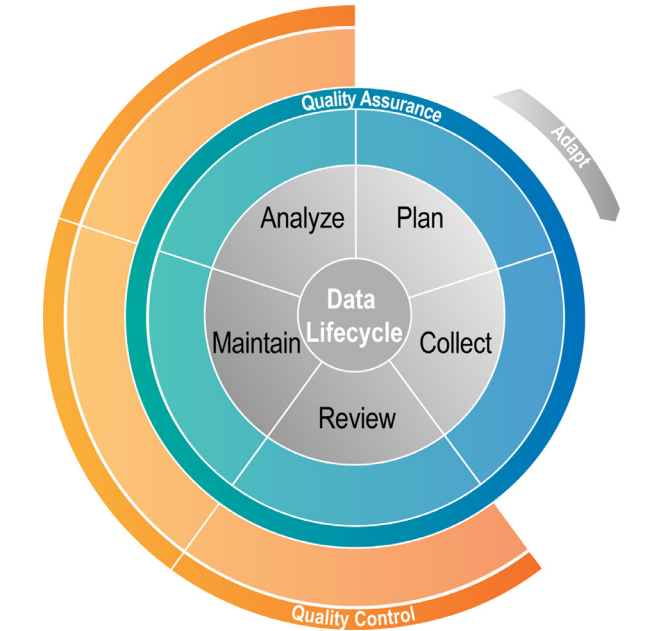
High-quality data are a critical component of rangeland research and management where short- and long-term implications of management decisions have significant policy,

cision, timeliness, reliability, completeness, and relevancy.<sup>15</sup> Data management is the process of collecting, annotating, and maintaining quality data so they are findable, accessible, interoperable, and re-usable.<sup>16</sup> Recent efforts to improve rangeland data quality have focused on improving the effectiveness of data management,<sup>17</sup> including describing the data lifecycle,<sup>18</sup> building data management plans,<sup>19</sup> following data standards,<sup>20</sup> using metadata,<sup>21</sup> and leveraging software for data management.<sup>22</sup> Although high-quality data are a consequence of good data management and good data management identifies data quality issues, data management is not the only process that contributes to data quality. Data quality is also the result of clear communication among team members, well-documented study objectives, careful selection of methods and sample designs, adequate training, and frequent calibration, and appropriate analysis.<sup>23</sup> All members of the rangeland community, including data managers and data collectors, have a role in improving and maintaining data quality.<sup>14</sup>

While the importance of data quality is broadly accepted in the rangeland community, specific steps for ensuring data quality are often unclear, overlooked, or considered synonymous with data management. To address data quality, many inventory and monitoring efforts refer to quality assurance (QA) and quality control (QC) as “QA/QC,” but the meaning of QA/QC can be highly variable between programs and individuals.<sup>12,24</sup> The purpose of QA/QC is to increase the repeatability, defensibility, and usability of data by 1) preventing errors whenever possible, 2) identifying errors that do occur, 3) fixing the error with the correct value if possible, and 4) describing and noting remaining errors that cannot be fixed so they can be excluded from analyses.<sup>23</sup> To achieve these goals, all members of a study or monitoring team, which includes data managers, must have a shared understanding of data quality and what actions they are responsible for to ensure the desired level of data quality is attained.

We find it useful to separate the term QA/QC into its different components: QA and QC (Fig. 1). QA is a proactive process to prevent errors from occurring<sup>12,23</sup> and includes the careful design of the monitoring programs (Stauffer et al. this issue), training and calibration of data collectors and sensors (Newingham et al. this issue), structured data collection (Kachergis et al. this issue), and active data management. QC is a reactive process where errors are identified and corrected if possible<sup>12,23</sup> and includes outlier, logical, and missing data checks and expert review of data that occur sometimes iteratively throughout the data life cycle. Although QA and QC are two distinct processes, both are question driven. QA asks “What could go wrong? How can we prevent it?” and QC asks “What is going wrong? What did go wrong? Where did it go wrong? Why did it go wrong? Can we fix it?” Because both sets of questions are important, we encourage the rangeland community to adopt “QA&QC,” rather than “QA/QC,” which implies that one can exist without the other and is frequently interpreted as a single process (QC).

Here we present 10 practical, overarching QA&QC questions for the rangeland community to adopt (Table 1). If asked regularly and answered thoroughly, these 10 questions can

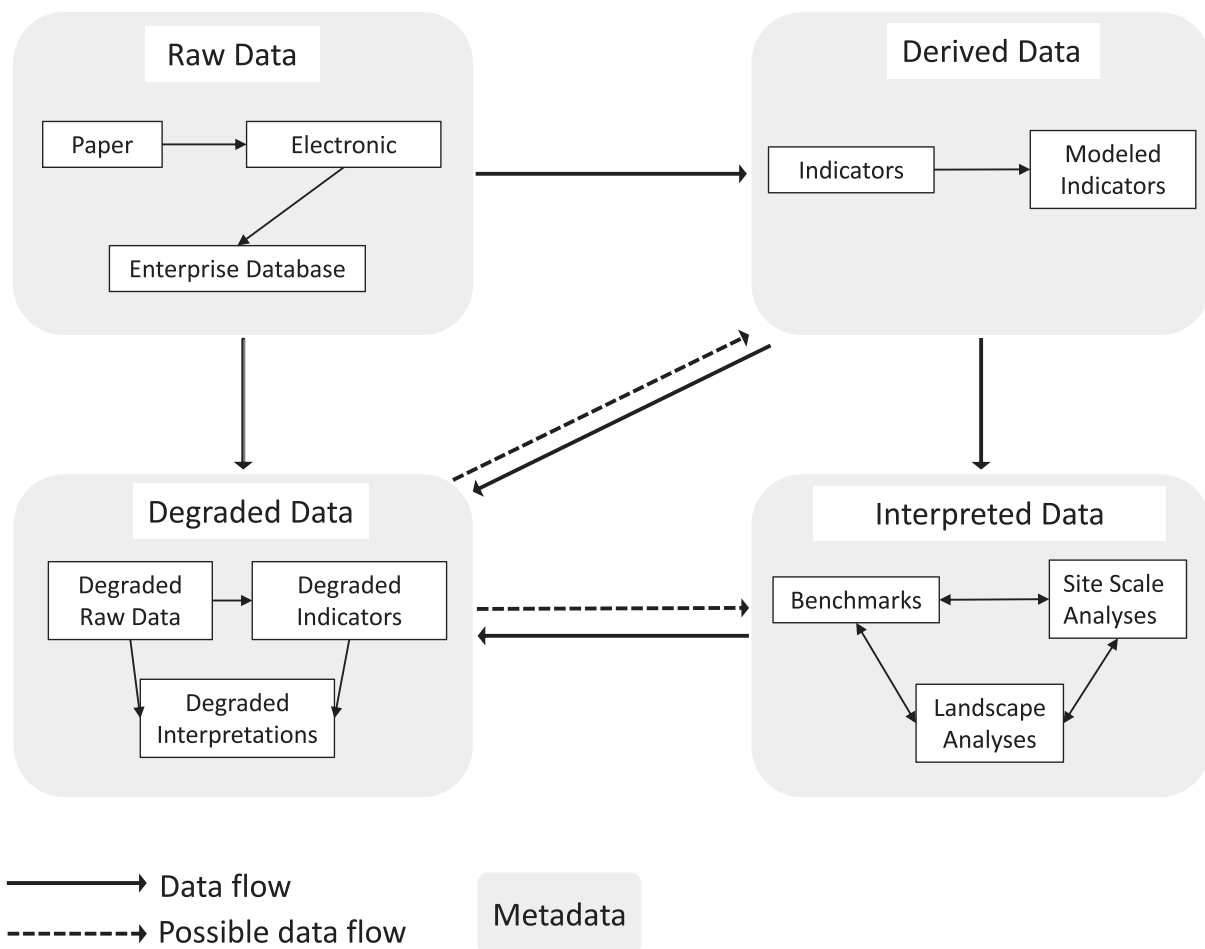


**Figure 1.** The data lifecycle documents the progression of data through planning, data collection, data review, data maintenance and storage, and data analysis and interpretation. Quality assurance occurs continuously throughout the data lifecycle, whereas quality control begins after data are collected. For simplicity we have only identified five lifecycle stages. However, this framework can easily be expanded or contracted to accommodate a different number of lifecycle stages.<sup>14,49</sup> Modified from McCord et al. 2021.<sup>14</sup>

**Table 1**  
Ten important questions to improve rangeland data quality

Question	Quality assurance	Quality control
1. What is my data ecosystem?	X	
2. What is my data quality plan?	X	
3. Who is responsible?	X	
4. How are data collected?	X	
5. How are the data stored and maintained?	X	
6. How will training occur?	X	
7. What is the calibration plan?	X	
8. Are the data complete, correct, and consistent?		X
9. What are the sources of variability?	X- for future data cycle	X- in current data cycle
10. How can we adapt to do better next time?	X- for future data cycle	X- in current monitoring cycle

help researchers and managers improve the quality of rangeland data. The questions build upon each other; however, any question can be revisited at any time. Questions 1 to 7 are QA steps to prevent errors. QC is addressed in Questions 8 to 10. Additionally, Questions 9 and 10 can be considered QC questions for the current data collection cycle and QA questions to adapt future data collection. These questions are used to establish projects, build data management plans, evaluate existing research and monitoring programs, prioritize limited resources, and improve collaboration within data collection efforts.



**Figure 2.** A general conceptual model of the data ecosystem and data flow. Monitoring data can exist in a range of states. Raw data include the original observations or values in paper format, personal electronic file (e.g., Excel, Microsoft Access database, and ESRI file geodatabase), or in an enterprise database (e.g., SQL or Postgres). Raw data may be transcribed from paper to an electronic file to a database. Indicators are derived from the raw data, which can be direct indicators (e.g., bare soil, vegetation composition) or combined with covariates to produce modeled indicators (e.g., dust flux). Data may also exist as interpretations of monitoring data using benchmarks, site scale analysis, or landscape analysis. For each data state, there is an opportunity for data to degrade due to errors of omission (i.e., missing data), commission (incorrect values or observations), or incorrect assumptions regarding the data. Once raw data are in a degraded state it is extraordinarily difficult to achieve a reference state again, although it may be possible to reverse degraded indicators and interpretations. For every type of data, metadata provide critical “data about the data” that enables the use and re-use of data. Rangeland managers and scientists who work with data can build a more detailed version of this conceptual model, appropriate to their data, to anticipate resource needs, potential weak points in the data flow, and where QA&QC steps can prevent or correct degraded data.

## 1. What is my data ecosystem?

Successful implementation of QA&QC is most effective when data collectors, data managers, and data users have a shared understanding of what kinds of data are being collected, how those data are collected and stored, how data will be used, and where there are opportunities for error.<sup>19</sup> To build this shared understanding, we recommend constructing a conceptual diagram of the data ecosystem (Fig. 2). In describing the data ecosystem, scientists and managers identify different kinds of data they are working with, how those data might be transformed from data collection to data storage, to data analysis, and how those data will be documented through metadata. This helps identify where personnel and technological (e.g., data collection applications, databases, and analysis

software) resources are needed and anticipate weak points and opportunities for preventing errors. Within the data ecosystem, it is useful to envision different states (e.g., raw data, calculated indicators or variables, and interpreted data) as well as what each of those states might look like when they are corrupted. If we can anticipate the conditions under which the data no longer accurately represent rangeland condition, it is easier to prevent those issues from occurring. For example, in building a conceptual model of a data ecosystem, a team might notice that they are planning to collect data on paper and store those data in a database. However, the team might note that they currently do not have a process for digitizing the data so that it can be ingested into the database, therefore additional staff time will be needed to enter and check those data to prevent transcription errors. Similarly, while describ-

ing the anticipated analyses, a team realizes that the planned database schema will require transforming data into another data format, so they are able to plan and automate that process.

Although calculated and interpreted data can often be re-stored with some effort as long as the raw data are sound, the opportunities for degraded raw data to be corrected are limited because it is difficult, if not impossible, to replicate field conditions from the raw data collection event.<sup>25</sup> The kind of data (e.g., qualitative vs. quantitative, sensor vs. observational) and available resources available will guide the selection of appropriate data quality actions.<sup>26</sup> The conceptual model of the data ecosystem also recognizes that errors will occur, and therefore includes a process for documenting errors in meta-data documentation when they do occur. It is incumbent upon land managers and researchers who collect and use rangeland data to have a detailed conceptual model of their data to enact a data quality plan that promotes a desirable data workflow, preserves data quality, and documents the data and any known issues.

## 2. What is my data quality plan?

A data quality plan, informed by an understanding of the data ecosystem (Question 1), can make it easier to anticipate where there are opportunities for error and how those errors can be prevented. A data quality plan describes 1) how sample designs and analyses are checked to make sure they meet objectives, 2) strategies for data collector training and calibration, 3) descriptions of the maximum allowable variability in the data, 4) how to detect errors, 5) how to correct those errors if possible, and 6) how to properly annotate the errors so the original value is still recorded and an explanation of the change is given. For instance, how will the team handle location coordinates that look incorrect? Where will the original value be recorded, and how will the change be described? This is necessary in case the updated value is later proven to be incorrect and an additional change based on the original data is needed.

A data quality plan should encompass the entire data life-cycle (Fig. 1), from sample design to analysis, and address the role of each team member in the data collection effort.<sup>27</sup> Because data quality tasks are often captured across a range of documents, it is important to plan how and where you will describe your data quality plans.<sup>28</sup> In addition to important QA&QC steps recorded in data management plans, other data quality plans might be described in protocol documents,<sup>12</sup> sample design documentation,<sup>29</sup> and analysis workflows.<sup>30</sup> We also encourage developing a process for revising the data quality plan in response to insights gained from collecting, managing, and analyzing data. Assigning version numbers and dates to data quality plans will help future data users understand the data ecosystem at the time data were collected. With a documentation strategy in place, Questions 3 to 10 can be used to populate and improve those data quality and data management plans.

## 3. Who is responsible?

Rangeland data collection is often a collaborative, interdisciplinary process.<sup>6</sup> Every member of the monitoring or study team who interacts with data is responsible for maintaining and ensuring the quality and integrity of those data. While in some cases the land manager, project leader, data collector, data manager, analyst, interpreter, and data QC specialist are the same person, often these roles are filled by multiple individuals with different levels of experience or even from different organizations. For instance, the data collector may have little connection to how the data are analyzed and interpreted, whereas the data manager and analyst sometimes are not intimately familiar with the data collection protocols. Within data collection teams, assigned roles and responsibilities also ensure that data quality tasks are appropriately distributed according to skillset. This is particularly important as data collectors also have the greatest power to detect and correct errors before they are embedded in the dataset. Without a shared understanding of how quality data will be collected and stored, errors are likely to occur. Therefore, clearly defining who is responsible for what, and when, is critical to successfully maintaining data quality.<sup>19</sup> Discretely identified roles that clearly tie to the broader monitoring or study objectives empower each member of the team to take ownership of preventing, detecting, correcting, and documenting any errors within their domain and toolset. Detailed timelines of when tasks are to be completed can help budget resources to complete data quality tasks and identify where there might be lapses in data quality due to heavy workload. The longer data stay in a file cabinet or hard drive, the more institutional knowledge is lost as data collectors leave and project leads focus on other projects. Clearly communicating roles has added benefits when multiple kinds of data are involved, as collecting and managing observational data may have different requirements compared with sensor data.<sup>31</sup>

## 4. How are data collected?

Data quality steps will differ depending on whether data are collected electronically or on paper data sheets. Electronic data collection applications provide a cost-efficient method of quickly capturing accurate data while at the same time reducing error rates.<sup>31,32</sup> For instance, hand-recorded geospatial coordinates are often transposed or erroneous. Electronic data capture of study locations can reduce this common error. While more and more data collection programs use electronic data collection,<sup>33,34</sup> considerable amounts of rangeland data are still recorded on paper datasheets. Although upfront costs of equipment purchase, training, and form design to support electronic data capture are greater than paper, these are up-front investments whereas the labor costs of data entry and error checking are continual (Table 2).<sup>32</sup> Initial knowledge required to design electronic forms for field data collection may take time, but once the skill is learned, subsequent forms can be developed quickly with minimal effort and eas-



**Table 2**

Properties and requirements of electronic and paper data.\*

Skill or capability	Electronic	Paper
Design knowledge required	Minimal to Advanced	Minimal
Field technician training (how to enter data)	Moderate	Minimal
Electronic field devices required	Yes	Sometimes <sup>†</sup>
Batteries required	Yes	Sometimes <sup>†</sup>
Customizable data entry forms	Yes	Yes
Data entered	Once <sup>‡</sup>	Twice <sup>‡</sup>
Data are handwritten	No	Yes <sup>‡</sup>
Required fields enforced	Yes	No <sup>‡</sup>
Data validation	Yes	No <sup>‡</sup>
Enforced field types (e.g., text or integer)	Yes	No <sup>‡</sup>
Automatically capture GPS coordinates	Yes	No <sup>‡</sup>
Hidden fields (appear only when necessary)	Yes	No
Scan and automatically enter Barcodes and QR Codes	Yes <sup>§</sup>	No <sup>‡</sup>
Interactive maps	Yes	No <sup>‡</sup>
Electronic backups in the field	Yes <sup>  </sup>	No
Near real time QC	Yes	No <sup>‡</sup>

Note: This table, together with Question 4, can be used to determine the best data collection systems for each monitoring program.

\*Based on the experience of the authors.

<sup>†</sup>Lab and simple field experiments may not require a GPS unit or camera, but most field experiments will require a GPS device or camera, which rely on batteries.

<sup>‡</sup>Indicates a source of additional error that may be introduced.

<sup>§</sup>Camera or laser reader and appropriate software required.

<sup>||</sup>A laptop is generally required for backups if wi-fi or cellular coverage is not available.

ily shared within the range community either through rangeland specific applications (e.g., Database for Inventory, Monitoring, and Assessment,<sup>33</sup> Vegetation GIS Data System,<sup>35</sup> and LandPKS<sup>34</sup>) or customizable survey software (ESRI Survey123 forms, <https://survey123.arcgis.com/>; Open Data Kit, <https://opendatakit.org/>). Electronic data capture also improves data quality through automated data quality checks (see Question 8), automated geospatial data capture, setting allowable data ranges, field standardization (e.g., only numbers allowed in number fields), and controlled domains or options (e.g., plant species name codes) for each field, and automatically linking different data types (e.g., photos and tabular data). Cloud-based data uploads from mobile devices to enterprise databases (e.g., ESRI's Survey123 to ArcGIS online workflow) and automated QC scripts (e.g., the Georgia Coastal Ecosystems sensor QC toolbox) enables real-time error checks that provide feedback to data collectors. This allows data collectors to correct issues if necessary during the field season.<sup>30,31</sup> We encourage the rangeland community to explore the many low-cost options for electronic data capture, but do recognize that paper data collection may be the appropriate solution for some data collection teams due to lack of resources, the size of the team, and some field settings (e.g., wet conditions where waterproof devices are unavailable or remote locations where recharging batteries is difficult). At a

minimum, it is important to have a paper data collection plan as a backup, as screen glare, extreme temperatures, low batteries, and lack of signal are all common challenges of using electronic data capture.

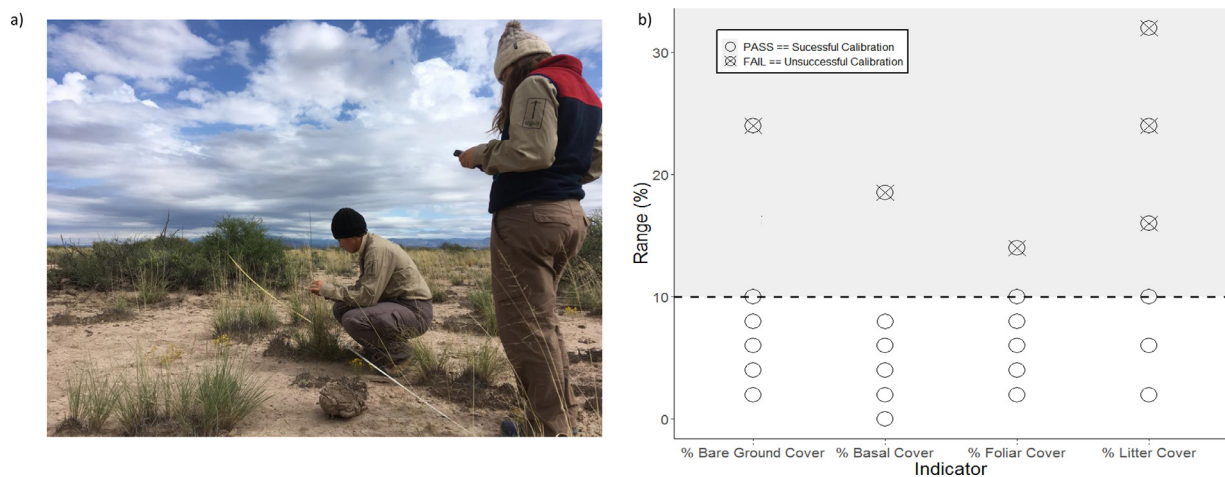
Raw data in an electronic format are also easily ingested into electronic data storage platforms or databases (see Question 5). Emerging data collection mobile platforms (e.g., ESRI Survey123, Open Data Kit) allow for cloud-based data upload and automated data submission. Additionally, a comprehensive data capture and data storage workflow can make rangeland data more readily available for use in data-supported decision-making and research. We anticipate that the availability of electronic data capture applications and central data repositories will continue to increase and become integral to rangeland data collection.

## 5. How are the data stored and maintained?

Proper data management before, during, and after a study is one of the most critical, and often overlooked, parts of data quality.<sup>36</sup> Improper data management can lead to loss of data, reduced inference, misleading conclusions, improper exposure of personally identifiable information, an erosion of trust in the data (by stakeholders or the public), and inability for others to use data in both the short- and long-term.<sup>27</sup> Rangeland data includes raw data (see Question 4), as well as calculated indicators or variables, sample design information, interpreted data, additional tables (e.g., crosswalk tables or those with site level information), geospatial data, and analysis datasets (e.g., benchmarks). Planning for data management includes identifying standard formats for field types (e.g., date, text, integer formats), creating naming conventions, and setting up file and folder structures, backup plans, and security for protected and personally identifiable information.<sup>20,27</sup>

Recent technological and practical advances enable data management to proceed more quickly and efficiently than ever before.<sup>32</sup> These advances include practical guidance on structuring data as “tidy data,” where each observation unit is a row, each variable is a column, and each observation is a cell.<sup>22</sup> Although open-source text files and spreadsheets like Microsoft Excel may be used for storing and visualizing rangeland data, relational databases, such as the ESRI file geodatabase and Microsoft Access, open source databases such as MySQL, and enterprise versions of these databases (e.g., SQL Server, Postgres) allow users to link different kinds of tidy data together in a coherent structure. Relational databases 1) improve storage and access to data by allowing users to efficiently organize and search the database, 2) support complex queries and calculations that present the data in different ways, 3) visualize the data from multiple different viewpoints to aid in the QA&QC and analysis processes, and 4) centralize data across data collectors and over time.<sup>37</sup>

Data management and storage systems also make it easier to share and standardize data, either directly with partners, via web services, or to data repositories. In addition to storing raw, calculated, and analyzed data, data management also in-



**Figure 3.** Calibration (Question 7) is an important process to minimize observer variability in the line-point intercept method (A), especially when the true value is not known or is difficult to measure.<sup>12</sup> For successful calibration in the BLM AIM and the NRCS NRI programs, the line-point intercept absolute range of variability among observers should be less than or equal to 10% (B).<sup>12,50</sup> Photo courtesy of Rachel Burke

### Box 1

#### Calibration among data collectors

Calibrating data collectors is the primary control on detecting and reducing observer variability in rangeland data collection (see Question 7). Calibration among data collectors, as used by the AIM program, addresses observer and measurement error during data collection. It acts as a mechanism of quality assurance by providing time for data collectors to discuss discrepancies in data and clarify differences in protocol interpretation. Data collection begins only after all data collectors are calibrated. Results of AIM calibration exercises (Fig. 3) are used to identify sources of error and protocol misinterpretations, which allows data collectors and project managers to improve training, protocols, and QA&QC practices to mitigate those specific issues. Calibration data from regional AIM training sessions helps observers and instructors identify areas of improvement prior to data collection (Fig. 3). Each observer records measurements on the same transect and those observations are compared. If the range of variability among observers is less than the tolerance range (e.g., 10% for line-point intercept), the calibration is successful and formal data collection may begin. If observers do not successfully calibrate on all indicators for a method, observers discuss the results, identify sources of confusion, and repeat the calibration exercise on a new transect.

cludes curating metadata. Metadata enables the reusability of data by providing managers and researchers with the needed information to interpret and use data. Standardized data formats and metadata documentation (e.g., FGDC, ISO, EML) are most useful when they include data history records, a data dictionary of field name meanings, documented known errors, spatial projection (e.g., NAD83), and date format (e.g., ISO 8601) to guide appropriate use of the data. Metadata provide a validation of data quality to others (see Question 8), thus metadata are a core component of any dataset.<sup>21</sup>

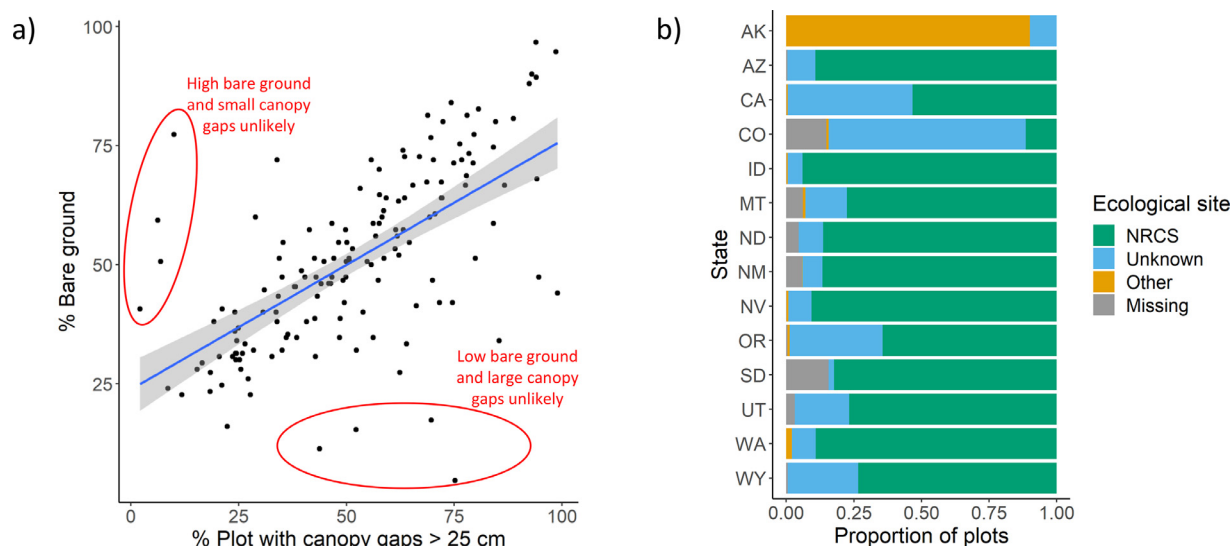
## 6. How will training occur?

Training is the primary opportunity to ensure that team members understand how to collect, manage, and use data properly and consistently. Frequent training, together with clear roles and responsibilities (Question 3), reduces errors due to personnel turnover and provides staff with updates to protocols and workflows. Rangeland monitoring courses are offered in many university programs to give young rangeland professionals exposure to the rangeland data collection and monitoring community (see Newingham et al., this issue). These university courses, as well as in-person national monitoring training programs, and web-based training resources all provide new and experienced users with further guidance (e.g., <https://www.landscapetoolbox.org/training>). Web-based training activities including manuals, courses, and

recorded presentations provide an introduction or brief refresher on how to collect data and use data collection tools (e.g., data collection apps, water quality instruments) when travel to in-person training is impractical. For field-based collection methods, we recommend in-person training as the primary learning method that is supplemented by web-based training. In the field, instructors can demonstrate techniques, answer questions, and provide feedback to data collectors in a more dynamic way than is possible in remote learning settings. Field trainings also should include data capture, either with electronic apps or using paper data sheets, so that data entry can be reviewed and field data workflows, such as daily backups to avoid data loss, are practiced. In these trainings, data collectors benefit from exercises that involve reviewing data for completeness, correctness, and consistency (Question 8) and making corrections as needed. Ideally, all data collectors would attend an in-person training at the beginning of each field season. Many monitoring programs, including AIM, NRI, and Interpreting Indicators of Rangeland Health, hold yearly, standardized field trainings to reach the rangeland data collection community.

## 7. What is the calibration plan?

Calibration by comparison of measurements to a standard or among data collection specialists helps data collectors identify and correct implementation and equipment errors be-



**Figure 4.** Visualizing monitoring data can be used to identify outliers, missing data, and other data errors (Question 8). Visual data checks can include looking for consistency or correlation between methods, such as bare ground estimates from the line-point intercept and canopy gap methods (A). Data visualization can also identify where and why incorrect values were entered. For instance, in the BLM AIM and NRCS NRI, data collectors are required to use the ecological site name recognized by the NRCS; however, in some instances those names are unknown to the data collectors and so the data collectors use a different name or leave the field blank (B). As a result, it may be assumed there is no ecological site ID available, which may not always be the case. In all cases, photos or site revisits are valuable in confirming or correcting errors.

fore they occur during data collection. Calibration is not to be taken lightly. A faulty sensor or uncalibrated field technician can result in incorrectly collected data. If calibration error is within the range of expected values, the error may never be detected resulting in erroneous conclusions. Depending on the data, calibration may occur between data collectors (Box 1, Fig. 3),<sup>12</sup> against a known value,<sup>38,39</sup> or through double-sampling (i.e., repeat sampling of the same attribute with two different methods to improve precision).<sup>40</sup> A calibration exercise is successful if the indicator estimated by data collectors is within an allowable range of variability.<sup>12</sup> If an indicator value falls outside the tolerance range, calibration results are reviewed by the team (data collectors, project leaders, and instructors) at the plot to identify the sources of variability and re-train data collectors. Sensor equipment calibration schedules should follow the factory-recommended calibration intervals. For observational data, we recommend that all data collectors calibrate early and often. For instance, following the *Monitoring Manual for Grassland, Shrubland, and Savanna Ecosystems*,<sup>12</sup> data collectors must successfully calibrate prior to data collection and then monthly or when entering a new ecosystem, whichever occurs first. Similarly, for species composition by weight and other production methods, recalibration may occur more frequently during early and rapid phenological changes when encountering a new precipitation pattern, landform, utilization rates, and changes in vegetation. If a new data collector joins the data collection team, a calibration event also is triggered.

Although it is not common practice to publish calibration results alongside rangeland data, we encourage the rangeland community to adopt this practice. Publishing calibration results can verify that calibration steps were taken and detail the

observer variability within the dataset (Question 9). Calibration data are also important when describing advantages and disadvantages between methods and prior to replacing an existing method with a new one.<sup>41</sup> Calibration results may provide opportunities for including observer variability as a covariate in analysis. Public calibration data can identify areas of improvement for teaching data collection methods (Question 6), where if one program is especially successful at calibration, the community can learn from those successful training and data collection practices.

## 8. Are the data complete, correct, and consistent?

Frequent review of rangeland data for completeness, correctness, and consistency will detect errors and missing data in a timely and efficient manner (Fig. 4). Errors detected in this review process are best addressed in the field, during data collector review. However, these checks are also important steps in data storage and analysis workflows. Many of these data checks can be automated using digital data collection forms and web-based dashboards (e.g., Tableau, ESRI ArcGIS Insights). Data are complete if they have every data element present so that every field in every data form is completed for every method required for that project. Data are correct if they are accurate and follow the data collection protocol. For instance, a correct application of the line-point intercept method requires accurate plant identification, proper pin drop technique, and consistent species code selection following a known taxonomic reference (e.g., USDA Plant codes, unknown plant protocol) in the correct location



**Table 3**

Summary of BLM AIM lotic core indicator crew and intra-annual variability (Question 9) as assessed by residual mean square error (RMSE), coefficient of variation (CV), and signal to noise (S:N) ratio

Category	Indicator	First study			Second study		
		RMSE	CV	S:N	RMSE	CV	S:N
Water quality	pH (SU)	High*	High*	Low*	High*	High*	Low*
	Specific conductance ( $\mu\text{S}/\text{cm}$ )	High*	High*	Moderate*	Low†	Moderate†	Moderate†
	Total phosphorous ( $\mu\text{g}/\text{L}$ )	High*	High*	High*	Low†	Moderate†	Moderate†,§
	Total nitrogen ( $\mu\text{g}/\text{L}$ )	Low‡	Low‡	Low‡	Moderate†	High†	Moderate†
	Macroinvertebrate O/E score (unitless)	High*	High*	Moderate*	High*	High*	Moderate*
Riparian function	Bank cover + stability (%)	High*	High*	High*	High*	High*	Moderate*
	Bank stability (%)	High*	High*	Moderate*	Low‡	Low‡	Low‡
	Bank cover (%)	Moderate†	High†	Moderate†	Moderate†	High†	Moderate†
	Bank angle (°)	Moderate†	High†	Moderate†	High*	High*	Moderate*
	Canopy cover - bank (%)	High*	High*	Moderate*	Moderate†	High†	Moderate†
	Riparian vegetative complexity (unitless)	Low†	Moderate†	Low†	Low‡	Low‡	Low‡
	Riparian vegetative complexity - woody only (unitless)	High*	High*	Moderate*	Low‡	Low‡	Low‡
	Non-native woody vegetation (%)	NA	NA	NA	High†	Low†	Moderate†
	Stream habitat/function	High†	Moderate†	Moderate†	Moderate†	Moderate†	Moderate†
	Pool frequency (#/km)	NA	NA	NA	Moderate†	High†	Moderate†
Stream habitat/function	Floodplain connectivity (unitless)	Moderate†	High†	Moderate†	Moderate†	High†	Low†
	Large woody debris (#/100 m)	High*	High*	High*	Moderate	Low	Low
	Depth coefficient of variation (%)	High*	High*	High*	Moderate†	Moderate†	Low†
	Instream habitat complexity (unitless)	Low‡	Low‡	Low‡	Low‡	Low‡	Low‡
	Bankfull width (m)	High*	High*	High*	High*	High*	High*
Covariate	Slope (°)	Moderate†	High†	Moderate†	High*	High*	High*
	Entrenchment (unitless)	NA	NA	NA	Moderate†	High†	Low†

Note: Each indicator is rated as having High, Moderate, or Low precision across the three measures. RMSE thresholds were based on published values and professional judgment of meaningful differences in measured indicators.<sup>47</sup> CV Values < 20% are characteristic of high consistency; 20–35% moderate consistency; and > 35% low.<sup>47</sup> Following Roper et al.,<sup>47</sup> we used S:N to assess indicator precision where S:N < 2 equals low precision;  $\geq 2.0$  to < 10 equals moderate precision, and  $\geq 10$  equals high precision.

\* Indicator rated as having high precision for at least two of the three measures.

† Indicator rated as having high precision for at least one measure and moderate for a second.

‡ Indicator rated as having low precision for two or more measures.

§ Outliers were removed from total phosphorous analyses for one pair of sites in the 2013–2015 study and two pairs in the 2017 study. Outlier inclusion resulted in Moderate/Low/Low and Low/Moderate/Low ratings, respectively.

on the datasheet.<sup>12</sup> Although data reviewers might find it difficult to check the pin drop technique later, we can infer that, if both plant identification and other elements of a pin drop are recorded correctly, the likelihood of other methodological errors is lower. It is also helpful to review data for likely spelling mistakes (e.g., squirrel, squirell, squirell), as typos and unclear handwriting result in species misidentification and erroneous values. Data checks might also find data to be correct if measured values fall within allowable ranges (e.g., percentages must be between 0 and 100%).

Correct data also can be verified by consistency checks to verify that data follow expected patterns<sup>16</sup> or logical relationships among data collection programs, between methods, over time, and within the ecological potential of the site.<sup>38</sup> Method consistency checks, for instance, might verify that stream bankfull channel width is greater than wetted width when sampling below flood stages or that total canopy gaps are equal to or less than bare soil cover (Fig. 4). Ecological consistency checks rely on local knowledge to ensure that range-

land data are consistent with our understanding of ecosystem processes and change. Specific checks include ensuring that species are consistent with ecological site potential and, where repeat measurements are available, that changes in species composition are likely given climate and management data. Where outliers exist, ecological checks can determine if those outliers are due to site heterogeneity, extreme conditions, or due to an error.<sup>42</sup> For instance, stream pH values below 6 or above 9 are only possible if substantial alteration has occurred (e.g., acid mine drainage). As rangeland ecosystems change, we urge extreme caution before removing outlier values from analyses, as it is possible that these values represent previously unobserved disturbances (e.g., fire, drought, and climate change) or novel ecosystems.<sup>43</sup> Therefore we recommend a “preponderance of evidence” approach, using photos and other datasets, to identify erroneous outliers.<sup>29</sup>

Quality assurance plans should contain data quality objectives that set desired levels of completeness, correctness, and consistency.<sup>23</sup> If data do not meet these objectives, corrective

## Box 2

### Studying variance decomposition in the BLM AIM Wadeable Stream and River Core Methods

The BLM Lotic AIM conducted a study to quantify the intra-annual variability (see Question 9) for two different iterations of the Wadeable Stream and River AIM field protocol. In this study, approximately 10% of the total monitoring locations were resampled, 25 locations for the first protocol iteration (2013–2015) and 37 for the second (2017). Locations were distributed proportionally among geographical regions and stream types to adequately characterize spatial variation and the types of streams data collectors encountered. Although, the study aims included separating sampling and nonsampling error, this proved difficult. To minimize within season temporal variation and attempt to isolate data collector bias, locations were sampled within 4 weeks of each other. The first study assessed crew variability among all possible pairs of data collectors and crews were not aware of repeat sampling. The second study assessed crew variability between a single crew and all other crews due to crew logistical constraints. Within season variability was quantified using residual mean square error (average deviation, in native units, among repeat measurements), the coefficient of variation (variability between repeat measurements scaled to the mean), and the signal to noise ratio (estimate of sample variability relative to site variability; Table 3). Each measure of variability was rated as corresponding to high, moderate, or low repeatability and then used as a line of evidence to determine overall repeatability of the BLM Lotic AIM Wadeable Stream and River Core indicators. As a result of these two studies, adaptive monitoring principles were applied.<sup>46</sup> Some indicators were omitted from the program (e.g., ocular estimates of instream habitat complexity), and protocol changes were made to others (e.g., floodplain connectivity) to improve consistency among data collectors (see Question 10). Measures of indicator precision were comparable to those of other monitoring programs.<sup>47</sup> This assures data users of the high quality of lotic AIM data and its comparability to other monitoring programs.

action is taken, if possible, and all data edits are tracked (see Question 2) with a clear rationale for the edit. If no corrective action is possible, data are omitted if they are clearly wrong or, if they are questionable but not clearly wrong, data are flagged as suspect with a clear comment about why they may not be appropriate to use in certain analyses. For example, a vegetation cover value deemed too high to be plausible that cannot be fixed would be excluded from an analysis looking at average cover but could still be included in an occupancy analysis. If electronic data capture is part of the data collection program (see Question 4), many checks for completeness, correctness, and consistency can be programmed into data collection applications to prevent common errors. However, ecological checks generally require manual review of data after collection and a level of expertise that individual data collectors may not have. Photos and data visualization also can assist with these ecological checks (Fig. 4).

## 9. What are the sources of variability?

Even if data are complete, correct, and consistent it is important to identify where there are general sources of variation in a dataset. In addition to spatial and temporal ecological variation, variability in rangeland data is due to variation in data collectors. Collectively, these factors add noise (uncertainty) to rangeland data that obscure our capacity to detect differences among locations or changes through time.<sup>44</sup> Sampling error occurs when your estimate differs from the true value because you have only sampled a portion of the entire population.<sup>12</sup> Sample design, stratification, and sample size can influence adequate characterization of ecological variation through space and time (see Stauffer et al. this issue for a review of this topic). Additionally, sampling and nonsampling variance components can be combined in power analyses to determine the size of changes the data collection effort can detect and assist with designing better studies.<sup>45</sup> Sampling error is an important source of variability and should be considered prior to collecting or analyzing data. Here we focus our discussion on variance components that are a result of nonsampling errors (i.e., errors not due to the lim-

itations of sample designs in measuring ecological variability), which can be addressed through QA&QC. Sampling and nonsampling variance components can be combined in power analyses to determine the size of changes the data collection effort can detect and can assist with designing better studies (Box 2). Describing variability across data collectors can identify which indicators data collectors struggle to measure consistently (Box 2, Question 7) and improve data collection protocols and training (Box 1, Question 6). Ultimately, certain indicators may not be measurable at desired levels of precision no matter how many replicates are taken or how well data collectors are trained. After careful consideration through the adaptive monitoring process,<sup>46</sup> new methods of measuring these indicators may be selected, the indicators may be omitted from the study, or the indicators may only be sampled in situations where the indicators are needed, and less precise data are acceptable.

Quantifying different components of indicator variability is time intensive and expensive. Thus, only a few monitoring programs and studies have conducted such analyses.<sup>47,48</sup> If similar data are collected across monitoring programs and studies, data may be used to quantify sampling and nonsampling error across locations and years, but estimates of within season variability could differ among programs. For example, the precision of stream indicators such as bankfull width, percent fine sediment, and percent stream pool habitat differs among monitoring programs that use relatively similar field methods.<sup>47</sup> Such field measurement variation, or intra-annual variability, can result from the combined effects of measurement variation among different field crews, within-season environmental variability, and changes in location. Intra-annual variability is likely the variance component of most interest to monitoring programs assessing trend across years so that they can make proper inferences in analysis. For example, if percent vegetative cover changes from 80% to 90% between year 1 and 2, but data collected within the same year by two different data collectors differs by 10% at a monitoring location, any changes in cover <10% could simply be due to observer bias rather than management changes. Ideally, monitoring programs and long-term studies would quantify variability among crews within a season for each major iteration of a protocol (Box 2).

## 10. How can we adapt to do better next time?

Improving rangeland data quality involves using QA&QC questions to evaluate data and adaptively manage monitoring and research programs. Data collection, especially within monitoring and long-term studies, is an iterative process, with continual improvements based on feedback from the team, metrics from training and calibration, implementation of data management systems, and results of data review.<sup>36</sup> Even in the best data collection systems, mistakes will be made throughout the data collection process. New situations or “edge cases” may be encountered that highlight opportunities for clarifying protocols. Successful data collection efforts identify and learn from those mistakes and adjust for the next field season or in the next study. Rangeland studies and monitoring programs can learn from each other by sharing these mistakes and lessons learned with the community. Through adaptive monitoring, QA&QC Questions 1 to 9 can be revised and refined in subsequent monitoring cycles to produce a higher quality dataset. For example, within the BLM AIM program, data management protocols, calibration protocols, training, and electronic data capture programs are updated and revised annually in response to feedback from data collectors, data users, and errors found during QA&QC. However, we caution against rapid changes in monitoring programs and long-term studies, as substantial shifts can limit power to detect change or differences over space and time. Therefore, when a comparative analysis is critical, care should be taken to ensure that any updates to the monitoring program and study are thoughtfully considered and other data sources (e.g., remote sensing<sup>11</sup>) are available to provide a preponderance of evidence in detecting trend.<sup>36</sup>

## Conclusions

High-quality rangeland data are key to data-supported decision-making and adaptive rangeland management. We have presented 10 QA&QC questions that managers, data collectors, and scientists can address to ensure data quality and thereby increase the efficacy of monitoring and other data collection efforts. The answers to our 10 questions can guide the appropriate personnel, data management tools, and analysis strategies to maintain data quality throughout the data lifecycle. Given the expense of collecting and managing rangeland data, improving data quality workflows will reduce the frequency of costly errors and ensure that rangeland data are fit for use in decision-making and in rangeland research and modeling. In the experience of the authors, high-quality data are also more likely to be collected once and used for many purposes, which increases the efficiency of rangeland monitoring. Research studies, assessment, monitoring, and inventory programs can improve data quality by thoroughly describing the data ecosystem, clearly defining roles and responsibilities, adopting appropriate data collection and data management strategies, identifying sources of error, preventing

those errors where possible, and describing sources of measurement variability. Ensuring data quality is an iterative process and improves through adaptive management of monitoring and inventory programs. The QA&QC questions posed in this paper apply to all members of the rangeland community and all data collected in experimental studies, inventories, short-term monitoring, and long-term monitoring programs. We encourage interagency and interdisciplinary partnerships to discuss these questions early so that data quality is ensured as a collaborative process. Improving data quality will improve our ability to detect condition, pattern, and trend on rangelands, which are needed to improve adaptive management and co-production of scientific research for natural resource management.

## Declaration of Competing Interest

S.E.M is a Guest Editor for the Special Issue and J.W.K. is Editor in Chief for Rangelands, but they were not involved in the handling, review, or decision process for this manuscript. The content of sponsored issues of Rangelands is handled with the same editorial independence and single-blind peer review as that of regular issues.

## Acknowledgments

The data for this paper are available on request from the corresponding author. This research was supported by the USDA NRCS (agreement 67-3A75-17-469) and the BLM (agreement 4500104319). This research was a contribution from the Long-Term Agroecosystem Research (LTAR) network. LTAR is supported by the United States Department of Agriculture. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US Government.

## References

1. HOLECHEK JL. An approach for setting the stocking rate. *Rangelands*. 1988; 10(1):10–14. <http://hdl.handle.net/10150/640265>.
2. METZ LJ, REWA CA. Conservation Effects Assessment Project: assessing conservation practice effects on grazing lands. *Rangelands*. 2019; 41(5):227–232. doi:10.1016/j.rala.2019.07.005.
3. KACHERGIS E, LEPAK N, KARL MG, MILLER SW, DAVIDSON Z. *Guide to Using AIM and LMF Data in Land Health Evaluations and Authorizations of Permitted Uses*. U.S. Department of the Interior, Bureau of Land Management; 2020 National Operations Center; Accessed July 28, 2020 <https://www.blm.gov/documents/noc/blm-library/technical-note/guide-using-aim-and-lmf-data-land-health-evaluations-and>.
4. HERRICK JE, LESSARD VC, SPAETH KE, ET AL. National ecosystem assessments supported by scientific and local knowledge. *Front Ecol Environ*. 2010; 8(8):403–408. doi:10.1890/100017.
5. TOEVS GR, KARL JW, TAYLOR JJ, ET AL. Consistent indicators and methods and a scalable sample design to meet assessment, inventory, and monitoring information needs across scales.



- Rangelands*. 2011; 33(4):14–20. doi:10.2111/1551-501X-33.4.14.
6. BESTELMEYER BT, BURKETT LM, LISTER L, BROWN JR, SCHOOLEY RL. Collaborative approaches to strengthen the role of science in rangeland conservation. *Rangelands*. 2019; 41(5):218–226. doi:10.1016/j.rala.2019.08.001.
  7. TRAYNOR ACE, KARL JW, DAVIDSON ZM. Using Assessment, Inventory, and Monitoring data for evaluating rangeland treatment effects in Northern New Mexico. *Rangelands*. 2020; 42(4):117–129. doi:10.1016/j.rala.2020.06.001.
  8. WEBB NP, VAN ZEE JW, KARL JW, ET AL. Enhancing wind erosion monitoring and assessment for U.S. rangelands. *Rangelands*. 2017; 39(3):85–96. doi:10.1016/j.rala.2017.04.001.
  9. JONES MO, ALLRED BW, NAUGLE DE, ET AL. Innovation in rangeland monitoring: annual, 30 m, plant functional type percent cover maps for U.S. rangelands, 1984–2017. *Ecosphere*. 2018; 9(9). doi:10.1002/ecs2.2430.
  10. VEBLEN KE, PYKE DA, ALDRIDGE CL, CASAZZA ML, ASSAL TJ, FARINHA MA. Monitoring of livestock grazing effects on Bureau of Land Management land. *Rangel Ecol Manag*. 2014; 67(1):68–77. doi:10.2111/REM-D-12-00178.1.
  11. BARKER BS, PILLIOD DS, RIGGE M, HOMER CG. Pre-fire vegetation drives post-fire outcomes in sagebrush ecosystems: evidence from field and remote sensing data. *Ecosphere*. 2019; 10(11):e02929. doi:10.1002/ecs2.2929.
  12. HERRICK JE, VAN ZEE JW, MCCORD SE, COURTRIGHT EM, KARL JW, BURKETT LM. *Monitoring Manual for Grassland, Shrubland, and Savanna Ecosystems*. 1. 2nd ed. USDA - ARS Jornada Experimental Range; 2018 Accessed November 3, 2018 <https://www.landscapetoolbox.org/manuals/monitoring-manual/>.
  13. BESTELMEYER B, BROWN J, FUHLENDORF S, FULTS G, WU XB, BRISKE D. A landscape approach to rangeland conservation practices. In: Briske DD, ed. *Conservation Benefits of Rangeland Practices: Assessment, Recommendations and Knowledge Gaps*. United States Department of Agriculture, Natural Resources Conservation Service; 2011:337–370.
  14. MCCORD SE, WEBB NP, VAN ZEE JW, ET AL. Provoking a cultural shift in data quality. *BioScience*. 2021; 71(6):647–657. doi:10.1093/biosci/biab020.
  15. WANG RY, STRONG DM. Beyond accuracy: what data quality means to data consumers. *J Manag Inf Syst*. 1996; 12(4):5–33. doi:10.1080/07421222.1996.11518099.
  16. WILKINSON MD, DUMONTIER M, AALBERSBERG IJ, ET AL. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016; 3. doi:10.1038/sdata.2016.18.
  17. BORER ET, SEABLOOM EW, JONES MB, SCHILDHAUER M. Some simple guidelines for effective data management. *Bull Ecol Soc Am*. 2009; 90(2):205–214. doi:10.1890/0012-9623-90.2.205.
  18. MICHENER WK, JONES MB. Ecoinformatics: supporting ecology as a data-intensive science. *Trends Ecol Evol*. 2012; 27(2):85–93. doi:10.1016/j.tree.2011.11.016.
  19. MICHENER WK. Ten simple rules for creating a good data management plan. *PLoS Comput Biol*. 2015; 11(10). doi:10.1371/journal.pcbi.1004525.
  20. BRINEY K. The problem with dates: applying ISO 8601 to research data management. *Journal of eScience Librarianship*. 2018; 7(2). doi:10.7191/jeslib.2018.1147.
  21. FEGRUS EH, ANDELMAN S, JONES MB, SCHILDHAUER M. Maximizing the value of ecological data with structured metadata: an introduction to Ecological Metadata Language (EML) and principles for metadata creation. *Bull Ecol Soc Am*. 2005; 86(3):158–168. doi:10.1890/0012-9623(2005)86[158:MTVOED]2.0.CO;2.
  22. WICKHAM H. Tidy data. *J Stat Softw*. 2014; 59(1):1–23. doi:10.18637/jss.v059.i10.
  23. MICHENER WK. Quality assurance and quality control (QA/QC). In: Recknagel F, Michener WK, eds. *Ecological Informatics: Data Management and Knowledge Discovery*. Springer International Publishing; 2018:55–70.
  24. U.S. EPA. National Coastal Condition Assessment Quality Assurance Project Plan. 2014. Accessed September 14, 2020. [https://www.epa.gov/sites/production/files/2016-05/documents/ncca\\_2015\\_qapp\\_version\\_2.1.pdf](https://www.epa.gov/sites/production/files/2016-05/documents/ncca_2015_qapp_version_2.1.pdf)
  25. SPECHT A, BOLTON M, KINGSFORD B, SPECHT R, BELBIN L. A story of data won, data lost and data re-found: the realities of ecological data preservation. *Biodivers Data J*. 2018; 6:e28073. doi:10.3897/BDJ.6.e28073.
  26. VAN SCHALKWYK F, WILLMERS M, MCNAUGHTON M. Viscous open data: the roles of intermediaries in an open data ecosystem. *Inf Technol Dev*. 2016; 22(sup1):68–83. doi:10.1080/02681102.2015.1081868.
  27. BRINEY K, COATES H, GOBEN A. Foundational practices of research data management. *Research Ideas and Outcomes*. 2020; 6:e56508. doi:10.3897/rio.6.e56508.
  28. MICHENER WK. Ecological data sharing. *Ecol Informatics*. 2015; 29:33–44. doi:10.1016/j.ecoinf.2015.06.010.
  29. HERRICK JE, VAN ZEE JW, HAYSTAD KM, BURKETT LM, WHITFORD WG, ET AL. *Monitoring Manual for Grassland, Shrubland and Savanna Ecosystems*. II. USDA-ARS Jornada Experimental Range; 2005 *Design, Supplementary Methods and Interpretation*.
  30. YENNI GM, CHRISTENSEN EM, BLEDSOE EK, ET AL. Developing a modern data workflow for regularly updated data. *PLoS Biol*. 2019; 17(1). doi:10.1371/journal.pbio.3000125.
  31. STURTEVANT C, FLAGG C, LEISSO N, ET AL. NEON Science Data Quality Plan. Accessed April 3, 2020. <https://data.neonscience.org/api/v0/documents/NEON.DOC.004104vA>
  32. THRIEMER K, LEY B, AME SM, ET AL. Replacing paper data collection forms with electronic data entry in the field: findings from a study of community-acquired bloodstream infections in Pemba, Zanzibar. *BMC Res Notes*. 2012; 5(1):113. doi:10.1186/1756-0500-5-113.
  33. COURTRIGHT EM, VAN ZEE JW. The Database for Inventory, Monitoring, and Assessment (DIMA). *Rangelands*. 2011; 33(4):21–26. doi:10.2111/1551-501X-33.4.21.
  34. HERRICK JE, KARL JW, MCCORD SE, ET AL. Two new mobile apps for rangeland inventory and monitoring by landowners and land managers. *Rangelands*. 2017; 39(2):46–55. doi:10.1016/j.rala.2016.12.003.
  35. DESPAIN DW, PERRY C. Vegetation GIS Data System. Accessed March 26, 2021. <https://vgs.arizona.edu/>
  36. LINDENMAYER DB, LIKENS GE. The science and application of ecological monitoring. *Biol Conserv*. 2010; 143(6):1317–1328. doi:10.1016/j.biocon.2010.02.013.
  37. CODD EF. A relational model of data for large shared data banks. *Commun ACM*. 1970; 13(6):377–387. doi:10.1145/362384.362685.
  38. CAMPBELL JL, RUSTAD LE, PORTER JH, ET AL. Quantity is nothing without quality: automated QA/QC for streaming environmental sensor data. *BioScience*. 2013; 63(7):574–585. doi:10.1525/bio.2013.63.7.10.
  39. SALLEY SW, HERRICK JE, HOLMES CV, ET AL. A comparison of soil texture-by-feel estimates: implications for the citizen soil scientist. *Soil Sci Soc Am J*. 2018; 82(6):1526. doi:10.2136/sssaj2018.04.0137.
  40. WILM HG, COSTELLO DF, KLIPPLE GE. Estimating forage yield by the double-sampling method. *Agronomy Journal*. 1944; 36(3):194–203. doi:10.2134/agronj1944.00021962003600030003x.



41. BARKER BS, PILLIOD DS, WELTY JL, ARKLE RS, G M, TOEVS GR. An introduction and practical guide to use of the Soil-Vegetation Inventory Method (SVIM) data. *Rangel Ecol Manag.* 2018; 71(6):671–680. doi:[10.1016/j.rama.2018.06.003](https://doi.org/10.1016/j.rama.2018.06.003).
42. ZUUR AF, IENO EN, ELPHICK CS. A protocol for data exploration to avoid common statistical problems. *Methods Ecol Evol.* 2010; 1(1):3–14. doi:[10.1111/j.2041-210X.2009.00001.x](https://doi.org/10.1111/j.2041-210X.2009.00001.x).
43. WILLIAMS JW, ST JACKSON. Novel climates, no-analog communities, and ecological surprises. *Front Ecol Environ.* 2007; 5(9):475–482.
44. VANDENBERGHE V, BAUWENS W, VANROLLEGHEM PA. Evaluation of uncertainty propagation into river water quality predictions to guide future monitoring campaigns. *Environ Model Softw.* 2007; 22(5):725–732. doi:[10.1016/j.envsoft.2005.12.019](https://doi.org/10.1016/j.envsoft.2005.12.019).
45. LARSEN DP, KAUFMANN PR, KINCAID TM, URQUHART NS. Detecting persistent change in the habitat of salmon-bearing streams in the Pacific Northwest. *Can J Fish Aquat Sci.* 2004; 61(2):283–291. doi:[10.1139/f03-157](https://doi.org/10.1139/f03-157).
46. LINDENMAYER DB, LIKENS GE. Adaptive monitoring: a new paradigm for long-term research and monitoring. *Trends Ecol Evol.* 2009; 24(9):482–486. doi:[10.1016/j.tree.2009.03.005](https://doi.org/10.1016/j.tree.2009.03.005).
47. ROPER BB, BUFFINGTON JM, BENNETT S, ET AL. A comparison of the performance and compatibility of protocols used by seven monitoring groups to measure stream habitat in the Pacific Northwest. *N Am J Fish Manag.* 2010; 30(2):565–587. doi:[10.1577/M09-061.1](https://doi.org/10.1577/M09-061.1).
48. WEBB NP, CHAPPELL A, EDWARDS BL, ET AL. Reducing sampling uncertainty in Aeolian Research to improve change detection. *J Geophys Res Earth Surf.* 2019; 124(6):1366–1377. doi:[10.1029/2019JF005042](https://doi.org/10.1029/2019JF005042).
49. MICHENER WK, ALLARD S, BUDDEN A, ET AL. Participatory design of DataONE—enabling cyberinfrastructure for the biological and environmental sciences. *Ecol Informatics.* 2012; 11:5–15. doi:[10.1016/j.ecoinf.2011.08.007](https://doi.org/10.1016/j.ecoinf.2011.08.007).
50. USDA NATURAL RESOURCES CONSERVATION SERVICE. National Resources Inventory Grazing Land On-Site Data Collection: Handbook of Instructions. 2020. Accessed September 15, 2020. <https://grazingland.cssm.iastate.edu/site-data-collection-handbook-instructions>

---

*Authors are from: USDA-ARS Jornada Experimental Range, Las Cruces, NM, USA (Sarah McCord, Catherine Dillon, Ericha M. Courtright, Justin W. Van Zee, Nicholas P. Webb); University of Texas-El Paso, El Paso, TX, USA (Sarah McCord, Craig Tweedie); US Geological Survey, Forest and Rangeland Ecosystem Science Center, Boise, ID, USA (Justin L. Welty); Utah State University, Logan, UT, USA (Jennifer Courtwright); New Mexico State University, Las Cruces, NM, USA (Catherine Dillon, Nicholas P. Webb); Bureau of Land Management, National Operations Center, Denver, CO, USA (Alex Traynor, Sarah H. Burnett); USDA-Natural Resources Conservation Service, Portland, OR, USA (Gene Fults); Department of Forest, Rangeland and Fire Sciences, University of Idaho, Moscow, ID, USA (Jason W. Karl)*