Challenges with secondary use of multi-source water-quality data in the United States

Lori A. Sprague [a,*], Gretchen P. Oelsner [b], Denise M. Argue [c]

[a] U.S. Geological Survey, Denver Federal Center, Box 25046, Bldg. 53, MS 415, Denver, CO, 80225, United States
[b] U.S. Geological Survey, 6700 Edith Blvd. NE, Suite B, Albuquerque, NM, 87113, United States
[c] U.S. Geological Survey, 331 Commerce Way, Pembroke, NH 03275, United States

ABSTRACT

Combining water-quality data from multiple sources can help counterbalance diminishing resources for stream monitoring in the United States and lead to important regional and national insights that would not otherwise be possible. Individual monitoring organizations understand their own data very well, but issues can arise when their data are combined with data from other organizations that have used different methods for reporting the same common metadata elements. Such use of multi-source data is termed "secondary use"—the use of data beyond the original intent determined by the organization that collected the data. In this study, we surveyed more than 25 million nutrient records collected by 488 organizations in the United States since 1899 to identify major inconsistencies in metadata elements that limit the secondary use of multi-source data. Nearly 14.5 million of these records had missing or ambiguous information for one or more key metadata elements, including (in decreasing order of records affected) sample fraction, chemical form, parameter name, units of measurement, precise numerical value, and remark codes. As a result, metadata harmonization to make secondary use of these multi-source data will be time consuming, expensive, and inexact. Different data users may make different assumptions about the same ambiguous data, potentially resulting in different conclusions about important environmental issues. The value of these ambiguous data is estimated at $US12 billion, a substantial collective investment by water-resource organizations in the United States. By comparison, the value of unambiguous data is estimated at $US8.2 billion. The ambiguous data could be preserved for uses beyond the original intent by developing and implementing standardized metadata practices for future and legacy water-quality data throughout the United States.

Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Substantial investments have been made in water-quality monitoring in the United States by Federal, State, and local governments, Tribes, water utilities, universities, and many others since the inception of water-quality monitoring in the early 20th century (Myers, 2015), but funds for monitoring have slowly been eroding over time (National Research Council, 2004). For example, the U.S. Geological Survey (USGS) began monitoring a network of 505 stream sites throughout the United States in 1991 through the National Water-Quality Assessment (NAWQA) project of its National Water-Quality Program. Since that time, reductions in funding have led to a reduction in the NAWQA network to 117 sites, limiting the spatial and temporal resolution of key findings from the program (Rowe et al., 2013). Two other USGS national monitoring networks—the Hydrologic Benchmark Network (focused on undeveloped streams) and the National Stream-Quality Accounting Network (focused on large rivers)—were reduced from 54 to 15 sites and from 518 to 39 sites, respectively, between the 1970s and the 1990s (Hooper et al., 2001). Funding decreases have affected other organizations as well. For example, the Temporally Integrated Monitoring of the Environment (TIME) and Long Term Monitoring (LTM) program—a collaborative program managed by the U.S. Environmental Protection Agency (USEPA)—was initiated in the 1980s to examine trends in surface-water chemistry in response to changing air emissions and acid deposition. The number of lakes and streams monitored by the integrated program has dropped

* Corresponding author.
E-mail addresses: lsprague@usgs.gov (L.A. Sprague), goelsner@usgs.gov (G.P. Oelsner), dmargue@usgs.gov (D.M. Argue).

over time due to funding cuts, leading to a loss of sites in Colorado, the Upper Midwest, and Vermont (U.S. Environmental Protection Agency, 2009). An evaluation of historical nutrient data collected by multiple organizations that reported data in the USGS National Water Information System (NWIS) database and the USEPA Storage and Retrieval (STORET) database found that the number of monitoring sites nationwide having at least 2 years of nutrient data and at least 20 samples increased throughout the 1970s, but widespread decreases were occurring by the early- to mid-1990s. The timing of the peak number of sites ranged from the mid- to late-1970s (in the Midwest and Pacific Northwest) to the early- to mid-1990s (in the South) (Saad et al., 2011).

As the collective resources for water-quality monitoring in the United States have decreased, it has become increasingly critical to leverage information from the disparate monitoring networks to address regional and national water-resource issues. Major monitoring organizations have made substantial progress in making their data publicly available (Myers, 2015), but fragmented data storage practices (including nomenclature, use of metadata, and data storage and dissemination platforms) continue to pose challenges when combining data from multiple organizations. Individual monitoring organizations understand their own data very well and are able to make use of those data locally. But problems can arise when their data are combined with data from one or more other organizations using a different dissemination platform and(or) using a different approach to describe the same common metadata elements.

In this study, we surveyed available nutrient data and metadata from 488 organizations that have conducted monitoring in U.S. streams at any point since 1899 (Supplementary Table 1). The survey included ambient monitoring data that were publicly accessible from Federal, State, regional, and local government agencies and non-governmental organizations, and included more than 25 million water-quality records from 321,927 sites. Nutrients are the focus of this survey because they were recently found to be the most widespread chemical stressor in U.S. streams—46% of U.S. streams have high levels of nitrogen, phosphorus, or both (U.S. Environmental Protection Agency, 2016a). Nutrients also are among the most commonly monitored water-quality parameters in the United States, and as a result, they provide a window into common challenges encountered for a wide variety of parameters when using data from multiple sources. The ultimate objective of this study was to identify major inconsistencies in metadata practices that limit the secondary use of multi-source nutrient data. The term "secondary use" is defined herein as distinct from the term "primary use". "Primary use" refers to use of data for the original intent determined by the organization that collected the data. "Secondary use" refers to the use of the same data for other purposes. Data users making secondary use of the data are defined herein as "secondary data users".

2. Material and methods

All available nutrient concentration records for streams and rivers were retrieved from NWIS and STORET, the two largest national water-quality databases in the United States. STORET serves water-quality data from the USEPA and multiple submitting organizations; NWIS primarily serves data from the USGS, but also contains a small amount of data collected by other organizations. Data were retrieved from NWIS in May 2013 and from STORET in October 2013; more recent data from both data bases were obtained from the Water-Quality Portal (which serves data from NWIS and STORET) between January and March 2015. Additional nutrient concentration records from major water-resource agencies in each State also were included in the evaluation. These data were

available from local agency databases or directly from staff within the agency. These data were obtained between June 2010 and May 2015; some organizations provided data on more than one date. Overall, approximately 70% of the final nutrient records came from NWIS and STORET.

The 488 organizations across the United States included in the survey comprised 19 Federal agencies; 6 regional (multi-State) organizations; 100 State water, natural resources, or environmental protection agencies; 130 tribal organizations; 108 county or subcounty organizations; 24 academic organizations; 17 nongovernmental organizations; 34 volunteer organizations; and 50 private organizations (Table 1). One-half of the data came from State water, natural resources, or environmental protection agencies; another one-third of the data came from Federal agencies.

Variations in nomenclature, terminology, and jargon in the water-resources community have long been problematic when aggregating water-quality data from multiple sources. In particular with nutrients, there are different chemical and physical forms in natural waters (notably, partitioning between organic and inorganic forms and dissolved and particulate phases) and different methods of field collection and laboratory analysis. Defined standards and universally accepted nomenclature for nutrients in their various forms have been unavailable or have not been widely adopted. Ambiguous and(or) inconsistent terminology, ambiguous and(or) incomplete metadata, and clearly incorrect data and metadata were all encountered in our survey. To evaluate these issues, we identified the key result-level metadata elements needed to unambiguously interpret each value. Result-level metadata applies to an individual numerical value from a sample, and includes metadata elements such as units of measurement. Often, there are multiple results in a sample (for example, nitrate, ammonia, and orthophosphate results together in a single sample). Sample-level metadata applies to all results in the sample, and includes elements such as date, time, site name, site location, and sampler type. This study evaluates only result-level metadata.

The key result-level metadata elements in our evaluation included parameter name, sample fraction (filtration status), chemical form (molecular or elemental), numerical value of the analysis, units of measurement, and remark codes indicating either poor quality or values detected below the laboratory reporting limit. The number of records that could be unambiguously interpreted were tallied, along with the number of records that had missing or ambiguous information for each of the key metadata elements.

3. Results and discussion

3.1. Records with missing or ambiguous metadata

Nearly 14.5 million of the original 25 million records surveyed—or over one half of the original records—had missing or ambiguous information for one or more of the key metadata elements. To make use of any of these 14.5 million records, a secondary data user would need to make some assumptions about the value.

3.1.1. Parameter name

Of the many different nutrient parameters reported by organizations throughout the United States, ten are most commonly used to characterize nutrient concentrations in streams and are most frequently reported in water-quality databases. These common parameters included ammonia, Kjeldahl nitrogen (ammonia and organic nitrogen), nitrite, nitrate, nitrite plus nitrate, nitrogen (mixed forms, including nitrite, nitrate, ammonia, and organic nitrogen), organic nitrogen, organic phosphorus, orthophosphate,

Table 1
The number of individual organizations surveyed and the percent of surveyed data associated with each major organizational group.

Organizational group	Number of individual organizations	Percent of surveyed data
Federal	19	33
Regional	6	0.71
State	100	50
Tribal	130	0.77
County or subcounty	108	11
Academic	24	2.7
Non-governmental organization	17	0.13
Volunteer	34	1.5
Private	50	0.27

and phosphorus (mixed forms, including orthophosphate, poly-phosphates, and organic phosphorus). Nitrogen (mixed forms, including nitrite, nitrate, ammonia, and organic nitrogen) and phosphorus (mixed forms, including orthophosphate, poly-phosphates, and organic phosphorus) are herein referred to as nitrogen (mixed forms) and phosphorus (mixed forms), respectively. These parameters often are reported as "total nitrogen" and "total phosphorus", names that are ambiguous for reasons we discuss below.

For these ten nutrient parameters, there were 1046 original parameter names in the surveyed data. Of those, 931 could be unambiguously associated with one of the ten common parameters; 115 could not be unambiguously associated. Because of wide variations in nomenclature and the use of abbreviations, this association had to be determined manually. For example, Table 2 shows a subset of the 130 original parameter names that were possible variations on "nitrite plus nitrate". Many possible variations could be clearly identified as "nitrite plus nitrate", as in examples 1 through 4; however, there are issues with the other examples. Example 4 clearly defines "inorganic nitrogen" as nitrite plus nitrate, whereas example 5 clearly defines "inorganic nitrogen" as ammonia plus nitrite plus nitrate. As a result, examples 6, 7, and 8 ("inorganic nitrogen" with no further detail) are rendered ambiguous. This is a clear illustration of a situation where the original data provider knows the exact definition of examples 6, 7, and 8, but a secondary data user will not, once the data from examples 4 through 8 are combined. Additionally, a single organization provided data for examples 7 and 8; the two parameter names are similar, but the existence of more than one name raises the possibility of a meaningful distinction that a secondary data user may not fully understand. Examples 9 and 10 are unclear because of their reference to oxidized nitrogen. While "nitrite plus nitrate" sometimes is referred to informally as "NOX"

Table 2
A subset of the 130 original parameter names that were possible variations on nitrite plus nitrate in the surveyed data.

Original parameter name	Example
Nitrate plus nitrite, water, filtered, field, milligrams per liter as nitrogen	1
NO2+3 (mg/L)	2
NITROGEN, NITRITE (NO2) + NITRATE (NO3),Dissolved	3
Inorganic nitrogen (nitrate and nitrite)	4
Inorganic nitrogen, water, dissolved, calculated as NH3+NO2+NO3, milligrams per liter as nitrogen	5
Inorganic Nitrogen	6
Nitrogen,Inorganic||Nitrogen,inorganic,total(ug/LasN)	7
Nitrogen,Inorganic||Nitrogen,inorganicasN	8
Total NOX mg/L	9
Nitrogen, oxidized	10

or "oxidized nitrogen," oxidized nitrogen technically can refer to different combinations of oxygen and nitrogen, or a mixture of such compounds, including nitrite (NO_2^-), nitrate (NO_3^-), or nitrite plus nitrate ($NO_2^- + NO_3^-$). This is another illustration of a situation where the original data provider knows the meaning of their parameter name, but a secondary data user would need more precise information. Of the 931 original parameter names that could be unambiguously associated with one of the ten nutrient parameters, orthophosphate had the greatest number of different names (147), followed by ammonia (141), and then phosphorus, mixed forms (119).

Two unique issues were identified with the parameter name harmonization for orthophosphate and phosphorus (mixed forms). The first issue was related to naming conventions and associated laboratory analyses of phosphates. For the purposes of this study, the parameters "soluble reactive phosphorus" (SRP), "orthophosphate", and "orthophosphate-plus-hydrolysable-phosphate" were associated with the name orthophosphate. Operationally, SRP and orthophosphate are equivalent when they both are determined using a spectrophotometric analysis that measures orthophosphate as well as a small amount of other polyphosphates that may be unavoidably hydrolyzed during the analysis (Jarvie et al., 2002). Orthophosphate may be different when it is determined using ion chromatography, because there is no inadvertent hydrolysis of other polyphosphates (Westland and Boisclair, 1974). But because the amount of hydrolysis during the spectrophotometric method typically is small, orthophosphate and SRP were considered to be equivalent for this study, regardless of analytical method. Operationally, orthophosphate-plus-hydrolyzable-phosphate is not the same as either SRP or orthophosphate, because it includes an additional deliberate hydrolysis step that can hydrolyze condensed polyphosphates (Jarvie et al., 2002). However, the amount of condensed polyphosphates typically is so low in natural waters that orthophosphate-plus-hydrolyzable-phosphate also was considered to be equivalent to SRP and orthophosphate for this study.

A second, more intractable, issue with the phosphorus parameters was that variations on the name "phosphate" (such as "phosphate", "total phosphate", or "phosphate-phosphorus") were used to represent either orthophosphate or phosphorus (mixed forms), depending on the source. This issue appears to stem from the way phosphorus (mixed forms) is analyzed in the laboratory. A water sample that is collected in the field from a stream for phosphorus (mixed forms) analysis can be analyzed in a laboratory by converting all forms of phosphorus in the sample to phosphate. As a result, some organizations report the resulting value as phosphorus (mixed forms) (based on the field perspective), but other organizations report the same value as phosphate or total phosphate (based on the laboratory perspective). In STORET, "phosphate-phosphorus" has been used to represent phosphorus (mixed forms) (U.S. Environmental Protection Agency, 2006). Outside of STORET, some organizations appear to follow this convention and use a variation on the name phosphate to represent phosphorus (mixed forms). Other organizations use a variation on the name phosphate to represent orthophosphate or total phosphate. In many cases, the intent is not clear, rendering the data ambiguous. When the laboratory analytical method is available and documented (for example "4500-P F. Automated Ascorbic Acid Reduction Method" [Rice et al., 2012]), that information can be used to help resolve this ambiguity. When the analytical method is not available or is available but not documented (for example, "Method 1"), then this is an issue that cannot be resolved by the data user.

Ultimately, the parameter name was ambiguous for 3,557,821 records, the third largest number of records affected among the critical metadata elements (Table 3). In addition, for those that could be unambiguously identified, substantial manual

Table 3

Count of nutrient records missing key metadata elements that are needed to unambiguously interpret the associated data value. Of the 25,125,379 original records, 14,453,492 had missing or ambiguous information for one or more of these key metadata elements.

Parameter name	Starting records	Affected records									
		Filtration status		Chemical form		Units of measurement		Remark codes		Data entry	
		Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Ammonia	4,305,527	2,473,751	57	1,074,957	25	456,060	11	22,304	0.52	223,668	5.2
Nitrate	1,898,557	272,968	14	662,006	35	72,478	3.8	9193	0.48	67,949	3.6
Nitrite	1,527,346	186,401	12	302,840	20	99,957	6.5	14,693	0.96	85,028	5.6
Nitrite plus nitrate	3,670,035	2,114,854	58	–	–	217,762	5.9	5945	0.16	55,124	1.5
Kjeldahl nitrogen	3,271,083	2,206,324	67	–	–	93,499	2.9	9223	0.28	61,051	1.9
Organic nitrogen	501,153	113,003	23	–	–	3035	0.61	162	0.032	9901	2.0
Nitrogen mixed forms	916,604	228,971	25	–	–	5937	0.65	753	0.082	3070	0.3
Organic phosphorus	53,902	24,472	45	–	–	74	0.14	689	1.3	3582	6.6
Orthophosphate	2,834,848	482,676	17	754,064	27	125,153	4.4	12,848	0.45	63,428	2.2
Phosphorus mixed forms	2,588,503	905,392	35	–	–	15,689	0.61	1810	0.070	8321	0.32
Ambiguous	3,557,821	2,937,643	83	1,471,748	41	221,452	6.2	46,903	1.3	55,332	1.6
Totals	25,125,379	11,946,455	–	4,265,615	–	1,311,096	–	124,523	–	636,454	–

[–, not applicable. Some records had missing or ambiguous information for more than one metadata element and are counted in more than one column. Therefore, the sum of the totals from each metadata column will be greater than 14,453,492.].

intervention and time was required to harmonize the original parameter names.

3.1.2. Filtration status

Filtration is the physical process used to separate the particulate and aqueous fractions of a water sample. Samples are filtered for several purposes. For example, samples may be filtered to identify the amount of a chemical associated with the particulate versus the aqueous fraction, to remove microorganisms in order to help preserve a sample before laboratory analysis, to remove suspended materials that interfere with some laboratory analyses, or to determine chemical speciation and fractionation of trace elements for geochemical studies (U.S. Geological Survey, variously dated). Filter pore sizes can vary; 0.2, 0.45, and 0.6 μm are common.

In a single stream sample, it often is possible to determine both unfiltered and filtered variations of the same analyte. Unfiltered results will include the amount of chemical associated with both the particulate and the aqueous fraction. Filtered results will include the amount of chemical associated with just the aqueous fraction. Knowledge of the filtration status of a sample is vital, because filtered and unfiltered values for the same analyte may be very different. This difference is illustrated by paired filtered and unfiltered values of Kjeldahl nitrogen, orthophosphate, and nitrate in Fig. 1. Paired values (filtered and unfiltered values of the same analyte from the same stream sample) indicated substantial differences and biases between filtered and unfiltered values of Kjeldahl nitrogen, orthophosphate, and nitrate in several rivers of the United States (Fig. 1a−c). Filtered and unfiltered values are not always different, however. Little difference was observed between paired filtered and unfiltered values of nitrate in another river (Fig. 1d). Because it can be difficult to accurately determine the propensity of nutrients to sorb to the particulate fraction in-river, a data user cannot universally assume that filtered and unfiltered values will be comparable enough to safely combine into a single data series. Similarly, a data user cannot safely disregard the filtration status of a sample when no filtration status is reported.

Water samples can be filtered in the field at the time of collection or in the laboratory prior to analysis. For some analytes, the same laboratory method can be used on both a filtered and an unfiltered sample. As a result, a description of the laboratory method (even if available) is not always a reliable determinant of filtration status.

In this study, both "filtered" and "dissolved" were assumed to mean filtered. "Unfiltered" and "wholewater" were assumed to mean unfiltered. "Particulate" or "nonfilterable" were assumed to mean particulate. On many occasions, these variations appear to have been reported in abbreviated form (for example, "F", "D", "U", "W", "WW", "P", etc.). Because the meaning of these abbreviations will not be clear to all data users—particularly those who are not already aware that filtration status can be an important determinant of a nutrient concentration value—the use of these abbreviations was considered to be ambiguous. If the parameter name contained the word "particulate" or "nonfilterable", and the fraction was reported as "dissolved" or "suspended" (or vice versa), then the fraction and parameter name were contradictory and the data values were considered to be ambiguous.

Another widespread source of ambiguity in both filtration status and parameter names was the use of the term "total", without supporting metadata specifying whether a sample was filtered or unfiltered. The ambiguity arises because "total" was used in two ways in the surveyed data: (1) to represent the inclusion of multiple species, such as when ammonia and organic nitrogen were summed to give total Kjeldahl nitrogen, or when all dissolved phosphorus species were included in the determination of total dissolved phosphorus; and(or) (2) to represent an unfiltered sample. Some values were clearly identified as being both summed and unfiltered (e.g. "total Kjeldahl nitrogen, unfiltered" or "total phosphorus, unfiltered"); others were clearly identified as being summed and filtered (e.g. "dissolved Kjeldahl nitrogen" or "total phosphorus, filtered"); others could only have been unfiltered and not summed, because they could not have been derived as a sum (e.g. "total nitrite"). However, many records were ambiguously identified (e.g. "total Kjeldahl nitrogen" or "total phosphorus"). As with all of the key metadata elements, these parameter names likely are unambiguous to the originating organization. But they become ambiguous when the data are combined by a data user with data from another organization using different nomenclature.

Depending on the eventual application, a data user might decide that for some parameters at some sites—such as nitrite plus nitrate data at the site shown in Fig. 1d—any differences between filtered and unfiltered data are so small and unbiased as to be negligible and therefore all data designated filtered, unfiltered, or unspecified could safely be combined. To support such decision making, however, it would be preferable to have filtration status available in the metadata. For other parameters—such as phosphorus (mixed
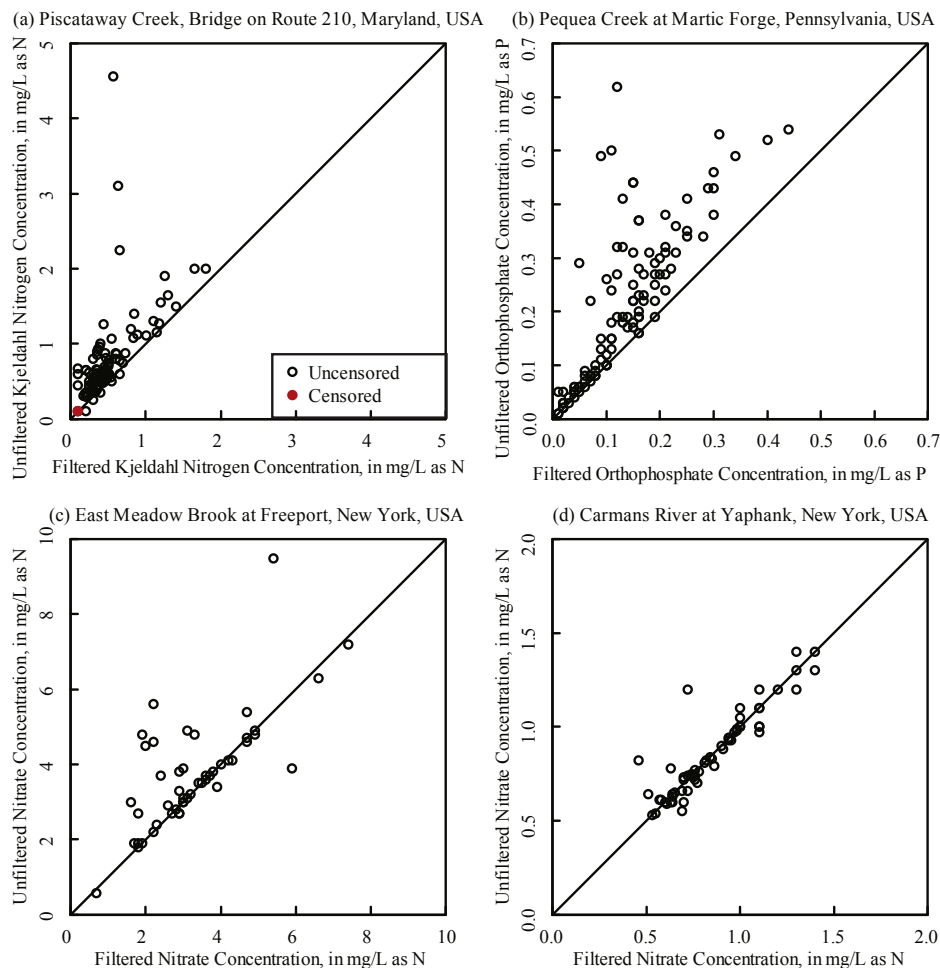
Fig. 1. Comparison of paired filtered and unfiltered nutrient concentrations in selected rivers of the United States.

forms) or Kjeldahl nitrogen—some data users may choose to assume the value represents an unfiltered sample based on an original parameter name provided by the source as "total phosphorus" or "total Kjeldahl nitrogen". The user assumes some risk with this practice, however, because of the aforementioned ambiguity in the use of the word "total" in parameter names.

Missing or ambiguous metadata for filtration status affected 11,946,455 records, by far the largest number of records affected among the critical metadata elements (Table 3). Excluding ambiguous parameter names, the parameters with the largest percentage of records affected by missing or ambiguous filtration status included Kjeldahl nitrogen (67%), nitrite plus nitrate (58%), ammonia (57%), organic phosphorus (45%), and phosphorus (mixed forms) (35%). Of the ten nutrient parameters, only three (nitrite, nitrate, and orthophosphate) had fewer than 20% of records affected by ambiguous reporting on filtration status. In large part, that was because these three parameters cannot be summed from multiple species and therefore it could safely be assumed that "total" meant unfiltered. For all other parameters, the use of the term "total" was a large contributor to the ambiguity in filtration status.

3.1.3. Chemical form

Some water-quality results can be reported in two chemical forms—molecular or elemental. Speciation is another, more imprecise, term sometimes used for chemical form. For example,

nitrate (NO_3^-) can be reported "as nitrate" (the molecular form, which includes the full set of nitrogen and oxygen elements in the nitrate molecule) or "as nitrogen" (the elemental form, with includes just the nitrogen element). The molar mass of oxygen and nitrogen are 15.9994 and 14.0067 g/mol, respectively; therefore the molar mass of nitrate is $(15.9994 \times 3) + 14.0067$, or 62.0049 g/mol. Nitrogen comprises only 22.5% of the weight of the nitrate molecule, and as a result, nitrate "as nitrogen" is 22.5% of nitrate "as nitrate" for the same amount of nitrate in a sample. A case in point: 45 mg/L of nitrate "as nitrate" is equivalent to 45 mg/L × 0.225, or 10 mg/L, of nitrate "as nitrogen". This is why the USEPA maximum contaminant level for nitrate is expressed either as 10 mg/L of nitrate "as nitrogen" or 45 mg/L of nitrate "as nitrate".

Because of the magnitude of this conversion factor, unambiguous and complete reporting of chemical form is critical to avoid misinterpretation of data. This is illustrated by nitrate concentrations in White Clay Creek, Pennsylvania, between 2002 and 2015 shown in Fig. 2a. The filled and open circles show the same data reported as two different chemical forms—nitrate in mg/L as nitrate (filled circles) and nitrate in mg/L as nitrogen (open circles). To make the conversion, each individual nitrate "as nitrate" concentration was multiplied by 0.225 to convert it to nitrate "as nitrogen". Both data series are compared to the USEPA maximum contaminant level of 10 mg/L of nitrate "as nitrogen", shown as a horizontal solid line in Fig. 2a. If a data user was to incorrectly assume the nitrate "as nitrate" concentrations were nitrate "as nitrogen"
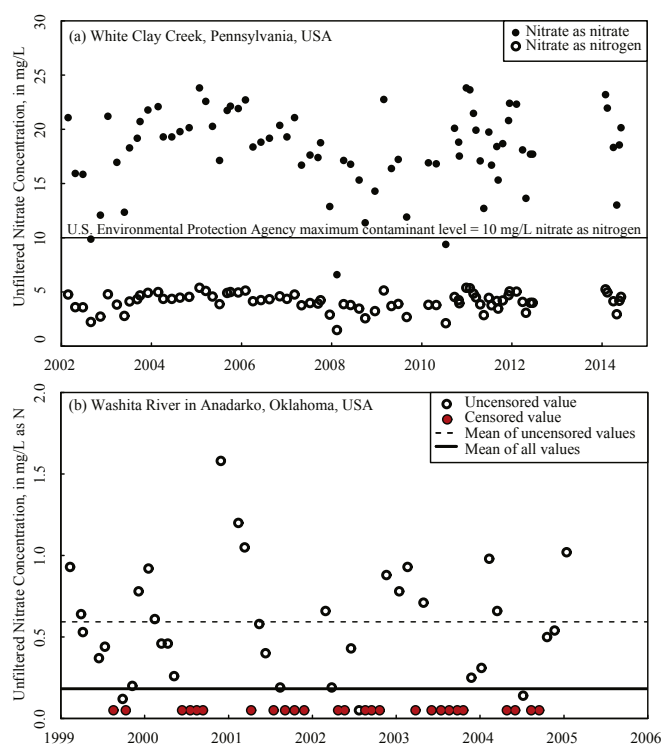
Fig. 2. (a) Comparison of nitrate concentrations in White Clay Creek, Pennsylvania, USA, with the same nitrate observations reported with different chemical forms (nitrate as nitrate and nitrate as nitrogen). (b) Comparison of mean nitrate concentrations in Washita River, Oklahoma, USA, with and without the inclusion of censored values.

concentrations, they would reach an incorrect conclusion that most nitrate concentrations in White Clay Creek exceeded the USEPA maximum contaminant level.

Similar reporting options are available for other nutrient parameters, including ammonia (NH_3), nitrite (NO_2^-), and orthophosphate (PO_4^{3-}). For these parameters, data values without a reported chemical form are ambiguous. Conversion factors between the two chemical forms are substantial and vary for each of these parameters (1.22, 3.28, and 3.06, respectively). As a result, misinterpretation of chemical form by a data user will result in using data that are incorrect by a factor of 1.22, 3.28, and 3.06, respectively. Because of the magnitude of these conversion factors, unambiguous reporting of chemical form is critical to avoid serious misinterpretation and misuse of data. For other parameters—including nitrite plus nitrate, nitrogen (mixed forms), organic nitrogen, organic phosphorus, phosphorus (mixed forms), and Kjeldahl nitrogen—chemical form can safely be assumed even when it is not reported, because there is only one common element in the molecule. As a result, we did not include counts due to missing chemical form for these parameters, even when chemical form was not reported. We did include counts due to missing chemical form for ambiguous parameter names, but some of those are likely to be nitrite plus nitrate, nitrogen (mixed forms), organic nitrogen, organic phosphorus, phosphorus (mixed forms), and Kjeldahl nitrogen, if their true meaning were clear.

Missing or ambiguous metadata for chemical form affected 4,265,615 records, the second largest number of records affected among the critical metadata elements (Table 3). Excluding ambiguous parameter names, the relevant parameters with the largest percentage of records affected by missing or ambiguous chemical form included nitrate (35%), orthophosphate (27%), and ammonia (25%).

3.1.4. Units of measurement

Many data values were reported without units of measurement, such that it could not be unambiguously determined whether concentrations were in mg/L, μg/L, or some other unit. Other values were reported with units that were clearly inappropriate for the associated parameter, such as nephelometric turbidity units (NTU) associated with a parameter like ammonia. Data without units or with inappropriate units are ambiguous.

There also were occasional values that were reported with units but that appeared to be typographical errors—for example, a phosphorus (mixed forms) concentration of 500 mg/L is extremely unlikely to occur in natural waters. It is far more likely that the units were reported incorrectly and that the true value is 500 μg/L, or 0.5 mg/L. In contrast, a phosphorus (mixed forms) concentration of 50 mg/L is more ambiguous, because while the concentration is very high relative to concentrations found in most natural waters, it could occur during locally extreme conditions such as an overflow of untreated sewage or a heavy runoff event immediately after nearby fertilizer application. The inclusion of proper units with sample metadata would help unambiguously identify data values and distinguish true outliers from typographical errors.

In some cases, an organization reported units for uncensored data but not for censored data (see "Remark codes" section below for more information on remark codes and censoring). The reasons for this practice are unknown, but it creates the potential for substantially biased analyses if only the censored data are excluded because the units are unknown.

Missing or ambiguous metadata for units affected 1,311,096 records, the fourth largest number of records affected among the critical metadata elements (Table 3). Excluding ambiguous parameter names, the parameters with the largest percentage of records affected by missing or ambiguous units included ammonia (11%), nitrite (6.5%), nitrite plus nitrate (5.9%), and orthophosphate (4.4%).

3.1.5. Remark codes

With any data value, there can be descriptive qualifiers critical to accurately interpreting the value. Of most interest for water-quality analyses are indications of poor data quality (for example, "Contamination present in the sample") and laboratory censoring (data values below a laboratory detection limit). In some databases, such as NWIS, these characteristics are stored as "remark codes" or "value qualifier codes". In other databases, such as STORET, these characteristics are stored as "qualifier codes" or "result detection condition". We refer to these characteristics collectively herein as "remark codes".

There were 587 unique remark codes assigned to nutrient values in the surveyed data. Many of these remark codes were not defined in an accessible location, and many others were ambiguously defined. When appropriate remark codes are available, data with one or more remark codes indicating poor data quality can safely be excluded, depending on the project objectives. Data with a remark code indicating laboratory censoring need to be retained and handled appropriately in subsequent analyses. The omission of censored data or the omission of remark codes indicating the data are censored (in effect substituting the detection limit for a value that actually falls somewhere between zero and the detection limit) can lead to substantial bias in data analyses (Helsel, 2005). This is illustrated by nitrate concentrations in the Washita River in Oklahoma between 1999 and 2006 shown in Fig. 2b. The mean nitrate concentration determined using all values is 0.182 mg/L as N (the solid line in Fig. 2b). If censored values were excluded from the data set, the lowest concentrations observed in the river would be discounted and the apparent mean would be much higher at 0.593 mg/L—a clearly biased result (the dashed line in Fig. 2b). Note that

calculation of the mean with all values (including censored values) was performed using maximum likelihood estimation (Helsel, 2005).

There were 63 unique remark codes that indicated laboratory censoring; these remark codes had to be manually identified. For some censored data, there was no remark code provided; rather, a comment field associated with the data value indicated that the value was censored. These comments had to be manually located and translated into a harmonized remark code, a step that some data users may not be aware is necessary. As described above, some sources provided units of measurement for uncensored data but not for censored data. The reasons for this practice are unknown, but it creates the potential for substantially biased analyses if only the censored data are excluded because of missing units of measurement.

Of the 488 sources surveyed, 118 did not provide any records containing censored remark codes. The total number of records from each of these 118 sources ranged from 1 to 396,002. For those sources with a small number of records, it is plausible that no censored data were ever collected. For those sources with a larger number of records, it is unlikely that no censored data were ever reported by the laboratory, given the prevalence of very low environmental concentrations for some of the nutrient parameters. When contacted, a number of these sources confirmed that they did in fact have censored data, but deliberately did not upload any data with remark codes to STORET and(or) did not make any data with remark codes otherwise publicly available through their own databases. The reasons for this practice are unknown, but it creates the potential for substantially biased analyses when only the remaining uncensored data are used. Further, this practice will not be transparent to data users. Data users will be unaware of any resulting bias in their analyses and may make crucial policy or management decisions on the basis of these biased analytical results.

Missing or ambiguous metadata for remark codes affected 124,523 records, the smallest number of records affected among the critical metadata elements (Table 3). Excluding ambiguous parameter names, the only parameter with more than 1% of records affected by missing or ambiguous remark codes was organic phosphorus (1.3%). Importantly, however, the totals in Table 3 do not include the unknown number of censored records that were deliberately withheld by some data providers. The number of records affected is likely much larger than reported here.

3.1.6. Zero, negative, and censored missing values

Among the surveyed data, 636,454 values were reported as zero, negative, or "censored" with no reported value (e.g., a censored remark code was given without an associated laboratory detection limit). An analytical determination of zero or negative concentration is not possible with nutrients, so any zero or negative values should be considered suspect. With the "censored" missing values, it is unclear if the remark code entry is in error and no analysis was ever performed, or if an analysis was performed and the resulting value was censored but no detection limit was reported. The true values are unknown in all of these situations. In some cases, zero or negative values also were reported with a remark code indicating they were censored, although no laboratory detection limit was provided. In other cases, no remark code was reported so it is unknown whether those zero and negative values also were intended to represent censored values. As a result, a data user cannot safely make a general assumption about any of these values.

Data entry issues involving the reporting of values as either zero, negative, or missing with a remark code affected 636,454 records, the fifth largest number of records affected among the critical metadata elements (Table 3). Excluding ambiguous parameter

names, the parameters with the largest percentage of records affected by these data entry issues included organic phosphorus (6.6%), nitrite (5.6%), and ammonia (5.2%).

3.1.7. Increasing data usability

This evaluation identified major inconsistencies in result-level metadata that limit the utility of multi-source data. Data usability could be increased through the adoption of several standardized metadata practices. First, consistent use of parameter naming conventions—such as those curated by the Open Geospatial Consortium Naming Authority (Larsen et al., 2016)—could help with identification and merging of data from multiple sources using a shared vocabulary. Second, restricting the use of the term "total" to instances where the parameter includes the sum of multiple species and using the terms "filtered", "unfiltered", and "particulate" to separately describe filtration status would avoid ambiguity in both parameter names and filtration status related to the use of the term "total". Third, complete and consistent reporting of filtration status, chemical form, and units of measurement would avoid the potential for serious misinterpretation of data values. Fourth, adoption of consistently defined remark codes (such as those curated by the Open Geospatial Consortium Naming Authority), encoding remark codes with a data value rather than using text narratives, and full reporting of all censored data could help ensure proper handling of censored and poor-quality data in subsequent data analyses. Finally, discontinuing the improper use of zero values, negative values, or missing values with a remark code could help ensure secondary data users are making full and appropriate use of all data.

3.1.8. Additional issues

Parameter name, filtration status, chemical form, units of measurement, remark codes, and accurate data entry are all metadata elements needed to unambiguously quantify a numerical data value at the result level. There are higher-level metadata elements needed to identify other important aspects of a value at the sample level. While we do not delve into these elements in detail here, they are important to keep in mind when storing, serving, and using multi-source water-quality data. These additional sample-level metadata elements include a site name that does not change over time and provides the stream or river name together with geographic references, an accurate latitude and longitude for the site, the date and time of sample collection, the method used to collect the sample from the stream (for example, a composite sample versus a grab sample), the analytical method used by the laboratory, and a flag for whether the data value is stored in another public database. The National Water-Quality Monitoring Council (NWQMC) previously has made detailed recommendations for these sample-level metadata elements (National Water-Quality Monitoring Council, 2006).

The NWQMC recommendations are one example of the growing awareness of the importance of metadata. In 2013, the U.S. Office of Management and Budget issued a new directive requiring Federal agencies to disseminate data used in scientific research in a way that supports reproducible research and future scientific endeavors, by using machine readable and open formats, data standards, and common metadata standards (U.S. Office of Management and Budget, 2013). However, this directive stops short of providing details on the critical metadata elements for specific types of data, saying only that the metadata must be sufficient for data users to understand the strengths, weaknesses, and analytical limitations of the data. In 2012, the USGS, the USEPA, and the NWQMC began sponsoring an on-line cooperative Water-Quality Portal for downloading water-quality data that are automatically linked and aggregated from NWIS, STORET, and the U.S.

Department of Agriculture's "Sustaining the Earth's Watersheds - Agricultural Research Database System" (STEWARDS) (National Water-Quality Monitoring Council, 2016). The Water-Quality Portal produces data formatted according to the Water Quality Exchange (WQX) Outbound XML schema, which has been developed collaboratively by the USGS and the USEPA. The WQX schema is a major advancement in providing a standard set of metadata elements that data providers can use to share discrete data in the Water-Quality Portal (Larsen et al., 2016; U.S. Environmental Protection Agency, 2016b). However, the WQX schema does not currently provide a complete framework for the metadata elements described in this study, and it is not used exclusively by all monitoring organizations in the United States. Until WQX or another data schema is fully developed and adopted by all monitoring organizations, the data in the Water-Quality Portal (and by extension, the data in NWIS, STORET, and STEWARDS) can still be rendered ambiguous when they are combined by secondary data users with data from other databases using a different data schema. Clearly there is still a need for standardization of detailed metadata specific to water-quality data.

3.2. The cost of missing or ambiguous metadata

Previous studies have estimated the cost to collect a single stream-quality sample, including the following items: salary, travel, supplies, equipment, laboratory analysis, administrative support, database support, and quality control and quality assurance management costs. While this list is more inclusive than just the costs for collection and laboratory analysis of a sample, none of these other items would be necessary if samples were not collected and analyzed. As such, these items are a necessary inclusion in the total cost estimate. Detailed cost estimates are not readily available from most of the monitoring organizations in this study, but limited published estimates indicate a range in sampling costs. Horowitz (2013) estimated that the total cost (including the supporting activities noted above) ranged between US$4000 and $6000 per sample in 2013, or between US$4108 and $6162 in 2016, accounting for inflation (U.S. Department of Labor, 2016). Betanzo et al. (2015) estimated that the cost ranged between US$2900 and $5900 in 2012, or between US$3022 and $6148 in 2016, accounting for inflation (U.S. Department of Labor, 2016). The range in costs reflects factors such as differences in distances required to travel between an office and a stream monitoring site, as well as differences in stream size and the requisite sampling time, equipment, and sampling methods (Betanzo et al., 2015). Third. Herrera Environmental Consultants and Aspect Consulting (2010) estimated that the cost was US$1968 in 2010; or US$2179 in 2016, accounting for inflation. To reflect the range in sampling costs, we used a minimum estimate of US$2,179, a maximum estimate of US$6,162, and an average estimate of US$3788 (based on the average of US$2,179, $3,022, and $6162).

The economic value of legacy nutrient data affected by missing or ambiguous metadata can be estimated by multiplying 3,928,774 (the number of unique samples with one or more affected records; note multiple affected records can be present in a single sample) by $US3,788 (the current average cost of collecting a stream-quality sample, as described above). We estimated that approximately 20% of the surveyed data were duplicated in multiple databases and reduced the number of unique affected samples accordingly to 3,143,019. The final calculation of the average economic value of affected samples was $US12 billion. Using the minimum and maximum costs of collecting a stream-quality sample, the respective range in the economic value of affected samples was $US6.8 billion to $US19 billion.

It is important to recognize the value of legacy nutrient data

unaffected by missing or ambiguous metadata. There were 2,706,136 unique samples with one or more unaffected records, which translated into an average economic value of $US8.2 billion (range $US4.7 billion to $US13 billion). Because some samples had both ambiguous and unambiguous records (for example, nitrate is unambiguous but orthophosphate is ambiguous in the same sample), some samples are included in both the unambiguous and ambiguous total.

Because a secondary data user would need to make unverifiable and possibly incorrect assumptions about the data values affected by missing or ambiguous metadata in order to use the data, the economic value of $US12 billion represents a substantial potential loss from a collective investment by water-resource organizations in the United States. This loss is only realized, however, when data from multiple sources are combined by a secondary data user for uses beyond their original intent, so the loss can in theory be avoided. Individual organizations are familiar with the metadata elements associated with their own data, but they are lacking standardized and commonly accepted metadata practices that they could employ to increase the usability of their data. If standardized metadata practices were adopted nationwide, theoretically most of the missing or ambiguous metadata could be addressed by the data providers. Modifications made by the original data providers would be accurate and could be carried over to multiple public data platforms serving those data.

While the economic value of the affected data is enormous, the costs to address metadata issues in legacy data and implement standardized metadata practices for future data are not trivial. In addition, recent and future metadata issues may be easier to address than historical issues; the records or documentation needed to address metadata issues in older data may no longer exist with the original data providers. Recently, however, the NWQMC has suggested that updating legacy water-quality data records to include minimum metadata elements would provide utility to data records beyond their original intent, thereby justifying the costs required to make the updates (National Water-Quality Monitoring Council, 2006). Larsen et al. (2016) recommend that these minimum metadata standards be high enough to maintain data integrity, but not so high as to eliminate participation by data providers.

One of the biggest challenges moving forward appears to be finding funding sources for such an effort. A reality of reduced funding for water-quality monitoring is that monitoring organizations have more narrowly defined objectives and outlets for communicating findings to help offset the reductions in funding. In many cases, this may lead monitoring organizations to prioritize maintenance of long-term monitoring networks over enhancements to data storage, interoperability, and access for secondary users.

The potential scientific value of multi-source water-quality data to secondary data users is evident when comparing the number of unambiguous nutrient records available from just the USGS NWIS database to those available from the 488 sources surveyed in this study (Fig. 3). Combining the unambiguous nutrient records from all sources makes available an additional 4 million records nationwide. The potential scientific value of all legacy multi-source water-quality data—which could only be realized for secondary data users by updating historical water-quality data records to include minimum metadata elements—is evident when comparing the total number of unambiguous nutrient records to the total number of nutrient records (unambiguous plus ambiguous) (Fig. 3). Combining all of the unambiguous and ambiguous nutrient records makes available an additional 14.5 million records nationwide.

The scientific value of multi-source water-quality data also has been established by the many previous studies made possible by
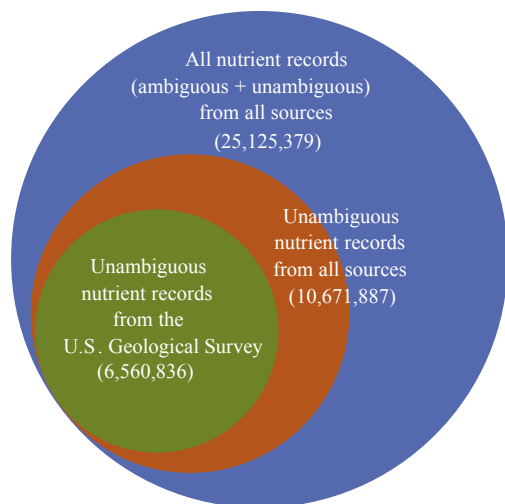
Fig. 3. Total number of nutrient records (unambiguous plus ambiguous) from all sources, number of unambiguous nutrient records from all sources, and number of unambiguous nutrient records from the U.S. Geological Survey.

leveraging data from different sources and sampling programs. These studies have addressed important societal issues, including changes in stream quality, the effectiveness of pollution control investments, threats to the health of humans and aquatic life, and the economic impacts of pollution. For example, Stets et al. (2015) used data from numerous Federal, State, and local water-quality databases and published reports to link increases in nitrate concentrations in 22 U.S. rivers between 1945 and 1980 to the increased use of fertilizer in agricultural production. Dodds et al. (2009) used nitrogen and phosphorus data from multiple Federal, State, and local organizations to document US$2.2 billion annual value losses in recreational water usage, waterfront real estate, spending on recovery of threatened and endangered species, and drinking water treatment as a result of anthropogenic eutrophication of freshwater systems throughout the United States. Monteith et al. (2007) used data from multiple sources to show that increasing trends in dissolved organic carbon in streams of North America and Europe between 1990 and 2004 resulted from reductions in sulfur emissions and the severity of coastal storms. The USEPA and other organizations have developed a chemical screening and prioritization program to catalog available toxicology information for hundreds of thousands of chemicals from over 200 public sources (Judson et al., 2009). Vidic et al. (2013) used water-quality data from five Federal, State, and non-governmental databases to show that concentrations of barium, strontium, and bromide were elevated in Pennsylvania rivers in areas of known brine effluents from centralized waste treatment plants that may have been receiving flowback and produced water from Marcellus Shale gas development. These examples just scratch the surface of our collective reliance on water-quality data from multiple sources to address important environmental issues; numerous similar studies have been published over the years.

4. Conclusions

Substantial investments have been made in water-quality monitoring in the United States by Federal, State, and local governments, Tribes, water utilities, universities, and many others since the early 20th century. Data from these monitoring efforts are invaluable for documenting the current status and long-term trends in the Nation's river quality. Combining water-quality data

from multiple sources can help counterbalance diminishing resources for stream monitoring in the United States and lead to important regional and national insights that would not otherwise be possible. Inconsistencies in metadata practices, however, pose substantial challenges when these data are used beyond their original intent by secondary data users. Harmonization of multi-source water-quality data can be time consuming, expensive, and inexact, requiring manual intervention and subjective professional judgement. The costs incurred to defensibly merge multi-source data can be prohibitive for many secondary data users. Multiple secondary data users end up working with and merging the same data, and collectively, such costs represent a large redundant expense that could be better directed toward developing and implementing standardized metadata practices across organizations. In addition, ambiguities in the metadata could lead secondary data users to assume different characteristics for the same data, potentially resulting in different conclusions about important policy or management issues. The value of these ambiguous data is estimated at $US12 billion, a substantial collective investment by water-resource organizations in the United States. Those data could be preserved for uses beyond the original intent by developing and implementing standardized metadata practices for future and legacy water-quality data in the United States.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.watres.2016.12.024.

References

Betanzo, E.A., Choquette, A.F., Reckhow, K.H., Hayes, L., Hagen, E.R., Argue, D.M., Cangelosi, A.A., 2015. Water Data to Answer Urgent Water Policy Questions: Monitoring Design, Available Data and Filling Data Gaps for Determining the Effectiveness of Agricultural Management Practices for Reducing Tributary Nutrient Loads to Lake Erie, Northeast-midwest Institute Report, p. 169. http://dx.doi.org/10.13140/RG.2.1.1102.5684.

Dodds, W.K., Bouska, W.W., Eitzmann, J.L., Pilger, T.J., Pitts, K.L., Riley, A.J., Schloesser, J.T., Thornbrugh, D.J., 2009. Eutrophication of U.S. freshwaters: analysis of potential economic damages. Environ. Sci. Technol. 43, 12—19. http://dx.doi.org/10.1021/es801217q.

Helsel, D.R., 2005. Nondetects and Data Analysis: Statistics for Censored Environmental Data. John Wiley and Sons, New York, p. 268. ISBN: 0471671738.

Herrera Environmental Consultants and Aspect Consulting, 2010. River and Stream Water Quality Monitoring Plan for the Hood Canal Watershed, Prepared for the Washington Department of Ecology WRIA 16/14b Planning Unit, p. 68. http://www.ecy.wa.gov/programs/eap/wrias/planning/docs/09-04512-000_River_and_Stream_WQ_Monitoring_Plan_2010_07_27.pdf.

Hooper, R.P., Aulenbach, B.T., Kelly, V.J., 2001. The National Stream Quality Accounting Network: a flux-based approach to monitoring the water quality of large rivers. Hydrol. Process. 15, 1089—1106. http://dx.doi.org/10.1002/hyp.205.

Horowitz, A.J., 2013. A review of selected inorganic surface water quality-monitoring practices: are we really measuring what we think, and if so, are we doing it right? Environ. Sci. Technol. 47, 2471—2486. http://dx.doi.org/10.1021/es304058q.

Jarvie, H.P., Withers, P.J.A., Neal, C., 2002. Review of robust measurement of phosphorus in river water: sampling, storage, fractionation, and sensitivity. Hydrol. Earth Syst. Sci. 6, 113—132. http://dx.doi.org/10.5194/hess-6-113-2002.

Judson, R., Richard, A., Dix, D.J., Houck, K., Martin, M., Kavlock, R., Dellarco, V., Henry, T., Holderman, T., Sayre, P., Tan, S., Carpenter, T., Smith, E., 2009. The toxicity data landscape for environmental chemicals. Environ. Health Perspect. 117, 685—695. http://dx.doi.org/10.1289/ehp.0800168.

Larsen, S., Hamilton, S., Lucido, J., Garner, B., Young, D., 2016. Supporting diverse data providers in the Open Water Data Initiative: communicating water data quality and fitness of use. J. Am. Water Resour. Assoc. 52 (4), 859—872. http://dx.doi.org/10.1111/1752-1688.12406.

Monteith, D.T., Stoddard, J.L., Evans, C.D., de Wit, H.A., Forsius, M., Høgåsen, T., Wilander, A., Skjelkvåle, B.L., Jeffries, D.S., Vuorenmaa, J., Keller, B., Kopácek, J., Vesely, J., 2007. Dissolved organic carbon trends resulting from changes in atmospheric deposition chemistry. Nature 450, 537—540. http://dx.doi.org/10.1038/nature06316.

Myers, D.N., 2015. Foundations of water quality monitoring and assessment in the United States. In: Ahuja, Satinder (Ed.), Food, Energy and Water—the Chemistry Connection. Elsevier, Amsterdam, pp. 21—79.

National Research Council, 2004. Confronting the Nation's Water Problems: the Role of Research. Committee on Assessment of Water Resources Research. National Academies Press, Washington, D.C., p. 324. http://dx.doi.org/10.17226/11031

National Water Quality Monitoring Council, 2016. Water-quality Portal. Accessed 23 August 2016. http://www.waterqualitydata.us/.

National Water Quality Monitoring Council, 2006. Data Elements for Reporting Water Quality Monitoring Results for Chemical, Biological, Toxicological, and Microbiological Analytes. NWQMC Technical Report No. 3, p. 55. http://acwi.gov/methods/pubs/wdqe_pubs/wqde_trno3.pdf.

Rice, E.W., Baird, R.B., Eaton, A.D., Clesceri, L.S. (Eds.), 2012. Standard Methods for the Examination of Water and Wastewater, twenty-second ed. American Public Health Association, American Water Works Association, Water Environment Federation, p. 1496. ISBN: 9780875530130.

Rowe, G.L., Belitz, K., Demas, C.R., Essaid, H.I., Gilliom, R.J., Hamilton, P.A., Hoos, A.B., Lee, C.J., Munn, M.D., Wolock, D.W., 2013. Design of Cycle 3 of the National Water-quality Assessment Program, 2013—23: Part 2: Science Plan for Improved Water-quality Information and Management. U.S. Geological Survey Open-file Report 2013—1160, p. 110. http://pubs.usgs.gov/of/2013/1160/.

Saad, D.A., Schwarz, G.E., Robertson, D.M., Booth, N.L., 2011. A multi-source nutrient dataset used to estimate loads, improve monitoring design, and calibrate regional nutrient SPARROW models. J. Am. Water Resour. Assoc. 47, 933—949. http://dx.doi.org/10.1111/j.1752-1688.2011.00575.x.

Stets, E.G., Kelly, V.J., Crawford, C.G., 2015. Regional and temporal differences in nitrate trends discerned from long-term water quality monitoring data. J. Am. Water Resour. Assoc. 51, 1394—1407. http://dx.doi.org/10.1111/1752-1688.12321

U.S. Department of Labor, 2016. Bureau of Labor Statistics, CPI Detailed Report—data for April 2016. Accessed on 6 June 2016. http://www.bls.gov/data/inflation_calculator.htm.

U.S. Environmental Protection Agency, 2016a. National Rivers and Streams Assessment 2008—2009—A Collaborative Survey. U.S. Environmental Protection Agency Office of Wetlands, Oceans and Watersheds Office of Research and Development Report EPA/841/R-16/007, p. 116. https://www.epa.gov/sites/production/files/2016-03/documents/nrsa_0809_march_2_final.pdf.

U.S. Environmental Protection Agency, 2016b. STOrage and RETrieval and Water Quality EXchange. Accessed on 23 August 2016. https://www.epa.gov/waterdata/storage-and-retrieval-and-water-quality-exchange.

U.S. Environmental Protection Agency, 2009. Evaluation of EPA's Temporally Integrated Monitoring of Ecosystems (TIME) and Long-term Monitoring (LTM) Programs. U.S. Environmental Protection Agency Office of Policy, Economics and Innovation Report, p. 97. https://www.epa.gov/sites/production/files/2015-09/documents/eval-epa-temporally-integrated-monitoring-ecosystems.pdf.

U.S. Environmental Protection Agency, 2006. SRS Names for Carbon, Nitrogen and Phosphorous Characteristics. ** Added October 23, 2006 **. Accessed on February 24, 2015. http://www.epa.gov/storet/wqx/wqx_downloads.html.

U.S. Office of Management and Budget, 2013. Memorandum for the Heads of Executive Departments and Agencies: Open Data Policy—managing Information as an Asset. Accessed August 3, 2016. https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf.

U.S. Geological Survey (variously dated). National field manual for the collection of water-quality data: U.S. Geological Survey Techniques of Water-Resources Investigations, book 9, chaps. A1—A10. http://pubs.water.usgs.gov/twri9A.

Vidic, R.D., Brantley, S.L., Vandenbossche, J.M., Yoxtheimer, D., Abad, J.D., 2013. Impact of shale gas development on regional water quality. Science 340, 6134. http://dx.doi.org/10.1126/science.1235009.

Westland, A.D., Boisclair, I., 1974. The analytical separation of phosphate from natural water by ion exchange. Water Res. 8, 467—470. http://dx.doi.org/10.1016/0043-1354(74)90078-5.