

Strengthening the practice and profession of statistics and data science using ethical guidelines

Rochelle E. Tractenberg

Reprint requests:

Rochelle Tractenberg  
Neurology Department, Building D, Suite 207  
Georgetown University Medical Center  
4000 Reservoir Rd. NW  
Washington, DC 20057 USA  
rochelle.tractenberg@gmail.com

**KEYWORDS:** Stewardship; Professional Identity; Professionalism; Mastery Rubric for Stewardship; Mastery Rubric for Ethical Reasoning; Data Science; ethical practice.

DOI: 10.31235/osf.io/93wuk  
[https:// osf.io/preprints/socarxiv/93wuk](https://osf.io/preprints/socarxiv/93wuk)  
CC-BY Attribution 4.0 International

Citation: Tractenberg, RE. (2019, May 1). Strengthening the practice and profession of statistics and data science using ethical guidelines. Retrieved from <https://doi.org/10.31235/osf.io/93wuk>

## Abstract

This article builds on the concept of disciplinary and professional stewardship, to discuss the ethical practice guidelines from two professional associations and a method that you can learn to use in order to implement those guidelines throughout a professional career. The steward is an individual who practices in a field in a manner that invites and warrants the trust of the public, other practitioners, and employers to uphold and maintain the integrity of that field. It is important to your sense of professional identity - and also your profession - to cultivate a sense of stewardship; and one of the foundational aspects of stewardly behavior is to understand professional practice guidelines and the types of behaviors that are expected by practitioners in a given field. Therefore, this article presents two sets of guidelines that can support professionalism, ethical practice, and the development of a coherent professional identity for the statistician and data scientist. The American Statistical Association (ASA) and the Association of Computing Machinery (ACM) are large professional organizations with international membership. An overall objective of each of these organizations is to promote excellence in and by their members and all those who practice in their respective – sometimes shared/joint – domains. It can be helpful to consider the field of ‘statistics and data science’ to be a hybrid of, or co-dependent on, these two fields, which is one reason why the two organizations are presented together. Another reason is that both organizations take ethical practice very seriously, and both engaged in lengthy projects to carefully revise their respective ethical guidelines for professional practice in 2018. Not only does engagement with the guidelines support you initiating, and beginning to demonstrate, your commitment to this particular professional identity, but also exploring the ethical guidelines for professional practice (through ASA or ACM) is a first step towards documenting your commitment to stewardly work as a data scientist. Ethical reasoning, the third focus of this article, helps deepen the understanding of the guidelines and can be useful to generate evidence of stewardly development.

## 1. Introduction

Most professional preparation – including undergraduate coursework, internships, and other experiences – is focused on learning all of the “tools of the trade”. These “tools” include methods, techniques, and how to apply these correctly and in the appropriate circumstances. Elsewhere (Tractenberg, 2019-a) I have asserted that “correct application” of these “tools” involves considerations beyond just whether a given tool of the trade will work to solve a specific problem – i.e., whether the application of some data science-specific knowledge is all that any given problem requires. A steward is defined as someone to whom “we can entrust the vigor, quality, and integrity of the field” (Golde & Walker 2006: p. 5). We have identified a core set of knowledge, skills, and abilities (KSAs) that, together, constitute stewardly behaviors (Rios et al. 2019). Stewardly practice – or the application of that discipline-specific knowledge in an ethical, professional, way - requires an understanding of professional standards (KSA 1, “requisite knowledge”; see Rios et al. 2019 and Tractenberg 2019-a). However, it is not very reassuring to entrust a field to someone who has simply memorized a list of professional practice standards or descriptions; greater trust is warranted for the steward who can demonstrate their ability to use the practice standards appropriately. For this reason (among others!), a developmental trajectory was published that discusses all of the stewardship KSAs together with how they can be demonstrated and improved upon over time, in a coherent whole called a Mastery Rubric (the Mastery Rubric for Stewardship, Rios et al. 2019; see Section 5 of this article and Tractenberg 2017 for greater detail about the Mastery Rubric construct). With the understanding that stewardship requires a set of KSAs, all being performed at a high level of sophistication by the independent practitioner (“journeyman”), reviewing the first stewardship KSA (of eight – see Section 2 and Tables 2 and 3) helps orient us to why “learning the ethical guidelines” is not sufficient to earn the distinction of “steward”.

Table 1. Performance level descriptors of the Stewardship KSA, “Requisite knowledge/situational awareness”

| Novice   | Apprentice  | Journeyman   | Master  |
|--|---|--|---|
| Largely unaware of the professional community and standards within which their academic or professional interests operate. | Learning to recognize when and how to demonstrate stewardship, that professional standards of practice involve both legal/illegal and ethical/unethical continua, and how to recognize and respond to these features. | Exercises professional practice standards and recognizes situations in which stewardship should be modeled and/or applied with respect to themselves and others, and to interactions within and outside of the profession or discipline. | Models, promotes, and teaches recognition of situations in which stewardship can and should be demonstrated; identifies strategies for how best to proceed when it isn't clear. |

Table 1 describes how stewardship KSA 1, “Requisite knowledge/situational awareness”, can develop over time as an individual becomes increasingly worthy of the designation of “steward”.

Notice that the differences between the novice and apprentice on this KSA do not hinge on the apprentice's memorization of the professional practice standards, but rather, on understanding that there are two dimensions, the legal/illegal one and the one that is a much more nuanced continuum, running from ethical to unethical. The apprentice steward is identified by their engagement in learning about these dimensions and how to both recognize them and respond to them in practice. Part of the recognition of illegal or unethical behaviors comes from an awareness that our professional decisions can lead to both harmful and beneficial effects or impacts. Clearly, most decisions that lead to illegal acts are unethical – although breaking traffic laws, committing illegal acts, in order to get an injured person to a hospital is a concrete example of illegal yet ethical decisions. While the relevant laws are usually clear (which does not prevent illegal behaviors!), we are focused on the more difficult circumstances surrounding – and complicating – decisions that are on the ethical to unethical *continuum*. In some cases, the harms that accrue following decisions that are contrary to ethical practice standards are obvious, but the exact point where behavior deviated from what the ethical practitioner should do/have done is less (or not) clear. Even if something is legal, it does not prevent harms from accruing to some as a result of our actions and decisions. In those cases, the harms must be mitigated or undone, if possible. In other cases, however, decisions have not yet been made – making it harder to figure out which guideline principles or ethics code elements should be followed. For more on harms and a method for determining harms and benefits, see Tractenberg (2019-b).

This article first explores the substance of the professional practice guidelines from the American Statistical Association (ASA) and Association of Computing Machinery (ACM). They may seem, on their faces, to be simple lists of desired/desirable behaviors. An inclination might be to just memorize them (and/or keep the lists handy “in case of emergency”). While these should definitely be bookmarked, the intention here is not to encourage that inclination! Instead, exploring the guidelines through the perspective of the evolving steward, and with the ethical reasoning framework that this article will introduce, will provide some evidence that ethical guidelines are essentially *useless if simply memorized*. These guidelines were not developed to be memorized - they are not lists of rules. The ASA and ACM are large professional organizations with international membership. An overall objective of each of these organizations is to promote excellence in and by their members and all those who practice in their respective – sometimes shared/joint – domains. It can be helpful to consider the field of statistics and data science to be a hybrid of these two fields, which is one reason why the two organizations are presented together. Another reason is that both organizations take ethical practice very seriously, and both engaged in lengthy projects to carefully revise their respective ethical guidelines for professional practice in 2018.

Guidelines describe the “best” professional practices that are agreed on by the practicing community. The ASA and ACM have articulated, and continue to refine and maintain, their respective statements of what constitutes ethical professional practice in their domains. The ASA Ethical Guidelines for Statistical Practice (GLs) and ACM Code of Ethics (CE) also specifically note that even those who do not self-identify as statisticians, or who are not members of the ASA, are obliged to follow these GLs whenever they work with data. Nearly every scientist (and many who are not scientists!) use statistics and increasingly, other aspects of data science, in their day to day work: biomedical research, policy, and business use statistical methodologies implicitly or explicitly to support decisions. While specific training and background is required

for specialists who develop or design computational machinery (algorithms, software, hardware, etc.), the ACM CE are also described as applying to *all uses/users of computation*. Both of these organizations want *all* applications of their disciplinary knowledge to be done ethically – and since many non-members use these methods, the GLs and CE are both described as pertaining to *all* who use or implement methods/technology from their domains.

Rather than serving as lists of behaviors to memorize, the ASA and ACM professional practice guidance documents are intended to be useful references over time; as experience and responsibilities change over time, individuals should also experience different levels of interaction with/relevance of the GL principles. Similarly, both the ASA and ACM documents entail performance descriptions relating to critical and fundamental aspects of most or all of our activities as practitioners, and are not limited to specific projects. Over time in a career, a wide variety of circumstances will occur and simply having memorized the GLs is not going to suffice for determining, and justifying, which GLs or GL/CE principles/elements are most applicable and lead to the most stewardly result for any given situation. That is why an increasingly sophisticated understanding of the professional practice standards is only *one* KSA in the repertoire of the steward.

As discussed elsewhere (Tractenberg 2019-a; see also Rios et al 2019), the novice performer of any given stewardship KSA is not only/always someone early in their training, but can also be someone further along in a career but new to the construct of stewardship. The apprentice is that individual who is actively engaged in developing the required foundation for stewardly practice in the domain (e.g., learning how to ethically utilize the tools of the trade), and the journeyman is that individual who is a recognized independent – ethical - practitioner in that field, domain, or trade. As we consider the first KSA of stewardship, we can see that the independent practitioner (journeyman steward) is the one who consistently exercises the professional standards: consistency of performance means following the practice standards before harms accrue, as well as identifying whenever they have not been followed so that harms can be identified and mitigated. Journeyman performance of this KSA requires far more than memorization and/or keeping the guidelines handy.

With 52 (ASA) or even 28 (ACM) elements, the GLs/CE contain a LOT to remember – but the point is that memorizing the CE/GLs and their content is not sufficient for stewardly professional practice. Recognizing when, and that, the practice standards from either or both professional perspectives may be needed (identifying that an ethical problem or challenge exists) and determining what alternatives exist are two very difficult – and essential – steps in utilizing these documents as they were intended to be used. This recognition is the crux of stewardship KSA #1: Requisite knowledge and situational awareness.

## 2. Stewardship and professional practice guidelines

Guidelines are established for “good” or ethical professional practice. Just as there is more to “learning statistics” than applying “the right” formulae or running software, there is also more to “being a stewardly or ethical statistician/data scientist” than learning the GLs/CE for either – or even both – of these professional associations. There are multiple methods for accomplishing many statistics and data science tasks, and the “best” method may depend on what resources are

available and the priorities that the data themselves cause/create (e.g., hard to collect; hard to verify; prone to measurement error; confounded with other variables in the data set, etc.). Similarly, there are often multiple ways to identify ethical problems or questions that can arise in practice, before or after they have arisen. There are definitely multiple ways to address ethical problems, including “ignore it”. Factual knowledge (i.e., memorizing the GLs/CE) is unlikely to help you develop the ability to formulate, and then choose from, multiple possible responses to a breach of ethics - particularly if *all* of the options are justifiable. Once you are familiar with the contents of the GLs/CE, including their specific elements, their use in your everyday decision-making will typically involve some prioritization of competing obligations or solutions. It is this *second* step – making decisions that lead to actions that are consistent with the GLs/CE – that characterizes the journeyman steward.

Unlike laws, guidelines are not rules. Thus, determining which practice principles apply – and which ones take precedence when two or more are applicable – and in which situations, is an important and complex type of reasoning. What is illegal is a lot clearer than what is “unethical” in the workplace - also, a shorter list, so it makes sense to tell people what *not* to do. However, a focus on these (what not to do – or, how to avoid getting caught doing what you shouldn’t) will not promote the integrity of the field. The eight different aspects of knowledge, skill, and ability (KSAs) that comprise stewardship are discussed in depth elsewhere (Rios et al. 2019; Tractenberg, 2019-a) and are listed at the end of this Section. As noted, the first KSA requires both a familiarity with, and an understanding of, professional practice guidelines. However, it is also true that “data science” as a career or field is in a very early stage of development, and as of March 2019, there are no professional practice standards for “data science” that can be used to learn to reason ethically. In the United Kingdom, the government agencies concerned with data ethics have worked towards a framework within which government practices can be aligned to promote ethical use and collection of data ( [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/524298/Data\\_science\\_ethics\\_framework\\_v1.0\\_for\\_publication\\_1.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/524298/Data_science_ethics_framework_v1.0_for_publication_1.pdf) ), giving an example of guidance about how to *practice* data science in an ethical way. Another example is the “Data Science Ethics Checklist”, <http://deon.drivendata.org/> (DSEC), which provides a list of things to consider – again, as a project goes on/in the practice of data science. Neither of these is a method for formalizing – for the learner/future practitioner – what it means to be an ethical or stewardly professional. While these lists can be useful in documenting the features of a project and whether ethical features were considered at each stage, these do not refer to professional practice standards, so cannot help orient you towards stewardly professional practice.

Tractenberg et al. (2015) pointed out that the development of professional guidelines is one way for a profession to clarify the qualifications to practice, and to solidify the formation of coherent professional identities across practitioners. Neither the UK nor the DSEC examples are interested in the professional identity of the practitioner; the GLs and EC clearly are. It is also important to recognize that many discussions of “ethical data science” end up focusing solely on protecting privacy and confidentiality of data, and security. However, there is a lot more to “professional practice with data” for the data scientist than data security and privacy. The DSEC includes items relating to data storage (and security and longevity) and both analysis and modeling of the data that are collected. The entire DSEC is 20 items – privacy and security of the data are only two of

these 20 items. Similarly, the ASA (52 items in total) and ACM (28 items in total) include only a few items relating to privacy and confidentiality – not because these are not important aspects of stewardly practice, but because there is a great deal more to professionalism than just privacy and confidentiality considerations. Moreover, harms do not accrue *solely* due to breaches in privacy and confidentiality. And, harms to the profession accrue whenever our actions and decisions decrement, rather than bolstering, the integrity of the field. Furthermore, unlike checklists both the ASA and ACM guidance recognize that in some situations, more than one of their constituent elements may come into play. While the ACM encourages prioritizing “the public good” to resolve conflicting ACM principles, the ASA stipulates that, whenever principles conflict, priority of considerations must be done “according to context”.

There is no set of *practice* guidelines specifically for data science, and there is no real rationale for a separate set that does not include ASA and ACM perspectives. We therefore turn to the guidelines of these organizations for the initiation of your development of stewardly behaviors. This article is entitled “strengthening your practice and your profession” because it seeks to not only promote ethical practice, but also to enable you to use your growing understanding of, and commitment to, ethical practice of data science to *strengthen the profession* as the steward you seek to become. As you saw in Table 1, the journeyman steward recognizes when to use the GL/CE *and does so*. This recognition plus action demonstrates for all observers that the steward *is* a person to whom the field can be entrusted. By accepting this responsibility, the steward not only practices ethically, s/he also helps to create both work conditions and a culture that support stewardship and ethical practice – this is one way that the steward strengthens their profession.

Tables 2 and 3, presented in the next two sections, each show a matrix of alignment between the Principles of the ACM CE (Table 2, Section 3) and ASA GLs (Table 3, Section 4) relating to the practitioner. In each of these tables, the *rows* represent the features of the respective organizations’ CE/GLs, while the columns capture the KSAs of the steward.

As you can see from the guidance documents themselves, or their representations in Sections 3 and 4 and Tables 2 and 3, the ASA GLs and ACM CE are complex. The 2018 revisions of the ACM CE took over a year to do, with multiple versions being circulated throughout the ACM membership for comment and revisions before the final was approved in February 2018. The 2018 ASA GL revisions by the Committee on Professional Ethics of the ASA (Chaired by this article’s author from 2016-2019) followed a major revision effort 2014-2016. The 2016 GLs were revised in 2018 to specifically note that bullying, sexual harassment/assault, and intimidation are not ethical - but also, that any “ethical statistician” who observes these should do what they can to stop them (again, promoting an ethical workplace and strengthening the profession by ensuring that the statistical practitioner can work in an ethical context). This raises an important point about the utility and complexity of the ASA and ACM guidelines: *they are not static*. As reflections of the ideals for ethical practice, such guidelines require ongoing attention – just as was described for the individual’s development and documentation of the stewardly mindset. The steward who “exercises professional practice standards” does so by understanding their dynamic nature, as well as how they interrelate – not by simply memorizing what they contain.

The KSAs of stewardship (S) are more fully discussed elsewhere (Tractenberg, 2019-a; Rios et al. 2019), so they are abbreviated in Tables 2 (ACM GL alignment) and 3 (ASA GL alignment) for simplicity’s sake, as follows:

Requisite knowledge /situational awareness: **S1**

Create and/or generate new methods/new knowledge: **S2**

Critically evaluate extant knowledge: **S3**

Conserve ideas (or not, if deemed rejectable & non-conservation is justified): **S4**

Responsibly write: **S5**

Responsibly teach/mentor/model: **S6**Responsibly apply disciplinary knowledge: **S7**Responsibly communicate: **S8**

The reader is strongly encouraged to become familiar with these KSAs of stewardship (e.g., Rios et al. 2019; Tractenberg 2019-a) to make Tables 2 and 3 more comprehensible. The Code of Ethics (CE) and Guidelines (GLs) are summarized in the following text and should be accessed and bookmarked, so that both Tables 2 and 3 and the discussion of those Tables, are more meaningful.

### 3. ACM CE

The ACM Code of Ethics can be found here: <https://ethics.acm.org/> and was initially published in 1992. The Code was updated in 2018. The code is motivated by the fact that “Computing professionals’ actions change the world”: “To act responsibly, they should reflect upon the wider impacts of their work, consistently supporting the public good. The ACM Code of Ethics and Professional Conduct (“the Code”) *expresses the conscience of the profession.*” (ACM 2018, emphasis added). The ACM Code is not intended just for ACM members: they specify that “(t)he Code is designed to inspire and guide the ethical conduct of all computing professionals, including current and aspiring practitioners, instructors, students, influencers, and anyone who uses computing technology in an impactful way.” Finally, the Code “serves as a basis for ethical decision-making”. In the final section of this article, a method for ethical decision making will be discussed.

Table 2 shows the alignment of ACM principles with KSAs of Stewardship. We know that stewardship KSA #1 involves understanding and exemplifying professional practice standards, but it can be helpful to explore how the ACM CE principles (and the ASA GLs in Table 3) are essentially in agreement with stewardship. The ACM Code is organized into four main areas: General Moral Principles; Professional Responsibilities; Professional Leadership Principles; and Compliance with the code. These four areas contain 2-9 specific elements yielding 28 principles, all listed in the rows of Table 2 together with their degree of alignment (the # of KSAs that “match” the ACM elements on some level) with the stewardship KSAs that appear in the columns. For simplicity, Table 2 shows only the heading of each CE element; the full Code contains explanatory narrative of each of these elements.

Table 2. Alignment of 2018 V3 ACM Code of Ethics (rows) with Stewardship KSAs (columns).

|                     | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 |                  |
|---------------------|----|----|----|----|----|----|----|----|------------------|
| Stewardship KSAs:   |    |    |    |    |    |    |    |    |                  |
| ACM Code of Ethics: |    |    |    |    |    |    |    |    | <b>alignment</b> |



| <b>1. GENERAL MORAL PRINCIPLES</b>  |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| <i>A computing professional should...</i>   |   |   |   |   |   |   |   |   |   |
| 1.1 Contribute to society and to human well-being, acknowledging that all people are stakeholders in computing.   |   | * |   | * | * | * | * | * | 6 |
| 1.2 Avoid harm.<br><i>In this document, "harm" means negative consequences to any stakeholder, especially when those consequences are significant and unjust.</i> | * | * | * | * | * | * | * | * | 8 |
| 1.3 Be honest and trustworthy.  |   | * |   |   | * | * | * | * | 5 |
| 1.4 Be fair and take action not to discriminate.  |   | * | * |   | * | * | * | * | 6 |
| 1.5 Respect the work required to produce new ideas, inventions, creative works, and computing artifacts.  |   | * | * | * | * |   |   | * | 5 |
| 1.6 Respect privacy.  | * | * |   |   | * |   | * | * | 5 |
| 1.7 Honor confidentiality.  | * | * |   |   |   |   | * | * | 4 |
| <b>2. PROFESSIONAL RESPONSIBILITIES</b>   |   |   |   |   |   |   |   |   |   |
| <i>A computing professional should...</i>   |   |   |   |   |   |   |   |   |   |
| 2.1 Strive to achieve high quality in both the process and products of professional work.   |   | * |   |   | * |   | * | * | 4 |
| 2.2 Maintain high standards of professional competence, conduct, and ethical practice.  | * |   |   |   | * | * | * | * | 5 |
| 2.3 Know, respect, and apply existing rules pertaining to professional work.  | * |   |   |   |   |   | * |   | 2 |
| 2.4 Accept and provide appropriate professional review.   | * |   | * |   |   |   |   | * | 3 |
| 2.5 Give comprehensive and thorough evaluations of computer systems and their impacts, including analysis of possible risks.                                      |   |   | * | * |   |   |   | * | 3 |
| 2.6 Have the necessary expertise, or the ability to obtain that expertise, for  | * |   |   |   |   |   | * |   | 2 |

|  |   |  |  |   |  |   |   |   |    |
|--|---|--|--|---|--|---|---|---|----|
| completing a work assignment before accepting it. Once accepted, that commitment should be honored.  |   |  |  |   |  |   |   |   |    |
| 2.7 Improve public awareness and understanding of computing, related technologies, and their consequences.   | ‡ |  |  |   |  |   | ‡ | ‡ | 3‡ |
| 2.8 Access computing and communication resources only when authorized to do so.  | ‡ |  |  |   |  |   | ‡ |   | 2‡ |
| 2.9 Design and implement systems that are robustly and useably secure.   | ‡ |  |  |   |  |   | ‡ |   | 2‡ |
| <p align="center"><b>3. PROFESSIONAL LEADERSHIP PRINCIPLES</b></p> <p align="center"><i>In this section, “leader” means any member of an organization or group who has influence, educational responsibilities, or managerial responsibilities. These principles generally apply to organizations and groups, as well as their leaders.</i></p> <p><i>A computing professional should...</i></p> |   |  |  |   |  |   |   |   |    |
| 3.1 Ensure that the public good is the central concern during all professional computing work.   |   |  |  |   |  |   |   |   | 0  |
| 3.2 Articulate, encourage acceptance of, and evaluate fulfillment of the social responsibilities of members of an organization or group.   | ‡ |  |  |   |  |   |   |   | 1‡ |
| 3.3 Manage personnel and resources to enhance the quality of working life.   | ‡ |  |  |   |  | ‡ | ‡ |   | 3‡ |
| 3.4 Articulate, apply, and support policies and processes that reflect the principles in the Code.   | * |  |  |   |  |   |   |   | 1  |
| 3.5 Create opportunities for members of the organization or group to learn and be accountable for the scope, functions, limitations, and impacts of systems.   | * |  |  |   |  |   |   | * | 2  |
| 3.6 Retire legacy systems with care.   |   |  |  | * |  |   |   |   | 1  |

|   |   |  |  |  |  |   |   |  |      |
|---|---|--|--|--|--|---|---|--|------|
| 3.7 Recognize when a computer system is becoming integrated into the infrastructure of society, and adopt an appropriate standard of care for that system and its users.  | ‡ |  |  |  |  |   | * |  | 1/1‡ |
| <b>4. COMPLIANCE WITH THE CODE</b>  |   |  |  |  |  |   |   |  |      |
| <i>A computing professional should...</i>   |   |  |  |  |  |   |   |  |      |
| 4.1 Uphold, promote, and respect the principles of the Code.  | * |  |  |  |  |   |   |  | 1    |
| 4.2 Treat violations of the Code as inconsistent with membership in the ACM.  |   |  |  |  |  | ‡ |   |  | 1‡   |
| (‡) The Guidelines and the stewardship KSAs are quite different; the language used in each case is also quite different. These ‡s indicate that, <i>given the correct context</i> , that KSA would be aligned with that Guideline Principle, but superficially the alignment is not apparent. |   |  |  |  |  |   |   |  |      |

As you can see in Table 2, there is a great deal of alignment between the GL principles and the KSAs of stewardship. But, as noted in the table, in some cases the ACM GL principles are explicitly worded for a computational context that makes it difficult to identify alignment exactly; the ‡ symbol is used to identify these difficulties. As an example, take ACM GL principle 4.2: “Treat violations of the Code as inconsistent with membership in the ACM.” Since the ACM is an organization with members but stewardship is a construct (and not an organization), it isn’t possible to match 4.2 with any of the KSAs of stewardship exactly. However, the spirit of KSA S7, “Responsibly teach/mentor/model”, suggests that if you are teaching, training, or mentoring people to be stewardly, you would logically treat violations of your stewardship training as inconsistent with the label “steward”. These conceptual similarities are indicated in the ACM alignment table with the ‡ symbol. By contrast, consider CE 2.2 (“Maintain high standards of professional competence, conduct, and ethical practice”), which is a clearer and more literal match to the stewardship KSAs S1 and S5-S8 (indicated with the asterisk \*).

This Table reinforces the claim that just memorizing the list of items on the ACM CE would not be sufficient even for a simple alignment exercise: without thoughtful consideration of the CE elements and stewardship KSAs, you would only be able to literally match words. Part of a deeper understanding of the practice guidelines (and of stewardship) entails moving from the superficial/literal to the more abstract intent of the CE and what it means, and looks like, to be stewardly as a data scientist. It is worthwhile for you to consider the alignment represented in Table 2: focus first on identifying the cells where it is most clear to you how the CE elements and stewardship KSAs “match”, and make sure you understand the alignment at that level. Then move on to the cells where alignment seems more abstract; understanding how all features of professional practice represent, or are consistent with, stewardship can help you to demonstrate apprentice level performance, and then move on to the journeyman level of stewardship KSA #1.

#### 4. ASA GLs

The ASA Ethical Guidelines for Statistical Practice (“ASA GLs”) began to take shape in the 1970s and were first published by the ASA in 1995. As reiterated in 2016 and 2018, the ASA GLs are “intended to help statistics practitioners make and communicate decisions ethically”. Like the ACM CE, *decisions* that the practitioner makes are highlighted as the points around which consideration and ethical practice are focused. The ASA GLs also specify that they exist “(t)o inform those relying on statistical analysis, including employers, colleagues and the public, of the standards that they should expect.” Just like the ACM CE, the ASA GLs are also relevant beyond those who are members of the ASA or who self identify as “statisticians and data scientists” – specifying that the GLs “...should guide both those whose primary occupation is statistics and those in *all other disciplines* who use statistical methods in their professional work.” (emphasis added). The ASA GLs promote observable professional behavior among practitioners, and trust in their work by employers/colleagues/collaborators.

The eight general principles for ethical statistical practice (<http://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx>) are shown below; the full 52 items appear at that URL:

##### **A. Professional Integrity and Accountability**

The ethical statistician uses methodology and data that are relevant and appropriate, without favoritism or prejudice, and in a manner intended to produce valid, interpretable, and reproducible results. The ethical statistician does not knowingly accept work for which he/she is not sufficiently qualified, is honest with the client about any limitation of expertise, and consults other statisticians when necessary or in doubt. It is essential that statisticians treat others with respect.

##### **B. Integrity of data and methods**

The ethical statistician is candid about any known or suspected limitations, defects, or biases in the data that may impact the integrity or reliability of the statistical analysis. Objective and valid interpretation of the results requires that the underlying analysis recognizes and acknowledges the degree of reliability and integrity of the data.

##### **C. Responsibilities to Science/Public/Funder/Client**

The ethical statistician supports valid inferences, transparency, and good science in general, keeping the interests of the public, funder, client, or customer in mind (as well as professional colleagues, patients, the public, and the scientific community).

##### **D. Responsibilities to Research Subjects**

The ethical statistician protects and respects the rights and interests of human and animal subjects at all stages of their involvement in a project. This includes respondents to the census or to surveys, those whose data are contained in administrative records, and subjects of physically or psychologically invasive research.

##### **E. Responsibilities to Research Team Colleagues**

Science and statistical practice are often conducted in teams made up of professionals with different professional standards. The statistician must know how to work ethically in this environment.

##### **F. Responsibilities to Other Statisticians or Statistics Practitioners**

The practice of statistics requires consideration of the entire range of possible explanations for observed phenomena, and distinct observers drawing on their own unique sets of experiences can arrive at different and potentially diverging judgments about the plausibility of different explanations. Even in adversarial settings, discourse tends to be most successful when statisticians treat one another with mutual respect and focus on scientific principles, methodology and the substance of data interpretations.

### **G. Responsibilities Regarding Allegations of Misconduct**

The ethical statistician understands the differences between questionable statistical, scientific, or professional practices and practices that constitute misconduct. The ethical statistician avoids all of the above and knows how each should be handled.

An 8<sup>th</sup> Guideline Principle is specific for *employers*:

### **H. Responsibilities of Employers, Including Organizations, Individuals, Attorneys, or Other Clients Employing Statistical Practitioners**

Those employing any person to analyze data are implicitly relying on the profession's reputation for objectivity. However, this creates an obligation on the part of the employer to understand and respect statisticians' obligation of objectivity.

Because this article is focused on the practitioner as steward, Table 3 does not include GL Principle H. We revisit GL Principle H in the final section, however, as it does relate specifically to how the steward contributes to a culture of ethical practice (strengthening the profession) by promoting an ethical workplace environment and, in the case where the steward employs or manages statisticians or data scientists, following ASA GL Principle H (and ACM CE 3).

Table 3. Alignment of ASA Professional Statistics Guidelines A-G (rows) with Stewardship KSAs (columns). Adapted from Rios et al (2019) with permission.

| Stewardship KSAs:  | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 |                  |
|--|----|----|----|----|----|----|----|----|------------------|
| ASA GUIDELINE:   |    |    |    |    |    |    |    |    | <b>alignment</b> |
| A. Professional Integrity and Accountability                           | *  |    | *  | *  | *  | *  |    |    | 6                |
| B. Integrity of data and methods                                       | *  |    | *  |    | *  | *  | *  | *  | 6                |
| C. Responsibilities to Science/Public/Funder/Client                    | *  |    | *  | *  | *  |    |    |    | 5                |
| D. Responsibilities to Research Subjects                               | *  | *  | *  | *  |    | *  |    |    | 5                |
| E. Responsibilities to Research Team Colleagues                        | *  | *  | *  | *  |    | *  | *  | *  | 7                |
| F. Responsibilities to Other Statisticians or Statistics Practitioners | *  | *  | *  | *  |    | *  | *  | *  | 7                |
| G. Responsibilities Regarding Allegations of Misconduct                | *  |    |    |    |    | *  | *  | *  | 4                |
| <b>alignment</b>   | 7  | 3  | 6  | 5  | 3  | 6  | 4  | 4  |                  |

The row marginals in Table 3 show that between 3-7 of the seven core ASA Ethical Guideline Principles are aligned with each stewardship dimension. Every ASA Ethical Guideline Principle *is* relevant for every quantitative practitioner - and practicing according to these Principles is clearly relevant for stewardship of the profession of statistics and the practice of data analysis/data science. Both Tables 2 and 3 show that professional behaviors that are consistent with the ASA Guidelines and the ACM CE (rows) can also generate evidence of stewardship of the discipline or profession (columns).

Note that there is a bit more overlap - alignment is stronger- between ASA GL Principles and the stewardship KSAs (i.e., the marginals are greater in Table 3 compared to Table 2). Because there are so many more ASA GL elements (52 compared with 28 for the ACM), there is a bit more material to align. The construct of stewardship was originally formulated to describe those completing the PhD, which is the research degree designating the holder to be qualified to do research in their field. Statistics is a discipline that all scientists, and most researchers, use, so ethical statistical practice happens to be described in terms that are highly consistent with those of stewardship. While computers are almost universally used in research, the level of reliance on computation (covered by the ACM CE) in research and science is not as high as it is for statistics (which is essential in both data science and any experimental design/analysis); thus, the language of the ACM principles may not be as compatible with that of the stewardship KSAs as the ASA GL elements are. It is important to note that all three of these (ACM CE, ASA GLs, and the construct of stewardship) were created completely independently of each other – so the fact that there is any overlap at all is evidence of the reliability of the alignment you see in Tables 2 and 3, as well as a consensus towards what constitutes “good practice” for stewardly and ethical statistics and data science.

Tables 2 and 3 show that, as you integrate stewardship KSAs into your training and your own self-directed professional development, you will create opportunities to demonstrate both the ACM CE and the ASA GL elements. Similarly, if you were to shift focus to a professional identity consistent with statistics and data science, and focus your self-directed development on the ASA GLs and/or ACM CE, you will create opportunities to demonstrate most, if not all, of the KSAs of stewardship. The KSAs of stewardship are consistent with the ASA GLs and ACM CE; these can be used together to reinforce one another – and to preserve/promote the integrity of data science - throughout a career.

It is also important to recognize that these practice standards can be useful to you in many different ways – they are not intended for your reference only when a problem arises (although it is hoped that they *will* be useful then as well!). You do not have to be a member of the organization for their ethical guidelines to apply to you: as both the ASA and ACM note, if you perform functions that constitute part of professional practice in these fields, then the professional practice standards *will apply to your behavior and performance*. There are additional considerations that must also be acknowledged:

- Knowing that there are these GLs/CE can help you start to formulate a professional identity. Committing to following and upholding these – which is actually one of the ACM CE elements (4.2) and is also threaded throughout the KSAs of stewardship (Rios et al. 2019) is the first step.

- It is not ethical to selectively follow some of the GLs or CE elements (the ones you prefer), but it is also true that some of the GL/CE elements may never, or only rarely, pertain to you – even if your job title is “statistician”, “data scientist” or “computational machinist”. The difference here is the choice: it is not ethical to choose which Principles or elements to follow, but it is consistent with stewardly practice if some of the elements never apply (e.g., “use care when modifying or retiring systems.” (ACM 3.6) might be limited to those with leadership roles).
- There are seven different ASA GL principles and four different ACM GL principle areas, but these are interrelated –within and across the standards. The practitioner (data scientist, statistician, computational machinist) *has responsibilities* in each of those areas. Although separately described, Principles or elements may actually come into play simultaneously in any situation – which underscores the assertion that simply memorizing the contents of these documents is not sufficient. The steward/ethical practitioner must also learn how to prioritize the different principles, in order to make ethical and stewardly decisions that can be justified. Both the ACM and ASA standards articulate this.
- There is one area in both the ASA and ACM GLs that relate to “the employer”. For the ASA, this area is H: “responsibilities of employers”, while for the ACM, this area is 3 (Professional Leadership Principles). These are important to note because while *you* are responsible to your discipline and the disciplinary community to practice ethically and in a stewardly manner, your employer has -and *you*, if you employ or manage others, have-a responsibility to promote and protect ethical and stewardly behavior.
- It is never acceptable to seek or create “loopholes” in the GLs/CE (unless you also seek to fix/close them). If an unexpected ethical challenge arises, the steward seeks guidance, not exceptions, in the GLs/CE. To try and justify unethical or underspecified behaviors (“no one said I couldn’t!”) is unprofessional, unethical, and not stewardly.

While professional practice standards exist and are maintained to enable professional identity formation and to promote a coherent view of the profession by the public and other stakeholders (Tractenberg et al. 2015), concerns about data safety, privacy and confidentiality have grown as technology has allowed more and more of our personal data to be collected and used. In the face of growing concerns about safety, privacy, and unethical commercial uses of personal data, many groups are creating tools and methodologies to “help”. One such tool is the Data Science Ethics Checklist (DSEC <http://deon.drivendata.org/>). This tool was designed to support decision making (i.e., it was intended to be “actionable”), but was developed from the perspective that “the primary benefit of a checklist is ensuring that we don't overlook important work”. Table 4 explores the alignment of the DSEC with the 8 stewardship KSAS.

|                     |    |    |    |    |    |    |    |    |                  |
|---------------------|----|----|----|----|----|----|----|----|------------------|
| Stewardship KSAs:   | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | <b>alignment</b> |
| Data Science Ethics |    |    |    |    |    |    |    |    |                  |





| C. Analysis   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|
| <b>C.1 Missing perspectives:</b><br>Have we sought to address blindspots in the analysis through engagement with relevant stakeholders (e.g., checking assumptions and discussing implications with affected communities and subject matter experts)? | * |   | * | * | * |   |   |   | 5 |
| <b>C.2 Dataset bias:</b> Have we examined the data for possible sources of bias and taken steps to mitigate or address these biases (e.g., stereotype perpetuation, confirmation bias, imbalanced classes, or omitted confounding variables)?         | * | * | * | * |   | * |   |   | 5 |
| <b>C.3 Honest representation:</b><br>Are our visualizations, summary statistics, and reports designed to honestly represent the underlying data?  | * |   | * |   | * | * | * | * | 6 |
| <b>C.4 Privacy in analysis:</b><br>Have we ensured that data with PII are not used or displayed unless necessary for the analysis?  |   | * |   |   | * |   | * | * | 4 |
| <b>C.5 Auditability:</b> Is the process of generating the analysis well documented and reproducible if we discover issues in the future?  | * | * | * | * | * | * |   |   | 6 |
| D. Modeling   |   |   |   |   |   |   |   |   |   |
| <b>D.1 Proxy discrimination:</b><br>Have we ensured that the model does not rely on variables or proxies for variables that are unfairly discriminatory?  | * | * | * | * |   | * |   |   | 5 |
| <b>D.2 Fairness across groups:</b><br>Have we tested model results for fairness with respect to   |   | * | * | * |   |   | * |   | 4 |

|   |   |   |   |   |  |  |   |   |   |
|---|---|---|---|---|--|--|---|---|---|
| different affected groups (e.g., tested for disparate error rates)?   |   |   |   |   |  |  |   |   |   |
| <b>D.3 Metric selection:</b> Have we considered the effects of optimizing for our defined metrics and considered additional metrics?  |   | * | * | * |  |  | * |   | 4 |
| <b>D.4 Explainability:</b> Can we explain in understandable terms a decision the model made in cases where a justification is needed?   |   | * | * | * |  |  |   | * | 4 |
| <b>D.5 Communicate bias:</b> Have we communicated the shortcomings, limitations, and biases of the model to relevant stakeholders in ways that can be generally understood?   |   | * | * | * |  |  |   | * | 4 |
| <b>E. Deployment</b>  |   |   |   |   |  |  |   |   |   |
| <b>E.1 Redress:</b> Have we discussed with our organization a plan for response if users are harmed by the results (e.g., how does the data science team evaluate these cases and update analysis and models to prevent future harm)? | * |   | * | * |  |  | * | * | 5 |
| <b>E.2 Roll back:</b> Is there a way to turn off or roll back the model in production if necessary?   |   |   |   |   |  |  | * |   | 1 |
| <b>E.3 Concept drift:</b> Do we test and monitor for concept drift to ensure the model remains fair over time?  |   |   | * | * |  |  | * |   | 3 |
| <b>E.4 Unintended use:</b> Have we taken steps to identify and prevent unintended uses and abuse of the model and do we have a plan to monitor these once the model is deployed?  | * | * | * | * |  |  | * |   | 5 |

Although the DSEC was developed very recently (e.g., in about 2018, inferred from the citations listed on the deon.org site) and independently of considerations of stewardship (or the ACM Code of Ethics), the DSEC does call on many of the stewardship KSAs. The deon objectives do not include a focus on professional identity but do include an acknowledgement that checklists enable quick(-er) decisionmaking that does include key ethical considerations. The alignment in Table 4 shows that key stewardly KSAs are at least potentially aligned – while the language is again, not specific to the scientific context (like the ASA GLs tend to be), contemplating the DSEC items can support your development of evidence of progress and growth in stewardly KSAs. However, simply using this or other checklists does not promote growth in stewardship; and there are key features of professional practice that fall outside of the “project” type that a checklist must necessarily focus on. For example, the ACM CE includes a section (with 7 items) focused on leadership, which entails different decisions specific to the role and quite distinct from those considered in the DSEC. A tool like the DSEC can be useful, but should not supplant or interfere with professional identity development. “I use a checklist!” is not a professional identity - “I am a steward of data science!” *is*.

## 6. Ethical Reasoning

In other articles on stewardship (Rios et al. 2019; Tractenberg 2019-a) you will have read about the Mastery Rubric (MR) construct, and specifically how it can be used to promote and facilitate the development of stewardship. Briefly, a MR is a curriculum development and evaluation tool. The MR describes the KSAs that are intended to be taught (for a specific curriculum or program of study), together with concrete and observable performance descriptions that allow the instructor as well as the learner to monitor (and encourage) progress towards articulated learning goals. Thus, for a given domain, a MR can be used by an instructor to help ensure that their course learning goals are consistent with the overall curriculum, but they are also useful to the learner both in and outside of formal educational programs. A key feature of a MR is that the learner can use them to make sure they are progressing from one performance level description to the next – similarly to how we have been revisiting the descriptions of stewardship KSA #1 (Table 1) to consider what a person needs to do in order to demonstrate that they are at the apprentice level, or to move up to the journeyman level. You can read more about the MR construct in Tractenberg (2017) or Rios et al. (2019) and Tractenberg (2019-a); the latter two papers specifically discuss the MR that was created to support teaching, learning, and development in stewardship.

As noted earlier, the MR for stewardship (MR-S, Rios et al. 2019) outlined the eight knowledge, skills, and abilities (KSAs) that, when practiced, can generate evidence of a commitment to being, and performance as, a steward of the discipline, field or profession. Another MR was created in 2012 outlining KSAs that are involved in reasoning ethically, the MR for ethical reasoning (MR-ER, Tractenberg & FitzGerald, 2012). Ethical reasoning is an approach to identifying “responsible” responses to an identified ethical challenge, and choosing among them in a defensible manner. Ethical Reasoning (ER) KSAs, like all KSAs in every MR, are learnable and improvable. That means that, like with stewardship, there are ER-specific KSAs to learn, and the level at which you perform each KSA also changes as you become more sophisticated and capable.

While the DSEC and the GLs/CE are not static – and are recognized in their respective contexts as being subject to change across time and as technology develops – ER KSAs *are* intended to be both learnable and improvable – just like the stewardship KSAs. This is a set of KSAs that you can deploy in a planning stage, i.e., before any unethical behaviors can be executed (similar to the DSEC), or to evaluate what may have gone wrong or what may be going on in an existing situation. The ER KSAs are actually steps you can walk through:

1. Identification and assessment of one's prerequisite knowledge;
2. recognition of an ethical issue;
3. identification of relevant decision-making frameworks;
4. identification and evaluation of alternative actions;
5. making & justifying a decision (about the moral issue); and
6. reflection on the decision.

Any practitioner in any field needs to be able to do the KSAs of ethical reasoning in order to both identify ethical challenges that arise and to make and justify decisions about how to respond to those challenges. Note that the DSEC referred to earlier (<http://deon.drivendata.org/>) may support you identifying and assessing your prerequisite knowledge (ER KSA #1): if you do not know how to determine if data were obtained with informed consent (Checklist item A. 1), then you recognize a gap in your prerequisite knowledge and that this gap must be addressed before any ethical decisions can be made. However, the items on this Checklist are all yes/no questions. If you answer no to any of them, does it constitute an ethical issue? “Any NO answer creates an ethical problem” is implied, but what to do if you answer NO to any of them is completely unspecified. In fact, the ACM CE and ASA GLs also do not direct you as to *how* exactly you should do any of the things on the lists – that is left up to your judgment, ability level, and the specific context in which you practice. This underscores the point that practice in statistics and data science involves decisions, and it is the practitioner's responsibility to recognize these decisions, make them ethically, and justify how these decisions were made. The last two KSAs (make and justify a decision; and reflect on the decision) are explicit about this, and this is part of what makes the ER KSAs so compatible with the ACM CE and ASA GLs: all are concerned with making ethical decisions.

Here are general features of the interactions between the CE and GLs and each ER KSA:

1. Identify and ‘quantify’ your prerequisite knowledge: Before you have done anything and are still planning, you can use the ASA GLs, ACM EC – and even the data science ethics checklist – to plan (using the items on all of these documents to make sure that you know, or have access to, all the information needed to go on in an ethical fashion). For ASA GLs, Principle A suggests you need a full understanding of the data and the methods, so that whatever you end up planning will be “valid, interpretable, and reproducible”. For ACM CE, 1.2 states that you need to avoid harm. Looking just at ACM 1.2 and ASA A here shows that even valid, interpretable and reproducible work that causes harm will be unethical. In fact, the ASA GL Preamble states, “using statistics in pursuit of unethical ends is inherently unethical.” So ER KSA #1 actually goes fairly far towards preventing unethical behavior, or not allowing decisions to do unethical things, when you bring the ASA and ACM guidance to bear. Just using the DSEC (or any checklist) does not address harm prevention, but rather, stipulates that the ethical data scientist plans a response to harms that may arise (DSEC E.1) or unintended uses to which technology may be put (DSEC E.4).

2. Identify or recognize the ethical issue. What about what you are doing, planning, or have observed is inconsistent with the GLs or EC? What seems “questionable”? Amazingly, this is the obvious crux of any ethical decision, but is also incredibly difficult! If you look at the prerequisite knowledge you pulled together for ER KSA #1 above, you might see where an ethical issue (i.e., a potentially unethical decision, or a behavior someone else has done that is unethical) can arise or has arisen. Importantly, if you observe someone using statistics or data science in pursuit of unethical ends, they are violating the purpose of the ASA GLs, ASA GL Principle A, and ACM CE Principle 1.2. For you to make a decision, however – about what you should do (or not do), you need to go back to ER KSA #1 above and make sure you identify all the ACM and ASA elements that you think are violated (or would be violated) by the specific behavior you are thinking of or observing. Note that these determinations are also required to identify harms (DSEC E.1) and unintended use/abuse (DSEC E.4).
3. Identify decision-making frameworks. The ASA GLs and ACM EC describe “the ethical practitioner”. In formal terms, this makes the guidelines “virtue ethics” decision making frameworks. When you choose or decide to behave or operate/practice in a way that is consistent with these frameworks, *you are* “the ethical practitioner” (so naturally, acting or deciding inconsistent with these frameworks makes that person unethical). When you are making decisions using these frameworks, you are generally going to be modeling your decisions according to “what would the ethical practitioner do in this case?” Another way to use these frameworks is called “utilitarian”. Formally, this framework can be helpful in identifying the positive and negative effects of a decision. For the utilitarian perspective, you need to have a firm idea of what those positive and negative effects would be *on*, and that can be very complex to conceptualize. The utilitarian perspective can generally be summarized as, “how can benefits be maximized while harms are minimized?” However, it is possible for a decision to maximize benefits to you, or your colleagues, while not minimizing harms to others; the opposite decision could minimize harms to others but not create any (or possibly even minimize) benefits to yourself and your colleagues (for more detail on harms and benefits and determining/prioritizing them, see Tractenberg 2019-b). Because a utilitarian perspective is so complicated, we focus here on the virtue approach, enabling us to just focus on what the “ethical practitioner” would do. This ER KSA (identify decision making framework) is absent from the DSEC by design (it is intended “to provoke conversations around issues”), whereas both ASA and ACM emphasize their intended uses in the professional’s decision making.
4. Identify and evaluate alternative actions (on the ethical issue). As noted, the GLs and CE are intended to support ethical decision making. That means there are at least two options, and you (the practitioner) must decide which one is most consistent with what the ethical practitioner would do in this case. Here is an example where the DSEC could really help: choose the method/approach or whatever (depending on where you are in your project or workflow) that leads to a “yes” answer on the corresponding DSEC item. If you’re deciding whether or not to obtain informed consent, or to provide all data contributors with “a clear understanding of the data uses to which they consent”, you can tell (from the DSEC) that NOT doing these things leads to a “no” answer on that DSEC item (A.1) – meaning, if you

do NOT do these things, then your data collection work will not be ethical. Every decision has to be between at least two things: it may be possible to abstract any choice you make in your work/workflow to something that matches what is on the DSEC. This would change a difficult-to-decide situation into more of a yes (I decide on something that creates a “yes” for the DSEC item) or no (I decide on something that creates a “no” for the DSEC item) decision. As we have discussed, “no” on a DSEC item suggests it is unethical. However, it might be impossible, or too costly (in time, effort, money, or some combination) to make a “yes” decision. The crucial thing is, you have *now identified that you cannot make a “yes” decision – and you’ve also discovered why (because of costs)*. Instead of saying to your boss or colleagues, “Hey! You can’t do that!”, you can say, “it looks like we need to X, but according to the ACM/ASA/CDSEC, that could lead to unethical practice or charges that we are unethical. How can we work around this problem so we can X without doing it unethically?” What would the ethical practitioner do if s/he had to also do X? If there is no way for the ethical practitioner to do X, then that is a 3<sup>rd</sup> option: do not do X.

Note that, if you had not made it this far in your ethical reasoning steps, you might have missed this – either done X without considering the negative effects, or done nothing because you guess X is maybe unethical but you’re not sure, so you just skip it or ignore the fact that it might actually be doable *without* being unethical. The next person who needs to do X might not think this carefully about it, and might just do it, even though it is unethical. When you behave in ways that are not stewardly, it either fails to notify others that there is a risk of non-stewardly behavior, or it communicates that there is no reason to be stewardly or ethical. This is an example of how *your* ethical practice strengthens the whole profession: modeling stewardly behaviors opens that option up to observation by others, modeling unethical behaviors makes these appear to be options, which *weakens* the profession.

5. Make and justify a decision. At work, you make hundreds of decisions all the time. Clearly you can’t take the time to think this carefully about every single one of them. The point here is that you have identified an ethical problem (with your ER KSA#2), but not only did you identify it, you also came up with *three* alternatives (1. do it, ignore what the ethical practitioner would have done; 2. find a way to do it in a way that the ethical practitioner would also do; 3. Determine that there’s no way for an ethical practitioner to do X – it can’t be done ethically, so do not do X). You might actually be able to justify all three of these decisions, but clearly two of them are not consistent with the ethical decision making framework you have chosen. You might think to yourself, “Hey, I can just go back to KSA#3 and choose a different decision making framework so what I want to do *is* consistent with that framework!” Remember from section 4 of this article, “It is never acceptable to seek or create “loopholes” in the GLs (unless you also seek to fix/close them). If an unexpected ethical challenge arises, the steward seeks guidance, not exceptions, in the ethical GLs. To try and justify unethical or underspecified behaviors (“no one said I couldn’t!”) is unprofessional, unethical, and not stewardly.”
6. Reflect on the decision. Depending on the case, or the decision you walked through these KSAs to make, there might be a huge amount of reflection. You might re-think the decision periodically throughout your career – new information or technology may come along that reduces the costs associated with doing X ethically, for example. However, in this case we

want to focus our attention on stewardship. Thinking about the next person who has to think through this same decision or a similar one, what additional information would be/would have been helpful to you (was there additional prerequisite knowledge you wish you had)? Thinking about your experience reasoning your way to an ethical decision, does a decision like the one you thought through help to create the culture that promotes fluency in ethical reasoning and/or a more ethical workplace? Obviously, the steward wants the next person who has to make this or a similar decision to think it through and choose the ethical, stewardly option. Reflecting on your decision and decision making process may increase the likelihood of the next decision maker doing so in a stewardly way.

If you encountered resistance at work (or school) as you tried to work through a difficult decision, as you attempted to ensure that *your* practice is consistent with the CE, GLs, the DSEC, or “what the ethical practitioner would do”, reflect on *that*. This article is intended to support your growth and your ethical practice. Not all workplaces, peers, or supervisors will be as interested in your stewardship as you should be, and just as employers can choose the “best” candidate for the data science jobs they have available, you can also choose to work in the “best” and most stewardly climate. The ASA GLs have one Principle that we mentioned in Section 4 but left out of the alignment table with stewardship KSAs (Table 3) - because it is not about you, the steward/ethical practitioner, but is instead about those who employ statisticians and data scientists. This is Principle H, which is presented here in its entirety:

#### **H. Responsibilities of Employers, Including Organizations, Individuals, Attorneys, or Other Clients Employing Statistical Practitioners**

Those employing any person to analyze data are implicitly relying on the profession’s reputation for objectivity. However, this creates an obligation on the part of the employer to understand and respect statisticians’ obligation of objectivity.

Those employing statisticians are expected to:

1. Recognize that the Ethical Guidelines exist, and were instituted, for the protection and support of the statistician and the consumer alike.
2. Maintain a working environment free from intimidation, including discrimination based on personal characteristics; bullying; coercion; unwelcome physical (including sexual) contact; and other forms of harassment.
3. Recognize that valid findings result from competent work in a moral environment. Employers, funders, or those who commission statistical analysis have an obligation to rely on the expertise and judgment of qualified statisticians for any data analysis. This obligation may be especially relevant in analyses that are known or anticipated to have tangible physical, financial, or psychological impacts.
4. Recognize that the results of valid statistical studies cannot be guaranteed to conform to the expectations or desires of those commissioning the study or the statistical practitioner(s).
5. Recognize that it is contrary to these Guidelines to report or follow only those results that conform to expectations without explicitly acknowledging competing findings and the basis for choices regarding which results to report, use, and/or cite.
6. Recognize that the inclusion of statistical practitioners as authors, or acknowledgement of their contributions to projects or publications, requires their explicit permission because it implies endorsement of the work.

7. Support sound statistical analysis and expose incompetent or corrupt statistical practice.
8. Strive to protect the professional freedom and responsibility of statistical practitioners who comply with these Guidelines.

Note that hiring you “... creates an obligation on the part of the employer to understand and respect statisticians’ obligation of objectivity”. Elements 7 and 8 are also specifically relevant here – not because they are binding on the employer, but because they describe what “the ethical employer would do”. Note that DSEC items A (A2, Data Collection), C (C1-C5, Analysis), D (D1-D5, Modeling) and E (E3, Concept Drift) *all* capture features of objectivity that are essential to ethical practice of statistics and data science. Whereas privacy arises in relatively few DSEC items, over half of them are focused on ensuring that the results are objective and robust, but fair and transparent. Features of the DSEC list ensure competent practice in data science, as do many of the ASA GL (e.g., Principle A among others) and ACM CE (e.g., Principles 1 and 2). You are encouraged to be a steward of data science (Tractenberg, 2019-a) and also work for people/organizations who encourage stewardship and ethical practice. When you become more familiar with ethical reasoning, and stewardship, you will also become better able to determine how likely a colleague, supervisor, or employer is to prevent, allow, or encourage you to engage with these essential features of ethical practice. Reflecting on your ethical decision making can lead you to understand your workplace, and the culture in which you practice; supporting your development as a steward of the profession.

## 7. Conclusions

The everyday *practice* of data science, as well as statistics and computing, involves *decisions*. These decisions are expected to be grounded in professionalism and responsibility – in fact, both the ACM and ASA state that their guidance for ethical practice are focused on promoting ethical decision making. The GLs/CE describe *professionalism* –what the ethical practitioner *should do* and not what behaviors to avoid - and outline the professional’s responsibilities.

You should look through the ACM CE and ASA GLs carefully, learning how checklists and other “shortcuts” that may continue to proliferate actually impact your stewardship. Some may be useful tools (like the DSEC) but others may be problematic – and your reflection and application of ethical reasoning KSAs can help you communicate strengths and weaknesses of these approaches to your colleagues. Future and current practitioners in statistics and data science should be familiar with the content of the GLs and CE, and should begin to consider how these GLs may come into play at work– whether they are at an early stage of career preparation or are well into their career. Ethical practice is something to continuously strive for. The practice standards describe the professional habits of all practitioners in statistics and data science; ethical practice in government (e.g., Bureau of Labor Statistics; Census), industry (e.g., business; pharmaceutical; biomedical), social science fields (decision making/marketing; industrial/organizational), and scholarship, among other areas. Ethical reasoning can be challenging, but its KSAs are learnable and improvable, just like the KSAs of stewardship are. Further reading that features applying the ER KSAs in practice with the ASA Ethical GLs specifically is Gunaratna & Tractenberg (2016). That walks the reader through the ethical reasoning in two cases that the ethical practitioner needs to respond to (other people behaving in ways inconsistent with the GLs).



The professional practice standards (ACM/ASA) are careful and complete, and also could be used for many different technological groups including computer science, math, statistics, informatics, and other disciplines where data are involved. As such, learning to reason ethically with either or both of these codes/guidelines will serve as an excellent basis to develop your stewardship as a data scientist.

## References

American Statistical Association (ASA) *ASA Ethical Guidelines for Statistical Practice*-revised (2018) downloaded from <https://www.amstat.org/ASA/Your-Career/Ethical-Guidelines-for-Statistical-Practice.aspx> on 30 April 2018.

Association for Computing Machinery (ACM). *Code of Ethics* (2018) downloaded from <https://www.acm.org/about-acm/code-of-ethics> on 12 October 2018.

Golde CM, & Walker GE. (Eds.). (2006). *Envisioning the future of doctoral education: Preparing stewards of the discipline-Carnegie essays on the doctorate*. San Francisco: Jossey-Bass.

Gunaratna N, Tractenberg RE. (2016). Ethical Reasoning for consulting with the 2016 Revised ASA Ethical Guidelines for Statistical Practice. *Proceedings of the 2016 Joint Statistical Meetings, Chicago, IL*. Pp. 3763-3787.

National Academy of Engineering. (2013). *Practical Guidance on Science and Engineering Ethics Education for Instructors and Administrators*. Washington, DC: National Academies Press.

Rios CR, Golde CM, Tractenberg RE. (2019). *Preprint*. The preparation of stewards with the Mastery Rubric for Stewardship: Re-envisioning the formation of scholars *and practitioners*. Published in the *Open Archive of the Social Sciences (SocArXiv)*, <https://osf.io/preprints/socarxiv/vw7j5/>

Tractenberg RE & FitzGerald KT. (2012). A Mastery Rubric for the design and evaluation of an institutional curriculum in the responsible conduct of research. *Assessment and Evaluation in Higher Education*. 37(7-8): 1003-21. DOI 10.1080/02602938.2011.596923.

Tractenberg RE, Russell AL, Morgan G, FitzGerald KT, Collmann J, Vinsel L, Steinmann M, Dolling LM. (2015) Amplifying the reach and resonance of ethical codes of conduct through ethical reasoning: preparation of Big Data users for professional practice. *Science and Engineering Ethics* 21(6): 1485-1507. DOI 10.1007/s11948-014-9613-1

Tractenberg, RE. (2017). *Preprint*. The Mastery Rubric: A tool for curriculum development and evaluation in higher, graduate/post-graduate, and professional education. Published in the *Open Archive of the Social Sciences (SocArXiv)*, [osf.io/preprints/socarxiv/qd2ae](https://osf.io/preprints/socarxiv/qd2ae)

Tractenberg, RE. (2019-a). *Preprint*. Becoming a steward of data science. Published in the *Open Archive of the Social Sciences (SocArXiv)*, <https://doi.org/10.31235/osf.io/j7h8t>

Tractenberg, RE. (2019-b). *Preprint*. Teaching and Learning about Ethical Practice: The Case Analysis. Published in the *Open Archive of the Social Sciences (SocArXiv)*, <https://doi.org/10.31235/osf.io/58umw>