Data-Centric Engineering (2020), 1: e17
doi:10.1017/dce.2020.13

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

DRAT: Data risk assessment tool for university–industry collaborations

Joanna Sikorska[1], Sam Bradley[2], Melinda Hodkiewicz[1] and Ryan Fraser[2]

[1]School of Mechanical Engineering, The University of Western Australia, Crawley, Western Australia, Australia
[2]CSIRO, Kensington, Western Australia, Australia
*Corresponding author. E-mail: melinda.hodkiewicz@uwa.edu.au
Joanna Sikorska equally contributed to this study.

Received: 19 December 2019; **Revised:** 14 September 2020; **Accepted:** 14 September 2020

Keywords: Data sharing; engineering asset management; industry–academia collaboration; risk assessment; risk management

Abstract

For research in the fields of engineering asset management (EAM) and system health, relevant data resides in the information systems of the asset owners, typically industrial corporations or government bodies. For academics to access EAM data sets for research purposes can be a difficult and time-consuming task. To facilitate a more consistent approach toward releasing asset-related data, we have developed a data risk assessment tool (DRAT). This tool evaluates and suggests controls to manage, risks associated with the release of EAM datasets to academic entities for research purposes. Factors considered in developing the tool include issues such as where accountability for approval sits in organizations, what affects an individual manager's willingness to approve release, and how trust between universities and industry can be established and damaged. This paper describes the design of the DRAT tool and demonstrates its use on case studies provided by EAM owners for past research projects. The DRAT tool is currently being used to manage the data release process in a government-industry-university research partnership.

Impact Statement

The process for obtaining a required dataset from these asset owners can be challenging, even when undertaking approved industry-academic projects. Approval processes differ between organizations. Often the decision is left to individual managers whose attitude to, and motivations for, releasing data for research work vary widely. When data are provided to the researcher, unilateral restrictions on subsequent publication of results and/or data are often imposed, irrespective of the data contents. In the past, these issues have delayed research commencement, limited the pool of academics willing to work on these projects and degraded research outcomes. The DRAT offers an alternative, transparent decision-making process that has been made publicly available under a Creative Commons Attribution 4.0 International License. The tool ensures that (a) recommended restrictions and controls are based on the actual risk posed by a dataset (rather than a one size fits all approach), (b) the data owner's needs for confidentiality are appropriately managed, and (c) the potential for research value is maximized.

1. Introduction

Engineering asset management (EAM) encompasses the processes, systems and human factors involved in managing the life cycle of engineering assets and the systems within which these assets operate. To

demonstrate the impact of EAM research it is necessary to change the way asset management is conducted in asset owning organizations. This necessitates interaction with asset owners, and in the case of data-centric engineering research work, access to an organization's data. This is often undertaken through formal University–Industry Collaborations.

EAM data are held in enterprise resource planning systems, manufacturing execution systems, computerized maintenance management systems, condition monitoring databases, asset registers and supervisory control and data acquisition systems, to name just a few. Access to this data is necessary to build and validate a wide range of models including those developed by reliability engineers, the prognostic health management community and maintenance optimization researchers. To achieve robust validation, data from at least one (preferably more than one) industry partner are required. However, a recent review of process system failure and reliability literature reports that "collaborative works between industry and academia are observed to be rare" (Amin et al., 2019) and we have shown elsewhere that few published models in this field are validated using industry data (Webb et al., 2020). We suggest that data-sharing in the field of EAM is sufficiently problematic as to be reducing the progress of the field by restricting researchers' access to data required to develop effective models for asset life prediction and to improve maintenance management practice.

Over the past 10 years, academia has been seeing an increasing push toward open science; this paradigm is reliant on "open data," with increasing encouragement from publishers and funding government institutions for authors to release the data from which their research conclusions are derived (Sikorska et al., 2016; European Commission, 2020; ODI, 2020). In some industries, such as genomics, astrophysics, epidemiology, and geospatial research, this move has been embraced because it has enabled research that could not have been performed otherwise; the effort and costs to acquire such large datasets are simply too prohibitive for any one institution. A number of initiatives have arisen, such as FAIR-sharing.org (FAIRsharing.org, 2020), EU Support Center for Data Sharing (SCDS, 2020), ISO/IEC JTC 1/SC 4, W3.org, Platform Industrie 4.0 and the European Data Strategy, to guide data producers and consumers on how to enable data sharing. However, most of the available documents prepared by these organizations pertain to the micro-details of how to structure data to enable its collation and use, rather than guides for what and how to share between independent organizations. The challenge of how to assess and manage the risks of sharing data pertaining to asset performance, specifically for EAM researchers, is the focus of this paper.

2. Background

2.1. Data-sharing practice in the engineering sector

All asset owning organizations have formal risk assessment processes and risk management is integrated into everyday practice in most corporate processes. There are six stages in the risk management process presented in ISO 31000 as shown in Figure 1 (Australia, 2018a): (a) scope, context, criteria, (b) risk assessment, (c) risk treatment, (d) monitoring and review, (e) communication and consultation and (f) recoding and reporting . However, there is no specific international standard or widely accepted guidance note on how to apply the processes described in ISO 31000 to sharing data. Information technology standards, such as the AS ISO/IEC 27000 series (Australia, 2018b) tend to focus on the safety of their information while it is under the data owner's control. A recent report by the European Commission claims that "currently there is not enough data available for innovative re-use, including for the development of artificial intelligence…[and]… data sharing between companies has not taken off at sufficient scale" (European Commission, 2020). Similar findings, published by the Open-Data-Initiative suggest that very few industrial companies are willing to share their data (ODI, 2020). Of 2060 industry workers surveyed, only 27% of their employing businesses shared data. Furthermore, the readiness of businesses to share data by having a data strategy which encompassed "a vision for how the company will collect store, manage, share and use data" varied significantly across sectors; the best performers were in the finance and accounting sector, where up to 82% of businesses had a formal strategy (either specific or built into other systems), while engineering design or architecture businesses were the worst performers,

https://doi.org/10.1017/dce.2020.13 Published online by Cambridge University Press

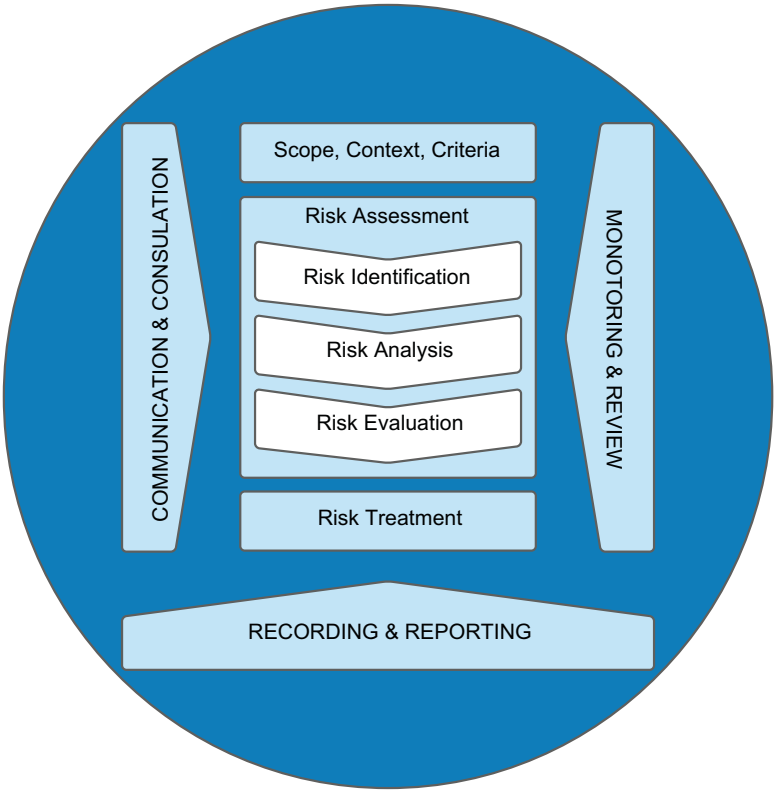

Figure 1. *ISO31000 risk management process (ISO, 2018a).*

with only 36% having any type of strategy. 67% and 60% of manufacturing and construction businesses respectively had some form of strategy. However, having a strategy is not sufficient to enact data sharing, as the survey reported that only 59% of British businesses with a strategy shared data with third parties. This is a problem for researchers, as rich and complete datasets with well-documented meta-data are a necessity for analytical modeling and decision-making. The Open-Data-Initiative survey suggests that engineering focussed businesses are less prepared to share data in other sectors (ODI, 2020). The impact of this is that there are very few examples of EAM data based on real industry data from operating sites publicly available to researchers (Sikorska et al., 2016).

2.2. Data sharing roles in university industry collaborations

Figure 2 considers the roles involved in the process of data release for EAM related data to a University-Industry (UI) research project. These are the *academic collaborator* (data scientist), and their assigned industry contact (*industry collaborator*). Often the industry collaborator is a research project manager and not an EAM subject matter expert (SME). The industry collaborator needs to locate an appropriate SME, identify the data that meets the needs of the academic collaborator and then make contact with the *data custodians*. Once the data sets have been located and retrieved, they need to be risk assessed and a decision made about their release. This decision requires a review by the corporate *legal group* who will then look for a *manager* to take accountability for, and sign off on, the data release. The SME, data custodian, manager and legal group will often have little knowledge of the UI collaboration and, depending on their maintenance and engineering experience, little direct understanding of the raw data. Yet by virtue of their roles in the organizational hierarchy, they all play a role in the decision to approve or reject release of the data.

https://doi.org/10.1017/dce.2020.13 Published online by Cambridge University Press

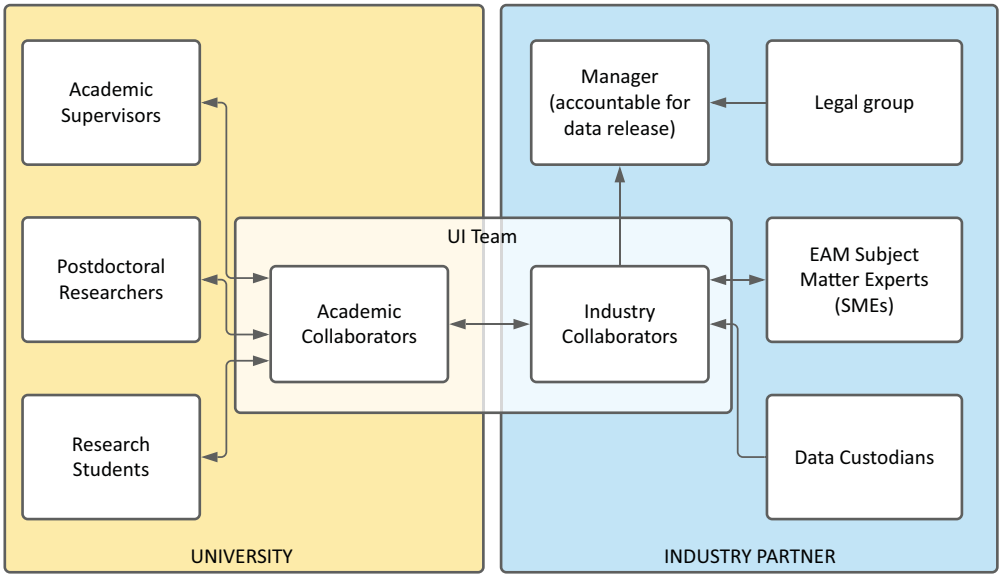

Figure 2. *Team roles in university and asset-owning industry collaborations.*

2.3. Security risks from shared data

Risk is commonly defined as a function of the probability of an adverse event (or threat) occurring and the consequences of that event. When sharing data, the data owner (industry partner) is vulnerable from their data: (a) being maliciously used by the researcher or a member of his/her institution who has access to the data; (b) being lost by the researcher and finding its way into the hands of a malicious third party who understands how to exploit the data; or (c) being stolen by a malicious third party intentionally accessing the information without the researcher's knowledge. In cases (b) and (c), the likely consequences will be less than if the data were accessed from the data owner's facilities, as the third party would have to know where the information originated and the volume of accessible information would usually be much less. The consequences of any of these events are also affected by the type of data, or what it contains. As discussed earlier, EAM research predominantly relates to information about equipment performance and maintenance history. Personal private data are rarely used and thus should be assessed on a case by case basis. For obvious reasons, there is no agreed list on what business drivers are important to all organizations as the potential economic impact of an information security incident must ultimately be speculative (Cashell et al., 2004).

There is no published research into the relative importance of potential threats to an organization from lost asset-related data. Instead, we propose that information relating to health, safety and environmental variables, which are often collected for regulatory compliance purposes, are deemed the most sensitive because they could pose the most serious consequences to the data owner; should they be found to be in breach of any health, safety, and environment (HSE) operating requirements, the company could be fined, suffer from bad publicity and lose customer confidence, or even potentially lose their license to operate. The next most damaging type of data would be information that pertains to current operational capabilities or financially sensitive data that may jeopardize a company's competitive advantage or its relationship with third parties. Older operational or financial data, that is, no longer current is of less importance, with maintenance data pertaining to common and nonoperationally critical assets being of least value to third parties. Information that can easily identify the company increases the consequences to the data owner, especially when coupled with other higher risk categories (HSE, current operational data) and hence increases the risk. Unfortunately, actual risks are not the only factor in decision-making.

https://doi.org/10.1017/dce.2020.13 Published online by Cambridge University Press

2.4. Concerns in data sharing

Reasons for reticence in data sharing between organizations, organization-to-organization and organization–researcher groups include: (a) inadequate economic incentives, (b) lack of trust, (c) fear of data misappropriation by third parties (d) loss of privacy, particularly the risk of re-identification, (e) lost or reduced intellectual property value of their data, (f) negative consumer/customer/user reaction and (g) risks to regulatory compliance associated with how researchers use their data (Bruneel et al., 2010; Harris and Sharma, 2017; European Commission, 2020).

Data associated with personal data are of specific concern to management as organizations operating in Europe collect and manage personal data in accordance with The General Data Protection Regulations (GDPR) (GDPR, 2018). Similar legislation exists in non-EU countries, most of which impose heavy fines for data breaches (Ponemon Institute, 2019). Consequently, the bulk of published literature pertains to security management to prevent loss of employee or customer private data. In this paper, however, we are explicitly concentrating on physical asset data and therefore considerations of GDPR are not considered.

Another cause for reticence in sharing data occurs at a more practical level. Costs associated with the time and effort required by company employees to locate and extract the data, often from disparate systems, as well as supply and document associated meta-data, can be considerable. Thus, industries in which data are mostly highly automated, well-structured and easily understood are more likely to share data (Couture et al., 2018; ODI, 2020). Similarly, individuals are less likely to want to share data if it is labor-intensive to do prepare (Borgman, 2012). As noted earlier EAM data come from numerous data sources and is often very difficult to aggregate and assimilate (Webb et al., 2020). Some of it can even be kept in proprietary systems and thus not possible to extract in an appropriate form.

ISO 31000:2018 is a risk management framework widely used across all industries (ISO, 2018a). Its main process is shown in Figure 3. The ISO/IEC 27000:2018 (ISO/IEC, 2018b) is a set of more than five standards dealing with Information security management and is based on the principles of ISO 31000. The appendix of ISO 27005 helps users scope their assessments in terms of assets, threats and best practice, and is seen as best practice in information security risk management (Wangen, 2017). NIST Special Publication 800-30 Revision 1 is often used in conjunction with ISO27005 to guide risk assessments (Stoneburner et al., 2002). Other approaches are also available and have been reviewed extensively in (Gritzalis et al., 2018). Most of these describe general principles and do not specify how the risk assessment should be performed (Sendi et al., 2010). Engineering techniques, such as Hazard and Operability Analysis (HAZOP), Failure Modes and Effects Analysis (FMEA) and quantitative risk assessment (Cameron et al., 2005; Smith, 2017), are also sometimes used by IT professionals for assessing information associated risk within their organizations and are incorporated into some of the processes listed. However, in such cases, the process for assessing risks is based on a detailed desktop study by experts that takes into account all the possible ways that the data could be accessed and used nefariously, perform probability assessments on each, and then predict the consequences of each event occurring. This is a lengthy and hence costly process, so is not normally performed on an individual piece of data. It is too complex and onerous a process for approving research data requests that must be done quickly with the limited imposition on the industry partner.

2.5. Establishing accountability for data release

The concept of responsibility is central to the notion of collective agency and organizations. If something goes wrong, individuals are identified as being responsible for the bad event; they can then be held "accountable" and "blamed" (Grossi et al., 2007). Releasing data to research partners requires one individual to take responsibility and be held accountable for approval. In the absence of a formal organizational process for this, an individual *manager* has to perform their own risk assessment. When unguided and without support, the accountable manager's decision will be determined by factors such as agency, expertise, risk profile, personality, prior experience working with academia, and level of trust in both the project and its academic collaborators (Dirks and Ferrin, 2001; Bruneel et al., 2010; Dingler and Enkel, 2016; Tan, 2016; European Commission, 2018). Figure 3 shows the accountable manager's

https://doi.org/10.1017/dce.2020.13 Published online by Cambridge University Press

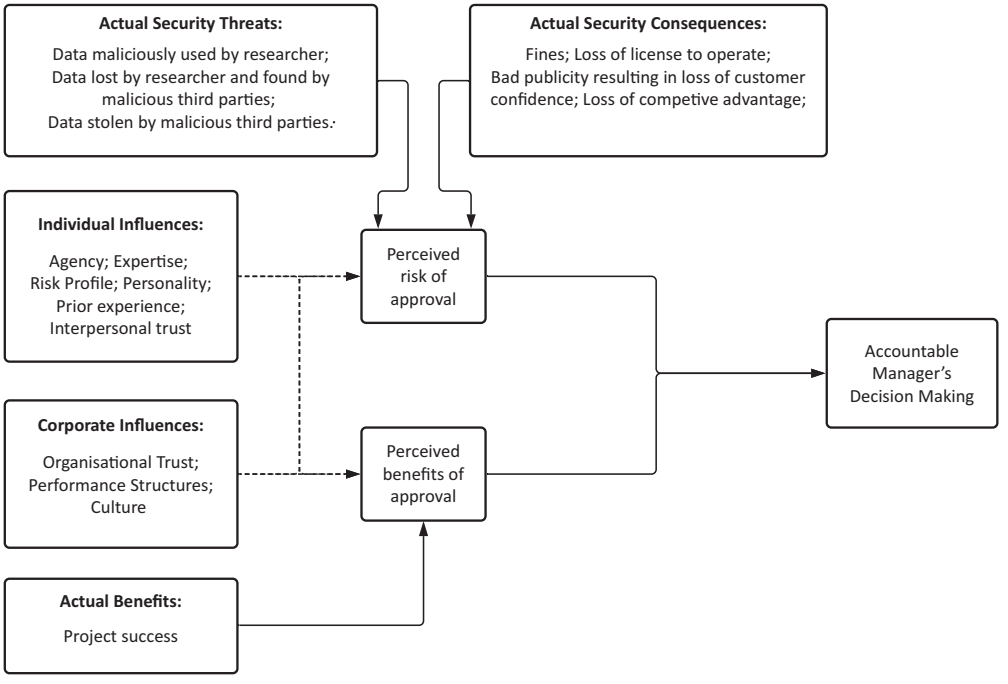

Figure 3. *Influences on the accountable manager's decision-making for the release of EAM data to researchers.*

perception of benefit and perception of risk as two separate constructs. These perceptions are driven by the individual factors mentioned and moderated by organizational factors such as intra-organizational trust and the performance management structure; a rigid performance management structure and low levels of intra-organizational trust will drive the individual manager to perceive that approving the data's release carries a higher level of risk (than in organizations with high trust and flexible performance structures).

2.6. Trust and control in university–industry collaboration

In collaborative projects, it is important that entities have a high degree of perceived certainty that their partners will cooperate satisfactorily and not behave opportunistically. This is known as "partner cooperation" and is comprised of two elements: inter-organizational trust and control (Das and Teng, 1998). In an alliance, control is the "regulatory process by which the partner's pursuit of mutually compatible interests is made more predictable" (Das and Teng, 1998). Contracts and legal agreements are an important part of this landscape. In UI projects, these agreements describe how IP is to be shared as well as the level of editorial input required prior to publication and whether data used therein can be released. Although time consuming to negotiate, these agreements provide some certainty for both parties and establish expectations.

Trust (be it inter-organizational, intra-organizational or inter-personal) is defined as the "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party" (Mayer et al., 1995, p. 712) and is critical in enabling cooperating behavior (Gambetta, 1988). It involves both choice and an element of risk (Vlaar et al., 2007). Trust is also highly reciprocal: trust evokes trust, distrust evokes distrust and as the level of trust reduces it is replaced with mistrust (Fox, 1976). Research has shown that higher levels of intra-organizational and inter-personal trust are associated with greater knowledge sharing, greater willingness to collaborate and more pertinently, reduced barriers to releasing datasets and supporting information (Dirks and Ferrin, 2001; Gopalakrishnan and Santoro, 2004; Bruneel et al., 2010; Dingler and Enkel, 2016; Tan, 2016; European Commission, 2020).

https://doi.org/10.1017/dce.2020.13 Published online by Cambridge University Press

Organizations, and the individuals within them, are more likely to share data with researchers and institutions they have worked with previously, or whose reputation is well known (Bruneel et al., 2010). Methods to increase trust include regular communication, team building activities and shared tasks so that the team can develop a shared culture as well as and an understanding of each other's abilities.

Whether trust and control are contradictory or complementary factors (Das and Teng, 1998; Cristina Costa and Bijlsma-Frankema, 2007) is a matter of some controversy among academics. Some researchers argue that high levels of control can be perceived as an indication of mistrust (Argyris, 1952; Sitkin and Roth, 1993), while others argue that control mechanisms if used properly, can help build mutual trust (Goold and Cambell, 1987; Sitkin, 1995; Edelenbos and Eshuis, 2012). Others suggest that this is because there are different types of control measures, and controls only erode trust if the method is not appropriate for the situation (Kirsch, 1996; Sitkin and Stickel, 1996; Evans, 2010). While output controls (prescribing specific performance goals) can negatively affect trust, process controls (rules, goals, procedures, and regulations that specify desirable patterns of behavior) and social controls (utilizing organizational values, norms, and cultures to encourage desirable behavior) improve trust levels (Aulakh et al., 1996). The latter (social control) seems to be the most effective at improving trust because these "soft" behavior guidelines, associated with terms such as "recommended" or "informal," require an element of personal judgement on behalf of the trustee; these controls do not stipulate what must be done but rather what should be done based on shared goals, values and norms. However, like trust, social controls can only be implemented slowly as it takes time to nurture a sense of shared responsibility, culture and community. It is thus likely that process controls can be particularly beneficial in the early stages of a relationship before social controls are possible and higher levels of trust have been developed.

The propensity of companies to enter UI collaborations has been shown to differ across sectors (Bekkers and Freitas, 2008). Companies belonging to sectors that do not change rapidly, such as mechanical and civil engineering, are in fact much less likely to collaborate with universities (Meyer-Krahmer and Schmoch, 1998; Schartinger et al., 2002), when compared to more R&D intensive manufacturing industries such as pharmaceuticals, biotechnical and aerospace. Companies in rapidly changing industries are much more likely to perceive a benefit from collaborating with universities and transferring up to date knowledge as they rely that on academic knowledge and basic scientific research for developing their next products or services (Bekkers and Freitas, 2008). Thus, it is expected that such companies would enter a new collaboration with higher levels of initial trust than those working in industries where the short-term benefits are more ambiguous.

Another reason for low levels of trust in a new UI collaboration may be due to the different motivations and environmental drivers of the collaborating parties. Universities (and government-funded research institutions) are primarily driven by the need to create new knowledge and share the results as widely as possible for the greater scientific good; or in the words of the UK Research Institute: "to ensure everyone in society benefits from world-leading research and innovation" (UKRI, 2020). As international university ratings are strongly affected by citation rates and research impact (QS World University Rankings, 2020). Regularly reducing a university's capacity to publish in high-quality journals will eventually impact its financial performance as it has been shown that the lucrative international student market is affected by a university's ranking (Griffith and Rask, 2007; Horstschräer, 2012; Broecke, 2015; Gibbons et al., 2015).

This is also true for individual researchers. University academics must publish prolifically in reputable journals in order to increase their personal reputation and standing within the scientific community, obtain promotion within their research institution and to obtain research funds for future projects (Piwowar et al., 2007; Evans, 2010; Kim and Zhang, 2015). Recent evidence also suggests that publishing supporting datasets increases academic citation rates (Piwowar and Chapman, 2010; Henneken and Accomazzi, 2011). Therefore, undertaking projects with publication restrictions can limit an academic's future career opportunities. Increasingly, journal publishers are requesting that supporting datasets be made available so research can be verified or replicated, although this is still not widely practised because individual reward mechanisms considerations for academics associated with keeping data for private use and selective distribution are stronger than the drive to open-science (Fecher et al., 2017). Data sharing

https://doi.org/10.1017/dce.2020.13 Published online by Cambridge University Press

requirements by publishers are also much less prevalent in engineering than in other sectors (Naughton and Kernohan, 2016; Wiley, 2018; Wiley, 2020). It should also be mentioned that even when journals require that data be shared and the authors do so, research has found that much of the data are not subsequently available (Nelson, 2009; Fecher et al., 2015; Fecher et al., 2017).

Businesses, on the other hand, need to maintain a competitive advantage, which necessitates that pertinent knowledge remains private long enough for it to be leveraged (Evans, 2010). Thus, by necessity, companies are usually reticent to publish results, or wish to do so in a way that maximizes the advantage to themselves as the funding (or provisioning) organization, while minimizing the benefit to competitors. They also need to obtain results much more quickly (Alexander et al., 2018). Industry researchers, unsurprisingly, publish much less frequently in journals and are more likely instead to publish their ideas as patents (Anthony et al., 2004), for which supporting data is never required. They are also less likely to share information with one another than their academic colleagues (Haeussler, 2011). This is not necessarily because of any inherent differences in the personalities of the researchers but because they are exposed to different constraints and competitive interest considerations (Anthony et al., 2004; Haeussler, 2011).

The attitudinal misalignment between industry and academic partners can result in some degree of mistrust between collaborating parties, especially when the collaborative team is newly established and the organizations have not worked together before (Bruneel et al., 2010). This naturally limits knowledge sharing and full disclosure until parties are assured their interests will be adequately satisfied by the partnership. We propose that process controls can be beneficial in establishing expected behavior early in the collaboration.

3. DRAT tool development

3.1. Requirements

Taking into account the issues described above and in Table 1, we propose the following requirements. These requirements are linked to the factors identified above. Specifically, the need for a process that provides a decision tree for managers based on a set of questions, a process, that is, familiar (in this case, ISO 31000), is independent of personalities and prior working relationships with industry researchers and provides an established and approved process for the manager to follow.

Table 1. Requirements for a process to support data sharing for EAM data between universities and industry.

Identifier	Requirement	Link to factors in Figure 2
T1	Enable an individual manager (the user), within an organization owning EAM data, to identify and assess the risks associated with sharing a specific data set with a research partner(s).	**Expertise**: enables a manager without direct expertise of the data or its context to ask specified questions and elicit answers in a known and approved framework.
T2	Enable risk controls to be applied in a consistent way.	
T3	Conform to the processes in the ISO 31000:2018 Risk Management Standard.	**Agency**: Use of a known and well-understood risk management framework (ISO 31000)
T4	Do not require the user to share any information about the data they are collecting with the researcher prior to risk assessment.	**Trust:** Does not require trust in the early project stage
T5	Enable the assessments made by the user to be captured and retained for auditing.	**Organizational performance management practice**: The decision is transparent and documented, and done in accordance with an established corporate procedure.

Abbreviation: EAM, engineering asset management.

https://doi.org/10.1017/dce.2020.13 Published online by Cambridge University Press

3.2. Process

We propose a tool to guide the manager and others involved in the data release process shown in Figure 3. The data risk assessment process is conceived as a flowchart that asks a series of yes/no questions. Flowcharts are widely used in business processes to guide decision-making. The DRAT tool follows the steps in the ISO 31000:2018 standard starting with risk identification, moving onto assessment treatment and recording, and envisages feedback loops for communication and consultation and monitoring and review.

3.3. Risk identification and assessment

Firstly, questions are asked about the type of data that has been requested.

(a) Does the dataset in the rawest available form (RAF) contain personal data about people? (This is to clearly exclude personal data from the analysis and highlight to the reviewer that this must be analyzed separately).
(b) (If no) Does, the RAF dataset contain data about equipment, people or processes required to ensure HSE or regulatory compliance?
(c) (If no) Does the RAF dataset contain highly sensitive cost or operations data?

Depending on the type of data, other questions are asked about the importance of the data to the project and/or the ability to anonymize the data without affecting the usefulness of the research work. If the data do contain sensitive information, the user is asked whether the data contain potentially less sensitive data as well (i.e., safety and cost data). Finally, questions are posed regarding the desire and ability to anonymize the specific data attributes. The final risk ranking is based on the cumulative residual risk. We have taken the approach that anonymization should be performed as late in the process as feasible, to maximize research analysis options.

This DRAT process classifies company dataset risk into five grades from Highly Safe (with no perceived risk of adverse events from data release) Corporate data, to Unsafe (very high risk of adverse events from data release) Corporate Data. An additional level is also identified for Potentially Unsafe Personal that should be analysed separately. The resulting classifications are summarised in Table 2.

The next stage is to establish risk treatment, also called controls. To do this we considered a number of approaches (such as NIST SP 800-53:Appendix D (Stoneburner et al., 2002), CRAMM (CRAMM, 1987), OCTAVE (OCTAVE, 2006), and settled on the Five Safes framework as it was the only one specifically designed for sharing data, and did not require a quantitative evaluation of risk factors. The Five-Safes framework is a well established system-based approach developed by the UK Office for National Statistics in 2003 (Desai et al., 2016; Ritchie, 2017) and is specifically designed for sharing personal data. Initially, it was used mostly by statistical agencies (e.g., the ABS) and social science academics but has since been more widely adopted by government, health and private sector bodies to share inter-agency data. For example, the approach was recently used by the Data Taskforce, led by the Australian Computer Society and the NSW Data Analytics Center (with input from various Australian government agencies and Corporations) as the basis for its white paper describing a framework to support automated data sharing between governments and institutions in an Australian privacy context (ACS, 2017). The Five Safe's framework is made up of five "safe" dimensions, each of which describes an independent but related aspect of disclosure risk as follows.

Safe data—Has appropriate and sufficient protection been applied to the data to prevent the identification of an individual company?

Increasing the safety (reducing the risk of identification) can be undertaken by ensuring that the minimum cohort size (number of individuals with the same features) is greater than one, aggregating the data so that attributes as less diverse (e.g., grouping attribute values into broader categories so that the minimum cohort size is increased) or anonymizing the data. This can be considered when working with multiple organizations within one research consortium or when undertaking private projects. However, often a company can be identified by association with a government grant that must be acknowledged; in

https://doi.org/10.1017/dce.2020.13 Published online by Cambridge University Press

Table 2. Summary of risk rankings acronyms used.

Safe data classification	Risks usually come from	General comments	Example data attributes
Potentially unsafe personal data.	Data attributes that could be used to identify individual persons either directly, indirectly or by combining the data with other reasonably available information. This may include persons who have died within the last 30 years.	Avoid if possible. If absolutely necessary (due to possible gleaned value) requires a separate risk assessment.	Average repair time of individual maintainers, age of maintainer, sex of maintainer, shift, certifications.
Unsafe (very high risk) corporate data.	Dataset contains attributes that could be used to identify data owner and could definitely cause significant losses (financial, reputation) if released unwittingly. Dataset may span multiple locations so could be advertised as relating to the entire company and not an isolated case. Type of data would be of interest to third parties so nefarious use could be expected if data was obtained by public in the rawest complete form.	Avoid if possible. Manage with utmost confidentiality if access to raw data is justified by the potential value of project.	Company-wide emissions data in sensitive locations, cost data and raw data relating to "bad actors" associated with HSE/regulatory trends that have not yet been reported to the public. Technical data of designs that could be replicated by a competitor.
Low level of safety (high risk) corporate data.	Data attributes that could be used to identify and could cause medium level losses (financial, reputation) if data released unwittingly and happened to be used nefariously. Data could be a combination of cost data and data relating to "bad actors" associated with HSE/regulatory problems that may not be satisfactorily anonymized. Data relates to multiple sites/locations at the company.	Avoid if possible and select other assets, processes or procedures. All efforts should be made to anonymize, obfuscate or aggregate the data by the researchers early in the project so that it could be more widely shared unless this compromises the analysis that can be performed.	Emissions data from multiple plants, sensitive supply agreement data, ore/feedstock properties (if not widely known), individual HSE incident data that has not yet been reported.
Moderately safe (low-medium risk)' corporate date	Data attributes that collectively might be used to identify data owner and could cause minor embarrassment to company if released carelessly and happened to be used nefariously.	Data would probably only be made publicly available in anonymized or semi-anonymized form so publications should be prepared accordingly.	Technical/O&M information of uncommon or slightly customized assets or assets not used by other researchers. Supply agreement cost data.
Safe (low risk) corporate data.	Minimal risk to data owner of either dataset containing information damaging to data owner or of interest to third parties, or of data owner being identified.	Data available to all researchers with no specific requirements on data storage or sharing among relevant parties (subject to contractual agreements).	Technical/O&M data of commonly used assets. Total job-specific maintenance cost data.
Highly safe (no risk) corporate data.	No or only positive effects are expected to data owner if data is released and data owner identified.	There should be no restrictions on the publication of analysis of data or data but researchers need to check with any contractual agreements by which they are bound.	Maintenance only data relating to common assets used widely in multiple industries.

Abbreviations: HSE, health, safety, and environment; O&M, operation and maintenance.

this, it must be assumed that identification will occur. As safety decreases (risk of identification increases), other "safes" need to be increased to manage the overall risk.

Safe projects—Is the data to be used for an appropriate purpose?

The less safe the data (i.e., more likely identification can occur), the more scrutiny needs to be undertaken in order to ensure that data is only used for pre-approved projects. Conversely, for data, that is, deemed completely safe (i.e., re-identification is not possible or is permitted and/or desired) then there is no risk of making data publically available and projects do not need to be preapproved (nor can they be once the data is public).

Safe people—Is the researcher appropriately authorized to access and use the data?

This element relates to the level of trust that the data owner places in a researcher to ensure that the dataset is managed in accordance with the agreed processes and procedures, both during the research and once it has been concluded. For projects and datasets deemed less safe, additional requirements or restrictions may be placed on who can access the dataset. This may include additional training, the signing of specific confidentiality agreements, or even government-approved checks (e.g., security clearances).

Safe settings—Does the access environment prevent unauthorized use?

This element refers to how the data can be obtained and the practical controls to ensure the data are not inadvertently or purposefully released. Data shared on publicly available internet sharing sites will have a very low safe setting level, while at the other extreme, data, that is, only available on a single computer in a locked room with no internet access and with armed guards at the door to check that mobile devices are not taken into the room, will have a very high safe settings level. The appropriate level will depend on the levels of the aforementioned elements.

Safe outcomes—Are the results nondisclosive?

This element considers the residual risk that the company owning the data are, and does not want to be, specifically identified as the source of the research outcomes. It is predominantly dependant on the safe data level. When less safe data are being analyzed, outcomes will most likely require aggregation, anonymization and obfuscation before they can be published; this may also require confirmation by expert statisticians or the data owner to verify the company cannot be recognized.

Table 3 describes for each level of data classification, minimum controls that should be considered against each of the Five Safe's dimensions. It is important to recognize that these suggested controls are the minimum standards proposed to manage the risk appropriately. If contracts between the research parties contain specific requirements that exceed these levels (e.g., that all outputs are approved prior to publication), then the more stringent control must of course be applied.

To facilitate easier use, as well as ensure traceability and repeatability, a web-based tool was developed to enact this process. It presents the questions as requiring YES or NO answers, and offers guidance in cases where the user is not sure of which answer to select. The results and recommendations are saved in a pdf format, which can then be submitted for final approval as per the company's usual risk management processes. This is available at https://drat-process.com.au/home.

4. Case study

The case describes a research project on the potential to set regionally specific asset planning performance targets for waste water blockages (Green et al., 2016). Current practice is to have one corporation-wide target but there are widely different situations in the regions with some regions always well below the target and others well above. The overall performance of the corporation in wastewater blockages per 100 km is reported to their economic regulator. Data considered necessary for the project included pipe age, type, network length, population, rainfall, capital spend, renewals program plans, blockage events, work orders to identify work planned and work executed and maintenance activity type/cost, failure causes on eight geographically dispersed business units.

We demonstrate the DRAT tool in this case, showing the results of the assessment in Table 4. This resulted in an assessment at the Unsafe Corporate Data level. The DRAT suggested controls are virtually identical to those imposed at the time of the research. In this case study, the primary risk had been due to

https://doi.org/10.1017/dce.2020.13 Published online by Cambridge University Press

Table 3. Minimum controls for each data risk classification against the Five Safe's dimensions.

Safe data classification	Minimum controls: safe projects	Minimum controls: safe people	Minimum controls: safe settings	Minimum controls: safe outputs
Potentially unsafe personal data	Data project to be preapproved by company and Research Center Leadership Team to confirm worth of project.	As determined by a separate risk assessment. This is outside the scope of the DRAT tool.		
Unsafe (very high risk) corporate data	Data project to be preapproved by Company and Research Center Leadership Team to confirm worth of project.	Individual researchers to be preapproved by Company. No unauthorized collaboration. Specific confidentiality agreements may be required. Training for students and supervisors on data security and confidentiality required.	Dataset only to be viewed on company premises or in Company approved locations and not put in the cloud, emailed or carried on USB.	Rerun DRAT on final anonymized/obfuscated dataset to identify whether data, i.e., required to present or explain results still poses the same risk ranking. If so, only present results to a limited set of data owner's employees, selected by the Company. Otherwise, senior researchers to verify that outputs cannot identify which Research Center Partner Company participated in work. Anonymized outputs to be vetted by Company AND Research Center leadership team prior to release or publication of any results. Dataset will never be released in the original form. No sharing with other Research Center partners without the specific agreement of Company.
Low level of safety (high risk)' corporate data.	Data project to be preapproved by company and Research Center Leadership team to confirm worth of project.	Data only made available to a small number of identified researchers in the Research Center. Training for students and supervisors on data security and confidentiality required.	Data to be stored on restricted-access servers as approved by the Research Center and not put in the cloud, emailed without encryption or carried on USB.	Outputs would unlikely be approved for publication without significant anonymization and obfuscation and supporting data would not be released to third parties for verification. DRAT should be replied after anonymization to confirm that risk ranking has reduced, prior to requesting permission to publish. Otherwise no sharing with other Research Center partners without specific agreement of Company.

Table 3. *Continued*

Safe data classification	Minimum controls: safe projects	Minimum controls: safe people	Minimum controls: safe settings	Minimum controls: safe outputs
Moderately safe (low-medium risk)' corporate date	Data project to be preapproved by Research Center Leadership Team.	Raw data available to all Research Center researchers (not other industry partners) but to be managed with confidentiality.	Should not be emailed (use link to secure location instead). Avoid carrying on USB or use encrypted USB.	Publications and anonymized/obfuscated datasets to be checked by data owner(s) prior to release, but data owner(s) would not usually restrict publication of either research outcomes or supporting data in anonymized/obfuscated form. Recommend rerunning DRAT after anonymization to confirm that risk ranking has reduced. Can be shared with Research Center partners after anonymization/obfuscation.
Safe (low risk) corporate data.	Subject to data owner's approval, data could be used for similar projects.	Data available to all Research Center researchers (subject to contractual obligations).	Password protect data files when emailing. Data can be analysed by researchers from any private location.	Depending on contractual obligations there is no reason to restrict publication of data analysis. Approval required to release anonymized/obfuscated/aggregated dataset but approval would be expected by researchers. Level of anonymization or obfuscation would depend on data owner, results of analysis and data type. Can be shared with Research Center partners after anonymization (unless Company agreed that it does not wish to remain anonymous) and subject to contractual obligations.
Highly Safe (no risk) corporate data.	Subject to data owner's approval, data could be used for any other project.	Subject to data owners approval data can be analysed by any student or researcher in the Center and may be released externally	Data can be used from any location.	Depending on contractual obligations, but there is no reason to restrict the publication of data analysis or even dataset. Right to publish freely would be expected by researchers, however, approval should still be confirmed with data owner prior to release.

Abbreviation: DRAT, data risk assessment tool.

Table 4. Case study demonstrating the DRAT process.

Assessment:	
1. Does the dataset in the RAF contain personal data about people?	No
2. Does the RAF dataset contain data about equipment, people or processes required to ensure HSE or regulatory compliance?	Yes
3. Have these assets, processes or procedures been associated with known bad events?	No
4. Is there a "greater good" to the company or community from analyzing these specific assets, processes or procedures?	Yes
5. Does the RAF contain highly sensitive cost or operations data?	Yes
6. Is the sensitive data necessary to obtain research value?	Yes
7. Will excluding recent records make it less sensitive and still be useful for the research intent?	No
8. Can the sensitive data be scaled or a differential privacy algorithm applied and it still provide research value?	No
9. Will the sensitive data be used with similar data from other companies?	No
10. Does the RAF dataset contain attributes that could identify the data owner specifically?	Yes
11. Does the owner want to potentially remain anonymous or keep some/all of the data anonymous?	Yes
12. Can the dataset eventually be anonymized without reducing the value of the research results?	Yes
Final assessment category: Unsafe corporate data Data attributes that could be used to identify data owner and could definitely cause significant losses (financial, reputation) if released unwittingly. Dataset may span multiple locations so could be advertised as relating to the entire Company and not an isolated case. Type of data would be of interest to third parties so nefarious use could be expected if data was obtained by public in rawest complete form. Avoid if possible. Manage with utmost confidentiality if access to raw data is justified by potential value of project.	
Suggested controls:	
Safe data:	Data should be anonymized as early as possible.
Safe projects:	Data project to be preapproved by Company and Research Center Leadership Team to confirm the worth of project.
Safe people:	Individual researchers to be pre-approved by Company. No unauthorized collaboration. Specific confidentiality agreements may be required. Training for students and supervisors on data security and confidentiality required.
Safe settings:	Dataset only to be viewed on Company premises or in Company approved locations and not put in the cloud, emailed or carried on USB.
Safe outputs:	Rerun DRAT on final anonymized/obfuscated dataset to identify whether data, that is, required to present or explain results still poses the same risk ranking. If so, only present results to a limited set of data owner's employees, selected by the Company.

https://doi.org/10.1017/dce.2020.13 Published online by Cambridge University Press

Table 4. *Continued*

Assessment:	
1. Does the dataset in the RAF contain personal data about people?	No
	Otherwise, senior researchers to verify that outputs cannot identify which Research Center Partner Company participated in work. Anonymized outputs to be vetted by Company AND Research Center leadership team prior to release or publication of any results. Dataset will never be released in the original form. No sharing with other Research Center partners without the specific agreement of Company.

Abbreviations: DRAT, data risk assessment tool; RAF, rawest available form.

the presence of town names in the original dataset; these were replaced with identification letters A–H very early in the project as there was no analytical disadvantage in doing so. Another source of risk had been cost data in the original dataset, however, as this was not used in the analysis, it was removed and permission was then provided to publish the results and share the dataset among other researchers at the university.

5. Testing

Testing involved an assessment of the risks and controls involved in a number of prior industry-funded collaborative projects, involving the release of data, undertaken by the authors in the last decade. The range of resulting risk rankings are shown in Table 5.. In all cases, the suggested controls were very similar to those that were agreed with the respective company partner through significant discussion and negotiation with the industry partner as the project progressed. In each case, there was already an established relationship (and trust) between the academic supervisor, industry collaborator and SME (roles shown in Figure 2) and each had a good knowledge of the data and subject matter. It is worth noting that the actual actions (shown below) were not decided a priori at the start of the project as they would be under the use of the DRAT tool. The aim of showing these projects is to say that the DRAT recommended actions broadly map onto actions taken in mature, consenting university industry collaborations. For obvious reasons, we are not permitted to talk about data and publication requests which have been denied by industry collaborators.

The second stage of testing involved introducing the DRAT tool to industry partners of the recently formed Australian Research Council Industrial Research Training Center (ITTC) for Transforming Maintenance Through Data Science. The industry partners are BHP, Alcoa and Roy Hill. BHP and Alcoa are global mining companies with revenue (2019) of USD 44.3 billion and USD 11.6, respectively. Roy Hill is owned by a consortia of private and publicly listed companies. The industry partners circulated the process within their organizations for comment, including to their legal departments. The tool is now in use within this ARC ITTC and is being used regularly to assess data sets for release to the research partners in the Center.

In practice, the DRAT tool has brought the following benefits to the industry and university partners in the ITTC.

- The web tool is easy and quick to use by industry employees without any training.
- Feedback from industry partners indicates that the DRAT process provides more consistent assessments and the summary information facilitates quicker approval by management.
- The datasets that have been are provided have not been significantly manipulated improving their value to researchers.

https://doi.org/10.1017/dce.2020.13 Published online by Cambridge University Press

Table 5 Application of the DRAT tool to prior research projects.

Project title	Raw dataset required	Uncontrolled risk ranking	DRAT Recommended action	Actual action
Classifying machinery condition using oil samples and binary logistic regression (Phillips et al., 2015)	Raw and expert classified oil analysis results for mobile mining equipment engines	Safe corporate data	Control by data owner: No restriction on the publication of results. Approval from data owner required prior to publication of dataset.	**Publication** approved. No company or specific asset identifiers mentioned. **Data** approved for release and publically available on prognostics data library
Setting asset planning performance measurement Targets (Green et al., 2016)	**Pipe** age, type, network length, population, rainfall, capital spend, renewals program plans, blockage events, work orders to identify work planned and work executed and maintenance activity type and cost, failure causes. Data for 8 geographically dispersed business units.	**Unsafe** corporate data	**Control** by data owner: **Company** to prune dataset prior to release to exclude most recent records but do not otherwise manipulate data. **Check** that no undesired identifiers remain prior to publication. No approval from data owner for publication of dataset. **Control** by researchers: Data access to researchers is provided in RAF. Researchers must anonymize/obfuscate/aggregate dataset (or only part thereof if agreed by data owner) prior to preparing any results for publication	Publication approved. No data released. Paper co-authored with company. Business units not named.
Mining company wear conveyor monitoring (Webb et al., 2020)	Data for each conveyor in the study: Operating hours and tonnages (movements), Conveyor belt design data, Conveyor usage, Wear thickness reports.	**Unsafe** corporate data	Control by data owner: Company to prune dataset prior to release to exclude most recent records but do not otherwise manipulate data. Check that no undesired identifiers remain prior to publication. No approval from data owner for publication of dataset. Control by researchers: Researchers must anonymize/obfuscate/aggregate dataset (or only part thereof if agreed by data owner) prior to preparing any results for publication	Publication approved. No company or specific asset identifiers mentioned. No data released.
Sewer blockage prediction (Xie et al., 2017)	Pipe attributes: pipe age, diameter, length, slope, depth, installation month, installation decade, joint type, pipe material (VC pipes only). Failure data: blockage date, cause of blockage Environmental context: groundwater level, submergence depth, pipe location, soil type, distance to various road types, distance to railway. Dates: 2006–2013 Data located in SAP and GIS systems	Safe corporate data	Control by data owner: No restriction on the publication of results.	Publication approved. No data released but data made available to the University to support other statistical projects. Paper co-authored with company.

Abbreviation: DRAT, data risk assessment tool.

- The process facilitates transparent communication between all parties regarding the control measures that must be implemented to manage the risk appropriately.

Here is a quote from one of the partners involved:

> "We have evaluated DRAT on data sets we had previously agreed to release and the risk mitigations suggested by the tool are coherent with our prior assessment. We suggest the tool has use wider applicability than just assessing data for use in university collaborations, such as for data released as part of contractual and consulting work as well."

6. Conclusions

This paper describes a process, called DRAT, that enables industry partners to assess and control the risks associated with releasing EAM datasets to university research partners. Existing risk management processes used by the sector do not specifically cover the sharing of data with universities. The tool is designed to enable a manager without direct expertise in the details of the data or the EAM process or familiarity with the university collaborator to use a well-known and risk management framework (ISO 31000) and an organisationally approved decision tree to decide on the risk posed by the release of the data and suggest controls based on the Five Safes approach. This process is transparent and documented, relieving the manager of risks that he/she will have been seen to have made an ad-hoc decision which cannot subsequently be justified. The DRAT is enacted using a web-tool, that is, freely available. It is tested on a number of data sets used for university-industry research work and publications.

Acknowledgements. The authors would like to thank industry members and academic partners in the Australian Research Council Center for Transforming Maintenance through Data Science (Industrial Transformation Research Program Grant No.IC180100030) who commented on the model and paper. This work would also have not been possible without funding from the BHP Fellowship for Engineering for Remote Operations—supporting community projects in areas in which BHP operates.

Authorship Contributions. J.S.: Conceptualization (Lead), Data curation (Lead), Formal analysis (Equal), Funding acquisition (Lead), Investigation (Equal), Methodology (Lead), Software (Supporting), Validation (Supporting), Visualization (Lead), Writing-original draft (Lead); S.B.: Software (Lead), Visualization (Supporting), Writing-original draft (Supporting), Writing-review & editing (Equal); M.R.H.: Conceptualization (Lead), Formal analysis (Equal), Funding acquisition (Lead), Investigation (Equal), Methodology (Supporting), Project administration (Lead), Supervision (Lead), Validation (Lead). Writing-original draft (Supporting), Writing-review & editing (Equal); R.F.: Funding acquisition (Lead), Project administration (Lead), Supervision (Supporting). Writing-original draft (Supporting), Writing-review & editing (Equal).

Data Availability Statement. DRAT is licensed under a Creative Commons Attribution 4.0 International License and is available via www.drat-process.com.

References

ACS. (2017). *Data Sharing Frameworks*. Technical White Paper 100.

Alberts CJ, **Behrens SG**, **Pethia RD and Wilson WR** (2006) *Operationally Critical Threat, As set and Vulnerability Evaluation (OCTAVE) framework, Version 1.0. CARNEGIE-MELLON UNIV PITTSBURGH PA SOFTWARE ENGINEERING INST*. 1999. Pittsburgh, PA: Carnegie Mellon University.

Alexander A, *et al.* (2018) University–industry collaboration: using meta-rules to overcome barriers to knowledge transfer. *The Journal of Technology Transfer 45*(2), 1–22.

Amin MT, **Khan F and Zuo MJ** (2019) A bibliometric analysis of process system failure and reliability literature. *Engineering Failure Analysis 106*, 104152.

Anthony JK, **Enno K, and Steven JV** (2004) Publications from industry. Personal and corporate incentives. *Plant Physiology 134* (1), 11–15.

Argyris C (1952) *The Impact of Budgets on People*. New York: Controllership Foundation.

Aulakh PS, **Sahay A and Kotabe M** (1996) Trust and performance in cross-border marketing partnerships: a behavioral approach. *Journal of International Business Studies 27*(4), 1005–1032.

ISO (2018a) *AS ISO 31000:2018: Risk Management—Guidelines*. SAI Global, p. 16.

https://doi.org/10.1017/dce.2020.13 Published online by Cambridge University Press

ISO/IEC (2018b) *ISO/IEC 27000:2018: Information Technology—Security Techniques—Information Security Management Systems—Overview and Vocabulary*. SAI Global, p. 31.

Bekkers R and Freitas (2008) Analysing knowledge transfer channels between universities and industry: to what degree do sectors also matter? *Research Policy 37*(10), 1837–1853.

Borgman CL (2012) The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology 63*(6), 1059–1078.

Broecke S (2015) University rankings: do they matter in the UK?. *Education Economics 23*(2), 137–161.

Bruneel J,**D'este P and Salter A** (2010) Investigating the factors that diminish the barriers to university–industry collaboration. *Research Policy 39*(7), 858–868.

Cameron IT and Raman R (2005) Process systems risk management. In *Process Systems Engineering*. Vol. *6*. San Diego: Elsevier Science & Technology.

Cashell BW, *et al.* (2004) *The Economic Impact of Cyber-Attacks*. Library of Congress. Congressional Research Service. The library of Congress https://archive.nyu.edu/bitstream/2451/14999/2/Infosec_ISR_Congress.pdf.

Couture JL, *et al.* (2018) A funder-imposed data publication requirement seldom inspired data sharing. *PloS One 13*(7), e0199789.

CRAMM (1987). *Risk Analysis and Management Method (CRAMM)*. Norfolk, UK: Central Computing and Telecommunications Agency.

Cristina Costa A and Bijlsma-Frankema K (2007) Trust and control interrelations: new perspectives on the trust—control nexus. *Group & Organization Management 32*(4), 392–406.

Das TK and Teng B-S (1998) Between trust and control: developing confidence in partner cooperation in alliances. *The Academy of Management Review 23*(3), 491–512.

Desai T, **Felix R, and Welpton R**. (2016) Five safes: designing data access for research. In *Economics Working Paper Series*. Bristol, UK: University of the West of England.

Dingler A and Enkel E (2016) Socialization and innovation: insights from collaboration across industry boundaries. *Technological Forecasting & Social Change 109*, 50–60.

Dirks KT and Ferrin DL (2001) The role of trust in organizational settings. *Organization Science 12*(4), 450–467.

Edelenbos J and Eshuis J (2012) The interplay between trust and control in governance processes: a conceptual and empirical investigation. *Administration & Society 44*(6), 647–674.

European Commission (2018) *Commission Staff Working Document: Guidance on Sharing Private Sector Data in the European Data Economy*. Brussels: European Commission.

European Commission (2020) In The European Economic and Social Committee and the Committee of the Regions(ed.), *A European Strategy for Data, T.C. Communication from the Commission to the European Parliament*. Brussels: European Commission.

Evans JA (2010) Industry collaboration, scientific sharing, and the dissemination of knowledge. *Social Studies of Science 40*(5), 757–791.

FAIRsharing.org (2020) Available at: https://fairsharing.org/ (accessed 20 June 2020).

Fecher B, **Friesike S, and Hebing M** (2015) What drives academic data sharing? *PLoS One 10*(2), e0118053.

Fecher B, *et al.* (2017) *A Reputation Economy: Results from an Empirical Survey on Academic Data Sharing*. London, England: Palgrave Communications, Nature Publishing Group.

Fox A (1976) Beyond Contract: Work, Power and Trust Relations. *The American Journal of Sociology 82*, 239–42.

Gambetta D (1988) Can we trust? In Gambetta D (ed), *Trust: Making and Breaking Cooperative Relations*. New York: Basic Blackwell, p. 246.

GDPR Rep (2018) *Complete Guide to GDPR Compliance*. Available at gdpr.eu. (accessed 13 August 2020).

Gibbons S, **Neumayer E and Perkins R** (2015) Student satisfaction, league tables and university applications: evidence from Britain. *Economics of Education Review 48*, 148–164.

Goold M and Cambell A (1987) *Strategies and Styles: The Role of the Centre in Managing Diversified Corporations*. Oxford: Basic Blackwell.

Gopalakrishnan S and Santoro MD (2004) Distinguishing between knowledge transfer and technology transfer activities: the role of key organizational factors. *IEEE Transactions on Engineering Management 51*(1), 57–69.

Green D, *et al.* (2016) Setting targets in an asset management performance measurement framework. In *2016 World Congress on Engineering Asset Management (WCEAM2016)*. Jiuzhaigou: WCEAM.

Griffith A and Rask K (2007) The influence of the US News and World Report collegiate rankings on the matriculation decision of high-ability students: 1995–2004. *Economics of Education Review 26*(2), 244–255.

Gritzalis D, *et al.* (2018) Exiting the risk assessment maze: a meta-survey. *ACM Computing Surveys (CSUR) 51*(1), 1–30.

Grossi D, **Royakkers L and Dignum F** (2007) Organizational structure and responsibility. *Artificial Intelligence and Law 15*(3), 223–249.

Haeussler C (2011) Information-sharing in academia and the industry: a comparative study. *Research Policy 40*(1), 105–122.

Harris L andSharma C (2017) *Understanding Corporate Data Sharing Decisions: Practices, Challenges and Opportunities for Sharing Corporate Data with Researchers*. Washington, DC: Future of Privacy Forum 20005, p. 21.

Henneken EA and Accomazzi A (2011) Linking to data-effect on citation rates in astronomy. arXiv preprint arXiv:1111.3618.

Horstschräer J (2012) University rankings in action? The importance of rankings and an excellence competition for university choice of high-ability students. *Economics of Education Review 31*(6), 1162–1176.

https://doi.org/10.1017/dce.2020.13 Published online by Cambridge University Press

Kim Y and Zhang P (2015) Understanding data sharing behaviors of STEM researchers: the roles of attitudes, norms, and data repositories. *Library and Information Science Research 37*(3), 189–200.

Kirsch LJ (1996) The management of complex tasks in organizations: controlling the systems development process. *Organization Science (Providence, R.I.) 7*(1), 1–21.

Mayer RC, **Davis JH and Schoorman FD** (1995) An integrative model of organizational trust. *The Academy of Management Review 20*(3), 709–734.

Meyer-Krahmer F and Schmoch U (1998) Science-based technologies: university–industry interactions in four fields. *Research Policy 27*(8), 835–851.

Naughton L and Kernohan D (2016) Making sense of journal research data policies. *Insights the UKSG Journal 29*(1), 84–89.

Nelson B (2009) Empty archives: most researchers agree that open access to data is the scientific ideal, so what is stopping it happening? Bryn Nelson investigates why many researchers choose not to share. *Nature 461*(7261), 160–4.

ODI (2020) In **T.O.D. Institute** (ed), *Data Sharing in the Private Sector*. The Open Data Institute and YouGov Plc. https://theodi.org/article/new-survey-finds-just-27-of-british-businesses-are-sharing-data (accessed July 2020).

Phillips J, *et al.* (2015) Classifying machinery condition using oil samples and binary logistic regression. *Mechanical Systems and Signal Processing 60*, 316–325.

Piwowar HA and Chapman WW (2010) Public sharing of research datasets: a pilot study of associations. *Journal of Informetrics 4*(2), 148–156.

Piwowar HA, **Day RS and Fridsma DB** (2007) Sharing detailed research data is associated with increased citation rate. *PLoS ONE 2*(3), e308.

Ponemon Institute (2019) *Cost of a Data Breach 2019*. Traverse City, MI: Ponemon Institute LLC.

QS World University Rankings. (2020). *Methodology*. Available at https://www.topuniversities.com/subject-rankings/methodology (accessed 20 July 2020).

Ritchie F (2017) The "five safes": a framework for planning, designing and evaluating data access solutions. In *Data for Policy 2017*. London, UK: Data for Policy.

SCDS (2020) *EU Support Centre for Data Sharing*. Available at https://eudatasharing.eu/ (accessed 20 June 2020).

Schartinger D, *et al.* (2002) Knowledge interactions between universities and industry in Austria: sectoral patterns and determinants. *Research Policy 31*(3), 303–328.

Sendi AS, *et al.* (2010) *FEMRA: Fuzzy Expert Model for Risk Assessment*. IEEE, pp. 48–53.

Sikorska J, **Hodkiewicz M**, **De Cruz A**, **Astfalck L**, **Keating A** (2016) A collaborative data library for testing prognostic models. In *3rd European Conference of the Prognostics Health Management Society*. Bilbao: PHM Society.

Sitkin S and Stickel D (1996) The road to hell: the dynamics of distrust in an era of quality. In Kramer RM and Tyler TR (eds), *Trust in organizations*. Thousand Oaks, CA: Sage Publishers, pp. 166–195.

Sitkin SB (1995) On the positive effect of legalization on trust In Bies RJ, Lewicki RJ and Sheppard BH (eds), *Research on Negotiation in Organizations*. Greenwich, CT: JAI, pp. 185–217.

Sitkin SB and Roth NL (1993), Explaining the limited effectiveness of legalistic "remedies" for trust/distrust. *Organization Science (Providence, R.I.) 4*(3), 367–392.

Smith DJ (2017) Reliability, maintainability and risk: practical methods for engineers. Oxford, England: Butterworth-Heinemann.

Stoneburner G, **Goguen A and Feringa A** (2002) *SP 800-30: Risk Management Guide for Information Technology Systems. Revision 1*. NIST (National Institute of Standards and Technology): Gaithersburg, Maryland, p. 56.

Tan CN-L (2016) Enhancing knowledge sharing and research collaboration among academics: the role of knowledge management. *Higher Education 71*(4), 525–556.

UKRI (2020). *UKRI Strategic Prospectus*. Available at https://www.ukri.org/about-us/strategic-prospectus/ (accessed 14 August 2020).

Vlaar PWL, **Van den Bosch FAJ and Volberda HW** (2007) On the evolution of trust, distrust, and formal coordination and control in interorganizational relationships: toward an integrative framework. *Group & Organization Management 32*(4), 407–428.

Wangen G (2017) Information security risk assessment: a method comparison. *Computer (Long Beach, California) 50*(4), 52–61.

Webb C, **Sikorska J**, **Khan RN**, **Hodkiewicz M** (2020) Developing and evaluating predictive conveyor belt wear models. *Data-Centric Engineering* 1.

Wiley C (2018) Data sharing and engineering faculty: an analysis of selected publications. *Science & Technology Libraries (New York, NY) 37*(4), 409–419.

Wiley C (2020) Data sharing: an analysis of medical faculty journals and article. *Science & Technology Libraries* 1–12. doi: 10.1080/0194262X.2020.1781740

Xie Q, **Bharat C**, **Nazim Khan R**, **Best A**, **Hodkiewicz M** (2017) Cox proportional hazards modelling of blockage risk in vitrified clay pipes. *Urban Water Journal 14*(7), 669–75.

Cite this article: Sikorska J., Bradley S., Hodkiewicz M. and Fraser R. (2020). DRAT: Data risk assessment tool for university–industry collaborations. *Data-Centric Engineering*, 1: e17. doi:10.1017/dce.2020.13