



Analytical modelling and UK Government policy

Marie Oldfield¹ · Ella Haig¹

Received: 20 May 2021 / Accepted: 26 June 2021
© The Author(s) 2021

Abstract

In the last decade, the UK Government has attempted to implement improved processes and procedures in modelling and analysis in response to the Laidlaw report of 2012 and the Macpherson review of 2013. The Laidlaw report was commissioned after failings during the Intercity West Coast Rail (ICWC) Franchise procurement exercise by the Department for Transport (DfT) that led to a legal challenge of the analytical models used within the exercise. The Macpherson review looked into the quality assurance of Government analytical models in the context of the experience with the Intercity West Coast franchise competition. This paper examines what progress has been made in the 8 years since the Laidlaw report in model building and best practise in government and proposes several recommendations for ways forward. This paper also discusses the Lords Science and Technology Committees of June 2020 that analysed the failings in the modelling of COVID. Despite going on to influence policy, many of the same issues raised within the Laidlaw and Macpherson Reports were also present in the Lords Science and Technology Committee enquiry. We examine the technical and organisational challenges to progress in this area and make recommendations for a way forward.

Keywords Modelling · Society · AI · Policy · Organisational culture · Organisational behaviour · COVID · Evidence based · Government

1 Introduction

In the last year, due to COVID, evidence-based modelling has been at the centre of government and decision-making. Although evidence-based modelling has been highlighted in the UK Government for decades, the circumstances and urgency of COVID have brought its importance to the fore. The policy of the UK Government is to use evidence from experts, through rigorous process, to underpin its decisions [1] on national policy. The UK Government takes evidence from modellers, experts and policy officials to decide on the best way forward. In 2012 the West Coast Mainline franchise disaster highlighted the serious consequences of not undertaking robust modelling or robust decision making. The legal challenges to the West Coast Franchise Decision caused a significant loss of taxpayers' money and serious reputational damage to the UK Government [1]. The Laidlaw Report

aimed to make recommendations on how to solve the failings moving forward. However, the Lords Science and Technology Committee of 2020 [2] again showed the same systemic issues showing up in Government Modelling. Indeed, in addition to those issues previously highlighted in [3], further issues were shown to have occurred. The problems indicated by this succession of reports point to challenges, both technically and organisationally, that prevent robust modelling from being undertaken by the UK Government. In 2020 Committee Members who sat on the inquiry boards for the West Coast Mainline franchise failure occupied senior positions in Government.

The Intercity West Coast Rail Franchise (ICWC) failed because the Department for Transport “did not get basic processes right and had failed to learn from mistakes made in previous projects” [4]. The House of Commons Committee for Public Accounts stated that this failing had not only occurred previously but that the recommendations made from the 2010 report *The failure of Metronet*¹ [5] “to

✉ Marie Oldfield
Marie.Oldfield@myport.ac.uk

Ella Haig
ella.haig@port.ac.uk

¹ University of Portsmouth, Portsmouth, UK

¹ The loss to the taxpayer arising from Metronet's poor financial control and inadequate corporate governance is some £170 million to £410 million [5].

prevent a lack of oversight and information were clearly not applied in this competition” [4]. The decision to drive cost savings during the Intercity West Coast Rail Franchise Competition and not employ external financial advisers cost taxpayers tens of millions of pounds, a substantial amount of this being used in compensation payments to bidders. The department were not able to quantify or understand the risk levels surrounding the bids and were therefore unable to correctly model the risk capital required to balance this. These errors led to the “Department asking First Group for a lower subordinated loan facility than was needed to protect itself from the recognised additional risk in the bid. A higher subordinated loan facility was requested from Virgin Trains. This opened the Department to the risk of legal challenge and ultimately led to the cancellation of the franchise competition” [4]. These failings were highlighted by the House of Commons Committee for Public Accounts and attributed to multiple issues such as lack of leadership, lack of ability to challenge, lack of transparency in modelling, a drive to cut costs, failure to apply common sense, failure to apply basic processes and lack accountability [4]. This failure would then be examined further by the Laidlaw Report and Macpherson Review.

The Laidlaw Report [6] addressed what went wrong in the Intercity Intercity West Coast Rail Franchise failure and produced recommendations for moving forward. On the 3rd of October 2012, the competition to run passenger trains on the West Coast Main Line had been cancelled “following the discovery of significant technical flaws in the way the Intercity West Coast Franchise process was conducted” [6]. The Laidlaw Report discusses the significant flaws in the modelling process used and lessons learned by the Department for Transport (DfT) from the InterCity West Coast competition.

The aim of the Macpherson Review was to identify any systemic issues cross-government and to identify best practice that could be implemented moving forwards. In particular, the aim was to “examine the quality assurance of Government analytical models which are used to inform policy” [1], as well as identifying best practice from industry and the private sector.

This paper examines the initiatives proposed by the Laidlaw Report and Macpherson Review and compares them to the findings of the Lords Science and Technology Committee [2] that took place in June 2020 in the context of discussing the models that informed policies related to COVID. This paper examines the need for empirically informed studies that are not only robust but based on technical and organisational best practice. Studies that are challenged, based on good methodology and a thorough understanding of aspects such as risk, impact, caveats and assumptions. Studies that are conducted in the best interests of the UK population. The need for evidence-based policy is discussed at length in UK Government

publications such as the AQuA book² [7] and in enquiries such as The Intercity West Coast franchise [6], the failure of the Metronet [5] and the Lords Enquiry of 2020 [2]. However, systemic failings have been highlighted for many decades with substantial impact on, not only the public purse but the lives of UK Citizens. This paper makes recommendations from the examination of multiple enquiries over multiple decades with the vision of a new future in evidence-based policy.

COVID, by 2020, was perceived as a potential public health issue and as such, modelling would need to be performed with existing international data to determine what steps might need to be taken by the UK [8]. In June 2020 an enquiry was commissioned into this modelling after numerous failings had been identified. These failings led to policy being implemented that was neither based on evidence-based modelling nor, in some cases, implemented prior to Government and Academic modellers being able to produce their modelling results. The Government changed policy designed to influence public behaviour prior to any modelling taking place and at a fast pace according to evidence within the Lords Enquiry [2]. Modellers raised numerous concerns around the modelling undertaken for COVID policy such as: lack of ability to challenge the modelling or parameters, lack of appropriate modelling for the situation, incorrect metrics, lack of validation and verification of models, lack of availability of data, incorrectly discounted modelling, lack of discussion around suggestions for parameters within models, lack of risk understanding or awareness, lack of leadership, and problematic resource allocation. Many of these failings had been found previously in the Macpherson Review, Laidlaw Report and the 2010 report *The failure of Metronet* [5] but in the Lords Enquiry of 2020 additional failings were also highlighted.

The recommendations found in this paper relate directly to the previous recommendations by the House of Commons Public Accounts Committee, Laidlaw Report and Macpherson Review. It is recommended that leadership must be accountable, communicative and fully trained for the post. This includes understanding the processes in the domain of modelling and area-specific challenges. Transparency and data in modelling should be addressed by providing caveats, risks and assumptions as well as model-relevant paperwork. Constructive challenge should be encouraged and facilitated. Guidance must be based on existing modelling best practice but where this is lacking in role and communication guidance, Prince II³ or

² The AQuA book provides guidance on producing quality analysis for Government. Please find further information in Sect. 2.3.

³ PRINCE II is a structured project management method and practitioner certification programme. PRINCE II emphasises dividing projects into manageable and controllable stages.

SCRUM⁴ techniques could be incorporated. Training and support must be provided for practitioners in modelling and robust data collection and use. Modelling, with the responsibility placed upon it to support evidence-based policy decision making, should be allocated as a skilled profession. In addition, gaps should be identified in primary, secondary and tertiary education in areas such as modelling the real world in an ethical and robust manner. This is likely to identify other aspects of note in education such as risk awareness, proportionality and measurement.

This paper is structured as in the following. The first section gives an overview of the terminology and life-cycle of models and their quality assurance. The second section discusses modelling in the context of the Macpherson Review and Laidlaw Report in the wake of the Intercity West Coast Franchise debacle. The third section discusses how effective the implementation of the recommendations from the Laidlaw Report and Macpherson Review were. The fourth Section covers the Lords Enquiry of 2020 in which the modelling undertaken to address policy requirements on COVID was examined. This paper then proceeds to examine organisational culture as a current barrier, but also as a fundamental factor to improvement in this area and discusses what can be done to ensure progress. Recommendations are then made on ways forward before concluding.

2 Modelling

As stated in the Macpherson Review, “Modelling is essential to the work of government” [1]. “From providing the evidence to support major investment decisions to predicting the spread of pandemic flu, models underpin decisions which affect people’s lives and have major financial implications” [9]. It is therefore vital that the models used are fit for purpose. “Balancing the tension between supporting innovation, so that society’s right to benefit from science is protected and limiting the potential harms associated with poorly designed modelling is challenging” [10].

In this paper a model is defined by the data journey, along with the outputs from this process [11]. This is because the inputs, outputs and data journey are one of the most important parts of the modelling journey. Models under the remit of the quality assurance process outlined in the AQuA book [7] are discussed. Indeed, as the analysis progresses through the analytical cycle, there are various checks performed to ensure that the analysis is fit for purpose. Checks

that confirm that the right analysis has been performed are known as validation and checks that the analysis has been carried out correctly are known as verification. An example of one of these checks is “the quality of any data inputs, and any assumptions that drive the analysis, including the estimation of parameters” [7], indicating that the data has to be of sufficient quality for the model to be robust.

UK policy can be based on models as varied as, financial models, defence operational models or scheduling models. The process of building a model comprises how the data is collected and processed, how the model is built to achieve the aim of the questions being asked, the robustness and transparency of the model, as well as outputs and interoperability. Methods to ensure fully robust models include ‘Validation’ and ‘Verification’ [7, 12] as well as such articles as data flow charts, model topology, version control, testing records and audit records.

Before outlining the life cycle of a model, we briefly discuss business-critical models in the following subsection given their importance and the need to manage high risks⁵ associated with such models.

2.1 Business-critical model definition

A business-critical model is defined as a model that, if built or altered or used incorrectly could lead to “serious financial, legal or reputational damage” [13]. These models must, according to AQuA and Laidlaw, under all circumstances, be quality assured under the AQuA book and audit processes within Government [6]. These models are required to be transparent due to potential impacts being so wide ranging and sizeable.

Business-critical models as defined by the Macpherson Review can ultimately cost lives. These models, as stated in the Laidlaw Report [6] and the Macpherson Review [1], should be subject to the highest level of scrutiny to ensure they are robust and transparent. The requirements around this were highlighted by the Macpherson Review and are discussed in the subsequent section on Quality assurance (QA).

2.2 The life cycle of a model

The life cycle of a model explains how a model is built and what is inherent in the process. “Typically, models progress through a four-step cycle, albeit often with significant cycling between steps during the life of a model:

⁴ Scrum is a framework utilizing an agile mindset for developing, delivering, and sustaining complex products, with an initial emphasis on software development, although it has been used in other fields including research, sales, marketing and advanced technologies.

⁵ High Risk is defined as likely to result in failure, harm, or injury. In the case of the UK Government this would also include reputational and financial damage.

scope, specify and design; build and populate; test; deliver and use” [13]:

1. In the scoping phase the client relates what conceptual task they would like to solve. This might be, for example, what aspect of a call centre needs to be optimised to increase revenue.
2. In the design phase the modeller takes the client’s specifications that have been discussed in full with the client.
3. The build phase is where the model is constructed.
4. Testing is normally done on an existing, potentially historical, data set and then validated by comparison to real life output.

This process follows the verification and validation process illustrated in the Macpherson Review [1] and the AQuA book []. The deliver and use phases involve creating the required documentation such as user guides, testing reports, version control, commented code, training materials and audit trails, and delivering this to the client. Something to note is that the client and modelling team are not always independent, especially in Government modelling. When the model builder and client are in house and potentially on the same team, the lines required for robust quality assurance can become blurred.

2.3 Quality assurance of models

To combat the aforementioned catalogue of issues raised in the Laidlaw Report and the Macpherson Review, the UK Government commissioned the AQuA book. This book was part of a series called the Rainbow Books and was written to guide analysts in the process of developing and using models. Subsequent to the Macpherson Review, an audit process was also implemented to ensure that models have the correct paperwork and training in place, and that custodians of models would have the correct information to use and implement the model they are charged with overseeing.

There are many types of quality assurance that can be used within modelling. The Macpherson Review [1] highlighted that “It is important that the design stage includes a clear understanding of the model structure and logic as well as the underlying assumptions, limitations, inputs required, and outputs expected”. Also, “The completed model should be available, together with a full set of quality controlled input data and details of the model’s inputs’ limitations or uncertainties” [1].

Several types of quality assurance were proposed by Macpherson [1]:

- *Developer testing*: use of a range of developer tools including parallel build and analytical review or sense check;
- *Internal peer-review*: obtaining a critical evaluation from a third party independent of the development of the model, but from within the same organisation;
- *External peer-review*: formal or informal engagement of a third party to conduct critical evaluation, from outside the organisation in which the model is being developed;
- *Use of version control*: use of a unique identifier for different versions of a model;
- *Internal model audit*: formal audit of a model within the organisation, perhaps involving use of internal audit functions;
- *Quality assurance guidelines and checklists*: model development refers to department’s guidance or other documented quality assurance processes (e.g. third party publications);
- *External model audit*: formal engagement of external professionals to conduct a critical evaluation of the model, perhaps involving audit professionals;
- *Governance*: at least one of planning, design and/or sign-off of models for use is referred to a more senior person. There is a clear line of accountability for the model;
- *Transparency*: the model is placed in the wider domain for scrutiny, and/or results are published; and
- *Periodic review*: the model is reviewed at intervals to ensure it remains fit for the intended purpose, if used on an ongoing basis.

According to Macpherson [1] the formal deliverables will vary depending on the model; however, there should be clear documentation for the model to ensure robustness and transparency. The recommendation was that documentation included in quality assurance processes needed not be cumbersome and, in some instances, more akin to a diary of design. The following were presented as best practice [1]:

- At the design stage, model design documentation to support the build phase describes the model and should include the quality assurance strategy for the build and testing phases. Some quality assurance may be performed at this stage to provide assurance that the model structure, logic and assumptions are robust before the model is built;
- Review by either internal or external reviewers should be considered for complex models and an assessment of the suitability and availability of the inputs and outputs should be made;
- At the build stage the documentation accurately describes the model as developed (noting any differences from the design), any verification testing done and the test results;
- Once the model is complete and has been subject to appropriate verification testing, a further validation testing phase should be conducted, and documented, to ensure the model is fit for purpose;

- At the test or ‘deliver’ stage the documentation includes: a description of the tests run; the test results; any issues identified; and corrections made. If user documentation is needed it should also be developed and reviewed at this stage together with any required training material.

In the following section, the findings of the Laidlaw Report and the Macpherson Review are discussed in detail and the areas of critical importance for the improvement of modelling practices are summarised from the two sources.

3 The Laidlaw Report and the Macpherson Review

Given the number of models underpinning critical Government decisions in 2012 it could be expected that a system of quality assurance already existed. However, a system of cross-government best practice was not in place at this point, and it was more siloed and department-led if it existed [1]. This is shown by different departments having different levels and volume of model guidance, but no uniform approach across Government. This approach currently continues [9].

The Macpherson Review also aimed to identify business-critical models across Government. These are the models with most potential to cause serious consequences whether legally, financially or reputationally. The Ministry of Defence declared around 60 such models [13], which constituted 13% of the overall models across Government at that time.

The Macpherson Review involved specialists from many different areas: “the review was multi-disciplinary, involving Operational Analysts/Researchers, Economics, Statistics, Policy Professionals, Software Experts and Social Science expertise as befits the very wide-ranging and diverse modelling stock that is used to underpin evidence-based decision making across Government” [13]. There is a wide variety in the types of models used in Government, from simple spreadsheet type quantitative models to very complex models using mathematical and statistical representations of the real world. There are also many new platforms to choose to model on, such as Python, R, Simul8, and a plethora of data science open source packages. Therefore, it is important that whatever package or platform is chosen, the relevant quality aspects required for a robust model can still be produced.

The Macpherson Review determined that “successful modelling is, therefore, not just a matter of modellers accurately building models. Decision makers also need to understand the strengths and limitations of the chosen modelling approach. Departments’ cultures should reflect this by minimising barriers between policy and analytical professions, and encouraging mutual understanding and respect, as well as emphasising the importance of

communication skills” [1]. The report highlighted the importance that a model be explained to the non-professionals, as they are key members and stakeholders in any model build. Not only is their context key to producing a robust model, but their feedback gives valuable information towards the understanding of the model. It is crucial that a model “design be informed by real world needs” [14].

3.1 Key findings of the Macpherson Review and Laidlaw Report

The Macpherson Review and Laidlaw Report were very thorough in their investigations and understanding of the issues. Many of the people involved in the Intercity West Coast Rail Franchise were interviewed and much data and correspondence was sifted through. What follows are the key points from both sets of investigation.

The Macpherson Review stated that “it is vital that all levels in an organisation understand the value attached to models and quality assurance” [1]; this includes everyone from the modeller to Ministerial level. In the Intercity West Coast Rail Franchise, “the Permanent Secretary was deliberately not allowed to see the details of the competition and commercially confidential information” [4]. Following the poor escalation of issues, a poor culture of best practice, a lack of opportunity to challenge and an absent Senior Responsible Owner (SRO) [6], meant that the analysis was not well understood or robust [6]. Any caveats or assumptions that are also attached to a model should be detailed out in accompanying paperwork so anyone can be aware of the space in which the model operates and what it can and cannot do. The Macpherson Review stated that these should be “clearly communicated, and if modelling is not possible within the given constraints, analysts should have the support and means to say so” [1].

The Macpherson Review outlines the importance of the Senior Responsible Owner (SRO) and that they should be “sufficiently senior” to take responsibility for the model: “the key requirement is that policy professionals and analysts work together closely to ensure the model SRO is able to ask the right questions, fully understands the uses and limitations of the model and is therefore able to sign-off to confirm it is fit for purpose” [1]. These were found to be contributory factors in the Intercity West Coast Rail Franchise competition. The SRO’s sign-off assures that [1]:

- The quality assurance process used is compliant and appropriate;
- Model risks, limitations and major assumptions are understood by the users of the model; and
- The use of the model output is appropriate.

The AQuA book also raises the requirement for education and training to ensure that users and developers of models, as well as SRO's are correctly trained to ensure the model, risks, assumptions and context are well understood [7].

However, the Macpherson Review states the following: “a fairly high proportion of models (around 50%) had outputs that were available to external scrutiny and so are classified as ‘transparent’” [1]. This is potentially misleading, as outputs alone do not contain the details required to indicate that a model is transparent. As detailed above, a full set of documentation, including inputs, outputs, assumptions, caveats, etc., would be required. Only then, with SRO sign-off, a full set of plain English documentation [7], and peer-review, should a model that is classified as business critical or that could have serious consequences when used, be signed off. Indeed, the Macpherson Review states that “If the model SRO cannot give their sign-off, this signals the model is not fit-for-purpose. In this case, the model should not be used until any specific issues are rectified. This may entail amending the model, undertaking further quality assurance, or producing a completely new model that better supports the policy need” [1].

Another common challenge within Government is that “there are challenges in preserving good quality assurance when a model’s scope and purpose shifts in response to often sudden change in policy and priorities” [1] and “machinery of government change can lead to legacy issues with models that started in one department and subsequently end up owned by another. It can be challenging to track the development of these models and update them” [1]. This can cause serious issues both through lack of knowledge preservation, staffing constraints and mothballing of models (i.e. retiring the model and then bringing it back in to use some time later) [7].

The AQuA book states that “the commissioner must be confident in the quality of the outputs and understand the strengths, limitations and contexts of the analysis so that the results are correctly interpreted” [7]

An empirical study, [15] found that explanations and the way they are delivered can substantially impact the end result of the interpreter of the model [16]. This would be a highly relevant point when speaking with Ministers who may not have a background or an in-depth understanding of the model at hand.

Among the many issues identified within the Laidlaw Report and the Macpherson Review a number of issues stand out as being of critical importance in modelling practices. These are discussed in the following subsections.

3.2 Resources and allocation of time

The Laidlaw Report recommends that “appropriate discipline is applied in the allocation and balance of time” [6,

17], while the Macpherson Review outlined that “Departments with the most developed quality assurance processes appeared to have sufficient specialist and experienced staff, but not all felt they had the staff with the right skills in place to match the demands on them. Retaining specialist staff and providing career progression for experts was highlighted as a challenge” [1].

3.3 The ability to escalate issues, to challenge and poor communication within modelling and project teams

The Laidlaw Report recommends that “a review is carried out of escalation policy and of the effectiveness of communication to staff of expectations and responsibilities in respect of line reporting” [6]. In the working culture of the project team, it was found that “one significant factor contributing to the flaws relates to the conduct of individual Department for Transport officials, including in relation to the opportunities that were missed to escalate or report information” [6]. This highlights a cultural issue within the team that prevented escalation of problems and allowed opportunities to be missed [18, 19].

3.4 Skills

Another recommendation of the Laidlaw Report is that “a skills review is carried out and a thorough needs assessment undertaken to establish whether there are capability, experience or leadership gaps” [6]. Education and training should be provided to close any gaps found. The Macpherson Review highlighted that “There should be appropriate capacity and capability where specialist staff have sufficient time built-in for quality assurance, and are able to draw on expertise and experience across Government and beyond” [1]. This is supported by research by Harrison [20].

3.5 Leadership

In the Laidlaw Report, changes in the leadership of the DfT “contributed to the flaws in the ICWC franchise process and adversely impacted the DfT’s effectiveness in identifying and/or resolving those flaws” [6]. The Macpherson Review recommended that “There should be visible leadership at the top of the organisation—backed by incentives—to create a culture that expects high-quality QA” [1]. Leaders have to be people-centric for success according to Peters [18], and Appelbaum discusses how leaders can implement a vision to allow external auditing to become a reality [21, 22].

3.6 Internal and external audit processes

The Laidlaw Report highlights shortcomings in performing internal/external reviews/audits: “The Inquiry team questions the basis and extent of review, checking and model auditing undertaken by the DfT. The Inquiry team notes that several internal reviews of the GDP⁶ Resilience Model were reportedly undertaken but there is no documentary evidence on the facts and nature of the review. There is also no record of any model audit or best practice review being undertaken” [6]. In the Macpherson Review, the internal and external model audit are mentioned as important types of quality assurance [1].

3.7 Lack of clarity around roles

The Laidlaw Report highlighted that there was a “lack of clarity in roles and responsibilities and in associated accountability, including a failure to get the SRO structure to work for the benefit of the project” [6]. This indicates that the SRO may not have been adequately qualified or may not have had time to be involved in the project, which would be a very large oversight considering that the expert on the model was not present in many project meetings [6]. The report also stressed the following: “It should be noted that the SRO role is an important one not only in ensuring that the required resources are available but also in providing overall oversight, quality control and risk review as well as, where appropriate, escalation to Ministers” [6]. In addition to this, there was a perceived lack of efficacy in the governance framework, and lack of clarity around the function, authority and interrelationship of committees and boards was also cited as a contributing factor. There appeared to be a link to non-independency of boards and committees, which caused some escalatory and quality issues [6].

3.8 Senior Responsible Owner (SRO)

The approach of allocating a Senior Responsible Officer to models was highlighted as a preferred way forward in the Macpherson Review. The responsibilities of the SRO and the importance of this role were outlined earlier in this section; they included: (a) the seniority of the SRO for the purposes of responsibility and accountability; and (b) the conditions required for an SRO sign-off for a model. Furthermore, in relation to the design stage, “The model SRO should at this stage check that the proposed design meets the organisation’s requirements. They should check the assumptions, limitations, inputs and outputs to make sure they remain

consistent with the intended use of the model and discuss the most appropriate approach to QA” [1].

3.9 Summary

In this section it can be seen that the majority of the issues concern organisational culture, learning and processes. “Knowledge loss causes challenges for organisations that wish to remain competitive. These organisations must identify the risks that could lead to knowledge loss and become aware of issues that affect knowledge retention” [23–25]. Indeed, it has been a common concept for decades that a business should “review their successes and failures, assess them systematically, and record the lessons in a form that employees find open and accessible”. Thus, “the knowledge gained from failures [is] often instrumental in achieving subsequent successes” [26]. This has been seen within IBM’s computer development programmes, Boeing’s progress from the 707 and 727 to its success with the 737 and 747 models and also within Xerox’s product development process [27].

Garvin [26] also gives a recipe for successful knowledge transfer in the following quote “For learning to be more than a local affair, knowledge must spread quickly and efficiently throughout the organization. Ideas carry maximum impact when they are shared broadly rather than held in a few hands”. Garvin [26] indicates that training programmes, personnel rotation and education can be some of the most effective ways, along with reports, to “ensure that knowledge is transferred across the organisation” [26].

However, this process is crucially interrupted when downsizing is brought into the equation, as happened during the last decade within the civil service. Fisher states that “even when downsizing is implemented without the intention of major re-structuring, the net result is the same number of employees left to do the same amount of work” [28]. This indicates that a compounding of factors, including lack of adequate organisational knowledge transfer and learning, could potentially have led to the same issues that were found in the ICWC enquiry being found again in the Lords Select Committee of 2020 nearly a decade later [2].

After the ICWC disaster, it was the AQuA book that was written to try to alleviate the issues found in the initial Laidlaw Report in 2012. In the next section, we examine what the AQuA book set out to do and whether these goals were achieved [29].

4 The AQuA book

As noted by Robinson and Glover [9], multiple publications were created in order to address the myriad of issues within Government modelling post ICWC; these are known as the Rainbow books and include:

⁶ Gross Domestic Product.

- The Green book, which look at the processes of appraisal and evaluation in Central Government;
- The Orange book focused on the management of risk, and covered principles and concepts;
- The AQuA book provided guidance on producing quality analysis for Government;
- The Magenta book covered Central Government guidance on evaluation.

The AQuA book [7] is the most relevant for the discussion here; this book is a compilation of key texts intended to guide modellers and interpreters, and also to give an outline of quality assurance processes. This book covers the following aspects of modelling [7]:

- Analysis, modelling and decision making in Government;
- Roles and responsibilities in analysis and model development;
- Risks in modelling and analysis;
- Quality assurance, verification and validation;
- Overview of common pitfalls.

The AQuA book follows much the same direction as the other Rainbow books but emphasises the implementation of analysis and modelling. More granular directions and guidance are given. The AQuA book provides advice for producing fit for purpose analysis and provides guidance on verification and validation of models. The AQuA book also provides guidance and useful templates to produce the relevant documentation to prepare for quality assurance.

The three pillars concentrated on within the AQuA book are uncertainty, stakeholders and fitness for purpose. This covers: (a) advice for working in the scoping phase and communicating with stakeholders involved in the model; (b) uncertainty and how to deal with assumptions or caveats which may produce a large amount of uncertainty in a model and (c) fitness for purpose, how the model works and how it can be verified and validated [30]. These three pillars aim to ensure modelling follows a process that would ensure a fit-for-purpose, robust and transparent result that can be audited and delivered with some degree of confidence.

Post AQuA book there was sporadic activity across Government in producing modelling guidance. The National Audit Office, in 2016, produced a short paper on a Framework for Modelling [31] and in 2018 the Department for Energy and Climate Change published a paper on Modelling Assurance [9]. The AQuA book does reasonably well in formalising the technical aspects of modelling but does not address the critical areas of leadership and resource allocation. This was one of the key outcomes of the ICWC enquiry. In itself, process design and implementation is an extremely valuable task but without the relevant soft skills and the support of leadership, in a manner that encourages

cultural change, it is of little use. This is because organisational behaviours are inextricably linked to the acceptance and use of processes. If employees do not understand or do not buy into the leader's vision, or have issues with the process, then no matter how good the process appears it will fail. Wu notes that process and organisational conflict result in negative outcomes for projects [32]. Wu also states that "process conflict and relationship conflict affected each other and were negatively related to project success, leading to poor communication among teams" [32]. This brings us to the year 2020 and the Lords Select Committee Enquiry which concentrates on modelling nearly a decade from the ICWC enquiry [2].

The Lords Select Committee for Science and Technology was called to question modellers on the robustness of their analysis that has been used to underpin Government policy in relation to COVID. Some of the members of this committee were also present on the Lords Committee on AI implementation in the UK in 2018 which examined explainability, transparency and modelling [33].

In the Lords Select Committee of June 2020, many of the issues highlighted by Macpherson and Laidlaw came again to the fore. Issues such as lack of time, lack of ability to challenge and lack of understanding of uncertainty were quoted as reasons why the modelling produced was not robust, did not follow guidance and was at some points, even discarded by stakeholders unnecessarily. Quality assurance processes quoted ranged from none to checking a press conference number against a model that may or may not have been verified and validated.

To enable comparison, the output from the Lords Select Committee is examined under the same headings as the Macpherson Review and Laidlaw Report, in addition to new emerging issues. As the output of the Lords Committee was in the form of oral transcript, it has been reported in this paper in this manner. In the transcript, A is the Interviewer and B the Interviewee.

4.1 Resources and allocation of time

It is stated that the team were not allowed the time to model despite this being one of the key outputs of the Laidlaw Report and the Macpherson Review. "All that we ask for as modellers is that, as these things are implemented, we have the time to get some data, even if it is just preliminary data" [2].

The interviewee commented on the short turnaround time between a request for modelling and the delivery deadline; they mentioned that the typical time is 2–3 days and that often this is over weekends. They further mentioned that this short turnaround does not allow the modellers to address anything other than what has been posed; additionally, they said: "There is not a whole lot of room for interpretation and

for addressing questions other than that which has been put down from on high and from SAGE” [2].

SAGE (Scientific Advisory Group for Emergencies) had driven the modelling despite there being qualified experts present to undertake this. Challenge appears not to have been possible here and the phrase ‘on high’ puts the conversation on a certain footing. This is similar to issues found in the Macpherson Review indicating that culture of challenge and communication continues to be an issue.

4.2 The ability to escalate issues, to challenge and poor communication within modelling and project teams

The following quotes demonstrate a lack of challenge and open communication channels with leadership. There is also a key issue with what models were asked to do versus what they were actually capable of doing.

B: “This modelling committee ... has been used to address a very narrow set of questions, which largely have been to do with population-wide social distancing: lockdown. It has not been used to explore other possible ways to respond to this epidemic. That is absolutely not the fault of the modellers; they are doing what they are asked to do” [2].

B: “We can make fairly simple, statistics-based forecasts for 2 weeks—I think you said that, [My Colleague], and I would agree. Longer term, there are too many uncertainties. I do not think we can make predictions” [2].

The interviewer asked to what extent the modellers have been able to change their approach, including parameters and input data to take into account uncertainty. Following on what the interviewee said about the use of the data being focused on social distancing, the interviewer asks what else it should have been focused on besides social distancing.

Having not used data from across Europe, the model was targeted at the wrong interventions despite the modellers having months of previous data from international sources:

B: “So far there has not been so much interest in targeted responses—except in very particular circumstances, which I am sure we will come back to—and the models have not been asked to, and have not really considered, those targeted responses in detail. Shielding, for example, a very appropriate strategy for people who are particularly vulnerable to this virus, has been part of Government policy across the UK Nations practically since day one, since the very early stages, but it is not included in any of the models. We are not aiming the models at the right target; we are aiming

them at everyone when in fact the burden of this disease is very concentrated” [2].

B: “Typically, we react to commissions that come from SAGE, which presumably come in turn from the Cabinet Office, and they tend to be targeted at population-level impacts”.

A: “In a sense, then, not all the right questions have been asked. Has there been any move among the modelling community to seek to persuade policymakers that there should be a greater focus on some of the sub-population?

B: “London saw if not a more rapid growth then a greater level of infection prior to lockdown, and correspondingly has had a much more rapid decline since so there is a question to be asked whether London could have a differential relaxation of measures to elsewhere in the country. The question of age is still not one that I think has been widely discussed” [2].

Here there is clear disagreement on modelling aims and not enough consideration of previous data, data that could be used to validate or verify the ensuing model. There is inconsistency apparent that SAGE wanted population level modelling when the data was showing that targeted interventions would have performed better. Therefore, Lord Hollick states that the right questions were not asked. Challenge does not seem to be an open channel to the modellers and they are asked to “persuade” policy makers when this critical exercise should be an open discussion. There was also a reported inconsistency in the ‘R measure’ meaning it was difficult to explain when the measure itself was not well understood [34, 35]. The understanding of risk, risk proportionality and subsequent communication to the public was not robust [2, 36]. Again, there is a lack of ability to challenge on this, despite the feeling that an alternative approach to risk would have been better. This is illustrated in the transcript where the Interviewer states that the R measure appears inappropriate. The Interviewee responds that the single measurement has been a distraction and that to use a single measure to drive policy would be misleading. The Interviewee further states that the general impression is that the R measure is a critical number but that this would be incorrect. By concentrating on the R number the Interviewee states that we have now lost sight of the real risk of the over 70’s and the lack of modelling around this cohort.

B: “(...) we should probably go back to old-fashioned public health and think about it in terms of risk: what is the risk to an individual in this location at that time? Apart from anything else, that is very helpful to the individual concerned, allowing them to make informed choices about how they behave. I am not sure whether the R number helps an individual decide how to behave, but it certainly does not help me” [2].

Here the modeller states that what they have been asked to model is not appropriate. Therefore, by extension it may not be useful to base policy on. There are additional issues with risk perception and management here that are crucial to understanding the level of appropriate risk for an individual. This is detailed out in the paper by Oldfield & McMonies submitted to the Lords Risk Planning Committee in 2021 [36].

4.3 Skills

The responses given below indicate that potentially the models used were simply not well enough understood by policy makers. The model being discussed simply gave potential futures, not a specific future. It appears to be a very vague model along the lines of multi-simulation type modelling. This was a pertinent point raised by Macpherson where policy makers were unable to challenge models as they did not have the understanding to do so.

The Interviewer asks about the predictions done by the models and whether they do indeed come true. The Interviewee states that the models simply simulate many hundreds of thousands of scenarios but without a likelihood attached to them. It is stated that the model just produces a long list of alternatives for what might happen [2].

4.4 Leadership

In the following exchange, it can be seen that the questions are potentially not the right set to be asking of the data and modelling team. However, again challenge does not appear to be possible.

A: “First, are you content that, from whichever source these questions are coming, the correct questions are being asked of the models and the data you have available, that the limitations are properly understood by those asking the questions, and that you and fellow modellers are in a position to dismiss those questions if the methodology for the modelling you have available to you is inappropriate to address them?” [2].

The answers to these questions were deflective. One answer was there was not as much oversight over models as there could have been and another answer stated that the wrong questions were being asked of the modellers in the first place. Challenge or discussion was not mentioned as a recourse to potential poor decision making. Another answer stated that there were no examples of validity checks on the models being used and that one team did not necessarily know what the other team were working on. The lack of interdisciplinary working along with distant stakeholders means that context and discussion, which are critical to developing and using models, is lost. This problem is

covered extensively in the AQuA book [7] and Laidlaw Report [6] where potential communication channels, team formation and relevant mandated paperwork are discussed.

4.5 Internal and external audit processes

No mention was made of quality assurance or validation and verification processes which are documented at length in Government publications and are mandated for business-critical models. The type of validation and verification discussed does not constitute any reasonable attempt at validation and verification and indeed is not along any lines that are prescribed or mandated in Government guidelines for modellers.

B: “I think the first batch was reported by Patrick Vallance at one of the press conferences. He reported a particular figure, so I immediately asked my team, “What’s our model saying about last week’s level of antibody positives?” They said, “It predicts about 6%”, which was exactly what Patrick Vallance said. So we felt reassured that our model had successfully captured something that it had not had any data inputs for. That sort of internal validation happens continuously with all these models” [2].

This is in stark opposition to documented and mandated best practice by Laidlaw [6], which is concerning. Modellers should be able to understand their model and output and use validated quality assurance techniques to understand their model. This was laid out very clearly in the Laidlaw Report [6] and the Macpherson Review [1]. When asked how often the models that policy is dependent upon are tested, the response was that verification and validation of the sorts of models that are used in real time have always been extremely challenging” [2].

This point raises a potential culture issue as Norling [37] states that often academic modelling is done in the abstract with no danger of significant decisions being made as a result, so the risks are minimal; the picture of assurance through development is in sharp contrast to much of the actual practice in academic modelling where a lot of the ‘assurance’ is done post-hoc after the analysis and just before publication.

In the Macpherson Review [1] it was stated that there was good practice to be found across Government. However, cross-government communication and the spread of best practice may not have been achieved as we saw previously with local attempts at guidance implementation that were mostly ineffective or failed altogether. There is also evidence from DSTL [7] and the Ministry of Defence that operational models that have to be used in real time can be validated and verified, especially in the operational analysis toolkit

for field guidance. This best practice could, as a minimum, have been used to ensure that models were correctly assured.

4.6 Senior Responsible Owner (SRO)

There did not appear to be any SRO mentioned, despite the emphasis in the Macpherson Review [1] on having an SRO for each model developed with the purpose to inform policy. This could be a potential culture issue due to academic modelling not being joined up with Government modellers.

4.7 Policy becoming divorced from the modelling

In the enquiry it was raised by the modellers that policy to change the public's behaviour was taking place before the modelling could be undertaken. The modeller then has to try to keep up with changes in policy and has no hope of modelling impacts given a changing environment. This raises large red flags as to what evidence these policy changes were made upon. Here, it seems that the Government is changing parameters and policy so quickly that modellers are not able to convene the necessary interdisciplinary experts. Consequently, policy decisions were made on, what was thought to be, incorrect questions and parameters being modelled.

4.8 Modelling being incorrectly discounted

In the following exchange, it is detailed that work has been undertaken around many aspects of COVID and related possible interventions but that this work was not fed into policy-making. Another issue here was that challenge was not possible to discuss the parameters of the model. As the parameters were of key importance but not added into the models, they were discounted.

B: "With regard to age, [and social distancing] there is an enormous amount of modelling activity going on around the world. (...) We have done lot of work on this in my own group and there is work out there, but this is not what has been fed down into SPI-M 5 by SAGE and the Cabinet Office" [2].

The modelling had been undertaken in order to answer the full scope of the problem, but SAGE and the Cabinet Office did not take this information on. Furthermore, it seems that challenge was not possible on the scope or type of questions to be asked or considered and interdisciplinary work and discussion fell by the wayside.

Having identified an at-risk group, it was not then modelled according to the Interviewee who states that the flu models were originally constructed to model schools. However, there is very little going on in schools and due to there never having been an outbreak in schools worldwide that we know of it is not critical to model this. The Interviewee

states, however, that Care Homes should have been modelled but were not, despite the risks being raised.

4.9 Lack of best practice and interdisciplinary working

The Interviewee states that other relevant models were not studied or examined by modellers on this team. Indeed, despite Spain and Italy having serious concerns within care homes, the care homes were not modelled. The Interviewee had not studied these models in detail and so could not say if there was any important information that should have been used in UK modelling [2].

4.10 Lack of availability of relevant data within the UK

The lack of cross-government communication seen here concerning open data sources is very disappointing. This data could potentially be critical for modelling but could not be obtained in a reasonable or timely manner.

B: "The data management systems we have in place through the NHS and NHS Scotland, which I am more familiar with, are frankly very cumbersome. It has been difficult to extract the right data at the right time for the right person in the right place. There is a lot of difficulty there. That is a historical problem; I have been complaining about it for about 10 years" [2].

4.11 Summary

This section not only illustrates the perpetuation of the issues found in the Laidlaw Report [6] and the Macpherson Review [1] but highlights new areas of concern within the availability of UK data, lack of best practice and interdisciplinary working, modelling being incorrectly discounted and modelling becoming divorced from policy. Some of these issues can be connected to leadership issues and the lack of communication and ability to challenge, i.e. the areas that AQuA guidance did not cover in a substantial way post the Laidlaw recommendations. Here it can be seen that, not only have the organisational knowledge transfer and cultural issues identified by Macpherson persisted, but further issues are highlighted where the processes prescribed and mandated within the AQuA book [7] and by Laidlaw [6] have not been implemented. The processes detailed in the AQuA book were designed and implemented to establish a route to transparency for modelling and a verified pathway to ensure models are constructed correctly and the validated and verified so that policy decisions have a sound basis.

In the Lords Select Committee on Science and Technology, 2020, it can be seen that previous steps that have been

taken have not had the intended effect. These may not have had an effect because the organisational issues were not embedded in the AQuA book or other documentation. Furthermore, this is compounded by issues of lack of relevant data, modelling being incorrectly discounted and policy diverging from the evidence-based modelling approach mandated by the UK Government for policy decision making. Addressing organisational issues seems to be key to progress in this area. One might argue that COVID modelling has an urgency that is not necessarily there for all modelling, however, in many departments, especially defence, modelling can be even more urgent than COVID. The urgency makes the requirement for robust modelling even more manifest. The next section outlines a number of recommendations with the aim to institute robust modelling across Government for both urgent and non-urgent circumstances. It is not the urgency here that drives the requirement for robust modelling but the method by which models should be kept up to date and in working order.

5 Recommendations and conclusion

Post the Laidlaw Report [6] and the Macpherson Review [1] much work was put into constructing a series of documentation to improve analytical modelling across Government. Throughout this paper, it can be seen that similar issues with analytical modelling remain and new problems both with technical processes and cultural problems have been raised by the Lords Select Committee [2]. As seen in the first section, Macpherson raised multiple issues with the ICWC franchise and Laidlaw went on to propose multiple ways forward both technically and culturally to ensure that the same mistakes were avoided in the future. As shown in the analysis of the Lords Enquiry of 2020, we not only see the same issues re-occurring but new issues coming to the fore. This could be partially due to AQuA failing to address the organisational and cultural issues. It could also be due to a lack of implementation of the AQuA book by multiple departments within the UK Government. It is clear that the processes laid out by Laidlaw and AQuA have not been successfully implemented from the testimony in the Lords Enquiry 2020.

Based on the findings we propose a unified modelling framework that takes the foundations provided by the AQuA book and addresses further the issues found within cultural and organisational contexts.

As the Macpherson Review states, adequate education and training should be provided to both leadership and technical staff to ensure fit for purpose analysis. We propose that this education should begin at university level where ethics, modelling and knowledge transfer are introduced within courses to establish a culture of knowledge transfer

and learning as well as technical best practice. Modelling is undertaken in technical sciences at university but the Benchmark Statements⁷ include this as part of a larger degree and so there can be a lack of focus on the teaching of modelling or ethics as a specific skill. The skills required to use a basic statistical model would not be sufficient to start from scratch and build a model reflecting real-world scenarios with which to inform policy. This is a skill in itself and includes such aspects as awareness of data quality, ethics, user implementation problems, context and an understanding of the environment that the model is being created in, for example, defence or public sector, where there can be a high price to pay for faulty analysis.

We also propose that, despite downsizing or austerity measures, even small organisations, with the right culture can embed best practice. Boeing, Xerox and IBM have shown that huge strides can be made once the organisation culture is developed to enable them. Therefore, despite downsizing having occurred within the civil service, processes are able to be adapted to ensure the continuation of robust analysis [28].

The recommendations from this paper are as follows:

- *Clear, accountable, communicative and fully trained leadership:* Leadership must be involved as no less than a cultural shift is needed to implement this. Leadership bears the responsibility of ensuring communication and challenge are facilitated as well as ensuring guidance is followed in terms of robust modelling. Leadership training in this area is crucial [38]. As highlighted by the analysis of the Laidlaw Report, the Macpherson Review and the Lords Select Committee from June 2020, leadership and management have been key to ensuring that communication and challenge are necessary for robust modelling to occur. Therefore, it is crucial to ensure that adequate training is given [1] in this area to those in management positions in departments where modelling occurs. This training should be developed beyond generic aspects related to leadership and management to include aspects related to modelling, understanding the whole process of modelling and the leadership and management implication at each stage in this process.
- *Transparency in both modelling, data and communication:* Transparency is a key element of the data collection [8, 16] and modelling process. When modelling is undertaken, the risks, caveats and assumptions must be explicit in the accompanying paperwork. The limitations must also be explicit. This means that the model and data can be used for the purpose it is currently validated and veri-

⁷ Benchmark Statements are the detail of a taught degree and justify content as well as assessment.

fied for. Any changes may then be tracked such that the model can be adapted for different circumstances until a limit of change is reached that triggers another validation and verification cycle [39]. Transparency then drives challenge due to the explicitly clear function and development of the model. Scrutiny can then be performed more efficiently either inside or outside the domain of the model.

- *Facilitate constructive challenge*: Foster an environment of constructive challenge, especially around Policy. This area should not be the domain of a few committee members to feed down narrow questions; the modellers should have the ability to use the data and model what they believe, in their expert opinion, to be a valid way forward [3, 13]. It must be a two way conversation. As we have seen in the Lords Enquiry, modellers have sometimes been denied the opportunity of challenge within their work which has led to experts producing modelling that they are ready to admit is not robust. Unfortunately, modelling can have a huge impact on UK Society (such as the resulting policy seen in 2020) and as such it is not appropriate to bypass expert opinion without sufficient communication. If a policy expert is unaware of the modelling being undertaken then the context and assumptions might not be passed on. Therefore, if a modeller is asked to model something specific and cannot challenge this then there may be large errors where context, reality and assumptions are not captured. The modelling asked for may not be appropriate. Therefore it is crucial that modelling experts be able to have open communication with those for whom they will be modelling. To create an open challenge environment leadership is key. Wu [32] outlines that a communication methodology is key and states that it is crucial in developing the team for success. Without open communication the team cannot learn from conflict or implement processes robustly [32]. Where Government departments do use JSP655⁸ for evidence-based modelling requirements, project and programme execution for investment approvals it does not specify communication channels, working roles, methods of risk discussion or routes for raising challenges. This could be ameliorated by taking best practice from such methodologies as SCRUM and Prince II that are very clear on aspects such as roles and communication channels.
- *Guidance must be based on existing modelling best practice*: There is a wealth of knowledge such as the AQuA book that can be built on very easily. This type of documentation can be used internally or industry wide as it

is somewhat model agnostic. The conceptual modelling journey overarches how to build models not the specifics of building them. This in turn is subject to leadership, availability of those with the right set of skills who are able to model well and the existing quality assurance procedures being used and developed further [1]. The right professionals must be recruited to construct and use models as modelling requires a huge amount of training, that is becoming more of a burden as machine learning and AI is being introduced [40–44]. This means that professionals involved in modelling are becoming even more specialised on tools and data. This should be reflected in recruitment.

- *Provide support and training on robust modelling process for practitioners*: Focus on supporting practitioners in collecting and analysing high-quality data sets with a focus on how the data is collected and the statistical robustness of the initial data. As detailed in Macpherson and AQuA, professionals must be supported with ongoing training and developments so that they can continue to be effective in their role and continue to build robust modelling. Training also should extend to policy makers and management within departments that produce modelling so that aspects such as the risks, assumptions and context of the model are well understood [1, 8, 36].
- *Provide support and guidance on robust data collection and use*: Understanding what data is present and how it relates to the real world. Understanding the purpose of the required modelling in collaboration with domain experts. Data quality and accessibility is key for model developers. If a modeller is not able to access the correct data then the model cannot be robust. Collecting new data might also be necessary and this should be done robustly to ensure the right data is collected for the model. Indeed, this guidance already exists in part, published by the UK Government Statistical Service [45] and the European Statistical Service [46].
- *Recognise modelling as a skilled profession*: Allocate modelling as a skilled profession with relevant recognition; alongside domain-specific educational resources, courses and tools [47–49]. If we wish to recognise experts in modelling and recruit them correctly for our needs, we must have an understanding of who is qualified. There are numerous professional charterships one can rely on to judge the experience of a professional but they do not indicate the particular areas of technology of modelling that the professional is qualified in. As we move towards an ever more complex environment of tools and methodologies it is crucial to understand the skills we need for a project and where to find them [50].
- *Address gaps at primary, secondary and tertiary educational levels*: Implement training in this area from school

⁸ Joint service publication (JSP) 655: Investment Approvals, which is the policy and guidance on modelling and evidence required for investment approvals and scrutiny.

level so that the complex concepts such as assumptions, caveats, quality assurance and answering the right questions with constructive challenge become a cultural fixture [51]. Subject Benchmark Statements can be updated in Higher Education so that it is clear that ethics [52], context and the modelling lifecycle must be covered. In addition, leadership and soft skills should be taught so that a cultural shift and continuous improvement mentality occurs downstream within organisations.

- *A suite of transparent crisis response models with guidance on adaptations:* In situations of urgency, it is advisable to have on hand a set of models that are validated and verified. This is advisable for high risk situations, or indeed business critical models, such as defence and health. These models can then be used by modellers to adapt to the current circumstances. In high risk situations it is crucial that an open and flowing communication channel is used.

The recommendations above form a pathway to more improved, robust modelling. It is crucial for the future of the UK that we develop ways to communicate and challenge so that the technical expertise can flourish, especially within policy. The situations discussed above affect the entire of UK and every inhabitant. Therefore, we owe it to them to implement the best solutions based on rigorous, ethical, evidence-based analysis.

Funding This work has not been funded.

Availability of data and material Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The corresponding author verifies on behalf of all authors that there is no conflict of interest.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will

need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Macpherson.: Review of quality assurance of government analytical models. [Online] <https://www.gov.uk/government/publications/review-of-quality-assurance-of-government-models> (2013). Accessed 20 Mar 2021
2. Lords: Lords Select Committee on Science & Technology. [Online] <https://committees.parliament.uk/committee/193/science-and-technology-committee-lords> (2020). Accessed 12 Apr 2021
3. Mashelkar, R.A.: Impact of science, technology and innovation on the economic and political power. *AI & Soc.* **32**, 243–251 (2017). <https://doi.org/10.1007/s00146-015-0639-y>
4. House of Commons: Committee of Public Accounts. [Online] <https://publications.parliament.uk/pa/cm201213/cmselect/cmpubacc/813/813.pdf> (2012). Accessed 20 May 2021
5. House of Commons: Committee of Public Accounts. [Online] <https://publications.parliament.uk/pa/cm200910/cmselect/cmpubacc/390/39003.htm> (2010). Accessed 25 May 2021
6. Laidlaw, S.: Report of the Laidlaw Inquiry: Inquiry into the lessons learned for the department for transport from the intercity west coast competition. [Online] <https://www.gov.uk/government/publications/report-of-the-laidlaw-inquiry> (2012). Accessed 20 Mar 2021
7. DSTL. AQuA book. London: HM Treasury. [Online] <https://www.gov.uk/government/publications/the-aqua-book-guidance-on-producing-quality-analysis-for-government> (2015). Accessed 12 Apr 2021
8. Oldfield, M., Smalley, J., McMonies, M.: [Online] <https://committees.parliament.uk/work/570/data-transparency-and-accountability-covid-19/publications/written-evidence/> (2020). Accessed 25 May 2021
9. Department for Business. Quality assurance: Guidance for Models. [Online]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/737293/BEIS_QA_Guidance_for_Models.pdf (2018). Accessed 3rd May 2021
10. Knoppers, B.M., Thorogood, A.M.: Ethics and big data in health. *Curr. Opin. Syst. Biol.* **4**, 53–57 (2017). <https://doi.org/10.1016/j.coisb.2017.07.001>
11. ICO: Guide to Data Protection. [Online]. <https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-artificial-intelligence/part-1-the-basics-of-explaining-ai/definitions/> (2021). Accessed 3 Mar 2021
12. ICO: Explaining Decision made with AI. [Online]. <https://ico.org.uk/for-organisations/guide-to-data-protection/key-data-protection-themes/explaining-decisions-made-with-ai/> (2020). Accessed 20 May 2021
13. Robinson, A., Glover, P.: Developments in the quality assurance of government models used to support business critical decisions. In Proceedings of the Operational Research Society Simulation Workshop, pp.176–181 (2014). https://www.theor-society.com/media/3591/sw14-proceedings-book-final-hw_20042016101209.pdf#page=181
14. Holstein, K., Wortman Vaughan, J., Daumé III, H., Dudik, M., Wallach, H.: Improving fairness in machine learning systems: What do industry practitioners need? In: Proceedings of the 2019 CHI conference on human factors in computing systems, pp.1–16 (2019). <https://doi.org/10.1145/3290605.3300830>

15. Dodge, J., Liao, Q. V., Zhang, Y., Bellamy, R. K., Dugan, C.: Explaining models: an empirical study of how explanations impact fairness judgment. In: Proceedings of the 24th International Conference on Intelligent User Interfaces, pp. 275–285 (2019). <https://doi.org/10.1145/3301275.3302310>
16. Grimm, V., Augusiak, J., Focks, A., Frank, B.M., Gabsi, F., Johnston, A.S., Liu, C., Martin, B.T., Meli, M., Radchuk, V., Thorbek, P.: Towards better modelling and decision support: documenting model development, testing, and analysis using TRACE. *Ecol. Model.* **280**, 129–139 (2014). <https://doi.org/10.1016/j.ecolmodel.2014.01.018>
17. Lenk, H.: Ethics of responsibilities distributions in a technological culture. *AI & Soc.* **32**, 219–231 (2017). <https://doi.org/10.1007/s00146-015-0642-3>
18. Peters, L.: The simple truths about leadership: creating a people-centric culture. Springer (2019)
19. Rezvani, A., Barrett, R., Khosravi, P.: Investigating the relationships among team emotional intelligence, trust, conflict and team performance. *Team Perform. Manag.* **25**(1–2), 120–137 (2019). <https://doi.org/10.1108/TPM-03-2018-0019>
20. Harrison, C., Burnard, K., Paul, S.: Entrepreneurial leadership in a developing economy: a skill-based analysis. *J. Small Bus. Enterp. Dev.* **25**(3), 521–548 (2018). <https://doi.org/10.1108/JSBED-05-2017-0160>
21. Acha, S., Mariaud, A., Shah, N., Markides, C.N.: Optimal design and operation of distributed low-carbon energy technologies in commercial buildings. *Energy* **142**, 578–591 (2018). <https://doi.org/10.1016/j.energy.2017.10.066>
22. Appelbaum, D.A., Kogan, A., Vasarhelyi, M.A.: Analytical procedures in external auditing: a comprehensive literature survey and framework for external audit analytics. *J. Acc. Lit.* **40**, 83–101 (2018). <https://doi.org/10.1016/j.acclit.2018.01.001>
23. Bessick, J., Naicker, V.: Barriers to tacit knowledge retention: an understanding of the perceptions of the knowledge management of people inside and outside the organisation. *S. Afr. J. Inform. Manag.* **15**(2), a556 (2013). <https://doi.org/10.4102/sajim.v15i2.556>
24. Hamon, R.J., Junklewitz, H., Sanchez Martin, J.: Robustness and explainability of artificial intelligence. Luxembourg, Publications Office of the European Union (2020). <https://doi.org/10.2760/57493>
25. Kazim, E., Koshiyama, A.: Explaining decisions made with AI: a review of the co-badged guidance by the ICO and the Turing Institute. SSRN (2020). <https://doi.org/10.2139/ssrn.3656269>
26. Garvin, D.: Building a learning organization. *Harv. Bus. Rev.* **71**(4), 78–91 (1993)
27. Gino, F., Staats, B.: Why organizations don't learn. *Harv. Bus. Rev.* **94**(1–2), 24–24 (2016)
28. Fisher, S.R., White, M.A.: Downsizing in a learning organization: are there hidden costs? *Acad. Manag. Rev.* **25**(1), 244–251 (2000). <https://doi.org/10.5465/amr.2000.2791613>
29. Goodman, B.W.: A step towards accountable algorithms? Algorithmic discrimination and the European union general data protection. In 29th Conference on Neural Information Processing Systems, Barcelona. NIPS Foundation (2016). <http://www.mlandthelaw.org/papers/goodman1.pdf>
30. Robinson, S.: General concepts of quality for discrete-event simulation. *Eur. J. Oper. Res.* **138**(1), 103–117 (2002). [https://doi.org/10.1016/S0377-2217\(01\)00127-8](https://doi.org/10.1016/S0377-2217(01)00127-8)
31. NAO: Framework or Review Models. [Online] http://www.nao.org.uk/wp-content/uploads/2016/03/11018-002-Framework-to-review-models_External_4DP.pdf (2016). Accessed 20 Mar 2021
32. Wu, G., Liu, C., Zhao, X., Zuo, J.: Investigating the relationship between communication-conflict interaction and project success among construction project teams. *Int. J. Project Manag.* **35**(8), 1466–1482 (2017). <https://doi.org/10.1016/j.ijproman.2017.08.006>
33. Lords.: Committee on Artificial Intelligence. [Online] <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf> (2018). Accessed 20 Mar 2021
34. Crane, T.: The mechanical mind: a philosophical introduction to minds, machines and mental representation. Routledge (2015)
35. Edmonds, B.: Review of The AQUA book: guidance on producing quality analysis for Government. [Online] <http://jasss.soc.surrey.ac.uk/19/3/reviews/7.html> (2016). Accessed 26 Mar 2021
36. Oldfield, M., McMonies, M.: Call for written evidence risk assessment and risk planning. Lords Select Committee, London. [Online]. <https://committees.parliament.uk/committee/483/risk-assessment-and-risk-planning-committee/publications/> (2021). Accessed 19 May 2020
37. Norling, E., Edmonds, B., Meyer, R.: Informal approaches to developing simulation models. In: Edmonds B., Meyer R. (eds.) *Simulating Social Complexity. Understanding Complex Systems*, pp 61–79, Springer, Cham (2017). https://doi.org/10.1007/978-3-319-66948-9_5
38. Yammarino, F.J., Salas, E., Serban, A., Shirreffs, K., Shuffler, M.L.: Collectivistic leadership approaches: putting the “we” in leadership science and practice. *Ind. Organ. Psychol.* **5**(4), 382–402 (2012). <https://doi.org/10.1111/j.1754-9434.2012.01467.x>
39. Berglund, J.: Why safety cultures degenerate: and how to revive them. Routledge (2016)
40. Coeckelbergh, M.: Language and technology: maps, bridges, and pathways. *AI & Soc.* **32**, 175–189 (2017). <https://doi.org/10.1007/s00146-015-0604-9>
41. Johannessen, J.-A.: Knowledge management for leadership and communication: AI, innovation and the digital economy. Emerald Group Publishing (2020)
42. McCarthy, J., Hayes, P. J.: Some philosophical problems from the standpoint of artificial intelligence. In: *Readings in artificial intelligence*, pp 431–450 (1981). <https://doi.org/10.1016/B978-0-934613-03-3.50033-7>
43. Morley, J., Floridi, L., Kinsey, L., Elhalal, A.: From what to how: An initial review of publicly available AI Ethics tools, methods and research to translate principles into practices. *Sci. Eng. Ethics* **26**, 2141–2168 (2020). <https://doi.org/10.1007/s11948-019-00165-5>
44. Sokol, K., Flach, P.: Explainability fact sheets: a framework for systematic assessment of explainable approaches. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp. 56–67 (2020) <https://doi.org/10.1145/3351095.3372870>
45. GSS. Code of Practice for Statistics. [Online]. <https://gss.civilservice.gov.uk/policy-store/code-of-practice-for-statistics/> (2018) Accessed 1 May 2021
46. ESS. Quality assurance Framework of the European Statistical System. European Statistical System Committee. [Online]. <https://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646> (2019). Accessed 1 Apr 2021
47. Gebru, T., Morgenstern, J., Vecchione, B., Wortman Vaughan, J., Wallach, H., Daumé III, H., Crawford, K.: Datasheets for Datasets (2018). <http://arxiv.org/abs/1803.09010>
48. Kwak, S.K., Kim, J.H.: Statistical data preparation: management of missing values and outliers. *Korean J. Anesthesiol.* **70**(4), 407–411 (2017). <https://doi.org/10.4097/kjae.2017.70.4.407>
49. Kallus, N., Zhou, A.: Residual unfairness in fair machine learning from prejudiced data. In: Proceedings of the 35th International Conference on Machine Learning, pp. 2439–2448 (2018). <http://proceedings.mlr.press/v80/kallus18a.html>
50. Borgmann, A.: Real American ethics: taking responsibility for our country. University of Chicago Press (2010)

51. Gkamas, V., Rigou, M., Paraskevas, M., Zarouchas, T., Perikos, I., Vassiliou, V., Gueorguiev, I., Varbanov, P., Sharkov, G., Todorova, C., Sotiropoulou, A.: Bridging the skills gap in the Data Science and Internet of Things domains: A Vocational Education and Training Curriculum. In: Proceedings of the 2019 ICDE World Conference on Online Learning, pp. 312–320 (2019)
52. Chae, Y.: US AI regulation guide: legislative overview and practical considerations. *J. Robot. Artif. Intell. Law* **3**(1), 17–40 (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.