

Please cite the Published Version

Edmonds, Bruce (2016) Book Review: The Aqua Book: Guidance on Producing Quality Analysis for Government. JASSS-THE JOURNAL OF ARTIFICIAL SOCIETIES AND SOCIAL SIMULATION, 19 (3). ISSN 1460-7425

Publisher: University of Surrey

Version: Published Version

Downloaded from: <https://e-space.mmu.ac.uk/621587/>

Usage rights: © In Copyright

Additional Information: This is an Open Access article published in JASSS, published by University of Surrey, copyright JASSS.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

© Copyright JASSS



The Aqua Book: Guidance on Producing Quality Analysis for Government

HM Treasury

HM Treasury: United Kingdom, 2015

ISBN 978-1-910337-67-7

Reviewed by **Bruce Edmonds**

Manchester Metropolitan University



The Aqua Book:
guidance on producing quality
analysis for government

March 2015

In 2012 the UK government ran a competition for who would run a particular service (the InterCity West Coast rail franchise). Civil servants modelled the various bids to assess its costs and benefits and a winner of the competition announced as a result. Unfortunately for the government, this decision was challenged in the courts and it turned out that there were mistakes in the modelling, and the decision was retracted. This caused the government some embarrassment as well as cost it quite a bit of money because it had to re-

run the competition. This illustrated the importance of analysis and modelling in delivering a major government project as well as what can happen if the modelling was of poor quality.

As a result the government instituted an extensive review of models and modelling practice with its decision-making processes. This review (MacPherson rep, interim, appendix) catalogued the models currently within use within the UK government and came up with a list of recommendations as to future modelling practice. On that basis of this review a committee was formed which developed mature recommendations in the "Aqua Book". This is a review of this book along with its associated resources.

Although these resources were designed for civil servant modellers they include a lot of advice that is highly relevant for other modellers. It is also significant in that it is the first set of recommendations for policy modelling developed by a government that I know of [1]. As a result this might develop into an established standard for modelling, and certainly for models that might be used within government as part of a policy-making process. In general I feel that there is insufficient dialogue between academic and government modellers. The UK government is starting to address this, with initiatives to involve more outside expertise in government modelling exercises. This review is partly to raise awareness among academics to the developments within Government but also to learn from these for non-government modellers.

The Aqua Book (AB, HM Treasury 2015a) starts off with its rationale (p.9):

1.1 Analysis is vital to the success of policy development and the delivery of programmes, projects and operational services. Analysis helps to shape and appraise options, provides insight into how complex systems work and behave, measures system performance and improves efficiency.

1.2 However, if analysis and any supporting models, data and assumptions are not fit-for-purpose then the consequences can be severe ranging from financial loss through to reputational damage and legal challenge. In the most severe of consequences, lives and livelihoods can be affected.

Thus it recognises that analysis and modelling in general is central to the efficient and intelligent delivery of services. Indeed the MacPherson review already lists hundreds of models already used in various departments for a wide variety of purposes and subject to different levels of quality assurance (HM Treasury (2013c)). However it also vividly illustrates the approach of the government to such models which is to seek to prevent modelling 'errors'.

The aim of AB is summarised as follows (p. 7):

...to extend best practice across the whole of government. They focus on quality assurance, governance and accountability, culture, capacity, capability and control. ... It outlines a sensible, achievable set of principles. These principles will help ensure that our work can be trusted to inform good decision making. ...we need to create an environment where the skills and

time to deliver analysis is respected, and a culture that values it is encouraged.

The book goes on to make recommendations in three areas: identifiable roles to cover different areas of responsibility, the right modelling environment, and the analytic process, ending with a summary of resources to aid modellers.

Unsurprisingly, in a hierarchical civil service where service flows upwards to the minister (an elected politician) and decisions flow downwards, there is a concern that identifiable responsibilities be established. Thus it recommends the creation of four roles for each model:

- the senior responsible officer (SRO) who is ultimately accountable,
- the commissioner who specifies and appoints the analytic team to do the modelling,
- the person who assures and checks the quality control,
- and the person(s) who do the analysis.

These four roles might be combined in the same person, but remain separate for the purposes of these guidelines. During the whole process a quite intensive collaboration is envisaged between these with almost continual communication between them being a prime recommendation of the report. Thus the AB book mirrors the more general policy development cycle [2] in which the modelling is embedded (Green Book, GB; HM Treasury 2013a). In contrast in academic modelling these are often not distinguished, being all performed by a single person. However, it might be much better if modellers did have a real 'customer' for whom the modelling is done rather than this being merely a narrative that is invented post-hoc to justify their effort, and the process of specifying and deciding the modelling approach be more independent of the model assurance as well as the model development and analysis itself.

The right modelling environment involves (p. 9):

...a culture where leaders value and recognise good quality assurance. It requires adequate capacity, including specialist skills and sufficient time to

conduct quality assurance effectively. It also needs a set of controls... and a route for challenge where analysts have concerns.

This is an area where academics have relatively high levels of time and available skill, as well as a free ability to challenge any modelling decisions. On the other hand the level of quality assurance can be low, with modellers basically making it up as they go along. Higher standards could definitely be developed in this regard.

The most interesting part of the AB for social simulators will be the guidance that is developed for the modelling process itself. Here the AB makes four broad recommendations (p. 10):

- Proportionality of response (in relation to risks and amount of use)
- Assurance throughout development (quality assurance throughout model lifecycle including documentation)
- Verification and validation (checking fit for purpose)
- Analysis with RIGOUR: quality analysis needs to be Repeatable, Independent, Grounded in reality, Objective, have understood and managed Uncertainty, and the results should address the initial question Robustly

Often academic modelling is done in the abstract with no danger of significant decisions being made as a result, so the risks are minimal. However, as social simulation becomes less experimental and more useful this will have to change as models get used as part of decision-making processes. 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 (Norling *et al.* 2013). Checking fit for purpose presumes that the exact purpose for a model is specified, which is often not the case in academic modelling [3].

The 'RIGOUR' criteria for quality modelling are all wonderful ideals, but differ wildly in how achievable they are. Repeatability can be greatly improved through adequate documentation and is closely associated with an ability to replicate models. All academic modelling should be independent, but is often highly biased by what is traditional in a field or what is easier to implement –

academic modelling might be relatively free of outside biases but it generates whole forests of internal ones [4]. I and others have long argued that social simulation should be more grounded in evidence rather than hoping abstract models will be 'roughly right' [5]. If we knew how to make a social simulation 'robust' we would be getting Nobel prizes.

The issue of uncertainty goes to the heart of much social simulation. Much in the AB about uncertainty assumes two connected things: (a) that they are dealing with the kind of non-chaotic mathematical/accounting models where uncertainty can be estimated and (b) what they are modelling is such that all the important factors can be accounted for. This is linked to the implicit purpose for many of the models the AB talks about which are to forecast the effect of possible decisions. For many social phenomena there are always unmodelled processes that could overwhelm the results of any analysis, and where analytic mathematical models [6] are clearly inadequate. Here forecasting is (almost always) not a purpose that can be fulfilled but rather the aim is to explain observed data with a complex process, or to do a kind of risk-analysis - discovering some of the complex processes which could result from a situation.

The Aqua Book is divided into three parts. Part A is aimed at the SRO's and commissioners of analysis, covering: how commissioning and specifying analysis relates to the decision-making processes it is embedded within (Chapter 2); how to ensure the modelling is fit for purpose through quality assurance (Chapter 3); Chapter 4 goes into how to implement the quality assurance; and Chapter 5 looks at documenting and assessing the uncertainty in what is modelled. Part B is aimed at the analytical assurer and the analyst. It seeks to clarify the responsibilities of the various roles within the modelling life cycle. In this part: Chapter 6 looks at verification and validation; how to check that the model has been checked (Chapter 7); and analysing the vulnerability of a model to sources of uncertainty. The final section, Part C is a single chapter (Chapter 9) listing associated helpful resources (HM Treasury *et al.* (2015b)).

I will not describe all that is of interest in this book, but only point out some of the highlights.

- It acknowledges a variety of modelling types and purposes. The purposes

that are distinguished (p. 14) are: testing systems under a variety of scenarios, carrying out investigations to understand a problem in more detail, enabling the monitoring of processes to facilitate risk management, comparing and appraising options, and understanding past behaviour to better prepare for the future. This is helpful because clarity in identifying and specifying the purpose of any particular modelling exercise is vital. A lack of clarity or even a conflation of purposes is, in my view, one of the major current weaknesses in social simulation modelling, allowing poor quality modelling to be published and deceiving potential modelling clients as to their power and usefulness. However this particular list does not cleanly distinguish different kinds of modelling project, and I suspect that many of these will simultaneously hold for many projects. Also the link from purpose to the kind of checks needed on the modelling is missing here – what is necessary to get a better understanding of the past is very different from that needed for anything requiring accurate forecasting.

- The list of types of model (p. 15) is clearer, probably because these have grown out of modelling practice within government. They are: policy simulation (to better understand the consequences of policy options, e.g. distributional impact of tax and benefit changes), forecasting the future (e.g. future energy demand), financial evaluation (e.g. working out costs of a project), procurement and commercial evaluation (e.g. evaluate bids for service franchises), planning (e.g. assess number of teachers that need to be trained), ‘science based’ (the modelling of the physical environment, e.g. likelihood of floods), allocation of funds (how they are distributed over regions and services), and conceptual (“understand the key influences that are important to a system”). These are distinguished by the history of their use, the areas they are applied to, their associated techniques and their purposes. Recognising and retaining their different identities is probably wise as their associated practices will have subtly developed over the years to suit their purposes and the processes the modelling is embedded within.

- The book goes into quite a lot of detail about which of the above roles should be involved at each stage of the modelling process and what their responsibilities are at these stages. Throughout it stresses the importance of communication, getting away from a model where the modellers go off for a period of time and come back with the ‘answers’ and towards one with continual engagement between the people involved. Thus this picture of

modelling is very different from the one found in academia, which tends to be a largely solitary activity (with the exception of those at the participatory end of the modelling spectrum, such as ‘companion modelling’ (handbook chap).

- The book spends a lot of time, one way or another, talking about the assessment and management of ‘uncertainty’ (within the context of its attitude to risk as described in the ‘Orange Book’ (OB; HM Treasury 2013b) [7]). This is clearly one of the main targets for improvement and (reading between the lines) one of the principle causes of friction between modellers and other policy actors. Here the AB basically recommends that: (a) a much greater proportion of the modelling attention and effort should be directed at this, (b) that the uncertainties that affect a modelling exercise should be communicated more often, more honestly and in non-ambiguous language and (c) to avoid communicating unwarranted confidence in the described outcomes. This is all good advice and should be adsorbed by social simulators. However the advice is all written from the point of view of what should be done by the various modelling roles, implying that the principle responsibility for good policy modelling lies with these. However, an equal responsibility lies with those that ‘use’ modellers and their analyses. Considerable anecdotal evidence suggests that policy actors have frequently made unrealistic demands upon modellers, seeking to offload some of the responsibility of decision making to some ‘science’ [8], as well as being impatient of the caveats and warnings that modellers include to accompany their results [9]. The book acknowledges the existence of ‘deep uncertainty’ but seeks to bracket it off. This kind of uncertainty will become increasingly relevant and will require a fundamental change in the way decision making is made (two chaps) – it will not be fixed or avoided by simply improving existing modelling practice (Edmonds & Gershenson 2015, Jager & Edmonds 2015).

- The book rightly places a big emphasis on verification and validation (Chapter 6). One of the interesting new developments in this regard is the continual documentation of the whole process including a log of: the verification and validation checks done, the changes and modifications and the reasons for making them, the various versions along with their differences in data and assumptions, as well as a list of issues that arise and risks identified. This compares with the suggestion that simulators of complex systems should publish ‘TRACE’ documents that log similar things

along with their results (Grimm *et al.* 2014).

- Chapter 6 also recommends the ‘independent’ scrutiny and critique of modelling. Whilst a variety of internal scrutiny is described (this being the point of the separation of roles described above) it also briefly mentions ‘peer review’. There are signs that the government is increasingly seeking some kind of independent views on its modelling, but *how* this is done is critical to its effectiveness. Much of this peer review would not meet current best practice within social simulation, where the model code and a complete description of the model implementation are required. Rather, often critique of only a broad overview or summary paper of the modelling is sought. I can understand why politicians or senior civil servants might wish to hide the basis of their decision making from the public that pays for it, but ultimately that is not in the interest of good policy development. Modelling should move with current trends towards transparency of governance and policy making and make more modelling completely transparent – publishing the code of their models (Edmonds and Polhill 2015). It might generate more ‘noise’ from academics as they seek to nit-pick at the model assumptions and techniques, but the public would become used to this and it would result in much sounder modelling (and hence the quality of policy developed as a result) in the longer-term.

- Finally, I would like to heartily recommend the associated set of resources that accompanies the AB (HM Treasury *et al.* 2015b) and is summarised in Chapter 9. This includes more detailed discussion and detail about verification & validation as well as a wonderful and succinct check list for reviewing models (National Audit Office 2016).

In conclusion there is much here to interest and inform the social simulation community in terms of modelling approach and practice. It also indicates that government modellers and academic modellers have a lot that they could usefully swap in terms of modelling techniques and practice. People do produce models for or with government departments and agencies will have to understand and take on board this guidance, but the rest of us could very usefully adopt much that is here.

Of course, even though this is very well considered and useful advice – there is very little that one could disagree with here – there are areas one could

criticise. The whole focus of the AB is on how modelling practice and management can be improved to provide a better service for policy makers. This will make modelling more onerous to do and hence increase its cost, maybe with a result in reducing the amount of modelling done in support of policy (at least in the short term), reflecting the precautionary principle implied in all of this. It might be concluded that it is generally better not to model than risk making a mistake, so if one was taking a blame-avoidance strategy one might deliver worse policy just because a perfect model was not possible [10].

The second main critique is that there is not equal weight on proving guidance with the policy 'clients' of modelling, the senior civil servants and politicians. Maybe this is simply not possible given the norms of the civil service, or maybe the AB is simply not the place for this, but the way models are used in a complex and radically uncertain world is at least as important as the modelling process itself. The MacPherson review (HM Treasury 2013c) was more explicit about this but these elements have largely disappeared in the development of the AB. I hope that this is, in fact, being addressed within government but I fear the AB marks a shift of responsibility for errors down towards the modellers.

Finally, the AB gives an impression of the pressure that government modellers are under. They often have to work to very tight deadlines and to increasingly high standards. Fortunately the UK government is now taking the development of modelling skills and capacity seriously by supporting them more and with its internal professionalization within the civil service. One can always decide policy on the basis of one's own intuition and qualitative advice especially when the development of quality modelling is so expensive in terms of time and resources. However a good model can deliver huge marginal benefits in the longer-term and even aid in the avoidance of major disasters, so it can be a very worthwhile investment. The next time someone complains about the level of their taxes and how it is all wasted on bureaucrats and civil servants, think about the sheer value that these modellers are delivering.

Notes

[1] Although the Joint Research Council of the EU is in the process of

developing something comparable for models used to assess ICT-enabled Social Investment Projects (Misuraca & Kucsera 2016).

[2] This cycle is captured by the acronym ROAMEF within its 'Green Book' (GB): "For any given decision, the Rationale and Objectives must be understood. Following Appraisal of the options and implementation of the decision the outcome should be Monitored, Evaluated and the original rationale reconsidered with the completion of the Feedback".

[3] For example a confusion between explaining known data and forecasting unknown data (Edmonds & Moss 2001).

[4] To the extent, as (Kuhn 1962) pointed out, that we don't even see many assumptions for what they are.

[5] e.g. (Moss & Edmonds 2005, Edmonds 2010).

[6] By which I mean those that can be solved, if they can't be solved one has to simulate them anyway.

[7] Surely no one can say the UK civil service is not colourful!

[8] In the worst case they are only looking for modeling that supports the decisions they have already made.

[9] This was one of the main points that came out during a workshop at the Royal Society on the relationship between modellers and civil servants. It was reported how the caveats that modellers wrote to accompany their conclusions were often edited out as their results were reported up the policy tree, leading to the modellers becoming increasingly reluctant to make definite conclusions.

[10] Of course, whether this is the right decision depends on a lot of factors, including the risk of mistakes and the danger of deceiving oneself with a model as the AB documents.



References

EDMONDS, B. (2010). Bootstrapping Knowledge About Social Phenomena Using Simulation Models. *Journal of Artificial Societies and Social*

Simulation, 13(1):8. <http://jasss.soc.surrey.ac.uk/13/1/8.html>.

EDMONDS, B. & Gershenson, C. (2015). Modelling Complexity for Policy: opportunities and challenges. In Geyer, R. & Cairney, P. (eds.) *Handbook on Complexity and Public Policy*. Cheltenham: Edward Elgar, 205–220.

EDMONDS, B. & Moss, S. (2001). The Importance of Representing Cognitive Processes in Multi-Agent Models, Artificial Neural Networks – ICANN'2001, August 2001, Vienna, Austria. In: Dorffner, G., Bischof, H. and Hornik, K. (eds.), *Lecture Notes in Computer Science*, 2130: 759–766.

EDMONDS, B. & Polhill, G. (2015). Open Modelling for Simulators. In Terán, O. & Aguilar, J. (Eds.) *Societal Benefits of Freely Accessible Technologies and Knowledge Resources*. Hershey, USA: IGI Global, 237–254.

GRIMM, V., *et al.* (2014). Towards better modelling and decision support: Documenting model development, testing, and analysis using TRACE. *Ecological Modelling*, 280, 129–139. <http://dx.doi.org/10.1016/j.ecolmodel.2014.01.018>.

HM TREASURY (2013a, updated 2015). *The Green Book: appraisal and evaluation in central government*. <https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-government>.

HM TREASURY (2013b). *Orange Book: Management of risk – Principles and Concepts*. <http://gov.uk/government/publications/orange-book>.

HM TREASURY (2013c). *Review of quality assurance of government models*. <http://www.gov.uk/government/publications/review-of-quality-assurance-of-government-models>.

HM TREASURY (2015a). *The Aqua Book: guidance on producing quality analysis for government*. <http://gov.uk/government/publications/the-aqua-book-guidance-on-producing-quality-analysis-for-government>.

HM TREASURY, Department for Business, Innovation & Skills, Department of Energy & Climate Change and + others (2015). *Aqua Book resources*.

<http://www.gov.uk/government/collections/aqua-book-resources>.

JAGER, W. & Edmonds, B. (2015). Policy Making and Modelling in a Complex world. In Janssen, M., Wimmer, M. and Deljoo, A. (eds.) *Policy Practice and Digital Science*. Heidelberg: Springer, 57–74.

KUHN, T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

MISURACA, G. & Kucsera, C. (eds.) (2016). *Proposed methodological framework to assess the social and economic impact of ICT-enabled social innovation initiatives promoting social investment in the EU-i-FRAME V1.5 (D2.2)*. Institute for Prospective Technological Studies, Joint Research Centre. http://ec.europa.eu/jrc/sites/default/files/IESI_D2.2_i-FRAME-V1.5_DRAFT_V1.0.pdf.

MOSS, S. & Edmonds, B. (2005). Towards Good Social Science. *Journal of Artificial Societies and Social Simulation* 8(4):13
<http://jasss.soc.surrey.ac.uk/8/4/13.html>.

NATIONAL AUDIT OFFICE (2016). *Framework to review models*.
<http://www.nao.org.uk/report/framework-to-review-models/>.

NORLING, E., Edmonds, B. & Meyer, R. (2013). Informal Approaches to Developing Simulation Models. In Edmonds, B. & Meyer, R. (eds.) *Simulating Social Complexity – A Handbook*. Heidelberg: Springer, 39–55.