International Conference on Hydroinformatics

2014

A Model To Structure Scientific Information For Decision Makers

Rebecca Kristine Schmidt

Damian Barrett

Simon Gallant

A MODEL TO STRUCTURE SCIENTIFIC INFORMATION FOR DECISION MAKERS

REBECCA K SCHMIDT (1), DAMIAN J BARRETT (1), SIMON GALLANT (1)
(1): Commonwealth Scientific and Industrial Research Organisation, PO Box 1666, Canberra ACT 2601, Australia

Scientific information for decision makers needs to be discoverable and transparent so that decisions are more easily based on scientific evidence. One way to achieve these goals is to specify a robust model that structures this scientific information. This paper presents a model that governs the scientific activities, inputs and products for bioregional assessments (BAs), which provide information on the ecology, hydrology, geology and hydrogeology of a bioregion with explicit assessment of the potential direct, indirect and cumulative impacts of coal seam gas and coal mining development on water resources. This scientific information will be available for all interested parties – including Australian federal and state decision makers, industry and the community – when considering coal seam gas and coal mining developments.

This model shows the flow of information – including uncertainty – through a BA; the linkages between disciplines, data and numerical models; and 'decision points' which can be used to adaptively manage the BAs. The model is illustrated as a 4-meter-long diagram that:

- specifies a standard structure and language for products, with codes for 'chunks' of content. This ensures consistency when BAs are undertaken in many different bioregions, and also enables moving away from traditional reports (with linear narratives) to modern delivery on multiple channels – ensuring that the information is discoverable
- specifies a standard structure for on-demand reporting of provenance – ensuring transparency.

This model is useful not only because it documents the logic behind a BA; it also visually communicates the linkages within this interdisciplinary programme, ensuring that the science – and the reported information – is better integrated.

INTRODUCTION

An information model can be generally described as a representation of concepts, relationships, constraints and operations to specify the semantics of a domain of discourse [1]. This paper presents an information model that governs the scientific activities, inputs and products for bioregional assessments (BAs). First, background is provided on the Bioregional Assessment Programme and its methodology. The model is then described, including how it is used to ensure that products are discoverable and transparent. The paper concludes by discussing the way in which is it used by researchers, the client and the products' audience.

BACKGROUND

The Bioregional Assessment Programme

A bioregional assessment is a scientific analysis of the ecology, hydrology, geology and hydrogeology of a particular geographic area, with explicit assessment of the potential direct, indirect and cumulative impacts of coal seam gas and large coal mining development on water resources [1][2]. The Bioregional Assessment Programme undertakes these assessments for its primary audience, the Independent Expert Scientific Committee on Coal Seam Gas and Large Coal Mining Development (IESC), but also for a range of stakeholders including state government regulators, coal seam gas and large coal mine proponents, and interested community members. The outputs are a suite of 12 distinct scientific products for each of the 13 geographic areas currently being studied. All unencumbered datasets will also be published.

The programme team spans both disciplines and research agencies with four main collaborators: the Commonwealth Scientific and Industrial Research Organisation (CSIRO); Geoscience Australia; the Bureau of Meteorology; and the federal Australian Government Department of the Environment. Nearly 200 people are working together to deliver over 150 products over the course of three years. In addition to scientists who specialise in the relevant disciplines, the programme also includes a products team that undertakes the quality assurance / quality control procedures with respect to content, format and delivery. These editors, mapmakers and technologists specialise in integration in interdisciplinary projects and have broad domain knowledge which they use to design products to suit their purpose and audience.

The methodology for bioregional assessments

The BA methodology [1] articulates the scientific and intellectual basis for a consistent approach to all BAs. This methodology provides guidance to research scientists and managers preparing BAs within research agencies.

A BA comprises five components of activities: contextual information, model-data analysis, impact analysis, risk analysis and outcome synthesis. Information generated during the contextual information and model-data analysis components accumulates to provide information used in the impact analysis and risk analysis components. The final component delivers a synthesis that the IESC uses to support scientific advice on impacts and risk of coal seam gas and coal mining development on water resources. The components are not sequential in time; rather they are largely overlapping, and information passes between components of the BA via interdisciplinary interactions spanning the key disciplines of ecology, hydrology, geology and hydrogeology.

Specific workflows will vary between BAs in response to the availability of existing data, information and fit-for-purpose models in different bioregions. While a BA ideally is a quantitative analysis of impacts, deficiencies in data, information and models may preclude this. In this case, semi-quantitative and qualitative methods are to be substituted − supported by multiple lines of evidence − to provide the best and most current scientific advice possible. Measures of uncertainty are also required to be reported – ideally expressed quantitatively.

A key goal of the programme is to provide discoverable and transparent products. This is challenging because the BA methodology is complex and is applied differently in different bioregions, using a standard sequence of modeling and analysis steps. The first step to developing discoverable and transparent products is to ensure that the researchers and audience understand the way all the information fits together, and to specify a structure to report the information, uncertainty and provenance. While the BA methodology articulated this structure in words, the information flow was not clear and needed to be communicated differently for greater understanding.

THE MODEL

The model is illustrated as a 4-meter-long diagram [4] that shows the flow of information (including uncertainty) through a BA; the linkages between disciplines, data and numerical models; and 'decision points' which can be used to adaptively manage the assessments. A simplified version is shown in Figure 1, and an excerpt is shown in Figure 2.

The first step was listing by hand, on a big piece of paper, all numbered components and subcomponents along with their input, intermediate and output datasets. Activities, information, models and decisions were added when transcribing the hard copy into a Visio diagram, with all entities connected by solid black lines indicating the flow of general information and dotted lines indicating the flow of uncertainty information. Coloured lines indicate where information will be published in a suite of products associated with each component and subcomponent.

Intermediate 'decision points' were included at places where the BA could be adaptively managed throughout the course of the BA, as guided by either management or scientific decisions. Individual activities can be assigned required staff resources (FTEs) and costs, which can be summed and analysed for the entire BA, so that the model can be used as a project management tool to cost an entire BA or to compare costs of taking one path over another.

The resulting diagram was printed and laminated, and formed the basis of planning meetings and stakeholder engagement meetings.

The following sections indicate how this model is used to ensure that the information presented in products (such as reports, web content, maps, charts and data registers) is discoverable and transparent.

Discoverability

In many bioregions, the BA products will be the first time that this wide range of information has been brought together to one publically accessible place. Given the breadth of the data, disciplines and geographic locations reported, careful thought must be given to how the products are structured and presented in order to ensure discoverability.

The model specifies a standard structure and language for products, with codes for 'chunks' of content. This ensures consistency when BAs are undertaken in many different bioregions, and provides easy pre-structured ways to answer questions such as 'Were similar groundwater models used for bioregion A and B?' or 'Which BAs included a quantitative (as opposed to qualitative) cumulative risk analysis?'.

In addition to classifying information according to the model, the information will also be classified according to other axes:

- code of the product where it is reported (e.g. a number such as '1.2' in Figure 1)
- geographic location (bioregion and subregion, or points, vectors or polygons therein)
- ownership, access and licences of datasets
- disciplines (e.g. ecology or geology)
- delivery medium (e.g. time series, databases, web services, raster files).

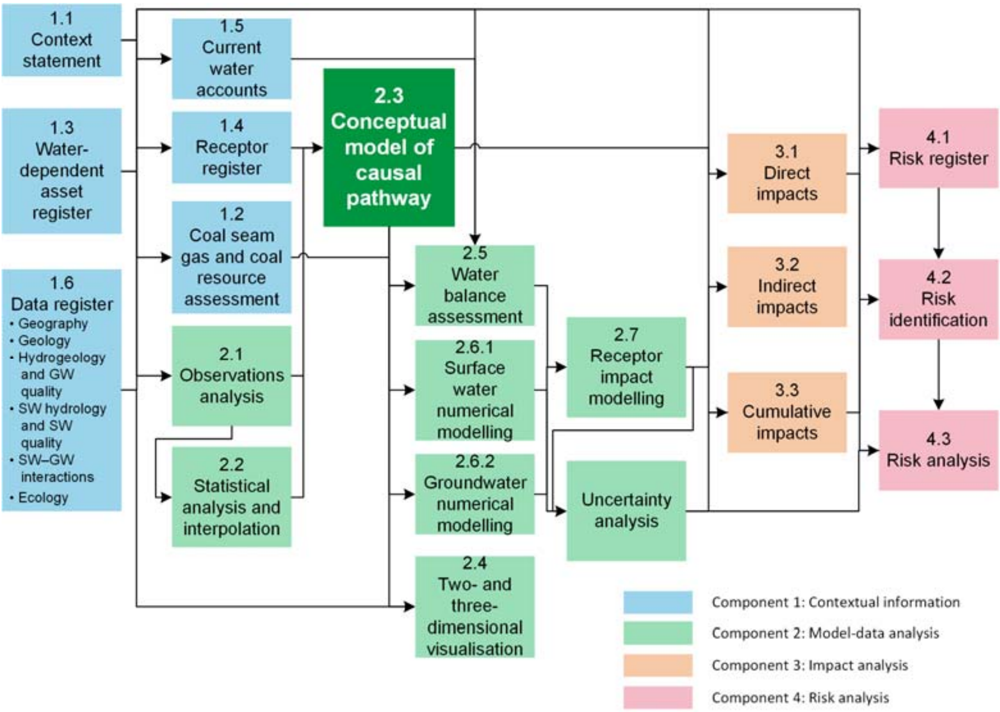

Figure 1. The simple model indicates the flow of information through a bioregional assessment. See [4] for the full model, and Figure 2 for an excerpt

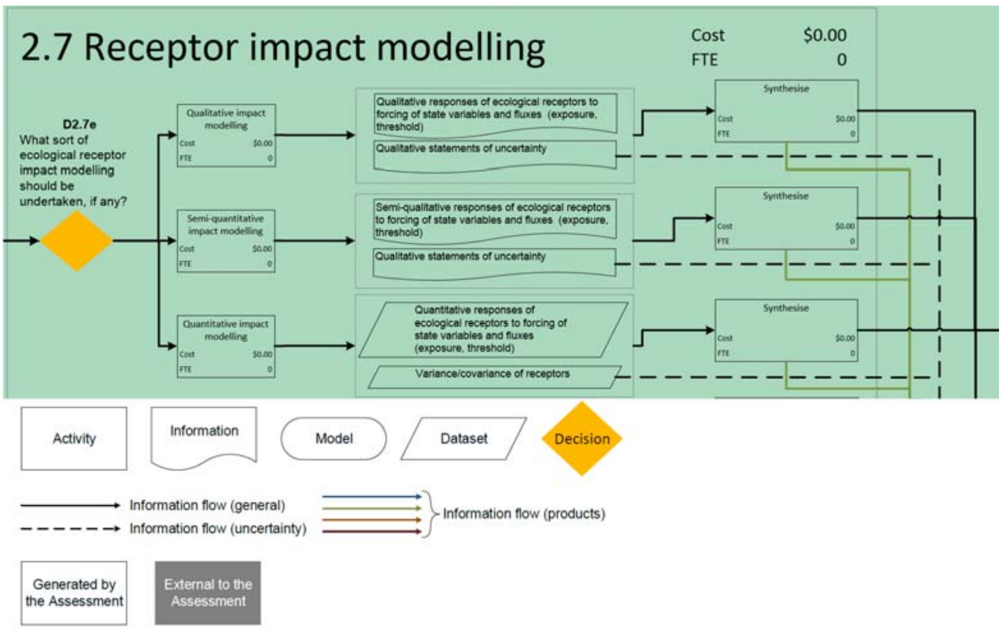

Figure 2 An excerpt from the full model [4]

Classifying information according to multiple axes enables moving away from traditional reports (with linear narratives) to modern delivery on multiple channels with a non-linear unspecified path through the information. Other questions can be answered such as ‘What ecological information is available for koalas across all BAs?’ or ‘What information is available at location X, where a coal seam gas development is proposed?’.

The information can also be broken up and reassembled into new reader-led products that reflect the audience’s interest. For example, the IESC provides information guidelines [5] that list the specific information that they need to deliver robust scientific advice. By formally mapping the information required (as listed in the guidelines) to the appropriate part of the model, the programme can automatically assemble relevant information from BAs in a form that the IESC requires, given the location of a proposed development.

Transparency

To be transparent, the programme needs to report information accompanied with both uncertainty and provenance (the lineage of data and processes). The provenance needs to be adequately captured at all stages of the development of products, including data processing and human decisions. This requires standardised reporting that relies on potentially complex methodologies for representation. At its most general, the programme will report provenance with respect to entities, activities and actors using the PROV-O standard [6][7][8]. The information model reported here will also be used as a more specific ontology to enable reporting of provenance in the same way that the results and uncertainty are provided.

This means that the programme can move beyond answering questions such as ‘How was this result reached?’ to questions such as ‘Was the BA in bioregion A conducted in a similar way to that in bioregion B?’ or ‘Was the BA undertaken as specified by the BA methodology?’. Both the simple (Figure 1) and the more complex full diagram [4] can be used to report results at several scales, thus meeting many of Sandve’s [8] ten simple rules for reproducible computation research, but particularly the rule related to hierarchies: ‘generate hierarchical analysis output, allowing layers of increasing detail to be inspected’.

In specifying a standard structure for information – as well as a map for how the information should be connected if a BA is performed correctly – the model is useful not only as a communication tool, but also for compliance, by offering a visual tool for comparison of different implementations of BAs. The pathway through the model results in a unique ‘fingerprint’ of a BA, and the similarity of two BAs can be assessed at a glance. Because methods go from simpler (at the top of the diagram) to more complex (at the bottom), the relative complexity of a BA can also be estimated (with more ‘diagonal’ pathways being more complex).

DISCUSSION AND CONCLUSION

In interdisciplinary projects, with complex linkages between disciplines, data and models, a diagram is often required to effectively communicate how the reported information is generated, linked and integrated (see, for example, Table 1.2 in [5] and Figure 1.3 in [11]). Researchers within individual disciplines require a roadmap to understand the broader context that their contribution fits into, encouraging better integration of the science and information. In the Bioregional Assessment Programme, the diagram of the information model has played a central role in project planning for the BAs, with activities forming the basis of project plans;

the subcomponents and codes dictating the design and outline of products; and the diagram itself framing discussions at meetings – thus improving integration of products and science.

The client and audience also benefit from the added understanding that a diagram provides. In this case, the client has singled out the diagram of the model as an output that is 'amazingly helpful' for resonating and connecting with people when communicating to stakeholders. This is initially surprising, given the complexity and size of the diagram, but the complexity may well be a key aspect, in line with Edward Tufte's advice: 'to clarify, add detail' [12]. With its simple underlying structure (Figure 1), and more complex details [4], the model allow readers to zoom between two different scales, thus enhancing communication.

Beyond contributing to effective integration and communication, the model plays a more fundamental role in providing a structure for the information, for not only the primary content, but also the associated uncertainty and provenance. The model translates easily into an ontology for provenance. Coupled with a system for recording provenance and an agreed provenance standard, the model enables on-demand provenance reporting, thus allowing the IESC and other stakeholders to judge the credibility of the information, ensuring transparency.

ACKNOWLEDGEMENTS

This work was funded by the Bioregional Assessment Programme, a scientific collaboration between the Australian Government Department of the Environment, Bureau of Meteorology, the Commonwealth Scientific and Industrial Research Organisation (CSIRO), and Geoscience Australia. The authors thank Nicholas Car and Peter Fitch for valuable discussions that contributed to this paper.

REFERENCES

[1] Lee TY (1999) Information modeling from design to implementation. National Institute of Standards and Technology. Viewed 31 March 2014, <www.mel.nist.gov/msidlibrary/doc/tina99im.pdf>.

[2] Department of the Environment (2013) Overview of the Bioregional Assessment Programme. Viewed 26 March 2014, <www.environment.gov.au/coal-seam-gas-mining/pubs/overview-bioregional-assessment-programme.pdf>.

[3] Barrett DJ, Couch CA, Metcalfe DJ, Lytton L, Adhikary DP and Schmidt RK (2013) Methodology for bioregional assessments of the impacts of coal seam gas and coal mining development on water resources. A report prepared for the Independent Expert Scientific Committee on Coal Seam Gas and Large Coal Mining Development through the Department of the Environment. Department of the Environment, Australia. Viewed 26 March 2014, <www.environment.gov.au/coal-seam-gas-mining/pubs/methodology-bioregional-assessments.pdf>.

[4] Barrett DJ, Schmidt RK and Gallant S (2014) The decision tree for the methodology for bioregional assessments. A product from the Technical Programme, part of the Bioregional Assessment Programme. To be published at <www.bioregionalassessments.gov.au> in April 2014.

[5] IESC (2013) Information guidelines for proposals relating to the development of coal seam gas and large coal mines where there is a significant impact on water resources. The Independent Expert Scientific Committee on Coal Seam Gas and Large Coal Mining.

Viewed 27 March 2014, <www.environment.gov.au/coal-seam-gas-mining/pubs/iesc-information-guidelines.pdf >.

[6] Lebo T, Sahoo S and McGuinness D (eds.) (2013) PROV-O: The PROV ontology. W3C Recommendation 30 April 2013. Viewed 31 March 2014, <www.w3.org/TR/prov-o/>.

[7] Car NJ, Francis W and Simons B (2014) Scientific process modelling using PROV-O and extensions to it for resource access and service description. 5th International Provenance and Annotation Workshop, 9–13 June 2014, Cologne, Germany.

[8] Francis W, Car NJ, Schmidt RK and Gallant S (2014) Leveraging quality assurance and quality control processes to deliver provenance as a first-order scientific output in large-scale environmental assessments. International Environmental Modelling and Software Society (iEMSs) 7th International Congress on Environmental Modelling and Softwarem, San Diego, California, USA, DP Ames, N Quinn (eds).

[9] Sandve GK, Nekrutenko A, Taylor J and Hovig E (2013) Ten simple rules for reproducible computational research. PLoS Comput Biol 9(10): e1003285. doi:10.1371/journal.pcbi.1003285

[10] CSIRO (2012) Assessment of the ecological and economic benefits of environmental water in the Murray–Darling Basin. CSIRO Water for a Healthy Country National Research Flagship, Australia. Viewed 31 March 2014, <www.mdba.gov.au/kid/files/2017-Assessment_Ecological_Economic_Benefits.pdf>.

[11] CSIRO (2012) Proposed project methods. A report to the Australian Government from the CSIRO Flinders and Gilbert Agricultural Resource Assessment, part of the North Queensland Irrigated Agriculture Strategy. CSIRO Water for a Healthy Country and Sustainable Agriculture flagships, Australia. Viewed 31 March 2014, <publications.csiro.au/rpr/pub?pid=csiro:EP122673>.

[12] Tufte ER (1990) Envisioning information. Graphics Press, Cheshire CT USA.