I have read with deep interest the Deakin University PhD thesis of Mr Andrew Simmons submitted to me for examination. This thesis encapsulates a rare cross-disciplinary research outcome and makes several worthwhile contributions to data processing pipelines, sports data analysis and geo-spatial data analysis. Although the broad field and contribution can be seen to be advancing software engineering with a focus on data processing, the publications arising from the research effort are in SIGSPATIAL (a premier geo-spatial conference), OZCHI (the leading human computer interaction conference in Australia), and MathSport (the premier conference in Australia related to Mathematics and Computers in Sports).

I am firmly of the opinion that the thesis demonstrate that Mr Simmons has an expert level knowledge about the field of spacio-temporal data processing in general and shows a confident application of this for sports data. This thesis shows the capacity to generate grounded new knowledge, and to engage in a scholarly discourse with research communities in multiple discipline areas.

The thesis demonstrates strong written communication skills, the work is well organised, the references are grounded and complete (to the context), and the work is presented in a writing style that is quick to parse. I recommend the award of the PhD degree.

A few minor suggestions are described below based on the reading of the thesis in its entirety. These suggestions offer a critique with minor suggestions and can be incorporated and checked by the supervisors; I see no reason to further examine the thesis as the suggestions do not impact on the depth or quality of the fundamental contributions.

The thesis motivates the core problem and gap that there is a need to make complex spatio-temporal analysis techniques more accessible to sports researchers and practitioners. Although the core research goals were addressed within the context of analysing GPS data feeds in supporting decision making, the exploration of this problem yielded a number of sub-problems; spanning the need to have data processing pipelines that are aware of the underlying domain context, a requirement of better data representation, an analysis method that allows comparative analysis (of spatio-temporal data), and visualisations and presentation of analysis to coaches in a form that helps improve decision making. Furthermore, the research effort also identified the need to be aware of both provenance while preserving privacy in this context – such side-tracks are natural in an open research exploration and these were also tackled.

The work proceeds carefully by setting the context of prior technological and analysis methods in Chapter 2. This chapter is quite long and can be trimmed with a more expansive Appendix. Although the content of the chapter itself is helpful to a reader with minimal background in this area, the knowledge captured is a summative exposition and not critical to the structure of the overall thesis.

Chapter 3 presents a summary of AFL jargon to contextualise and help a reader not fully immersed in the game better appreciate the narrative. The more helpful part of this chapter and the core contribution is in the meta-model that captures a data structure (or schema) that identifies the informational grammar needed in the context of AFL game data

analysis. The information theoretical perspective is an interesting lens in terms of motivating how performance analysis works within sports. Although this is theoretically interesting, the underlying knowledge is not directly novel. This aspect can be improved by providing a stronger motivation of how this lens can be better utilised either in practice or how this can lay the foundation of further works.

Chapter 4 is a deep and thought out unit of work. Surprisingly, the content here has not been put for publication as this will make a great stand-alone contribution by itself, and I strongly encourage that the candidate considers this recommendation. This chapter motivates data provenance in the context of sports, evaluates current tools & techniques against a framework and offers a notation and approach to help sports analysts consider these issues better. The evaluation undertaken does not directly show if the ideas are directly helpful to a sports analyst. However, the fundamental concepts are rigorously presented, and this chapter should aim to present a short outline of an evaluation protocol as part of further work to be done.

Chapter 5 presents findings on what appears to be a sub-problem that was discovered as a natural process of doing the research. That is, it does not come out as something that was known ahead of the PhD work starting. This is a natural effect and it is good to see such components included within the work. This chapter presents the need for a better de-identification protocol in this domain. Given that the protocol developed has been used and evaluated in practice there is confidence that it is helpful. Although not a gold standard, the primary efficacy seems to be there and hence it can be deemed to be a worthwhile contribution upon which further refinement is possible. The chapter can be improved by outlining briefly, how the protocol can be further refined and improved or areas that need specific future attention.

Chapter 6 is a novel method that shows the value in treating spatio-temporal reference frames as geographical objects. The novelty and value of this work is validated by the SIGSPATIAL work arising from this unit of research.

Chapter 7 brings the overall thesis together with a baseline evaluation to demonstrate the value of the computation pipeline in its totality. There are a number of new ideas, concepts and models used and hence it was a difficult read. This chapter can be improved by captions and legends in the diagrams.

The overall thesis presents a set of new contributions that are novel and all of which together more than make the case for a PhD award. The evaluations undertaken can all be improved and at a minimum effort should be made within this thesis on how these evaluations can be undertaken.

I would like to congratulate Mr Andrew Simmons and his supervisory panel on contributing a rare cross-disciplinary body of work. The work demonstrates Mr Simmons has the skills and knowledge appropriate to be awarded a PhD.