Author's response to reviews

Title: Evaluation of PROforma as a language for implementing medical guidelines in a practical context.

Authors:

David R Sutton (dsutton@brookes.ac.uk)
Paul Taylor (p.taylor@chime.ucl.ac.uk)
Kenneth Earle (Ken.Earle@stgeorges.nhs.uk)

Version: 3 Date: 1 February 2006

Author's response to reviews: see over
Revisions

There are only minor changes in the wording of the abstract

The Background is introduced using material previously in the Methods section, with some additions, to explain better the purpose of the project and the context of the work. This material now includes answers to the questions raised by Silvana Quaglini about the healthcare organisation and the clinical aims of the study.

The material on clinical guidelines and their representation, including an account of Proforma is then presented, with minor revisions. The section on knowledge engineering and principles of knowledge presentation has been moved to the end of the Background and substantially revised to provide a better introduction to the new Methods section.

The Methods section has been completely rewritten to address comments from all reviewers, most particularly David Kaufman and J. F. Arocha. It now includes a clear statement of the objectives and how they are operationalised. New material is included giving details of decisions taken in the course of implementing the system. The rationale and approach taken for the evaluation is explained.

Following a suggestion from David Kaufman the Results section is organised according to the three phases of the project and the four characteristics of knowledge representation languages. The points made are broadly those from the original paper but the presentation is altered and additional material included, responding to comments from the reviewers.

The Discussion section now contains material relating to the most interesting of the observations now in the Results section. Much of this material is new and included in response to suggestions from the reviewers, who particularly requested detail and examples illustrating these points, esp J.F. Arocha. These have been included for example in the section dealing with support for information hiding on p21 (requested by Samsun Tu and Silvana Quaglini) and the use of UML Activity Diagrams (requested by Samsun Tu and Silvana Quaglini). An additional figure is included here showing an activity diagram.

Minor changes have been made to the conclusions. The helpful comments of the reviewers, which we feel have greatly improved the paper, are acknowledged.
Specific responses to comments are listed below. (Note that the numbering of observations does not correspond to that used by the reviewers)

**Points made by reviewer 1 (Samson Tu).**

1. *p. 13 What kinds of logical distinctions should be encoded in PROforma?*
   
The authors state that “One of the aims of the PROforma language is to describe the logic of processes in a way that allows them to be examined and validated by domain experts. Thus one possible answer...is to say that knowledge needs to be represented in PROforma if it needs to be validated by domain experts.”

   It’s not clear what’s the argument being made here. Is the authoring saying that, because PROforma allows encoded knowledge to be validated by domain experts while other software components do not, it’s a good idea to encode the knowledge that requires validation? It would be highly desirable for the authors to give specific examples. It’s not clear what formalized knowledge doesn’t require some kind of validation.

   The argument that we were trying to advance was that a PROforma process description describes knowledge is intended to be read by particular domain experts. Therefore when deciding what to include one should ask oneself whether those experts are in a position to validate the knowledge encoded. We deal with this topic in the Results and in the Discussion using the example of the knowledge required to assess cardiovascular risk, which is done using regression equations derived from epidemiological data.

2. *The authors found the essential logic of the medical process involved to be relatively simple. This reviewer, having created a guideline-based decision-support system for the management of hypertension, finds that hard to believe. If the system gives advice about how a patient’s medication should be modified, the necessary knowledge is quite complex. It is very frustrating that the authors didn’t give more details about the kinds of advice and knowledge available in their system.*

   We stated that “the choices that were made in these processes were relatively simple”. What we meant by the word “simple” is that these choices could be represented using arithmetic comparisons and propositional logic. We did not mean that they were simple in the sense of being easy to encode or validate. However the process description we used for this study is possibly also simpler in a more general sense than the guidelines created for the comparative study written by Peleg et al. This is because some of the decisions that are encoded in these guidelines would, in the system that we have created, be left to the discretion of prescribing GPs. For instance if a patient is to be prescribed
more than one anti-hypertensive agent then the choice of the second agent is left to the GP. The reason for leaving these decisions to the GP is not that they cannot be represented in PROforma – the PROforma process description produced for the Peleg study demonstrates that they can. However since the process description is intended for use in a real study, the requirements for testing and validation of its logic are more stringent and it was felt it would be easier to gain the confidence of the various stakeholders in the system if the more detailed decisions were left to GPs. This point is explained in the revised paper.

3. **Support for abstraction and information hiding - please give examples.**

The point could be illustrated by considering the top-level plan shown in Figure 3. It would be possible to modify this plan and yet argue that the essential logic of the process had not changed. For instance the three actions of the form “Return in N weeks” could be replaced by a single action containing a parameter that varied so as to express the value of N. Similarly the three decisions in the plan could be combined into a single decision with more complex logic. Ideally the Process Description would contain some sort of guarantee that such changes could not “break” any component of the system that accessed the process description. In order to do this it would be necessary to prevent “inessential” details of the process being revealed to other components of the system so as to provide a guarantee that these details could not affect their behaviour. At present there is no way of incorporating this kind of information hiding into a process description.

The significance of this weakness becomes apparent when a complete system is created because the knowledge encoded in a process description often needs to be refined and one needs to know whether consequent changes will affect other components of the system.

4. “…the Tallis implementation of PROforma provides no support for record structures or for collections other than ordered lists” Is it a problem with Tallis implementation or is it a more general problem in PROforma? In (Peleg et al., 2003), it’s said that “PROforma, GUIDE, and Asbru do not constrain the possible classes of complex objects.”

Our comments refer specifically to the Tallis implementation of PROforma. The Arezzo implementation does contain some facilities for representing complex objects but, as the Peleg paper observes, does not place any constraints on the classes of these objects.
5. “The PROforma language proved adequate...may be because the reasoning requirements of the study were relatively simple” Again, it’s very frustrating not being told what the reasoning requirements were.

We hope that the comments we make in response to point 2 go some way to meeting the reviewers concern. We will revise our paper so as to clarify the reasoning requirements and to outline where they are simpler, and where they are more complex, than those of guidelines produced in other publications such as Peleg et al.

6. “UML has more complete graphical representation. This was one of the main reasons why UML activity diagrams were used during the knowledge acquisition phase of the project.” This statement is surprising. UML activity diagrams models processes similar to those of PROforma. The authors really need to describe in detail why UML activity diagrams are more convenient. For example, it is not clear at all that UML activity diagrams can support graphical representation of arguments and candidates that the authors say PROforma lacks. My understanding is that arguments are Boolean expressions for or against a particular candidate (which is a pointer to the next task). Please argue for your assertions and give concrete examples!

UML is defined as a graphical language, therefore everything that it represents has a standard graphical representation. The PROforma specification sets out the syntax for a text representation of the language but does not describe any standards for its graphical representation. There are some de facto standards for representing PROforma process descriptions (decisions represented as circles, actions as squares etc.). However many key PROforma concepts have no standard graphical representation at all. Not only is there no standard way of representing candidates and arguments (the reviewer is right to point out that these cannot be represented in UML either) but also there is no standard for describing task preconditions or parameters in a diagram. This point is dealt with in the revised paper.

7. “In particular we considered that strengths and weaknesses of the PROforma language are revealed by: ....” The authors didn’t really argue for the assertions made in this paragraph.

In the revised paper we describe how the observations made about PROforma arise from the needs described in the paragraph to which the reviewer refers. For instance the lack of graphical standards for PROforma emerges as an important issue when modelling knowledge acquired from experts because, at this stage of a project, it is necessary to be able to represent knowledge as diagrams on paper that can be read without the use of software tools. Similarly the need to hide information and create
abstractions emerges when process definitions are incorporated into larger systems because they provide means of “decoupling” the process definition from other components of the system.

8. Very little is said about the knowledge needed to provide the advice for the users.

The revised paper provides more detail on the knowledge acquisition process and examples of the representation.

9. No example is given on why activity diagrams are better than PROforma processes in terms of usability by domain experts. Are the difficulties intrinsic difficulty with PROforma or is it a problem of tool support? Suppose you model the PROforma task ontology in Protégé, would PROforma be more usable? Much of the complexity in encoding computer-interpretable guidelines is in writing the decision criteria. That’s why in the EON system, for relatively simple criteria, we use a set of object templates that clinicians can just fill in the fields.

The top-level plan set out in Figures 3 and 5 should help. When this plan is set out as a PROforma diagram it can be seen that there are circumstances under which a patient would, for example, be asked to return in two weeks time. However there is no way, in Fig 3, of indicating graphically what these circumstances are (although this information is encoded in the text of the PROforma process description and can be made visible through the use of appropriate software tools). In an activity diagram, Fig 5, the conditions under which each action is taken can be expressed graphically as guards on transitions between activity states.

The difficulties mainly lie with deficiencies in the standards for representing process descriptions as diagrams and, possibly, to the limited facilities for attaching conditions to transitions between PROforma tasks, rather than to the tasks themselves.

10. Little is said about how the needs for distributed system and user interface impact the use of PROforma.

We hope that we have gone some way to addressing this concern in our answers to points 7 and 1.

11. It seems that the way to simulate guards in PROforma is to “reify” the condition by creating a decision task where the guard condition is a “conclusive” argument for the subsequent task.
One would also need to ensure that the subsequent task has appropriate preconditions. This can be rather complex if there is more than one transition into this task.

**Points made by reviewer 2 (Silvana Quaglini).**

12. *The background section must contain also some information about the healthcare organisation of the country where the study has been carried. Which are the responsibilities of the pharmacists and of the GPs? Are each patient addressed to a single pharmacist or does he address more than one?*

We have included additional material dealing with this at the start of the background section.

13. *Some essential data about the study are lacking: How many pharmacists accepted to be part of the study? How many patients? Which is the perceived utility if the system? It is not clear if the web interface is a utility provided by PROFORMA or it has been built by the authors from scratch. To this concern, the reader should be greatly facilitated by some pictures showing the user interface.*

We have included additional material dealing with this at the start of the background section and in the account of the implementation.

14. *The authors say that pharmacists get the patients data. There are no privacy problems? Which subset of the EPR is shared between GP and Pharmacists?*

The Pharmacists will have access only to data acquired for use in the study. This data is held on a database that contains no data that could be used to identify patients. Pharmacists do not have access to any data held by the GPs. This material is included in the Background.

15. *The paper lacks from examples: when they say “The study revealed a number of weaknesses in PROforma’s logical adequacy ...”, the authors should provide at least one example, and the same for the other weaknesses encountered.*

One example is given in our response to point 3. We will examine the other weaknesses that we have described and make sure that these are illustrated with examples.

16. *About the heuristic power, the authors say that probably the implemented guideline was too simple: they should try with a more complex one!!*
The strengths (and to certain extent the weaknesses) of this paper result from the fact that PROforma is being used in real study with clinical as well as informatic objectives. This means that we cannot simply choose a more complex guideline and there would be strong ethical arguments against making the existing PROforma process description more complex than it needs to be if the only reason for doing this was to provide a more exacting test of the adequacy of the language.

17. About the comparison of PROforma with Java and Pascal, I don’t understand this comparison: PROforma is not a programming language. This point should be clarified, i.e. which is the role of the PROforma language within the guideline representation.

We mean that the expression language is not Turing Complete: that is to say that there are functions (in the mathematical sense of a mapping from one set of symbols to another) that could be represented by a Turing Machine but which could not be represented in the PROforma expression language. We express the point by making a comparison with Java or Pascal because the readership of the paper will be more familiar with these languages. We do not intend to draw a comparison with other facilities that Java or Pascal possess but which are irrelevant to the question of Turing Completeness (for example the facilities provided for the construction of graphical user interfaces).

We have attempted to make our meaning clearer in the paper by giving a more precise explanation of what we mean by a “function”.

18. About UML, again I suggest to make an example (with pictures) showing a situation in which UML diagram performs better than PROforma graph.

We have provided, in Figures 3 and 5 a Proforma graph and a UML Activity Diagram. A detailed examination of the advantages, for us, of the UML is included in the revised Discussion.

Points made by reviewer 3 (David Kaufman).

19. Many of the most serious concerns are raised in the General section. The lack of detail in the methodology and the informal reporting of results are perhaps the most problematic. In addition, there should be a clear statement of the objectives and how they are operationalized in the context of this study. The general objective is to analyze the adequacy of PROforma in a specific clinical context. However, the majority of claims make only tenuous reference to specific contexts.
We have included new material in the Methods and Results section to address these criticisms. We have also re-organised the Background to provide a better rationale for our work. We have included more detail about the specific context in the observations made in the Results and Discussion.

We would point out that the degree of rigour and formality that we can bring to bear on this paper is limited by considerations such as: the unavoidable subjectivity of criteria such as notational convenience; the fact, where informatics is concerned, any results that emerge from the exercise we report are interesting only to the extent that they could not be anticipated at the outset of the study; and the practical and clinical concerns associated with a real study. These considerations all limit the extent to which we could devise a methodology designed to answer questions about the PROforma language set out in advance of the study.

20. **PROforma was evaluated in the context of 3 phases: knowledge acquisition, analysis design and system implementation. The foci of the analysis are logical adequacy, heuristic power, notational convenience and explanation support. A well organized presentation would provide an explicit breakdown of results according to the 3 phases and on the basis of the 4 criteria.**

We have adopted this suggestion.

21. **It would be extremely useful, if the authors could connect the clinical aims of the study (e.g., satisfactory outcomes) with the stated informatics aims (expressiveness of the Performa language) in some fashion. Although an extended treatment of clinical outcomes is likely beyond the scope of this paper. In general, some combination of quantitative (Tables), qualitative and anecdotal evidence could make for a compelling story. Though the selection of methods is of course entirely up to the authors and would suffice as long as they are presented with greater clarity and adhere to proper standards.**

It is rather difficult to connect the clinical aims of the study to the informatics aims because a system constructed using PROforma could, in principle, be replaced by another system that behaved in exactly the same way but which did not use PROforma (the same could be said of any other knowledge representation language). It is therefore unlikely that one would be able to point to any particular clinical outcome and claim that it occurred because of the use of a particular knowledge representation language.

This does not mean that the choice of language is irrelevant (if it did then we would still be constructing systems in machine code). However the
effects of the choice of language are likely to affect the ease with which the system can be developed rather than its eventual behaviour.

22. The methodological description is deficient in important respects. What are the units of data? Presumably, they differ according to the different phases. How is their adequacy judged? How were the criteria for determining (for example) logical adequacy arrived at? Is there a consensus process for making such determinations?

The evaluation of knowledge representation languages is, of necessity, a somewhat subjective and qualitative process. We are not aware of any method that could reasonably be described as a consensus process or that would meet the reviewers’ desire for units of data.

23. I am particularly interested in the implementation phases and specifically would like to know about events or instance in which the system contributed to a sound decision or an unsound one or proved to be problematic in some important respect. For example, what are the factors that contributed to a false urgent referral? Can they be traced back to some limitations in the modeling language? I appreciate that some of these factors may be a result of human error or some other aspect of the system, but some diagnostic reporting could be very informative. I also realize that evaluating a knowledge representation language may not neatly fit into an experimental package the way a controlled clinical study would. But it is clear that the paper could use much greater systematicity and rigor in articulating the method and in the reporting of the results.

As we mention in our response to point 21, a system constructed using PROforma could, in principle, be replaced by another system that behaved in exactly the same way but which did not use PROforma. It would therefore be virtually impossible to point to any particular event arising during the use of the system and attribute this event to the choice of PROforma as an implementation language.

24. There are some contradictions in the paper. For example, the authors conclude that “there was no difficulty in heuristic power, but in the results discusses problems with computational completeness and the inability to express certain functions. The contradictions may be more apparent than real, but should be addressed nonetheless.

The reason why the lack of computational completeness did not create difficulties in heuristic power is that, although there are some functions that cannot be represented in PROforma, the process description reported in this paper did not require the use of any of these functions. We will attempt to rewrite the paper so that this point is made with greater clarity.
Points made by reviewer 4 (J. F. Arocha).

25. I would say that the major problem I have with the paper is that it is highly abstract, which makes it difficult to determine how the different PROforma "functionalities" examined were evaluated. Although the title of the paper hints at an "evaluation in a practical context," the description remains highly abstract and does not tie to actual application of the system. I think that this can be remedied by providing more concrete examples, for instance, of the strengths and weaknesses of the system. The authors describe some examples, but given that the study was done within the specific application of hypertensive patient monitoring, more illustrative examples would help the reader actually understand better the consequences of such strengths and weaknesses or other characteristics examined.

We hope that our response to points made by other reviewers (for instance points 15 and 3) will address this problem.

26. Special attention should be given in the Methods section to provide a more explicit description of how PROforma was evaluated. What methods were used? How were the methods applied?

We have provided a completely re-organised and substantially rewritten Methods and Results section to meet this concern including new material explaining our Methods and how they were applied.