Modelling Clinical Goals: a Corpus of Examples and a Tentative Ontology

John FOX, Alyssa ALABASSI*, Elizabeth BLACK, Chris HURT, Tony ROSE
Advanced Computation Laboratory, Cancer Research UK

Abstract. Knowledge of clinical goals and the means to achieve them are either not represented in most current guideline representation systems or are encoded procedurally (e.g. as clinical algorithms, condition-action rules). There would be a number of major benefits if guideline enactment systems could reason explicitly about clinical objectives (e.g. whether a goal has been successfully achieved or not, whether it is consistent with prevailing conditions, or how the system should adapt to circumstances where a recommended action has failed to achieve the intended result). Our own guideline specification language, PRO*forma*, includes a simple goal construct to address this need, but the interpretation is unsatisfactory in current enactment engines, and goals have yet to be included in the language semantics. This paper discusses some of the challenges involved in developing an explicit, declarative formalism for goals. As part of this, we report on a study we have undertaken which has identified over 200 goals in the routine management of breast cancer, and outline a tentative formal structure for this corpus.[1]

1 Introduction: Modelling Clinical Guidelines

The medical informatics community is developing a variety of methods to support the computerisation of clinical guidelines. Overviews of the leading current approaches, with associated publications and links, can be found at www.openclinical.org. The longest established approach to providing computer-based guidelines is the Arden Syntax for medical logic systems, a rule-based format that has been widely taken up by industry (particularly in the USA). The limitations of rule-based formalisms are, however, being increasingly appreciated; Tu and Musen [25] have identified five general capabilities that computerised guidelines and guideline representation methods should be able to support: interpreting data; making decisions; sequencing actions; refining actions (i.e. breaking tasks up into sub-components) and setting goals (e.g. specific patient states to be achieved). The difficulty of formalising such high level processes has stimulated another paradigm based on "task networks" [18] in which clinical tasks are viewed as object-like structures that represent steps in clinical processes. Task networks offer greater capabilities than simple rules, including the ability to model clinical workflow and time-oriented schedules.

Peleg et al [18] compared six prominent task-network approaches to modelling clinical guidelines which address some or all of the capabilities identified by Tu and Musen [25]: Asbru [21]; EON [25, 26]; GLIF [16, 17]; GUIDE [20]; PRODIGY [19, 14] and PRO*forma* [9, 7, 23]. Peleg et al's review [18] compares the approaches on eight dimensions. Four dimensions are concerned with representation and interpretation of data:

*also at Guy's Hospital Breast Unit, London

[1] We would like to thank Samson Tu and referees for comments that have led to improvements in this paper.

- Patient information modelling
- Representation of medical concepts
- Data abstraction and interpretation
- Expression language for reasoning about data;

Four are concerned with modelling the structure of a guideline:

- Representation of guideline actions
- Organisation of guideline plans
- Models of decision-making
- Representation of goals/intentions.

Peleg et al conclude that the different task network approaches have similar capabilities on a number of these dimensions. Indeed, they identify a set of common features that could be the basis of a consensus model for representing and executing guidelines, including common tasks like actions, plans and decisions, and control features like sequential and parallel task enactment, and task cycling.

However, there is at least one important area where the capabilities of all these systems remain weak. This is in the representation and management of the *goals* that underpin clinical tasks. Clinical goal setting is identified by Tu and Musen [25] as a significant requirement for guideline systems, but there has been little work in the area (an exception being the work of Shahar et al [21], which is discussed later). In particular, there has been little investigation of the requirements for representing and processing goals, and no studies to our knowledge of the relative strengths of alternative approaches. This paper begins an analysis of these problems in the domain of breast cancer.

The organisation of the paper is as follows: first we describe the domain in which we are attempting to develop a scheme for representing and processing clinical goals, the diagnosis and treatment of breast cancer. Next, we discuss some of the benefits that would be gained from an improved understanding and formalisation of goals and describe work carried out to date on goal processing in our own PRO*forma* guideline technology (in which some significant shortcomings are identified). We then present an empirical study of clinical goals in breast cancer care, and introduce a provisional classification model for the corpus of examples obtained from this study. We conclude with some comparisons with other approaches to this problem.

2 The Study Domain: Breast Cancer

The CREDO project at Cancer Research UK (www.acl.icnet.uk/credo.html) is working towards a multi-centre trial of decision support and workflow technologies in cancer care. The aim is to determine whether such technologies can significantly improve the effectiveness of service delivery in complex, multidisciplinary care pathways. The trial will address a number of questions, including whether computerised support for care planning, decision-making and workflow management can improve consistency, quality and safety of cancer care, and whether such services can be offered in an integrated form that is acceptable to and valued by clinicians.[2]

[2]The project is being developed in collaboration with three specialist centres at Guy's Hospital, London, Addenbrookes Hospital, Cambridge (UK) and the DOD funded Centre of Excellence in Breast Cancer Care which is being led by the Carol Franc Buck Breast Cancer Center at the University of California in San Francisco.

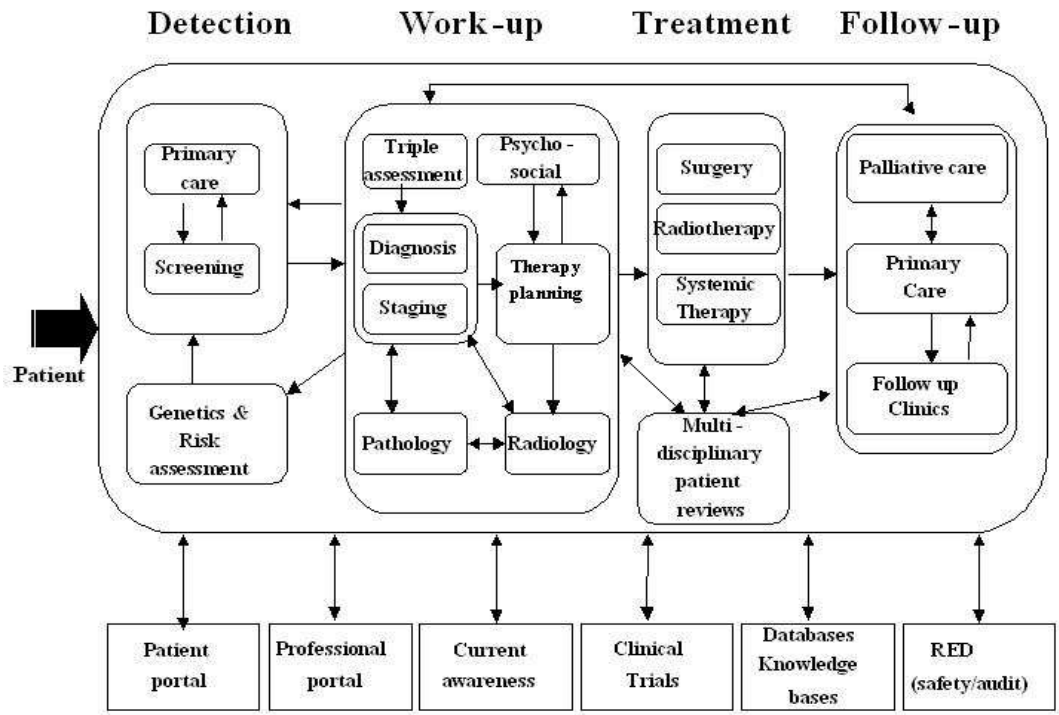

Figure 1: CREDO core service model for breast cancer, ©Cancer Research UK.

CREDO is intended to evaluate the use of decision support and workflow services at many points in the "patient journey", from the *initial encounter* where cancer may be first suspected (including the family physician; screening service) and *work-up*, where diagnosis and disease staging are established and treatment is planned, followed by delivery of *treatment*, and long term *follow-up*. These processes are carried out over months or years, and involve many professionals in different disciplines (general practitioners, surgeons, radiologists, oncologists, pathologists etc.) in different places at different times. Consequently, there are many different services to be considered, and important challenges regarding effective communication and coordination of the many different tasks to be carried out.

An early objective of the project has been to define its scope and develop service requirements with our clinical colleagues as well as building prototypes to facilitate these activities. Figure 1 shows the "core service model" we have developed for the project. Each rounded rectangle represents a CREDO service (and will contain many component services). A service is defined as a collection of guidelines, clinical pathways and other machine-interpretable resources that support a particular clinical goal. The CREDO system can be thought of as a collection of communicating agents, which individually and collectively help to support timely, effective and safe patient care. The CREDO system is also intended to provide a range of interfaces to external services that assist patients, clinicians and researchers in achieving their objectives (represented by the rectangles at the bottom of Figure 1).

The breast cancer domain includes many instances of the requirements identified by Tu and Musen [25]: interpreting data, making decisions, sequencing actions, refining plans into workflows of actions, and goal setting. In addition, we wish to support various kinds of communication between the many professionals involved in this complex multidisciplinary process, and, for reasons explained later, in actively managing clinical goals.

2.1 Example of a Credo Service: Triple Assessment

To get a clearer picture of the kinds of services that CREDO is intended to support, we will briefly look at one such service in detail: "Triple Assessment" of a patient with suspected cancer. This typically takes place at a "one-stop" or "same day" clinic where the aim is to achieve a provisional diagnosis and assessment of the grade and spread of the cancer if any. Triple assessment includes:

1. Clinical Examination;
2. Various forms of imaging: (mammography, ultrasound, MRI, CT);
3. Histo-pathology: Fine needle aspiration or Core needle biopsy.

The definition of each CREDO service has been specified by an experienced breast surgeon, Alyssa Alabassi. Component services have been organised into 3 categories: clinical services, patient services and communication services.

2.1.1 Clinical Services

- Support for eligibility decisions for investigations and further follow up.
- Support for eligibility decisions for genetic risk assessment.
- Management of follow up or discharge back to primary care.
- Recruitment into clinical trials.
- Support for psychological/psycho-social risk assessment.
- Tracking results and investigations.

2.1.2 Patient Services

- Provision of relevant information about cancer.
- Providing personalised schedules.
- Identifying support groups through "Patient matching".

2.1.3 Communication services

- Notifying patient's physician of results, management, discharge plan.
- Notifying patient of results and management plan.
- Requesting further investigations where needed.
- Inviting patients for follow up and investigations.
- Recalling patients for follow-up assessment.
- Communicating findings to multidisciplinary team.
- Referring for specialist assessments (psycho-social, genetic).

In all some 222 specific services have so far been identified for routine patient management in the breast cancer domain.

3 Why Should we Formalise CREDO Service Goals?

Obviously, a clinical procedure is generally carried out *for a reason*, to bring about or prevent some desired or undesired state, for example. If the procedure fails to achieve the desired outcome then the clinical goal will not be achieved and, worse, patient safety may be compromised. With serious conditions, care should therefore be planned with possible failure of

clinical interventions in mind (e.g. a drug may not achieve its intended objective) or a patient may unexpectedly deteriorate for reasons having nothing to do with the original disease (e.g. a dangerous reaction to a medication). Consequently, when we design and enact clinical processes such as CREDO services we should allow for the possibility of urgent changes to the care plan, or "plan repair" [7]. If we are to do this, we need to be explicit about the reason for each service to enable the system to be capable of recovering when goals are not achieved.

Unfortunately, most published medical guidelines do not make the goals underlying recommended actions explicit: authors simply assume that the reasons for recommendations will be obvious to the trained audience at whom the guideline is aimed. Moreover, authors of guidelines frequently assume that the trained professional will know when a guideline's recommendations are inappropriate, by realising that in some specific clinical circumstances the consequences of a recommended intervention may in fact be inconsistent with the well-intentioned purposes of the author (who, the rationale goes, could not possibly foresee all the contingencies that might arise in every clinical setting where the guideline is used).[3]

In contrast, Shahar [21] argues that a guideline should "explicitly capture the design rationale (process and outcome intentions) of the author, while leaving flexibility at application time to the attending physician and their own preferred methods". Shahar identifies three main dimensions to be considered when trying to capture a clinical intention (Shahar and his colleagues typically use the term "intention" rather than goal, but for present purposes the terms are taken to be synonymous):

- Whether the intention is to *achieve*, *maintain* or *avoid* some situation;
- Whether the intention refers to a clinical *state* or *action*;
- Whether the intention holds during enactment of the clinical process (*intermediate*) or after it has been completed (*overall*).

An important challenge for guideline representation research is to formalise goals and intentions rigorously. If we can formalise the concepts in such a way that a guideline engine can reason about them there are a number of potential benefits. For example, complex sequences of tasks such as those involved in cancer treatment are frequently interrupted or have to accommodate unexpected situations which may make the normal routine impractical (e.g. lack of staff, equipment faults, overloaded services etc. etc). Guideline enactment systems should be able to respond flexibly in the face of adverse circumstances. Furthermore, if a guideline engine can reason explicitly about clinical intentions, and whether or not its proposals have succeeded or failed in achieving them, then it may be able to adapt its plans to ensure the safety and efficacy of the care process [10, 15].

A further benefit of explicitly representing clinical goals concerns improved accountability of decision support systems. Since the introduction of "expert systems", it has been widely accepted that explanations for any advice given are desirable. However, explanations need to be grounded with respect to the goal that the advice is intended to achieve - traditional rule-based explanations are often unsuccessful precisely because they are not grounded in this way. Another aspect of communication is between members of a care team. If guideline systems are used to support complex, multidisciplinary processes like breast cancer care, the software agents that implement guideline services will need to be able to explain, discuss and even negotiate their tasks with human users and other agents. Such capabilities are not con-

[3] An exception to this is in the protocol documents of clinical trials, which often lay out chapter and verse of the criteria for particular clinical procedures, and for departing from normal practice. Although CREDO is itself a trial, the trial is of the value of decision support in routine clinical care—we are trying to support existing routines, however, not trying to impose a novel protocol.

ceivable without an explicit representation of the goals that the agent is pursuing. An explicit representation of goals may also be needed for critiquing a clinician's decisions or actions (e.g. showing an action to be inconsistent with one or more clinical intentions [2]) and form an important focus for the quality assessment of clinical practice [1].

4 Formal Models of Goals

Some logicians, computer scientists and others have recently attempted to understand formally the concept of an agent's goal. We cannot review the literature in this area here, but will limit ourselves to mentioning a few topics that have been helpful in understanding some of the issues in modelling clinical intentions.

AI research starts with the premise that to build intelligent systems, such as planning systems and robots which must cope with complex and unpredictable environments, it is necessary to separate the behavioural aspects of intelligence (what the robot should do) from the cognitive aspects (what the robot should believe and how it should decide to act with respect to its beliefs and goals). Research in the seventies put goal-directed behaviour at the centre of intelligent behaviour and many ad hoc schemes were implemented.

Wellman and Doyle [28] were early critics of these ad hoc approaches, proposing as an alternative the decision-theoretic concept of "utility", which views goals simply as states which have *positive utility* for an agent and goal-oriented decision making as the selection of actions that *maximise expected utility*. While accepting the criticism of ad hoc methods, it is questionable whether a single number can capture the semantics of goals like "ensure that blood sugar has been within safe limits for several hours before initiating therapy" and there is growing interest in developing a richer axiomatisation of goal semantics.

Winikoff et al [29] stress an important distinction between procedural and declarative aspects of goals and in particular the need to decouple the successful completion of a plan from the successful achievement of the goal that the plan was designed to achieve. They suggest that there is a set of declarative properties that the goals of a rational agent must have, i.e. goals must be:

- Persistent (goals exist so long as their success conditions are not satisfied)
- Unachieved (a goal is dropped precisely when its success conditions are true)
- Possible (a goal is dropped with failure when it becomes impossible to achieve)
- Consistent (an agent should not adopt goals that conflict)
- Known (goals must be explicit if an agent is to reason about them)

The questions for formalists are: what (formal) properties do goals have, and what normative constraints should be placed on them and the behaviour they control? These are unsettled questions throughout AI, and it is unclear whether a complete or practical scheme can be developed purely on the basis of formal analysis and mathematical intuition. The medical informatics community is in a unique position to explore such questions by investigating how clinicians, whose expertise is highly goal-directed, manage patient clinical goals in one of the most complex and uncertain domains that confront human beings.

5 Current Approaches to Goal Management in CREDO

Decision-making, workflow management, communication and other CREDO services are being implemented using the PRO*forma* process modelling language[4] (see [7] for a detailed review). PRO*forma* is a first-order logic language that has been extended to represent *tasks*, where task classes (decisions, plans, actions and enquiries) are viewed as processes that are intended to achieve some defined goal. The language and associated tools have been used to implement a range of clinical applications.[5] In most respects, the language has been stable since its first definition in 1996; the formal definition of the syntax and operational semantics of the stable subset has recently been published [23]. However, the syntax and semantics of the goal component is an important area where we have not yet managed to achieve a stable language definition.

Two authoring environments are currently available for building PRO*forma* applications: Arezzo ®[6] (a commercial system) and Tallis (an experimental toolset which is available for research use). These tools are being used to explore different aspects of goal processing experimentally.

Arezzo has a uniform goal-processing scheme in that all PRO*forma* tasks have a specific slot for defining the goal of the task. This scheme is a generalisation of the context schema for clinical decision-making proposed by Huang et al [12]. In this schema *Context = (Task, Focus)* the Task represented the goal of a particular decision (diagnose, treat, investigate, etc.), and the Focus was the medical focus of the decision (joint pain, inflammation, etc.). In Arezzo the goal of a task is separated from the rest of the task definition, but is modelled in an analogous fashion, using the syntax *Goal = Verb:Object* (e.g. "treatment_goal = manage: hypertension"). The Arezzo engine manages these goals according to a continuous control cycle in which a task that is in progress will be automatically terminated if its goal state becomes true (or omitted if the goal state is already true when the task is considered for enactment).

Winikoff et al's [29] distinction between declarative and procedural semantics of goals, and proposal to decouple goal failure from plan failure, are to some extent achieved in the Arezzo engine. The *known*, *persistent* and *unachieved* properties of goals are captured directly (declaratively) through the goal attribute of all PRO*forma* tasks and the Arezzo engine implements goal persistence in a manner that fits with Winikoff et al's proposals. The other goal related properties, *possible* and *consistent*, can only be implemented indirectly (procedurally). For example a PRO*forma* task can respond to significant clinical situations or events by means of its trigger conditions and preconditions, and from then on the goal state determines its persistence. In addition, the termination and abort conditions can be used to bring active plans to an end in response to situation changes (e.g. a goal is no longer achievable or relevant).

While these are useful behaviours, the syntax of Arezzo goals is too limited to describe the range of clinical goals we expect to have to manage in CREDO, and the semantics do not deal with the problem of how a PRO*forma* agent should recover from failure.

The Tallis toolset is intended to support the complete lifecycle of design, implementation,

[4] Given the theme of the workshop we would also describe PRO*forma* as a guideline modelling language but since it is not exclusively for medical applications we use this more neutral description.

[5] CAPSULE: Decision support for prescribing in primary care [27]; ERA: Compliance with cancer referral guidelines [3, 8]; RAGs: Risk assessment in cancer genetics [5, 6]; IRIDIUM: integration of decision support with Clinical Evidence published by British Medical Journal [22]; Retrogram: a system to advise on management of HIV+ patients [24]; LISA, for helping clinicians comply with the dosage rules of the MRC 2002 ALL trial [4, 13].

[6] Available from *Infer*Med (www.infermed.com)

deployment and maintenance of applications, (a small collection of demonstrators can be found at www.openclinical.org/kpc/). Goal slots in Tallis representations are currently limited to text descriptions, but since the toolset is designed for assembling PRO*forma* applications from standard components, Tallis is an attractive platform for exploring the plan repair aspect of goal processing. Currently, Tallis only supports manual assembly of applications from components held in a component store but the aim is to implement an automated method for run-time plan repair in response to goal failure.

Every PRO*forma* task has a set of declarative properties that can be used for this purpose viz: goals, preconditions and post-conditions. Suppose a medication has been administered as part of a treatment plan, with the goal of bringing a patient's blood pressure within normal limits, but this method has not been successful. An algorithm that will achieve a basic form of repair is as follows.

1. Search the repository of PRO*forma* tasks for a task component whose goal term matches the current goal (reducing blood pressure). This is a candidate for repairing the plan.
2. Check the preconditions of the candidate component, if any, to ensure there are no features of the current clinical situation that make the component unacceptable in this context.
3. Check the post-conditions of the candidate component to see that it has no consequences that are inconsistent with the preconditions of any tasks that are already scheduled later in the currently active plan.

The correct semantics of goals in PRO*forma* (or any other formalism so far as we are aware) cannot be finalised yet, in part because we do not have enough knowledge of the kinds of goals that are likely to be encountered in the clinic, and in part because we do not have enough experience of plan repair. Without a clear solution to this, we are taking an empirical approach to the problems. The first step has been to carry out an analysis of the range of goals that occur in the breast cancer domain. These are described in the next section.

6 Modelling Clinical Goals in Breast Cancer Care

In order to understand the requirements for modelling and managing goals in this domain we have carried out an extensive review and analysis of the whole of the CREDO core service definition. This was initially carried out in terms of the PRO*forma* task types which could be used to implement each service, and the clinical goals which are implicit in each service description. For example, consider the following triple assessment tasks:

- Eligibility decisions for genetic risk assessment;
- Invite/recall patients for follow up and investigations;
- Follow up or discharge back to primary care.

Each of these services can be mapped to a particular type of task in the PRO*forma* task ontology, respectively:

- *Decision*, of type "eligibility";
- *Enquiries*, of type "invitation";
- *Plan*, of type "workflow", which is a sequence of scheduled *actions* of type "discharge"

and *enquiries* of type "follow-up".

PRO*forma* task networks that will manage such services are straightforward to construct in most cases (most of the complexity appears to be at the platform or middleware level). However, the definition of the intentions that lie behind these tasks is more troublesome because, unlike PRO*forma* tasks and service components, we do not have an established ontology of types with which to analyse clinical goals. In the absence of a classification we therefore decided to try to create an empirical ontology of goal types as a tool for understanding those that arise in the breast cancer domain. A *classification system* is traditionally the first step in carrying out a scientific analysis of any natural phenomena (in this case the phenomena are clinical intentions). Refining the classes into a *hierarchy* would further illuminate the domain by recognising similarities and differences between goal types. The first attempt resulted in an *a priori* taxonomy, defined before the CREDO service model had been developed (see Figure 2).

Once the CREDO service definition had been agreed we used it to explore the goals underlying the 222 services in CREDO service definition, and revise the goal taxonomy in light of observations. The overall process was as follows.

1. For each CREDO service we wrote a short English sentence that we felt captured the clinical goals that the service was intended to bring about.
2. Each goal instance was assigned to one of the ontological categories in Figure 2. If the goal did not appear to fit into any of the "leaf" classes we introduced a new class into the hierarchy.
3. At completion of this step duplicate goals were removed, equivalents merged and, where initial assignment appeared mistaken, the goal was assigned to another class.
4. Where goals naturally grouped together within an ontological class we considered introducing a new subclass (recursively).
5. Classes in the a priori ontology that had no entries were deleted.

The resulting revised ontology is shown in Figure 3.

7 Discussion

The revised ontology in Figure 3 has the merit of being grounded in observations of a large clinical domain, as against a structure constructed on purely theoretical grounds, but we would make a number of comments and reservations.

1. The specific ontological classifications that we have assigned should not be thought of as final or even unique. It appears that many if not all statements of clinical "intentions" can be paraphrased into statements that one might reasonably put in a different class. At this point, therefore, our scheme has primarily heuristic value.
2. Goals are interrelated in ways that are not shown in a simple class hierarchy. For example:
    a. Goal-subgoal relationships (e.g. in order to achieve successful management of a patient we would normally need to successfully achieve a correct diagnosis and then successfully carry out treatment).
    b. Goal priority relationships (e.g. curing the patient is more important than ensuring the patient is comfortable)
    c. Goal precedence relationships (e.g. a goal to collect data may be needed before a

Root class

- □ Knowledge goals (or *internal* or *epistemic* goals)
  - ◇ Acquire information/knowledge about setting
    - ● *Example: get clinical history, measure clinical parameters*
  - ◇ Decide between alternative hypotheses about world
    - ○ Detection
      - ● *Example: determine the presence or absence of an abnormality or monitor for the occurrence of an abnormal event*
    - ○ Classification
      - ● *Example: determine which of N possible conditions is present or which stage a disease has reached*
    - ○ Stratification
      - ● *Example: establish level of risk for a clinical condition*
    - ○ Predict unknown state from current known states
      - ▷ Diagnosis
        - ● *Example: predict the aetiology of a clinical problem*
      - ▷ Prognosis
        - ● *Example: Predict future prognosis of patient from current state*
- □ Action goals (or *external* or *practical* goals)
  - ◇ Achieve
    - ○ Eradicate
      - ● *Example: eradicate an infectious organism*
    - ○ Create
      - ● *Example: create a sterile site*
  - ◇ Control
    - ○ Prevent
      - ● *Example: prevent side-effect of treatment*
    - ○ Limit goals
      - ● *Example: maintain physiological parameter within limits*
  - ◇ Communicate
    - ○ Enquire
      - ● *Example: ask for an appointment*
    - ○ Inform
      - ● *Example: tell colleague results of test*

Figure 2: Goal ontology version 1.

Goals
- □ Knowledge goals
  - ◇ Acquire information/knowledge about specific setting [15 instances]
    - ○ Acquire information about setting
    - ○ Acquire comparison data in setting
  - ◇ Decide between alternative hypotheses about the world [52 instances]
    - ○ Detect
    - ○ Classify
      - ▷ Staging
      - ▷ Eligibility (determine whether criteria are satisfied)
        - • {Investigation, Referral, Therapy, Research trial}
      - ▷ Assess level of some parameter
        - • {Urgency, Risk, Need, Quality}
    - ○ Predict
      - ▷ Diagnosis
      - ▷ Prognosis
- □ Action goals
  - ◇ Achieve some state of world [65 instances]
    - ○ Limit changes to current state
    - ○ Bring about required future state
      - ▷ Empower staff
      - ▷ Prevent unwanted future state
      - ▷ Ensure compliance with plan
    - ○ Decide between alternative interventions
      - ▷ Decide whether to carry out intervention or not
      - ▷ Decide type of intervention
      - ▷ Decide scheduling of intervention
  - ◇ Enact tasks [90 instances]
    - ○ Arrange service
    - ○ Investigate
    - ○ Communicate
      - ▷ Educate
      - ▷ Inform

Figure 3: Goal ontology version 2.

goal to take a decision can be implemented)

How does the revised taxonomy compare with Shahar's scheme for classifying intentions [21]? The three main distinctions of his scheme are considerably broader than the classes in the a priori and empirical ontologies in boxes 1 and 2, but are generally compatible. The table below shows the 222 CREDO goals reclassified according to our understanding of Shahar's scheme. There are some obvious features of this summary.

State: 67	Achieve: 24	Intermediate: 47
Action: 155	Maintain: 24	Overall: 175
	Avoid: 3	

First there are far more action-oriented goals (e.g. "Ensure proper discharge back into primary care") than state-oriented goals (e.g. "Establish eligibility for psychosocial counseling") in the corpus. This may be a feature of a cancer domain, where care is relatively procedural rather than judgmental (in fact 90 of the action-oriented goals we classified were simple "to-do" items like "arrange access to community-based services") and this may be less typical of domains in which protocols are not so heavily used. On the other hand it may be that there is some ambiguity about these assignments. As noted above it is frequently possible to paraphrase goal statements in multiple ways (e.g. the procedural statement "arrange access to community-based services" can be paraphrased as "ensure that patient has proper access to community-based services").

Second, we expected that among the practical goals many would be concerned with controlling (achieving, maintaining or avoiding) a specific clinical state. Not only was this not the case (only 48 goal descriptions make a desired clinical state explicit) but only 3 examples were concerned with avoiding undesirable situations. Again we suspect that this has to do with protocol-based care which tends to emphasise routine tasks rather than the clinical rationale for those tasks (a feature that makes clinicians understandably concerned that protocol-based care may encourage carers to lose sight of the reasons for their actions thereby leading to inflexible, "algorithmic" behaviour). This is underlined by the small number of examples of "avoid" goals, since it is surely the case that avoiding adverse events and unintended negative effects of treatment is central to all of medicine, including cancer care where avoiding potentially life-threatening consequences of surgery, radiotherapy, cytotoxic drugs etc is fundamental. It is quite possible, of course, that our failure to elucidate control goals has something to do with our methodology but either way it suggests that we need to improve our ability to elicit and formalise the medical rationale for clinical tasks.

Third, there are far more goals classified as applying "overall" than at "intermediate" points. For example once a patient's diagnosis or cancer staging has been established it would be the same throughout the patient's subsequent treatment, while goals to administer systemic therapy and minimise toxicity will only hold during treatment. Again, this asymmetry may be idiosyncratic: the management of hypertension or diabetes, for example, involves many repeating clinical assessments. Furthermore, Samson Tu identified a lot of intermediate goals in cancer protocols in the form of avoiding excessive toxicity (*personal communication*), which are notably absent from our corpus.

Tu has also raised the question of whether we can talk about goals without specifying the agent who has set the goal. He notes that Shahar is concerned with the intentions of a guideline's authors, while Stefanelli's group has discussed goals in terms of the organisation that is providing care. In contrast our analysis is not *prescriptive* in the sense of defining clinical practice norms (which guideline committees and care providers are concerned with)

but *descriptive* in that we are trying to classify the goals that are implicit in the tasks of the CREDO service model. Our aim is to understand the types of clinical goals that the CREDO system will need to support as a step towards defining an operational semantics for a guideline engine that is to manage breast cancer care. We see our classification as a useful step towards a practical understanding of clinical goals and their management but more formal analysis is clearly needed before we can develop a normative model.

During the preparation of this study we have had a number of discussions with Aziz Boxwala who has also been carrying out an empirical study of goal types with his colleagues. Based on a study of four guidelines (Management of Asthma; Diagnosis of Asthma; Diagnosis and Management of Sinusitis; and Treatment of Acute Myocardial Infarction), they report that they have developed an ontology consisting of 53 classes of goals (c.f. our 35) with the highest classes being Assessment, Communication, Management, and Decision goals. Their scheme is broadly compatible with ours at the upper levels in our taxonomy. Hashmi et al [11] propose a formalisation of goals as 5-tuples $\langle$ C, I, T, TC, P $\rangle$ where:

- **C** is the initial state or context in which the goal applies.
- **I** is the intention verb that specifies whether the target function is to be achieved, avoided, etc Over 27 distinct verbs have been identified.
- **T** is the description of the target function and applies to state of anatomical structures, the diseases or disorders, physiological functions. Over eight major categories that can be further expanded were identified.
- **TC** is a set of temporal constraints.
- **P** is the priority of the goal. A decision-support system can use this to select and rank from among competing goals.

We have used Hashmi et al's scheme [11] to reclassify our corpus (so far as we can, given the available details). The table below gives an analysis of the number of goals in our corpus that refer to each of the elements of the Hashmi et al model.

Context: 222	Verb: 222	Target: 47	Temporal: 47	Priority: 0

An informal goal statement which appears to fit with this scheme would be "if a patient presents with symptoms of possible breast cancer then it is obligatory that the patient is referred to see a specialist oncologist within two weeks", where Context = patient presenting with symptoms, Target = breast cancer, Verb = refer, Temporal = within two weeks and Priority = obligatory.

We have assumed that every goal in our corpus is associated with a Context, as Hashmi proposes. Part of this context is defined by the CREDO service the goal is associated with but in the PRO*forma* task implementation of each service more context detail will be provided (e.g. preconditions, triggering events). Our first PRO*forma* goal model (see discussion of Arezzo earlier) also requires a verb, again consistent with the Hashmi proposal. In this provisional comparison the goals in the breast cancer corpus that specifically refer to or entail temporal information are the same as those identified in the Shahar comparison (48 examples).

Two elements of the Hashmi proposal are more problematic. First the Target element of the model is described as a "desirable target state that is to be achieved within temporal constraints". If we limit this to patho-physiological systems then the only explicit examples in the corpus are the same 48 "achieve-maintain" goals we identified in the comparison with the Shahar model. If we broaden this out (e.g. allowing a target to be other kinds of systems,

such as external clinical services as in "refer to specialist oncologist within two weeks") then far more examples can be found in our corpus. We need to see a more complete description of the Hashmi et al scheme before we can take this question further.

A second notable feature of our corpus is that we found no examples of Priority statements, concerning which goals should take priority over others in conflict resolution. This seems surprising. It may be that such information is common sense to clinicians and not worth mentioning, or perhaps in a field like breast cancer most clinical goals are regarded as mandatory. Whatever the reason, the Hashmi analysis indicates that this requires further investigation.

8 Conclusion

There are many reasons to make the goals of clinical procedures explicit, both as part of conventional documentation and formally as a basis for achieving greater flexibility and adaptability in point-of-care decision support and workflow systems. However, current technical proposals are at best incomplete and theoretical proposals have not been validated empirically. Using breast cancer as an example domain we have developed a corpus of examples of clinical goal statements and designed a tentative summary in the form of a simple ontology. Comparisons with other proposals in this area suggest that there may be some promising convergence in this area, but further work is required before a final scheme can be settled.

References

[1] A. Advani, M. Goldstein, and M. Musen. A framework for evidence-adaptive quality assessment that unifies guideline based and performance-indicator approaches. In *Proceedings of American Medical Informatics Association Symposium (AMIA 2002)*, 2002.

[2] A. Advani, K. Lo, and Y. Shahar. Intention-based critiquing of guideline-oriented medical care. In *Proceedings of American Medical Informatics Association Symposium (AMIA 1998)*, 1998.

[3] J. Bury, M. Humber, and J. Fox. Integrating decision support with electronic referrals. In *Proceedings of the Tenth World Congress on Health and Medical Informatics (Medinfo2001)*, 2001. see also www.openclinical.org/aisp_era.html.

[4] J. Bury, C. Hurt, C. Bateman, J. Atwal, K. Riddy, J. Fox, and V. Saha. Lisa: A clinical information and decision support system for collaborative care in childhood acute lymphoblastoc leukaemia. In *Proceedings of the AMIA Annual Sympsomium*, 2002.

[5] J. Emery, R. Walton, A. Coulson, D. Glasspool, S. Ziebland, and J. Fox. Computer support for recording and interpreting family histories of breast and ovarian cancer in primary care (rags): Qualitive evaluation with simulated patients. *British Medical Journal*, 319:32–36, 1999.

[6] J. Emery, R. Walton, M. Murphy, J. Austoker, P. Yudkin, C. Chapman, A. Coulson, D. Glasspool, and J. Fox. Computer support for interpreting family histories of breast and ovarian cancer in primary care: Comparative study with simulated cases. *British Medical Journal*, 321:28–32, 2000.

[7] J. Fox and S. Das. *Safe and Sound: Artificial Intelligence in Hazardous Applications*. MIT Press, 2000.

[8] J. Fox, H. Humber, J. Bury, A. Rahmanzadeh, and R. Thomson. Publets: Clinical judgement on the web? In *Proceedings of American Medical Informatics Association Symposium (AMIA 2001)*, 2001.

[9] J. Fox, N. Johns, A. Rahmanzadeh, and R. Thomson. Pro*forma*: A method and language for specifying clinical guidelines and protocols. In J. Brender, J. P. Christensen, J-R. Scherrer, and P. McNair, editors, *Medical Informatics Europe '96*, pages 516–520. IOS Press, 1996.

[10] P. Hammond, A. Harris, S. Das, and J. Wyatt. Safety and decision support in oncolgy. *Methods of Information in Medicine*, 33(4):371–381, 1994.

[11] N. Hashmi, A. Boxwala, D. Zaccagnini, and J. Fox. Formal represesntation of medical goals for medical guidelines. Personal communication, 2003.

[12] J. Huang, J. Fox, C. Gordon, and A. Jackson-Smale. Symbolic decision support in medical care. *Artificial Intelligence in Medicine*, 5:415–430, 1993.

[13] C. Hurt, J. Fox, J. Bury, and V. Saha. Computerised advice on drug dosage decisions in childhood leukaemia: A method and a safety strategy. In *Proceedings of the 9th Conference on Artificial Intelligence in Medecine, Europe (AIME 2003)*, 2003. see also www.openclinical.org/aisp_lisa.html.

[14] P. Johnson, S. Tu, N. Booth, B. Sugden, and I. Purves. Using scenarios in chronic disease management guidelines for primary care. In *Proceedings of American Medical Informatics Association Symposium (AMIA 2000)*, 2000. see also www.openclinical.org/gmm_prodigy.html.

[15] S. Modgil. Linking rules to termiologies and applications in medical planning. In *Proceedings of the 9th Conference on Artificial Intelligence in Medecine, Europe (AIME 2003)*, 2003.

[16] L. Ohno-Machado, J. Gennari, S. Murphy, N. Jain, S. Tu, D. Oliver, and E. Pattison-Gordon. The guideline interchange format: A model for representing guidelines. *Journal of the American Medical Informatics Association*, 5(4):357–372, 1998. see also www.openclinical.org/gmm_glif.html.

[17] M. Peleg, A. Boxwala, O. Ogunyemi, P. Zeng, S. Tu, R. Lacson, E. Begnstam, and N. Ash. Glif3: The evolution of a guideline representation format. In *Proceedings of American Medical Informatics Association Symposium (AMIA 2000)*, 2000. see also www.openclinical.org/gmm_glif.html.

[18] M. Peleg, S. Tu, J. Bury, P. Ciccarese, J. Fox, R. Greenes, R. Hall, P. Johnson, N. Jones, A. Kumar, S. Miksch, S. Quaglini, A. Seyfang, E. Shortliffe, and M. Stefanelli. Comparing computer interpretable guideline models: A case-study approach. *Journal of the American Medical Informatics Association*, 10(1):52–68, 2003.

[19] I. Purves, B. Sugden, N. Booth, and M. Sowerby. The prodigy project—the iterative development of the release one model. In *Proceedings of American Medical Informatics Association Symposium (AMIA 1999)*, 1999. see also www.openclinical.org/gmm_prodigy.html.

[20] S. Quaglini, M. Steffanelli, A. Cavallini, G. Micieli, C. Fassino, and C. Mossa. Guideline-based careflow systems. In *Proceedings of the 6th Conference on Artificial Intelligence in Medecine, Europe (AIME 2000)*, 2003. see also www.openclinical.org/gmm_guide.html.

[21] Y. Shahar, S. Miksch, and P. Johnson. The asgaard project: A task-specific framework for the application and critiquing of time-oriented clinical guidelines. *Artificial Intelligence in Medicine*, 14:29–51, 1998. see also www.openclinical.org/gmm_asbru.html.

[22] R. Steele and J. Fox. Enhancing conventional web content with intelligent knowledge processing. In *Proceedings of the 9th Conference on Artificial Intelligence in Medecine, Europe (AIME 2003)*, 2003.

[23] D. Sutton and J. Fox. The syntax and semantics of the pro*forma* guideline modeling language. *Journal of the American Medical Informatics Association*, 10(5):433–443, 2003.

[24] C. Tural, L. Ruiz, C. Holtzer, J. Schapiro, P. Viciana, J. Gonzalez, P. Domingo, C. Boucher, C. Rey-Joly, and B. Clotet. Clinical utility of hiv-1 genotyping and expert advice: The havana trial. *AIDS*, 16:209–218, 2002. see also www.openclinical.org/aisp_retrogram.html.

[25] S. Tu and M. Musen. A flexible approach to guideline modeling. In *Proceedings of American Medical Informatics Association Symposium (AMIA 1999)*, 1999. see also www.openclinical.org/gmm_eon.html.

[26] S. Tu and M. Musen. From guideline modeling to guideline execution: Defining guideline-based decision-support services. In *Proceedings of American Medical Informatics Association Symposium (AMIA 2000)*, 2000. see also www.openclinical.org/gmm_eon.html.

[27] R. Walton, C. Gierl, P. Yudkin, H. Mistry, M. Vessey, and J. Fox. Evaluation of computer support for prescribing (capsule) using simulated cases. *British Medical Journal*, 315:791–795, 1997.

[28] M. Wellman and J. Doyle. Preferential semantics for goals. In *Proceedings of the AAAI 9th National Conference on Artificial Intelligence*, 1991.

[29] M. Winkoff, L. Padgham, J. Harland, and J. Thangarajah. Declarative and procedural goals in intelligent agent systems. In *Proceedings of the 8th International Conference on Principles of Knowledge Representation and Reasoning (KR2002)*, 2002.