Contribution Structures

ORLENA GOTEL
oczg@doc.ic.ac.uk

ANTHONY FINKELSTEIN
acwf@doc.ic.ac.uk

Imperial College of Science, Technology and Medicine
Department of Computing, 180 Queen's Gate, London SW7 2BZ

Abstract

The invisibility of the individuals and groups that gave rise to requirements artifacts has been identified as a primary reason for the persistence of requirements traceability problems. This paper presents an approach, based on modelling the dynamic contribution structures underlying requirements artifacts, which addresses this issue. We show how these structures can be defined, using information about the agents who have contributed to artifact production, in conjunction with details of the numerous traceability relations that hold within and between artifacts themselves. We describe a scheme, derived from work in sociolinguistics, which can be used to indicate the capacities in which agents contribute. We then show how this information can be used to infer details about the social roles and commitments of agents with respect to their various contributions and to each other. We further propose a categorisation for artifact-based traceability relations and illustrate how they impinge on the identification and definition of these structures. Finally, we outline how this approach can be implemented and supported by tools, explain the means by which requirements change can be accommodated in the corresponding contribution structures, and demonstrate the potential it provides for "personnel-based" requirements traceability.

Key words: Contribution structures; pre-requirements traceability; requirements engineering; requirements traceability.

Classification: 3.1D.

1. Introduction

Requirements traceability (RT) has been defined as: "*the ability to describe and follow the life of a requirement in both a forwards and backwards direction (i.e., from its origins, through its development and specification, to its subsequent deployment and use, and through all periods of on-going refinement and iteration in any of these phases)*" [Gotel & Finkelstein, 1994a]. RT is fundamental for the management of change and evolving requirements. With the introduction and enhancement of tools that provide RT, such as ARTS [Flynn & Dorfman, 1990], DOORS [QSS, 1994], RDD-100 [Alford, 1993], RTM [MST, 1993], and Teamwork/RqT [Cadre, 1992], the mechanics are now in place for establishing lifecycle-wide RT. Despite the many advances, RT remains cited as a key problem area confronting industry.

Findings from recent work, which investigated the actual problems practitioners experience

when they claim to have RT problems, indicated that the majority of these are *informational* in character [Gotel, 1992; Gotel & Finkelstein, 1993]. It found that problems mainly occur when the above tools are not used to control information about requirements that practitioners want to trace. In particular, inadequate *pre-requirements traceability*, caused by the paucity and unreliability of information maintained about requirements production, was uncovered as a likely reason for longer-term RT problems. These informational issues are beginning to be addressed by the development of numerous RT meta-models, which delineate the information to record and link types to establish, in order to meet the RT needs of practitioners. The reader is referred to [Harrington & Rondeau, 1993; Laubengayer & Spearman, 1994; Pohl *et al.*, 1994; Pyle *et al.*, 1993; Ramesh & Edwards, 1993] for examples of such work.

However, significant findings from our analysis of the problem were: (a) the lack of agreement regarding the quantity and type of information that practitioners wanted to trace about requirements; and (b) the extreme importance that practitioners attached to personal contact and informal communication. The latter finding was not just a consequence of the first, to cope with those inevitable situations in which required information is absent, but was also found to be essential to account for the situated character of information needs. This practice enables any information which is available to be consolidated, supplemented, or questioned. It reflects the fact that people are often the final authority about requirements and, as such, are frequently able to prevent potential RT problems. Nevertheless, the ability to locate appropriate individuals and groups was reported to be problematic in practice. We suggest that this is because contemporary RT-related work, in its strive to supplant the need for human contact with evermore extensive and traceable project histories, rationales, decision records, and the like, does not prepare appropriate foundations to actively facilitate this most basic of working practices.

In [Gotel & Finkelstein, 1994a], the inability to locate, and so access, the human sources of actual requirements, requirements-related information, and requirements-related work, was concluded to be the crux of the multifaceted RT problem. Here, we recommended making details about the social setting that gave rise to the artifacts produced in requirements engineering (RE) explicit, and so traceable. In [Gotel & Finkelstein, 1994b], we proposed an approach to do this, based on modelling the *contribution structures* underlying requirements artifacts. In this paper, we provide more details of this approach. In Section 2, we describe the problem we are addressing, and provide an outline of the approach for orientation. Fuller details of the approach are given in Sections 3 through to 5. In Section 6, we describe how the approach has been made operational, illustrate how contribution structures extend conventional forms of artifact-based RT with the traceability of associated personnel, and explain how this helps ensure that the results of RE remain modifiable and maintainable. In Section 7, we critique the approach, and make reference to our research agenda.

2. Social infrastructure

In this section, we explain why there is a need to capture relevant information about agent participation that can be used to model the *social infrastructure* underlying RE. By "social infrastructure", we refer to the overall system of agents in the process, along with the various relationships they are involved in. We describe the deficiencies with prevailing practice, which make informed traces of agent participation untenable, summarise the basic requirements which arise from these, and outline the approach and its assumptions.

2.1. Scope & rationale

We restrict our concern to the issues of pre-requirements traceability, and so limit our scope to the traceability of information relating to the tangible artifacts produced in RE. Traceability needs to be maintained between such artifacts to prevent what eventually ends up a requirement being "black-boxed" in a formal requirements document. It provides the ability for such documented requirements to be re-examined and re-worked, from their source(s), and through their chain(s) of production. In this way, requirements are able to emerge throughout a project's life, in a more informed and controlled manner than is possible with *post-requirements traceability* alone (motivation for these 2 basic types of RT can be found in [Gotel & Finkelstein, 1994a]). To be more specific, we are concerned with the pre-requirements information that illuminates the social infrastructure underlying artifact production.

We limit our scope because our empirical studies found that practitioners predominantly claim to have experienced RT problems when, being unable to retrieve particular information about requirements artifacts they want from a project repository, they have further been unable to identify those agents who would be in a position to supply it. This inability was found to be caused by the way in which details of agent participation are currently recorded and maintained; no doubt a reflection of the absence of recommended guidelines to achieve this in the recognised RE standards (see [Dorfman & Thayer, 1990] for a representative collection of standards). So, although there have been advances in the techniques and tools used to collect, structure, and retrieve as much information as possible about RE activities, there has been little real focus on the RE participants.

2.2. Problems

In practice, we found that information about the RE participants, where not absent, was commonly inadequately described and maintained. Typical records of participation usually consisted of a list of names in an "author/owner" field of a document. Those documents which had been changed were generally characterised by the addition of further names, appended notes, or by official change request forms. This practice was found to compound RT problems, especially as the size and longevity of a project increased, as such records soon become unstructured, unwieldy, and inaccessible for analysis purposes. We found that it was not unusual for the end products of RE to have lost details about who originally generated a requirement and who was involved in all phases of its refinement. This meant that important questions were often unanswerable, like: "Who is responsible for this piece of information?"; "To whom should I refer for more information?"; "Within the remit of which group do decisions about this piece of information lie?"; and "Who was responsible for copying this information into this document?" These shortcomings lead to questions that are answered by, and defects that are addressed by, agents who are not necessarily best placed to do so. A repercussion which effects quality is that agent commitment to developing artifacts, as well as to each other, becomes fragmented and lost over time. This suggests that details about the RE participants is crucial pre-requirements information to retrieve, and that a dedicated approach is needed to collect, structure, and handle this information.

Simply appending an "author" label to a document results in relatively coarse and static notions of ownership. Moreover, such labels are conventionally used to refer to those agents who wrote the documentation, as opposed to those who inspired or formulated the content therein. So, they neither account for those situations in which many agents may have participated, either

directly or indirectly, nor do they account for the nature and scope of their participation. In addition, they do not provide a suitable structure in which to represent any changing patterns of participation as the document contents evolve and are used elsewhere. These shortcomings mean that questions regarding the origin of a requirement can only be handled rather simplistically at present. This suggests that participation details need to indicate the status of those agents who are party to the production of artifacts, along with the mode of their participation. Such details also need to be evolvable.

2.3. Requirements

All the above issues point to a need to maintain a detailed and dynamic model of those agents who have participated in the production of requirements artifacts. We refer to such a generative model as the "contribution structure" underlying the requirements. In addition, these issues imply that a dedicated approach is needed to guide the definition, redefinition, and use of this model. So, the basic requirements are:

- A means to differentiate the various ways in which agents can contribute to requirements artifacts, which also supplies the building blocks with which to model contribution structures.

- A way to account for the numerous relations that exist between the requirements artifacts themselves, to provide further information about contribution structures, and to allow the agents and artifacts to co-evolve.

- A basis for reasoning with and about the information modelled by the contribution structures, so selective information about agents can be retrieved to extend artifact-based RT, amongst many other utilities.

2.4. Approach & assumptions

Figure 1 provides an overview of the approach. The approach basically involves linking tangible RE artifacts (i.e., *contributions*), to details of those agents who have contributed to their production (i.e., *contributors*), using *contribution relations*.

By the term "artifact", we refer to any communicative occurrence in the RE process with a physical existence of its own, which can either be:

- *Primitive* (i.e., composed of no other artifacts).

- *Composite* (i.e., composed of other artifacts).

We assume that all the tangible artifacts that are produced and exchanged in RE are held in an on-line artifact repository which handles conventional artifact-based RT. We make this assumption due to: (a) the increasing maturity of digital imaging and optical character recognition [Reinhardt, 1994]; (b) the sophistication of the techniques and applications available for document management [Dewire, 1994]; and (c) the existence of environments which deal with product interrelations and their traceability [Pohl & Jacobs, 1994].

By the term "agent", we refer to the human participants in the RE process, which are either:

- *Individual* (i.e., non-decomposable).
- *Group* (i.e., decomposable into further groups and/or individuals).

We assume that various agent details, such as names, positions, rights, and duties, are held in an organisational repository. These details could be configured to carry out any required forms of organisational modelling, and so show formal power relations, responsibility relations, and so forth.

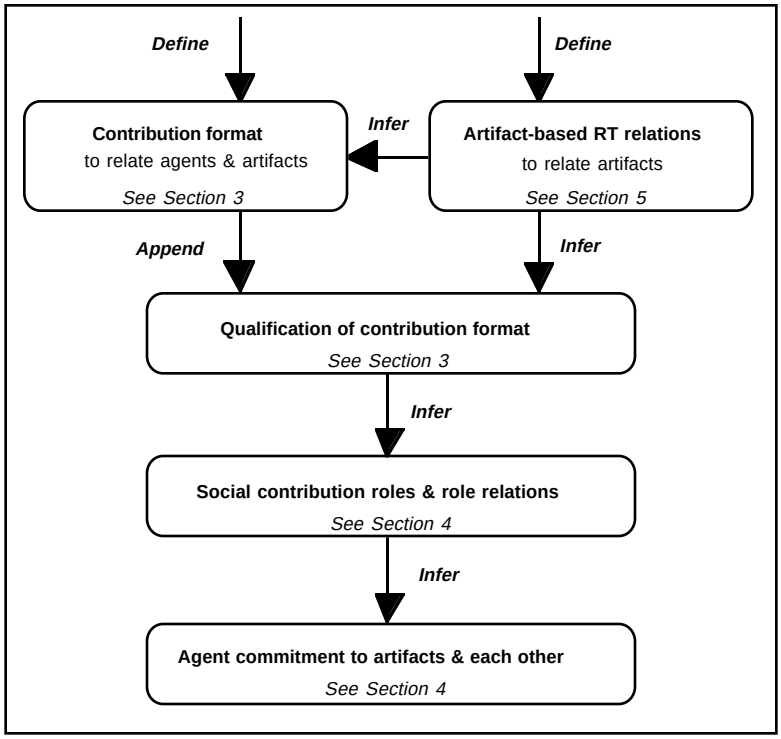

Figure 1: *Steps of the approach.*

The term "contribution structure" refers to all the contribution relations that have been defined for an artifact. The potential richness with which this contribution structure can be described depends upon the way in which the link between agents and artifacts is defined. As artifacts often depend on the existence of other artifacts, or are themselves decomposed into component artifacts, this description also depends upon how well these artifact-based relationships are defined and taken into account. The social infrastructure underlying RE is therefore described in terms of the contribution structures of all the artifacts it produces.

2.5. Related work

Although we are unaware of other research explicitly directed at the above issues, our work has been influenced by research in a number of areas. Research in the sociology of science and technology has proved insightful. For instance, various approaches have been advanced which examine how scientific knowledge, facts, and artifacts, are both related to, and influenced by, the social structures from which they arose (see [Callon *et al.*, 1986; Crane, 1972; Barnes & Edge, 1982; Bijker *et al.*, 1987; Latour, 1987; Latour & Woolgar, 1979; Law, 1991; Law, 1994] for examples of such approaches). In addition, we have been encouraged by work in the

area of information systems development, especially that which has been examining the implications of viewing development as a social process. For instance, there has been some evidence that an enhanced understanding of the development process is obtained when it is interpreted from a social action perspective (see [Hirschheim *et al.*, 1991]). Current research in software process modelling is also relevant, particularly that which explores the nature of the relationships between agents, their activities, and their products (see research by the NATURE project into the development of a comprehensive RE process meta-model [Jarke *et al.*, 1993; Jarke *et al.*, 1994a; Jarke *et al.*, 1994b; Moreno *et al.*, 1994; Rolland, 1994a; Rolland, 1994b]).

3. Relating agents & artifacts

Although the relation between agents and artifacts could be defined using terms like "contributed_to" and "contributed_by", these do not distinguish different types and degrees of participation, and so would not meet the basic requirements we listed in Section 2.3. Instead, they would lead to flat and coarse network models of the contribution structure. The crux of our approach is therefore to define this contribution relation in a way which differentiates the precise nature of each of the contributions, and which also provides a basis for modelling more granular and layered contribution structures. In this section, we present our scheme for doing this, and discuss how its relations can be further qualified. We start with an account of the work upon which the scheme is based.

3.1. Foundations

The scheme is derived from work in the area of *sociolinguistics*, and in particular, descriptive models of the interaction between language and social life (see [Hymes, 1972a]). Such models aim to provide finer-grained schemes through which to describe and analyse the components of communicative situations than those provided by the traditional dyadic models of communication theory (like that of [Shannon & Weaver, 1949]). More specifically, our scheme is based on Goffman's work on the nature of participation in social encounters [Goffman, 1974; Goffman, 1981], and has been motivated by related work on *framing* and *involvement strategies* [Tannen, 1993].

Goffman's work is concerned with placing the production and reception of talk within an interactional framework, so that it can be studied as a component of the full physical, social, and cultural environment in which it occurred. To enable such an analysis, he decomposes the crude concepts of "hearer" and "speaker" into their underlying constituents, which he refers to as *participant roles*. These roles provide smaller elements for identifying, codifying, organising, and referring to participants. More specifically, he refers to the set of categories obtained from the decomposition of "hearer" as the *participation framework*, and those obtained from the decomposition of "speaker" as the *production format*.

As we are primarily interested in modelling those agents directly involved in requirements production, in a manner which is amenable to further analysis and traceability, Goffman's notion of "production format" provides insight as to how this could be done. Here, he suggests 3 analytical capacities in which participants can "speak", which together clarify the notion of "speaker". He refers to these as:

- *Animator* - the transmitter or talking machine.

- *Author* - the composer of the lines.

- *Principal* - the motivator of the words or whose position they establish.

By layering and embedding these capacities, he further describes how situations of information dependency, such as retelling, can be accommodated.

	Production	**Reception**
Participant	Those agents **directly** involved in producing RE artifacts	Those agents who make use of RE artifacts, for whom they **have** been explicitly produced
Non-participant	Those agents **indirectly** involved in producing RE artifacts	Those agents who make use of RE artifacts, for whom they **have not** been explicitly produced

Table 1: *Partitioning the social dimension of RE.*

3.2. Contribution format

Requirements artifacts are produced and used within a social environment. By applying Goffman's frame analytic method to study the social organisation of this process, the RE space would be partitioned according to Table 1. In this paper, we only focus on the first quadrant of this figure, but we later anticipate examining the others, along with their interdependencies, since they frequently coincide and directly influence each other in critical ways. For now, and in the spirit of Goffman, we use the concept of a *contribution format* to define the nature of the contribution relations between agents and artifacts. Our scheme therefore delineates 3 fundamental capacities in which agents can contribute to artifacts, as shown in Figure 2, which together clarify the broad notion of "contribution". We refer to these as:

- *Principal* - the agent(s) who motivated the production of the artifact and whose position and/or belief is established by the information therein (i.e., committed to what it expresses and responsible for its effect or consequences).

- *Author* - the agents(s) who chose, formulated, and organised the content and structure of the information in the artifact (i.e., responsible for its syntax and semantics).

- *Documentor* - the agent(s) who captured, recorded, or transcribed the information in the artifact (i.e., responsible for its physical manifestation).

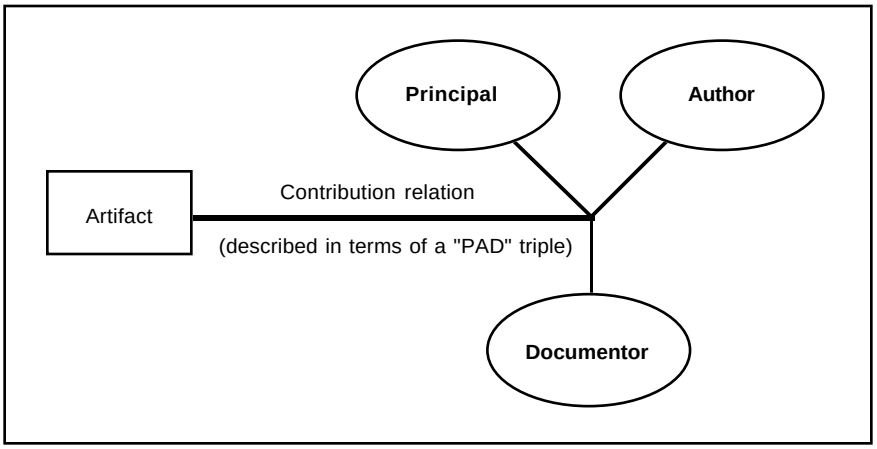

Figure 2: *The contribution format of an artifact.*

3.3. Further qualification

The approach recommends using the attributes given in Figure 3 to specify further details about the nature of each of the above capacities. As we are presently examining ways in which such qualification can be obtained as a by-product of the approach, we only briefly mention these below.

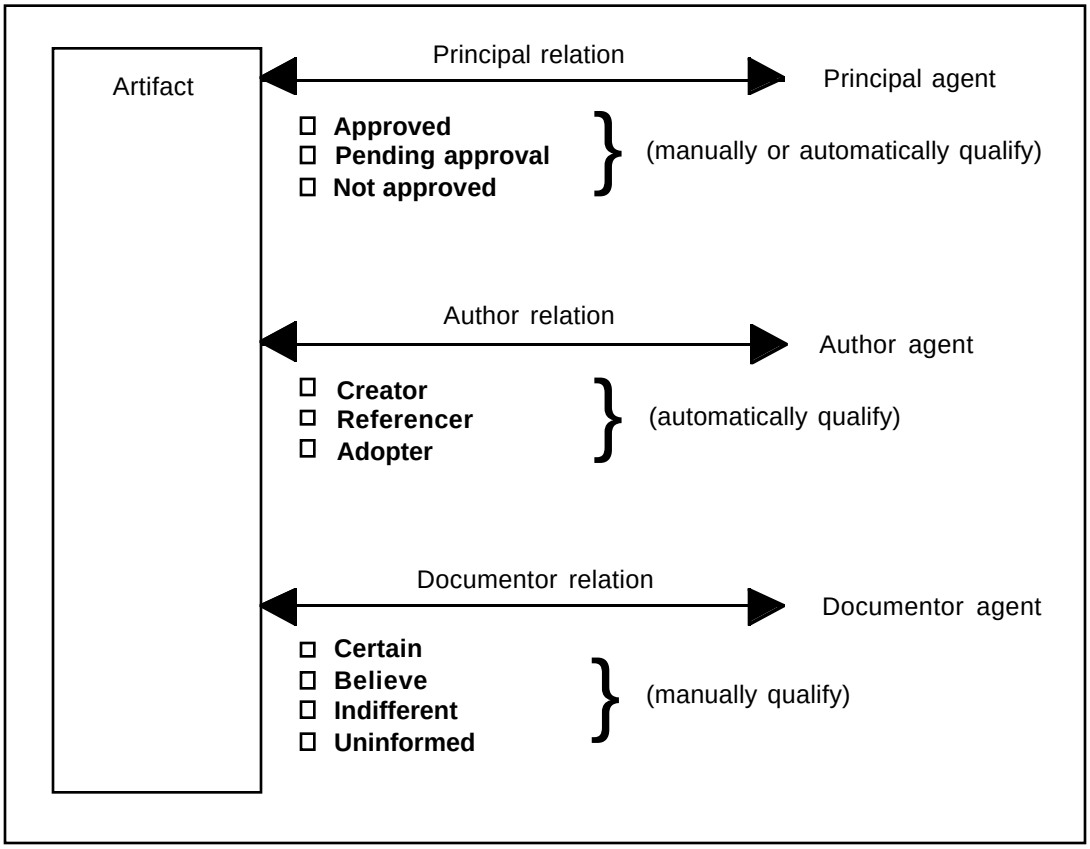

Figure 3: *Attributes to qualify the 3 capacities of the contribution format.*

Signatures are all-pervasive in systems and software development. They signify authorisation, stabilisation, and the transfer of commitments between agents. Therefore, the principal capacity is qualified to reflect *sign-off* procedures, so indicating whether the artifact is:

- *Approved* by principal.
- *Pending approval* by principal.
- *Not approved* by principal.

Those who document information usually have various types and degrees of commitment towards it, which can be reflected in the end result, and carried through to any subsequent encounters. Therefore, the documentor capacity is qualified to reflect the *mood* of the documentor, so adapting the mood types given in [Matthews, 1965]:

- *Certain* that the content of the artifact is true (i.e., *emphatic* mood).
- *Believe* that the content of the artifact is true (i.e., *period* mood).
- *Indifferent* to the truth value of the artifact's content (i.e., *quotative* and *report* moods; as either from a second-hand and indefinite source, or from a second-hand and known source, respectively).
- *Uninformed* about the truth value of the artifact's content (i.e., *indefinite* and *question* moods; as either nobody knows its value, or somebody knows its value, respectively).

The author capacity is qualified according to the relations that the artifact in question has to other artifacts. There are 2 possible scenarios, the details and implications of which are discussed further in Section 5.3:

- No relations exist, so the authorial status is that of *Creator*.
- Relations exist, so the authorial status is determined relative to the broad *communicative function* of each of these, and further distinguished by their *communicative purpose*.

4. Developing contribution structures

In this section, we describe how the information captured using the above scheme can be manipulated to infer, and so model, richer details about contribution structures. In particular, we show how it can be used to provide a clearer picture of how the agents are related to both artifacts and to each other. We explain how it can be used to determine details about the social contribution roles of agents and their resulting role relations. We also explain how this imparts detail about the *individual* and *collective commitment* of agents to artifacts, and about their *social commitments* to each other. These 3 commitment types are differentiated in [Castelfranchi, 1993].

4.1. Foundations

Levinson points out that, when an agent "speaks" in one of the analytical capacities defined by Goffman, they are also active in a particular *social role* from which the words take their authority [Levinson, 1988]. He maintains that these roles need to be distinguished because, whereas an agent's analytical capacity is likely to remain relatively constant, the social role in which they are active is likely to alter rather more frequently, which has associated implications for the granularity of any analysis that is possible. Furthermore, knowledge of such roles provides essential information about *person* and *social deixis*, which assists in the interpretation of communicated information [Levinson, 1983]. Levinson's extensions that we are mainly concerned with here are the distinctions he makes between *basic* and *derived* production roles. He regards Goffman's 3 capacities as basic production roles. He then proceeds to suggest numerous ways in which these could be re-assembled to derive more complex production roles which reflect those attended to, and distinguished in, actual language use.

4.2. Contribution roles & commitments

Following Levinson, we distinguish between the basic contribution roles that are defined in the contribution format (namely principal, author, and documentor), and the derived social contribution roles that can be inferred from these basic ones. This results in finer-grained, and hence more changeable social distinctions, and in information about the actual roles that are dynamically assigned between agents in practice. These extensions are important because: (a) the notion of "social role" is central to the study of social structures [Nadel, 1957]; (b) they provide a handle with which to explore the network of relations that exist between participants using social network analysis [Scott, 1991]; and (c) they can reveal information that assists with issues relating to *communicative competence* [Hymes, 1972b] and *social accountability* [Buttny, 1993]. A simple example of such a derivation is shown in Figure 4.

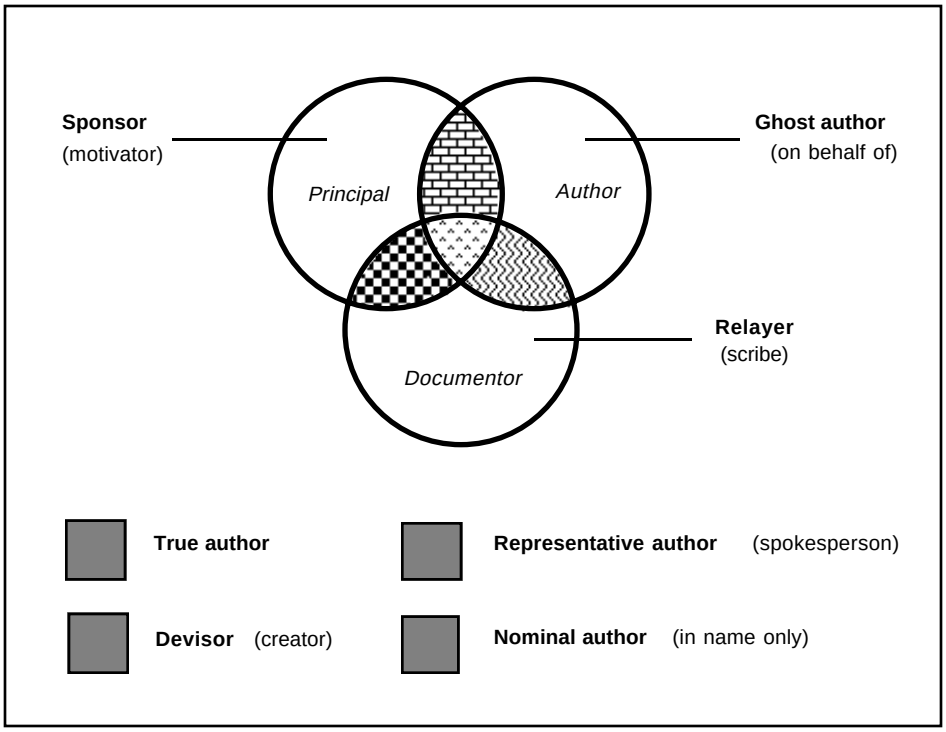

Figure 4: *An example set of derived social contribution roles.*

The relationships that exist between the agents themselves provides information about the role relations that have been dynamically formed and reinforced in practice. They also tell us about the ensuing social commitments, information that is rarely captured by formal organisational structures and pre-assigned project roles. The type, intensity, and directionality of these relations, and how they vary with respect to different artifacts, provides useful material for the analysis of informal organisational structures and project roles. For instance, it enables us to determine which agents act as direct contributors, indirect contributors, mediators, or third parties. It also tells us about local power and solidarity, recurrent or occasional collaboration amongst agents, emerging group alignments, substitute agents, reciprocation, and so forth.

Knowledge of an agent's social contribution role with respect to an artifact also tells us some rudimentary information about their commitments to it, by which we mean about those aspects of an artifact that specific agents can be called to account for. Such information is directly useful in filtering the retrieval of agent sources to reflect particular types of query or change proposal. It can help directly locate the source of motive, source of format, and so on. Table 2 provides a simple example set of commitments based upon the previous set of social contribution roles.

Committed to: (source of)	Physical appearance	Anticipated or realised effect	Structural form	Semantic content
True author	✔	✔	✔	✔
Devisor	✘	✔	✔	✔
Relayer	✔	✘	✘	✘
Sponsor	✘	✔	✘	✘
Representative	✔	✘	✔	✔
Nominal	✔	✔	✘	✘
Ghost	✘	✘	✔	✔

Table 2: *An example set of rudimentary commitments to artifacts.*

4.3. Example

To illustrate these extensions, consider Scenario 2 of Figure 5. Here, assume Olly has decided *the sensor needs to be polled once every twenty microseconds*, and Dave has written this down as a requirement in the requirements specification. Olly is both the principal and author of the written requirement, whereas Dave is its documentor. From this, we can infer that Olly and Dave stand in a devisor/relayer role relationship with respect to the requirement. As relayer of the requirement, all Dave is committed to is its physical appearance, so he can deal with any typographical queries or change requests. Queries like "Why twenty microseconds?", or change proposals like "Why not make it once every ten?", need recourse to Olly, as she is

committed to the actual content, and she is the one whose position would be challenged by any change. Now, if Olly and Dave stand in a devisor/relayer relationship for all the artifacts they jointly contribute to, the information that can be inferred about the social relationship and commitments between these 2 agents will contrast with that inferred if this were only a one-off. By adopting such an approach, a variety of social details can be disclosed that would otherwise remain hidden.

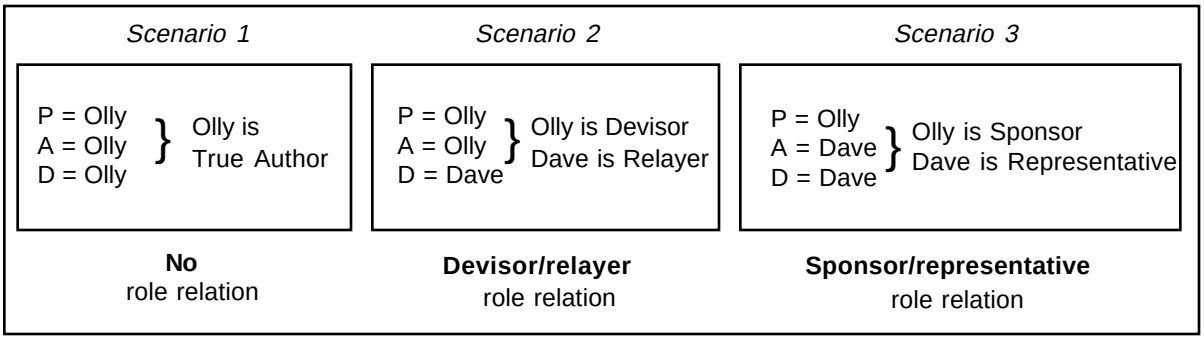

Figure 5: *An example set of derived social contribution roles & role relations.*

5. Relations between artifacts

If one requirements artifact is a subsequent specialisation of another, it seems reasonable to assume that some responsibility for the resulting artifact is retained by the original contributor(s). Our approach therefore needs to recognise, and deal with, the relations that exist within and between the artifacts themselves if it is to account for the linked and embedded nature of contributions. In this section, we propose a categorisation for artifact-based traceability relations to account for these observations, and we indicate how knowledge of these relations can be used to obtain a better understanding of the relationships between agents.

5.1. Categories

The relations an artifact has to other artifacts defines its *artifact space*. With respect to the artifacts of the requirements production process, such relations define its *artifact production base*. They make it possible to distinguish "original" artifacts from other varieties. We suggest there are 3 broad categories of relation, which describe alternative artifact-based structures, and provide the basis for different types of RT:

- *Temporal* relations, which describe the historical structure of development, and provide the means to trace requirements history.

- *Developmental* relations, which describe the logical structure of development, and provide the means to trace requirements flow-down.

- *Auxiliary relations*, which describe the additional ways in which information both within and across artifacts is related, and provide supplementary structures and forms of traceability.

The first 2 categories of relation capture the *macrostructure* of the RE process. The relations

used in these categories are fairly well established and are used to provide traditional forms of RT, so we go into no further detail here. The third category of relation captures the *microstructure* of the RE process. We are mainly concerned with developing a suitable set of relations within this category because: (a) there is no well established set of such relations in use; and (b) we suggest it is these relations that have subtle, though crucial, effects on the determination of contribution structures. The 2 types of auxiliary relation we are most concerned with we call *containment* relations and *connectivity* relations.

5.2. Containment relations

By recording the relation between a composite artifact and those other artifacts which are its components, we can make the task of assigning the contribution format much easier. Though clearly a composite artifact may have different agents acting in identical capacities with respect to its components, it is a default assumption that they are the same, until explicitly declared otherwise. As a containment relation is purely structural, no further leverage is to be gained from clarifying the precise reason for the containment. The presence of containment relations means that multiple contribution formats can be defined, interrelated, and managed. They lead to the layering effect shown in Figure 6.

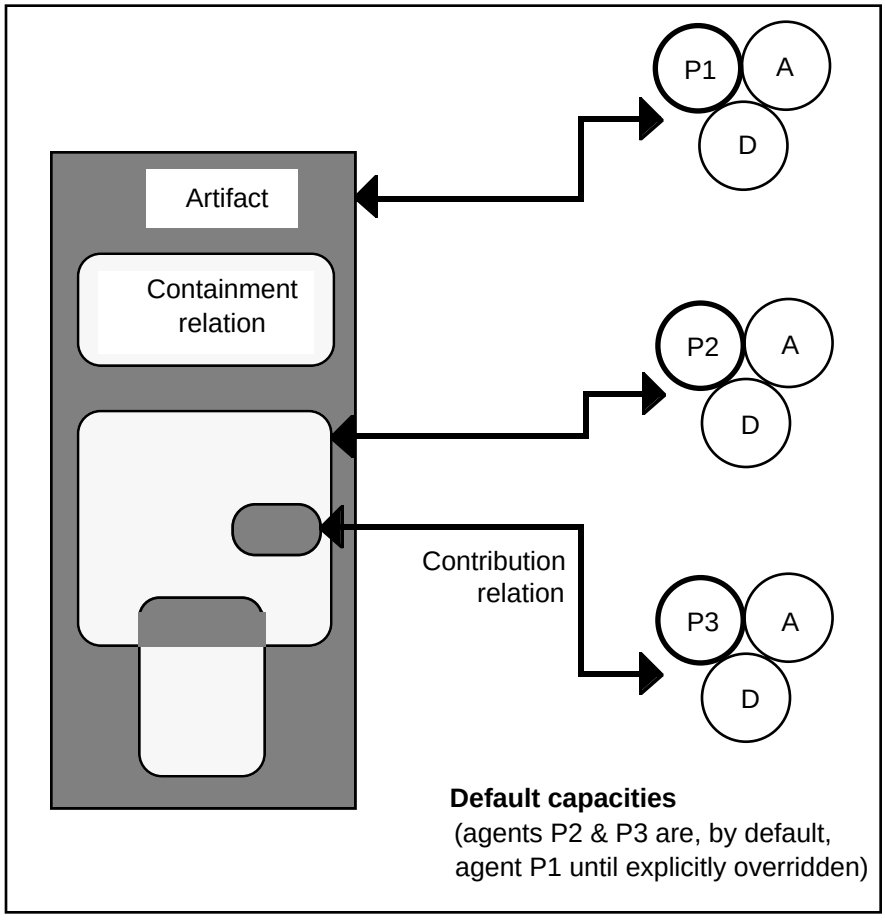

Figure 6: *The layered contribution formats of a composite artifact & containment relations.*

5.3. Connectivity relations

In order to inform the development of a useful set of connectivity relations, one which highlights the different ways they impinge on the determination of contributions, we look to work in *text linguistics* which examines the various ways in which textual occurrences can be related (see [De Beaugrande & Dressler, 1981] for an introduction to this area). In particular, we focus on the relations of *cohesion* and *coherence*, since these are purely text-centred notions.

Cohesion relations are those which deal with how the components of a surface text are mutually dependent and "stick together", so deal with *connectivity at the surface*. Many sets of cohesion relations have been proposed (as seen in [Halliday & Hasan, 1976; Crystal, 1987]). Coherence relations are those which deal with how the components of a text are mutually accessible and relevant, so deal with *connectivity of the underlying content*. Work on text coherence includes: (a) theories of discourse relations, which elucidate the implicit relations that exist between sentences of a text and bind it together (see [Grimes, 1975; Martin, 1983; Hobbs, 1985]); (b) theories of discourse structure, which explain the underlying hierarchical structure of texts and relations they embody (see [Polanyi, 1986; Grosz & Sidner, 1986]); and (c) theories which combine the previous ideas (see [Mann & Thompson, 1988]).

Our set of connectivity relations therefore draws upon the above work, which deals with connectivity at the sentential level of text, and extends the underlying concepts to encompass our definition of "artifact". For this reason, we do not claim to have an exhaustive and conclusive set, a contentious issue anyway according to [Knott & Dale, 1993], but rather a working set to examine the impact of such relations on modelling contribution structures. Our working set of relations is divided into 2 groups, which reflect the broad *communicative function* that a connectivity relation can serve:

- *References* - connectivity relations which function to reference exist when the physical content of the source and target artifacts does not overlap (functions to link existing content).

- *Adopts* - connectivity relations which function to adopt exist when the physical content of the source and target artifacts overlaps in some way (functions to embed existing content).

Tables 3 and 4 show how these broad groups are further decomposed.

In this paper, we only illustrate one way in which connectivity relations can be used to inform the modelling of contribution structures. As discussed in Section 3.3, the way in which we qualify the authorial status of an artifact depends on the connectivity relations it has to other artifacts. An example of this qualification, which uses the above tables to identify the type of the connectivity relation, is given in Figure 7. Such qualification provides information about the agent chains of dependency that emanate from the artifacts they have in common.

Purpose of relation (reason for referencing)	Function of referenced artifact B with respect to referencer artifact A (e.g.s of subsumed cohesion & coherence relations)
(a) To frame - B provides a framework in which to understand or interpret the information in A	To give preparatory information (background, circumstance, setting, locate) To give motivation (purpose, reason) To give the particular case (component, representative, decompose, subordinate) To give the general case (abstract, compose, collect, superordinate) To enable analysis (argue, evaluate, explain, interpret, critique, resolve) To enable inference (induct, deduct, abduct, refute)
(b) To match - A and B are juxtaposed for a specific reason	To compare with (liken, relate, analogy, similarity, resemblance) To contrast with (antithesis, difference) To coordinate with (synchronise, associate, alternative, option, branch)
(c) To substantiate - information in B strengthens that in A	To illustrate (exemplify, demonstrate, show) To support (solidify, assist, consolidate, justify, evidence, backing)
(d) To show causality - cause/consequence pairs	B (causes, enables) A A (replies to, result of, answers, responds to) B

Table 3: *The "references" group of connectivity relations.*

Purpose of relation (reason for adopting)	Function of adopter artifact A with respect to adopted artifact B (e.g.s of subsumed cohesion & coherence relations)
(a) To copy - use existing information as is	To copy information that exists in B (repeat, reassure, orient, resume, emphasise)
(b) To add - use existing information with change	To add to information that exists in B (define, describe, qualify, elaborate, develop, extend)
(c) To remove - use existing information with change	To remove information that exists in B (delete, dismiss, reject, repudiate, subtract, replace)
(d) To alter - use existing information with change	To alter the wording or structure of information that exists in B (refine, summarise, rephrase, rename, clarify, correct)

Table 4: *The "adopts" group of connectivity relations.*

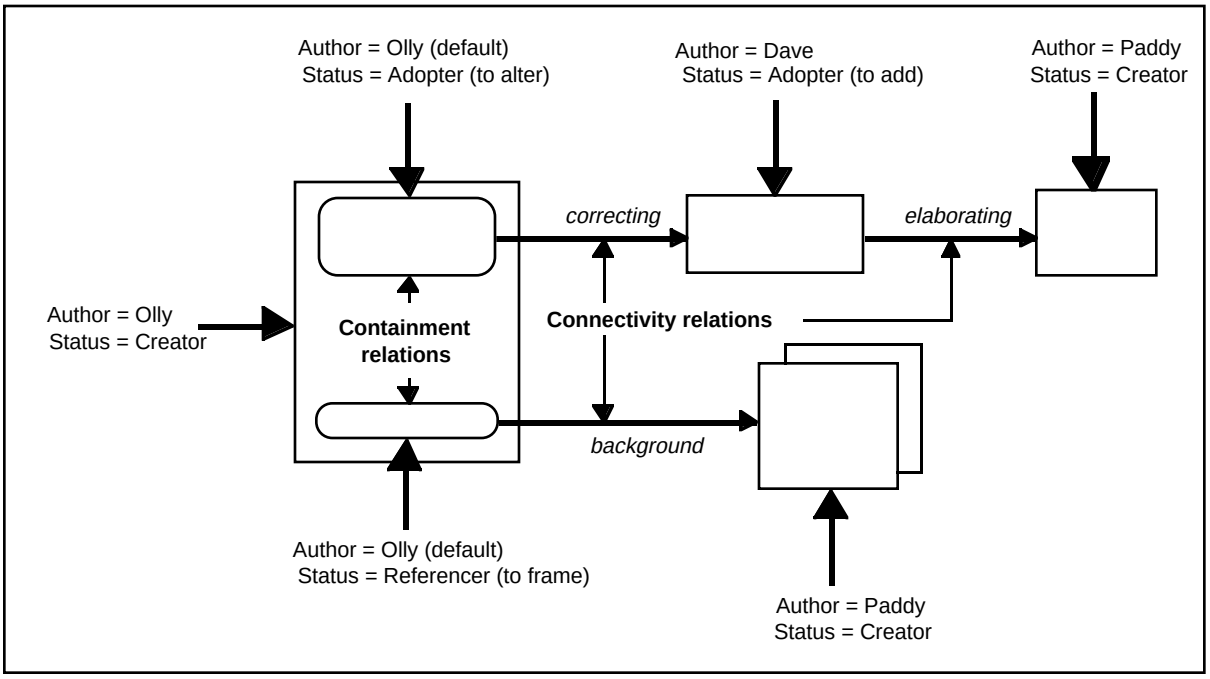

Figure 7: *An example set of connectivity relations used to determine the authorial status.*

6. Implementation of the approach

In this section, we describe the tool we have prototyped to demonstrate and evaluate our approach. In so doing, we highlight how the approach could be supported by minimal extensions to existing tools, notably to standard document preparation systems. We provide a simple example scenario to illustrate its operation.

6.1. Tool support

We have developed a prototype tool in which conventional notions of artifact-based RT can be extended with associated contribution structures. A schematic of this tool is given in Figure 8. The tool has been implemented using a combination of HyperCard and MacPROLOG, as the front-end and back-end respectively. It assumes that requirements artifacts are held in an on-line repository of some form, which manages the artifact-based traceability relations we identified in Section 5, as this is not meant to be its focus. We assume that the connectivity and containment relations are defined in the artifacts themselves, for use of the markup extensions we describe below, and that the source and target of these relations are hypertextually linked. For instance, this could be done using the *Hypertext Markup Language* (HTML) instantiation of the *Standard Generalized Markup Language* (SGML), by introducing high level link semantics of "references", "adopts", and "contains" for the hypertext relations (refer to [Goldfarb, 1981; ISO, 1986; Berners-Lee & Connolly, 1993] for specifications of these languages).

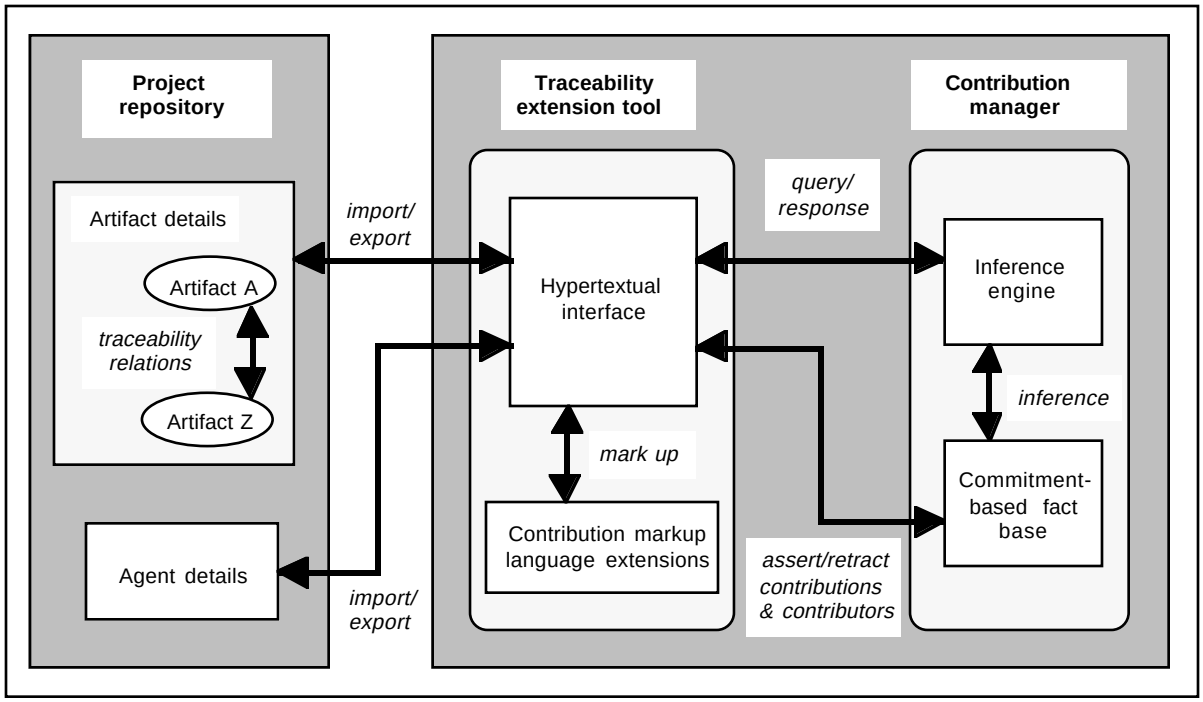

Figure 8: *The architecture of the tool.*

The *traceability extension tool* is implemented in HyperCard. It provides a hypertextual interface to the information held in the project repository. It enables additional artifact-based traceability relations and agent details to be interactively defined. Details about an artifact's contribution format can be similarly defined. These details are reflected in the underlying descriptive markup of the artifact using the primitive elements shown in Figure 9. These elements could be implemented as extensions to the HTML Document Type Definition.

Artifact elements	**Tags for start & end markers**
Principal	<ContribP=[Agent,Qualification]> </ContribP=[Agent,Qualification]>
Author	<ContribA=[Agent,Qualification]> </ContribA=[Agent,Qualification]>
Documentor	<ContribD=[Agent,Qualification]> </ContribD=[Agent,Qualification]>

Rules to interpret and use these elements are defined in the contribution manager's inference engine

Figure 9: *The basics of the contribution markup language extensions.*

The *contribution manager* contains rules that use the information captured by this markup to infer details regarding default agent capacities, the social roles of agents, their commitments, and so forth. It is therefore responsible for modelling and maintaining the contribution structures. By selecting and querying a section of an artifact, its *Artifact Profile* can be displayed. This describes all the artifacts and agents related to the artifact, each of which is a hypertextual trace anchor, along with the details of every relation. *Agent Profiles* can be displayed in a similar way. These profiles are shown in Figure 10, and act as only one of the many navigational springboards the tool provides to instigate various forms of traceability.

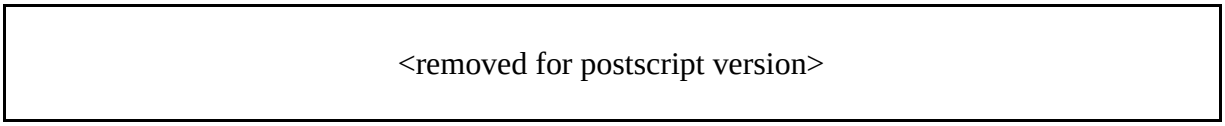

Figure 10: *Artifact & agent profiles.*

6.2. Example scenario

We use the following scenario to illustrate how our approach uncovers details about the social dimension of requirements production, and to indicate how access to these details can inform practice. It shows how, by the addition of 3 contribution tags in the artifact production process, relevant information becomes available that would otherwise remain obscured.

SCENARIO

A software project began with a wish list, reporting the needs from a **group of users**, which was written up by a **scribe** and authorised by a **project leader**. The project leader then held a meeting, of which an audio tape record was made, to discuss the wish list with **stakeholders**. A direct transcript of the meeting was subsequently made by a **couple of secretaries**. From the transcript and wish list, along with numerous other input documents, an initial requirements specification (RS) was written by a **group of requirements engineers**. Following circulation to, and comments from, various interested parties, a revised version of the RS was written. In particular, an alteration had been made to paragraph x as a result of an email message from the Managing Director's **Personal Assistant** (the M.D.'s P.A.) to the project leader. In this message, the **M.D.** passed on a verbal change request she received from **user 1** (a member of the group of users above). This corrected version of paragraph x becomes paragraph y in the revised RS.

Unfortunately, **member 2** of the group of requirements engineers inadvertently introduced an error when carrying out this change, largely because he did not acknowledge the subtlety of the wording in a fragment of the email message. This was because he had not been involved in the original discussion about the requirement at issue and had assumed that the M.D. was being unnecessarily fussy with wording. In checking the revised RS, **member 3** of the group of requirements engineers noticed the problem with the requirement specified in paragraph y.

In the following discussion of the example, it may be helpful to refer to Figure 1. We only intend to illustrate some of the information that can be obtained from modelling contribution structures, to show how the approach provides a way for member 3 to get an overall picture in which to understand the change, locate those involved, and so address the problem in conjunction with the most suitable agents.

Firstly, the artifact-based RT relations are defined. Figure 11 shows the standard temporal and developmental relations between the artifacts produced. Note that these are the relations specified and maintained by conventional project repositories providing RT. Figure 12 shows the additional containment and connectivity relations between the artifacts produced. For example, since paragraph y corrects paragraph x, they are linked by an adopts (to alter) connectivity relation. Our tool represents and handles all of these relations. So, if member 3 queries paragraph y, these artifact-based RT relations locate the email message as the reason for the change, and they retrieve the various paths back to the origin of paragraph y in the wish list.

Secondly, Figure 12 illustrates how the contribution format has been defined for these artifacts. The underlining signifies those capacities which can be automatically determined from the containment relations. This means that each of the relevant artifacts retrieved as a result of the above query can be augmented with its associated contributors, further indicating the capacities

in which they have contributed. This information points out: (a) when member 2 first became involved, and in what capacity; (b) who was involved in the same capacity with the previous version of paragraph y, namely member 1; (c) member 2's relationship to, and previous involvement with, member 1; and so on. For clarity, we have not qualified the contribution format in Figure 12. However, we can see that although member 2 is the author of paragraph y, he is in fact altering member 1's authored contribution of paragraph x because of the M.D.s authored contribution in the email message, and we can thereby extract the authorship dependencies between agents.

As explained in Section 4, the tool can determine the social contribution roles and role relations of all those involved, by manipulation of the above details. It can further infer details about individual, collective, and social commitments. From this extra information, a query of paragraph y results in additional information related to the social infrastructure underlying its production. For instance: (a) we are alerted to the fact that the M.D. was acting on behalf of user 1 when requesting the change, so only superficially the change instigator, and their respective roles of ghost author and motivator delineate where their commitments lie; (b) we are alerted to the basis for this role relationship in their joint collaboration in the meeting; (c) we are alerted to the fact that user 1 is the original human source of content and motive for paragraph y, as the original devisor in the wish list, so is the agent ultimately committed to its realisation and effect; and so on.

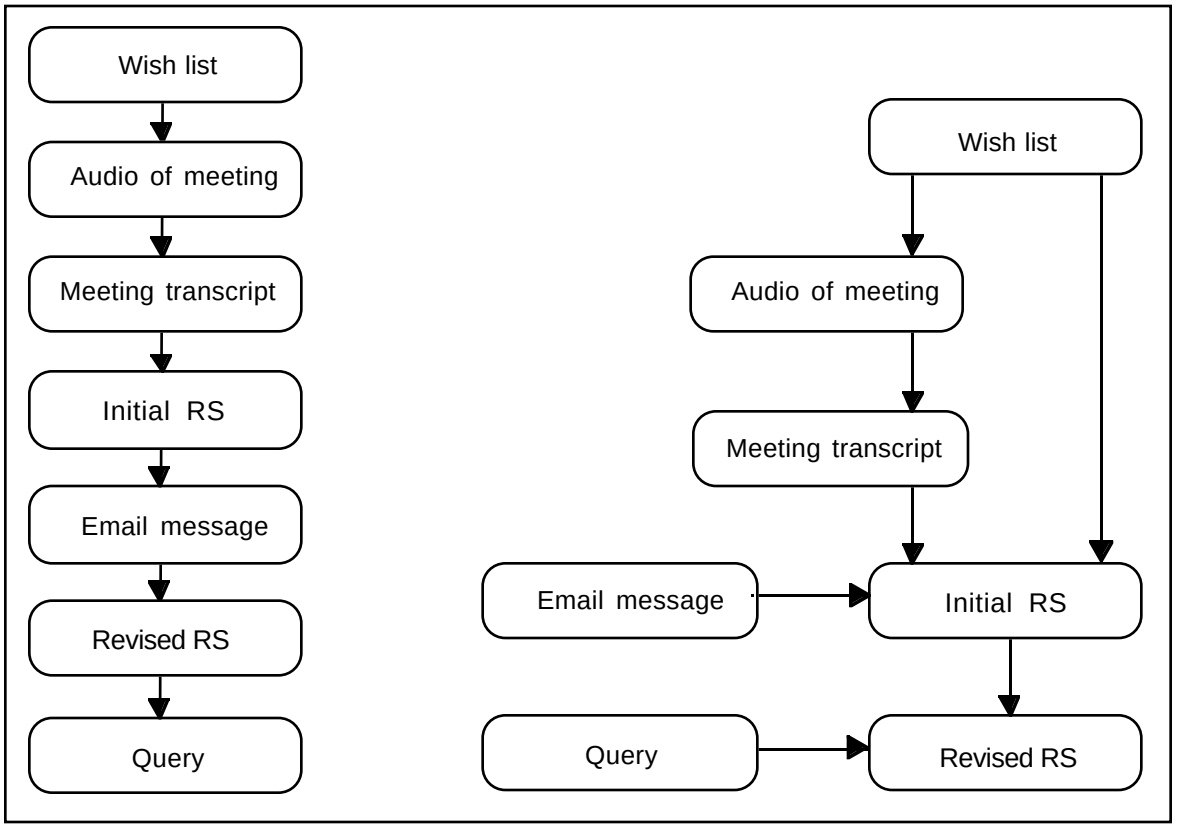

Figure 11: *The historical & developmental order of artifact production.*

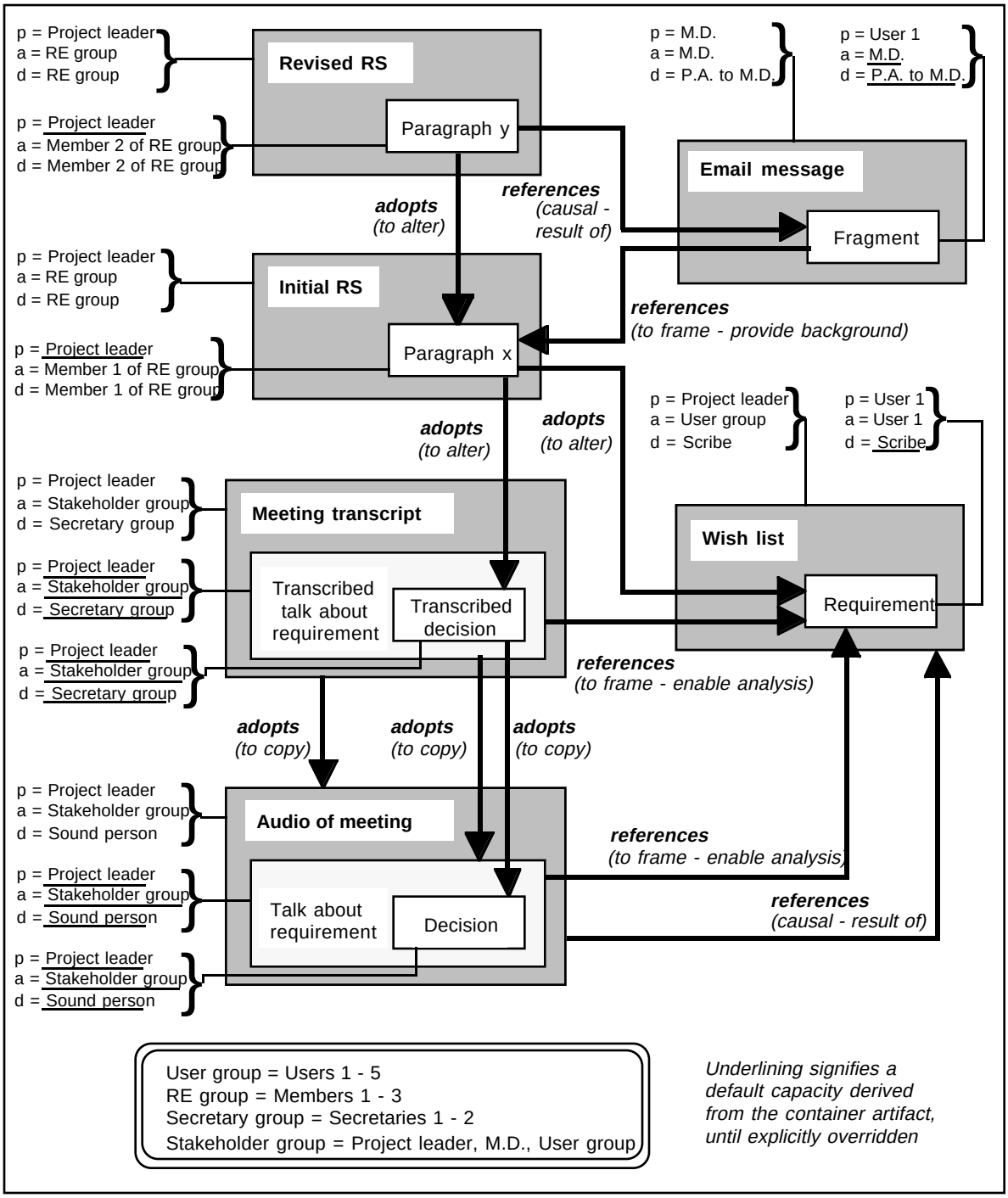

Figure 12: *The connectivity relations, containment relations, & contribution capacities.*

7. Discussion

In this section, we provide a critique of the approach, based upon preliminary evaluation with the above tool, and refer to our research agenda.

7.1. Strengths

The approach provides a way to deal with the absence of required information, supplement any documented information, and deal with the human side of the requirements change process. This is because it makes it possible to selectively identify the most appropriate agents to provide information or to involve in the change process. This helps ensure that requirements stay modifiable and maintainable. In addition, priority access structures can be constructed to guide these activities, since the complex nature of the underlying contributions has been handled in a disciplined way. Also, we believe that the approach provides a better basis for the many speech-act-based forms of analysis that are often carried out to examine the communication that has taken place in development. This is because knowledge of the underlying social network is a prerequisite for such analyses; agents communicate as the incumbents of social roles, which obviously impacts the illocutionary force. For example, to distinguish a "representative" speech act (as defined in [Searle, 1969]) as one of "asserting" from one of "summarising", details about both the informational and social arrangements need to be known a priori. Furthermore, the ability to identify implicit and derived group contributions, in addition to the explicit ones, means that more suitable forms of group-based analysis can be invoked in such situations (see [Hughes, 1984] for a discussion of such issues).

7.2. Limitations

A potential issue, and one we are looking into in some detail, is that of organisational resistance to its application. This is because the provision of clearer patterns of accountability has both positive and negative aspects ([Nissenbaum, 1994] provides an overview of some of these issues). In addition, we have yet to properly examine the issues involved in the scaleability of the approach to problems of industrial size. Other problems may arise from the reliance on people to instantiate the contribution format, as well as from their ability to characterise it according to our scheme. However, this scheme is only an initial one which has been designed to evaluate the basic ideas, so the actual terms we have chosen are not critical. An alternative approach would be to uncover the contribution categories that the participants themselves orient to whilst in the process of producing requirements artifacts, in order to make sense of this process, and use these in the contribution format. Similarly, to appeal to any institutional roles, role-relations, and commitments that are found to be prevalent in the domain of RE (like managers, consultants, elicitors, facilitators, user representatives, and so forth). To acquire such categories, in-depth field studies of actual working practices and arrangements would be essential.

7.3. On-going & future work

The approach requires more refinement and further critical evaluation. To do this, we are currently enhancing the tool described in Section 6, so that case studies can be carried out with practitioners. Future work will involve: (a) examining ways in which the contribution format could be automatically captured during artifact production; (b) examining how suitable communication tools could be integrated to automatically instigate any required communication

with agents, and in accordance with their preferred protocols; (c) investigating how the approach could be coupled with schemes that support discussion about requirements; and (d) looking at the possibilities that materialise for project management (i.e., by linking contribution structures to organisational models, predefined and actual organisational structures could be compared, dynamic details could be obtained about power relations and alliances, and strategies could be uncovered for integrating new personnel and dealing with the ramifications of those that leave).

7.4. Summary

RT is a key technology for managing development in the face of evolving requirements. In this paper, we have explained the added value that can be gained by tying people into the RT equation, particularly since this provides the firmest of foundations for dealing with the many issues relating to pre-requirements traceability. We have outlined an approach to model and keep track of the contribution structures underlying evolving requirements artifacts. We have further indicated how the approach provides the ability to extend conventional forms of artifact-based RT with accompanying contribution structures, which thereby offers a way to accommodate the diverse forms of personnel-based RT that practitioners were found to need in our empirical studies. Finally, we have described how this approach is currently being implemented, refined, and evaluated.

References

[Alford, 1993]

Alford, M. (1993). Requirements Analysis Using the RDD-100 System Designer, *DRA Requirements Colloquium Proceedings: Analysis of Requirements for Software Intensive Systems*, 19-20 May, pp. 90-103.

[Barnes & Edge, 1982]

Barnes, B. & Edge, D. (eds.) (1982). *Science in Context: Readings in the Sociology of Science*, Open University Press.

[Berners-Lee & Connolly, 1993]

Berners-Lee, T. & Connolly, D. (1993). *HyperText Markup Language*, Specification, http://info.cern.ch/hypertext/WWW/ MarkUp /HTML.html.

[Bijker *et al.*, 1987]

Bijker, W. E., Hughes, T. P. & Pinch, T. (eds.) (1987). *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, MIT Press.

[Buttny, 1993]

Buttny, R. (1993). *Social Accountability in Communication*, SAGE Publications.

[Cadre, 1992]

Cadre Technologies, Inc. (1992). *Teamwork/RqT*, Marketing Literature.

[Callon *et al.*, 1986]

Callon, M., Law, J. & Rip, A. (1986). *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*, Macmillan.

[Castelfranchi, 1993]

Castelfranchi, C. (1993). Commitments: from individual intentions to groups and organizations, *AAAI'93 Workshop*.

[Crane, 1972]

Crane, D. (1972). *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*, The

University of Chicago Press.

[Crystal, 1987]

Crystal, D. (1987). *The Cambridge Encyclopedia of Language*, Cambridge University Press.

[De Beaugrande & Dressler, 1981]

De Beaugrande, R. A. & Dressler, W. U. (1981). *Introduction to Text Linguistics*, Longman.

[Dewire, 1994]

Dewire, D. T. (1994). *Text Management*, McGraw-Hill.

[Dorfman & Thayer, 1990]

Dorfman, M. & Thayer, R. H. (1990). *Standards, Guidelines, and Examples on Systems and Software Requirements Engineering*, IEEE Computer Society Press.

[Flynn & Dorfman, 1990]

Flynn, R. F. & Dorfman, M. (1990). The Automated Requirements Traceability System (ARTS): An Experience of Eight Years, in Thayer, R. H. & Dorfman, M. (eds.) *System and Software Requirements Engineering*, IEEE Computer Society Press, pp. 423-438.

[Goffman, 1974]

Goffman, E. (1974). *Frame Analysis: An Essay on the Organisation of Experience*, Harper & Row, New York.

[Goffman, 1981]

Goffman, E. (1981). *Forms of Talk*, Basil Blackwell, Oxford.

[Goldfarb, 1981]

Goldfarb, C. F. (1981). A Generalized Approach to Document Markup, Proceedings of the ACM SIGPLAN SIGOA Symposium on Text Manipulation, *SIGPLAN Notices*, Volume 16, Number 6, June, pp. 68-73.

[Gotel, 1992]

Gotel, O. C. Z. (1992). *Requirements Traceability*, BT/Oxford University Research Initiative Report, Centre for Requirements and Foundations, Oxford, December.

[Gotel & Finkelstein, 1993]

Gotel, O. C. Z. & Finkelstein, A. C. W. (1993). *An Analysis of the Requirements Traceability Problem*, Technical Report TR-93-41, Department of Computing, Imperial College, August.

[Gotel & Finkelstein, 1994a]

Gotel, O. C. Z. & Finkelstein, A. C. W. (1994). An Analysis of the Requirements Traceability Problem, *Proceedings of the IEEE International Conference on Requirements Engineering* (ICRE '94), Colorado Springs, Colorado, 18-22 April, pp. 94-101.

[Gotel & Finkelstein, 1994b]

Gotel, O. C. Z. & Finkelstein, A. C. W. (1994). Modelling the Contribution Structure Underlying Requirements, *Proceedings of the First International Workshop on Requirements Engineering: Foundation of Software Quality* (REFSQ '94), Utrecht, The Netherlands, 6-7 June.

[Grimes, 1975]

Grimes, J. (1975). *The Thread of Discourse*, Mouton, The Hague.

[Grosz & Sidner, 1986]

Grosz, B. J. & Sidner, C. L. (1986). Attention, Intentions, and the Structure of Discourse, *Computational Linguistics*, pp. 175-203.

[Halliday & Hasan, 1976]

Halliday M. & Hasan, R. (1976). *Cohesion in English*. Longman.

[Harrington & Rondeau, 1993]

Harrington, G. A. & Rondeau, K. M. (1993). *An Investigation of Requirements Traceability to Support Systems Development*, Naval Postgraduate School, Monterey, California, September.

[Hirschheim *et al.*, 1991]

Hirschheim, R., Klein, H. K. & Newman, M. (1991). Information Systems Development as Social Action: Theoretical Perspective and Practice, *OMEGA International Journal of Management Science*, Volume 19, Number 6, pp. 587-608.

[Hobbs, 1985]

Hobbs, J. A. (1985). *On the Coherence and Structure of Discourse*. Technical Report CSLI-85-37, Centre for the Study of Language and Information, Stanford University.

[Hughes, 1984]

Hughes, J. (1984). Group Speech Acts, *Linguistics and Philosophy*, Volume 7, pp. 379-395.

[Hymes, 1972a]

Hymes, D. (1972). Models of the Interaction of Language and Social Life, in Gumperz, J. & Hymes, D. (eds.) *Directions in Sociolinguistics*, Holt, Rinehart & Winston, New York.

[Hymes, 1972b]

Hymes, D. H. (1972). On Communicative Competence, in Pride, J. B. & Holmes, J. (eds.) (1972). *Sociolinguistics: Selected Readings*, Penguin Books.

[ISO, 1986]

ISO/DIS 8879. (1986). *Information Processing Systems - Text and Office Systems: Standard Generalized Markup Language (SGML)*, Geneva.

[Jarke *et al.*, 1993]

Jarke, M., Bubenko, J., Rolland, C., Sutcliffe, A. & Vassiliou, Y. (1993). Theories Underlying Requirements Engineering: An Overview of NATURE at Genesis, *Proceedings of the IEEE International Symposium on Requirements Engineering*, San Diego, California, 4-6 January.

[Jarke *et al.*, 1994a]

Jarke, M., Nissen, H. & Pohl, K. (1994). *Tool Integration in Evolving Information Systems Environments*, NATURE Report Series 94-01, RWTH Aachen, Germany.

[Jarke *et al.*, 1994b]

Jarke, M., Pohl, K., Rolland, C. & Schmitt, J. R. (1994). *Experience-Based Method Evaluation and Improvement: A Process Modeling Approach*, NATURE Report Series 94-15, RWTH Aachen, Germany.

[Knott & Dale, 1993]

Knott A. & Dale R. (1993). *Using Linguistic Phenomena to Motivate a Set of Rhetorical Relations*, Technical Report HCRC/RP-39, University of Edinburgh.

[Latour, 1987]

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society*, Open University Press.

[Latour & Woolgar, 1979]

Latour, B. & Woolgar, S. (1979). *Laboratory Life: The Social Construction of Scientific Facts*, SAGE Publications.

[Laubengayer & Spearman, 1994]

Laubengayer, R. C. & Spearman, J. S. (1994). *A Model of Pre-Requirements Specification (pre-RS) Traceability in the Department of Defense*, Naval Postgraduate School, Monterey, California, June.

[Law, 1991]

Law. J. (ed.) (1991). *A Sociology of Monsters? Essays on Power, Technology and Domination*, Sociological Review Monograph 38, Routledge.

[Law, 1994]

Law, J. (1994). *Organizing Modernity*, Blackwell Publishers.

[Levinson, 1983]
Levinson, S. C. (1983). *Pragmatics*, Cambridge University Press.

[Levinson, 1988]
Levinson, S. C. (1988). Putting Linguistics on a Proper Footing: Explorations in Goffman's Concepts of Participation, in Drew, P. & Wootton, A. (eds.) *Erving Goffman: Exploring the Interaction Order*, Polity Press and Basil Blackwell.

[Mann & Thompson, 1988]
Mann, W. C. & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a Functional Theory of Text Organization, *TEXT: An Interdisciplinary Journal for the Study of Discourse*, Volume 8, Number 3, pp. 243-281.

[Martin, 1983]
Martin, J. R. (1983). The Logic of English Text, in Petofi, J. S. & Sozer, E. (eds.) *Micro and Macroconnexity of Texts*, Helmut Buske Verlag, Hamburg.

[Matthews, 1965]
Matthews, G. H. (1965). *Hidatsa Syntax: Papers on Formal Linguistics*, Volume 3, Mouton, The Hague.

[Moreno *et al.*, 1994]
Moreno, M., Rolland, C. & Souveyet, C. (1994). A Generic Approach to Support a Way-of-Working Definition, *CAiSE '94 Proceedings*, Utrecht, The Netherlands, 6-10 June, pp. 367-379.

[MST, 1993]
Marconi Systems Technology. (1993). *Requirements Traceability and Management Product Description v2.9307*, GEC Marconi Ltd.

[Nadel, 1957]
Nadel, S. F. (1957). *The Theory of Social Structure*, Cohen & West.

[Nissenbaum, 1994]
Nissenbaum, H. (1994). Computing and Accountability, *Communications of the ACM*, Volume 37, Number 1, January, pp. 72-80.

[O'Sullivan *et al.*, 1989]
O'Sullivan, T., Hartley, J., Saunders, D. & Fiske, J. (1989). *Key Concepts in Communication*, Routledge.

[Pohl & Jacobs, 1994]
Pohl, K. & Jacobs, S. (1994). *Traceability between Cross-Functional-Teams*, NATURE Report Series 94-13, RWTH Aachen, Germany.

[Pohl *et al.*, 1994]
Pohl, K., Domges, R. & Jarke, M. (1994). *PRO-ART: PROcess based Approach to Requirements Traceability*, NATURE Report Series 94-07, RWTH Aachen, Germany.

[Polanyi, 1986]
Polanyi, L. (1986). *The Linguistic Discourse Model: Towards a Formal Theory of Discourse Structure*, Technical Report 6409, BBN Labs Inc., Cambridge, MA.

[Pyle *et al.*, 1993]
Pyle, I., Hruschka, P., Lissandre. M. & Jackson, K. (1993). *Real Time Systems: Investigating Industrial Practice*, John Wiley and Sons.

[QSS, 1994]
Quality Systems & Software. (1994). *DOORS: Management Through Requirements v2.1*, Marketing Literature.

[Ramesh & Edwards, 1993]
Ramesh, B. & Edwards, M. (1993). Issues in the Development of a Requirements Traceability

Model, *Proceedings of the IEEE International Symposium on Requirements Engineering*, San Diego, California, 4-6 January.

[Reinhardt, 1994]

Reinhardt, A. (1994). Managing the New Document, *BYTE*, August, pp. 91-104.

[Rolland, 1994a]

Rolland, C. (1994). Modeling the Evolution of Artifacts, *Proceedings of the IEEE International Conference on Requirements Engineering* (ICRE '94), Colorado Springs, Colorado, 18-22 April, pp. 216-219.

[Rolland, 1994b]

Rolland, C. (1994). *A Contextual Approach for the Requirements Engineering Process*, NATURE Report Series 94-04, RWTH Aachen, Germany.

[Scott, 1991]

Scott, J. (1991). *Social Network Analysis: A Handbook*, SAGE Publications.

[Searle, 1969]

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press.

[Shannon & Weaver, 1949]

Shannon, C. & Weaver, W. (1949). *The Mathematical Theory of Communication*, University of Illinois Press, Urbana.

[Tannen, 1993]

Tannen, D. (ed.) (1993). *Framing in Discourse*, Oxford University Press.