Evaluating the Effectiveness of a Collaborative Requirements Engineering Modeling Notation for Planning Globally Distributed Projects

P. Laurent¹, A. Steele¹, J. Cleland-Huang¹ and P. Mäeder²

¹Systems and Requirements Engineering Center, DePaul University, Chicago, IL, USA
²Technical University of Ilmenau, Ilmenau, Germany

Abstract - In many software projects, stakeholders are distributed across different time zones, organizations, and geographical locations. This creates challenges for conducting people-intensive activities such as requirements elicitation, analysis, and prioritization. To address these problems we previously introduced a visual modeling notation to help project managers plan the collaboration infrastructures needed to support requirements-related activities in globally distributed projects. In this paper we present a refined version of the notation and report on an observational study we conducted in which project managers used our notation to plan globally distributed projects. Results show that the modeling activity and the resulting diagrams helped the project managers to better understand the communication needs for the project, to identify potential communication and collaboration problems, and to proactively address the infrastructure and communication needs for the project.

Requirements, global projects, visual notation

1 Introduction

In globally distributed projects stakeholders are often separated across time-zones and geographical boundaries. This creates numerous challenges for eliciting, analyzing, negotiating, specifying, and managing requirements [1], especially in conducting activities that are typically performed in face-to-face meetings. Herbsleb’s study on communication problems identified several impedences in distributed projects, such as cultural differences, incompatible support environments, and disparities in domain expertise across sites [2], while Taweel observed that communication and coordination challenges resulted in delayed projects, poorly-defined requirements, and repetition in the software development effort [3]. Finally, Damien et al [4, 5] studied the ways in which development teams coordinated their efforts when working on interrelated requirements.

Results from a series of interviews we conducted with requirements engineers from six globally distributed projects [6] showed that failing to clearly identify critical stakeholders and their interactions, and to establish the necessary communication and tooling infrastructures negatively impacted the success of the project and led to disorganized stakeholder interactions, data overload, increased travel requirements, and inefficient processes for supporting specific requirements engineering tasks [7].

To address these challenges we developed the Collaborative Global Requirements Engineering Notation (CGREN) which equips project managers to plan, analyze, and optimize their distributed requirements engineering processes, so that they can better understand their existing processes, identify weaknesses and problems, and establish essential processes and infrastructures [6]. As an additional benefit, CGREN provides a common notation for modeling distributed requirements projects and activities, and thereby facilitates comparisons across projects. These comparisons make it possible to identify recurring patterns of collaboration, common obstacles, and best practices used for collaborative requirements engineering activities. Such observations enable researchers to propose new techniques or improve existing methods to handle the specific challenges of global requirements processes.

In this paper we present a refined version of the CGREN, and also describe a participatory study we conducted in which requirements analysts were asked to interactively use the CGREN to plan requirements engineering activities for distributed projects. The study was designed to evaluate the effectiveness of the CGREN taxonomy, notation, and process for supporting stakeholders in the process of planning distributed requirements engineering activities. The results of this study led to some improvements in the model and clearly show the benefits of using the CGREN.

The remainder of this paper is structured as follows. Sections 2 and 3 present the taxonomy and notation of the CGREN. Sections 4 and 5 describe the study we conducted and the subsequent modifications to CGREN. Section 6 describes related work and section 7 summarizes our findings.

2 CGREN Taxonomy

The initial CGREN taxonomy focused around entities of: roles, sites, and artifacts; as well as three general types of relations: houses, accesses, and communicates, that were observed between the entities.
• **Roles:** The Rational Unified Process (RUP) defines a *role*, as a “hat” which can be worn either by an individual or a group of people [8]. Our study results identified a set of commonly occurring roles including a Subject Matter Expert (SME), Requirements Analyst (RA), Customer, Location Spokesperson (LSP), Project Manager (PM), Developer, Tester, and User. The SMEs took on a number of domain specific roles such as Artist and Sales Person. The most commonly identified roles were the SME, RA, and LSP. Most projects we investigated did not officially have RAs; however the RA responsibilities were assigned to a variety of job titles such as project managers and lead developers. The RA role is responsible for overall management of the requirements elicitation process. Several projects also included the LSP role, which was responsible for coordinating the requirements-related processes at a specific location. In some cases the LSP also served as a language translator between local and remote stakeholders. The LSP role was assumed by personnel holding a variety of job titles such as technical lead and designated regional representative.

The role entity has two attributes of *subtype* and *multiplicity*. The *subtype* attribute can be set to any predefined role type (i.e. SME, RA etc), while the *multiplicity* attribute documents the number of stakeholders assuming the given role. CGREN adopts the counting concept used by Amazon’s Pirahã tribe by constraining multiplicity values to *one*, *few*, or *many*.

• **Locations:** By definition, a distributed project includes two or more distinct locations. We refer to each location as a *site*, and define it as a place at which at least one project stakeholder is situated. A site is characterized by the close proximity of stakeholders, and their ability to meet together frequently to engage in face-to-face meetings. Stakeholders at a single site are normally able to communicate using a shared primary language. The metamodel shows that a site is defined by location, (primary) language, and time zone attributes.

• **Artifacts:** An artifact is defined as the specification of a physical piece of information that is used or produced by a software development process, or by deployment and operation of a system [8]. The primary goal of the requirements elicitation process is to discover and document requirements for the system. These requirements may be represented textually in structured or unstructured formats and/or graphically. Documentation can therefore assume multiple formats including but not limited to: Word documents, databases, UML models, dataflow diagrams, and/or spreadsheets. Some artifacts are associated with a specific location and reside permanently on a shared drive, online library, or in a repository at a specific site; while other artifacts are frequently moved from stakeholder to stakeholder across multiple locations, primarily via email. As a result, the *artifact* entity in the meta-model is specialized into *Stationary* and...
Travelling artifacts. A Stationary artifact belongs to exactly one site and is accessed at that site by both local and remote stakeholders, while a travelling artifact has no persistent site and is passed between distributed stakeholders using some kind of ownership token.

- **Means of Communication**: The study also identified three commonly occurring communication patterns involving various roles and artifacts. A communicate distributed relationship represents direct communication between two specific roles. For example, in one project, SMEs in North America communicated with SMEs in Asia primarily using email; while in another project the RA in North America held regular teleconferences with developers in Europe. This type of communication was characterized by the medium used (i.e. telephone, web-conference, or email), and also by the multiplicity of participating roles (i.e. 1:N, N:M etc). The Communicate distributed relationship is represented in the meta-model as an association between roles, while the communication medium is represented in individual models as a stereotype. The multiplicity of participating roles is captured through the previously discussed multiplicity attribute.

A communicate co-located relationship also connects two roles; however it represents the case that the associated roles are co-located and can communicate face-to-face. Roles can engage in a communicate co-located relationship either by being situated at the same Site, or when one or more of the participating stakeholders travel to the other site. For example, in one project an RA was responsible for traveling to two North American sites and a European site in order to interview SMEs. Finally, the accesses relationship associates roles with artifacts and means that stakeholders adopting that role contribute to the construction or maintenance of the associated artifact. Access is defined as read (R), write (W), and read/write (RW). The meta-model depicts the accesses relationship as an association between role and artifact entities, while the type of access (R, W, or RW) is modeled as a stereotype and not visible in Figure 11.

### 3 Visual Notation

The purpose of our work was to develop a visual modeling notation that could be used by stakeholders to plan, evaluate, and manage the requirements process in a distributed project [6]. We evaluated the icons used to represent entities and communication in the meta-model through conducting online surveys of 50 Software Engineering students from DePaul University. In the first phase of the study, participants were given a description of the role or relationship and were presented with 3-5 icon options. They were then asked to select the most representative icon and to optionally provide a rationale for their choice. We conducted this phase of the study in two rounds, using the second round to present additional icons, and/or to narrow down choices for controversial elements of the notation. In a second complimentary study we presented participants with the entire set of role icons and a list of the specific stakeholder roles, and asked them to associate each icon with a role. To increase the readability of our models we decided to label each role icon with the specific role. The notation, depicted in Figures 1-6, and presented throughout the remainder of this paper is the new notation developed as a result of this series of studies.

#### A. Basic Elements

Stakeholder roles are depicted as human shapes (Figure 1) and shown as one, few, or many stakeholders (Figure 2). Various adornments are used to represent specific roles, for example the RA is given a pencil, the LSP is assigned a bullhorn, and the customer is given a paper currency. Sites are depicted as containers (Figure 3). Artifacts are represented using well recognized symbols such as a file folder, spreadsheet, or text document (Figure 4). Finally, relationships are depicted intuitively using arcs (Figure 5). A solid line represents co-located communication between roles, a dashed line represents distributed communication between roles, and a dotted line represents the relationship between a role and an artifact. Arcs are adorned by symbols (Figure 6) representing various media of distributed communication, such as email or phone.

#### B. Examples

The CGREN notation can be used to model a variety of concepts at varying levels of abstraction. For example, a general view of the project may show sites, key roles, primary communication paths, and artifacts visible at the global level. In contrast a more concrete view might map out the specific communication and infrastructure needed to support the elicitation phase of the requirements
process. Figure 7 depicts communication between an RA and a few SMEs at a single site. Figure 8 depicts communication between an RA in the USA and a LSP in Pakistan. The LSP is responsible for internal communication with SMEs at her site. Inter-site communication is supported by teleconferencing and through a document shared via email.

4 A Participatory Study

Our observatory study was conducted using a tactile approach in which icons were printed onto small cards, and the participants utilized a white board to construct their models (Figures 9-10). The study was designed to address three research questions (RQ):

- **RQ1**: To what extent are project managers able to utilize the CGREN to model distributed requirements engineering processes in their projects? Are any important concepts missing or in need of improvement?
- **RQ2**: Does the CGREN help analysts identify problems and/or improve the infrastructure of their projects?
- **RQ3**: What is an effective process model for utilizing CGREN to model a project?

C. Study Design and Execution

Each observation of a requirements analyst using CGREN involved a training and enactment phase.

- **Participants**: Three professional requirements analysts (RA), from technical consulting, research, and healthcare fields participated in this study. Their specific job titles were consultant, business analyst and director, respectively. Each observation was conducted individually with only the RA and one researcher present.

- **Training**: At the start of each session, the researcher presented several examples of CGREN models and demonstrated the modeling of a project at the whiteboard. Each participant was given a notation guide which included icons depicted in Figures 1-6 and was given the opportunity to ask clarification questions.

- **Design and Procedure**: Each participant created a CGREN model for a specific distributed requirements elicitation project in which they had recently engaged. Project meta-data such as domain, size, duration, and geographical locations, was also collected. The study involved a ‘think-aloud’ protocol augmented by specific questions from the researcher, and an exit survey based on the questions depicted in Table 1.

D. Case Study Example

To illustrate the kind of modeling activities that were conducted during the participatory study, we describe the diagrams that were constructed by the first RA.

RA1 worked as a requirements analyst for a technical consulting company that had been engaged to develop an epidemiology tracking tool. The consulting company had leased office space in the same city as their client. RA1 was one of several RAs who communicated with distributed SMEs and developers. The group of RAs collectively authored and managed the requirements using a commercially available requirements management tool. The RAs elicited and gathered requirements from the SMEs using a combination of individual phone calls and through email exchanges. Each of the RAs was assigned a specific topic area and interacted with the respective SME to elicit requirements. The lead developer and a couple of the managers had read and write access to the requirements repository. Specific project sites and stakeholder roles are depicted in Table 2.

E. Evaluation

Our study was qualitative in nature. Research questions were systematically answered as a result of observing the participants utilizing CGREN, reviewing transcripts of the sessions, and through evaluating the answers to the open-ended exit survey questions.

**RQ1**: To what extent can project managers use CGREN to model distributed requirements engineering processes? All of the participants were able to successfully
model the roles, locations, communication methods, and artifacts of their selected projects. When asked “were you able to model all the concepts from your project?” all three participants responded positively. Furthermore the models produced during each of the three sessions demonstrated that all three RAs developed models which they claimed fully represented their projects, and which were correct with respect to the metamodel. However, when specifically asked if any graphical symbols were missing, two of the participants mentioned the need for the notation to allow stakeholders to assume multiple roles, sometimes simultaneously, and sometimes at different phases in the project. RA2 also pointed out the need to “denote frequency of communication” in order to differentiate between varying communication frequencies along different communication channels. In general, the results of this study confirmed that CGREN provided the ability to model all the concepts from your project. RA1 stated that she gained “A better understanding of the project (and a) better understanding of the stakeholders, the access they had, and ... their reach (impact in the project).” Using the communication diagram (Figure 10) she identified a specific problem that occurred because of the distribution of the major stakeholders. In this case the lead developer was located in Knoxville, while most of the communication to establish requirements took place in Atlanta. As a result of modeling these interactions, the RA commented “Wait a minute, all this communication is happening here (while) we have this one person who has to do all of these things, but they’re doing it remotely.” She stated that if CGREN had been available to her earlier in the project, this observation would have led to restructuring of communication patterns.

RA2 noted that for their project “the model is helpful for showing that ... in some of my locations I don’t really have a Spokesperson. And so there’s (sic) multiple SMEs that I’m going to... and (it is unclear) to what extent are they truly the authority.” She also stated that as a result of modeling the stakeholder roles, this reinforced that it would be helpful for her to have a designated spokesperson for each site who would be responsible for identifying SMEs. She further commented that “there’s multiple SMEs that I’m going to. And so that’s a lot of people I’m communicating with. ... I feel like it would be helpful to have fewer people and more people that were kind of designated as Spokespeople,” which echoes the findings of Turner and Boehm that stress the importance of finding CRACK (Collaborative, Representative, Accountable, Committed, Knowledgeable) people during the requirements elicitation phase of project planning[9].

**TABLE 1. EXIT SURVEY QUESTIONS**

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How useful was the modeling notation?</td>
</tr>
<tr>
<td>a. Were you able to model all of the concepts from your project?</td>
</tr>
<tr>
<td>b. Any problems using the graphical symbols? Any concepts missing?</td>
</tr>
<tr>
<td>c. What was easy to model? What was difficult?</td>
</tr>
<tr>
<td>d. Were the stakeholder types and roles sufficient?</td>
</tr>
<tr>
<td>e. Was the one-few-concept effective?</td>
</tr>
<tr>
<td>2. What, if anything, did you gain from using CGREN?</td>
</tr>
<tr>
<td>3. Did CGREN help you identify any potential issues?</td>
</tr>
<tr>
<td>4. Would you use a software version of this tool during your next project?</td>
</tr>
<tr>
<td>If so, at what phase?</td>
</tr>
</tbody>
</table>

**TABLE 2. RA1’S PROJECT SITES AND STAKEHOLDERS**

<table>
<thead>
<tr>
<th>Site</th>
<th>Stakeholder Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA0 – Atlanta, GA</td>
<td>1 PM/Lead RA and 5-7 RAs (from consulting company)</td>
</tr>
<tr>
<td>Consulting company</td>
<td>10+ Developers</td>
</tr>
<tr>
<td>USA1 – Atlanta, GA</td>
<td>1 Higher-level Manager</td>
</tr>
<tr>
<td>Customer site</td>
<td>1 Manager/LSP</td>
</tr>
<tr>
<td></td>
<td>10+ SMEs</td>
</tr>
<tr>
<td>USA2 – Virginia/D.C area</td>
<td>1 Manager/executive</td>
</tr>
<tr>
<td></td>
<td>10+ Developers</td>
</tr>
<tr>
<td>USA3 – California</td>
<td>5+ Developers</td>
</tr>
<tr>
<td></td>
<td>10+ Testers</td>
</tr>
<tr>
<td>USA4 – Knoxville, TN</td>
<td>1 Lead Developer</td>
</tr>
</tbody>
</table>

Finally, RA3 pointed out that CGREN would “shed some light on what some of the possible constraints and limitations could be” with respect to the current project configuration. In particular she pointed out that in her project all communication was via email, and that planning in advance would enable better infrastructure setup that could include video-conferencing technology and other techniques to support communication between stakeholders. One of the key results of the modelling activity for the RAs was that the method and quantity of communication during the planning and execution phase of the project was highlighted. For example, RA1 commented “...I never really noticed that I didn’t talk to the testers, even though they definitely wrote their test cases and complained sometimes about the way we wrote our requirements... after doing this [exercise] now I notice it.” Both RA1 and RA2 noted that the exercise made them painfully aware of the complexity of their communication needs.

We noticed the paucity of different elicitation techniques used in the three projects. All of the RAs relied on individual interviews and group meetings either conducted in face-to-face meetings or using phone or video-conferencing technologies. There were no examples of more creative elicitation techniques such as Joint Application Design (JAD) sessions, creativity workshops, or even basic scenario-writing using storyboarding or other similar techniques [10].
As a result of this observation we noted that if CGREN were extended to include the notion of meeting types and/or elicitation techniques, it could serve to inspire and educate project stakeholders about new techniques, and encourage them to think beyond their previous planning experiences.

RQ3: What is an effective process model for utilizing CGREN to model a project? Based on our previous experiences and our observations during the participatory study we developed the following guidelines that can be used in conjunction with CGREN.

1. Identify primary locations and model them as sites.
2. Identify project-level organizational roles and assign them to specific sites as the organizational plan evolves.
3. For each site, identify key local roles and communication patterns between roles within the site.
4. Establish basic communication patterns between critical roles across sites and assign communication responsibilities to specific roles. In CGREN add appropriate relationship arcs and attach applicable communication media to each of the relationship arcs. Decide how each communication path will be supported by technology.
5. Determine the key artifacts that are to be created collaboratively, and model them along with each role’s access and privileges. Include the applicable tooling/version control infrastructure.
6. Revisit project-level organizational structures and ensure that all roles are assigned to specific sites.
7. Model specific elements of the requirements engineering process by mapping task-specific roles, artifacts, and communication mechanisms onto the previously identified sites.

This process can be supported through the use of exemplar project templates from previous projects. Ideally the CGREN modeling exercise would be conducted as part of the kick-off event, but it can also be revisited throughout the project. One of the RA’s in our study specifically mentioned that she saw the CGREN models as part of a “living document.”

5 Refining the Model

As a result of the study we extended the meta-model to support the notions of communication volume, multiple hats, and elicitation techniques. Figure 11 presents the new meta-model uses classes to model access with associated type and frequency attributes, and the communication class with frequency and media attributes. The frequency attribute addresses our study participants’ request to model the volume of communication between two roles. Three additional classes are added to the meta-model to depict the notion of elicitation techniques used with specific collaborative events. To this end, an event is modeled as a collaboration between participants. A collaboration is associated with meeting type (i.e. JAD, Storyboarding, etc), a meeting name, and an outlook-style schedule depicting actual meeting times and duration. The associated icons are shown in Figure 12.

Each participant has a role in the meeting and each collaboration is assigned to a primary site. Communication and participation elements are represented as associations in instantiated models, while the collaboration type is modeled using one of the meeting type entities in Figure 12. To support the extended taxonomy, we also added an additional “many hats” icon, and introduced the visual notation that the width of the communication arc is approximately proportional to the estimated communication frequency. In addition, we introduced the icons shown in Figure 12 to represent a variety of elicitation techniques.

Figure 13 provides an illustrated example of how the new taxonomy and related notation could be used to plan a globally distributed JAD session. In this session the JAD meeting is being organized at Location-1 by a project stakeholder wearing dual hats of JAD Facilitator and RA.
Many participants, including SMEs, a developer, and a tester all physically participate in the JAD session, while SMEs from Location 2 and an LSP from Location 3 participate remotely using video-conferencing. The Location2 LSP communicates with local developers if issues arise during the JAD session. Finally, a report is sent to the manager at Location 4 at the end of the session.

6 Related Work

Other techniques exist for modeling stakeholders and their communication channels within an organization and/or project. Organizational charts identify project participants but are rather rigid in nature and often fail to capture the realities of how information is disseminated in an organization, and how roles and responsibilities are assigned in real projects [11]. Damian et al [4] described the communication paths between stakeholders in distributed projects using requirements-centered social network (RCSN) models. CGREN adopts several concepts from the RCSN, namely identifying stakeholders, and modeling communication paths between sites. However, the CGREN provides a more expressive approach for modeling communication media, stakeholder collaborations by role, artifact types, and other requirements engineering activities. Other visually oriented methods for describing large projects and their interactions fall short of capturing the details of a globally distributed requirements engineering project.

7 Conclusions

This paper describes our observations of the use of CGREN by requirements analysts to plan distributed requirements engineering processes. In general, CGREN helped the analysts to identify important locations, roles and communication mechanisms. Furthermore, new aspects of CGREN introduced in this paper, such as the inclusion of icons for specific elicitation activities, introduce the potential for stimulating greater creativity and improving the effectiveness of the requirements elicitation process. Our future work will involve augmenting our previously created CGREN tool with the new and modified icons, and testing CGREN in industrial settings.

8 References