Analytical Evaluation of Notational Adaptations to Capture Location of Activities in Process Models

Technical report M3W-1

Sundar Gopalakrishnan and Guttorm Sindre

Department of Computer and Information Science

Norwegian University of Science and Technology

August 2011, version 1.0

Abstract

When modeling work processes in mobile and multi-channel information systems by means of languages such as BPMN or UML activity diagrams it might be relevant to capture the location or other aspects of the user context of the activities performed. This report tries first to give an overview of various ways location could be captured in mainstream diagram notations, and then to evaluate analytically the advantages and disadvantages of these various notation alternatives. This evaluation is largely based on nine principles for visual notations stated by Moody (2009). While UML activity diagram notation is used as the basis for many of the examples, the notational means explored are of a fairly generic nature and would be usable also with other notations, such as BPMN.

This work was partially funded by the Norwegian Research Council through the project M3W (Model-driven support for Multi-channel Mobile Work), which lasted from 2006-2010, but partially also conducted after the end of the M3W funding.

1. Introduction

For mobile and multi-channel information systems it is often relevant to model where something is supposed to take place. Even if geographical location is included in some enterprise architecture frameworks (Zachman, 1987), business process modeling notations seldom capture location. In BPMN (Wohed, van der Aalst, Dumas, ter Hofstede, & Russell, 2006), UML activity diagrams (Dumas & ter Hofstede, 2001) or similar process notations, so-called swim-lanes and pools may be used to indicate who (person/role/organizational unit) performs a certain task, but not where.

As emphasized in (Walderhaug, Stav, & Mikalsen, 2008), the location and context of activities performed is of much higher importance in mobile information systems. For instance, whether a certain information processing task is to be performed in the office, or while driving, or while doing a repair job in difficult terrain and weather conditions, will have a large impact both on quality, efficiency, and job satisfaction, and is therefore an important process design decision. Another point is that ubiquitous information systems often have problems with information overload. As pointed out by (Deiters, Löffeler, & Pfenningschmidt, 2003) information should be delivered just-in-time based on the user's context rather than all at once, and different types of information would be relevant in different locations. This makes it important to capture the various possible locations of the users at an early stage, to determine when the system can feasibly deliver various types of information.

Since diagrams often play an important part in early stage IS development, it is interesting to investigate the possibility of capturing the location and context of activities in process models. To establish which is the better notation for a certain purpose, one will need empirical data, for instance from experiments and case studies with practical usage of the modeling techniques in question. However, experiments and case studies are costly and time-consuming and the number of possible notational tricks that could be used to show location is huge, easily yielding a combinatorial explosion if all had to pitted against each other in comparative experiments or case studies. Hence, it also makes sense to employ analytical evaluations from a broader viewpoint, to identify notational options that would be particularly promising, and then go on to experiment further with this more limited number of notations.

The scope of this report is therefore limited to the analytical evaluation of notations, and there are two research objectives to be addressed:

1. establish an overview of different ways (i.e., how?) location could be captured in process diagrams
2. evaluate the candidates from step 1 with respect to some set of given criteria

The rest of the report is structured as follows: Section 2 provides an overview of various possible adaptations (whether smart or not) for capturing location in process diagrams, using adapted UML activity diagrams for the visual examples. Section 3 then discusses what criteria to use for the evaluation, and Section 4 thereafter presents the results of the evaluation. Section 5 makes a brief discussion and conclusion to the work.

2. Possible ways of showing location in diagrams

The typical way of using diagrams in conceptual modelling is by means of graphs, i.e., nodes connected by edges. Different types of nodes are most typically distinguished by having different shapes and/or different labels, and similarly edges may be distinguished by labels and/or using different line types. However, in general there is a much wider range of visual variables that can be used to convey meaning in diagrams. (Bertin, 1983) explains 8 different variables that can be used:

- planar variables: horizontal position, vertical position. An example of using vertical position would be e.g. in an organization chart where a unit placed higher up in the diagram is on a higher level in the organization hierarchy. An example of using horizontal position might be that an activity placed to the left in a process diagram is meant to precede an activity placed further to the right.
- retinal variables: size, value (a.k.a. brightness), texture (a.k.a. pattern), color, orientation, and shape. Shape is one of the most commonly used, e.g. nodes that are rounded rectangles are activities, whereas diamond shaped nodes are decisions. Theoretically, the same distinction could be shown even with all nodes having the same shape and using one of the other retinal variables instead, e.g. size (activity nodes bigger, decision nodes smaller), value (activity nodes brighter, decisions darker), texture (activity nodes dotted, decisions hatched), color (activities yellow, decisions blue), or orientation (activities horizontal, decisions tilted 45 degrees). As will be understood from the subsequent sections, however, many of these would not be good notations.

As Bertin points out these variables can be used alone or in combination to give different meanings to diagrammatic constructs. Using planar variables to indicate location of activities would mean that an activity node's placement in the diagram would denote the location where the activity takes place in the real world. Instead using retinal variables, variations in the visual appearance of the activity node would indicate the place where the activity is performed. Figure 1 illustrates more concretely how these variables could be used for showing location, either just adding text (first row), adding an extra location node (next three rows), or providing variations of the activity node itself, e.g., in the bordering line (last four rows), fill (three rows above those), shape, size, or orientation. Many of these alternatives would not provide good notations, but are included just for completeness. Finally, the last row indicates a possible usage of planar variables, a practical diagram in this case might contain several pools arranged horizontally or vertically, and the placement of an activity node within one of these pools would indicate its location.

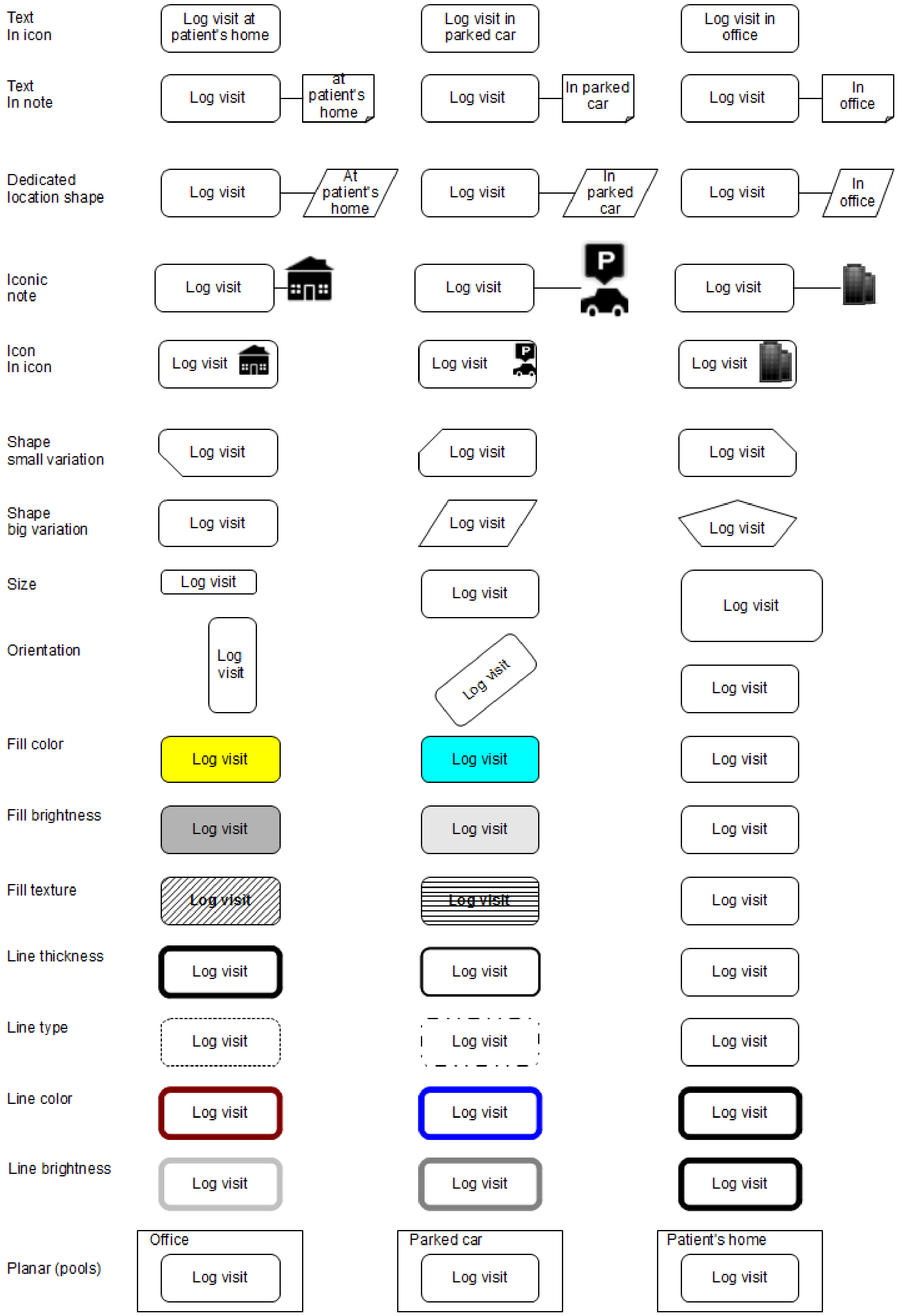

Figure 1: Possible ways to indicate location with activity nodes

In the following we illustrate some of the above possibilities with example diagrams, for all notations using the same case which shows part of a work process in home care. The process begins in the morning by the shift leader distributing patient visits on the available home care assistants. Each home care assistant then decides on the sequence in which patient's should be visited, possibly using some computer application calculating the quickest driving route, within the solution space given by time constraints for the visits, for instance that certain patients must be visited at specific times or in specific time intervals. The home care assistant then embarks on the trip, and while driving gets some preparatory information about the next patient on the visiting list. Getting this information while driving (which would typically mean it must be delivered as audio from a device in the car) would be a way of saving time compared to having to read the information textually before embarking on the visiting round, or in the parked car before entering each patient's home. During the visit itself, the home care assistant helps the patient with daily tasks. In case there is some special concern about the patient's health situation, a nurse may be contacted. The nurse will then evaluate the need and take further action if deemed necessary. Whether medical assistance was contacted or not, the home care assistant will anyway log some information about the visit performed, this is done when coming back to the parked car outside the patient's home before driving on to the next patient. The log will contain information about when the visit took place, the condition of the patient, and if there are any special needs related to the next visit (e.g., special items that must be shopped, light bulbs that must be replaced). Finally, when the day's visits are all done, the homecare assistant returns to the office and finalizes a daily report. Since a lot of information was recorded already along the route (esp. in the activity "Log info about visit") this should not require too much work, and part of the purpose of using a mobile information system here would be both to save time on this final activity (e.g., much of the report already done based on information input already, rather than writing it from scratch or copying it from handwritten notes made during the visits). Another benefit could be higher quality of the reports, because each visit is logged upon finishing it, rather than later in the day based on a perhaps less detailed memory of what happened.

Figures 2-3 shows the home care example by two different notations, one using the retinal variable texture, i.e., different locations are indicated by different pattern fills in the nodes, the other one using the alternative "icon in icon" from Figure 1. The latter can be said to make use of the retinal variable shape, although the shape of the activity icon itself is not changed, instead another and more intuitive icon is inserted inside this shape to indicate the location / context where it is performed.

In Appendix A we show the same example using a number of other possibilities from Figure 1.

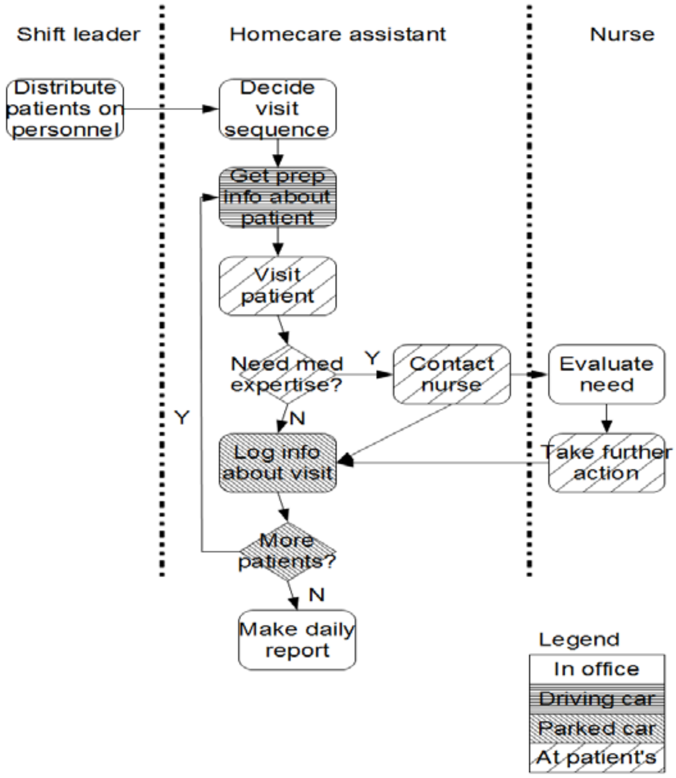

Figure 2: The home care example, with pattern-fills for location

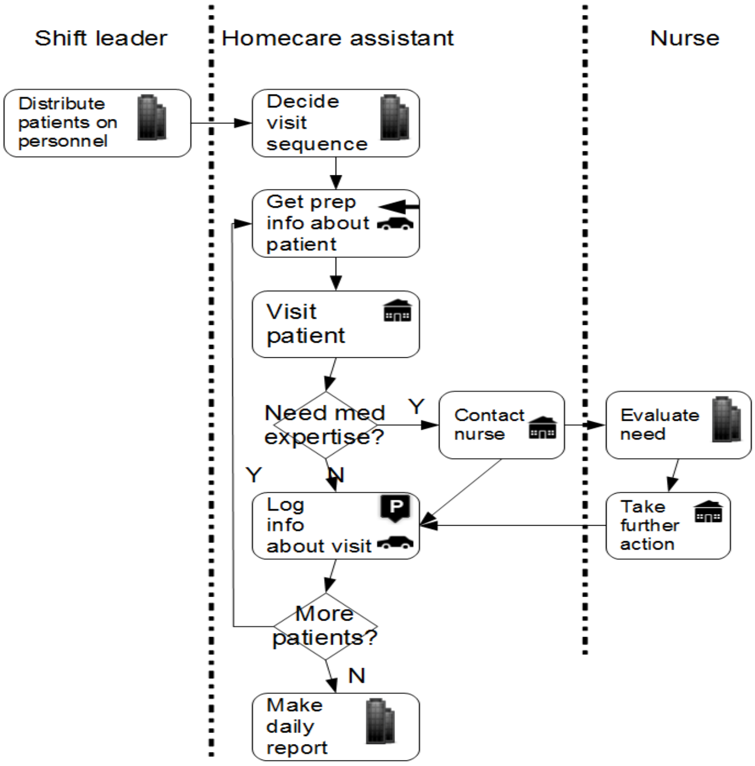

Figure 3: The home care example, with small icons inside the node to capture location

3. Criteria to be used for the evaluation

There has been a lot more research discussing what concepts modelling languages should include than how they should be represented visually. For instance, a large bulk of conceptual work has centered around the Bunge-Wand-Weber (BWW) ontology (Yair Wand & Weber, 1990; Y. Wand & Weber, 1993) and papers evaluating various modelling languages by means of this ontology, e.g., (Opdahl & Henderson-Sellers, 2002). There has been less research on the how the languages should best be represented visually, largely it seems visual representations have been chosen in a fairly ad hoc manner, for instance based on what was easily achieved with available drawing tools.

There has been some early efforts on defining criteria for good and bad visual representations, for instance (Tamassia, Di Battista, & Batini, 1988), but then typically focussing on layout of diagrams within the syntax rules given by a certain notation, rather than addressing the question of what symbols to choose in the first place. Typical goals for this kind of diagram layout is to reduce the number of crossing lines in the diagram, reduce the length of lines connecting various nodes, reduce the number of knees on such lines, achieve better balance of the diagram, etc. All of these can be related to aesthetics (making the diagram more pleasing to look at) and understandability, e.g. crossing lines will make it more difficult to see the relationships between nodes and make the diagram more messy, and the longer the links are, the more demanding it will be for the reader to follow them and see what nodes they actually connect. (Sindre, Gulla, & Jokstad, 1993) discuss similar metrics especially for decomposed graphs, i.e., where nodes can again be put inside other nodes, for instance due to a part-of or specialization relationship.

(Sindre, 1990), ch. 8.3, gives some criteria for visual representations, grouping these in four categories:

- perceptibility
- expressive power
- expressive economy, and
- method/tool potential

Perceptibility is by far the most elaborated criterion in this classification, with a number of subcriteria mentioned: the symbols used should be intuitive, discriminating between different symbols should be easy, usage of symbols should be consistent, symbols should be as simple as possible, and visual emphasis should be used in accordance with the relative importance of the concepts in the model. As for **expressive power**, it is observed that this is mainly a property of the conceptual basis, not its visual representation, but that a poorly chosen visual representation can destroy the expressive power, for instance if some concepts do not have a visual representation, or if a visual symbol is ambiguous in what concept it represents. As for

expressive economy this concerns how many symbols one needs to use in the external representation to convey various meanings in the conceptual basis. It is noted that the expressive economy of the external representation may be better than for the conceptual basis, for instance by introducing special symbols for groupings of underlying concepts that often occur together, or by omitting symbols that are understood from the context. Also it is noted that diagrams often have a better expressive economy than text or tables because a concept can be represented as one node and then linked to all other nodes that it relates to, whereas in textual statements (e.g. predicate logic) a concept would have to be mentioned for every relationship it participates in. Finally, for **method/tool potential** mentioned sub-criteria are whether the chosen notation lends itself to man-machine communication and information filtering, whether diagrams fit into standard paper and screen sizes, and some issues of ergonomy (e.g., avoiding features that are exhausting to look at, such as excessive blinking).

Another quite early effort to discuss quality of conceptual modelling languages is the SEQUAL framework. This was originally an effort to discuss the quality of conceptual models (Krogstie, Lindland, & Sindre, 1995; Lindland, Sindre, & Sølvberg, 1994), but later also extended into discussing the quality of the languages provided to make the models, (Krogstie, 2001, 2003; Krogstie & Arnesen, 2004; Nysetvold & Krogstie, 2006). Criteria used for the language are comprehensibility appropriateness (how easy it is for the stakeholders to interpret statements in the language), technical actor interpretation appropriateness (how easily an automated tool can interpret the statements), participant language knowledge appropriateness (how well the language is aligned with the stakeholders' previous knowledge of modelling languages), domain appropriateness (how well the language is able to capture the domain knowledge), knowledge externalizability appropriateness (how well the language allows the stakeholders to externalize their knowledge), and organisational appropriateness (how well the chosen language fits the goal of the modelling effort). The problem with these criteria is, however, that they are fairly vague and not specialized into any subcriteria that are easy to measure. Also, they are mostly focused on the conceptual level, although comprehensibility appropriateness would of course also cover the choice of visual notation.

A newer effort to define criteria for assessing the quality of a visual representation is found in (Moody, 2009), presenting 9 principles proposed for diagram notations:

1) *semiotic clarity* means that there should be a 1:1 mapping between graphical symbols and concepts. Defects might be symbol redundancy (several symbols for the same concept), symbol overload (the same symbol can refer to several concepts), symbol excess (symbols which do not denote any concept), and symbol deficit (concepts for which there is no symbol). Overload is considered the worst, causing ambiguity and misinterpretation. Deficit may sometimes be a necessity rather than a problem, because the diagram will become too cluttered if trying to show everything graphically (Moody, 2009).

2) *perceptual discriminability:* How easily and accurately can symbols be differentiated from each other? It will be easier for the user to distinguish between shapes that are obviously different from a quick glimpse (e.g., squares, circles, triangles) than between shapes that are different only in subtle details (e.g., rectangles with a varying height/width ratio, with or

without rounded corners, or with different border types or textual fonts). The redundant usage of several visual variables together (e.g., both colour and shape) may improve discriminability even further (e.g., green square, yellow circle, red triangle)

3) *semantic transparency:* How well does a symbol intuitively reflect its meaning? According to Moody, symbols can be either <u>immediate</u>, having a nice intuitive relationship with their corresponding concepts, <u>opaque</u>, having only an abstract relationship, or even worse, <u>perverse,</u> its intuitive interpretation being misleading vs. the represented concept.

4) *complexity management:* What constructs does the diagram notation have for supporting different levels of abstraction, information filtering, etc.? Decomposition is a keyword here, also it is preferable for a notation to avoid an explosion in the number of lines between nodes, since many, long and crossing lines will make a diagram much easier to read than fewer and shorter lines which are not crossing.

5) *cognitive integration:* Does the notation provide explicit mechanisms to support navigation between different diagrams? This may be navigation between different abstraction levels, or between different types of diagrams, e.g., from an activity diagram to a class diagram, to see the structural properties of the information objects processed in some activity.

6) *visual expressiveness:* To what extent does the notation utilize the full range of visual variables available? As argued in (Moody, 2009) most notations mainly use shape - plus maybe various line types (e.g. full lines, dotted lines) and textual annotations inside diagrams - whereas e.g. colour and texture are little used.

7) *dual coding:* Using text to complement graphics. While the usage of only text could give poor visual discrimination (as indicated in principle 6), the usage of both text and visual cues could be an advantage, as this again gives more redundancy of information and reduces the likelihood for misunderstandings.

8) *graphic economy:* Avoiding a too large number of different symbols, which makes the notation hard to learn and understand.

9) *cognitive fit:* Trying to adapt the notation to the audience, i.e. possibly using different dialects with different stakeholder groups.

Moody's 9 principles have much in common with the criteria presented in (Sindre, 1990), e.g. both state that symbols should be easily discriminated, that they should be as intuitive as possible, that there should be support for complexity management, etc. We choose to use Moody's list of criteria here rather than Sindre's, for the following reasons:

- Moody's work is newer, and based on a lot of sources, while Sindre's work from 1990 appears more ad hoc
- Since Sindre is one of the co-authors of this technical report, and hence involved in the analytical evaluation to be done, it appears in a way more scientifically objective to use criteria developed by somebody else

However, we chose to make some adaptations to Moody's list of principles for the purpose of this particular evaluation. Cognitive Integration investigates the relationship between several sub-languages within one bigger language (e.g., UML). Our purpose in this technical report is to look only at process models, and the possibilities for cognitive integration should not be much affected by the choice of one or other visual variable for showing location. Hence this principle would not cause different scores between our various alternatives, and is therefore discarded. Cognitive Fit concerns the usage of different dialects of a language for different stakeholders, based on the specific needs and preferences of these stakeholder groups. To evaluate this criteria it would be an advantage to have a clear definition of the stakeholder groups in question, and have representatives from these stakeholder groups involved in the evaluation, rather than just having researchers perform the evaluation. Hence this criterion is also considered beyond the scope of this work reported here. Instead, we add another criterion: *Minimal deviation from mainstream notation.* This, of course could be assumed to be related to cognitive fit for technically skilled stakeholders (if they are already familiar with the mainstream notations), but one could easily imagine also some notations which deviate radically from the mainstream, but which are still much more appropriate for technically skilled stakeholders than for other stakeholders. Moreover, the point of the criterion of minimal deviation is not only in the increased likelihood that stakeholders already familiar with the mainstream notation will also understand small adaptations more easily, there are also other advantages, such as leveraging the investment already made in tools supporting the mainstream notation, especially if an adapted notation deviates so little from the original one that these tools may still be fully usable.

4. Defining a grading scale

It was decided to use a 5 point grading scale for evaluating the notations according to each of the criteria, with the grades - - (big disadvantage, meaning that the notation alternative does very poorly with respect to satisfying the criterion), - (disadvantage), blank (neutral), + (advantage), and ++ (big advantage). Of course, numeral grades for instance from 1 to 5 could have been used instead with the same results when comparing alternatives, but it was felt that plus and minus signs would appear more intuitive when displayed in a table. Also, it is of course open to debate whether a 5 point scale is the best, other numbers (both smaller and bigger granularity) would of course have been possible. A 5 point scale was chosen because this is quite common in various types of questionnaires and evaluations, moreover it would be easy to collapse a 5 point scale to 3 points instead for analysis if finding this suitable.

Given the grading scale it is also of interest to discuss the interpretation of this scale for each of the criteria used in the evaluation.

4.1 Semiotic clarity

Moody states that symbol overload is the worst problem for semiotic clarity, followed by symbol redundancy, excess, or deficit. For deficit it is even said that this might be a necessary trade-off with respect to a later criterion (graphical economy), but in that case the score should still be negative for semiotic clarity and then positive for graphical economy, to maintain orthogonality between these two criteria. Hence, it would seem plausible to use the following scale:

- **- -:** the notation alternative introduces symbol overload in its attempt to capture location,
- **-** : the notation alternative has problems with symbol redundancy , excess, or deficit for capturing location
- " " : the notation has a semiotically clear visual construct for locations, but no graphical means to distinguish between different locations
- **+:** the notation has semiotically clear visual constructs also for distinguishing graphically between different locations
- ++: the notation is semiotically clear in all respects (i.e., also for all other constructs, not only those dealing with locations); this last alternative possibly meaning that the adapted notation not only is semiotically clear in its own adaptations, but that the adaptations even fix problems that were present in the original notation.

4.2 Perceptual discriminability

This depends on how easy it is to distinguish different alternatives using the visual variable in question. There are several sub-questions that could be relevant here.

- *Small scale* or *Symbol-by-symbol discriminability*: how easy is it to correctly identify each symbol individually, i.e., discriminate it from other symbols. Let us say alternatives for this is poor (hard to discriminate, easily confused), medium (may require closer inspection, but not easily confused), and good (easy discrimination).
- *Large scale discriminability:* Putting two diagrams side by side, e.g., two versions of the same process model with some minor variations, how easy is it to spot these variations without closer inspection of the diagrams. Again with alternatives poor (hard to discriminate, easily confused), medium (may require closer inspection, but not easily confused), and good (easy discrimination).
- *Number of alternatives that can reliably be discriminated:* This varies a lot depending on visual variables. Users would be able to distinguish a large number of different icons with intuitive figurative meanings (cf. the symbols inside traffic signs) but fewer abstract shapes, and even fewer different sizes or angular orientations, since the distinctions between these gradually get more difficult when the number of necessary alternatives reaches some barrier. Similarly, most people (except those with reduced color vision) would be able to distinguish reliably between more alternatives with colors (maybe approx. 10, e.g. white, black, red, orange, yellow, green, turqouise, blue, violet, brown) than if restricted only to grey-tones or texture patterns (maybe 5). So let's say: Poor = difficult to discriminate between 5 alternatives. Medium = should be possible to distinguish between at least 5 alternatives, but not much more. Good = should be possible to distinguish between at least 10 alternatives.
- *Effects on the readability of text*: Essentially, discriminability concerns the chosen visual variable. However, nodes in diagrams typically have to contain some textual labels, and if so it might also be of interest if the choice of visual variable has some

effect on the readability of this text. Poor = readability of text is made much more difficult. Medium = readability is somewhat reduced. Good = readability is not affected. This choice of alternatives because it is hard to imagine notational choices that would instead enhance the readability of text beyond what you already have for plain black text in a plain white activity node.

We now have four sub-questions here, for a 5 point grading scale. A possible way to do this is to say that each "poor" counts for half a minus sign, and each "good" for half a plus sign, e.g., a notation alternative which is poor for all four sub-questions would end up with **- -,** if good for all four it will have ++, if good for two and medium for two, or good for three and poor for one, it will have +, etc. A problem with this, of course, is that some notations will end up in between grades, for instance with 1,5 +'es (e.g., if being good for three sub-questions and medium for the fourth). To use the grading scale as much as possible, we choose to round upwards in absolute values in all such cases, e.g. 1,5 is rounded up to 2, and -0,5 similarly rounded down to -1.

4.3 Semantic transparency

The question for semantic transparency is how the graphical elements intuitively correspond to the concepts they are trying to represent. Here we have the following scale:

- **- -:** the graphic is semantically perverse, most people would intuitively connect this visual variable with a meaning completely different from the one sought
- - : for many people the graphic is semantically perverse, most naturally associated with a different meaning, but there are also many for whom it would be opaque (cf. next item)
- " ": the graphic element is opaque, only having an abstract relationship to the underlying concept (but no intuitive confusion, as in the two items above)
- +: for many people, the graphic is opaque, but there are also many for whom it does have an intuitive association with the underlying concept
- ++: the graphic is semantically immediate, i.e., for most people it intuitively reflects the underlying concept, e.g., the location to be captured

4.4 Complexity management

There are various metrics for measuring the complexity of diagrams, for instance based on the number of nodes, number of edges, number of crossing edges, long egdes, egde knees, etc. (Sindre, et al., 1993; Tamassia, et al., 1988). To go into this in detail would be too lengthy here. Since diagrams contain textual labels, the complexity or length of these might also be an issue. We suggest the following five point scale:

- **- -:** the notation alternative causes a huge number of new nodes and edges in the diagram
- - : the notation alternative causes some new nodes or edges
- " ": no new nodes or edges, but textual labels become longer
- +: no new nodes or edges, no increase in textual labels

- ++: no new nodes or edges, no increase in textual labels, in addition the notation alternative introduces some new possibility for composition / decomposition which may be used for effective information hiding when working with large models

4.5 Visual expressiveness

According to Moody this promotes the usage of as many different visual variables as possible in a notation, rather than overusing just a few variables.

- **- -:** the notation alternative does not at all increase the visual expressiveness of the notation, instead promoting increased use of a variable already used before in this type of diagram
- - : while still using a visual variable used from before, the alternative at least does this in a quite different way
- " ": the notation alternative associates meaning to a variable which has not had meaning before, but which may yet have been accidentally used (without a particular meaning) in such diagrams previously
- +: the notation alternative opens up the usage of a variable not used before
- ++: the notation alternative opens up the usage of several visual variables not used before

4.6 Dual coding

Dual coding is about the combination of graphics and text, or the combination of several redundant visual elements, to further enhance the understanding of the model. We use the following scale:

- **- -:** there is no dual coding of the constructs to capture location, these always rely on one single element (either text or a graphic, not both), and the notation is designed in a way which makes it impossible to introduce dual coding
- - : there is no dual coding of the constructs to capture location, and the way the notation is designed it is hard (but not impossible) to introduce it
- " ": there is no dual coding, but the notation is designed such that it can easily be added
- +: there is some dual coding of the location constructs, but not complete or systematic
- ++: dual coding is systematically used for all constructs in the diagram dealing with location

4.7 Graphical economy

Graphical economy advocates that the number of different symbols in the diagrams should be limited - from the motivation that the more symbols a language has, the more demanding it will be for a user to learn and remember these symbols. An interesting question here, of course, is exactly what is counted as different symbols. Obviously, in a UML activity diagram the rounded rectangle shape (activity nodes) and diamond shape (decision nodes) are different symbols. But should a big vs. small, red vs. green, or continuous bordering line vs. dotted bordering line activity nodes likewise be counted as different symbols. And should an activity node with the label "Send receipt" and another with "Receive order" also be counted as different symbols? The naive answer might be yes to all such questions, since there is of course a visual and meaningful difference between all of these. However, given that the motivation for the principle in the first place is the user's challenge of learning and remembering a big number of different symbols, it must be taken into account that these cases have different implications on learning and recall. The activity node and decision node are clearly different symbols. Activity nodes in different icons inside, on the other hand, could be considered a family of symbols (just like traffic signs), and it is easier to learn a bigger number of symbols if they are organized in families with a clear relationship. Coloured (or textured, etc.) activity nodes could be seen as composed symbols, the shape indicates that it is always an activity node, and the colour only modifies it with some more detail (i.e., the location). Similarly, textual labels do not change the meaning of the symbol itself, only provides some more detail. Given that people have already learnt to read, textual labels would be assumed to have the smallest learning burden, while colour or other symbolic distinctions would have to be learnt specifically from the notation. (This would of course be different if a modelling language was supposed to be used by people without reading skills, but that would be rare in IS development projects). Hence, text would be better from the perspective of graphical economy than the introduction of new visual distinctions. All in all, this gives the following scale:

- **- -:** a big number of entirely new and separate symbols are introduced
- - : some new and separate symbols are introduced
- " ": new symbols are introduced but only in a family with already existing symbols in the notation
- +: new visual elements are introduced, but only in a composed manner, slightly modifying the meaning of existing symbols
- ++: no new visual elements are introduced, at most additional text

4.8 Minimal deviation

Minimal deviation from the mainstream notation reflects the wish for adaptations to be minor, i.e., staying as close to the original notation as possible. One motivation for this would be in terms of ease of learning, i.e. employees already having invested a lot of time in learning the original notation, and this effort could be wasted if radically changing the notation. Another motivation is the existence of tool support for the original notation. With minor deviations the tool might still be usable, possible only requiring the change of some user-available settings to allow for the modification, while large modifications could require reprogramming of the tool, which would be time consuming if the source code is available to the organization in question, and impossible if the source code is not available. For minimal deviation we envision the following scale:

- **- -:** the notation is so different from the original that stakeholders will most likely think of it as a new notation, and new tools would have to be developed to support it

- - : the notation is close enough to the original to be thought of as a dialect, but still a radical deviation, for which a substantial learning effort and tool modifications are necessary
- " ": persons who know the original mainstream notation that the adaptation is based on, should be able to use it with approximately half the effort of learning a new notation, and with tool modifications which are, at most, half the effort of making a new tool
- +: persons who know the original mainstream notation that the adaptation is based on, should be able to use it after a small learning effort, and with some minor modifications of existing tools
- ++: persons who know the original mainstream notation that the adaptation is based on, should be able to use the adapted notation right away, with minimal learning effort, and using existing tools with no reimplementation of these

5. Performing the evaluation

Table 1 gives an evaluation of the various alternatives shown in Fig.1, the rows in the table equaling the rows in that figure, and the columns the various principles just discussed (SC = semiotic clarity, PD = perceptual discriminability, ST = semantic transparency, CM = complexity management, VE = visual expressiveness, DC = dual coding, GE = graphical economy, CF = cognitive fit, and finally MD = minimal deviation from mainstream notation). The marks indicate the range from a strong disadvantage (--), through neutral (blank), to a strong advantage (++). Two of the researchers used the criteria independently, then compared their results and performed a consensus process to arrive at the final marks in table 1. The correlation before consensus discussions was 0.79. (Nunally, 1978) (p.245) argues that an evaluation instrument used in basic research should have a reliability of at least 0,7 so the above-mentioned result seems good in this respect, especially since the criteria are not straightforward to interpret or give scores for.

Table 1: Evaluation of alternatives, assuming small adaptation of UML AD notation

Variable	SC	PD	ST	CM	VE	DC	GE	MD	Sum
Text in icon	-	-			- -	-	++	+	-2
Text in note	-	-		- -	- -	+	++	++	-1
Dedicated location shape				- -	- -	++	-		-3
Iconic note	+	++	++	-	-		- -		+1
Icon in icon	+	++	++	+	-	-	-		+3
Shape, small variation	+	-		+	- -		-		-2
Shape, big variation	+	+		+	- -		- -	- -	-3
Size	+	- -	- -	+		-		- -	-5
Orientation	+	- -	- -	+	+	-		- -	-4
Fill color	+	+	-	+	+		+		+4
Fill brightness	+	+	-	+	+		+		+4
Fill texture	+			+	+	-	+		+3
Line thickness	+	- -	- -	+			+	-	-2
Line type	+	- -	- -	+			+	-	-2
Line color	+	- -		+			+	-	0
Line brightness	+	- -	-	+			+	-	-1

The explanation for the grades given in the various columns is provided below:

SC: the two text alternatives (two first lines) get a minus here because of no symbol for location (i.e., deficit), the other alternatives do show it graphically, thus getting a plus (but not a double plus, since it does not fix any other problems in UML activity diagrams). There is a single minus rather than a double minus in the first two rows because symbol deficit is arguably less of a problem than symbol overload (Moody, 2009).

PD: it is deemed most easy to see the difference between icons (the alternatives "iconic note" and "icon in icon"), indicating visually both that it is a location and distinguishing between different locations, these therefore get double plus. The alternatives "shape, big variation" and different fills (color, brightness, texture) also give clear visual distinctions between different locations visually, but one cannot distinguish between as many different locations as with icons. There is a clear limit to how many radically different shapes one can introduce for various locations, especially since some basic shapes (e.g., rectangles, diamonds, circles, triangles, ovals) are already taken for other purposes. With pattern-fills, grey-tones, or color there is also a limit to how many alternatives the user could reliably distinguish between, maybe around 5 for patterns and grey-tones (incl. black and white) and 7-9 for color as long as users do not have reduced color vision. Thus, these alternatives only get a single plus. The alternative "dedicated location shape" clearly shows that something is a location, but does not distinguish visually between different locations, this therefore requires closer inspection of the text inside the node, hence a blank. The alternatives "text in icon" and "text in note" similarly require closer inspection of the text, but these have the additional disadvantage of failing to signal that it is a location in the first place, hence a minus. Similarly, the alternative "shape, small variation" requires closer inspection, thus getting a minus. The four last rows with variations of the bordering line of the activity node, are deemed to give very poor perceptual discriminability. Although these employ similar means as the three alternatives using fills, the visual distinctions become much more subtle when only the bordering line of the node is used, hence a double minus. Similarly, "size" or "orientation" are also deemed hard to distinguish, thus getting double minus. To allow for reliable distinction between different node sizes, the sizes would have to be clearly different, so the smallest would have to be quite tiny, making it hard to read the text label inside. For orientation, the user might distinguish easily between horizontal, vertical and 45 degrees diagonal nodes, but not much more, and again text becomes hard to read in the rotated nodes.

ST: The alternatives "iconic note" and "icon in icon" are the only ones with a clear advantage here - as long as one is able to find icons that intuitively reflect the locations in question. These therefore get a double plus. The three uppermost lines will mention the location as text, slightly less intuitive but still making it possible for the user to know the location directly from the node rather than looking up a legend or remembering something more abstract. These three alternatives therefore get a single plus. Most other alternatives have only an opaque relationship with the locations indicated, therefore getting a blank. The alternatives using size, orientation, line type or thickness would go as semantically perverse in Moody's terminology, therefore getting a double minus, i.e., they would intuitively be thought to represent something else that location. Size could intuitively be thought to represent the size, complexity, frequency or importance of an activity, not location. The same could go for line thickness or type, e.g., thicker line or full line meaning a more prominent activity than thin or dotted line. Orientation could intuitively be thought to represent direction, increase or decrease, but not easily location. Color and brightness get a single minus because, although not generally perverse, they could be so in some situations. E.g., colors red, yellow, and green could for instance be believed to be associated with risk (e.g., red = high risk, yellow = medium, green = low risk / ok), similarly black / inverted nodes could intuitively be thought to represent something negative or forbidden, cf. misuse cases (Sindre & Opdahl, 2005) and mal-activity diagrams (Sindre, 2007)) − i.e., these visual variables could have other interpretations which are more natural than location.

CM: For complexity management the major disadvantage is with notation alternatives which require the introduction of new nodes to hold locations, and therefore also extra links to connect these locations to activities, thus increasing the structural complexity of the diagram. "Text in note" and "Dedicated location shape" both introduce an extra node and text label, plus links to this node, thus getting a double minus. "Iconic note" also introduces an extra node and links but at least avoids the extra text label, thus getting a single minus. "Text in node" does not create any new node, but the text labels become substantially longer when locations must be included, hence getting a blank. The other alternatives, which neither cause new nodes or longer texts, are all given a single plus. They have advantages vs. the notations that got negative scores, but to deserve a double plus there should also have been a possibility for information hiding by composing and decomposing locations, which none of them offer.

VE: According to Moody's explanation of this principle, positive scores should here be given to notations which employ visual variables not otherwise used in UML, i.e., orientation, color, brightness, and fill texture get a plus (but not a double plus, since each notation alternative introduces only the usage of one new variable). Size does not have a meaning in UML, but diagrams often contain nodes in different sizes, for instance just because of variations in the length of text labels to put inside - hence this only gets a blank, and the same goes for line thickness and brightness, as for instance bigger nodes may also have been equipped with thicker lines in many UML examples, although it did not have a particular meaning. Line type and color are also classified as blank; line color could be said to introduce a new variable (color) but only very subtly so. Those visual means that are already used in UML and having a meaning get a double minus here, i.e. alternatives using text or shape, since these do not enhance the visual expressiveness of the notation. Icons could be seen as advanced shapes but are somewhat different from the typical shapes of UML, thus getting away with a single minus.

DC: The alternative that best combines graphics and text is the dedicated location shape (having an own symbol for locations, and explaining in text what the location is), this therefore gets a double plus. The "text in note" alternative also uses text, but the graphic symbol could also be something else than a location, this therefore gets a single plus. Most of the remaining alternatives do not use text as shown in Fig. 2, but it would be easy to add a textual explanation of the location if preferable. These therefore get a blank. Some alternatives get a minus because addition of text would be more difficult: "Icon in icon" because there is less space left for extra text due to the inserted icon, and size similarly because the smaller nodes would hardly have place for additional text. For "orientation" added text in rotated nodes would be difficult to read, and similarly, "fill texture" also gets a minus because text is often more difficult to read over a pattern fill, especially if there is more text so that a smaller font must be used. The alternative "text in icon" can be said to use only text for the location, no graphic at all, so this also gets a minus.

GE: From the point of view of graphical economy, new symbols are considered a disadvantage. Hence the most positive alternatives (getting a double plus) are those that introduce no new symbols, i.e., the two uppermost rows. The alternatives using fills or modifications of the bordering line are also considered good (a single plus) since these do not introduce new symbols, only variations on existing symbols, using an orthogonal visual variable. Size is a little more problematic, although a big and a small activity node could be said to be the same symbol, some users may also feel they are different, cf. the difference between big diamonds (used for decision points in UML activity diagrams) and smaller diamonds (used for aggregation / composition in UML class diagrams), this therefore gets a blank. The same goes for orientation, although it is the same symbol only rotated, these will easily appear as different symbols for many users. The "dedicated location shape" gets a single minus, as it creates one entirely new symbol in the notation, while the four rows below that one (two with icons and two with new shapes) get double minuses, since each causes the introduction of a bigger number of new symbols, one per location type that must be distinguished.

MD: The minimal deviation from standard UML notation is to have no deviation at all, e.g., the "text in note" alternative, since UML notes already exist in the notation and can be used for anything the user likes, including the description of locations. "Text in icon" is also a no-deviation alternative when it comes to the diagram notation as such, but it does break the normal naming convention for activity nodes (verb+object), instead requiring longer names with verb + object + location phrase. Hence, this gets only a single plus. Using node fills (color, brightness, texture) is deemed a fairly small deviation because it uses a visual variable not used before, otherwise leaving the notation the same, these therefore get a blank. The same goes for dedicated location shapes and icons; they represent an addition to the notation, but does not mess up the interpretation of the existing notation in any way. Variations on the bordering line is seen as a slightly bigger deviation. Although not used for anything else in UML activity diagrams, variations of line types are for instance used in UML class diagrams, and a change to use a notational trick already used for one purpose for another purpose, too (or instead) is a bigger leap than applying a trick not used before. Hence, these get a single minus. Finally, the alternatives "shape, big variation", "size", and "orientation" represent the biggest deviations from mainstream notation, thus getting double minus here. Using several widely different shapes for activities in various locations would be a big leap from a notation where all activities have the same shape. Size, as mentioned before, has been used accidentally in UML for instance because of diagram space concerns or needs to fit long textual labels, changing this to having meaning would be quite different. Using orientation of nodes would also make diagrams look very different from before.

The rightmost column provides the summed scores from the grades given. All in all, the most positive candidates from the analytical evaluation are those using fill in the activity nodes, either in the form of color, grey-tones ("brightness"), or texture, tightly followed by the "icon in icon" alternative. These have advantages over alternatives using changes in size, orientation, or the bordering line when it comes to perceptual discriminability, and over alternatives using text or separate nodes for locations (first four rows) when it comes to complexity management and/or visual expressiveness. Also, they do reasonably well when it comes to small deviation from the standard notation since the basic shapes remain unaltered, and fills in activity nodes are not used for any other purposes in the original notation (whereas different line types, for instance, may already be used in some cases in UML), thus not having any big drawbacks. An assumption underlying this, however, is that the number of locations that must be distinguished in a diagram is not too big. In the examples we looked at, there were up to 5 different locations to be differentiated between in the same diagram, and in many work processes there would not be many more different locations (places) than this (e.g., in office, at the client's site, or travelling in between). If the number becomes bigger, patterns or grey-scales would soon have trouble with reliability. Color would go a little longer (except for users with reduced color vision), but would also have trouble if there are 10 or more different locations to distinguish. In such cases, the alternative "icon in icon", might be better, since the users would be able to distinguish reliably between a much larger number of icons, at least as long as one is able to find intuitive locations, cf. traffic signs, where only a limited number of shapes and colors are used, but with a large number of different icons that could be inside these shapes, e.g. to indicate different types of dangers. Another assumption behind the sums of scores, of course, is that all our evaluation criteria are equally important. If instead arguing that the three first criteria (semiotic clarity, perceptual discriminability, semantic transparency) are more important than the other ones, the "icon in icon" alternative would again come out better because it has a double plus for each of these.

6. Discussion and Conclusion

Since black/white has already been used for misuse case diagrams (Sindre & Opdahl, 2005) and related mal-activity diagrams (Sindre, 2007), we chose color and pattern-fills as our first try, rather than grey-tones. To have a baseline to compare with, we picked the alternative with the smallest possible deviation from standard UML, namely providing the location in a note (cf. ++ for the "text in note" alternative in the MD column), since this only uses symbols already available in UML. Hence, our experiment plan was as follows:

- first compare the "Fill color" alternative with the "Text in note" alternative
- if "Fill color" showed advantage, compare this with the "Fill texture" alternative

A natural next step, according to the argument above, would be to make an experimental comparison also with the "icon in icon" alternative, but this has not been done so far and is therefore a topic for further work. The experiments mentioned in the two bullet items will now be presented in the next sections, starting with the experiment design, and then the results and analysis.

It can be noted that sometimes variables might be used in diagrams without really being officially part of the language definition. E.g., colour is not defined to have any meaning in UML, and its usage may be discouraged due to accessibility concerns for people with impaired colour vision. Still, it is easy to find on the web examples of UML diagrams that contain colour, either for purely aesthetic purposes, or for emphasizing some nodes. Similarly, there are many graph notations where horizontal or vertical placement does not have any meaning, but where one is free to choose the layout which appears most suitable, for instance to reduce the number of crossing lines or satisfy other aesthetical criteria for graphs. A number of such aesthetical criteria are defined in (Tamassia, et al., 1988) and (Sindre, et al., 1993) discuss some further extension for decomposed graphs.

Although a language officially does not assign any meaning to a certain visual variable, it is of course still possible that people who interpret diagrams may associate some meaning with this variable if it is somehow used in the diagram. Even for such "ad hoc" variables it is therefore important that they are not used in a misleading manner.

An important question not addressed in any detail by (Moody, 2009) and neither in this report is whether the 9 principles should bear equal weight in an evaluation of a notation, or whether some are more important than others. Empirical investigations will be important also to cast light on this question, checking if an observed advantage according to one of the principles also implies an advantage in practical usage of a notation. But since there are many plausible notation alternatives and many principles with possible interaction between them, a huge mass of experiments may be needed to get a clear picture.

References

Bertin, J. (1983). *Semiology of Graphics: Diagrams, Networks, Maps*. Madison, WI: University of Wisconsin Press.

Deiters, W., Löffeler, T., & Pfenningschmidt, S. (2003). The information logistics approach toward a user demand-driven information supply. In D. Spinellis (Ed.), *Cross-Media Service Delivery* (pp. 37-47). Dordrecht: Kluwer.

Dumas, M., & ter Hofstede, A. (2001). UML Activity Diagrams as a Workflow Specification Language. In M. Gogolla & C. Kobryn (Eds.), *«UML» 2001 — The Unified Modeling Language. Modeling Languages, Concepts, and Tools* (Vol. 2185, pp. 76-90): Springer Berlin / Heidelberg.

Krogstie, J. (2001). Using a Semiotic Framework to Evaluate UML for the Development of Models of High Quality. In K. Siau & T. Halpin (Eds.), *Unified Modeling Language: Systems Analysis, Design, and Development Issues* (pp. 89-106): IDEA Group Publishing.

Krogstie, J. (2003). Evaluating UML Using a Generic Quality Framework. In L. Favre (Ed.), *UML and the Unified Process* (pp. 1-22). Hershey, PA: Idea Group.

Krogstie, J., & Arnesen, S. d. F. (2004). Assessing Enterprise Modelling Languages using a Generic Quality Framework. In J. Krogstie, K. Siau & T. Halpin (Eds.), *Information Modelling Methods and Methodologies* (pp. 63-79). Hershey, PA: Idea Group.

Krogstie, J., Lindland, O. I., & Sindre, G. (1995, March 28-30). *Defining Quality Aspects for Conceptual Models.* Paper presented at the Proceedings of the IFIP 8.1 working conference on Information System Concepts (ISCO3): Towards a consolidation of views, Marburg, Germany.

Lindland, O. I., Sindre, G., & Sølvberg, A. (1994). Understanding Quality In Conceptual Modelling. *IEEE Software, 11*(2), 42-49.

Moody, D. L. (2009). The "Physics" of Notations: Toward a Scientific Basis for Constructing Visual Notations in Software Engineering. *IEEE Transactions on Software Engineering, 35*, 756-779.

Nunally, J. (1978). *Psychometric Theory (2nd edition)*. New York: McGraw Hill.

Nysetvold, A. G., & Krogstie, J. (2006). Assessing Business Process Modelling Languages Using a Generic Quality Framework. In K. Siau (Ed.), *Advanced Topics in Database Research* (Vol. 5, pp. 79-93). Hershey, PA: Idea Group.

Opdahl, A. L., & Henderson-Sellers, B. (2002). Ontological Evaluation of the UML Using the Bunge-Wand-Weber Model. *Software and Systems Modelling, 1*(1), 43-67.

Sindre, G. (1990). *Hicons: A General Diagrammatic Framework for Hierarchical Modelling.* Unpublished PhD thesis, Norwegian Institute of Technology (NTH), Trondheim, Norway.

Sindre, G. (2007, 11-12 June). *Mal-Activity Diagrams for Capturing Attacks on Business Processes.* Paper presented at the International Working Conference on Requirements Engineering: Foundations for Software Quality (REFSQ'07), Trondheim, Norway.

Sindre, G., Gulla, B., & Jokstad, H. G. (1993). *Onion graphs: Aesthetics and layout*. Paper presented at the IEEE Workshop on Visual Languages.

Sindre, G., & Opdahl, A. L. (2005). Eliciting Security Requirements with Misuse Cases. *Requirements Engineering, 10*(1), 34-44.

Tamassia, R., Di Battista, G., & Batini, C. (1988). Automatic graph drawing and readability of diagrams. *Systems, Man and Cybernetics, IEEE Transactions on, 18*(1), 61-79.

Walderhaug, S., Stav, E., & Mikalsen, M. (2008). *Experiences from Model-Driven Development of Homecare Services: UML Profiles and Domain Models*. Paper presented at the Model-Based Design of Trustworthy Health Information Systems (MOTHIS), workshop at MODELS'08.

Wand, Y., & Weber, R. A. (1990). An Ontological Model of an Information System. *IEEE Transactions on Software Engineering*, 1282-1292.

Wand, Y., & Weber, R. A. (1993). On the Ontological Expressiveness of Information Systems Analysis and Design Grammars. *Journal of Information Systems*.

Wohed, P., van der Aalst, W., Dumas, M., ter Hofstede, A., & Russell, N. (2006). On the Suitability of BPMN for Business Process Modelling. In S. Dustdar, J. Fiadeiro & A. Sheth (Eds.), *Business Process Management* (Vol. 4102, pp. 161-176): Springer Berlin / Heidelberg.

Zachman, J. A. (1987). A framework for information systems architecture. *IBM Systems Journal, 26*(3), 276-291.

A. More diagram examples

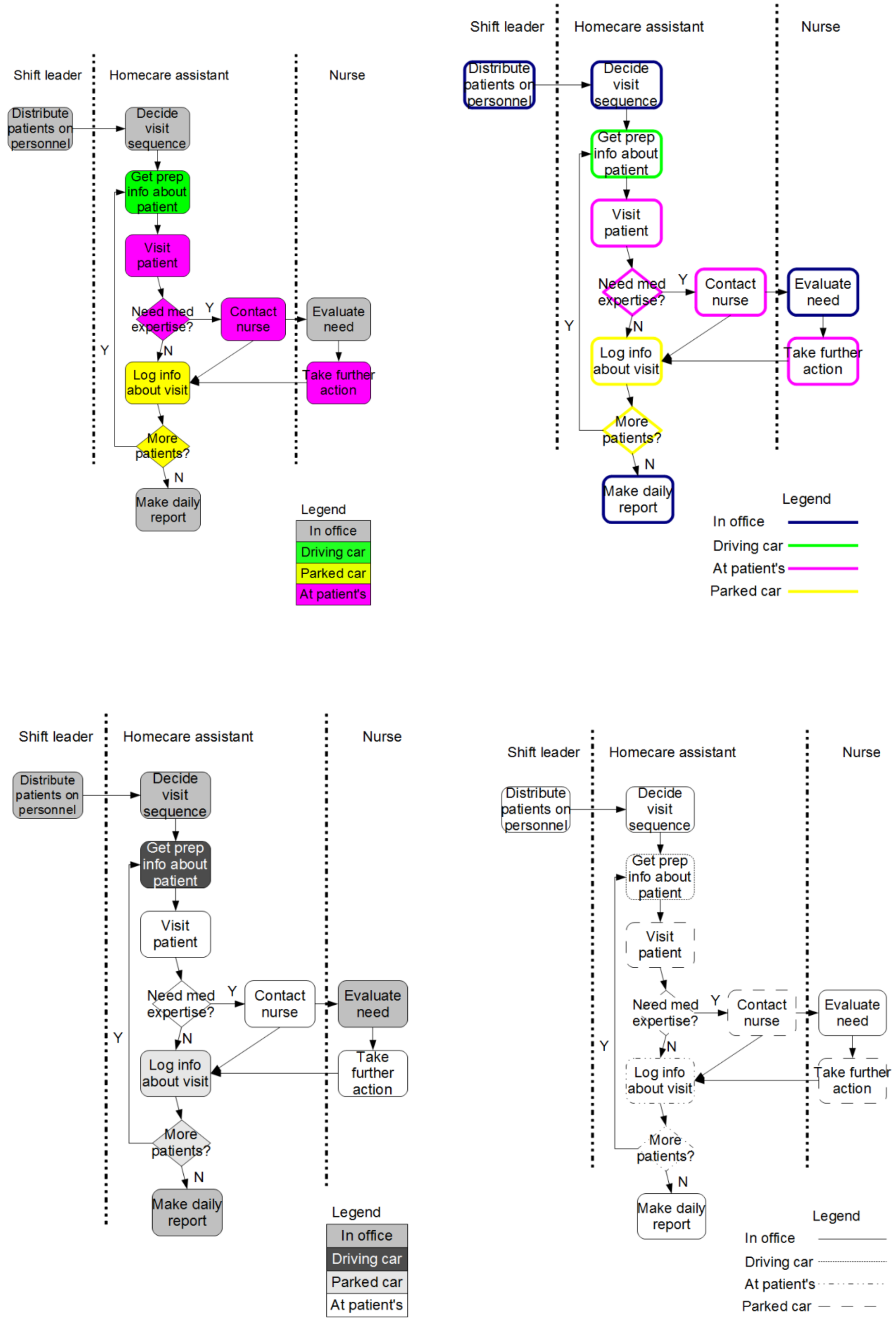

Figure A1: Showing location with node fill colour (upper left), line colour (upper right), node fill value (greytone, lower left) or line type (lower right)

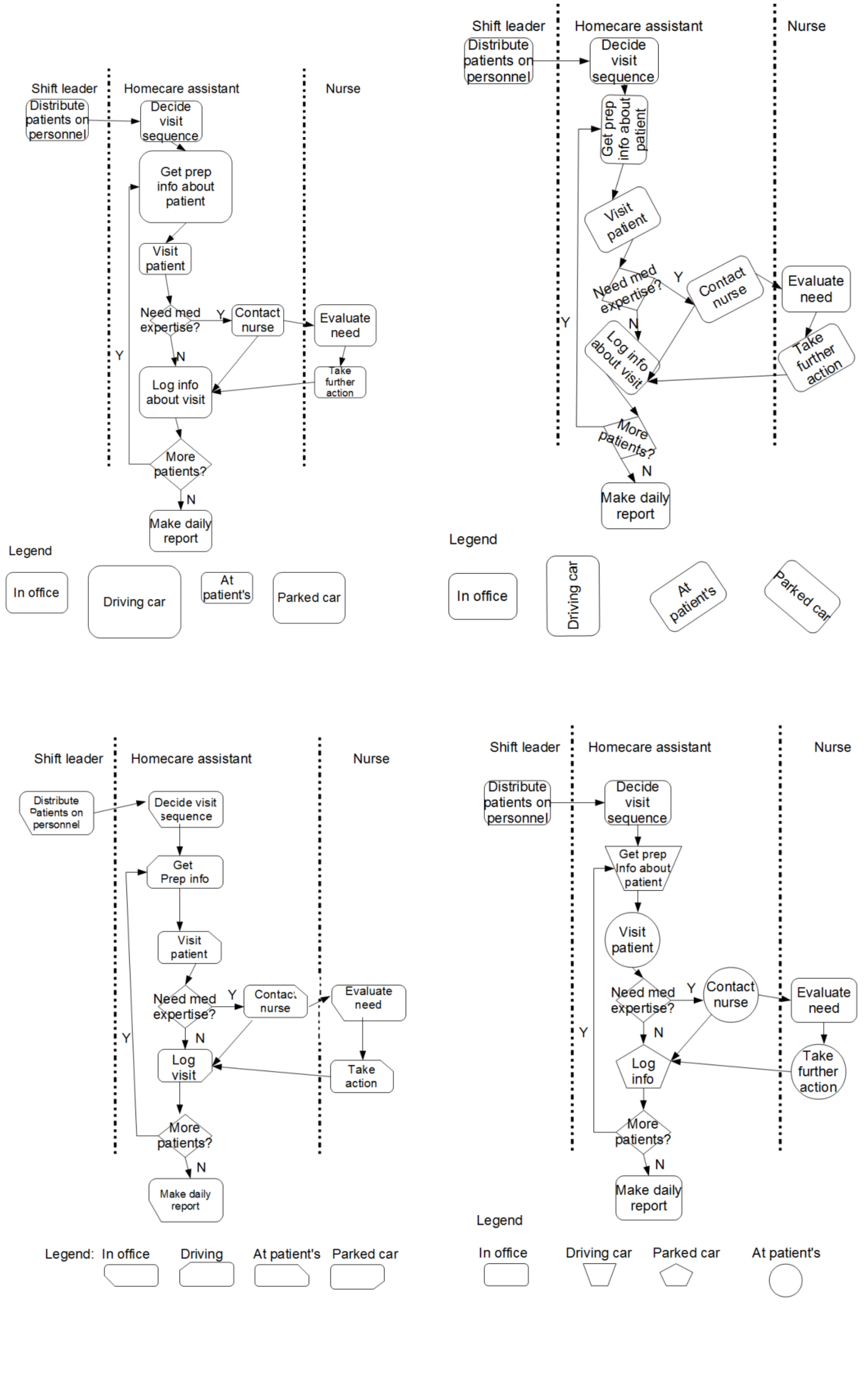

Figure A2: Showing location by node size (upper left), node orientation (upper right), node shape, small variations (greytone, lower left) or big variations (lower right)