Structuring Qualitative Data for Agent-Based Modelling
Ghorbani, Amineh; Dijkema, Gerard P.J.; Schrauwen, Noortje

Published in:
JASSS - The Journal of Artificial Societies and Social Simulation

DOI:
10.18564/jasss.2573

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 11-09-2020
Structuring Qualitative Data for Agent-Based Modelling

Using ethnography to build agent-based models may result in more empirically grounded simulations. Our study on innovation practice and culture in the Westland horticulture sector served to explore what information and data from ethnographic analysis could be used in models and how, MAIA, a framework for agent-based model development of social systems, is our starting point for structuring and translating said knowledge into a model. The data that was collected through an ethnographic process served as input to the agent-based model. We also used the theoretical analysis performed on the data to define outcome variables for the simulation. We conclude by proposing an initial methodology that describes the use of ethnography in modelling.

Keywords: ethnography, institutional analysis, survey, qualitative data, MAIA, conceptual modelling

Introduction

1.1 Building empirically-grounded artificial societies of agents requires qualitative and quantitative data to inform individual behaviour and reasoning, and document macro level emerging patterns (Robinson et al. 2007). While quantitative data can be collected through surveys, literature and other available sources, gathering qualitative data to design the behaviour of the agents, their decision making processes and their forms of interaction is not a straightforward task (Janssen & Ostrom 2006). Likewise, macro-level data for model validation requires theoretical analysis about the system that is being modelled (Robinson et al. 2007).

1.2 Modellers commonly use behavioural and social theories, and desk research to cover the qualitative aspects of agent-based models. They may also use surveys and statistical analysis to understand the decision making behaviour of individuals (Sanchiz & Lucas 2002; Dia 2002).

1.3 One field of research that can also be used to collect data for agent-based models is ethnography (Bhawan 2004). Ethnography is a research method covering many approaches in anthropology. The data is gathered through interviews and field surveys which are then coded for theoretical analysis. The collected data is a rich set for understanding human behaviour and interaction which is also a good source to build artificial humans or agents. Furthermore, the theoretical analysis that is performed on ethnographic data could be a good source of macro-level data for model validation by observing whether the same mechanism and patterns concluded from the analysis result from the simulation (Robinson et al. 2007).

1.4 Since ethnography provides a rich set of data about the system and its entities, we anticipate it can be used to make richer agent-based models populating them with empirically grounded data. However, this data, although coded for theoretical analysis, is difficult to interpret and decompose in order to build agents and their behavioural rules. Ethnographic data is normally in textual format obtained from interviews, fieldwork, participant observation, and documents (Yang & Gilber 2008).

1.5 The difficulty in making use of ethnographic information for agent-based modelling and simulation (ABMS) is due to the fact, that in qualitative ethnographic research the interviewees are normally allowed to talk about their concerns in an open manner, which may lead to an overload of information that may also be immensely rich and diverse in terms of content. In addition, the researcher and the interviewees each have their own world-view, which leads to bias, as abstraction and generalization is required to arrive at specifications of behaviour and characteristics suitable for building agent-based models.

1.6 The most complete research in the intersection between ABMS and Ethnography is Bhawan (2004). Bhawan (2004) provides a detailed procedure for the fieldwork process which describes how ethnographic data is collected and formalized. Bhawan (2004) used knowledge engineering techniques in the process, allowing a continued engagement with the interviewees. She designed a specific ontology (i.e., architecture) for her particular domain namely, Agro-Climatic systems, to decompose the ethnographic information into a model. Yang and Gilbert (2008) discuss the differences and similarities between ethnographic data and ABMS and propose recommendations for modellers when using ethnographic data. They emphasize on the requirement for computer-aided qualitative analysis to manage and structure the data. Another requirement indicated by them is a model of data to represent relationships among actors (Yang & Gilber 2008).

1.7 There are also case specific examples of using qualitative data in agent-based models. Geller and Moss (2008) present a model of solidarity networks in Afghanistan, informing agents’ structures, behaviour and cognition by qualitative data. They use an evidence-based approach following rules according to which agents’ behaviours are drawn directly from empirical studies. Moore et al. (2009) use a combination of ethnography and ABMS to study psychodrillent use and related harms. They also indicate the difficulty in generalizing ethnographic information to build agent-based models. They built a model called SimAmph and used the ontology to build agents. They also noted it is very useful to project the two domains as well as in facilitating collaborative model development and analysis.

1.8 Thus, from the literature, it appears that a shared ontology or a conceptual framework is one of the main requirements for generalizing and structuring qualitative information, especially ethnographic data for ABMS. To address this requirement, in this research, we use an ABMS framework called MAIA (Othman et al. 2013) which provides a shared ontology for social systems, covering a diversity of social, institutional, physical and operational concepts that are required for building agent-based models. Using MAIA as a template of required concepts may help collect and structure ethnographic data for building agent-based models. Therefore, in this research, we explore this possibility by using this modelling framework to structure ethnographic data collected from interviews, fieldwork and formal documents to build an agent-based model. To underpin this possibility, we use a case study on innovation practices in the Dutch horticulture sector.

1.9 The remainder of this paper is as follows. In Section 2, we give a brief overview on ethnography and introduce the MAIA framework. In Section 3, we introduce the horticultura case study. In Section 4, we explain the methodological process of integrating ethnographic processes into ABMS. In Section 5, we discuss the lesson learnt from this process and analyse our methodological process. Finally, we conclude in Section 6.

Background

2.1 The goal of this research is to propose a methodology for using ethnography to build agent-based models. In this section, we will first explain ethnography. Then, we will introduce the MAIA framework, which will be used as the tool for this methodological process.

Ethnography

2.2 Ethnography is a field of science that spans many methods and schools of approaches in anthropology. The power of ethnographic research is that real people are studied at the level of small communities/groups or individuals, and at the societal level, while the mutual interaction is also considered. This qualitative research aims to address complex phenomena by analysing and interpreting the system from the participants’ point of view. Ethnography is often exploratory in nature, using observations to construct the analysis from “bottom-up”. Together, this appears to be what is needed for developing agent-based models, in order to characterize the intellectual of the individual and the system:

Ethnographic research can range from a realist perspective in which behaviour is observed to a constructivist perspective where understanding is socially constructed by the researcher and subjects. Research can range from an objectivist account of fixed, observable behaviours to an interpretivist narrative describing "the interplay of individual agency and social structure." Critical theory research addresses "issues of power within the researcher–researched relationships and the links between knowledge and power" (Ythema et al. 2010).

2.3 In ethnography there are several types of methodologies, which can broadly be categorized as either inductive or deductive. An inductive approach to ethnography formulates theories from the "bottom-up" rather than from the "top-down". This means that the researcher starts by observing the community and by looking for repeated patterns of behaviour. If certain themes continue to appear, the researcher can develop a tentative hypothesis that is then verified and which may be turned into a theory. This may require the collection of more corroborating data from other communities within the same social system. Grounded theory is an inductive method of analysis commonly applied in ethnography to help scientists generate theories (Corbin & Strauss 2008). Unlike other theories, grounded theory does not start by hypotheses for social behaviour but concludes with them. The grounded theory approach is an iterative process where the analysis of the data may raise new questions that stimulate new data collection (Nunnan 2014). While this describes inductive research, some anthropologists also take the deductive approach, using predefined questionnaires, hypothesis, quantitative data and statistics etc.

2.4 The inductive approach is more flexible, however, when it comes to addressing human societies, as it helps the researchers let go of their preconceived (and often culturally biased) ideas of what the society they are studying is like. While the inductive approach is still used in cultural anthropology today, currently this theory has shifted from "start fieldwork and wait for answers" to "start field work with a few general questions to answer". This would provide enough frameworks to focus the research, but would leave the questions general enough to allow for the flexibility that studying human culture needs. Some methods play a central role in this inductive approach.
The interviews are recorded and coded in Atlas.ti for later analysis.
Using MAIA for field observation

4.5 During field observation, it is important to identify the relevant properties of the entities (i.e., agents and physical components) that are addressed during the interviews. The composition of the physical entities and their connections may be observed in the field and defined as physical components in the physical structure of MAIA. Thus, in a fashion similar to setting up the general structure for the semi-structured interviews, the MAIA structures can be used as a template for collecting data during field observation.

Using MAIA for studying formal documents

4.6 The formal documents are collected according to the information provided by the subjects. To collect the right information for modeling institutions, the ADICO structure (see Section Background) is used as the template.

Building a MAIA model

4.7 Upon completion of the previous steps, the collected data is used to build an agent-based model. This process is conducted by extracting relevant information from the data by using the MAIA framework. Again, we look at the structures one-by-one to clarify the process [47].

Collective Structure

4.8 The interviewed subjects can be defined as agent-types. Each subject can be defined as one separate agent-type. If the simulation is limited to the people interviewed, alternatively, one may group the agents according to some criterion and use each category to define a separate agent-type. In the greenhouse case, the 15 growers that were interviewed were divided into five categories distinguished by their stated priorities, their physical assets and characteristics. The first category is the niche growers whose greenhouse is relatively small in size and whose innovation activities are mainly marketing- and product-oriented. The other four categories are large bulk growers, the innovative bulk growers, moderate bulk growers and shop growers (see Schrauwen 2012).

4.9 Agents in the simulation are not limited to the interviewees; there may also be social entities that were addressed during the interviews. For example, the European Union was a social entity addressed by the growers, who influences their innovation strategies. This entity is, therefore, also defined as an agent in the simulation.

4.10 From the qualitative data, whether in the form of field observation or interview, the properties, personal values, intrinsic behaviours and decision-making of the actors are extracted to build the agents in the model.

Constitutional Structure

4.11 The main aspect of the constitutional structure is the institutions. These can be formal institutions extracted from legal documents, or informal institutions, namely, norms of behaviour and shared strategies extracted from the interviews or field observations. The patterns of behaviour observed from interviews can be the result of rules imposed by the society. These are defined as norms or shared strategies. If the rule of behaviour contains an obligation or prohibition by definition, the rule is considered to be a norm. If the actors perform the same routine without any obligation from the system, that routine can be considered as a shared strategy. All the formal and informal institutions are modelled as ADICO statements as defined in Section Background. Table 1 shows some of the institutions extracted from the interviews and legal documents.

<table>
<thead>
<tr>
<th>#</th>
<th>A</th>
<th>D</th>
<th>C</th>
<th>O</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Growers</td>
<td>may</td>
<td>get a subsidy from the EU for max 50% of an investment</td>
<td>if they invest in an accepted innovation and follow the rules 4</td>
<td>Type</td>
</tr>
<tr>
<td>2</td>
<td>Growers</td>
<td>must</td>
<td>join one of the 6 sales cooperations</td>
<td>if they want to get the GMO subsidy from the EU</td>
<td>rule</td>
</tr>
<tr>
<td>3</td>
<td>Growers</td>
<td>may</td>
<td>market under their own brand</td>
<td>if they want to get the GMO subsidy from the EU</td>
<td>rule</td>
</tr>
<tr>
<td>4</td>
<td>EU</td>
<td>may</td>
<td>fine the grower</td>
<td>if growers don’t follow the rules attached to the subsidy</td>
<td>rule</td>
</tr>
<tr>
<td>5</td>
<td>EU</td>
<td>not</td>
<td>increase the interest on loans</td>
<td>when growers are a less optimal financial situation</td>
<td>rule</td>
</tr>
<tr>
<td>6</td>
<td>Growers</td>
<td>copy the successful innovations of their colleagues</td>
<td>if the colleague is more successful</td>
<td>norm</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Growers</td>
<td>cooperate together with other growers</td>
<td>if performing similar practice</td>
<td>shared strategy</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Growers</td>
<td>adopt an innovation</td>
<td>when it has shown to be working at other greenhouses</td>
<td>shared strategy</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Growers</td>
<td>invest and modernize in the organization</td>
<td>if there is a successor</td>
<td>shared strategy</td>
<td></td>
</tr>
</tbody>
</table>

Physical Structure

4.12 Similar to building agents, the physical entities that are addressed by the interviewees are extracted from the text and defined as physical components in the MAIA model. These include energy, greenhouse and machinery (i.e., the innovative technology they adopt). The properties of these components are identified through field observation in addition to interviews. For example, during field work, it became clear that two properties, namely, the size of the greenhouses and their type of crops, mainly distinguish growers from each other.

Operational Structure

4.13 The events that were described by the interviewees are defined as actions in MAIA. The condition for performing those actions and the outcomes of the actions is to be extracted from the descriptions and to finalize the model. The modeller has to make a decision about the time loop and the actions that take place per tick. For this study, we decided that in each tick, seven action situations take place according to the following sequence:

- Daily life: In this action situation, the intrinsic capabilities of actors take place: being born, die, have a child, learn and start relationships.
- Cooperating: Within the action situation of cooperating, growers can group together and make a joint decision on investments in innovations. Also, knowledge, norms and values are shared amongst growers that are cooperating, adding up to the social capital of the growers.
- GMO: In this action situation, growers request GMO (Gezamenlijke Markt Ordening - collective market structuration) subsidy when they may recover half of the investments. GMO applications can either be accepted or rejected. Previous subsidy receivers may also be punished in this action situation, based on their previous actions.
- Loan: In this action situation, the grower can apply for a loan. He has to pay back his loan and report his money level to the bank, who may take over, when the grower is in trouble.
- Innovating: In the innovation situation, the decisions are made by the growers to invest in one of the categories of innovations. They invest their money in that innovation, while adopting a new physical component (i.e., technology) in their greenhouse with specific characteristics.
- Cultivation: In the cultivation situation, all horticulture-related activities are performed such as cultivation, employing technologies, and increasing efficiency. The investments of the previous round of innovations affect the cultivation process and produce outcomes, in terms of products, efficiency, use of inputs, etcetera. Also, the money level is checked and reported to the bank (if the grower is a member).
- Selling: In the selling situation, growers calculate the costs and value of their products and calculate a market price. They sell their products to the merchandisers. Products are exchanged with money.

Evaluative Structure

4.14 To build the evaluative structure of MAIA, not only the data collected was used, but also the anthropological analysis. We defined a set of variables that can be used to measure and study the possible emergent system elements from the simulation according to this analysis.

4.15 The theoretical analysis showed that a phenomenon called ‘isomorphism’ steers companies towards the same characteristics which gives rise to similar innovation practices that are not effective in the long run and may even harm the sector. To explore this phenomenon in the simulation, we defined the variable ‘homogenization’ to calculate the variation in innovation types. This value would be measured through time. The correlation between subsidies and this variable is also identified as a parameter of interest according to the ethnographic analysis.

4.16 One other issue in the analysis was ‘decreasing product value’. Many products, especially bulk products, are sold with little margin. This means that the income flowing back to the grower is at risk of being less than cost, which decreases their capital. With just one innovation not giving good returns, this may put them in danger. This may even cause bankruptcy. Therefore, another variable to keep track of in the simulation is the developments of product value (i.e., product price) in relation to time and different innovation types.

4.17 The sector’s sustainability is another point of interest in the study. This issue stands on three different pillars, namely, economical, ecological and social. To experiment with these pillars in the simulation, for the economical part, the ratio between product value and bankruptcy is calculated in relation to subsidies, loans and time. For the ecological aspect, the relation between water, energy and nutrient, and amount and value of products is defined as a metric. Finally, to track the social influence, we define two variables: social capital and bankruptcy.

4.18 In this section, we presented an overview of the process of ethnographic data collection and analysis used for conceptualizing an agent-based model of the horticulture sector. We explained how MAIA concepts can be used to inform data collection, and to build an agent-based model. In the next section, we will generalize this methodological procedure, to make it applicable to other social studies.

Generalizing the Process

Table 1: Some of the identified institutions in greenhouse case study.

http://jasss.soc.surrey.ac.uk/18/1/2.html
4.19 Figure 1 shows the general process of using ethnographic data to build an agent-based model using MAIA. Some concepts in the MAIA structures, as illustrated on the left side of the figure, are primarily used to semi-structure the data collection process. The collected data is then decomposed into an agent-based model, again, using the MAIA structures.

4.20 As Figure 1 shows, there is a cycle between the ethnographic research and the building of a MAIA model. Although semi-structuring data collection minimizes the need to redo interviews, it may still be required to collect further information for the model. This would especially hold for field observations and document collection.

4.21 Besides building the conceptual model, the ethnographic data is also used to perform theoretical analysis. Not only can this analysis be used to further enrich the model, specifically in the evaluative structure (see previous section), it is also used to draw conclusions. These conclusions can be used independently or in combination with the simulation results. Some sort of triangulation can thus be completed, comparing the social analysis with the dynamics generated by running the model. What may be an issue here, however, is that the same input data is used for both methods, so they are not completely independent.

Discussion

5.1 Building an agent-based model requires both quantitative and qualitative data. Although much of the information can be represented in the form of numeric values, the actual context of the model which shows the order of the events, and how agents make decisions and interact, requires ethnographic information. Ethnography can provide rich data for building agent-based models both at micro and macro levels. However, it needs structure and interpretation to be actually applicable to this simulation approach (Yang & Gilbert, 2008). In this paper we presented MAIA as a tool to collect and structure ethnographic data for ABMS. The process of building an agent-based model for the horticulture sector helped us to identify several benefits of using this tool.

5.2 First, the MAIA framework ensures consistency and coherence between the features extracted from the ethnographic process. Since MAIA is constructed as software meta-model, its soundness, completeness and parsimony have been verified (Ghorbani, 2013). Therefore, the modeller can be confident that the collected and structured data is by default consistent in the model.

5.3 Second, as Dey (2000) indicates, analyzing qualitative data also involves an abstraction process which may not be a straightforward task given the immense amount of data provided by ethnography which mostly concerns individuals. Since MAIA is an abstract template or ‘ontology’ for a set of concepts, it proved to be highly instrumental for facilitating and documenting this abstraction process.

5.4 Third, another contribution of MAIA in making use of ethnographic data is that it helps to identify the normative aspects of the system. The insights people provide about their view of the world through interviews are not based on external reality but are culturally generated and emergent. With the ADICO statements in MAIA, the modeller can extract the norms and shared strategies from the interviews in order to add a cultural/institutional dimension to the simulation.

5.5 Fourth, an important contribution of using MAIA is that not only the collected ethnographic data can be used to build an agent-based model, the theoretical analysis performed on the data is also put to use. The theoretical ethnographic analysis helps define the variables that measure the outcomes of the simulation. These variables are covered in the evaluative structure of MAIA. Therefore, besides informing agent behaviour, the methodological process introduced in this paper can help measure the possible outcomes of interest, i.e., macro-level patterns for the simulation.

5.6 Fifth, when an ethnographic researcher uses MAIA, her activities become more structured and tractable. We anticipate this will facilitate the interpretation and discussion of field research, and lead to a growing body of empirically grounded information that can be re-used for modelling and research studies.

5.7 Finally, linking the body-of-knowledge of anthropology and agent-based modelling of social systems may be mutually beneficial. We believe, the proposed method supports non-computing anthropologists in building agent-based models in order to complement their research methods. To explore the feasibility of this claim, an anthropologist performed the whole process starting from ethnographic fieldwork to the development of the conceptual model. We observed that MAIA can indeed bring ABMS within the reach of anthropologists who even have no familiarity with modelling.

5.8 Indeed, to build agent-based models from such data, a major difficulty is the step from a limited number of individuals interviewed to the creation of a whole society. The stories and decision-making are usually personal and related to personal incidents; it is hard to draw certain ‘types’ of agents from that, because those coincidental incidents in life have a large influence. While estimating the percentages of the type of people forming the society is hard, in the eventual ABM, these can become parameters for variation.

5.9 Finally, it is important to emphasize that the structuring of collected data although highly facilitated with MAIA, still depends on the creativity of the modeller. There are many choices and interpretations that the modeller has to make to transform qualitative data into an agent-based model. When MAIA is used, however, there will be both a unambiguously language to communicate about the decision taken, and a traceable track record of how the researcher arrived from empirical data to interpreted model results and model.

Conclusion

6.1 Managing and structuring data, especially qualitative, is a major challenge for agent-based modelling. This research presented a method to effectively use ethnographic data for building agent-based models.

6.2 We used the MAIA framework to semi-structure the data collection procedure and later on used the same framework to decompose the information and build a conceptual agent-based model. The conceptual model is then used to produce running simulations.

6.3 Although MAIA facilitated the structuring of qualitative information, another phase of data collection is required, namely one to complete the quantitative aspects of the simulation. This phase is not yet supported by the methodological process presented here. Therefore, the next step of this research is to extend the MAIA framework to support the quantitative data collection process.

Acknowledgements

This work was supported by the European Regional Development Fund, Duurzame Greengroep Westland-Oostland Task Force (http://greenportduurzaam.nl).

Notes

1 In ethnography, coding is the process of organizing the collected data for analysis. See http://www.helium.com/items/948634-explaining-the-inductive-approach-cultural-anthropology
2 Unstructured interviews are conversations that can take place anywhere and anytime; structured interviewees are completed while strictly adhering to the predefined interview protocol. 3 http://www.atlas.ti.com
4 More about MAIA and the modelling environment can be found at maia.tudelft.nl.
5 To setup interviews and questions using the MAIA framework, it may also be helpful to start conversations with a life history/narrative of the interviewees before diving into specific questions. While such conversations may take hours, such ‘off the record’ conversations can be very helpful. MAIA may help to strike the right balance between such talks and more to-the-point conversation and interview.
6 Ethnographic field work, the full MAIA model and the ethnographic analysis can be found in (Schrauwen, 2012).
Figure 2. The UML class diagram for the MAIA meta-model (Ghorbani et al. 2013)


http://jasss.soc.surrey.ac.uk/18/1/2.html