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

Abstract

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.

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 (Sanches & Lucas 2002; Dia 2002).

1.3 One field of research that can also be used to collect data for agent-based models is ethnography (Bharwani 2004). Ethnography is a research method covering many approaches in anthropology. The data is gathered through interviews and field surveys which are then "codified"[1] 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 semi-structured (Sanches & Lucas 2002).

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 can 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 Bharwani (2004). Bharwani (2004) provides a detailed procedure for the fieldwork process which describes how ethnographic data is collected and formalized. Bharwani (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 is to use a model of data to represent relationships among actors (Yang & Gilbert 2008).

1.7 There are also case specific examples of using ethnographic 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 directly drawn from empirical studies. Moore et al. (2009) use a combination of ethnography and ABMS to study psychotherapy use and related harms. They also indicate the difficulty in generalizing ethnographic information to build agent-based models. They built a model called SimAmph as a shared ontology to combine ethnography and ABMS for their particular case, which proved to be useful in making the connection between 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 (Dharmat 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.

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 horticulture case study. In Section 4, we explain the methodological processes of integrating ethnographic processes into ABMS. In Section 5, we discuss the lesson learnt from this process and analyse our methodological processes. Finally, we conclude in Section 6.
3.3 Structuring interviews with MAIA

Structuring interviews with MAIA involves the formalization of the Institutional Analysis and Development (IAD) framework of Eloron Ostrom (2009), extended with concepts from other social science theories (Structuration [Giddens 1984], Social mechanisms [Hedström & Swedberg 1999] and Actor-centered institutionalism [Scharpf 1997]).

2.8 The framework provides a guideline to arrive at a comprehensive overview if not model of a social system by defining five interrelated structures: (a) social problem, (b) evaluation of the problem, (c) institutional strategy, and (d) action arena, which are in turn ordered by actions. The evaluative structure provides concepts with the help of which the modeller can indicate what patterns of interaction, evaluation, and outcomes she is interested in. The modeller identifies those variables that can serve as indicators for model validity (is it sufficiently realistic?) and model usability (will its implementation help me to explore the question(s) I set out to address?).

Figure 2 at the end of this article shows the concepts in MAIA Extensive information on MAIA can be found in Ghobani et al. (2013).

Case Study: Horticulture Innovation

3.1 The key objective of our study of the horticulture sector is to elucidate the effects social institutions have on innovation practices in Westland, a region that is home to about 70% of all greenhouse acreage in the Netherlands.

3.2 The horticulture sector in the Netherlands at large is facing economic difficulties, which have become more severe since the crisis began in 2008 (Schrauwen 2012). The dominant presence of innovation strategies that target cost-reduction and volume-increase brings down the costs of products. They fail to bring the growers sustained benefits however, which causes serious problems in the sector. Due to mechanisms in the market, the growers only benefit financially from their innovations for a relatively short period. When their innovations spread in the sector, the market price of their products drops rapidly, because it is subject to fierce price competition, a characteristic of ‘cost leadership’ market segments. Few growers attempt to increase the value of their products by developing niche product-market combinations, or expand their activities in the value-chain by developing new channels to the market to capture a greater share of the value created between growers and consumers. Such innovation strategies beyond process innovation for unit cost-price reduction are less popular in the sector, despite their potential to counteract the effect of downward spiraling prices in competitive markets.

3.3 The goal of this study is to investigate the innovation practices in the Westland horticulture sector to obtain an understanding on how this observed pattern of innovation has emerged and how the underlying behavior of growers is shaped and maintained. We use grounded theory as our methodology to perform ethnographic field work. Besides using MAIA for data collection and model development, we perform a theoretical analysis using the Bathtub model of Coleman (1988) and several other theories (see Schrauwen 2012). The rationale for adopting a fieldwork approach (rooted in cultural anthropology) is that the organizations and innovation practices are socially embedded, and can be studied as such. Furthermore, the Westland is said to be home to Westlanders who share a common identity with respect to social and business culture, which is shaped by and has shaped their core business for centuries (Kazmi et al. 2013).

The Modelling Process

4.1 The purpose of our methodological practice is to guide the collection of data for building an agent-based model using an ethnocentric approach. This process is divided into two parts. The first part uses MAIA as a template for information collection, which includes field observation, interviews and the study of formal documents. For each of these methods, we make use of the MAIA framework to semi-structure the data collection process. The second part uses the collected information to build a MAIA model.

Collecting data using MAIA

Structuring interview with MAIA

4.2 In inductive ethnographic research, interviews are normally semi-structured. Therefore, it is common practice, to develop a general structure or guideline for the interviews, to ascertain that at least all relevant aspects are addressed. We use MAIA as the general structure for the interviews in order to cover all the information required to build an agent-based model. At the same time, we leave the questions open-ended so that the interviewees feel free to talk about what they may seem relevant to them.

4.3 The interviews were conducted with various stakeholders in the Westland horticulture sector (Schrauwen 2012):

- Experts: Experts were interviewed to gain better insight into the sector as a whole and also to evaluate the assumptions that were being made during the analysis and modelling phase.
- Growers: Fifteen growers were visited at their organization. Each interview took between two to five hours. The growers were either contacted directly or introduced by other respondents.
- Organizations: The bank, churches, educational institutes, municipality, LTO GlasKrocht and supermarket were the other actors interviewed in order to find out their influence on the social network of growers, their individual capital and investment, and their knowledge and background.

4.4 The concepts that were used to structure the interviews and direct the questions are:

- Collective Structure
  - Agent Decisions: What decisions do the growers make regarding their innovation practices? The growers are allowed to talk about their decisions freely without being forced to explain how they make those decisions.
  - Agent personal value: The growers are asked about what they care about most when they are making those decisions.
  - Related Agents: During the interviews, the growers are asked about other social entities they may be interacting with. These can be individual actors, such as other growers, or composite actors (i.e., organizational type) such as the bank, or the municipality.

- Operational Structure
  - Actions and Plans: The growers are asked about what their general activities are and how often they perform these activities. In this case study, they were asked about their daily, monthly and yearly activities. If each of these practices constitutes a process, they were also asked about the events that take place in that process. For example, if a grower decides to apply for a subsidy, what actions does he have to perform during the application process?

- Constitutional Structure
  - Roles: The growers are implicitly asked about the different roles they take in their activities. This is not a straightforward question, but one that would rather need to be extracted from the explanations thegrowers provide. For example, a grower explains that he has to be a client of the bank to apply for a particular subsidy or he emphasizes that he would only expand his greenhouse if he has a child who is willing to take over. From these remarks we can identify ‘bank client’ and ‘having a father’ as two of the roles, the growers may assume under certain conditions.
  - External Institutions: While asking about the operational activities and decisions, the subjects are also asked about the formal procedures, rules and regulations they need to go through. This is later used to collect relevant institutional documents.

- Physical Structure
  - Physical Components: During the interviews, the subjects are asked about the physical entities they use in their activities, the ones they own or the ones that influence their actions. It is important to ask about this aspect; while the interviewees are talking about the activities he performs in order to limit the information to what is relevant.

The interviews are recorded and coded in Atlas. i 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 [7].

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 criteria 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 interviewees. This entity 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 agents 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 actions 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>Type</th>
<th># A</th>
<th>Growers</th>
<th>C</th>
<th>O</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.a</td>
<td>Growers</td>
<td>may 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 1.b, 1.c</td>
<td>O</td>
<td>Type</td>
</tr>
<tr>
<td>1.b</td>
<td>Growers</td>
<td>must join one of the 6 sales cooperations</td>
<td>if they want to get the GMO subsidy from the EU</td>
<td>2</td>
<td>rule</td>
</tr>
<tr>
<td>1.c</td>
<td>Growers</td>
<td>may not market under their own brand</td>
<td>if they want to get the GMO subsidy from the EU</td>
<td>2</td>
<td>rule</td>
</tr>
<tr>
<td>2</td>
<td>EU may fine the grower</td>
<td>if growers don't follow the rules attached to the subsidy</td>
<td>2</td>
<td>rule</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Rabbank may not increase the interest on loans when growers are a less optimal financial situation</td>
<td>rule</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</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>5</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>6</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>7</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 can be extracted from the descriptions the subjects provide. The described sequence of actions helps to define agent plans in MAIA. Finally, 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 (Gedantechnieken Mark Ordening - collective market structuration) subsidy where 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.

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 is 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

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

22/10/2015
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. Unstructured interviews are conversations that can take place anywhere and anytime; structured interviews are completed while strictly adhering to the predefined interview protocol. http://www.atlas.ti.com. More about MAIA and the modelling environment can be found at maia.tudelft.nl. To setup interviews and questions using the MAIA framework, it may also be helpful to start conversations with a life history/narrative of the interviewee 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. The 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)

References


