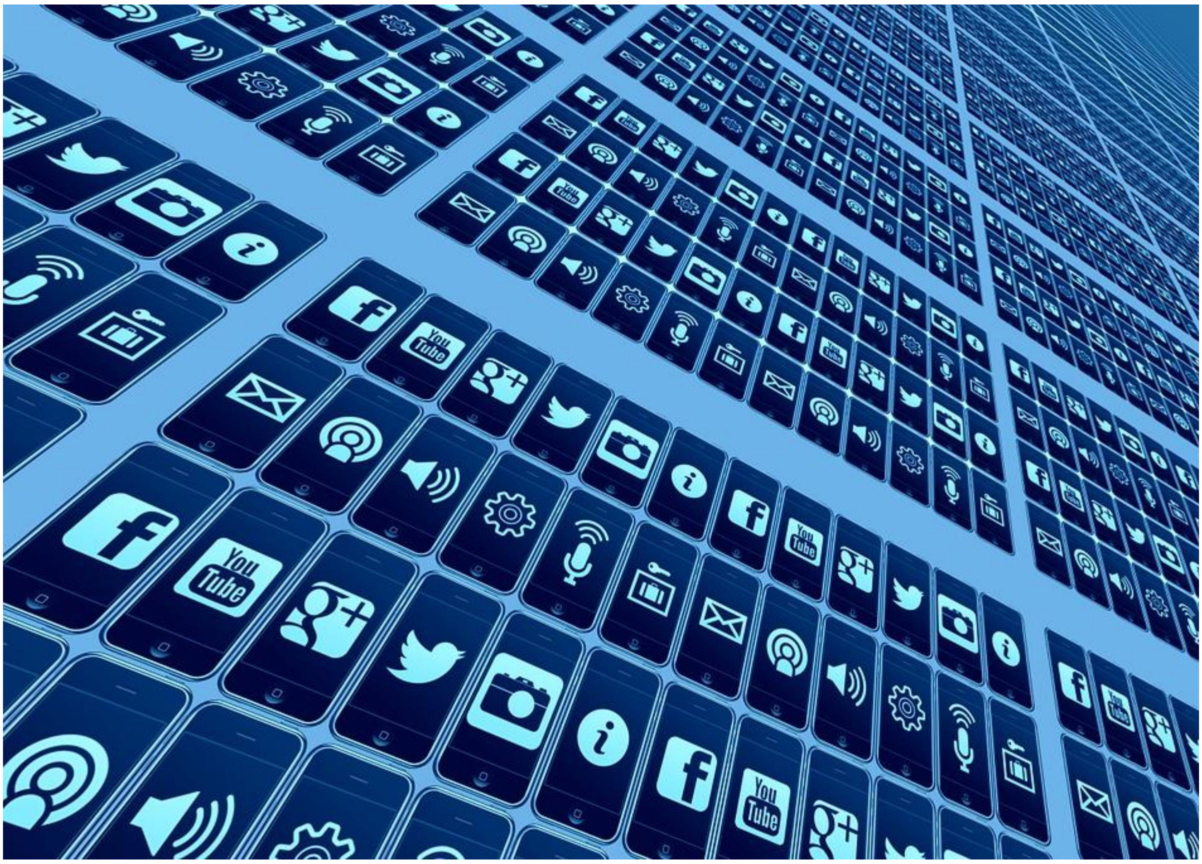
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 763912

Report No. SMARTEES 7.1 – D7.1 Report on the conceptual model of the SMARTEES simulation and data types to be included

Report on the conceptual model of the SMARTEES simulation and data types to be included



Picture under creative common licence. © Ag Ku. at pixabay.com

REPORT ON THE CONCEPTUAL MODEL OF THE SMARTEES SIMULATION AND DATA TYPES TO BE INCLUDED

KEYWORDS:
Simulation concept, CONSUMAT, FEARLUS, Agent-based modelling, Decision trees, Data mining, Decision-making, Artificial Intelligence, Metadata, Provenance, Semantic.

VERSION
01

DATE
31.08.2018

AUTHOR(S)
Patrycja Antosz, Wander Jager, Gary Polhill, Jiaqi Ge, Doug Salt, Amparo Alonso-Betanzos, Noelia Sánchez-Maroño, Bertha Guijarro-Berdiñas

PROJECT NO.
763912 (H2020)

NUMBER of PAGES/APPENDICES
38/1

ABSTRACT

SMARTEES REPORT ON THE CONCEPTUAL MODEL OBJECTIVE

This paper reports on the conceptual model of the SMARTEES simulation and data formats to be included. It is based on Tasks 7.1 (Conceptual modelling of the linkages between the different structural layers) and 7.2 (Developing a checklist for a data structure that matches the conceptual models). It provides conceptual solutions to integrate the different ABM traditions regarding the different layers in processes of social innovation (FEARLUS), human decision-making (CONSUMAT) and the habitual patterns of behaviour (decision trees or alternative machine learning methods). To enhance compatibility between the modelling infrastructure and the data collection activities, it also provides a description of the data types that can be incorporated in the project.

REPORT NO.	ISBN	CLASSIFICATION	CLASSIFICATION THIS PAGE
D7.1	NA	public	public

Document history

VERSION	DATE	VERSION DESCRIPTION
1	31.08.2018	Release of first version to EU

Table of contents

1 INTRODUCTION

SMARTEES aims at studying social innovation in a variety of cases. Not only do these cases differ concerning their location in Europe and type of innovation, such as transitions in traffic or investment in insulation, but they also differ concerning their behavioural context. Where some cases address investments in insulation, and hence almost refer to a single adoption decision (in a social context), other cases deal with changes in daily behavioural patterns (e.g. transportation habits), and sometimes successful changes create the conditions for further developments, resulting in cascading effects to a more sustainable community culture.

The conceptual model we develop for SMARTEES will have to be capable of covering the behavioural drivers, processes and social dynamics as present in this variety of cases. It will provide an integrated theoretical framework of behaviour that serves to connect the different modelling methodologies and tools within the project. Where it is obvious that the modelling of different cases will have to focus on different behavioural aspects, e.g. one-shot investment decisions versus changes in more complex habitual patterns, all simulation models developed within SMARTEES will be clearly positioned within the connecting theoretical framework. This integrated theoretical framework will also support connecting different modelling approaches in a consistent manner when this is needed for a particular case.

This report serves two main purposes. One, it describes a conceptual approach to agent-based modeling activities in SMARTEES. Two, it outlines how the conceptual approach maps into the requirements for software platforms and data types used in the project.

The conceptual approach to modeling stems from three perspectives described in the document: FEARLUS (see Section 2), CONSUMAT (see Section 3) and data-driven ABM (see Section 5). FEARLUS and CONSUMAT are theoretically-driven perspectives to modeling, which focus on different aspects of social phenomena. FEARLUS comprehensively shows micro-macro linkages between social and environmental systems. CONSUMAT outlines decision-making strategies of agents in the contexts of their need satisfaction and uncertainty. Section 4 of the report highlights commonalities and differences of the two perspectives. Data-driven ABM refers to a general concept of using quantitative and qualitative data to inform agent-based models. The description focuses specifically on possibilities of deriving assumptions for modeling decision-making processes and handling data pre-processing.

The conceptual approach to modeling taken in SMARTEES has particular consequences for data definition and collection. Four general rules guiding data collection were identified: (1) tailoring theoretical concepts and data collection processes to modelling cases of social innovation, (2) utilising secondary data to the greatest possible extent, (3) Establishing timelines, and (4) starting with descriptive models, which might be more complex but are understandable to decision-makers (see Section 6). Principles of metadata are defined in the last section of the document (see Section 7). Initial, more detailed specification of metadata, which will be updated throughout the project, is presented in Appendix 1.

2 SOCIAL INNOVATION IN FEARLUS

(Framework for the Evaluation and Assessment of Regional Land Use Scenarios)

FEARLUS is a modelling system that has been used to explore scenarios of agricultural land use change and various versions have featured in publications from 2001 through to the most recent in 2016. It has not specifically been applied in contexts where the 'social innovation' label has been applied, however, there are various dimensions of social innovation that are relevant to studies to which FEARLUS has been applied:

- It has been used to explore the dynamics of the spread of innovations in the sense of the land use decisions that the land manager agents in the modelling system make, most specifically around issues with experimentation versus habit (Gotts et al. 2003) and dynamics of imitation (Polhill et al. 2001; Gotts and Polhill 2009; Gotts and Polhill 2010);
- With respect to systemic change, it has been applied as an illustration of ways in which models can move beyond their designed envelope of operation, and to demonstrate how systemic change can occur through the collapse and reformation of regimes of land manager populations over time (Polhill et al. 2016);
- Most relevantly to social innovation as "a change in social relations, involving new ways of doing, organizing, framing and/or knowing" (Loorbach et al. 2016), FEARLUS was coupled with a biodiversity model (SPOMM) to explore new ways of incentivising land managers to adopt practices that promote biodiversity, including rewarding for biodiversity outcomes rather than for specific activities that are believed to do so, and 'clustering rewards' to incentivise managers to work together to create larger areas of the landscape providing contiguous habitat for species to be protected (Gimona and Polhill 2011; Polhill et al. 2013).

FEARLUS is applied to agricultural contexts, though Cioffi-Revilla and Gotts (2003) show how the representation of the agricultural context in FEARLUS is sufficiently abstract that its dynamics can be linked to a model of military conflicts between nation states (Cederman 2003). FEARLUS has also not typically been used in empirical contexts, with the intention instead to explore theoretical ideas. It was classified as a 'typification' by Boero and Squazzoni (2005) – a model designed to explore a class of systems rather than a specific one. The exception is Iturrioz and Polhill (2014), who applied FEARLUS to a case study in Argentina.

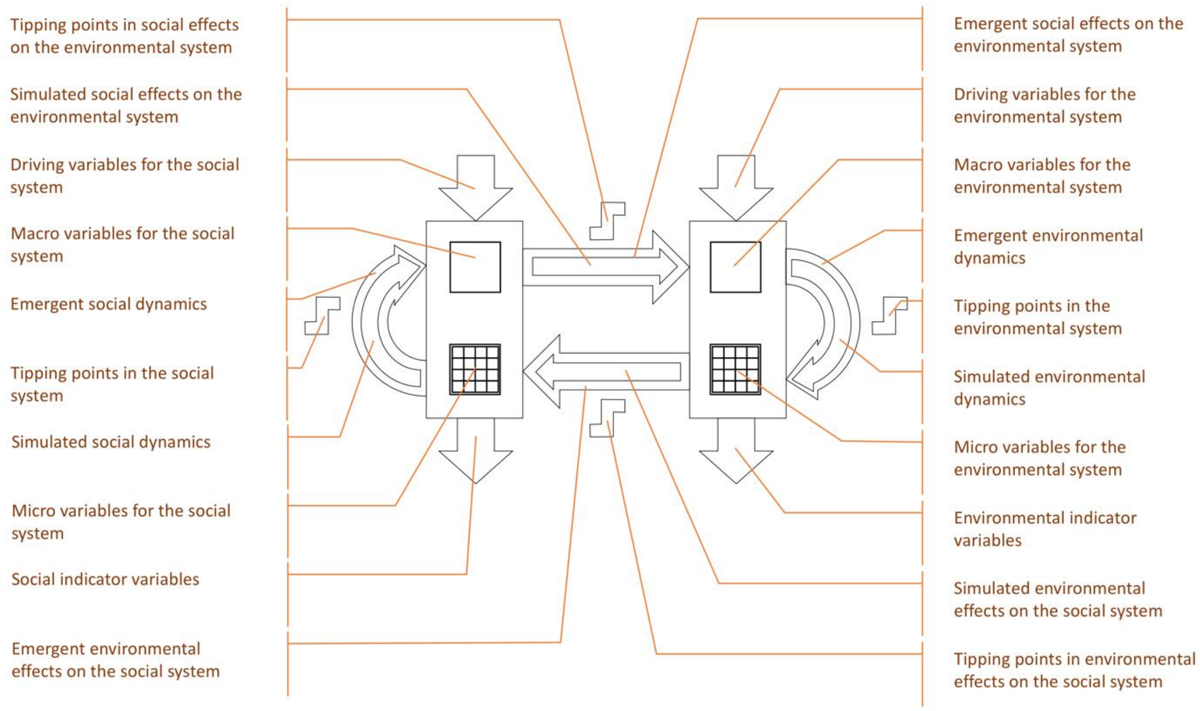
FEARLUS is perhaps better conceived as a modelling framework than a single model because the software is capable of switching on and off various components, and has been extended over the years since its first publication in 2001. However, the principal agents are land managers, who have to choose how to manage (use) their land. They have a discrete set of land use options among which to choose. The nature of the algorithms varies from publication to publication, with early work using heuristic algorithms, and later work using a simplified version of case-based reasoning (Aamodt and Plaza 1994), which is a symbolic AI algorithm that simulates the decision-making of domain experts selecting the option to use based on the similarity of the current context with those they have experienced previously.

The decision-making of land-manager agents is discussed in more detail in a subsequent section of this document, where it is compared with the approach taken by CONSUMAT (Jager 2000). The model simulates the consequences of the decisions in terms of profits returned to land managers, and in later versions of the model, financial incentives issued by an agent representing the government.

With FEARLUS having application in agriculture rather than the largely urban contexts to which SMARTEES is directed, it is not worth going into detail about the specific structure and functioning of the model. The interested reader is referred to the papers cited above. Rather, here, we consider the socio-environmental structure underpinning FEARLUS and FEARLUS-SPOMM at an abstract level, applying a modified version of the framework adopted in Polhill et al. (2016).

Figure 1 shows schematically various aspects of a coupled complex social-environmental system that could be documented when being simulated. The social system is represented in the left hand rectangle, which contains a square and a grid representing the macro and micro levels. The pair of nested arrows on the left-hand side show simulated processes (inner) and emergent dynamics (outer) within the social system, with the 'S' shape next to it documenting tipping points. Tipping points occur at the boundaries between one 'metastable state' (or regime) and another. As Polhill et al. (2016) describe it, tipping points can simply apply to a change in the arrangement of an existing regime, such as a new democratic government being elected. More radically, they can lead to the switching on of new processes that were not in operation previously, or the cessation of existing processes – an example might be the introduction of free healthcare. Most radically of all, the new regime requires a completely different vocabulary to describe what is happening, such as the shift from a feudal to a democratic system of governance. Either of the latter two are consistent with social information; the last being a full transformation – an alteration or replacement of established formal and informal institutions in the terminology of Loorbach et al. (2016). The right-hand side of the diagram shows the biophysical environment providing the context in which the social activities are taking place. The environment may have its own dynamics, just as the social side does, and there may be similar processes, emergent effects and tipping points across the social and environmental subsystems. Each of the social and environmental subsystems may have driving variables and/or measured indicators, as represented by the wide vertical arrows entering and exiting the subsystem rectangles.

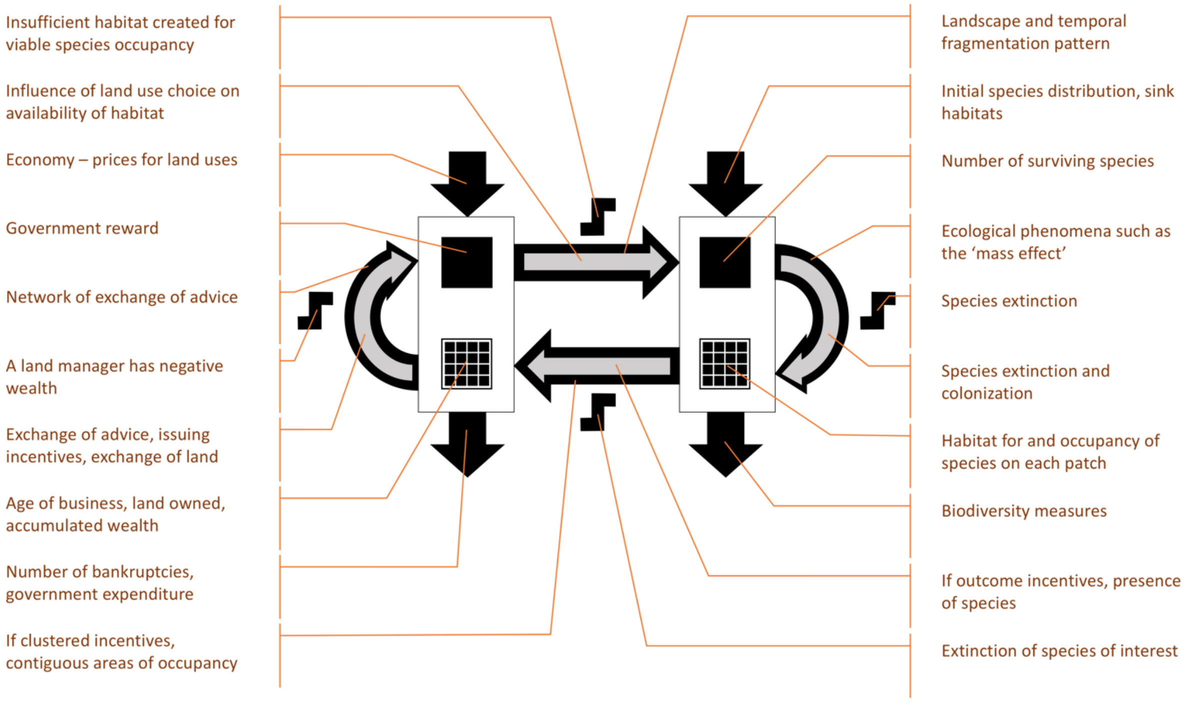
Figure 1. A general framework for conceptualizing complex interactions between social and environmental systems.



In Figure 2, the framework from Figure 1 is applied to FEARLUS-SPOMM. The various parts of the framework are highlighted to show that they are active. Not all components need to be highlighted in every case; in FEARLUS on its own, for example, environmental dynamics are not simulated particularly richly. In FEARLUS-SPOMM, however, the environmental dynamics are essential in understanding whether key indicators of biodiversity, such as the survival of particular species of interest, are achieved.

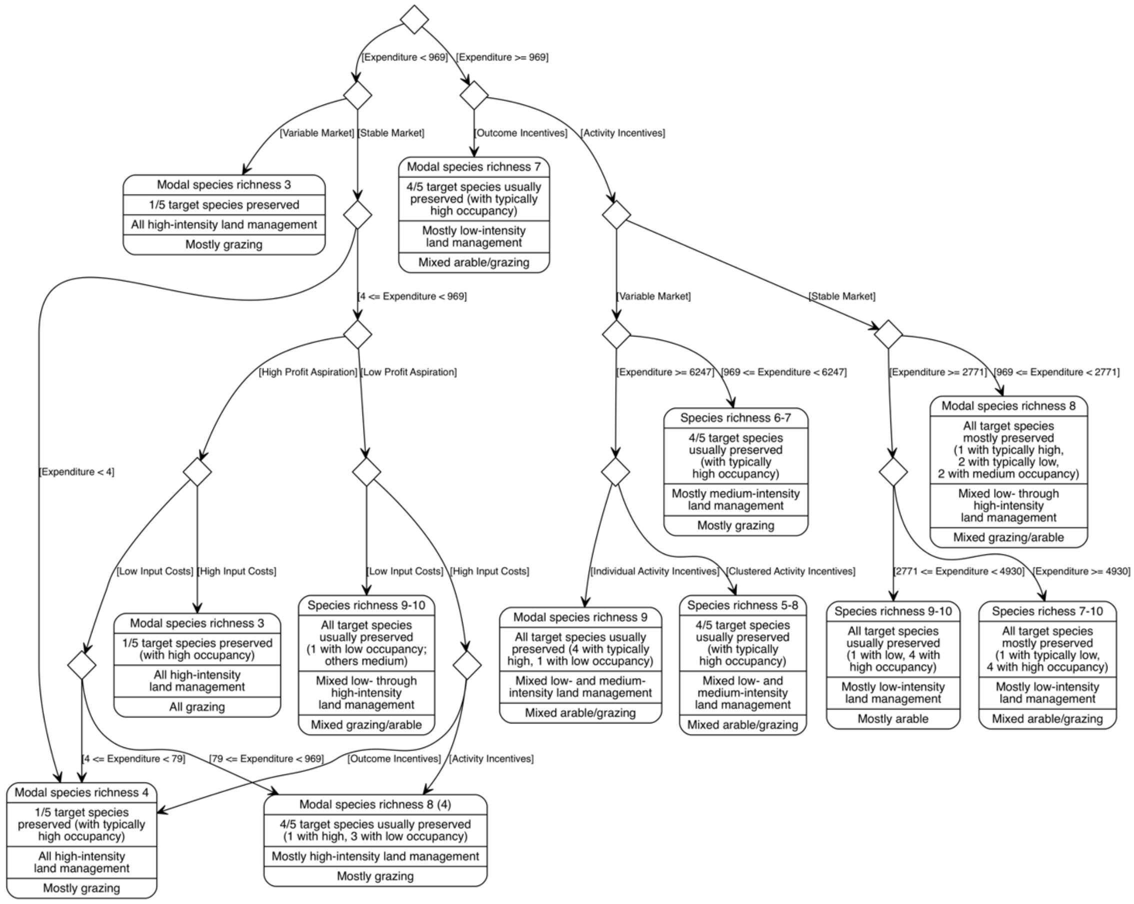
Typical usage with models such as this, particularly given their more stylized representation (no specific empirical case study is modelled), is to run them several thousand times under different conditions to get an impression of the various regimes that emerge. Figure 3 reworks a diagram in Polhill et al. (2013; fig. 5, p. 83), and shows (from the route to the rightmost box at the top of the diagram) that provided the government spends enough money, outcome-based incentive schemes robustly deliver reasonably good species richness, with most of the target species preserved with high typical levels of occupancy. This happens despite other influences on behaviour, such as input costs, market variability and aspirations of land managers. In comparison with activity-based incentives, outcome-based incentives are less sensitive to expenditure and market variability.

Figure 2. Applying the framework in Figure 1 to FEARLUS-SPOMM as described in Gimona and Polhill (2011) and Polhill et al. (2013).



Though decision trees will be discussed later on in the context of developing decision-making algorithms for agents based on empirical evidence, Figure 3 also shows how recursive partitioning classification trees (Breiman et al. 1984) can be used for the analysis of results. Here, the driving variables of the model are used as explanatory variables, and species richness as the response variable. The leaf nodes of the resulting decision tree correspond to distinct regimes of system behaviour with respect to land management intensity, predominance of arable or grazing, and the combinations and levels of occupancy of target species preserved in the landscape.

Figure 3. Variety of outcomes from FEARLUS-SPOMM simulations described in Polhill et al. (2013).



3 INDIVIDUAL DECISION-MAKING WITH CONSUMAT

Energy related consumer behaviour touches upon a wide variety of choice and usage processes, ranging from strong daily habits, e.g. in transportation choice, occasional elaborate action, e.g. when considering energy providers, and highly involved and socially relevant decisions, e.g. when purchasing a car. Variations in citizens' involvement, existing knowledge and social susceptibility cause heterogeneity in behaviour. As a result, either for individual or social reasons, some consumers may be more innovative than others. Processes of innovation diffusion, when new technologies and practices spread through a society, involve sharing experiences and development of norms in the social fabric of a society (e.g., Van Eck et al. 2013).

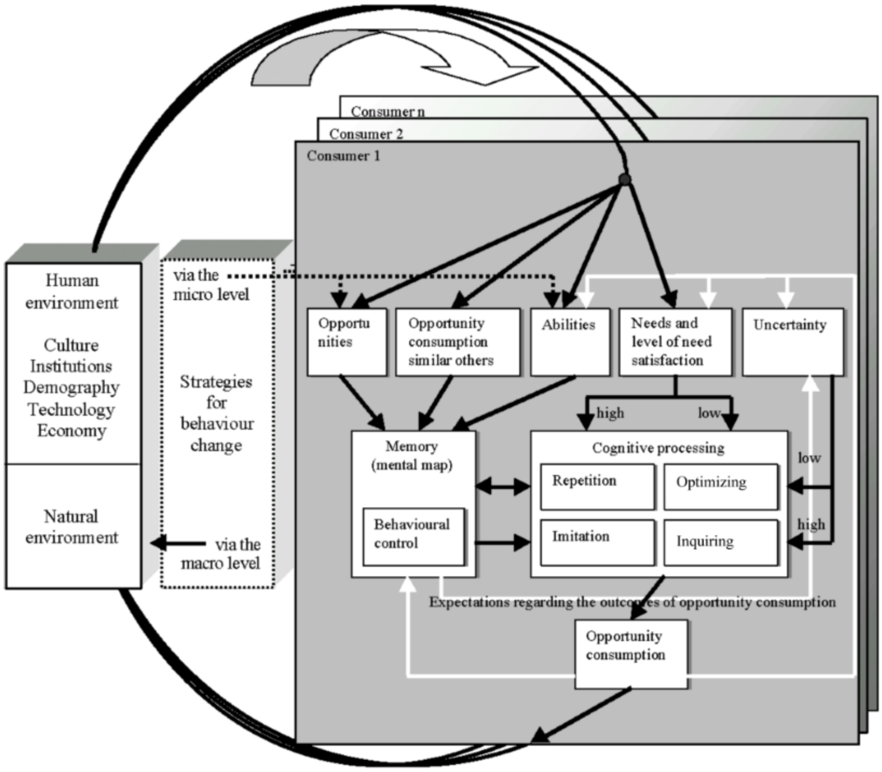
Many behavioural theories provide insights in the different causal mechanisms of consumer choice and usage processes, however, the challenge is to implement this knowledge in an integrated framework that allows for exploring how these operate in a societal context over time (e.g., Schlüter et al. 2017). Agent based models offer the methodological tool to construct networks of heterogeneous consumers equipped with different decision mechanisms, and hence allow studying the complexities of behavioural change in a society.

The CONSUMAT model (Jager 2000) has been developed to provide a conceptual framework that serves the development of agent based models capturing different decision mechanisms in a network of heterogeneous consumers, and has been applied to a variety of behaviours, including policy experiments on the diffusion of electric cars (Kangur et al. 2016). The principle of the CONSUMAT is based on organizing decision strategies on the basis of (1) the level of need satisfaction of a consumer, driving the cognitive effort, and (2) the uncertainty of the consumer, where personality and the decision context drive the individual versus social orientation of consumer decision making. Hence, for a satisfied and certain consumer the principles of habitual behaviour apply, whereas an uncertain and dissatisfied consumer will invest more cognitive effort in acquiring information from other people. The crux of this approach is that it allows for modeling changes in the decision making of a consumer, for example how a reduced satisfaction causes a consumer to reconsider a habit, and how opinion leaders in her/his network may influence the choice for an option that may develop into a new habit. This opens the possibility to model how new behaviours diffuse through a society, where the innovators may have different motives and engage in different decision processes than later adopters. Especially when the social context is relevant, this approach allows for the identification of tipping points in social systems (e.g., Nyborg et al. 2016).

3.1 The CONSUMAT model of artificial consumers

The aim of the CONSUMAT approach (Fig. 4) is to support the development of domain-specific social simulation models of consumer behaviour. It provides a simplified and, at the same time easily implementable, conceptual framework to model human actions. Agents in models are seen as consumers, as they consume behavioural opportunities present in their environment. The choice of a particular strategy/behavioural opportunity is driven by individual need satisfaction and experienced level of uncertainty via cognitive processes of decision-making and memory access. Applied to a population of heterogeneous simulated agents, this results in population behaviour that aggregates into macro-level outcomes, both in terms of the human environment (e.g., consumptive culture and norms) and the natural environment (e.g., emissions). The CONSUMAT model "closes the loop" by feed-forwarding this aggregated population behaviour towards the decision context of individual agents at the next moment in time. This allows for modelling individual processes, e.g. habit formation, as well as micro-macro processes, e.g. the emergence of norms, over time. It also allows for individual agents to switch between cognitive strategies when they experience (un)certainty and/or (dis)satisfaction. The CONSUMAT approach aims to provide a simple structure for simulated consumers to determine which type of decision strategy is used under which conditions.

Figure 4. Overview of the CONSUMAT framework.



Need satisfaction and **uncertainty** drive the type of decision making in which an agent engages (see Jager 2000 for an extensive description).

3.2 Need satisfaction

While we acknowledge the possibility of including more elaborated needs or goals in a model, in the interest of keeping our model transparent we start with three behaviour-driving forces: 1) existence/sustenance (related to safety); 2) social belonging and status (group position); and 3) personal preferences (taste, beliefs). The need to **exist** is related to having means to prevail, e.g. food, income, housing. Agents act in order to avoid depletion of these resources over time. **Social belonging and status** needs are associated with having interactions with others, belonging to a group, and maintaining/achieving social status. **Personal preferences** refer to satisfying one's personal taste with respect to overall life values and norms, e.g. environmental protection, altruism, or enjoyment of life. Agents balance the importance of these needs. As a result, some agents may be mostly motivated by the drive to manage their resources (existential need), while others may be more susceptible to the influences of other agents (social need). The importance of needs may also be related to the context of a situation, which may provide a so-called frame determining which need will be at focus in a certain context (e.g. Lindenberg & Steg, 2007). In any case, the fulfilment of needs results in **satisfaction**. The needs can be satisfied by consuming certain behavioural opportunities, e.g. purchasing products, engaging in particular practices or harvesting natural resources. A high degree of satisfaction suggests that the agent has made gratifying choices in the past and that it is doing well, so there is no need to engage in extensive decision-making at that moment. Dissatisfaction, however, requires extensive scrutiny of alternatives to increase the agent's satisfaction.

3.3 Uncertainty

Uncertainty is a psychological state influenced by insecurity concerning the expected results of performing behaviour, e.g. in situations where many alternatives are available and choice options are composed of many attributes. Also, when one's behaviour deviates from the norm, uncertainty may arise. In these circumstances using the experiences of other people and observing their behaviour is an effective strategy. Theory on similarity shows that people have a stronger tendency to interact with similar others (see e.g., Byrne 1961; McPherson et al. 2001); correspondingly, in the CONSUMAT framework the chances of interaction can be based on similarity concerning agent characteristics and behaviour.

3.4 Decision making

Depending on the satisfaction and uncertainty levels of the CONSUMAT agent, it will engage in one of the four cognitive strategies (illustrated in Figure 4, in the part of Consumer 1 labelled *Cognitive processing*):

1. Low uncertainty and high satisfaction prompt agents to engage in **repetition**, which is the script-based mechanism driving habitual behaviour (e.g., Wood & Ruenger 2016).
2. High uncertainty and high satisfaction results in **imitation**, which is e.g. an important driver of fashion dynamics (e.g., Bandura 1977).
3. When satisfaction is low, the agents are more motivated to invest effort in improving their situation. Hence when they are certain but dissatisfied, they will engage in **deliberation** - an assessment of available options implemented as expected utility maximization (see e.g. Anand 1993).
4. Low satisfaction and high uncertainty results in **inquiring**, where the behaviour of *comparable/similar* others is evaluated and copied if it increases expected satisfaction (see e.g. Ellison & Fudenberg 1995; Rosenbaum 1986). Whereas the imitation strategy just copies behaviour, the inquiring strategy is aimed at obtaining information from (relevant) others, and making a more deliberate choice on the options that have been identified using this social informative strategy. On the basis of similarity, a fixed social network for each agent can be constructed. Yet, when agent characteristics change over time, similarity is recalculated. This allows for simulating a dynamic network, which may be relevant in studying the development of consumer segments over time. Both fixed and dynamic networks can be implemented using this approach. Especially for modelling normative influences and social informative strategies the network is the vehicle through which information travels. Policy strategies aimed at connecting certain key individuals thus may have a profound effect on the diffusion of ideas and practices.

3.5 Individual representations of reality

The agents have a **memory** that serves as a mental map (Fig. 4, to the left of *Cognitive processing*). For future decisions, memory stores information on behavioural opportunities, as well as information on other agents' behaviour and abilities obtained from deliberation and inquiring. Hence, memory is updated only if the agent uses cognitively demanding decision strategies. As a consequence, a satisfied agent can continue to habitually perform particular behaviour (repetition) without updating its memory with information on newer or potentially better opportunities. **Abilities**, i.e. the agent's capacity to actually use particular behavioural opportunities, allow agents to take action. Combining information about agent's own abilities with the requirements for using a certain behavioural opportunity results in the formalisation of behavioural control in the memory, e.g., the agent knowing whether it can financially afford a certain product.

3.6 Micro-macro links

The behaviour of individual agents aggregates into collective impacts, which may affect the human and/or natural environment, depending on the domain being modelled. For example, if many agents follow a certain fishing strategy, this will impact the market price of fish (economy) and fish-stock (ecology, see e.g. Jager et al. 2000). In a same vein, if a cluster of agents in a network starts performing a new behaviour, others at the periphery of this cluster may experience the changing of the norm, and adopt as well, thus contributing to a “social contagion” process spreading through the network.

3.7 Model application

The CONSUMAT approach focuses on providing a framework for positioning theoretical mechanisms in a causal loop that is required for formal modelling. Consumers display different decision strategies in selecting behaviour, such as relying on habits, imitating peers or role-models, making deliberate comparisons and asking friends for advice. The CONSUMAT model offers a generic conceptual framework that combines and connects different decision strategies and their underlying drivers. Also, the switching between different decision strategies, e.g. when a short period of deliberation may result in a change of habit, is explicitly being targeted by the CONSUMAT model.

The CONSUMAT provides a generic framework that can be applied to different domains of environmentally relevant behaviour, e.g. the diffusion of electric cars (Kangur et al. 2016), consumer life styles (Bravo et al. 2013), farmers' interaction with climatic change (Van Duinen et al. 2015), and integrated models of consumer behaviour, economic markets and ecological systems (Jager et al. 2000). Depending on the domain and available data, the CONSUMAT approach can guide the development of a specific social simulation model.

3.8 CONSUMAT and SMARTEES – what do we do this for?

As a result of the modeling efforts, we aim to produce a theoretically grounded and empirically representative tool that allows us to experiment with simulated conditions. This experimentation will first explore the model behaviour. Systematic variation of parameter values will reveal under which (empirically based) conditions the model produces more stable population behaviour, and under which (deviating from empirics) conditions behaviour will change. Here we will focus in particular on discriminating between linear effects (e.g., price-demand functions) and non-linear effects (e.g. tipping points). The model will produce data that allow for statistical analysis of time series (changes in variances, autocorrelations) that are indicative for transitions (tipping-points) between linear and non-linear behaviour regimes.

The aim of performing simulations on the agent-based models of successful social innovations is to **identify conditions where tipping points** in particular social systems occur, i.e. conditions leading to large-scale changes in habitual behaviours of citizens, who adapt to innovations. Ecological science demonstrates that major transitions in ecological systems towards a different regime (transition) are often preceded by increased variances, slower recovery from small perturbations (critical slowing down) and increased return times (Boettiger & Hastings 2012; Dai et al. 2012; Dakos et al. 2012). The expectation is that also for other complex systems such indicators may identify the approach of a tipping point and a regime shift or transition (Scheffer et al. 2009). A social simulation model allows for following the behaviour of the parameters over time, and hence makes it possible to measure variations and autocorrelations over time. As such, the simulation model may identify tipping points in systems, and the critical variables driving these tipping points.

Having an understanding of the model behaviour, more systematic policy experiments can be envisaged. In particular, we want to identify situations where combinations of sequenced policy measures (financial, infrastructural, technical, informational) can move a system from a linear to a non-linear regime. This helps identifying tipping points towards a smarter energy use, be it how different people transport themselves, buy, use and share appliances, and manage their energy in an increasing connected and smart energy-grid.

4 INTEGRATION OF THEORETICAL INPUTS

Several characteristics of CONSUMAT and FEARLUS make an integration between the two approaches possible. In general, the two approaches are similar, as they both take two perspectives: systemic and individual. However, FEARLUS emphasizes the systemic perspective to a greater extent, and CONSUMAT describes individual decision strategies in greater detail.

4.1 The systemic perspective

Both approaches recognize the importance of the systemic perspective and distinguish between two systems: the social system (human environment in CONSUMAT), and the environmental system (natural environment in CONSUMAT). In FEARLUS the relationships between the two systems are explicit and focus is on the interactions between them.

4.2 The individual perspective

Decision-making in FEARLUS has similarities with CONSUMAT in the sense that different algorithms are employed in different contexts, and that there is a dimension of satisfaction that determines that context. However, FEARLUS does not incorporate the dimension of uncertainty (Table 1).

Table 1. Satisfaction and uncertainty in FEARLUS and CONSUMAT.

Satisfaction	FEARLUS	CONSUMAT
High	Determined by whether aspirations for profit (and in some versions, social approval) are met. If aspirations are met, then a Habit strategy is used that simply repeats what was previously done. Gotts et al. (2003) explore the dynamics of aspiration in early versions of FEARLUS.	Sub-check Uncertainty: High Uncertainty: Imitation Low Uncertainty: Habit
Low	A random probability determines whether an 'imitative' or 'innovative' strategy is used. Imitative strategies are limited to choices selected among those used by neighbours. Innovative strategies are open to the full range of available options.	Sub-check Uncertainty: High Uncertainty: Inquiring Low Uncertainty: Deliberation

FEARLUS implements a range of options for the imitative and innovative strategies, which are chosen according to the context in which the simulation is run. In the case of imitation in FEARLUS, strategies cover the range of options covered by the union of imitation and inquiring in the CONSUMAT sense of these terms. Some imitative strategies in FEARLUS do not consider whether an option used by a neighbour is suitable for use by the deciding agent; others do. Gotts and Polhill (2009) give attention to the various options for the copying (used here to mean a generalization of inquiring and imitation in CONSUMAT) algorithms in FEARLUS and the suitability of each. Hence, there are contexts in which FEARLUS uses copying strategies where CONSUMAT would not, and vice versa. However, the set of cases in which CONSUMAT uses habit is a proper subset of the cases where FEARLUS does. Additionally, CONSUMAT makes a distinction between different types of social processing that can be further elaborated upon within SMARTEES to address social innovation in community networks. Relevant here is the difference between strategies that can rely on generic observation of community behaviour (imitation) and strategies that involve specific information processing from opinion leaders in the community (inquiring).

Deliberation in CONSUMAT covers a proper subset of the options that are implemented as innovative strategies in FEARLUS. FEARLUS does have strategies effectively implementing utility maximization (or other optimization approaches), as well as other algorithms that might be called 'deliberative', most notably (reduced form) case-based reasoning (as was the case in FEARLUS-SPOMM, e.g. Polhill et al. 2013), in the sense that they are based on a careful evaluation of all the available options. However, FEARLUS also has random experimentation, which is at least at a surface level, something not structurally available in CONSUMAT (but random exploration has been implemented in some models). Here, the agents just choose an option at random, with no evaluation or consideration of whether one will be superior to another. Early work with FEARLUS found that a mixture of habit and random experimentation was surprisingly successful in some contexts (Gotts et al. 2003).

Though there are key differences between FEARLUS and CONSUMAT, there is sufficient similarity that FEARLUS need not be used further as a principle in this project. Rather, we might draw on decision-making algorithms from symbolic AI and computational psychology, such as case-based reasoning or concept learning algorithms, in accordance with context and available data. Many of these symbolic AI algorithms do not include copying elements – a key weakness – however, in case-based reasoning, agents can exchange experiences and store them in their episodic memory for use when considering their own options.

Of more relevance may be other agent-based modelling work done by the James Hutton Institute team, and the decision algorithms used there. For example, Ge and Polhill (2016) used Dijkstra's algorithm for route finding by commuters, whilst Ge et al. (2018) calibrated a linear model on questionnaire data to form the basis of householder decision-making about where to move. The latter could be seen as empirically-derived deliberation based on agent attributes, triggered by dissatisfaction with the present location. As such, parts of the model in Ge et al. (2018) could to some extent fit with CONSUMAT, as they both have a potential to use machine learning algorithms (e.g. decision trees) to uncover 'habitual/default' behavioural patterns,

As discussed, much of the FEARLUS decision-making framework is subsumed by CONSUMAT – or at least, there is a considerable overlap. The key contributions from FEARLUS to SMARTEES thus pertain largely to conceptual and methodological matters:

- Use of symbolic AI in decision-making algorithms,
- Principles of simulation model studies – using multiple runs and analysing simulation output,
- Importance of a dynamic biophysical environment contextualizing agent behaviour.

5 DATA-DRIVEN AGENT BASED MODELLING

Agent-based modelling (ABM) is a way of representing complex systems of autonomous agents or actors, and of simulating the multiple potential outcomes of these agents' behaviours and interactions in the form of a range of alternatives or futures. Following the idea used in the paper by Sánchez-Maroño et al. (2015) for the LOCAW project, where ABM was used as a synthesis tool for representing everyday practices in the workplace pertaining to the use of energy and materials, management and generation of waste, and transport, whenever relevant for research questions in SMARTEES, we proposed to use a data-driven agent-based model. In the experience of LOCAW project, decision trees, which perform relatively well with limited data size, were used for modelling decision-making procedures, because questionnaires provided a relatively small data sample.

5.1 Algorithms for representing decision-making strategies of agents and for pre-processing data

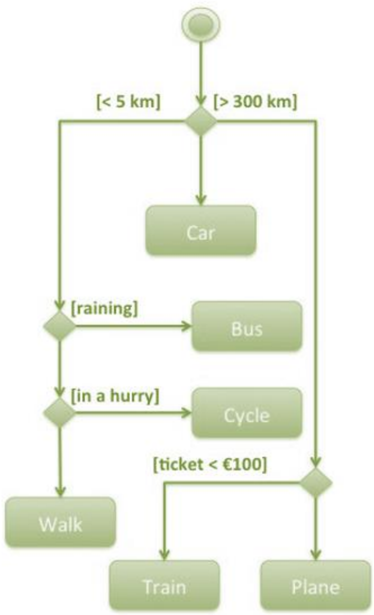
5.1.1 Decision making

Decision trees are data structures comprising a series of nested conditional expressions, each of which is a node in the tree. Outcomes are represented as 'leaf' nodes in the tree. Directed edges connecting nodes correspond to whether the conditional expression at the start of the edge evaluates to 'true' or 'false'. An example related to mobility is illustrated by the Figure 5. Various machine learning algorithms can be used to construct decision trees from data, but the C4.5 algorithm is probably one of the best known (Quinlan 1993).

In the context of social sciences, decision trees are a way to empirically construct simple agent decision-making algorithms from a suitably designed questionnaire. Sánchez-Maroño et al. (2015), for example, document a case study, in which the questionnaire features questions assessing psychological constructs and questions related to frequencies, with which certain behaviours of individuals are performed. Using the psychological constructs (and demographic variables) as explanatory variables, and the behaviours as response variables, decision trees that aim to predict behaviour given psychological and demographic features of individuals can be constructed.

One of the main advantages of decision trees is their transparency, that allows checking if the derived decision trees are theoretically consistent with the knowledge of the experts (e.g. psychologists and sociologists). However, if the transparency requirement is not necessary, and more data is available, there are other machine learning algorithms that may provide a more precise behaviour of the agent, such as is the case for artificial neural networks (ANN; Schalkoff 1997) or support-vector machines (SVM; Cortes & Vapnik 1995). In any case, the objective of both previous techniques is to establish a correspondence, or derive a function f, that relates a set of inputs (explanatory variables), x (x1, x2, x3, … xd), and a set of outputs (response variables) t (t1, t2, t3, … tc). There are different types of algorithms that allow learning this relationship, and many of them are categorized as supervised learning. In supervised learning, the correspondence between inputs and outputs is known for N patterns, that is, the output pattern (t1n, t2n, t3n, … tcn) is associated with the input pattern (x1n, x2n, x3n, … xdn). Once this correspondence has been learned by the algorithm from a training (data)set, the algorithm works by establishing a previously unknown output for new cases not included in the training set. For the learning procedure to be accurate, the sample comprising the training set should be as wide and varied as possible (in features/variables and samples/cases). If N has a small size, the function learned is not assured to behave correctly. In cases, in which the "general" pattern was not represented in the training set, the output result is not guaranteed to be adequate (poor generalization capabilities). The same problem occurs if we have a large amount of data samples but all of them represent the same type of data (poor variability).

Figure 5. Example of decision tree for an agent type in a superblock.



5.1.2 Data pre-processing

In cases where it will not be possible to collect enough data, different machine learning techniques may be used to pre-process the available data and make the set *N* suitable to learn the correspondence *f*. Sánchez-Maroño et al. (2017) investigate various uses of clustering, feature selection and discretization to develop decision-trees with acceptable expected generalization abilities. Clustering can simplify multidimensional cardinal spaces into a series of classes. Sánchez-Maroño et al. (2017) use it to identify categories of respondents in a questionnaire, based on their answers to the standard questions on Schwartz's values (1992). They then develop decision trees for performing sustainable behaviours using a separate calibration process for each of the identified categories. Also, feature selection can facilitate the construction of decision trees by reducing the number of explanatory variables employed. There are iterative and single-shot feature selection algorithms, the former exploring the space of variables to include iteratively with the decision tree algorithm, using strategies that increase the probability of finding the variables that yield a decision tree with the best estimated generalization ability. Single-shot feature selection algorithms use statistical techniques to recommend variables with the best chance of generating high-quality decision trees.

Machine learning algorithms (specifically decision trees and, at a less extent, ANN) do not necessarily use all of the variables available for the problem, even when used in combination with feature selection algorithms. As contexts change for agents, it is possible that variables ignored by feature selection or machine-learning algorithms would ideally have been included. This could occur if, for example, a variable was not a notable driver of behaviour at the time of the questionnaire, but becomes so later on, due to changing environmental or social factors. Ideally, when using questionnaire data, there would be a significant volume of data taken from diverse contexts to ensure that generalities are captured. This point is even more valid in situations relevant to social innovation.

5.2 Elements of a data-driven model

Whenever possible, a data-driven ABM will include three key components: a realistic population, a social contact network among the individuals in the population, and the environment.

5.2.1 Realistic population

The population could be generated with demographic attributes and household structure consistent with documentary data (census data, for instance). However, the population would have to take decisions, therefore data collection to obtain data regarding the individual factors that affect agents' decisions is planned in SMARTEES. The idea is to simulate the behaviour of persons involved in each case of study, according to the tasks they perform and the available options to be implemented. For instance, considering the case study of superblocks and focusing on mobility, we will need to gather data that could be relevant to determine the selection of a car as means of transport among others available for commuting.

5.2.2 Social networks

The potential of ABM is in the direct representation of each of the actors in a social system and their behaviours, as actors in their social and/or physical environment. Within an ABM, agents must interrelate and change their behaviour in time, based on the observation of the behaviour of others and/or the environment. To explicitly represent a relationship between two agents it is necessary to add a link between them: an agent will have several links, one for each other agent to which it is related. In that sense, the whole society will be interconnected generating a social network. When developing agent-based models is important to consider the following three points:

- Creation of the initial social network: to create links between different agents, i.e., to establish the network topology,
- Evolution of the social network: in real life the strengths of ties among people change: in some cases, links can be broken completely, whilst in others new links are created, often via existing relationships (e.g. friend-of-a-friend). Consideration of this fact in the model entails the design of mechanisms to evolve network (break and create links) over time. Similarity is one of the identified key drivers in connections.
- Influence of social network: agents make use of these links to socially interact, share views and opinions and convey habits of behaviour between them. We define as social influence the way in which an agent perceives the behaviour of others, modifies their own, and seeks to modify that of others using these links.

It will be necessary to collect data in these three points, using the appropriate means, to design a model that reflects the real topology of the network.

5.2.3 The environment

Finally, it is important to develop the environment (context), in which these agents interact and make decisions. This requirement implies the need to collect information about environment. Following the example of the superblocks (Figure 5), it is profitable to know if there is a bike lane for the bike to be a transport option, or to know the distances an agent has to commute (if it is a large distance and no train is available, car is the only option). Note that the bike lane information can be obtained through document analysis, but the distance to commute might be subject to primary data collection. Moreover, when collecting relevant information, one should plan the process so that relevant goals of the study are addressed, e.g. one should consider not only a current situation, but, whenever relevant, also the past events or future scenarios.

6 GUIDELINES FOR DATA COLLECTION

6.1 Tailoring theoretical concepts and data collection processes to modelling cases of social innovation

To better understand why social innovations in energy transition succeed or fail, different theoretical angles and associated data have to be combined to derive at a more complete picture of the dynamics of change in communities. As a consequence, within the SMARTEES project we adhere to an integrated mixed methodology approach, where different theoretical approaches are combined with different data (ranging from hard behavioural data to qualitative narratives) to develop an understanding of these transformational dynamics in communities. Feeding theoretical and empirical insights together with data into agent-based models is critical in constructing an inherently dynamic perspective on processes of change, especially in understanding how critical events may impact the courses of developments within a community. By combining the conceptual modelling frameworks of FEARLUS and CONSUMAT with other relevant theoretical approaches, in particular on networked social influences and habitual behaviour, we offer modellers readily available, coherent sets of assumptions to choose from when modelling a specific case. Whereas the basic principles of social innovations will be relevant for all cases, it is our understanding that each case of social innovation has unique properties and may require a specific application of a particular theoretical and/or empirical framework, such as attitude formation and opinion dynamics, habitual structures or networked cooperation. Therefore, our aim is to create an integrated, general and flexible theoretical approach that allows for focussing on those behavioural processes, networks and theories that are relevant for the development of the different case models. This integrated theoretical approach serves in informing data collection through: (1) identifying relevant research questions, (2) proposing valid indicators for phenomena of interest, (3) aiding in identifying relevant existing data sources, and (4) guiding primary data collection via helping in diagnosing gaps in secondary data.

6.2 Utilizing secondary data to the greatest extent possible

The devised data collection strategy emphasizes the possibilities of using secondary data (qualitative and quantitative) to the greatest extent possible. Recreating histories of successful innovations can utilize existing quantitative data collected in the process of representative research, e.g. international surveys such as European Social Survey, European Quality of Life Survey, EU-Silk, or data provided by national statistical offices. International surveys, containing indicators for both reference and follower cases may prove a valuable source of information used to parametrize agent-based models. Moreover, several international surveys are realized as panel studies or are repeated for a number of times with the use of representative sampling schemes, which enables tracking dynamics of certain processes. In the process of identifying relevant quantitative data sources it is important that the data is available on the level of the modelled entity (i.e. city, island, municipality). Please note that, whenever necessary, all applications for external data should be centralized in SMARTEES. If a partner identifies a database, which contains valuable indicators, there are chances that other partners might find valuable data for the cases that they are working on. Sharing information on intent to apply for data sources with partners assures that application only takes place once. This central application for access to data and data-bases will we coordinated by WP1.

We also emphasize that qualitative data could be a valuable source of information, especially in cases when the social innovation was popularized in a relatively distant past, when surveys were not as popular as they are now. Qualitative data are often available as narratives from the period when the social innovation was in a decisive phase. For example, stories on public debates, the role of different opinion leaders in the community and collaborations, and responses to public policy can be found in articles in journals or on websites. Qualitative data may also compose of documentations of the social innovation popularization process. For example, available documentation of the public intervention archived by the municipality (such as relevant legal documents in power when intervention was implemented, notes from meetings, strategies, plans, previous evaluation reports), local/national newspaper articles that describe the process of change, or previous scientific or journalist analyses of social innovation. Whenever relevant for research questions, information that may contribute to the modelling of social networks that capture the role and position of opinion leaders is of special interest.

Independently of the type of secondary data, it is important to assure proper data documentation and storage. Also, please note that any primary data collection should respond to gaps identified by analysing secondary data sources, and, to the greatest extent possible, should make use of already existing indicators for the purposes of comparative analyses.

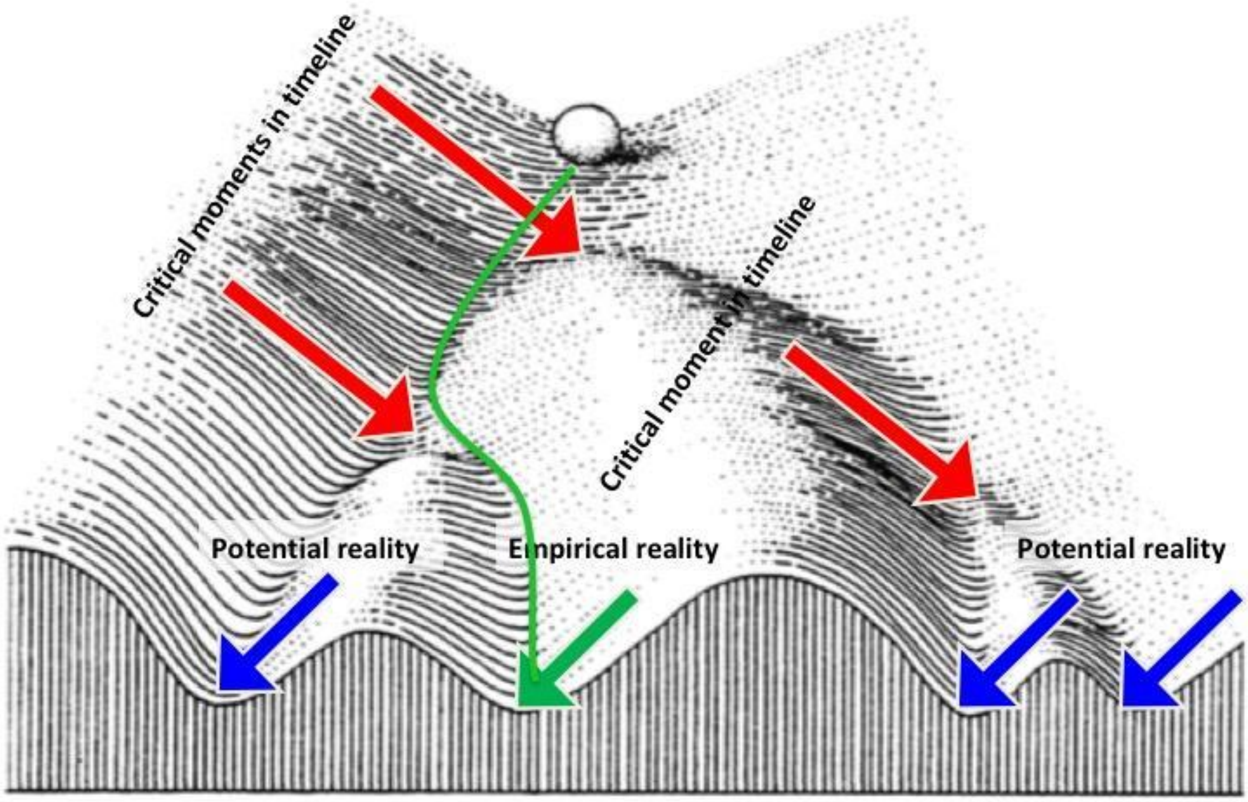
6.3 Timeline: organising case data on dynamics and transformations

SMARTEES aims to analyse and model cases that demonstrate social innovation processes over time. Hence organising data along a timeline seems to be a sensible backbone for the data collection & organisation. Data such as cross-sectional surveys is explicitly time-specific, identifying the state of affairs in a defined point in time, whereas qualitative data on specific narratives may address time-intervals. This data variety emphasizes the importance of defining secondary data with respect to time and of ascribing time stamps to any primary collected information (e.g., designing IDI protocols with respect to time). Moreover, we may acquire different types of data for different moments in time. In practice, already existing and newly acquired cross-sectional data points (e.g. questionnaire data, regular measurements of observed behaviour, community demographics, policy implementation documents), and fuzzy (and perhaps conflicting) narratives about how the process developed will have to be combined. Narratives from different people may reveal the different responses to this policy in terms of changing outcomes (needs) and adapting behaviour. Hence, following the case timeline, different types of data will have to be placed in relation to one another.

If our case simulation models succeed in recreating narratives present in investigated success stories this may help to identify tipping points, i.e. events that were crucial in influencing whether the community behaviour transforms. However, not all transformations will take a critical form, when a system changes to an alternative stable state (Scheffer et al. 2009). It is also possible, that the changes are non-critical, when resilience in the system forces a degree of continuity (Walker et al. 2004). The critical moments in the process of change deserve special attention in SMARTEES. At what time decisions were made and implemented? Who was involved? What behavioural and opinion dynamics happened in the community? Were there disagreements or protests? Answers to those questions can enable modelling of opinion leaders. Hence, in the narratives it is of critical importance to identify the people that made a difference at decisive moments, i.e. who these people were, what position did they have in the (informal) community network, and what their formal networks were. Identifying critical tipping point moments will also allow for learning about the factors influencing chances of success or failure at those critical moments, in particular with respect to the emergence of social support and/or resistance. To simulate under what conditions the empirical reality could have resulted in an alternative (potential) reality we need a variety of data sources describing the critical periods in the case. This will help us both parameterising the conditions as validating the processes as observed in the case.

In Figure 6, the timeline of the empirical reality that unfolded in the cases can be seen as the green line. In more traditional research, statistical models would be developed to fit this “data-line” as closely as possible. On the basis of such models, predictions are often made concerning future developments. However, once we realise that the tipping points in the systems could potentially have resulted in a different empirical reality, it can be understood that statistical approaches do not serve the purpose of identifying landscapes of potential, counterfactual developments. As statistical models do not capture the complex dynamics critical at the tipping points, they do not allow for exploring the sensitivity of the case for the social complexity in many processes of innovation. In SMARTEES we can use agent based models to explore the social dynamics and critical tipping points to account for this possibility of multiple outcomes. Hence in the data collection process it is of special interest to acquire data/narratives that shed a light on the volatilities and chance events that played a role in the development of the case in a particular direction.

Figure 6. Critical moments determine which of the potential realities turns into an empirical reality (background landscape from Waddington's (1957) classic model of an epigenetic landscape).



Transformations of social systems to a different state may be preceded by *early warning signals* (EWS, hereafter), reflecting increased sensitivity of a complex system to external or internal perturbation. Previous research shows that EWS, which can occur before critical and non-critical transformations of social systems, may be indicated by increased variances of variables, reflecting amplifications of small shocks as the system approaches critical transition (Spielmann et al. 2016). Identifying EWS and tipping points requires that collected data reflects not only the process of intervention and its effects, but defines the beginning of the transition process earlier.

6.4 Starting point: Descriptive models

Investigating the spread of social innovations in different communities is inevitably related to the question of how social norms, i.e. predominant behavioural patterns within a group, supported by a shared understanding of acceptable actions and sustained through social interactions within that group (Ostrom, 2000), change. With this respect, communities can be segmented into distinctive roles of e.g. leaders/promoters, followers and opposition. Each of those groups has unique motives to (not) carry out specific behavioural patterns in a given sequence. Theory on innovation diffusion (Rogers 2003) describes how a critical mass of connected people adopting a new behaviour can spread a norm change through a social network. Moreover, public interventions may stimulate tipping points by various means (e.g. providing new behavioural opportunities, decreasing utility of old behaviours or enhancing visibility of preferred behavioural patterns).

To achieve the goals of SMARTEES, we start with building theoretically supported conceptual models of selected cases of social innovations. These conceptual models are driven by relevant theoretical approaches and available empirical data. Translating the theories of cases into computational models will pose several challenges, so we should be prepared for a situation where the implemented model reflects the intended theory imperfectly. Nonetheless we propose, that modelling efforts are guided by the principles of EROS (Enhancing Realism Of Simulations, Jager 2017) and KIDS (Keep It Descriptive Stupid, Edmonds & Moss 2005), which focus on starting with a descriptive model (which may be quite complex) and, in the process of modeling, simplifying it where this turns out to be justified. Such an approach has two important advantages. First, it allows for making as much use as possible of various types of data (individual anecdotal stories, expert opinions, distribution of population characteristics, classifications provided analytical techniques (e.g. decision trees)) for building the conceptual models of social innovation and of public interventions, formalizing and validating them. Second, it enhances understanding of the models among policy makers/implementers and other interested stakeholders, who do not have a modelling background. A key challenge in SMARTEES is therefore the development of an artificial population that is (1) conceptually rich enough and empirically validated to conduct meaningful scenario experiments on, and (2) simple enough to offer a transparent view on the processes leading to developments in the energy domain, and the mechanisms behind the dynamic effect of certain policies.

7 METADATA

7.1 Metadata types

Metadata is data to obtain information about other data ("Definition of Semantic" 2018a), and has been in use since about 300 BCE when used to organize scrolls at the library of Alexandria (Gartner 2016). The reason we mention this is to show that metadata has always been a requirement from the beginning of organized data, and always will be necessary, even though most see it as unnecessarily burdensome. This kind of data arises for any electronic project artefacts such as a model, website, webpage, document, or data set; in fact, more generally, any kind of data that can be stored electronically. We refer to this collection as the project's corpus. This could be extended to all hard-copy documentation as well, and other physical artefacts, but this is probably impractical, and moreover the vast majority of the meaningful artefacts of the project will have some form electronic representation, if only for facilitating the access to such entities for the purposes of study to geographically dispersed individuals.

There are two key aims of collecting such metadata. These are:

- provenance, and
- project semantics.

Provenance is the source and history of an object (Freire et al. 2008). Any meaningful manipulation of the project corpus, such as the automatic extraction of glossary and key-terms is made extremely difficult without proper provenance metadata. Without such metadata, then even simple tasks such as identifying the correct, current and principal version of a given document can be incredibly difficult. In GLAMURS, the only way such documents could be identified was by human judgement. This may seem the obvious solution, rather than providing what can be onerous metadata, but when there are thousands of documents in a project (as is typically the case), this task becomes somewhat overwhelming for the knowledge engineers and moreover there is still no guarantee that the latter will select the correct document, especially if a revision includes minimal changes, or ambiguous changes such as several deletions.

Such information is needed for the purposes of analysing the contents of the corpus of a project, which is used to identify key terminology and key relationships between such terminology, and possibly project effectiveness by the development of indicators from the ordered collection of the corpus. We have demonstrated in GLAMURS that the analysis of such metadata allows the semi-automated linking of specialist vocabularies between experts' sets of vocabularies. These may be thought of as dictionaries between areas of expertise. Without such data, any terminology or glossary will be based on the judgement and subjective view of the compilers of sets of terminology and glossary, based on their selection and assessment of the seeming importance and clarity of such terms. What has obvious meaning to one expert does not necessarily contain meaning for others. This inevitably leads to the inclusion of the compilers' predilections such as, for example a bias towards the compilers' disciplines.

Some provenance metadata can be collected automatically by modern content management systems such as Sharepoint. However, it is important that metadata collection facilities are enabled in the software, and moreover that the project's standards for metadata insist on the correct use of such facilities in the software, and that team members adopt the standards. In addition, provision needs to be made to monitor artefacts that are not curated by the project's content management system; such as web-sites or the code repositories, data sets in data repositories, media stored electronically, and actual software such as instances of models. This will probably necessitate the development of automatic methods. For instance, this could be done with read access to the artefact, then "pulling" metadata using scripting, and then automatically generating any additional metadata required where this can be inferred, and finally storing the obtained metadata in a provenance database. However, an activity such as this needs to be coordinated, not only from the point of developing such a facility, but also the discipline of change management for project resources needs to be enhanced such that it should be possible to set up procedures to reliably notify those curating the metadata, or the scripts "pulling" such artefacts. This might be done automatically by monitoring scripts, but more realistically needs coordination between anybody making a release, the management of the project and those responsible for curating the project metadata.

We contend that the second set of the metadata required for a project is the project semantics. Semantics is generally taken to be the study of meaning (“Definition of Semantic” 2018b), but it is in the usual computer-science sense of the term that we use the idea of semantics. These semantics are the identification of meta-behaviours, i.e. rules that govern the interaction between sets of system regularities such as formal structures (Euzenat & Shvaiko 2007). The usual sense of semantics, the study of meanings, is somewhat catered for by manipulation of the project's provenance metadata, described above, in order to provide glossary and terminology translation, or relationships between concepts. As stated, the aim of the elicitation of the project semantics is to elicit meta-behaviours. It is somewhat suggestive that in some studies of information theory, such meta-behaviours are called "theories" (Barwise and Seligman 1997). These meta-behaviours, optimistically and ambitiously may represent emergent behaviours of the project in its entirety. That is, in the best-case scenario, such a study would uncover non-obvious rules about the interactions among team members, disciplines and stakeholders within the SMARTEES project.

In order to identify interactions among regularities, then these regularities first have to be identified. These will consist of the assumptions and premises of the project. Actors within the project will also have to be identified, along with their properties. For instance, a lot of the modelling in the SMARTEES project is to be agent-based. Thus, it seems evident that beyond the usual provenance metadata, then types of agents will have to be identified. One example is the question, "Is an agent an organization or a person?" Properties of these agents also have to be specified, e.g. most people have a name, sex and age; most organizations have a name and age, but (debatably) not a sex. It becomes apparent that definition of such regularities will evolve as the project proceeds, and hence not be known in advance. Any such initial framing of such metadata is thus provisional and will be constantly refined by a continual iteration of revision, addition and deletion to the specification of such metadata until the end of the project. It is hoped that in this manner, the metadata not only record the structure of the project but actively aids in framing theoretical conceptualisation within the project, and as an added bonus self-documents the SMARTEES project.

Each of the two primary aspects of the project's metadata is shortly described in Appendix 1.

7.2 Data curation

All data curation procedures and the choice of software used in various stages of the project should facilitate royalty-free availability of foreground data and knowledge to all partners for the duration of the project. To ensure equal access of all the partners to data and analytical results, and to provide possibilities of performing additional analyses of gathered materials at later stages of the project, whenever possible, chosen analytical software should be non-proprietary, with open source licences. As SMARTEEES implements a multi-method approach, several software packages, appropriate for various data types and analytical purposes, will be used. For data formats, non-proprietary, text-based formats should be used as much as possible. All files will be named according to the rules described in the Data Management Protocol (section 4.4.5 File name standards).

8 REFERENCES

Aamodt, A. and Plaza, E. (1994) Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Communications* **7** (1), 39-59.

Anand, P. (1993) *Foundations of rational choice under risk*. Oxford University Press, Oxford.

Bandura, A. (1977) *Social learning theory*. Prentice Hall, Englewood Cliffs, New Jersey.

Barwise, J. and Seligman, J. (1997). *Information Flow: The Logic of Distributed Systems*. Cambridge University Press, New York.

Boero, R. and Squazzoni, F. (2005) Does empirical embeddedness matter? Methodological issues on agent-based models for analytical social science. *Journal of Artificial Societies and Social Simulation* **8** (4), 6. http://jasss.soc.surrey.ac.uk/8/4/6.html

Boettiger, C. and Hastings A. (2012) Quantifying limits to detection of early warning for critical transitions. *Journal of the Royal Society Interface* **9** (75), 2527–2539. (doi: 10.1098/rsif.2012.0125)

Bravo, G., Vallino, E., Cerutti, A. K. and Pairotti, M. B. (2013) Alternative scenarios of green consumption in Italy: An empirically grounded model. *Environmental Modelling and Software* **47**, 225-234. (doi:10.1016/j.envsoft.2013.05.015)

Breiman, L., Friedman, J., Olshen, R. A. and Stone, C. J. (1984) *Classification and Regression Trees*. Chapman & Hall/CRC, Boca Raton, Florida.

Byrne, D. (1961) Interpersonal attraction and attitude similarity. *The Journal of Abnormal and Social Psychology* **62** (3), 713-715 (doi:10.1037/h0044721).

Cederman, L.-E. (2003) Modeling the size of wars: from billiard balls to sand piles. *American Political Science Review* **97** (1), 135-150.

Cioffi-Revilla, C., Gotts, N. (2003) Comparative analysis of agent-based social simulation: GeoSim and FEARLUS models. *Journal of Artificial Societies and Social Simulation* **6** (4), 10. http://jasss.soc.surrey.ac.uk/6/4/10.html

Cortes, C., Vapnik, V. N. (1995). Support-vector networks. *Machine Learning* **20** (3), 273–297. (doi:10.1007/BF00994018).

Dai, L., Vorselen, D., Korolev, K. S. and Gore, J. (2012) Generic Indicators for Loss of Resilience Before a Tipping Point Leading to Population Collapse. *Science* **336** (6085), 1175–1177. (doi: 10.1126/science.1219805)

Dakos, V., Carpenter, R. A., Brock, W. A., Ellison, A. M., Guttal, V., Ives, A. R., Kéfi, S., Livina, V., Seekell, D. A., van Nes, E. H. and Scheffer, M. (2012) Methods for Detecting Early Warnings of Critical Transitions in Time Series Illustrated Using Simulated Ecological Data. *PLoS ONE* **7** (7), e41010. (doi:10.1371/journal.pone.0041010)

“Definition of Semantic.” 2018a. https://www.merriam-webster.com/dictionary/metadata.

Edmonds, B. and Moss, S. (2005) From KISS to KIDS – an 'anti-simplistic' modelling approach. In Davidsson, P., Logan, B. and Takadama, K. (eds.) *Multi Agent Based Simulation 2004. Springer, Lecture Notes in Artificial Intelligence*, 3415, pp. 130–144.

Ellison, G. and Fudenberg, D. (1995) Word-of-mouth communication and social learning. *The Quarterly Journal of Economics* **110** (1), 93–125. (doi:10.2307/2118512)

Euzenat, J. and Shvaiko, P. (2007). *Ontology Matching*. Springer, New York.

Freire, J., Koop, D., Santos, E., and Silva, C. T. (2008). Provenance for Computational Tasks: A Survey. *Computing in Science & Engineering* **10** (3), 20-30. (doi:10.1109/MCSE. 2008.79)

Gartner, R. (2016). *Metadata. Shaping Knowledge from Antiquity to the Semantic Web*. Springer, Basel.

Gimona, A. and Polhill, J. G. (2011) Exploring robustness of biodiversity policy with a coupled metacommunity and agent-based model. *Journal of Land Use Science* **6** (2-3), 175-193. (doi:10.1080/1747423X.2011.558601)

Gotts, N. M. and Polhill, J. G. (2009) When and how to imitate your neighbours: Lessons from and for FEARLUS. *Journal of Artificial Societies and Social Simulation* **12** (3), 2. http://jasss.soc.surrey.ac.uk/12/3/2.html

Gotts, N. M. and Polhill, J. G. (2010) Size matters: large-scale replications of experiments with FEARLUS. *Advances in Complex Systems* **13** (4), 453-467. (doi:10.1142/S0219525910002670)

Gotts, N. M., Polhill, J. G. and Law, A. N. R. (2003) Aspiration levels in a land use simulation. *Cybernetics and Systems* **34** (8), 663-683. (doi:10.1080/716100277)

Iturrioz, G. and Polhill, G. (2014) Agent-based modelling of rural land use changes in the Atreucó area. In Ames, D. P., Quinn, N. W. T., and Rizzoli, A. E. (eds.) *Proceedings of the 7th International Congress on Environmental Modelling and Software, June 15-19, San Diego, California, USA*. ISBN: 978-88-9035-744-2.

Jager, W. (2000) *Modelling consumer behaviour*. Thesis Behavioural and Social Sciences, University of Groningen.

Jager, W. (2017) Enhancing the Realism of Simulation (EROS): On Implementing and Developing Psychological Theory in Social Simulation. *Journal of Artificial Societies and Social Simulation* **20** (3), 14. http://jasss.soc.surrey.ac.uk/20/3/14.html

Jager, W., Janssen, M. A., De Vries, H. J. M., De Greef, J., and Vlek, C. A. J. (2000) Behaviour in commons dilemmas: Homo economicus and Homo psychologicus in an ecological-economic model. *Ecological Economics* **35** (3), 357-379. (doi:10.1016/S0921-8009(00)00220-2)

Janssen, S., Athanasiadis, I. N., Bezlepkina, I., Knapen, R., Li, H., Domínguez, I. P., Rizzoli, A. E., and van Ittersum M. K. (2011) Linking models for assessing agricultural land use change. *Computers and electronics in agriculture* **76** (2), 148-160. (doi:10.1016/j.compag.2010.10.011)

Kangur, A., Jager, W., Verbrugge, R. and Bockarjova, M. (2017) Hybrid or full electric, that's the question: An agent-based model exploring the diffusion of electric vehicles. In Jager, W. and Ernst, A. (eds.) *Special issue Social*

Simulation in Environmental Psychology, Journal of Environmental Psychology **52**, 166-182. (doi:10.1016/j.jenvp.2017.01.002)

Knuth, D. E. (1984) Literate Programming" (PDF). *The Computer Journal. British Computer Society* **27** (2), 97–111.

Lindenberg, S., Steg, L. (2007). Normative, gain and hedonic goal frames guiding environmental behaviour. *Journal of Social Issues* **63** (1), 117–137. (doi:10.1111/j.1540-4560.2007.00499.x)

Loorbach, D., Avelino, F., Haxeltine, A., Wittmayer, J., O'Riordan, T., Weaver, P. and Kemp, R. (2016) The economic crisis as a game changer? Exploring the role of social construction in sustainability transitions. *Ecology and Society* **21** (4), 15. (doi:10.5751/ES-08761-210415)

McPherson, M., Smith-Lovin, L., and Cook, J. M. (2001) Birds of a feather: Homophily in social networks*. Annual review of sociology* **27**, 415-444. (doi:10.1146/annurev.soc.27.1.415)

Nyborg, K., Anderies, J. M., Dannenberg, A., Lindahl, T., Schill, C., Schlüter, M., Adger,W. N., Arrow, K. J., Barrett, S., Carpenter, S., Chapin III, S. F., Crépin, A. S., Daily, G., Ehrlich, P., Folke, C., Jager, W., Kautsky, N., Levin, S. A., Madsen, O. J., Polasky, S., Sche_er, M.,Walker, B., Weber, E. U., Wilen, J., Xepapadeas, A. and De Zeeuw, A. (2016) Social norms as solutions: Policies may influence large-scale behavioral tipping. *Science* **354** (6308), 42–43. (doi:10.1126/science.aaf8317)

Ostrom, E. (2000) Collective Action and the Evolution of Social Norms. Journal of Economic Perspectives **14** (3), 137-158. (doi:10.1257/jep.14.3.137)

Polhill, J. G., Filatova, T., Schlüter, M. and Voinov, A. (2016) Modelling systemic change in coupled socio-environmental systems. *Environmental Modelling and Software* **75**, 318-332. (doi:10.1016/j.envsoft.2015.10.017)

Polhill, J. G., Gimona, A. and Gotts, N. M. (2013) Nonlinearities in biodiversity incentive schemes: A study using an integrated agent-based and metacommunity model. *Environmental Modelling and Software* **45**, 74-91. (doi:10.1016/j.envsoft.2012.11.011)

Polhill, J. G., Gotts, N. M. and Law, A. N. R. (2001) Imitative versus nonimitative strategies in a land-use simulation. *Cybernetics and Systems* **32** (1-2), 285-307. (doi:10.1080/019697201300001885)

Polhill, G., Milazzo, L., Dawson, T., Gimona, A., and Parker, D. (2017) Lessons learned replicating the analysis of outputs from a social simulation of biodiversity incentivisation. In Jager, W., Verbrugge R., Flache A., de Roo G., Hoogduin L. and Hemelrijk C. (eds.) *Advances in Social Simulation 2015*, pp. 355-365. Springer, Cham.

Quinlan, R. (1993) *C4.5: Programs for Machine Learning*. Morgan Kaufman, San Francisco, CA, USA.

Rogers, E.M. (2003) *Diffusion of Innovations*. Free Press, New York.

Rosenbaum, M. E. (1986) The repulsion hypothesis: On the nondevelopment of relationships. *Journal of Personality and Social Psychology* **51** (6), 1156-1166. (doi:10.1037/0022-3514.51.6.1156)

Sánchez-Maroño, N., Alonso-Betanzos, A., Fontenla-Romero, O., Brinquis-Núñez, C., Polhill, J. G., Craig, T., Dumitru, A. and García-Mira, R. (2015) An agent-based model for simulating environmental behavior in an educational organization. *Neural Processing Letters* **42** (1), 89-118. (doi:10.1007/s11063-014-9390-5)

Sánchez-Maroño, N., Alonso-Betanzos, A., Fontenla-Romero, O., Polhill, J. G. and Craig, T. (2017) Empirically-derived behavioral rules in agent-based models using decision trees learned from questionnaire data. In Alonso-Betanzos, A., Sánchez-Maroño, N., Fontenla-Romero, O., Polhill, J. G., Craig, T., Bajo, J. and Corchado, J. M. (eds.) *Agent-Based Modeling of Sustainable Behaviors*. Cham, Switzerland: Springer. pp. 53-76. (doi:10.1007/978-3-319-46331-5_3)

Schalkoff, R. (1997) *Artificial neural networks*. McGraw-Hill, New York.

Scheffer, M., Bascompte, J., Brock, W.A., Brovkin, V., Carpenter, S.R., Dakos, V., Held, H., van Nes, E, H., Rietkerk, M. and Sugihara, G. (2009) Early-warning signals for critical transitions. *Nature* **461** (7260), 53–59. (doi:10.1038/nature08227)

Schlüter, M., Baeza, A., Dressler, G., Frank, K., Groeneveld, J., Jager, W., Janssen, M. A., McAllister, R. R. J., Müller, B., Orach, K., Schwarz, N. and Wijermans, N. (2017) A framework for mapping and comparing behavioural theories in models of social-ecological systems. *Ecological Economics* **131**, 21–35. (doi:10.1016/j.ecolecon.2016.08.008)

Schwartz, S. H. (1992) Universals in the content and structures of values: theoretical advances and empirical tests in 20 countries. *Advances in Experimental Psychology* **25**, 1-65. (doi:10.1016/S0065-2601(08)60281-6)

Spielmann, K. A., Peeples, M. A., Glowacki, D. M. and Dugmore, A. (2016) Early Warning Signals of Social Transformation: A Case Study from the US Southwest. *PLoS ONE* **11** (10): e0163685. (doi:10.1371/journal.pone.0163685)

Van Duinen, R., Filatova, T., Jager, W. and Van der Veen, A. (2015) Going beyond perfect rationality: Drought risk, economic choices and the influence of social networks. *The Annals of Regional Science* **57** (2-3), 335-369. (doi:10.1007/s00168-015-0699-4)

Van Eck, P. S., Jager, W. and Leeflang, P. S. (2011) Opinion leaders' role in innovation diffusion: A simulation study. *Journal of Product Innovation Management* **28** (2), 187-203. (doi:10.1111/j.1540-5885.2011.00791.x)

Waddington, C. H. (1957). *The Strategy of Genes: a discussion of some aspects of theoretical biology*. Allen & Unwin, London.

Walker B., Holling, C. S., Carpenter, S. R. and Kinzig, A. (2004) Resilience, adaptability, and transformability in social- ecological systems. *Ecology and Society* **9** (2), 5. http://www.ecologyandsociety.org/vol9/iss2/art5/

Wood, W., and Ruenger, D. (2016) Psychology of habit. *Annual Review of Psychology* **67**, 289-314. (doi:10.1146/annurev-psych-122414-033417)

APPENDIX 1

The project requires two sets of distinct metadata. The first describes how and by whom electronic documents have been created and modified. This is crucial in verifying data sets are what they are believed to be and contain what they should. This data is also absolutely fundamental and a requirement for reproducibility of experimental results. This data also allows the construction of project terminology (glossary). The second set of data is what we have denoted the "project-semantics" of the SMARTEES projects. These are classifications and relationships between those classifications, that are unique to the SMARTEES project, and will provide the common framework in which models and case-studies will be able to interact.

This appendix is divided into two sections. The first details the minimum metadata required to establish the modification and history of an electronic resource. This is referred to as *provenance* metadata. We have also specified an extended set of metadata, which would be the ideal metadata specification for such documents and would allow very accurate curation of any electronic datasets and documents. The collection of such comprehensive metadata would represent the ideal. The second set of metadata describes a very simple set of metadata required to describe basic classifications, specific to the SMARTEES project and the simple relationships between such categories. This an initial specification and should be considerably greater in size and complexity by the end of the project. This specification should evolve and grow predicated by increase in knowledge of the case-studies and the results of experiments conducted.

Provenance metadata

We suggest the following metadata always be recorded for any electronic artefact produced by the SMARTEES project. The below minimally guarantees that it is easy to identify "the" current version of a document,

- Title
- Created by - this could be a person, group, organization or process
- Creation date
- Data Sensitivity indicator and responsibility of a person, group, organization or process
- Previous Version (if extant) - there may be zero, one or more of these.
- Modified by (if applicable) of a person, group, organization or process
- Modified when (if applicable) - some kind of time specification
- Dependent upon (if applicable) - there may be zero, one or more of these.

The above also implies agency, so we will also have to allow for agents and processes.

We should be looking to map directly to use existing and accepted standards such as the PROV-O ontology (Belhajjame et al. 2013) and the Dublin core metadata element set for documents (DCMI 2012), rather than re-inventing vocabularies or ontologies to do so. Table 1A below shows a mapping of each attribute above to one entity in each of these networks. IRI, international resource identifier is the internationalized version of URI, which is a uniform resource locator. This uniquely specifies an entity on the Internet - the Internet being a proper superset of the World Wide Web. This table represents the absolute minimal amount of provenance data we should be collecting. Without what we have below then it will become increasingly difficult to consistent identify primary documentation. This will not only apply to automated recognition, but given our experience in previous projects, it even becomes difficult for humans to identify.

Table 1A. Minimum provenance metadata required.

Property	Dublin Core	PROV-O	Domain	Range
Agent	IRI	prov:Agent	NA	NA
Process	NA	prov:Activity	NA	NA
Document	IRI	prov:Entity	NA	NA
Title	dc:title	NA	Agent/Document	NA
Created by	dc:creato r	prov:wasGeneratedBy	prov:Entity	prov:Activity
Creation date	dc:date	prov:generatedAtTime	prov:Entity	NA
Data Sensitivity indicator	NA	NA	prov:Entity	N/A
Responsibility of	IRI	NA	prov:Entity	prov:Agent
Previous Version	IRI	prov:wasDerivedFrom	prov:Entity	prov:Entity
Modified by	IRI	prov:wasAttributedTo	prov:Entity	prov:Activity
Modified when	dc:date	prov:sttarted	prov:Entity	N/A

Dependent upon	dc:source	prov:used	prov:Activity	prov:Entity

Table 1A above represents the minimum we believe should be necessary to ensure that provenance metadata within the project is correctly and sufficiently tracked.

The correspondence in the Table 1A is not one-one. The provenance ontology, PROV-O tends to use "3-relations" (Sowa 1999) and reifications of these in that an activity will link to documents by reification, but this activity will have its own particular properties such as times, and who or what performed the activity. It would be our recommendation that this PROV-O model of change takes precedence over the Dublin core. Even though it is initially more difficult to understand, it offers a much greater degree of freedom when describing provenance.

Indeed, the PROV-O ontology and Dublin-core vocabulary are a great deal richer than the subset of terms and entities we have used above. However, for the purposes of clarity and ease of use we have tried to illustrate the absolute minimum number of entities and the properties that are required, for consistent and reliable primary document identification.

It should be noted that there are fields above which are not part of the standard ontologies and vocabularies utilised herein. This is the "data sensitivity indicator" and the "responsibility of" fields. We have introduced these fields as part of awareness for GDPR requirements. Some data are sensitive in that they must not be published in open environments, such as web forums, and this flag would provide a means of verifying that such data is not likely to break the law if it were so published. Moreover, we will make explicit the who is responsible for such data. Although these fields are unlikely to be able to be filled in automatically, we propose that these become required provenance fields. These fields are an attempt to preserve the privacy of our project participants and stakeholders and moreover minimise any legal consequences to members of the project team.

Shown below is an expansion is a link containing an expansion of the minimal plan described above. We have abstracted the information into a separate document as the table becomes difficult to read when included in a word processed document, such as this. This specification of extended metadata also includes the friend of a friend vocabulary, FOAF (Brickley and Miller 2010). This is accepted ontology for linking individuals who make use of a network, such as the World Wide Web. This is an attempt to utilise as many facilities of the PROV-O ontology, Dublin-core metadata vocabulary, and the FOAF vocabulary for precise provenance. In this table, tbl. 2 are detailed what the aspect of the ontology or vocabulary does, what the corresponding entity or property is in the 3 ontologies mentioned, what entities it applies to, why it might be needed and finally the possible method of acquisition. This is the ideal scenario in terms of consistent and comprehensive provenance metadata collection. The latest version of the document may be found here.

Project semantics

This obviously will be decided by a process of evolution of the course of the project. However, given the needs of the modellers, then it is apparent that we need at least the following two classes of data:

- Agent, and
- Environment.

The process of elucidating this ontology will be iterative and non-conformal, in that it evaluation or theoretical will not be set until the final ontology is published at the conclusion of the SMARTEES project.

Agents

We initially propose that agents will minimally allow at least the following properties: (these are derived from the GLAMURS work):

- subclass of;
- equivalent to;
- represents/similar to;
- modifies/writes;
- uses/requires/reads, and
- part of/proper part of

These properties would be extended over the period of the project, as the key drivers for energy related decision making agents or properties are uncovered. For example, wealth may be a key decision influencer.

Agents could have data properties such as Name, Title, Age, Description etc. These are properties that can generally be described by data primitives. Data primitives have only one property. For instance, an integer has only one property, its magnitude. Consequently, an Agent's name or title would fit into this category (even though the Agent may well have more than one, and may not vary due to change of language).

We also note that many of the suggestions for the minimal or extended provenance data described above might also be reused in the project semantics ontology. To see how this might work then please refer to the example ontology used to model the grant proposal and work plan for the SMARTEES project, which makes use of integration ontologies originally derived from the GLAMURS project, the friend of a friend RDF vocabulary (FOAF), the Dublin Core Metadata Initiative Vocabulary (DCMI ES), PROV-O provenance ontology, and the Simple Knowledge Organization System (SKOS). All these may be found in the ontologies sub-folder in the WP4 folder of the SMARTEES content management system here. A list of these ontologies contained in this directory is shown in Table 3.

Table 2A. The proposed metadata specification, standard and example ontologies and vocabularies.

File Name	Ontology	Purpose	Imports
skos.rdf	Simple Knowledge Organization System	Classification and terminology vocabulary	
prov.owl	PROV-O	Provenance ontology	
foaf.rdf	FOAF	Friend of a friend ontology describing connections between agents on an electronic network	
smartees.owl	SMARTEES plan ontology	The SMARTEES work plan and grant proposal as an ontology	
metadata.owl	Metadata specification ontology	The project metadata owl proposal	

h2020.owl	Vocabulary needed to describe an h2020 project	An ontology that contains things above and beyond the other component ontologies	h2020.owl

Environment

Various properties about the environment as object and data properties that will emerge as the project progresses. This differs from agency in that it is in the environment that agency takes place.
Both environment and agents might have the following object or data properties.

- at time/existence period, and
- location.

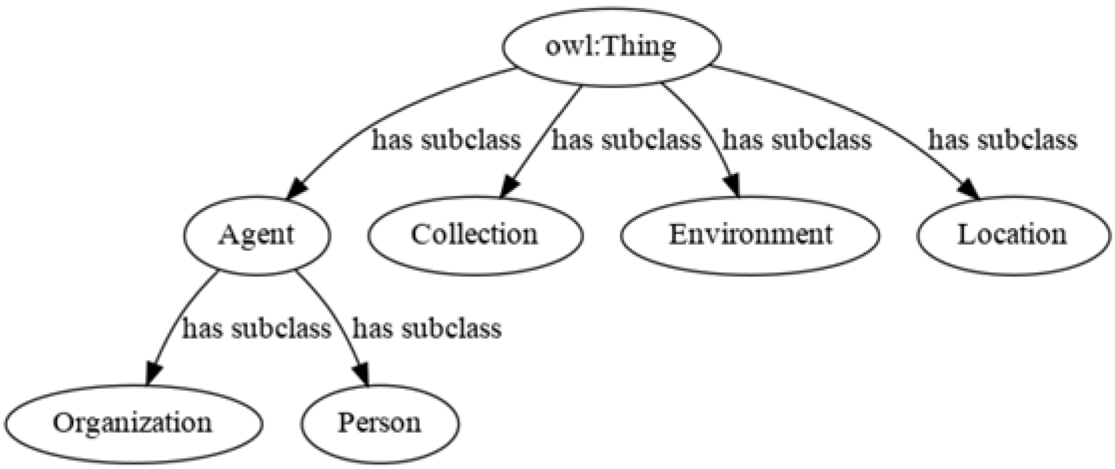
Project semantics ontology

To record this metadata, then we propose the use of a computerised ontology (Sowa 1999), and moreover we will make use of widely used technology of specifying ontologies, OWL (Horrocks et al. 2005). Moreover, as a standard means of actually encoding ontologies we suggest the OWL functional syntax (Motik et al. 2009) using UTF-8. This (or turtle) is the most human readable form an ontology in our opinion, and far more comprehensible than the usual XML based formats.

The proposed initial project-semantic ontology is listed in "An initial project-semantics ontology" subsection and also may be found in the WP4 sub-folder ontologies here.

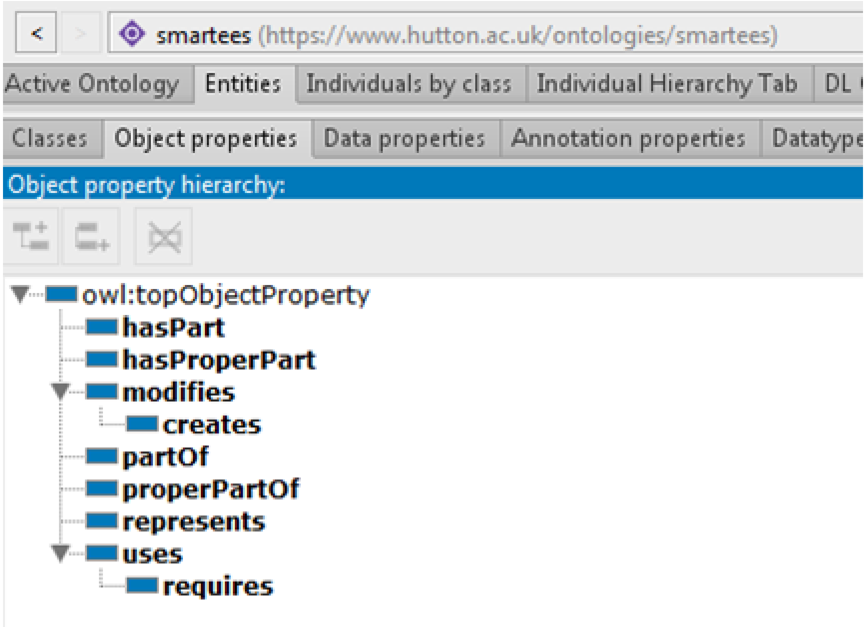
A diagram of the very simple class structure is shown in Figure 1A.

Figure 1A. A very simple class structure.



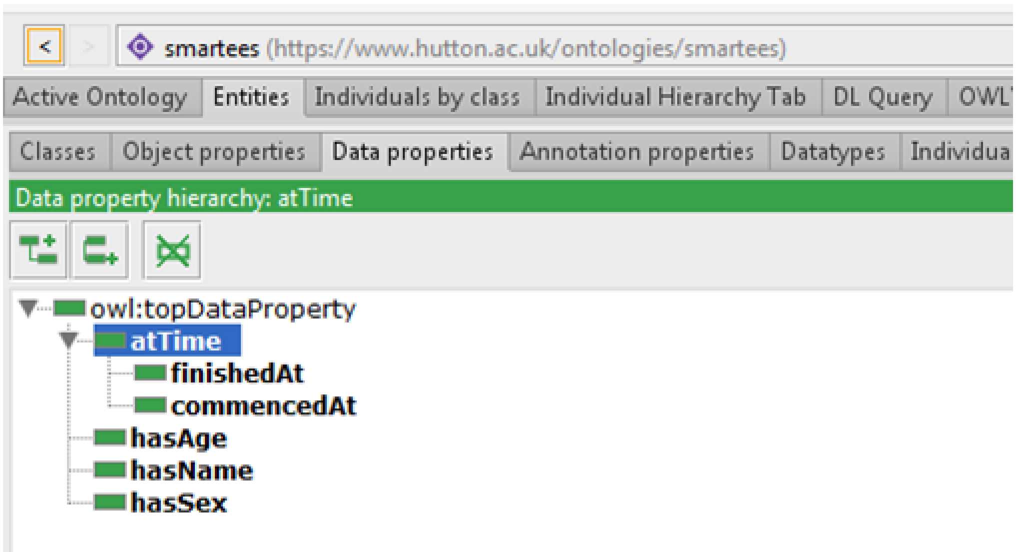
The object properties are shown in Figure 2A and the initially proposed data types are shown in Figure 3A.

Figure 2A. A Very simple set of object properties for the metadata ontology.



This is by no means authoritative, or final, but is very basic so far and we are looking forward to receiving suggestion on how such an ontology should be expanded and what it should include. However, it should be noted that each additional class, object property, or data type necessarily increase work to record that metadata. People are notoriously bad at supplying, or understanding the purpose of metadata (Doctorow 2001), so care should be exercised in what is declared mandatory.

Figure 3A. Very simple set of data properties for the metadata ontology.



The resultant metadata ontology will be published in an accessible code repository, we might also set up some kind of cloud environment which will allow the interactive investigation of the metadata ontology. This might be achieved using a thin client attaching to a graphical user-interface running on a virtual machine instance in the cloud, or utilising an existing commercial cloud presentation framework, such as Sandbox.

An initial project semantics ontology

The following is a very simple project-semantics ontology rendered as an OWL ontology using triples from the OWL functional syntax ontology specification language.

```
Prefix(:=<https://www.hutton.ac.uk/ontologies/smartees#>)
Prefix(owl:=<http://www.w3.org/2002/07/owl#>)
Prefix(rdf:=<http://www.w3.org/1999/02/22-rdf-syntax-ns#>)
Prefix(xml:=<http://www.w3.org/XML/1998/namespace>)
Prefix(xsd:=<http://www.w3.org/2001/XMLSchema#>)
Prefix(rdfs:=<http://www.w3.org/2000/01/rdf-schema#>)


Ontology(<https://www.hutton.ac.uk/ontologies/smartees>

Declaration(Class(:Agent))
Declaration(Class(:Collection))
Declaration(Class(:Environment))
Declaration(Class(:Location))
Declaration(Class(:Organization))
Declaration(Class(:Person))
Declaration(ObjectProperty(:creates))
Declaration(ObjectProperty(:hasPart))
Declaration(ObjectProperty(:hasProperPart))
Declaration(ObjectProperty(:modifies))
Declaration(ObjectProperty(:partOf))
Declaration(ObjectProperty(:properPartOf))
Declaration(ObjectProperty(:represents))
Declaration(ObjectProperty(:requires))
Declaration(ObjectProperty(:uses))
Declaration(DataProperty(:hasAge))
Declaration(DataProperty(:hasName))
Declaration(DataProperty(:hasSex))
############################
#   Object Properties
############################

# Object Property: :creates (:creates)

SubObjectPropertyOf(:creates :modifies)

# Object Property: :hasPart (:hasPart)

InverseObjectProperties(:hasPart :partOf)
TransitiveObjectProperty(:hasPart)
ReflexiveObjectProperty(:hasPart)
```

```
# Object Property: :hasProperPart (:hasProperPart)

SubObjectPropertyOf(:hasProperPart owl:topObjectProperty)
InverseObjectProperties(:hasProperPart :properPartOf)
TransitiveObjectProperty(:hasProperPart)
IrreflexiveObjectProperty(:hasProperPart)

# Object Property: :partOf (:partOf)

TransitiveObjectProperty(:partOf)
ReflexiveObjectProperty(:partOf)

# Object Property: :properPartOf (:properPartOf)

SubObjectPropertyOf(:properPartOf owl:topObjectProperty)
TransitiveObjectProperty(:properPartOf)
IrreflexiveObjectProperty(:properPartOf)

# Object Property: :requires (:requires)

SubObjectPropertyOf(:requires :uses)

# Object Property: :uses (:uses)

SubObjectPropertyOf(:uses owl:topObjectProperty)


############################
#   Data Properties
############################

# Data Property: :hasAge (:hasAge)

SubDataPropertyOf(:hasAge owl:topDataProperty)


############################
#   Classes
############################

# Class: :Agent (:Agent)

DisjointClasses(:Agent :Collection)
DisjointClasses(:Agent :Environment)

# Class: :Collection (:Collection)

SubClassOf(:Collection ObjectSomeValuesFrom(:hasPart ObjectUnionOf(:Agent :Location)))
DisjointClasses(:Collection :Environment)
DisjointClasses(:Collection :Location)
```

```
# Class: :Environment (:Environment)

DisjointClasses(:Environment :Location)

# Class: :Organization (:Organization)

SubClassOf(:Organization :Agent)

# Class: :Person (:Person)

SubClassOf(:Person :Agent)

)
```

Glossary for appendix

Belhajjame, K., Cheney, J., Corsar, D., Garijo, D., Soiland-Reyes, S., Zednik, S. and Zhao, J. (2013) *PROV-O: The PROV Ontology. W3C Recommendation 30 April 2013*. (https://www.w3.org/TR/prov-o/)

Brickley, D., and Miller, L. (2010) *FOAF Vocabulary Specification 0.91*. (http://xmlns.com/foaf/spec/)

DCMI (2012) *The Dublin Core Metadata Element Set*. (http://dublincore.org/documents/dces/)

Doctorow, C. (2001) *Metacrap: Putting the Torch to Seven Straw-Men of the Meta-Utopia*. (http://chnm.gmu.edu/digitalhistory/links/pdf/preserving/8_17.pdf)

Horrocks, I. (2005) OWL: A Description Logic Based Ontology Language. In Gabbrielli, M, Gopal, G. (eds.) *Logic Programming, 21st International Conference, ICLP 2005, Sitges, Spain, October 2-5, 2005, Proceedings.* ISBN: 978-3-540-31947-4.

Motik, B., Patel-Schneider, P. F., Parsia, B., Bock, C., Fokoue, A., Haase, P. and Hoekstra, R. et al. (2009) OWL 2 Web Ontology Language. *W3C Recommendation* 27 October 2009. (https://www.w3.org/2009/pdf/REC-owl2-overview-20091027.pdf)

Sowa, J. (1999). *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Course Technology. (http://www.jfsowa.com/krbook/)