Hypergraph of Massive Digital Traces as Representation of Human Activities: A Way to Reduce Energy Consumption by Identifying Sustainable Practices

Eddie Soulier
Tech-CICO
University of Technology of Troyes (UTT)
Troyes, France
email: eddie.soulier@utt.fr

Philippe Calvez
Crigen – Digital Lab
ENGIE
Saint-Denis, France
email: philippe.calvez1@engie.com

***Abstract*— On one hand, the ecological transition and sustainable development issues are today a reality that can not be ignored given the negative impacts of human activities on their environments. On the other hand, an increasingly important digitization of these environments results in the generation of massive volumes of digital traces, which are all signs of actors' activities. A significant challenge is to understand the ins and outs of environmental impact due to activities and considering Energy Impact (EI) as a key indicator and how this indicator can strongly change from an activity to another. Our approach considers the Practices recognition on the basis of these digital traces generated by human and non-human entities during specific activities. Practice (instantiation of activity) uses more or less resources (physical and virtual) during their existence. Being able to identify which one is more resources dependent would help to better understand how to promote ecological transition. Promoting, or at least identifying on the basis of indicators (i.e., Energy Impact), practices that have a low impact on the environment could be an innovative approach. These practices, defined as coordination of multiple heterogeneous entities in time and space, can be formalized in the form of multidimensional activities structures – Activities's Hypergraph – using the Assemblage Theory ("Agencement" in French) and using a set of mathematical tools (Simplicial Complexes, Hypernetworks). This research attempts to model the phenomenon of human and non-human activity based on the characterization of the context (massive contextual data). These Assemblages are represented and computed in a research platform (IMhoTEP), which aims to build these complex structures not based on a priori entities' classification, but by focusing on the relationships they maintain in several dimensions. The main goal is to offer a decision tool, which supports actors' ecological transition by understanding activities inducing consumption or production of resources. This academic research in the field of computer science is based on continuous digitization of physical and virtual spaces, particularly highly connected urban areas (Smart City, Internet of Everything).**

Keywords*— *Activities; Assemblage; Digital Traces; Energy Impact; Simplicial Complexe; Theory of Graph; Theory of Practice.

I. INTRODUCTION

It is observed that daily human activities have an increasingly impact – mostly negative – on the environment. They are mainly not sustainable due to excessive consumption and uncontrolled using of multiple resources (mostly non-renewable). Electrical energy (Figure 1) is a leading resource in human activities.

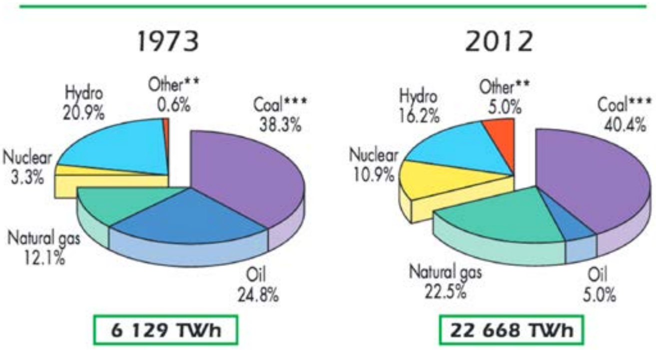

Figure 1. 1973 and 2012 fuel share of electricity generation.

Energy companies, institutional, provide several indicators to measure electric energy consumption including the "Energy Impact". We consider that each human activity has an Energy Impact (in French Emprise Energétique or EmE). We have chosen to use this indicator (EI) as a measure of electrical energy consumption in human activities. EI is the only indicator that takes into account the indirect consumption (not visible to the final consumer) in addition to the direct consumption visible to the final consummer. EI concept is based on research work of Pourouchottamin [1] and Figure 2 describes it.

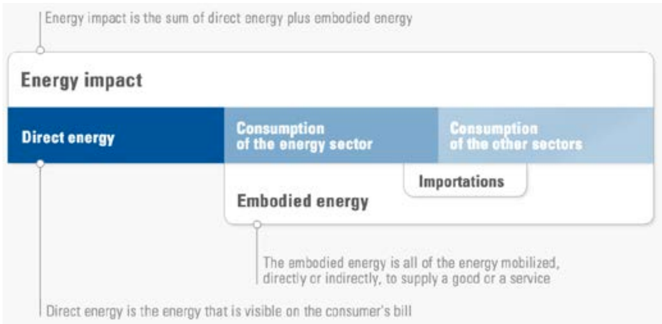

Figure 2. Energy Impact (Emprise Energétique (EmE)).

Direct energy refers to the energy visible to consumers, i.e., their fuel bills for individual vehicles, electricity, gas and other fuels for housing, etc. Direct energy corresponds to the final energy consumed by households. Embodied energy is the energy necessary for the provision of property or the service offering to the end consumer. Embodied energy is needed to manufacture and delivery of equipment at home, or produce food, energy required to the construction of the house, etc. It is also called content of energy for goods and services. This embodied energy must itself be decomposed according to its use. Embodied energy of the non-energy sectors means that it is used by economic actors in France and in the world to imagine, test, produce, and transport goods and services ultimately consumed by households. Embodied energy of direct energy corresponds to the share of energy required to develop, produce the "final" energy from natural resources and make it available to the consumer. The process requires extraction, conversion, transportation, manufacturing plants and infrastructures, etc.

Today, the calculation of energy impact is not based on the actual activities of people. It is difficult to identify what human actors really do for daily activities. Besides the question of measuring of energy impact, the functional representation of these activities do not correspond to the image the people have of their own activities. This makes it difficult to build a supervisory activity system that would help to reduce the consumption of energy by managing activities. There is in this sense a real need to answer this lack. This classic definition of human activities based on statistical study bring two basic problems:

- The first problem is that there is no overlap between the functional separation and real activities of people.
- The second problem is that the data used, are only primarily the result of statistical surveys that provide aggregated results from data, which are not directly related to activities.

Human activity is a very complex phenomenon to be observed except:

- In the case of very trivial activity (or for which semantics is well controlled)
- When it is based on pre-defined categories or declarative items.

In all other cases, we are confronted to an activity recognition problem.

Activity Recognition in computer science is a research area in strong development [2].

In this research field, there are two main approaches:

- Activity Recognition based on Vision
- Activity Recognition based on Sensor and

two types of algorithms/representation:

- Machine Learning
- Representation of knowledge.

We consider that these approaches do not allow to solve the problem as we consider it. Indeed, the computational approach of the recognition of activities do not take full account of the context of activities. For environment, lifestyles (context) significantly affect how activities are performed. That is why we have chosen to rely on specialized research work, which considers the influence of lifestyle on human activities. In particular, we study the academic work of Spaargaren [3-5], Røpke [6-8] and Shove [9-14]. Their conclusion is that changing the behavior of agents can not come from individual incentive micro level (micro) or institutional macro level (i.e., laws, ...). The basic and core assumption of our approach is that the behavior in the activity is determined by practice, in which activities are embodied (shown in Figure 3).

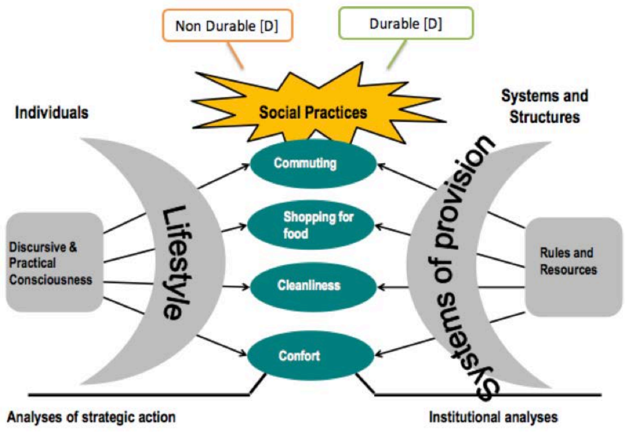

Figure 3. Influence of Lifestyle and System of provision on Social Practices [15].

Practice is one of the main theoretical artefactsf of our approach and one of our key contribution is to develop a modeling framework of practices and/for building a computer system (IMhOTEP Platform) to calculate the energy impact (EmE) related to these practices based on massive digital traces (Big Data) related to human activities.

The main problematic of our research work is how to identify (automatically and within computer science), in highly digitized environments, actors' practices in which activities are encapsulated. How to recommend low environment impact practices without using a priori or declarative lists?

II. FUNCTIONAL MODEL FOR BUILDING STRUCTURES OF HUMAN ACTIVITIES

We present our research model and our hypotheses in Figure 4.

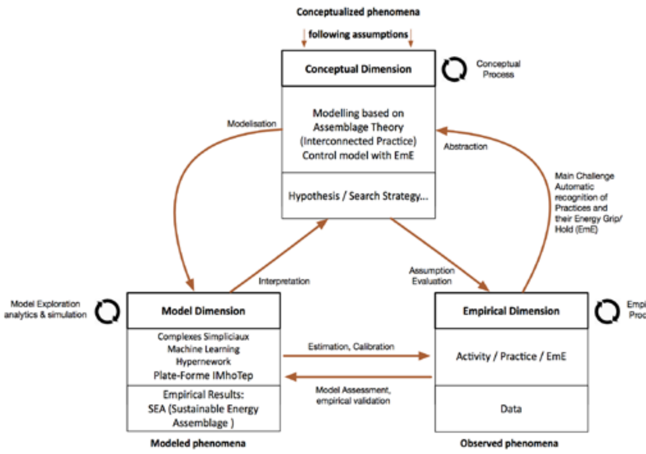

Figure 4. Global Research Model for building Sustainable Energy Assemblage (SEA).

The Empirical dimension of our model is related to the phenomenon observed, it is to say, activities and practices. Specifically, the observed phenomenon is an interconnected set of practices of various fields (housing, transportation, leisure...) and their energy impact (EmE). In this area, it seeks to observe and measure the phenomenon, by taking advantage of aggregating heterogeneous digital traces. It is in this empirical dimension that we establish a first form of the observed facts (basic activities) presented later in this work. We consider practice as the core conceptual artefact of our approach. Practice is defined by Reckwitz [16] as “a routinized type of behavior which consists of several elements, interconnected to one other: forms of bodily activities, forms of mental activities, ‘things’ and their use, a background knowledge in the form of understanding, know-how, states of emotion and motivational knowledge. A practice – a way of cooking, of consuming, of working, ...”

Our hypothesis is that practices are interconnected. For this key assumption, we suggest the concept of Assemblage (“Agencement” in French) based on previous academic works of Soulier with Delalonde [17], and with Bugeaud [18][19].

Assemblage is a methodological tool [18] for the study of any phenomenon consisting of:

- a large number of heterogeneous entities, autonomous and active,
- multidimensional network of relationships in which they associate,
- forms of organization and ability to act that emerge from their interactions.

In our field, we define a Sustainable Energy Assemblage (Agencement Energétique Soutenable [AED] in French), which is a multidimensional structure of entities in relationship in one or more dimension. In this structure – view as a Hypergraph – we aim to identify a cluster that represents Practice (group of heterogeneous entities in interaction). These practice have, as defined previously, an energy grip/hold (EmE).

The conceptual dimension of our research Model, helps us to characterize relevant factors of the observed phenomenon of the empirical dimension. That is why we consider significant concept for our approach such as Activity, Practice, Energy Impact.

We assume that an Assemblage is a multidimensional connectivity system and that we we could calculate this system and the connectivity of heterogeneous elements of a practice in several dimensions. For this we use the mathematical tool of Simplicial Complex developed by Atkin [20, 21] and generalized by J. Johnson [22] as Hypernetworks.

The last part of our model describes the model dimension. In this area, we seek to operationalize previous cited dimension of our model, through a consistent and unambiguous formal system. In our case, this formal system is based on mathematical tools that are Simplicial Complex that help operationalize the structure as SEA in a specific information system platform called IMhOTEP (susta**I**nable **MO**bility and **E**nergy social-**P**ractices inside Smart Cities), which we developed as a proof of concept to build and represent the dynamic structures of activities. This area of the model should confirm or not our theories from the conceptual dimension (theory of sustainable practices / Sustainable Energy Assemblage through the Energy Impact model) and confirm as well observation from the empirical dimension (Practice and Influence Energy). At this point, we transform our previous model into the model shown in Figure 5, in order to display our three core concept such Energy grip/hold (EmE), Practice and Assemblage.

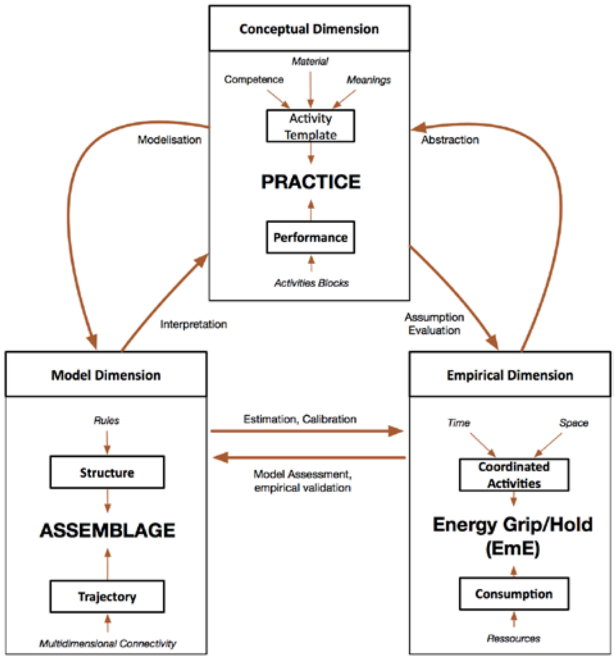

Figure 5. Research Model with Core Concept of the approach.

III. BUILDING STRUCTURES OF PRACTICES FROM DIGITAL TRACES

Calculating the Energy grip/hold (EmE) of practice from massive digital traces of activities (Big Data) generate many problems in data acquisition, data processing and modeling.

Digital traces are modeled based on ActivityStream protocol developed by Messina [23], recently upgraded in a new version of protocol by Snell [24]. We consider multiple dimensions of digital traces, as shown in Figure 6. All these dimensions aim to define more accurately characteristics of human activities.

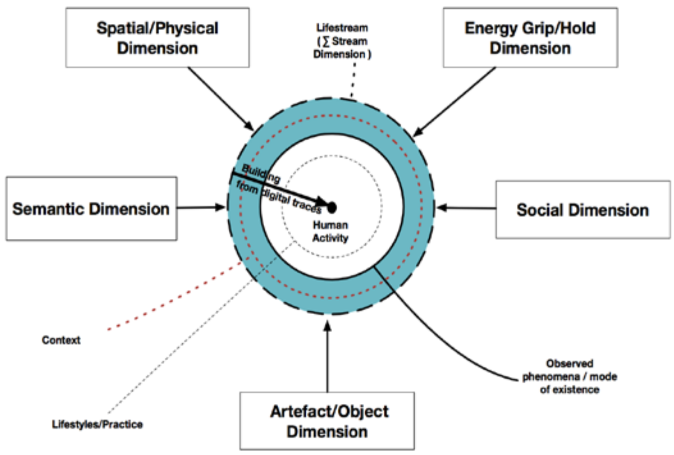

Figure 6. Digital Dimension of Human Activities.

The conceptual model takes into account the activity, location and mobility triangle, as shown in Figure 7. Activity can be localized or distributed and independent or dependent.

 ISBN: 978-1-61208-524-1

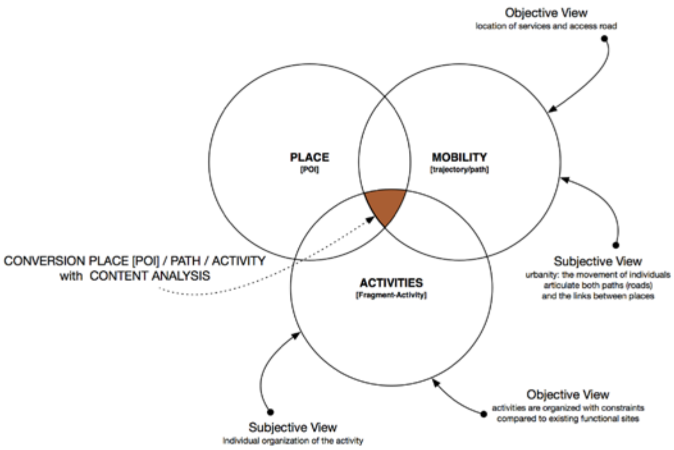

Figure 7. Triptych of Pratice.

The platform IMhOTEP developed for this academic research aims to build, from digital traces from multiple data sources (Figure 6), through automatic computing process based on Simplicial Complex mathematical tool, structures of activities in which Practice will be searched and identified. The model that we have chosen (Hyperconnectivity in multiple dimensions, paradigm of massive heterogeneous data, use of Natural Language Processing, automatic recognition patterns of activities with Machine Learning algorithms) creates numerous technical challenges. Figure 8 presents the architecture of IMhOTEP platform and how from Lifestream of data, we build structures of activities as a Hypergraph in which we seek, with Machine Learning algorithm (K-Means), to identify Practices and those who are sustainable.

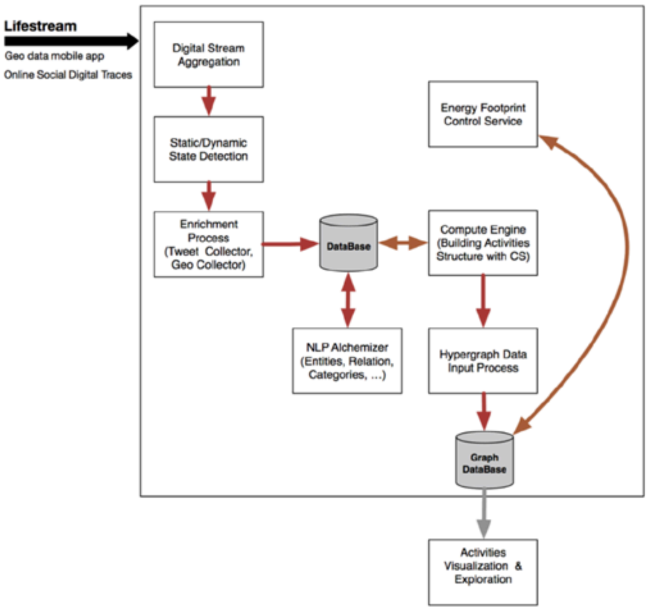

Figure 8. IMhOTEP Architecture.

IMhOTEP could become in the near future an industrial solution with automatic recognition of activities from massive digital data.

IV. EXPERIMENTATION WITH IMhOTEP PLATFORM

We present in this section our framework for experimentation. We started with a hypothetical situation with two people who agree to install the application on their smartphone and share their data for a specific duration, inside two different geographical areas (France and Germany). The goal of our experiment is to validate the Assemblage model (conceptual validation), the IMhOTEP platform (technical) and calculating the Energy grip/hold (EmE) for each practice using massive digital traces and the Energy Control Model (ECM) that we develop for this experimentation.

Hypothesis of our experimentation are that an unknown activity is caught by the system IMhOTEP, when someone is staying at the same spatial position (latitude/longitude) for at least 5 consecutive minutes. Then, the process of aggregation of digital traces can start, based on the localization of the person and the surrounding context. The nature of the activity is determined by the context analysis. Depending on the location of the subject of reference, we consider the nearest Point Of Interest (POI) as a reference to aggregate digital traces around it. We consider categorization of the POI based on services such Foursquare or Google Place. We then extract from these provided information keywords related to POI's. At the same time, we start to aggregate the social dimension information, with Tweets from Twitter service. We define a boundary Box based on the position of the POI and the person. These Tweets with use of NLP tools help us to identify more entities by extracting most cited keywords. These entities will be used later in the IMhOTEP Platform to build the structures of Activities. By using Simplicial Complex calculation, we use matrices of adjacency and matrices of incidences as shown in Figure 9.

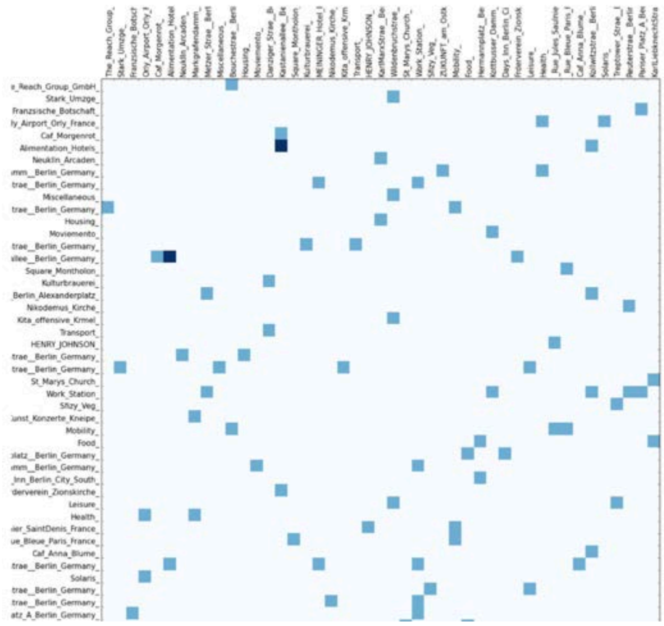

Figure 9. Use of Simplicial Complex mathematical tool – Example of Adjacenct Matrix.

The key point is the allocation of a category of activities automatically by grouping activities, which are closed. To do this we apply K-Means Algorithm to define clusters of category of activities in which we could later identify Practices. Once these clusters are defined we use graph query to search for specific practice and its Energy Impact evaluation. Figure 10 show the structure of activities.

Figure 10. Structure of Activities – Practice searching based on Graph Queries.

The scale of the following experiment was limited to 3 people over one month, mostly due to limitation of Smartphones using a data/sensor for following people over time, and location and the willingness of people to share their constant spatial position during a long period of time. The mobile application energy consumption of the battery, due to work all day long and push every Lat/Lon position and other data every minute, was clearly a disadvantage at this point of the experiment. Regarding the acquisition of data from different context sources (Tweets, Google Place…), it depends on how many "activities" (defined stops over 5 minutes and considered as activity) are done by user. The average volume of data per day, per user, if we consider all digital traces needed for the IMhOTEP approach acquired for this period of time, reached 12 Mo of data/per user/per day. This could be considered as relatively low. But if we scale up to hundreds or thousands of people, the volume could lead to a computational challenge, especially when considering to manage a large graph with hundreds, thousands, even millions of nodes with multidimensional relationships.

This implementation of IMhOTEP was experimental in order to acquire data to build an Hypergraph of activities. Computational challenges were present, for instance to connect all open source tools (mobile application, Database, code to extract data, API access). This element was a heavy constraint because when you want to access to API like Twitter or Google, or others such Yelp, you only get a specific amount of query from the API. Moreover test and implementation suffer due to API's limits.

Another limit directly linked to the Matrix Calculation (Incidence and Adjacency) was the time required to calculate the experiment. It was not a huge problem at this point, but size of matrix (for hypergraph calculation / Simplicial Complex) could lead to combinatorics computation challenge.

A data, collected every minute for Geo Position, starts to be acquired when someone stops doing activities. This "starting" point leads to aggregation of data. As already written, we came to an average of data per day per user of 12 Mo. We can imagine that if you acquire more data to better characterize the context of an activity and build a representation with Simplicial complex, you will have to scale and move to a probably more resilient infrastructure (such TitanDB/SPARK, which we are actually testing).

Regarding computational challenge, it will depend mostly on the objective of the approach and use cases. Because collecting real time or/and predictive activity directly from a user and his mobile phone will represent a huge challenge (computing / hardware, i.e battery). If the ultimate goal is to produce a daily dashboard, computational challenge will still be present but definitively in a different scale.

Based on our Energy Control Model, and the IMhOTEP platform, we have a 63% accuracy for identifying practices with the right categorization. When we ask people who participate to the experimentation if the practice found by the IMhOTEP platform has the right category and reflect what they did, they answer positively to more than 6 of 10. Regarding the initial objective to identify almost automatically complex practices, this intermediary result is relevant.

V. CONCLUSION

Our approach is based on massive digital traces aggregation and how from these pieces of heterogeneous information, we are able to build automatically and backward a structure of human activities. The other objective was to identify practice related to person without an a priori categorization of practices analyzed. Recognition rate could be improved by being able to better localize people in space and time and use more advanced machine learning algorithm to infer over the time lifestyles of people. IMhOTEP provides a new way of building large multidimensional Hypergraph to represent dynamic structures of heterogeneous entities. Perspective and next research should focus on how to manage large amount of users, which can lead to combinatory calculation limits due to large incidence and adjacency matrices.

REFERENCES

[1] P. Pourouchottamin, C. Barbier, L. Chancel and M. Colombier, "New representations of energy consumption," IDDRI, UMR 8568 CIRED, EDF R&D, 2013.

[2] L. Chen and I. Khalil, "Activity recognition: Approaches, practices and trends," Activity Recognition in Pervasive Intelligent Environments, Springer, 2011, pp. 1-31.

[3] G. Spaargaren and B. Van Vliet, "Lifestyles, consumption and the environment: The ecological modernization of domestic consumption," Environmental Politics, vol. 9, no. 1, 2000, pp. 50-76.R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[4] G. Spaargaren, "Theories of practices: Agency, technology, and culture: Exploring the relevance of practice theories for the governance of sustainable consumption practices in the new world-order," Global Environmental Change, vol. 21, no. 3, 2011, pp. 813-822.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[5] G. Spaargaren, "Sustainable consumption: a theoretical and environmental policy perspective," Society &Natural Resources, vol. 16, no. 8, 2003, pp. 687-701.

[6] I. Røpke, "The dynamics of willingness to consume," Ecological Economics, vol. 28, no. 3, 1999, pp. 399-420.

[7] I. Røpke, “Theories of practice—New inspiration for ecological economic studies on consumption,” Ecological Economics, vol. 68, no. 10, 2009, pp. 2490-2497.

[8] I. Røpke, “New technology in everyday life–social processes and environmental impact,” Ecological Economics, vol. 38, no. 3, 2001, pp. 403-422.

[9] E. Shove, The design of everyday life, Berg Publishers, 2007.

[10] E. Shove, H. Chappells, L. Lutzenhiserc and B. Hackettd, “Comfort in a lower carbon society,” Building Research & Information, Volume 36, Issue 4, 2008.

[11] E. Shove, “Beyond the ABC: climate change policy and theories of social change,” Environment and planning. A, vol. 42, no. 6, 2010, pp. 1273.

[12] E. Shove, M. Watson and M. Pantzar, The dynamics of social practice: Everyday life and how it changes, SAGE Publications Limited, 2012.

[13] E. Shove and G. Walker, “Governing transitions in the sustainability of everyday life,” Research Policy, vol. 39, no. 4, 2010, pp. 471-476.

[14] E. Shove, “Converging conventions of comfort, cleanliness and convenience,” Journal of Consumer Policy, vol. 26, no. 4, 2003, pp. 395-418.

[15] G. Spaargaren and C.S.A.K. Koppen, “Provider strategies and the greening of consumption practices: Exploring the role of companies in sustainable consumption,” The new middle classes, 2009, pp. 81-100.

[16] A. Reckwitz, “Toward a Theory of Social Practices A development in culturalist theorizing,” European journal of social theory, vol. 5, no. 2, 2002, pp. 243-263.

[17] E. Soulier, C Delalonde and O Petit, “ Subjectivation and singularizing process in the perspective of situated learning and actor network theory,” International congress AREF, 2007.

[18] F. Bugeaud, “Isamsara: for an engineering-based services and systems mereology hypergraphs ",” Troyes, 2011.

[19] E. Soulier, F. Rousseaux, F. Bugeaud, J. Legrand and H. Neffati, “Modeling and simulation of new territories projects using agencements theory, mereological principles and simplicial complex tool,” Proc. 2011 IET International Conference on Smart and Sustainable City (IET ICSSC2011), Shanghai, China July, 2011, pp. 6-8.

[20] R.H. Atkin, Combinatorial Connectivities in Social Systems: an application of simplicial complex structures to the study of large organizations, Birkhäuser - 3764309121, 1977.

[21] R.H. Atkin, Mathematical structure in human affairs, Heinemann Educational London - 0435820257, 1974.

[22] J. Johnson, “Hypernetworks of Complex Systems,” Complex Sciences, 2009, pp. 364-375.

[23] M. Atkins, et al., “Atom Activity Streams 1.0,” Abgerufen am, vol. 22, no. 08, 2011, pp. 2013.

[24] J. Snell, “JSON activity streams 2.0,” 2014.

 ISBN: 978-1-61208-524-1