

Open Ontology-Driven Sociotechnical Systems: Transparency as a Key for Business Resiliency

N. Guarino¹, E. Bottazzi¹, R. Ferrario¹, G. Sartor²

Abstract Most business and social organisations can be seen nowadays as complex *sociotechnical systems* (STSs), including three components: *technical artifacts*, *social artifacts*, and *humans*. Within social artifacts, a special role have *norms*, which largely influence the overall system's behavior. However, norms need to be understood, interpreted, negotiated, and actuated by humans, who may of course deviate from them, or even decide to change them. STSs are therefore essentially prone to failure: critical situations are part of STS's life, and may sometimes lead to tragic outcomes. That's why *resilience* to failure must be built into such systems, and is a crucial parameter to determine their quality. We argue in this paper that, to achieve a high level of resilience, *transparency* is the key: actors within the system need to take a reflective stance toward the system itself. In other words, an STS must be *open* to its actors, which by observing and understanding its dynamics can take the appropriate initiatives in presence of unforeseen problems, possibly modifying the system at run time. Ontological models can play a crucial role in this context. However, we need to make a radical change in our modelling approach, shifting the focus of analysis from ontology-driven *information systems* to ontology-driven *sociotechnical systems*.

Introduction

Most business and social organisations can be seen nowadays as complex *sociotechnical systems* (henceforth STSs), including three components: *technical artifacts*, *social artifacts*, and *humans*. The specific nature of STSs with respect to other sorts of systems has been studied recently in (Kroes et al. 2006), where the differences and the mutual interactions between technical and social aspects are analysed. Technical systems are physical systems designed to achieve some human purpose. Accordingly, they are modelled and analysed as physical systems, consisting of interconnected components, whose behaviour is completely described by natural laws. Knowledge of these laws can be obtained to great

¹ ISTC-CNR, Trento, Italy. {Guarino,Bottazzi,Ferrario}@loa-cnr.it

² European University Institute, Law Department, Florence, Italy.
giovanni.sartor@eui.eu

levels of precision by careful experimenting, allowing the system's architecture to be predictable and controllable to the extent that is required for the desired functionality. Social systems, in contrast, are made up of human persons and social organisations and institutions, which can themselves be analysed as social systems. The behaviour of persons and institutions is not entirely determined by natural laws, but is also guided by private decision-making rules internal to persons as well as 'public' norms guiding social behaviour. The latter may pertain to different institutions, may or may not be legally enforceable, and may compete with one another and with the pursuit of individual or collective goals.

In *designed* STSs, intended to realize a certain pre-defined function, both technical and social artifacts – as well as human operators – are crucial for the overall functioning. Technical artifacts, like tools and machines, determine what *can* be done, amplifying and constraining opportunities for action; social artifacts, like norms and institutions, determine what *should* be done, governing obligations, goals, priorities, and institutional powers. Since institutions, in turn, are created by norms, STSs can be seen as *norm-governed* systems, whose structure and behaviour largely depend on norms. However, norms need to be understood, interpreted, negotiated, actuated by humans, who may of course deviate from them, or even decide to change them.

Designed STSs are therefore essentially prone to failure: critical situations are part of STS's life, and may sometimes lead to tragic outcomes. That's why *resilience* to failure must be built into such systems, and is a crucial parameter to determine their quality. We argue in this paper that, to achieve a high level of resilience, transparency is the key: actors within the system need to take a reflective stance toward the system itself. In other words, an STS must be *open* to its actors, which by observing and understanding its dynamics can take the appropriate initiatives in presence of unforeseen problems, possibly modifying the system at run time.

Ontological analysis and ontology-driven conceptual modelling (Guarino 98) can play a crucial role in this context. However, we need to make a radical change in our modelling approach, shifting the focus of analysis from a piece of software (however *embedded* in an external system modelled as separated from the information system) to the *embedding* system, including the information system itself. In other words, we need to move from ontology-driven *information* systems to ontology-driven *sociotechnical* systems, where the ontology becomes the key for making the whole system transparent to itself and to the external environment, facilitating communications within the various components and helping the concerned actors to make the required choices pertaining to system design, management and use. This paper aims to be a *manifesto* for this radical change of perspective. In the following, we shall briefly discuss the social and scientific implications of ontology-driven sociotechnical systems, and suggest some methodological direction lines for future research.

Sociotechnical systems

From technical artifacts to sociotechnical systems. Concerning technical artifacts, modern design, development and management methodologies have increasingly recognised the need to take human and social aspects into account, embedding the technical dimension in the broader social context. Back in the sixties, this was advocated by the *sociotechnical systems theory* (STS theory) (Emery and Trist 1960). Although originally focusing on labour organisation, the STS theory has had a substantial impact on information systems research, as well as on agent-oriented software engineering and multi-agent interaction systems, where the need to consider people and organisations as not just users but *actors* has clearly emerged (Yu 2009). More recently, the principles of STS theory have been applied to service-oriented computing, where a systemic approach to “service science” has been proposed (Chesbrough and Spohrer 2006). However, the sociotechnical approach is not yet widely and effectively practiced. According to (Baxter and Sommerville 2011), one of the reasons of this is the lack of a *systematic engineering methodology*. In other words, it is not enough to propose sociotechnical *principles* urging engineers to adopt them: generic principles must be translated into formal engineering techniques (Coiera 2007), which in our understanding means we need first of all comprehensive formal *models*.

Legal and institutional aspects. Concerning social artifacts, disciplines such as philosophy, sociology, economics and law have provided theories and methods for analysing, modelling, and designing them (Searle 1995, Coleman 1990, Williamson 2000, MacCormick 2007). In particular, to capture the diverse ontological forms and functions of norms (stating obligations, providing permissions and rights, allocating roles, providing ways for achieving individual and social objectives) we need to take into account the work of philosophers, jurists (Hohfeld 1964) and legal theorists (Kelsen 1967, Ross 1968, Hart 1994) that have anticipated in many regards recent ontologies of normative entities, providing theories of legal norms and acts, normative positions and entitlements, and legal systems. Legal theory provides indeed a rich conceptual frameworks for approaching normative phenomena, though often lacking the precision required by more analytical approaches. Moreover, only to a very limited extent studies on social-institutional systems pay due attention to the role of technological components in norm-governed and institutional action. No legal theory is available to deal with the commonalities and differences between technical and legal artifacts and their mutual integration, even though it is now apparent that in many domains the objectives of the law (protection of individual rights, prevention of antisocial behaviour, facilitation of beneficial activities) can only be obtained by regulating the way in which technological objects and systems are designed (Lessig 2006). We need therefore an adequate theory of STSs to address the ways legal norms may affect the use of technical artifacts, and how the

intertwining of technical artifacts and laws can affect human behaviour, and the overall dynamics (and resilience) of STSs.

The role of social components. An important aspect that formal models of STSs should clarify concerns the role played by social components with respect to the technical components. For instance, focusing on human components, clearly we should distinguish the case where humans are just *users* of technical artifacts, being therefore *external* to the artifacts themselves, from the case where human operators play a functional role *internal* to technical artifacts, which are therefore *sociotechnical artifacts*. Similarly, we should distinguish between usage norms and internal organization norms, which are constitutive of sociotechnical artifacts. Indeed, such clarification would help addressing the still open terminological ambiguities concerning the exact nature of STSs (Baxter and Sommerville 2011), especially with regard to their actual *boundaries*: while everybody agrees on the fact that an STS includes both a technical and a social subsystem, for many people (e.g., Alter 2006) the social subsystem includes the users, while some recent work (Kroes et al. 2006) suggests a stricter notion, holding that the social components of an STS are just those which are necessary for its functioning, namely human operators plus (indirectly) the norms that define their specific roles. We suggest to use the term *sociotechnical artifact* for this stricter notion, keeping the term *sociotechnical system* for the broader notion.

Failures and impasses. The presence of social components in STSs exposes them to failures in a peculiar way, very different from merely technical systems. Despite the functioning of human operators as system components is usually optimised by meticulous training and instruction, still they may not comply with the rules defining their functional role, as such rules may compete with other rules characterising their individual behavior. In addition, external users are much more loosely guided by system's rules than human operators internal to the system, and are more likely to experience conflicts between system's norms and external norms, often concerning the interaction between users and operators. This creates risks of system failure absent from technical systems, and puts severe constraints to the optimal design and control of STSs. In this perspective, understanding the nature and the different kinds of failures within STS is of utmost importance. Failures cannot be always avoided or mitigated by constraining human behaviour and limiting human intervention by means of laws. In fact, some failures are caused by mistaken human behaviour (as in the Chernobyl disaster) and others are caused by the rigidity of the system, which does not provide enough feedback or completely excludes human intervention in critical circumstances (as in Kubrick's Dr. Strangelove). Yet other failures originate from simple technical faults (a broken connection) or from unpredictable external circumstances (a natural disaster). Without always resulting in failures, such events can lead to situations of *impasse* in which no further step can be readily imagined or taken.

Sociotechnical Systems and Information Systems

In information technology, adaptation and robustness have become central for software-intensive systems accounting for technological, social and legal aspects, such as social computing and socially-aware software applications, service-oriented applications, computational models of normative systems and norm-governed behaviour. The role of social organization and interaction has been addressed in agent-oriented software engineering (Yu 2009) and in artificial intelligence, especially in the multi-agent systems community. Considering in particular organisational models, the approach inspired to the language/action perspective (Dignum 2004, Colombetti et al. 2002) has shed new light on previous work on enterprise engineering approaches based on the CIMOSA methodology (Kosanke et al. 1999), as well as on the early enterprise ontologies (Ushold et al. 1998). A further recent trend focuses on systems which not only can reconfigure themselves, but whose purposes can evolve to comply with the changing external constraints and stakeholders' needs (Kephart 2003). This is a challenge in particular for *embedded systems*, especially if different environmental, technological, social and legal aspects have to be taken into account.

Altogether, we can conclude that several research trends, both from the technical side and from the social and institutional side, advocate the need for a unitary perspective that takes the social aspects as seriously as the technical ones (see also Kroes and Meijers 2005 and Ottens et al. 2006). Despite these efforts, however, no comprehensive theoretical approach provides for the integration between methods and theories supporting analysis and design of social artifacts and those addressing technological ones. In other words, no overarching "science of the artificial" (Simon 1969) bridging the two dimensions is yet available. We believe that this is partly due to insufficient communication and cross-fertilisation between different groups of people, both within computer science, and between computer science, social science and legal science. In applied computer science, many researchers advocating formal engineering techniques (such as logic-based formalisms or semantic technologies), tend to neglect or over-simplify social aspects, while sustainers of the sociotechnical approach sometimes fail to appreciate the importance of formalisation, or are however still searching for robust and comprehensive formal techniques. Between technological and social disciplines there is an even broader gap, since researchers working in the two fields often ignore methods and results of their counterpart. In particular, technologists fail to understand how norms and institutions shape human behaviour (and thus the design and use of technical artifacts), while social scientists and jurists fail to capture opportunities and constraints embedded in technological architectures.

The inability to capture, in an overarching model, the subtle interactions between the social and the technical components makes it difficult to enable the overall governance of STSs, as we cannot fully assess their benefits, risks and

costs. In particular, without formal enough comprehensive models, we cannot anticipate potential crises leading to impasses or failures, and establish technical and institutional mechanisms able to cope with them, avoiding or mitigating dangerous or even tragic outcomes.

Research Challenges and Methodological Suggestions

We believe that only by precisely understanding the complex structure and dynamics of STSs we will be able to adequately design and manage them, and that only by making an STS open and transparent to the reflection of its agents we can make it resilient to unforeseen crises. Therefore, the main research challenge is to develop a *comprehensive, well-founded theory of sociotechnical systems which embeds failures, impasses and recovery attempts at its very heart.*

Right now, we only have separate modelling techniques for *isolated components or aspects* of STSs, such as design specifications for the technical components, interaction and organisational models for inter-agent communication and collective behaviour, deontic models for the normative component, theories of legal norms and institutions. On the contrary, we need a comprehensive theory that *integrates together*:

1. an ontological analysis of STSs' nature and structure, in terms of their internal components and their mutual interactions, covering technical, social and legal aspects;
2. a declarative model of STSs' dynamics, accounting in particular for the constraints on expected behaviours and for the different kinds of anomalous behaviour, including critical situations and recovery patterns;
3. the identification of techno-institutional mechanisms enabling the self-governance of STSs, in particular providing them with the capacity to sustain, avoid or mitigate failures and impasses.

To integrate the above components into a comprehensive theory, we suggest to develop a methodology based on a combination of different approaches:

1. *Failure-oriented approach*, to focus on the most important practical need concerning the actual deployment of STSs, namely understanding, controlling, and living with organisational *failures*, technical *malfunctionings*, *misconceived* rules or decisions, or overall system's *impasses*.
2. *Formal ontological analysis*, to establish a rigorous basis for understanding the *nature and structure* of STSs. Building on previous work in applied formal ontology, we can leverage on established results in analytic philosophy from the theory of essence and identity, the theory of parts, the

theory of unity and plurality, the theory of dependence, the theory of composition and constitution, the theory of properties and qualities.

3. *Open declarative systems approach* (Montali 2010), to model and design STSs while taking into account flexibility, adaptability, and transparency. In our view, the *openness* choice means that i) agent interaction protocols and rules are dynamically *modifiable*, in order to cope with unpredictable and dynamic environments; ii) agents can *violate* the system's norms in an emergency response to a crisis situation; iii) agents have *transparent cognitive access* to the system's structure, goals and governing norms. The *declarative* choice means that norms and behavioural constraints are expressly stated, in order to enable reasoning about unexpected behaviour and malfunctioning cases.

We believe that the vision depicted above has a high social impact. While standard approaches to STSs focus on models, architectures and general recommendations to be adopted at the *design phase*, we insisted on the importance of having the system open also at *operational time*. In other terms, the structure and the functioning of an STS should be *transparent* and accessible to its users and stakeholders, not just to its designers or controllers. In this participatory perspective, where the ontology plays the role of a mediator of design, governance and social participation, the challenge is to give to both the STS's designers and participants the conceptual tools to reflectively understand and discuss the system's structures and operations. In particular, we envision the possibility for participants to anticipate or at least timely detect crisis situations, and identify ways to recover from a system's failure or impasse, while at the same time significantly shape the evolution of the STSs they live in.

Finally, of course, transparency needs to be compatible with privacy. Although we don't have any specific suggestions in this respect, we believe that the design of privacy measures will be facilitated by our approach, which enables to anticipate privacy threats as one of the problematic aspects of STSs, to be countered by a combination of technical, legal and social measures. In particular, our method can provide new support to the idea of privacy by design, whose effective and useful application requires understanding the various communication channels within STSs, the different roles involved and the nature of the data exchanged.

Acknowledgements. This work has been carried within the project ICT4Law (ICT Converging on Law), funded by the Piedmont Region. The authors are indebted to Stefano Borgo, Maarten Franssen, Claudio Masolo, Marco Montali, and Laure Vieu for their precious contribution.

References

- Alter, S. (2006). *The Work System Method: Connecting People, Processes, and IT for Business Results*. Work System Press, Larkspur, CA.
- Baxter, G., and I. Sommerville (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with Computers*, 23(1):4–17.
- Chesbrough, H., and J. Spohrer (2006). A research manifesto for services science. *Communications of the ACM*, 49(7):35–40.
- Coiera, E. (2007). Putting the technical back into socio-technical systems research. *International Journal of Medical Informatics*, 76:98–103.
- Coleman, J. S. (1990). *Foundations of Social Theory*. Harvard University Press, Cambridge, MA.
- Colombetti, M., N. Fornara, and M. Verdicchio (2002). The role of institutions in multiagents systems. In *Workshop on Knowledge based and reasoning agents, VIII Convegno AI*IA*, Siena, Italy.
- Dignum, V. (2004). *A Model for Organizational Interaction: based on Agents, founded in Logic*. PhD thesis, Utrecht University.
- Emery, F. E., and E. Trist (1960). Socio-technical systems. *Management Sciences: Models & Techniques*, 2:83–97.
- Guarino, N. (1998). Formal Ontology in Information Systems. In Guarino, N. (ed.), *Formal Ontology in Information Systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998*. Amsterdam, IOS Press: 3-15.
- Hart, H. L. A. (1994). *The Concept of Law*. Oxford University Press, Oxford, 2nd edition.
- Hohfeld, W. N. (1919). *Fundamental Legal Conceptions*. Yale University Press, New Haven, Conn.. (2nd ed. 1964.).
- Kelsen, K. (1967). *Pure theory of law*. University of California Press, Berkeley.
- Kephart, J., and D. Chess (2003). The vision of autonomic computing. *IEEE Computer*, 36(1):41–50.
- Kosanke, K., F. Vernadat, and M. Zelm (1999). Cimos: Enterprise Engineering And Integration. *Computers in Industry*, 40(2-3):83–97.
- Kroes, P., M. Franssen, I. van de Poel, and M. Ottens (2006). Treating Socio-Technical Systems As Engineering Systems: Some Conceptual Problems. *Systems Research and Behavioral Science*, 23(6):803–814.
- Kroes, P. and A. Meijers (2005). *Philosophy of technical artefacts*. Joint Delft-Eindhoven research programme 2005-2010. Technical report.
- Lessig, L. (2006). *Code V2*. Basic Books, New York.
- MacCormick, D. N. (2007). *Institutions of Law*. Oxford University Press, Oxford.
- Montali, M. (2010). *Specification and Verification of Declarative Open Interaction Models: A Logic-based Approach*. Springer.
- Ottens, M., M. Franssen, P. Kroes, and I. van de Poel (2006). Modelling Infrastructures As Socio-Technical Systems. *International Journal of Critical Infrastructures*, 2(2):133–145.
- Ross, A. (1968). *Directives and Norms*. Routledge, London, 1968.
- Searle, R. (1995). *The Construction of Social Reality*. The Free Press, New York.
- Simon, H. A. (1996). *The Sciences of Artificial*. MIT, Cambridge, Mass., 3rd edition, 1996.
- Uschold, M., M. King, S. Moralee, and Y. Zorgios (1998). The Enterprise Ontology. *The Knowledge Engineering Review*, 13(1).
- Williamson, O. E. (2000). The new institutional economics: Taking stock, looking ahead. *Journal of Economic Literature*, 38:595–613.
- Yu, E. S. (2009). Social modeling and i*. In A. Borgida, editor, *Conceptual Modeling: Foundations and Applications, Lecture Notes in Computer Science*. Springer.