University of Bath

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Governance of Autonomous Agents on the Web: Challenges and Opportunities

TIMOTHEUS KAMPIK, Umeå University, Sweden
ADNANE MANSOUR, Mines Saint-Étienne, France
OLIVIER BOISSIER, Mines Saint-Étienne, France
SABRINA KIRRANE, Wirtschaftsuniversität Wien, Austria
JULIAN PADGET, University of Bath, UK
TERRY R. PAYNE, University of Liverpool, UK
MUNINDAR P. SINGH, North Carolina State University, USA
VALENTINA TAMMA, University of Liverpool, UK
ANTOINE ZIMMERMANN, Mines Saint-Étienne, France

The study of autonomous agents has a long tradition in the Multiagent System and the Semantic Web communities, with applications ranging from automating business processes to personal assistants. More recently, the Web of Things (WoT), which is an extension of the Internet of Things (IoT) with metadata expressed in Web standards, and its community provide further motivation for pushing the autonomous agents research agenda forward. Although representing and reasoning about norms, policies and preferences is crucial to ensuring that autonomous agents act in a manner that satisfies stakeholder requirements, normative concepts, policies and preferences have yet to be considered as first-class abstractions in Web-based multiagent systems. Towards this end, this paper motivates the need for alignment and joint research across the Multiagent Systems, Semantic Web, and WoT communities, introduces a conceptual framework for governance of autonomous agents on the Web, and identifies several research challenges and opportunities.

CCS Concepts: • **Computing methodologies → Multiagent Systems**; **Intelligent agents**; • **Information systems → web services**.

Additional Key Words and Phrases: autonomous agents, norms, policies, preferences, governance

Authors' addresses: Timotheus Kampik, Umeå University, Umeå, Sweden, tkampik@cs.umu.se; Adnane Mansour, Mines Saint-Étienne, France, adnane.mansour@emse.fr; Olivier Boissier, Mines Saint-Étienne, France, olivier.boissier@emse.fr; Sabrina Kirrane, Wirtschaftsuniversität Wien, Vienna, Austria, sabrina.kirrane@wu.ac.at; Julian Padget, University of Bath, Bath, UK, j.a.padget@bath.ac.uk; Terry R. Payne, University of Liverpool, Liverpool, UK, t.r.payne@liverpool.ac.uk; Munindar P. Singh, North Carolina State University, Raleigh, North Carolina, USA, 27695, mpsingh@ncsu.edu; Valentina Tamma, University of Liverpool, Liverpool, UK, v.tamma@liverpool.ac.uk; Antoine Zimmermann, Mines Saint-Étienne, France, antoine.zimmermann@emse.fr.

1 INTRODUCTION

Over the last two decades, the Web has evolved extensively in response to a variety of different requirements. From originally providing a distributed information dissemination architecture, it has encompassed support for publication, discovery, consumption and aggregation of information, knowledge, and services, thereby interconnecting the digital, social and physical worlds. The Web's ubiquity, as well as the simplicity of its underlying communication protocols has resulted in it becoming the de facto standard for communication between services, and more recently, connected things. With the rise of the Internet of Things (IoT), the combination of the Web of Things (WoT) (as an extension of the IoT with metadata expressed in Web standards), traditional web services, and a knowledge dissemination infrastructure that is both machine navigable and machine understandable, has facilitated a new generation of applications that utilise the Web.

Berners-Lee et al. [14] outline how autonomous agents could comprehend and exploit this machine-readable knowledge to achieve a variety of tasks. Thus, the notion of *autonomy* provides a framework whereby individual agents (e.g., those representing or controlling services, things, or applications) may plan, collaborate, and cooperate to achieve complex but disparate goals. Such multiagent systems avoid centralised control, which is the bane of business process management [129]. By seeking mutually beneficial interactions, agents of heterogeneous construction (potentially originating from different developers) can evolve a mutually supportive economy across the Web, performing a multitude of tasks for Web users. However, to achieve this notion of collaborative agents that use the Web infrastructure, it is crucial to consider a *governance* perspective, which defines how agents should act in a given situation (also considering the consequences of their potential actions) and defines how frameworks that govern communities of agents should be designed, interoperate, and evolve. This perspective is of particular importance for the Web, where usage may cross social contexts and jurisdictions, and where no centralised control over the different agents is possible. Indeed, the need for intelligent system governance is, at the time of writing, a focus point of legislative and regulatory efforts; e.g., by the European Commission [74].

Therefore, what is needed is a new governance framework supported by a review of the related literature on the use of norms, policies, and preferences for autonomous normative agents, as well as contextualising these with respect to the notion of the Web (of Things). Towards this end, this paper makes three main contributions. Firstly, it motivates the need for norms, policies and preferences for autonomous agents on the Web by means of a simple motivating scenario. Secondly, it proposes a new governance conceptual framework and gives an overview of the state of the art on norms, policies, and preferences for autonomous normative agents (restricted to those efforts that provide the theoretical background to our proposed framework). Finally, it identifies several challenges and opportunities, for the MAS, Semantic Web and WoT communities, underlining the need for better integration and joint research across the different communities. Each challenge is motivated by a concise review of the state of the art, followed by several opportunities for future investigation. For each of the identified challenges, we discuss its maturity in terms of research and technological approaches, ranging from nascent solutions to those that have received community adoption, whereas the opportunities take the form of open research questions that need to be explored.

The remainder of the paper is structured as follows: we start by outlining our motivating use case scenario (Section 2) and presenting the relevant background (Section 3). A conceptual framework is then proposed (Section 4), accompanied by an instantiation based on our use case scenario (Section 5), after which we subsequently identify several challenges and opportunities (Section 6). We then conclude by proposing a research roadmap for the governance of autonomous agents on the web (Section 7).
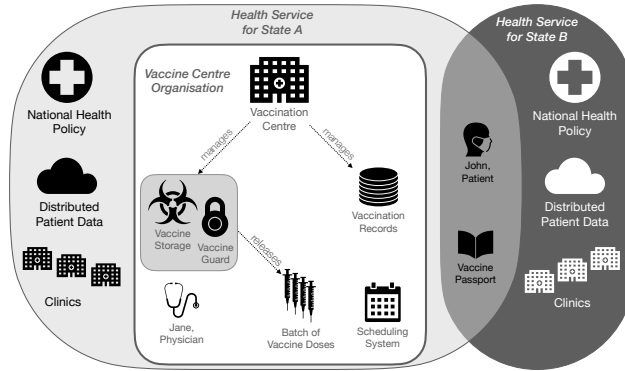
Fig. 1. Organisations, agents, things, and services in the scenario.

2 MOTIVATING SCENARIO

The synergy between autonomous agents that leverage Web technologies, and the resources (i.e., things, services, information) that they can exploit to achieve their goals can be illustrated through a motivating scenario that demonstrates the need for governance through the use of norms, policies, and preferences. Consider a scenario whereby there is a vaccination roll-out (for example, for the COVID-19 pandemic), where patients who request vaccinations may have differing personal circumstances. For example, John, the patient in Figure 1, may ask to be vaccinated early as he is the care giver for a vulnerable member of his family. As the demand for vaccines outstrips supply, policies exist that determine vaccination eligibility. Furthermore, as vaccines are available from different manufacturers (e.g., AstraZeneca and Pfizer-BioNTech) and can be of different types (e.g., mRNA or inactivated vaccines), these vaccination policies may vary depending on the recipient's personal health record and/or their preferences, as well as vaccine availability.

Patients may be registered to different clinics or health centres that follow local or national policies or guidance on health care. In this case, John is registered at a clinic in his country (labelled State B in Figure 1), but has a preference for vaccination near his current residential address in State A. Each country or state can be seen as having an *organisation* of different health centres (clinics, hospitals, and vaccination centres), following their own national health policy that prescribe a specific specification/format for patient medical records, which may be held under disparate data models and access policies. Patient medical records are available (subject to appropriate authorisation) via web services using secure protocols across the web infrastructure [135], and are encoded using established medical ontologies and vocabularies to facilitate record exchange within and across different national health organisations.

Vaccination centres store batches of vaccines within one or more temperature-controlled vaccine storage systems, where each storage system is responsible for both inventory management and the dispensation of the different COVID-19 vaccine batches from a specialised cold store via a robotic arm. The release and retrieval of vaccine batches is guarded by policies that must be satisfied to ensure appropriate use by authorised personnel (i.e., the vaccine guard in Figure 1). Once a batch of vaccines has been released, the vaccine doses should be used within a given time-frame to avoid spoilage and wastage, as they have a short shelf-life once thawed. Furthermore, a scheduling system determines which patients can be vaccinated in a given time-slot, based on vaccination demand and patient requirement (determined by the current vaccination policy that may change frequently). This scheduling system should ensure that no vaccines are wasted, whilst ensuring

that the policies determining which patients can receive which vaccines is adhered to. Thus, the vaccination centre could be considered as an organisation that coordinates and exploits a variety of disparate information technology (IT) systems integrated through a Web infrastructure, including data management, scheduling, patient-facing services, and IoT-based physical assets such as the robot arm and the automated vaccine stores. Typically, however, the task of orchestrating and using these different systems requires costly and time-consuming human intervention. Finally, once a vaccine has been administered, the patient's medical records should be updated, and the patient should be able to prove their vaccination status if required (e.g., using a vaccine passport [55]). The vaccine records should ideally be resilient to forgery whilst being privacy preserving and easy to administer [55]; thus they may utilise a passport mechanism that itself exploits web-based resources such as *verifiable credentials*[1], decentralised data platforms [142], blockchains [135], etc.

This scenario raises challenges due to the decentralised and dynamic characteristics of the involved organisations, policies, services, and stakeholders. Patients can request vaccination based on their interpretation of eligibility, which should then be validated by the vaccination centre. The handling of requests may require the collection of patient data from multiple sources and the mapping to a shared data model. The vaccination eligibility policy can change frequently due to, for example, the emergence of a new *variant of concern*, that may accelerate the need for vaccinating a specific population cohort or demographic. Changes to vaccination administration guidance may prioritise the use of certain types of vaccine over others for specific sub-groups (e.g., prioritising Pfizer-BioNTech over AstraZeneca, where possible, for certain patients based on medical risk assessments, or prohibiting certain vaccines for users where safety data is not available). Thus, the verification of vaccination eligibility for patients may rely on the aggregation of multiple policies, and on resolving inconsistencies between them. A further challenge involves ensuring that the process for adhering to the national prioritisation criteria is fair and transparent.

Additional legal and ethical challenges arise when considering the complete socio-technical system, including electronic health record access [81] and supply chains [127]. Finally, vaccination scheduling needs to take into account patient availability (to avoid no-show cases and thus avoid vaccine wastage), as well as stock availability. Scheduling is therefore a collaborative process involving factors such as the vaccination centre capacity, vaccine availability, and patient availability. However, availability data may be distributed across multiple sources and, for privacy reasons, cannot be held centrally.

This scenario underlines the need for systematic and scalable approaches for the governance of the different IT systems and IoT-based physical assets, taking into account the need to operate under different governance institutions, as well as interact across organisational boundaries (e.g., between countries). Such interactions must comply with applicable norms and policies encountered at different stages of the vaccination roll-out. For example, the European Commission recently proposed a Digital Green Certificate, recognised by all EU member states, that facilitates the safe free movement of citizens within the EU during the COVID-19 pandemic.[2]

Given the intrinsic openness of the Web, coupled with the fact that autonomous agents can act on behalf of both patients and medical practitioners that need access to critical medical applications, the need for regulation, security, and privacy are of utmost importance. Additionally, there is a need to facilitate coordination between stakeholders and ensure that relevant regulatory requirements are adhered to throughout.

---

[1]https://www.w3.org/TR/vc-data-model/
[2]https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/safe-covid-19-vaccines-europeans/covid-19-digital-green-certificates_en

3 BACKGROUND

The vaccine administration scenario detailed in Section 2 relies on the availability of a uniform access layer that sits on top of several different systems (e.g., data management, services, and IoT platforms). The Web provides the necessary infrastructure to integrate and make accessible all of these systems, effectively becoming an application architecture for the proposed MAS [39], on top of which autonomous agents may interact and cooperate to achieve common goals. In the following subsections, we present the relevant background in multiagent systems, the Semantic Web, and the WoT, followed by a discussion of the related work in norms, policies and preferences, with a focus on the governance of autonomous agents, both within and spanning those communities.

3.1 Multiagent Systems

A multiagent system is composed of a (dynamic) set of *agents* interacting inside a shared, possibly distributed, *environment* which itself comprises a dynamic set of *artefacts*. Agents are goal-oriented autonomous entities, encapsulating a logical thread of control, that pursue their tasks by communicating with other agents and by perceiving and acting upon artefacts within the environment. In essence, a MAS addresses the challenges of how agents may coordinate their efforts and cooperate in light of their autonomy [145]. Artefacts model any kind of (non-autonomous) resource or tool that agents can use and possibly share to achieve their goals. An agent perceives the observable state of an artefact, reacts to events related to state changes, and performs actions that correspond to operations provided by the artefact's interface. The coordinated and organised activities taking place in the system result from the concurrent and complex tasks handled by *groups of agents* interacting with each other, or acting within an environment. Such activities may lead to recurrent patterns of cooperation captured by agent *organisations*. Changes in the state of the environment may also lead agents to react and possibly affect the state of the organisation.

Research into multiagent systems has led to a number of concrete programming models.[3] These models[4] are concerned with agent-oriented programming [20], interaction and protocol languages [123], environment infrastructures [146], and agent organisation model and management systems [58]. The results produced so far have clearly demonstrated the importance of these concepts and abstractions for the development of multiagent applications. Additionally, a variety of languages, tools, and platforms for agent-oriented programming (MAOP) have been developed and application success stories exists (e.g., [49]). This type of research is often referred to under the umbrella of *Engineering Multiagent Systems (EMAS)*. An overview and a comparative analysis of several prominent MAOPs can be found in [91]. One of the most prominent underlying architectures used by many agent-oriented programming systems is the *Belief-Desire-Intention (BDI)* architecture, which models: knowledge (i.e., *beliefs*) that the agent knows about, either through observation of the environment or interaction with other agents; goals (i.e., *desires*) that the agent would like to bring about; and goals and plans of action (i.e., *intentions*) that the agent is currently focused on.

From an agent development environment perspective, the Jade platform [12] provides a variety of behaviours (one-shot, cyclic, contract net) and is still available, although the last release dates back to 2017. Although Jade does not directly provide support for BDI-based agents, they can be added through extensions such as Jadex [22]. Jack [27] is an example of a closed source BDI architecture, whereas the practical Agent Programming Language (2APL) is another open source language that retains BDI semantics [47]. GOAL [75] offers a further BDI architecture which is actively maintained, whereas SPADE[5] is a recently introduced Python-based BDI platform. The

---

[3]Refer to the proceedings of the EMAS or PAAMS series for broad overviews.
[4]The models presented here reflect the relevant state of the art with respect to different MAS and are by no means exhaustive.
[5]https://spade-mas.readthedocs.io/en/latest/index.html

JaCaMo MAOP framework, based on the JaCaMo conceptual meta-model [17], offers first-class abstractions to program the agents working environment and their organisation, in addition to offering the Jason interpreter for the BDI-based *AgentSpeak* language [20].

Whilst MAOP is thriving within the academic community, industrial adoption of MAOP technologies is in its infancy, and standardisation efforts such as FIPA [63] (that superseded KQML) have received little attention in recent years [99].

3.2   Agents and the Semantic Web

Attempts to tightly integrate autonomous agents and Web technologies date back to the vision of the *Semantic Web* of the early 2000s. Berners-Lee et al. [14] originally envisioned "a web of data that can be processed directly and indirectly by machines", in which intelligent agents act on behalf of humans, by searching for and understanding relevant information published on the web or acquired via services. Such information could potentially be made available by multiple sources, using alternative ontologies, often with different provenance. Autonomous agents rely on communication languages and protocols to exchange data and coordinate their behaviour and thus collaborate. Early approaches based on speech acts [8], focused on message types or *performatives* (e.g., *request*, *inform*, and *promise*) based on a folk categorisation of the intended meaning of the communication. This evolved through the DARPA funded Knowledge Sharing Effort (KSE) resulting in a communication language, the *Knowledge Query Manipulation Language (KQML)*, defining the mechanism by which agents communicated; and an ontology language, the *Knowledge Interchange Format (KIF)*, describing the knowledge that the performative referred to [62]. Although agents could perform services on behalf of their peers, discovered through capability registries [51], service invocation occurred as a by-product of requesting information. This contrasts with the notion of *web services* and *things*, which use web-based communication protocols, whereby the invocation of services could be requested explicitly (in a similar manner to calling methods or functions within a programming language) by providing the relevant input parameters, as data or knowledge fragments.

The prominent view from a Semantic Web perspective is that multiagent systems operate on the Web through the provision of services, using HTTP as the de facto standard transport protocol. Additionally, the Semantic Web community have developed standards, protocols, vocabularies, ontologies, and knowledge representation formalisms to facilitate the integration of machine-processable data from diverse sources at scale, using the existing web infrastructure. As such, the two communities diverged due to different priorities, though there is increasing recognition [39] that the Web is a natural application architecture for MAS and can support different types of interactions between agents and resources.

From a knowledge representation perspective, standards such as RDFS [24] and OWL [69] facilitate the representation of complex knowledge about agents, services, things and their relationship in an explicit and processable way. An example is the Provenance ontology (PROV-O), a data model for workflows expressed using agents, their actions, and other assets.[6] Additionally, reasoning engines have been developed that are capable of reasoning over OWL ontologies, albeit often with some restrictions (cf., Pellet[7], HermiT[8], FACT++[9], Racer[10], and RDFox[11]). However, the use of ontologically grounded annotations for services within agent communication pre-dates the

---

[6]https://www.w3.org/TR/prov-o/
[7]https://www.w3.org/2001/sw/wiki/Pellet
[8]http://www.hermit-reasoner.com/
[9]http://owl.cs.manchester.ac.uk/tools/fact/
[10]http://www.ifis.uni-luebeck.de/ moeller/racer/
[11]https://www.oxfordsemantic.tech/

Semantic Web [57, 77], and in some cases the Web itself [76]. Semantic Web service research exploited both F-Logic [84] as used by WSMO [119], and DAML-S [6] (based on the DARPA Agent Markup Language) which evolved into OWL-S [98]. Other approaches to support service utilisation were developed using OWL, e.g., the OWL ontology for protocols, OWL-P [52], or using federated service discovery mechanisms such as the semantically annotated version of UDDI [109]. These frameworks and ontologies were key in facilitating the discovery and use of services by autonomous agents, and provided an alternative communication paradigm built on web-based infrastructure. In addition, from the knowledge perspective, bespoke protocols were developed to support the decentralised management and exchange of knowledge and information amongst networks of agents or peers [131].

Other efforts include the provision of infrastructures for supporting the cleaning and valida-tion of the data published on Linked Open Data Platforms; e.g., LOD Laundromat [11][12] and OOPS [113].[13] Such techniques help detect errors in the data exchanged between agents and things. The SPARQL [72] query language facilitates federated querying over distributed data sources accessible via the web, whereas the Linked Data Platform [130] can be used to manipulate RDF data via HTTP operations. Approaches have also been proposed to enrich SPARQL with qualitative and quantitative preferences [70, 111] to select query results that satisfy user-defined criteria.

In recent years the Semantic Web community has broadened its focus beyond knowledge repre-sentation, reasoning, and querying to include knowledge extraction, discovery, search, and retrieval. However, many of the proposed tools and techniques have yet to be used extensively within MAS or by the MAS community. A recent survey [85] identified several open research challenges and opportunities in relation to the suitability of existing proposals for autonomous agent use cases, the combination of symbolic and sub-symbolic AI techniques for enhancing agent learning, and the development of tools and techniques for validation and verification.

3.3 Agents and the Web of Things

The *Web of Things (WoT)* [90] refers to the Internet of Things (IoT) with an application of Web standards and technologies for improving interoperability of IoT devices and infrastructure. Things are resources that can be acted upon or queried via APIs (e.g., WoT scripting API [88]); *autonomous goal-driven agents*[14] thus can make use of a WoT environment via WoT technologies and become part of the WoT ecosystem. Indeed, bringing agents to the Web requires more than simply exploiting Web protocols (such as HTTP [61]) and data formats (e.g., XML [23], RDF [44]). The communication infrastructure used by agents should comply with an architectural style based on well-defined principles, such as *Representational State Transfer (REST)* [60] as instantiated in the Architecture of the World Wide Web [79].[15] Furthermore, for things to be used without human intervention, they must be formally described. To this end, the W3C published the *Thing Description* [80] standard, which specifies how a JSON-LD representation of thing affordances (i.e., properties or actions) via Web APIs can be provided. In addition, the *WoT Discovery* [35] standard provides a mechanism for the automatic discovery of thing descriptions (thus obviating the need to hard-code the location of such descriptions beforehand). These standards support improved heterogeneity by decoupling agents from thing implementation details.

The WoT activity highlights the importance of metadata with clear semantics, and made their standards, especially thing descriptions, compatible with RDF and Semantic Web technologies. In

---

[12]https://github.com/LOD-Laundromat/LOD-Laundromat

[13]http://oops.linkeddata.es

[14]Here, we refer to agents in the sense of multiagent systems as discussed in Section 3.1.

[15]Note that the specification of the Web architecture defines the concept of *Web agents* as "a person or a piece of software acting on the information space on behalf of a person, entity, or process".

fact, even before a standardisation effort for the WoT started, multiple initiatives suggested the use of the Semantic Web to improve IoT systems [118]. More precisely, in REST style hypermedia systems such as the WoT, things and agents are resources that interact by producing and consuming hypermedia about their state and the artefacts surrounding them [38]. All resources are identified through IRIs[16] to support global referencing, irrespective of contextual information. Therefore, resources can be represented through semantic descriptions that are expressed in a uniform data exchange format such as RDF using terms from some standardised and interlinked vocabulary expressed in OWL [69]. This standardised knowledge model hides the specifics of the implementation and facilitates interconnected resources that can be queried by exposing SPARQL endpoints. Of particular interest to WoT environments are the vocabularies that describe sensors and actuators (SOSA/SSN [71]), provenance (PROV [92]), and temporal entities (OWL-Time [40]).

The WoT provides a natural substrate for multiagent systems based on the vision that systems of interconnected things should be open and easily reconfigurable, and therefore such systems should comprise autonomous and collaborative components. This notion was supported by Singh and Chopra [125] who argue that IoT systems need the kind of decentralised intelligence that MAS provides. Likewise, Ciortea et al. [39] recommend integrating the Web and MAS to leverage the proven benefits of hypermedia systems for MAS. Importantly, these papers emphasise governance as a major challenge.

The technologies that emerge from the WoT community are often industry-oriented and paralleled by standardisation efforts. A recent example is the abstract *WoT architecture* design document [90], supported by the *Thing Description* [80] and the *WoT Scripting API* [88] specifications, for which a reference implementation is provided.[17] Although these technologies are more mature than MAOP technologies from an engineering perspective, and have a clear path to industry adoption, they lack the rich abstractions related to agents and autonomy that MAOP technologies provide. For example, the notion of a *servient*, as introduced in the WoT architecture design document can be considered an evolutionary step from a stricter server-client separation; a notion that is considered simplistic within the MAS community. Recent approaches have sought to form a bridge between the MAOP and WoT technology ecosystems [36, 37]; however, this line of research is young and the corresponding technologies are nascent.

3.4 Norms, Policies and Preferences

Norms, policies, and preferences can help govern autonomous agent behaviour. The term *norm* has several meanings in natural language and is used widely in economics and social science. In MAS, the term "norm" typically expresses a deontic concept (e.g., a prohibition, permission, obligation, or dispensation). A coherent set of norms, i.e., created and evaluated as a unit, is referred to as an institution [103]. The same understanding of norms is found in the Semantic Web literature, where there is also a body of work focusing on policy specification and enforcement. Here, *policy* is an overarching term used to refer to a variety of system constraints, whereas the term preferences is primarily used in connection with privacy and personal data protection.

The study of norms is a long-running and active line of research within the MAS community, as evidenced by numerous Dagstuhl seminars [5, 48], and a handbook on the topic [32]. Normative MAS [16] are realised and characterised in multiple ways, including those based on: (1) the agents reasoning capabilities; (2) whether norms are implicit or explicit; and (3) whether or not the architecture includes monitoring and enforcement mechanisms.

---

[16]Internationalised Resource Identifiers [54]
[17]https://github.com/eclipse/thingweb.node-wot

Agent capabilities vis-à-vis norms typically fall into three categories: (i) *norm unaware*, whereby agents may be regimented by external agencies to enforce norm compliance [7]; (ii) *norm-aware*, where agents may choose whether or not to comply with norms, depending on the alignment of their goals with those norms, the penalties for non-compliance, and the likelihood of enforcement [122]; and (iii) *value aware*, whereby agents, in addition to being norm-aware, are able to participate in norm creation and norm revision, by reasoning about the values supported (or not) by particular norms [41]. Thus, compliance in normative systems depends on how individual agents reason and adapt to norms at both design and run time [93, 136, 141]. Implicit norms that reside within the agents themselves are expressed through agent behaviour, but are not otherwise externally discernible, whereas explicit or referenceable norms may have an abstract representation involving variables and a grounded (detached) representation in an entity such as a contract [124], institution [53, 56, 103], or organisation [17, 139]. Agents that are norm or value aware should be able to: (i) recognise norms; (ii) decide whether they want to follow them; and (iii) adapt their behaviour according to the norms, if they decide to do so. Such agents may additionally be able to engage in norm revision processes. Norms, and more broadly *conventions* or *social norms* [94], are established in an agent society in one of two ways, namely top-down and bottom-up [101, 149].

In *top-down* systems, norms are identified as part of the MAS design process and are either: hard-coded into the agents' behaviour (implicit representation), eschewing any form of normative reasoning and narrowing the scope for behavioural adaptation; or are prescriptive and explicitly represented, and thus external to the agents, typically represented in the form of abstract regulations (for example, ungrounded terms over variables) that, as a result of agent actions, become detached (for example, grounded terms over literals). The n-BDI variant [43] is a BDI-based agent architecture that allows for the internalisation of norms where the design suggests an agent-internal process that synthesise norm-style rules based on observed behaviour, whereas N-Jason [93] agents perceive institutional facts, which they internalise as beliefs and hence incorporate in their reasoning. Norms designed offline, however thoughtfully crafted for the long-term, are at risk of losing relevance in open, always-on, environments such as the Web, because it is not possible to anticipate all eventualities at design time. Furthermore, drift in the agent demographic or in systems goals, are likely to make norm revision essential over any sufficiently long system lifetime. With explicit norms, any norm change will affect the entire population. Such changes can be effected through a human-in-the-loop approach, where human designers revise the norms and then switch the system over at some suitable point; such as through a shutdown/reboot sequence, or the use of norm-aware planning [122]. In the latter case, an agent must manage a plan sequence that although initially compliant, may cease to be part of the way through the plan due to the change in norms. Such an agent must also be able to check that its learned way of achieving a goal is compliant with the new norms, perhaps by means of some oracle [107], or by being able to acquire a fresh plan that is compliant.

In *bottom-up* systems, an individual agent decides whether or not to adopt a norm: with implicit norms, it may seek advice from others or apply indirect reinforcement learning over its observations, as a basis for prediction, possibly in combination with a *strategy update function* [149]. In such systems, norms are deemed to have emerged once they have been adopted by a sufficiently large fraction of the population; this is typically 90% in most of the literature, and 100% in some cases (which is hard to achieve), or assumes a simple majority, which can risk oscillatory outcomes. However, convergence (this term appears to be used interchangeably with emergence in the literature [102]) is a function of the *capabilities* of the agents. Emergence with explicit norms depends on agent reasoning capabilities. An agent might inform the regulator that it wants to take a particular action in a particular state (without sanction) – the agent knows what it wants but not how to get it—as a request to change the norms without having to reason about norm

representation. A more difficult approach is that an agent might propose a new (abstract) norm – the agent knows how to define a new norm to get what it wants [73, 102]. As above, changes have to be actioned, which could be as outlined previously, although pluralist approaches are possible, as put forward by Ostrom [106], or by using one of the many voting mechanisms. The challenge for an agent then becomes how to decide which way to vote, which depends on their reasoning capabilities: are they able to evaluate the consequences of the norm change; and are they selfish (i.e., vote "yes" if the change is individually beneficial, e.g., increases their utility) or altruistic? (i.e., vote "yes" if the change is collectively beneficial). More sophisticated still would be the use of argumentation to determine if the revision is consistent with the population's values [122, 134].

In the early days, Semantic Web researchers proposed general policy languages, such as KAoS [21], Rei [82] and Protune [19], which cater for a variety of different constraints (access control, privacy preferences, regulatory requirements, etc.). A prominent early attempt to provide a semantic model of polices as soft constraints for agents was OWL Polar [121], an OWL DL explicit policy representation language. OWL Polar aims to fulfil the essential requirements of policy representation, reasoning, and analysis, where policies are system-level principles of ideal activity that are binding upon the components of that system, and thus are used to regulate the behaviour of agents [121]. Over the years the Semantic Web community have also proposed policy languages that are tailored to better cater for access control, privacy preferences, licensing, and regulatory governance requirements, including detailed surveys, for example, of the various policy languages, and the different access control enforcement strategies for RDF [87]. From a privacy perspective, the Platform for Privacy Preferences Project (P3P) [42] specification, deemed obsolete in 2018, aimed to allow websites to express their privacy preferences in a machine readable format that could be interpreted by agents that could automate decision making on behalf of humans. The P3P initiative, despite having failed, inspired subsequent work on representing and reasoning over privacy preferences, such as using OWL [65], catering to more expressive privacy preferences [89], and representing consent for personal data processing [18].

Many existing proposals rely on WebID [120], a community-driven specification that offers an identification mechanism making use of Semantic Web technologies to provide password-less authentication. An extension of WebID (specifically WebID-OIDC that relies on OpenID Connect[18]) is used in the *Solid* project. Solid[19] is an ongoing initiative, lead by Tim Berners-Lee, aimed at deploying a distributed Linked Data infrastructure for governing one's personal data, which is built on top of Linked Data Platforms. Additionally, there has been work on *usage* control in the form of licensing [28, 66–68, 143], and more recently, policy languages have been used as a means to represent regulatory constraints [50, 108]. The Open Digital Rights Language [64, 78], although primarily designed for licensing, has been extended to cater for: access policies [133]; requests, data offers and agreements [132]; and regulatory policies [50]. Usage control, however, often proves challenging for organisations and users, and any constraints imposed on the use of data need to ensure that policies are applied consistently across organisations and that there are robust propagation mechanisms preventing policies from becoming invalid [45, 46]. The notion of FAIR ICT Agents [86] is based on FAIR (Findable, Accessible, Interoperable and Reusable) principles [147], where ICT denotes *interactive intelligent agents that are constrained via goals, preferences, norms and usage restrictions*. Thus far, the WoT standards offer only limited support for norms, policies and preferences, which are currently described in guidelines targeted at human developers rather than as declarative, machine-readable statements usable by agents [117].

---

[18]https://openid.net/connect/
[19]https://solidproject.org/

Although research on norm-aware agents has made reasonable progress to date, much remains to be done to elevate human oversight to align with the three categories [74]: *human-in-the-loop*, where there may be human intervention in each decision cycle; *human-on-the-loop*, where there is human intervention in the design cycle and operation monitoring; and *human-in-command*, where there is human oversight of the overall system, including the means to decide when and how to engage the AI system. The motivated scenario presented herein draws on human-on-the-loop and human-in-command, and indeed it is these levels of abstraction that inspire the governance framework introduced in Section 4, since those are the system characteristics we aim to facilitate.

4 CONCEPTUAL FRAMEWORK

The overarching goal of this section is to identify governance entities, their relations and their purpose, with no aim to be prescriptive in their instantiation. In doing so, we propose a blueprint for the governance of socio-technical systems that can be instantiated in a variety of ways, using a variety of concrete software components. Thus, this section aims to provide guidance for developers on how different parts of an agent governance system fit together and the functions that they contribute. Our objective here is to enable a range of solutions, fit for different purposes, realisable through available (rather than prescribed) software, but still coherent through the framework set out in the three layers shown in Figure 2.

4.1 Global View

In order to provide something actionable for designers and implementors, we ground our framework for the governance of autonomous agents on the Web through three layers that structure the various entities and abstractions needed for the development of socio-technical systems on the Web. Each layer is assigned concepts that are necessary for governance: norms, policies, and preferences (as illustrated in Figure 2). The way in which these different parts are realised, and how they interact is dependent on various design decisions. In setting out this framework, we draw on and organise existing work on norms, policies, and preferences (as described in Section 3) to cater for abstract requirements for the governance of socio-technical systems. This gives rise to the following three layers:

Reactive Things & Services Layer. This layer comprises non-autonomous entities in the environment. As seen in Section 3, such entities are key notions of the WoT architecture [79] for which
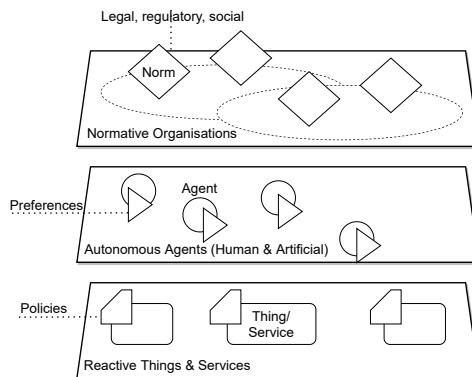

Fig. 2. Conceptual Framework for Governing Agents on the Web structured along three layers. Interactions (not represented here in this Figure) take place within each of the layer and between the layers.

first-class abstractions are proposed for specifying and discovering them and other entities within the MAOP approach (e.g., artefacts in the JaCaMo meta-model [17]). Such entities are perceived and acted upon by agents. We propose the use of *policies* for dealing with the governance of such non autonomous entities, following the same approach adopted by Semantic Web community. These policies state who can access them, and constraints on their usage (if any). Semantic Web technologies such as OWL-POLAR [121] can provide a means to implement, manage, and enforce policies that constrain access to things and services, and the affordances they provide.

Autonomous Agents (Human & Artificial) Layer. The agents layer is composed of entities that autonomously perceive and act upon their environment (i.e., things and services) and interact with the other entities. Agents are the main abstractions for specifying and managing autonomous behaviours. In contrast to the conventional model of programs on the Web as *servers* or *clients*, the WoT architecture introduces a *servient* that can both pro-actively access other things and services and reactively respond to requests from other things and services. In addition, servients can host one or several things. Whilst the Web architecture does not provide first-class abstractions for autonomy, it is possible to distinguish between "agentified" things that exhibit pro-active behaviours and reactive things by introducing custom properties into the W3C WoT *Thing Description* [83]. Agents have *preferences* that inform and constrain their actions with respect to things, web services, and other agents. Preferences control the local reasoning and decision-making undertaken by the agents, and can thus support governance. In traditional deliberation architectures for autonomous agents, preferences are specified (or emerge) as part of the often complex reasoning cycles. Hence, the management of these preferences given the presence of norms and policies can be challenging. Semantic Web approaches that consider preferences (e.g., SPARQL with preferences [112]) can enable declarative preference management, especially when an agent's preferences are to be considered.

Normative Organisations Layer. In MAOP, organisations are first-class abstractions [17] that group agents and their governance (i.e., norms). Although the WoT architecture does not provide such abstractions, its security and privacy guidelines reflect similar notions to organisational norms. Whilst the previous two layers (discussed above) included governance concepts dedicated to the local governance of each entity (e.g., policies for thing, preferences for agent), this layer addresses the governance of autonomous entities *participating in the system*. This layer manages abstractions for the logical grouping of agents with a particular purpose, and the provision of legal, regulatory, and social norms that may possibly span multiple organisations. However, organisations are entirely virtual and passive (i.e., shaped by their members), thus it is up to these member agents to stipulate, comply with (or violate), enforce, and evolve organisational norms. Semantic Web technologies such at ODRL [64] allow for the formalisation of norms for specific domains and purposes; hence, they can be integrated seamlessly with the more abstract MAOP abstractions for organisations and norms that are agnostic to these details.

From a MAS perspective, this framework is coherent with the JaCaMo meta-model [17]; from a WoT perspective, it is coherent with the WoT architecture [79]; and furthermore, it is coherent with the Semantic Web perspective, although with enhancements with respect to policies, preferences, and norms. It is worth noting that our conceptual framework provides *software engineering abstractions*. Analogously to the Web architecture, we do not recommend a one-to-one mapping of software abstractions to physical entities (devices). Considering Web architecture standards, the WoT Scripting API supports, for example, the instantiation of multiple things as part of one *servient*, which may represent a single physical machine.

4.2 The Layers

In this section, we return to each of the layers introduced in Section 4.1 and detail their composition and their governance.

4.2.1 Reactive Things & Services. As defined in Section 3.3, things are physical objects that are endowed with network capacities that allow one to make use of their functions in a digital environment. For example, they can be sensors that provide measurement data through the Internet, or actuators that can be triggered from a Web API. Therefore, in the WoT context, things resemble and are sometimes assimilated within web services. These web services are normally purely digital entities that simply exchange data via their input parameters and output results. In the WoT architecture, things may be autonomous, whereas, in contrast, our conceptual framework distinguishes between autonomous agents (which may be things in the WoT architecture) and *reactive* things.

When it comes to things and services, policies serve many purposes. Access control policies ensure that only authorised agents use specific things and services. Here, there is a need to provision both authentication and authorisation mechanisms, and policies may help resist security threats. Additionally, policies may govern the use of data that is produced by things and services; e.g., to ensure personal data protection or intellectual property rights.

From a policy governance perspective, it is useful to distinguish between *enforcement* and *compliance* of the autonomous agents acting on these things and services given their respective policies. Enforcement means that any violations are prevented, whereas compliance means there is a need for retrospective conformance checking.

4.2.2 Autonomous Agents (Human & Artificial). In contrast to (reactive) things and services, agents are entities that pursue their own goals autonomously. They determine the necessary actions that should be executed on the things and services situated in the environment. In the MAS literature, several agent architectures that are based on the different properties exhibited by the agents have been proposed [148]. They range from purely reactive (i.e., those that respond to stimuli without complex symbolic reasoning to reason about future actions) to deliberative ones (those that maintain a symbolic world model for reasoning about plans and decision making) [29]. A notable example of deliberative architectures is the BDI architecture [115], where agents are programmed using their mental attitudes such as beliefs, desires, wishes, etc. [20] and that is one of the mainstream architectures for cognitive agents in MAS.[20] This contrasts with reactive architectures (such as the subsumption architecture [25, 137]) typically used by robotic systems, whereby *behaviours* define the actions a robot should perform as a consequence of some stimuli (e.g., from sensor data or direct communication). Many hybrid agent architectures [26, 148] combine elements of both reactive and deliberative ones, where prominence is often given to the reactive aspect over the deliberative aspect (such as obstacle avoidance versus goal deliberation). Our conceptual model focuses on governance and is agnostic with respect to any particular architecture, and thus cater for the heterogeneity of agents.

In addition to taking decisions on their own, agents may also coordinate with humans or with other agents to adjust and align their goals with the other agents' goals and identify joint goals, and as such, they may communicate with other agents or human users by exchanging messages. We address the various means of interaction among agents in Section 4.3 below.

Each agent maintains a representation of its internal state that is built from the agent's internal reasoning, from its perceptions of the environment, i.e., the observable state of the things and services deployed in the system, and from its interactions with other agents. Acting on behalf of human users (e.g., assistant agents) or abstract entities (e.g., service agents), agents manage

---

[20]A number of different MAOP frameworks that adopt a BDI architecture were discussed previously in Section 3.1.

preferences that guide their decision process. It is important to differentiate between agents developed by the application designer and those that enter the system at run-time. This differentiation emphasises the level of control the application designer has over the agent with respect to its internal state. It also justifies the proposition of two levels of governance within our conceptual model: preferences for local and individual control; and organisations for global and collective control. In our conceptual framework, preferences cover many dimensions, ranging from privacy preferences to moral values or ethical principles. Additionally, there can be either agreement or conflict between preferences and access control as defined in the previous layer, due to the fact that an agent may need to verify someone's identity, and based on this determine what information to disclose. Being part of a MAS, the reasoning and decision mechanisms of the agents are enriched with mechanisms to reason over several factors including: norms; regulatory requirements coming from the organisation in which the agent participates; and over policies or access control rules attached to the resources, things, and services with which the agent interacts.

4.2.3 Normative Organisations. Organisations act as coordination mechanisms by which agents work together to achieve their joint goals. The design of agents within an organisation focuses mainly on the agents' capabilities and constraints, as well as on organisational concepts such as *roles* (or functions, or positions), *groups* (or communities), *tasks* (or activities) and *interaction protocols* (or dialogue structures); therefore on what relates the structure of an organisation to the externally observable behaviour of its agents [58]. Organisations usually have a structure defined by: (i) groups, whereby agents are classed together and possibly organised hierarchically; and (ii) roles, whereby agents assume various duties. For example, agents can belong to multiple organisations, be part of various groups, assume different roles (possibly at the same time), and join or leave organisations at will. Organisations can be formed at design time or emerge due to interactions between agents at run-time.

The dynamics of the organisational structure, for example an agent changing its role or joining a group, is governed by rules that are formalised as the norms of the organisation. Norms define what communication is possible, allowed, or forbidden between agents. An organisation is a means to regulate agent behaviour, and such organisations may be governed by norms, including laws and regulations adopted from the social setting or jurisdiction and those legislated within the organisation. The organisation structure and its normative part are described in such a way that agents can autonomously take part in the organisation and regulate themselves automatically with the aim of achieving their (individual or collective) goals. However, a formal, explicit encoding of norms is necessary to facilitate automated compliance and conformance checking.

4.3 Interactions Within and Among the Layers

Within a MAS that is fully aligned with our layered conceptual framework, several interactions may take place within each of the layers and across them (i.e., both inter and intra-layer interactions). Similarly to the Web architecture, the conceptual framework is protocol-agnostic. Some protocols may be chosen based on the underlying things and services to be used; whereas other protocols would be custom to the desired agent-agent interactions; and some of these latter protocols may be designed whereas others evolve. We identify the following types of interactions in the conceptual framework:

Agent-to-agent interactions. Agents can interact with other agents *directly*, by exchanging messages or acting upon each other, or *indirectly*, by observing each other's actions on the reactive things and services of their environment. Because agents are autonomous, the requests that one agent sends to another are handled at the discretion of the receiving agent. In comparison to the interaction with Web services, interaction with an agent may imply a higher likelihood that the

response deviates in a complex and nuanced manner from the requested resource.

Agent-to-thing/Thing-to-agent interactions. Agents *proactively* interact with things and services by *acting* upon them, *accessing* their properties, and by listening to (perceiving) events that things and services emit.

Thing-to-thing interactions. While things and services are purely *reactive*, they may interact with other things or services as part of a reaction chain. In this context, existing standards that are part of the Web architecture can be applied for basic communication, but more expressive approaches may be required to manage norms, policies, and preferences, for example when a thing communicates on behalf of an agent across organisational boundaries.

Agent-to-organisation/Organisation-to-agent interactions. An agent's preference depends on the norms of the organisations that the agent is a part of. However, because the agent is an autonomous entity, it may choose to not adopt an organisational norm. At the same time, the agent may attempt to change an organisation's norm, for example by proposing a norm update that then requires approval by a majority of the organisation's agents.

Thing-to-organisation/Organisation-to-thing interactions. In contrast to agents, things and services cannot directly affect organisations. Things and services can be implemented to dynamically adopt policies that reflect organisational norms, and the state of a thing or service can be considered by an organisation, but in both directions, the organisation is the *leading system*.

Organisation-to-organisation interactions. Several organisations may have (unidirectional or bidirectional) dependencies. For example, in a hierarchy of organisations, the norms of lower ranking organisations may depend on norms that are specified on a higher level in the hierarchy; still, a higher ranking organisation may have some norms that depend on the norms of multiple lower ranking organisations (consider dependencies between a federated state and its federal entities).

In the MAS community, interaction protocols are typically designed from a global perspective and aim to facilitate interaction and coordination between agents. A protocol specifies the permitted enactments; i.e., the possible sequences of message exchanges. Proposals for languages for interaction protocols include process algebra [59], Petri Nets [116], and information protocols [123]. Petri nets may then be mapped to models that are more accessible to human users, such as Business Process Model and Notation (BPMN) diagrams. In practice, protocol design and protocol discovery can go hand-in-hand: in particular, Petri Net-based protocols (*processes*) can be *mined* from IT system event logs [140], which, for example, can be used for organisational compliance checking [30]. Recently, *agent system mining* has been proposed as a novel process mining variant that focuses on the agents that participate in one or several (organisational) processes, i.e., on the micro-level instead of on the macro-level process view that an organisation imposes [138]. In the service-oriented community, the notion of a *choreography* is similar to an interaction protocol [9], in that a choreography describes interactions between services from a global perspective.

Human-Agent interactions are typically modelled as conversations between the different agents, i.e., *dialogues* [144]. A dialogue has a normative aspect: it is regulated by norms, and can establish new norms. In normative systems, dialogue protocols are specific notations for norms that specify the violation contexts. Utterances in a dialogue can be seen as moves in the underlying protocol that create obligations and permissions for the participating agents. Of particular interest are persuasive dialogues, where an agent can *convince*, *suggest*, or *command*. Agents can use persuasive dialogues to convince other agents to add new beliefs; to enter into some form of negotiation; or, in the case of the command, a new violation rule is introduced thus creating a new obligation [15].

Besides the protocols themselves and their logical organisation in process choreography, the Agent Communication Language (ACL) and the agreed vocabularies are crucial when it comes to the interaction and co-ordination of agents in a MAS (Section 3.2). The Web architecture specification lists *properties*, *actions*, and *events* as the central abstraction of its interaction model.
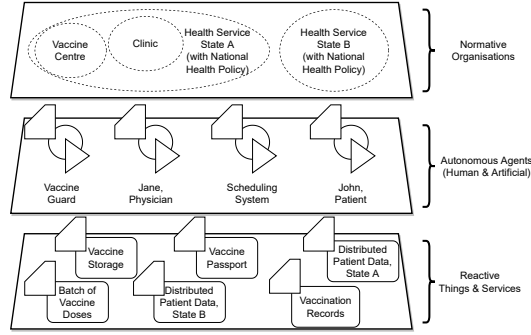
Fig. 3. Normative organisations, autonomous agents, and reactive things and services in the scenario.

In our conceptual framework, organisations, agents, things and services may expose properties and generate events, but only agents may execute actions. From a governance perspective, there is a need for policy, preferences, and norm-aware interaction protocols. For example, agents may need to authenticate themselves to other agents as well as to things and services, whereas collaborating agents may need to engage in preference elicitation and negotiation, and norms may need to be communicated and possibly agreed upon by agents that form part of an organisational structure.

5 USE CASE REVISITED

In this section, we demonstrate how a *normative* MAS that leverages web services, things, and Semantic Web technologies could be used to realise our motivating scenario (Section 2). We show how several example situations can be modelled using the proposed conceptual framework (Section 4) and highlight technologies that could be used to instantiate our governance framework.

5.1 The Global Setting

Agents encapsulate knowledge, goals, and preferences corresponding to the autonomous entities involved in the vaccination process. The resulting conceptual model is illustrated in Figure 3.

An assistant agent is in charge of managing personal data on behalf of a patient (e.g., the patient John). A physician agent is in charge of managing administrative tasks to act on behalf of the physician (e.g., Jane). Other types of agents access the things and services (i.e., a *vaccine guard* agent controls access to the freezer), and to manage the vaccination process by collecting patients' data and checking their eligibility (i.e., *scheduling system* agents). It is worth noting that, contrary to the other agents, the first two kinds of agents (i.e., assistant agents and physician agents) may not be under complete control of the stakeholders who develop and own the application.

We introduce a *vaccination centre organisation* to delimit the *vaccination* application and to provide scope for the adherence to regulations and behaviours for both artificial and human agents that are part of this structure. To this end, the organisation specifies roles and norms, whereby the roles are used to structure agent responsibilities, and the norms (i.e., duties, rights, and interdictions) regulate the vaccination application. Agents with a given role are expected to fulfil the corresponding norms. The vaccination centre is, in turn, part of the *health service organisation* of a particular state, in which *clinic organisations* complement the normative framework provided by state and vaccine centres. In addition to norms, the definition of the organisation may impose hard constraints on its composition that should be enforced by service policies. For example, by stating an upper limit on the number of agents that can adopt specific roles, the vaccination application may consequently limit the number of patients or physicians that may enter the organisation.

5.2 Illustrative Use of the Conceptual Framework with our Motivating Scenario

The following paragraphs describe the use of the conceptual model in situations derived from the motivating scenario. For narrative convenience, we use the terms *obligation*, *permission*, and *authority* in an informal sense.

5.2.1 Initialising the Vaccination Application. At the launch of the application, the *vaccination centre organisation* is created, by endowing the agents that support the business processes within the *vaccination centre organisation* with the roles necessary to fulfil their goals. The definition of the organisation (e.g., the roles and distribution of norms on the roles) is published as a web resource in a machine readable and understandable format, accessible to any agent wishing to become a member of that organisation. The current state of the organisation (i.e., which agent is assigned the various roles) is published and updated as necessary, over the entire lifetime of the organisation. The *freezer agent* adopts the *guard role*, which results in it being assigned the duty of managing access to the inventory of COVID-19 vaccine doses stored in the freezer. It obtains the permission to use the robotic arm to retrieve a vaccine dose when asked, and to deliver it to the staff. The *manager agent* is assigned the *organiser role*, and consequently inherits the obligation to compile lists of eligible patients based on the patient data and the vaccination eligibility policy. The *data agent* adopts the *collector role* and obtains the authority to collect personal information about each patient requesting a vaccination appointment; it also has the obligation to verify the patient's eligibility for receiving the vaccine as well as the obligation to solicit patients through dissemination channels when vaccine doses and scheduling slots are available.

5.2.2 Joining the Vaccination Centre Organisation. When a patient obtains the credentials to access and use the vaccination application, the *assistant agent* acting on behalf of the patient is provided access to the web resource describing the organisation. After reasoning over its obligations and authorities, as imposed by the *vaccination centre organisation*, the agent decides to adopt the role. The *assistant agent* subsequently acquires the obligation to provide access to the patient medical data. This role may create internal conflicts between preferences provided by the patient and the obligations assumed when the agent took on the *patient assistant role*. After accessing and reasoning about the description of the *vaccination centre organisation*, the *physician agent* discovers that it has the obligation to coordinate with agents that are assigned to other roles. To assume the *medical practitioner role*, the *physician agent* must authenticate itself; upon adopting the role, it captures the associated permissions, obligations, and authorisation for further decision making. The same process of role adoption applies to other agents.

5.2.3 Assessing Patient Eligibility. While assigned to the *organiser role*, the *manager agent* takes into account its preferences in defining the patient information collection policy, and sends it to the agent with the *collector role* (as stated by the organisation definition). Fulfilling its obligation, the agent checks the eligibility of all arriving patients so that each dose is only administered to an eligible patient and that doses are administered before their expiry date.[21] To fulfil its goals, the agent therefore requests that agents adopting the new *patient assistant role* share the necessary patient personal data. It is worth noting that agents with the *collector role* need to consider the obligations stated by the organisation as soft constraints, and identify contexts in which these constraints may be relaxed. Further complications may arise if any of the agents attempt to negotiate relaxations of these obligations, either in anticipation of, or after a (perceived or factual) violation. For example, an *assistant agent* can negotiate an exception for a potential obligation violation by using computational

---

[21]In a practical scenario, we would not expect 100% compliance with this constraint, but rather that the number of excess or wasted doses (relative to administered doses) does not exceed a specified threshold.

models and algorithms of formal argumentation where the *assistant agent* believes that the data it has for its patient satisfies the eligibility criteria. On behalf of the organisation, the agent that has adopted the *organiser role* is in charge of the definition of the eligibility policy, and consequently may interact with the agent in charge of the data collection by granting or denying the request for an exception, or even by updating the organisation's norms in order to accept the request.

5.2.4 Administering the Vaccine. When administering the vaccine, the agent with the *physician role* must respect the priority order for vaccine administration as defined by the agent in charge of the collection; for example, the elderly and vulnerable population must be vaccinated first, unless respecting the priority order implies wasting the dose. Importantly, the physician must not violate this priority order by, for example, preferentially vaccinating friends or relatives. In some cases, the *physician agent* must choose between either administering a vaccine dose despite the eligibility status being uncertain, or allowing the dose to expire. For example, existing regulations indicate that administering a vaccine dose to close relatives is impermissible, but the agent may conclude that in accordance with practitioner norms (such as the Hippocratic Oath), it is preferable if vaccines are not wasted. In such cases, the agent may need to prepare a defence strategy to avoid sanctioning, for example, via argumentation approaches [10, 13]. These issues merely relate to the permissions needed for the obligations implied by opening the fridge. Additional challenges arise when considering the complete socio-technical system, including electronic health record access [81] and supply chain integration [127].

6   CHALLENGES AND OPPORTUNITIES

We now present research challenges (i.e., technical limitations of existing proposals) and opportunities (i.e., open research questions) related to normative agents on the Web. The four horizontal challenges that characterise the contributions of norm-based multiagent systems for the Web are described below, in addition to two orthogonal challenges that need to be tackled in order to address the horizontal issues (and illustrated in Figure 4). For each area, we label the challenges in the context of limitations of the state of the art and subsequently identify future research opportunities. Figure 4 also indicates the practical maturity of each challenge, from *nascent* (blue sky challenges with basic research potential, using a white background) via *developing* (basic research with immediate practical potential, using a white-grey gradient background) to *practical* (challenges that can be addressed primarily from an engineering perspective using a grey background).

6.1   Relating Norms and Interaction Protocols

Challenges. A broad challenge in engineering normative MAS is that we need a way to operationalise norms in the sense of giving them a computational interpretation. Interaction protocols characterise interactions based on message order and occurrence – that is, in operational terms. However, it is nontrivial to produce protocols that are as flexible as necessary, yet enactable in a decentralised manner, while at the same time being verifiably correct. Although the W3C provides Web-based standards for retrieving and querying machine-readable data, these standards do not cater for usage constraints, such as access policies, intellectual property rights, and privacy preferences. In our scenario, an agent may, for example, want to decide with whom and in which context it shares its vaccination status. Existing work on interaction protocols [63] largely focuses on request-response interactions and imposes restrictions on computation for scenarios involving the interaction of three or more parties [33, 59]. In particular, traditional approaches entwine control flow details into the protocol, thereby making it difficult to separate them from the content, for which a declarative meaning can be specified. Prior work on specifying protocols based on norms (commitments) [31, 97] was hindered by the lack of declarative specification of the constraints on messages. More
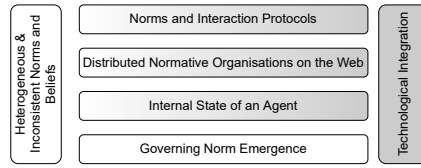
Fig. 4. Overview: Research challenges for normative MAS and the Web

recent approaches describe causality and integrity constraints on messages declaratively [123]; such protocols (whilst sufficiently flexible to support all enactments of the stated norms) can grow quite large [126], but emerging verification approaches aim to tackle this challenge [128].

Opportunities. When it comes to operationalising norms, from a service-provisioning perspective, there is a need to develop policy-aware querying and data retrieval protocols; whereas from an agent interaction perspective, norms should be mapped both to the agent platform and the environment. This raises several important questions, including: (i) How can we design norm-aware dynamic interaction protocols? (ii) How can existing querying and data retrieval protocols be extended, such that they are policy aware? (iii) What new languages are needed to facilitate norm governance? (iv) How do we model and reason with respect to norm changes and temporal validity?

When operating in deployed systems modelled according to the conceptual framework described in Section 4, these interaction protocols also need to take into account that the autonomous agent layer includes human agents which might have conflicting requirements, and therefore strategies might need to be employed in order to resolve conflicts. This raises a number of additional questions that need to be addressed, including: (i) How do we ensure protocol compliance by human agents? (ii) How do we model protocols that implement persuasion? (iii) What mechanisms do we use to resolve conflicting requirements?

6.2 Distributed Normative MAS and Open Organisations on the Web

Challenges. Organisations, institutions, and contracts are useful abstractions to structure norms and make them accessible to agents. Although agents have the choice of joining such structures, they may be subject to conditions that regulate their admission (and exit), as well as there being an expectation to comply with the organisation's norms. Due to the scale of the Web, numerous permanent, ephemeral, or evolving structures may exist. Consequently, an agent needs to be able to discover and reason about such organisations and the corresponding norms. In the vaccination scenario, for example, an agent may need to be able to discover organisations that model the healthcare systems of other jurisdictions that potential patients may need to refer to, when claiming that they are eligible to receive a specific type of vaccine dose. Ontologies facilitate the discovery of services [109], and their use as a means to represent organisations is promising. A major challenge related to the distributed management of such structures [17] is to monitor and enforce norm compliance, and to instantiate organisations, agents, or complete multiagent (sub-)systems at run-time on the Web, which is an emerging line of research in the MAS community [2–4]. Another challenge is that agents require abstractions and mechanisms to build and adapt organisations on the fly [34]. Additionally, an agent may participate in multiple structures that operate at different timescales and scopes, and hence accommodating their diversity is nontrivial.

Opportunities. Addressing the above challenges requires answers to the following research questions, in the context of Web-based and WoT-based technology ecosystems: (i) How can agents discover organisational norms on the Web? (ii) How can norm compliance be monitored and

enforced in dynamic scenarios, in which agents, organisations, or entire (sub-)MAS are instantiated at run-time? (iii) How can autonomous agents create and change organisations *on-the-fly*? (iv) How can normative organisations accommodate agents that participate in multiple organisations, with potentially inconsistent norms, and partial semantic interoperability? (v) How can compliance checking and enforcement approaches that are prevalent in the information systems literature be adapted and applied to normative MAS on the Web?

6.3 Internal State of an Agent and Norms

Challenges. An agent should be able reason about norms, taking into consideration its internal state (e.g., its beliefs, goals, and intentions), and explain its normative reasoning to others. This is, for example, important for the *administering the vaccine* scenario (Section 5.2.4), when a decision is to be made about whether to administer a vaccine to a patient whose eligibility status is uncertain, using for example qualitative methods [110], argumentation [10] or sub-symbolic techniques such as classification or Bayesian networks. Several languages express and support automated reasoning about agent internals, such as beliefs, desires or goals, and intentions. However, challenges exist when it comes to reasoning not only with respect to goals, but privacy preferences, regulatory constraints, and norms. The problems here not only relate to knowledge representation – how to represent all these aspects so they are accessible to an agent – but also to the impact such enrichment may have on deliberation performance. The most popular (symbolic) agent architecture is practical reasoning by means of beliefs, desires and intention (BDI), but this offers no guarantees about how long the deliberation cycle may take. Indeed this falls to the developer, in as much as they can endeavour to keep rules relatively simple whilst limiting the number of overlapping conditions, but the actual response time is out of their hands. Additionally, there is the more significant problem that arises due to the fact that agent architecture research has primarily focused on agent internals. Irrespective of whether they are comprised of symbolic or sub-symbolic aspects (such as reasoning, reinforcement or probabilistic learning), such architectures are not normally conceived or designed for interfacing with non-agent technological frameworks and their underlying abstractions (for example as provided by the W3C Web architecture and related standards).

Opportunities. From a norms perspective, important open questions in specification and enforcement include: (i) How can we ensure consistent representation of and adherence to norms? (ii) How should a governance architecture be designed in which rational agents are incentivised to comply with norms? (iii) Can a norm violation be excused, based on explanation or argumentation? (iv) Could transparency facilitate the persuasiveness of the explanation or argument? (v) How do we ensure that an agent is aware of the implications of violating a norm? (vi) How do we cater for agents that are not rational (as such agents may not be designed for criteria such as rationality)? (vii) How should performance limitations affect normative decision-making in practical reasoning architectures? (viii) How can normative reasoning be implemented in real-world agent architectures (which may only be agent-oriented in the broader sense)? (ix) How can agent architectures be better integrated with web technologies and standards?

6.4 Governing Norm Emergence

Challenges. Approaches for the governance of norm emergence are dependent on the capabilities of the agents in a MAS, bearing in mind that population properties may not be homogeneous. In our example scenario, the governance of norm emergence is, for example, important to facilitate vaccinations (i.e., the belief that getting vaccinated is, while typically not mandatory, good for one's health and more broadly for the public health at large), and to balance "hard" rules and "soft" recommendations to decrease the spread of COVID-19. The challenges here include modelling

and managing the spread of beliefs and counter-beliefs, the potential resolution of contrary positions through argumentation, and how to make hard and soft policies accessible to different agent architectures with different reasoning capabilities. We differentiate between a decentralised approach to norm emergence with implicit norms, where the norms emerge through the interactions of agents – [1] is one example of such a scheme – and various centralised approaches to the governance of norm emergence [101], which latter we classify by adapting the oversight terminology put forward by the High-Level Expert Group on Artificial Intelligence (AI HLEG) [74, §B.II.1.1]: (i) an external agency observes the behaviour of the population to identify patterns of behaviour and revise norms imposed by that agency to optimise for system goals (external agent/human on-the-loop) – for example Morales et al. [100] look at individual norms in isolation – while the general case of revising a consistent body of norms remains open; (ii) agents propose norm revisions to an external agency, which then implements them subject to an assessment of how those revisions contribute towards system goals (external agent/human in-command), which also remains open; and (iii) agents propose norm revisions and system participants, which may include humans, use an internal decision-making mechanism to establish which changes will be implemented (internal agents/humans in-the-loop). The uHelp system illustrates some preliminary steps in this direction [104], but relegates software agents to a supporting role. The human-in-the loop, human-in-command, human-on-the-loop approaches are closely related to the different strategies that regulate persuasive dialogues [15], where commands introduce new obligations, whilst suggestions introduce new beliefs. However, open challenges relate to devising persuasion strategies and the corresponding obligations.

Opportunities. In order to govern norm emergence in Web-based MAS and their socio-technical contexts, one needs to answer – for example – the following questions:

(i) How can the emergence of norms in conjunction with governance decisions be monitored and managed in Web-based MAS? (ii) What roles do human-on-the-loop, human-in-the-loop, and human-in-command approaches play in the context of the preceding question, and what are the engineering implications of these different human interaction approaches? (iii) What is necessary to maintain alignment of the (evolving) value preferences of participants with the norms that govern them: when does one norm change become many changes? (iv) Which collective decision-making mechanisms are best suited for all agent and for mixed human-agent systems? (v) What is the appropriate capacity for agent reasoning about (self-)governance? Is wanting to do something that is prohibited, or not wanting to do something that is obliged a sufficient statement of intention? (vi) How can an agent and/or a human evaluate the consequences of norm revisions? Will they create a fresh problem while resolving the current one? (vii) How can oscillatory norm change be prevented? An agent-dominated system could potentially change faster than a human-in-command can evaluate the changes.

6.5 Heterogeneous and Inconsistent Norms and Beliefs

Challenges. In heterogeneous information systems, we cannot reasonably assume that norms and policies are globally accepted and thus agents may hold inconsistent beliefs about them. For example, in our vaccination scenario, vaccine administration policies, eligibility requirements, and IT system landscapes may differ between two federated states A and B. However, a patient who moves permanently from A to B should ideally be able to receive the first vaccine dose in state A, and a dose of a matching or complementary type in state B after a reasonable time interval. For aligning norms and policies of sub-entities, reaching (partial) agreements in the face of conflicting beliefs regarding norms and policies is an important challenge that needs to be tackled to enable normative distributed MAS on the Web; using, for example, long-running lines of research on

agreement technologies [105] and formal argumentation [10]. Currently, the body of research on belief revision and argumentation-based reasoning is, however, poorly integrated with practical engineering perspectives; standardisation efforts like the specification of an *argument interchange format* [114] exist, but have not found substantial adoption.

Opportunities. There are several open questions when it comes to enabling reasoning and decision-making in the face of inconsistent norms and beliefs of agents on the Web, including: (i) To what extent is there a practical need for engineering abstractions that treat conflict and inconsistency in the context of a normative Web (of Things)? (ii) What systematic approaches to drawing inferences and making decisions in a Web (of Things) governance context can be designed, implemented, and standardised as software engineering abstractions? (iii) How can existing research on belief revision and argumentation-based reasoning be made more accessible both from an engineering and a standardisation perspective?

6.6 Technological Integration

Challenges. In order to facilitate the practical applicability of research on norms and policies for autonomous agents on the Web, it is crucial to build bridges across the technology ecosystems of the different communities. Section 3 provides an overview of the technology ecosystems that have emerged from the MAS, Semantic Web, and WoT communities. To summarise, in the WoT community, engineering-oriented work has been conducted in a highly practice-oriented manner, in close alignment with industry practitioners as well as standardisation bodies such as the W3C. An example of practice-oriented work can be observed though W3C IoT standards that feature an abstract architecture [90] and an interface specification (W3C WoT Scripting API) [88], supported by a JavaScript reference implementation.[22] Research on engineering autonomous agents and MAS has primarily gained traction within the academic community [99], and standardisation attempts such as FIPA[23] have lacked significant adoption. Adjusting agent-oriented programming and software engineering approaches to better serve the Semantic Web and WoT communities is a way for the MAS community to move their engineering research closer to practice. This lets us conclude that while each of the communities has its own thriving technology ecosystem, a key challenge lies in integrating these ecosystems, which exhibit different degrees of practical maturity.

Opportunities. The above observations raise two questions: (i) How can the technology ecosystems of (normative) MAS and the Semantic Web be integrated with the WoT, and in particular with the W3C Web architecture? (ii) How can issues of practical maturity be mitigated (by the integration strategy)? With respect to *(i)*, we argue that an integration strategy can employ a combination of two approaches across two dimensions that requires pragmatic trade-offs, considering the discrepancies between the technology ecosystems and their underlying conceptual abstractions.
Approach 1: Full-Fledged Framework Adoption. In order to facilitate implementations that build on research in the different communities, interfaces that integrate Semantic Web, MAOP, and WoT technologies can be devised that either re-implement their abstractions or integrate technology frameworks and specification languages [36–38]. A benefit of this approach is that it facilitates the adoption of powerful abstractions and technology ecosystems developed in these communities. A disadvantage, however, is that this approach can cause a high technological overhead.
Approach 2: Modular Abstraction Adoption. In order to facilitate implementations that build on technology stacks established in industry, *minimally viable abstractions* on norms and autonomous

---

agents can be implemented as reusable modules in mainstream programming languages; or alternatively, specific features of complex technology platforms can be exposed as service-oriented interfaces. This strategy resembles a call to action as made in Logan's *Agent Programming Manifesto* [96] and allows for deliberate trade-offs between conceptual richness and practical feasibility by avoiding the overhead (on conventional developers) of having to learn unfamiliar programming paradigms. For example, one may adopt JaCaMo's capabilities for modelling organisations and artefacts via a Java-based technology stack, and defer adopting Jason [20] since it involves a custom language for agent-oriented programming.

Broadly, the two approaches can be considered analogous to the *integrated system* (Approach 1) and *best of breed* (Approach 2) strategies for implementing large scale enterprise systems [95]. In an actual implementation scenario, these approaches represent the extremes of a scale with valuable trade-offs in between. We suggest that this trade-off is initially made from a conceptual perspective (*Which programming abstractions are useful in a given scenario?*) and followed using a technology perspective (*Which technologies do I want to use to implement these abstractions?*).

With respect to the second question, we argue that the strategy should prioritise mature technologies, and if necessary re-implement the requisite abstractions in technology stacks that are established in industry practice. Specifically, we might consider the WoT standards and technologies as a mature foundation on which to place Semantic Web and MAS technologies. An example of a synergy is in extending WoT servients to autonomous agents without necessarily committing to a BDI architecture for those agents.

7 CONCLUSION

This paper discusses the relevance of norms, policies, and preferences for governing complex socio-technical multiagent systems on the Web. The key challenge – the conceptual and technological integration of normative concepts with WoT abstractions and systematic evaluation of the practical usefulness of the integration results – is aligned with the general challenge for autonomous agents on the Web to transfer the rich theoretical achievements of the broader MAS community to the practical and engineering-oriented WoT community, and to facilitate real-world applications at scale. While the challenge of transferring research on normative agents and multiagent systems into engineering practice is well-known and generally acknowledged, this paper has taken the emergence of new Web standards, as well as the increased research interest in Web-based MAS, as a starting point to provide a new and broad perspective on it, with a focus on the Web and Web of Things Architecture standards.

In this context, the paper proposes a conceptual framework that serves to define the role played by various norms, policies and preferences when it comes to complex socio-technical MAS on the Web, and demonstrated it via a simple but realistic scenario.

In addition, the paper provides a research roadmap outlining the technical and theoretical research challenges and opportunities to support complex socio-technical MAS governance on the Web. In particular, this roadmap calls for: (i) relating norms and interaction protocols; (ii) incorporating normative organisations and norm governance approaches into WoT architectures and standards; (iii) combining agent reasoning to relate policies, preferences, and norms; (iv) tackling the emergence of norms for flexible governance; (v) designing reasoning methods about norms in the face of inconsistency; and (vi) cautiously advancing Semantic Web and (normative) MAS tools and frameworks into practice via the WoT.

REFERENCES

[1] Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. 2018. Robust Norm Emergence by Revealing and Reasoning about Context: Socially Intelligent Agents for Enhancing Privacy. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*. IJCAI, Stockholm, 28–34. https://doi.org/10.24963/ijcai.2018/4

[2] Huib Aldewereld, Julian Padget, Wamberto Vasconcelos, Javier Vazquez-Salceda, Paul Sergeant, and Athanasios Staikopoulos. 2010. Adaptable, Organization-Aware, Service-Oriented Computing. *IEEE Intelligent Systems* 25, 4 (2010), 26–35. https://doi.org/10.1109/MIS.2010.93

[3] Cleber Jorge Amaral and Jomi Fred Hübner. 2020. Jacamo-Web is on the Fly: An Interactive Multi-Agent System IDE. In *Engineering Multi-Agent Systems*, Louise A. Dennis, Rafael H. Bordini, and Yves Lespérance (Eds.). Springer International Publishing, Cham, 246–255.

[4] Cleber Jorge Amaral, Jomi Fred Hübner, and Timotheus Kampik. 2020. Towards Jacamo-rest: A Resource-Oriented Abstraction for Managing Multi-Agent Systems. In *Proceedings of the 14th Workshop-School on Agents, Environments, and Applications*, Gleifer Vaz Alves, Gustavo Guiménez Lug, André Pinz Borgeso, and Carlos Eduardo Pantoja (Eds.). UTFPR, Ponta Grossa, Parana, Brazil, 140–151.

[5] Giulia Andrighetto, Guido Governatori, Pablo Noriega, and Leon van der Torre. 2012. *Normative Multi-Agent Systems (Dagstuhl Seminar 12111)*. Technical Report 3. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

[6] Anupriya Ankolekar, Mark Burstein, Jerry R. Hobbs, Ora Lassila, David Martin, Drew McDermott, Sheila A. McIlraith, Srini Narayanan, Massimo Paolucci, Terry Payne, and Katia Sycara. 2002. DAML-S: Web Service Description for the Semantic Web. In *The Semantic Web — ISWC 2002*, Ian Horrocks and James Hendler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 348–363.

[7] Josep Lluís Arcos, Marc Esteva, Pablo Noriega, Juan A. Rodríguez-Aguilar, and Carles Sierra. 2005. Engineering open environments with electronic institutions. *Engineering Applications of Artificial Intelligence* 18, 2 (2005), 191–204. https://doi.org/10.1016/j.engappai.2004.11.019

[8] John Langshaw Austin. 1962. *How to do things with words*. Oxford University Press, Oxford, England.

[9] Adam Barker, Christopher D. Walton, and David Robertson. 2009. Choreographing Web Services. *IEEE Transactions on Services Computing* 2, 2 (2009), 152–166. https://doi.org/10.1109/TSC.2009.8

[10] Pietro Baroni, Dov M. Gabbay, Massimiliano Giacomin, and Leendert van der Torre. 2018. *Handbook of formal argumentation*. College Publications, Rickmansworth.

[11] Wouter Beek, Laurens Rietveld, Hamid R. Bazoobandi, Jan Wielemaker, and Stefan Schlobach. 2014. LOD Laundromat: A Uniform Way of Publishing Other People's Dirty Data. In *The Semantic Web – ISWC 2014*, Peter Mika, Tania Tudorache, Abraham Bernstein, Chris Welty, Craig Knoblock, Denny Vrandečić, Paul Groth, Natasha Noy, Krzysztof Janowicz, and Carole Goble (Eds.). Springer International Publishing, Cham, 213–228.

[12] Fabio Luigi Bellifemine, Giovanni Caire, and Dominic Greenwood. 2007. *Developing Multi-Agent Systems with JADE (Wiley Series in Agent Technology)*. John Wiley and Sons, Inc., Hoboken, NJ, USA.

[13] Trevor Bench-Capon, Henry Prakken, and Giovanni Sartor. 2009. *Argumentation in Legal Reasoning*. Springer US, Boston, MA, 363–382. https://doi.org/10.1007/978-0-387-98197-0_18

[14] Tim Berners-Lee, James Hendler, and Ora Lassila. 2001. The Semantic Web. *Scientific American* 284, 5 (2001), 34–43.

[15] G. Boella, J. Hulstijn, and L. van der Torre. 2004. Persuasion Strategies in Dialogue. In *Procs. of CMNA Workshop at ECAI'04*. Online Proceedings, University of Liverpool, Valencia, 29–32.

[16] Guido Boella, Leendert Van Der Torre, and Harko Verhagen. 2006. Introduction to normative multiagent systems. *Computational & Mathematical Organization Theory* 12, 2-3 (2006), 71–79.

[17] Olivier Boissier, Rafael H. Bordini, Jomi Hübner, and Alessandro Ricci. 2020. *Multi-agent oriented programming: programming multi-agent systems using JaCaMo*. MIT Press, Cambridge.

[18] Piero A Bonatti, Sabrina Kirrane, Iliana M. Petrova, and Luigi Sauro. 2020. Machine Understandable Policies and GDPR Compliance Checking. *KI-Künstliche Intelligenz* 34, 3 (2020), 303–315.

[19] Piero A. Bonatti and Daniel Olmedilla. 2007. *Rule-Based Policy Representation and Reasoning for the Semantic Web*. Springer Berlin Heidelberg, Berlin, Heidelberg, 240–268. https://doi.org/10.1007/978-3-540-74615-7_4

[20] Rafael H. Bordini, Jomi Fred Hübner, and Michael Wooldridge. 2007. *Programming Multi-Agent Systems in AgentSpeak Using Jason (Wiley Series in Agent Technology)*. John Wiley & Sons, Inc., Hoboken, NJ, USA.

[21] Jeffrey M. Bradshaw (Ed.). 1997. *Software agents*. MIT press, Cambridge, MA.

[22] Lars Braubach, Alexander Pokahr, and Winfried Lamersdorf. 2005. Jadex: A BDI-Agent System Combining Middleware and Reasoning. In *Software Agent-Based Applications, Platforms and Development Kits*, Rainer Unland, Monique Calisti, and Matthias Klusch (Eds.). Birkhäuser Basel, Basel, 143–168.

[23] Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, and François Yergeau. 2008. *Extensible Markup Language (XML) 1.0 (Fifth Edition), W3C Recommendation 26 November 2008*. W3C Recommendation. World Wide Web Consortium (W3C). https://www.w3.org/TR/2008/REC-xml-20081126/

[24] Dan Brickley and Ramanathan V. Guha. 2014. *RDF Schema 1.1, W3C Recommendation 25 February 2014*. W3C Recommendation. World Wide Web Consortium (W3C). http://www.w3.org/TR/2014/REC-rdf-schema-20140225/

[25] Rodney A. Brooks. 1991. Intelligence without representation. *Artificial Intelligence* 47, 1 (1991), 139–159.

[26] Joanna Bryson and Lynn Andrea Stein. 2001. Modularity and Design in Reactive Intelligence. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI 2001, Seattle, Washington, USA, August 4-10, 2001*, Bernhard Nebel (Ed.). Morgan Kaufmann, Palo Alto, California, 1115–1120.

[27] Paolo Busetta, Ralph Rönnquist, Andrew Hodgson, and Andrew Lucas. 2019. JACK intelligent agents - Components for intelligent agents in JAVA. Available from https://aosgrp.com/media/research/jack/busetta99jack.pdf. AgentLink Newsletter Issue 2.

[28] Elena Cabrio, Alessio Palmero Aprosio, and Serena Villata. 2014. These Are Your Rights - A Natural Language Processing Approach to Automated RDF Licenses Generation. In *The Semantic Web: Trends and Challenges - 11th International Conference, ESWC 2014, Anissaras, Crete, Greece, May 25-29, 2014. Proceedings (Lecture Notes in Computer Science, Vol. 8465)*. Springer, Cham, 255–269.

[29] Roberta Calegari, Giovanni Ciatto, Viviana Mascardi, and Andrea Omicini. 2021. Logic-based technologies for multi-agent systems: a systematic literature review. *Autonomous Agents and Multi Agent Systems* 35, 1 (2021), 1. https://doi.org/10.1007/s10458-020-09478-3

[30] Filip Caron, Jan Vanthienen, and Bart Baesens. 2013. Comprehensive rule-based compliance checking and risk management with process mining. *Decision Support Systems* 54, 3 (2013), 1357–1369. https://doi.org/10.1016/j.dss.2012.12.012

[31] Amit Chopra and Munindar P. Singh. 2004. Nonmonotonic Commitment Machines. In *Advances in Agent Communication: Proceedings of the International Workshop on Agent Communication Languages (ACL 2003) (Lecture Notes in Artificial Intelligence, 2922)*, Frank Dignum (Ed.). Springer, Melbourne, 183–200. https://doi.org/10.1007/978-3-540-24608-4_11

[32] Amit Chopra, Leendert van der Torre, Harko Verhagen, and Serena Villata. 2018. *Handbook of Normative Multiagent Systems*. College Publications, Rickmansworth.

[33] Amit K. Chopra, Samuel H. Christie V, and Munindar P. Singh. 2020. An Evaluation of Communication Protocol Languages for Engineering Multiagent Systems. *Journal of Artificial Intelligence Research (JAIR)* 69 (Dec. 2020), 1351–1393. https://doi.org/10.1613/jair.1.12212

[34] Amit K. Chopra and Munindar P. Singh. 2016. From Social Machines to Social Protocols: Software Engineering Foundations for Sociotechnical Systems. In *Proceedings of the 25th International World Wide Web Conference*. ACM, Montréal, 903–914. https://doi.org/10.1145/2872427.2883018

[35] Andrea Cimmino, Michael McCool, Farshid Tavakolizadeh, and Kunihiko Toumura. 2020. *Web of Things (WoT) Discovery, W3C First Public Working Draft 24 November 2020*. W3C First Public Working Draft. World Wide Web Consortium (W3C). http://www.w3.org/TR/2020/WD-wot-discovery-20201124/

[36] Andrei Ciortea, Olivier Boissier, and Alessandro Ricci. 2018. Engineering World-Wide Multi-Agent Systems with Hypermedia. In *Engineering Multi-Agent Systems - 6th International Workshop, EMAS 2018, Stockholm, Sweden, July 14-15, 2018, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 11375)*, Danny Weyns, Viviana Mascardi, and Alessandro Ricci (Eds.). Springer, Cham, 285–301. https://doi.org/10.1007/978-3-030-25693-7_15

[37] Andrei Ciortea, Olivier Boissier, Antoine Zimmermann, and Adina Magda Florea. 2018. Give Agents Some REST: Hypermedia-driven Agent Environments. In *Engineering Multi-Agent Systems*, Amal El Fallah-Seghrouchni, Alessandro Ricci, and Tran Cao Son (Eds.). Springer International Publishing, Cham, 125–141.

[38] Andrei Ciortea, Simon Mayer, Olivier Boissier, and Fabien Gandon. 2019. Exploiting Interaction Affordances: On Engineering Autonomous Systems for the Web of Things. In *Second W3C Workshop on the Web of Things The Open Web to Challenge IoT Fragmentation*. W3C, Munich, Germany, 0–4. https://hal.archives-ouvertes.fr/hal-02196903

[39] Andrei Ciortea, Simon Mayer, Fabien Gandon, Olivier Boissier, Alessandro Ricci, and Antoine Zimmermann. 2019. A Decade in Hindsight: The Missing Bridge Between Multi-Agent Systems and the World Wide Web. In *AAMAS 2019 - 18th International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems., Montréal, Canada, 5. https://hal-emse.ccsd.cnrs.fr/emse-02070625

[40] Simon Cox and Chris Little. 2017. *The Time Ontology in OWL, W3C Recommendation 19 October 2017*. W3C Recommendation. World Wide Web Consortium (W3C). http://www.w3.org/TR/2017/REC-prov-o-20171019/

[41] Stephen Cranefield, Michael Winikoff, Virginia Dignum, and Frank Dignum. 2017. No Pizza for You: Value-based Plan Selection in BDI Agents. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence,*

(IJCAI), Carles Sierra (Ed.). ijcai.org, Melbourne, 178–184. https://doi.org/10.24963/ijcai.2017/26

[42] Lorrie Cranor, Marc Langheinrich, Massimo Marchiori, Martin Presler-Marshall, and Joseph Reagle. 2002. *The Platform for Privacy Preferences 1.0 (P3P1.0) Specification*. W3C Recommendation. World Wide Web Consortium. Obsolete recommendation since 30 August 2018.

[43] N. Criado, E. Argente, P. Noriega, and V. Botti. 2010. Towards a Normative BDI Architecture for Norm Compliance. In *COIN@*. Citeseer, Informal Proceedings, Lyon, France, 65.

[44] Richard Cyganiak, David Wood, and Markus Lanthaler. 2014. *RDF 1.1 Concepts and Abstract Syntax, W3C Recommendation 25 February 2014*. W3C Recommendation. World Wide Web Consortium (W3C). http://www.w3.org/TR/2014/REC-rdf11-concepts-20140225/

[45] Enrico Daga, Mathieu d'Aquin, Alessandro Adamou, and Enrico Motta. 2016. Addressing exploitability of Smart City data. In *2016 IEEE International Smart Cities Conference (ISC2)*. IEEE, Piscataway, New Jersey, United States, 1–6. https://doi.org/10.1109/ISC2.2016.7580764

[46] Enrico Daga, Mathieu d'Aquin, Aldo Gangemi, and Enrico Motta. 2015. Propagation of Policies in Rich Data Flows. In *Proceedings of the 8th International Conference on Knowledge Capture* (Palisades, NY, USA) *(K-CAP 2015)*. Association for Computing Machinery, New York, NY, USA, Article 5, 8 pages. https://doi.org/10.1145/2815833.2815839

[47] Mehdi Dastani. 2008. 2APL: A Practical Agent Programming Language. *Autonomous Agents and Multi-Agent Systems* 16, 3 (2008), 214–248.

[48] Mehdi Dastani, Jürgen Dix, Harko Verhagen, and Serena Villata. 2018. Normative Multi-Agent Systems (Dagstuhl Seminar 18171). *Dagstuhl Reports* 8, 4 (2018), 72–103.

[49] Fernando De la Prieta, Sara Rodríguez-González, Pablo Chamoso, Juan Manuel Corchado, and Javier Bajo. 2019. Survey of agent-based cloud computing applications. *Future Generation Computer Systems* 100 (2019), 223–236. https://doi.org/10.1016/j.future.2019.04.037

[50] Marina De Vos, Sabrina Kirrane, Julian Padget, and Ken Satoh. 2019. ODRL Policy Modelling and Compliance Checking. In *Rules and Reasoning*, Paul Fodor, Marco Montali, Diego Calvanese, and Dumitru Roman (Eds.). Springer International Publishing, Cham, 36–51.

[51] Keith Decker, Katia Sycara, and Mike Williamson. 1997. Middle-agents for the internet. In *IJCAI-97: Proceedings of the Fifteenth International Joint Conference on Artificial Intelligence* (Nagoya, Japan). Morgan Kaufmann, San Francisco, California, 578–583.

[52] Nirmit Desai, Ashok U. Mallya, Amit K. Chopra, and Munindar P. Singh. 2006. OWL-P: A Methodology for Business Process Development. In *Agent-Oriented Information Systems III, 7th International Bi-Conference Workshop, AOIS2005, Utrecht, Netherlands, July 26, 2005, and Klagenfurt, Austria, October 27, 2005, Revised Selected Papers (Lecture Notes in Computer Science, 3529)*. Springer, Berlin, 79–94. https://doi.org/10.1007/11916291_6

[53] Mark d'Inverno, Michael Luck, Pablo Noriega, Juan A. Rodríguez-Aguilar, and Carles Sierra. 2012. Communicating open systems. *Artificial Intelligence* 186 (2012), 38–94. https://doi.org/10.1016/j.artint.2012.03.004

[54] Martin J. Dürst and Michel Suignard. 2005. *Internationalized Resource Identifiers (IRIs)*. Technical Report. Internet Engineering Task Force. http://tools.ietf.org/html/rfc3987

[55] Marc Eisenstadt, Manoharan Ramachandran, Niaz Chowdhury, Allan Third, and John Domingue. 2020. COVID-19 Antibody Test/Vaccination Certification: There's an App for That. *IEEE Open Journal of Engineering in Medicine and Biology* 1 (2020), 148–155. https://doi.org/10.1109/OJEMB.2020.2999214

[56] Marc Esteva, Julian Padget, and Carles Sierra. 2002. Formalizing a Language for Institutions and Norms. In *Intelligent Agents VIII*, John-Jules Ch. Meyer and Milind Tambe (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 348–366.

[57] Dieter Fensel and Enrico Motta. 2001. Structured Development of Problem Solving Methods. *IEEE Trans. Knowl. Data Eng.* 13, 6 (2001), 913–932. https://doi.org/10.1109/69.971187

[58] Jacques Ferber, Olivier Gutknecht, and Fabien Michel. 2004. From Agents to Organizations: An Organizational View of Multi-agent Systems. In *Agent-Oriented Software Engineering IV*, Paolo Giorgini, Jörg P. Müller, and James Odell (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 214–230.

[59] Angelo Ferrando, Michael Winikoff, Stephen Cranefield, Frank Dignum, and Viviana Mascardi. 2019. On Enactability of Agent Interaction Protocols: Towards a Unified Approach. In *Proceedings of the 7th International Workshop on Engineering Multi-Agent Systems (EMAS) (Lecture Notes in Computer Science, Vol. 12058)*. Springer, Montréal, 43–64. https://doi.org/10.1007/978-3-030-51417-4_3

[60] Roy Thomas Fielding. 2000. *Representational State Transfer (REST)*. Ph.D. Dissertation. University of California, Irvine. Chapter 5 of Roy Fielding's thesis.

[61] Roy Thomas Fielding and Julian Reschke. 2014. *Hypertext Transfer Protocol (HTTP/1.1): Semantics and Content*. Technical Report. IETF. http://tools.ietf.org/html/rfc7231

[62] Tim Finin, Richard Fritzson, Don McKay, and Robin McEntire. 1994. KQML as an Agent Communication Language. In *Proceedings of the Third International Conference on Information and Knowledge Management* (Gaithersburg, Maryland, USA) *(CIKM '94)*. Association for Computing Machinery, New York, NY, USA, 456–463. https://doi.org/10.1145/

191246.191322

[63] FIPA. 2003. FIPA Interaction Protocol Specifications. http://www.fipa.org/repository/ips.html FIPA: The Foundation for Intelligent Physical Agents.

[64] Nicoletta Fornara and Marco Colombetti. 2019. Using Semantic Web technologies and production rules for reasoning on obligations, permissions, and prohibitions. *AI Communications* 32 (2019), 319–334. https://doi.org/10.3233/AIC-190617

[65] Diego Zuquim Guimarães Garcia and Maria Beatriz Felgar de Toledo. 2008. A web service privacy framework based on a policy approach enhanced with ontologies. In *2008 11th IEEE International Conference on Computational Science and Engineering-Workshops*. IEEE, IEEE, Piscataway, New Jersey, United States, 209–214.

[66] Guido Governatori, Lam Ho-Pun, Antonino Rotolo, Serena Villata, and Fabien Gandon. 2013. Heuristics for Licenses Composition. In *Frontiers in Artificial Intelligence and Applications*, Vol. 259. IOS press, Amsterdam, The Netherlands, 77–86.

[67] Guido Governatori, Ho-Pun Lam, Antonino Rotolo, Serena Villata, Ghislain Auguste Atemezing, and Fabien Gandon. 2014. LIVE: A Tool for Checking Licenses Compatibility between Vocabularies and Data. In *Proceedings of the 2014 International Conference on Posters & Demonstrations Track - Volume 1272* (Riva del Garda, Italy) *(ISWC-PD'14)*. CEUR-WS.org, Aachen, DEU, 77–80.

[68] Guido Governatori, Antonino Rotolo, Serena Villata, and Fabien Gandon. 2013. One License to Compose Them All. In *The Semantic Web – ISWC 2013*, Harith Alani, Lalana Kagal, Achille Fokoue, Paul Groth, Chris Biemann, Josiane Xavier Parreira, Lora Aroyo, Natasha Noy, Chris Welty, and Krzysztof Janowicz (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 151–166.

[69] W3C OWL Working Group. 2012. *OWL 2 Web Ontology Language Document Overview (Second Edition), W3C Recommendation 11 December 2012*. Technical Report. World Wide Web Consortium (W3C). http://www.w3.org/TR/2012/REC-owl2-overview-20121211/

[70] Marina Gueroussova, Axel Polleres, and Sheila McIlraith. 2013. SPARQL with Qualitative and Quantitative Preferences. In *Proceedings of the 2nd International Conference on Ordering and Reasoning - Volume 1059* (Sydney, Australia) *(OrdRing'13)*. CEUR-WS.org, Aachen, DEU, 2–8.

[71] Armin Haller, Krzysztof Janowicz, Simon Cox, Danh Le Phuoc, Jamie Taylor, and Maxime Lefrançois. 2017. *Semantic Sensor Network Ontology, W3C Recommendation 19 October 2017*. W3C Recommendation. World Wide Web Consortium (W3C). https://www.w3.org/TR/2017/REC-vocab-ssn-20171019/

[72] Steve Harris and Andy Seaborne. 2013. *SPARQL 1.1 Query Language, W3C Recommendation 21 March 2013*. W3C Recommendation. World Wide Web Consortium (W3C). http://www.w3.org/TR/2013/REC-sparql11-query-20130321/

[73] Chris Haynes, Michael Luck, Peter McBurney, Samhar Mahmoud, Tomas Vitek, and Simon Miles. 2017. Engineering the emergence of norms: A review. *Knowledge Eng. Review* 32 (2017), e18.

[74] High-Level Expert Group on Artificial Intelligence (AI HLEG). 2019. Ethics Guidelines for Trustworthy AI. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

[75] Koen V Hindriks and Jürgen Dix. 2014. GOAL: A Multi-Agent Programming Language Applied to an Exploration Game. In *Agent-Oriented Software Engineering*, O Shehory and A Sturm (Eds.). Springer International Publishing, Cham, 235–258. https://doi.org/10.1007/978-3-642-54432-3_12

[76] Michael N. Huhns, Nigel Jacobs, Tomasz Ksiezyk, Wei-Min Shen, Munindar P. Singh, and Philip E. Cannata. 1993. Integrating Enterprise Information Models in Carnot. In *Proceedings of the International Conference on Intelligent and Cooperative Information Systems (ICICIS)*. IEEE, Rotterdam, 32–42. https://doi.org/10.1109/ICICIS.1993.291772

[77] Michael N. Huhns and Munindar P. Singh. 1997. Ontologies for Agents. *IEEE Internet Computing (IC)* 1, 6 (Nov. 1997), 81–83. https://doi.org/10.1109/4236.643942 Instance of the column *Agents on the Web*.

[78] Renato Ianella and Serena Villata. 2018. *ODRL Information Model 2.2, W3C Recommendation 15 February 2018*. W3C Recommendation. World Wide Web Consortium (W3C). http://www.w3.org/TR/2018/REC-odrl-model-20180215/

[79] Ian Jacobs and Norman Walsh. 2004. *Architecture of the World Wide Web, Volume One, W3C Recommendation 15 December 2004*. W3C Recommendation. World Wide Web Consortium (W3C). http://www.w3.org/TR/2004/REC-webarch-20041215/

[80] Sebastian Kaebisch, Takuki Kamiya, Michael McCool, Victor Charpenay, and Matthias Kovatsch. 2020. *Web of Things (WoT) Thing Description, W3C Recommendation 9 April 2020*. W3C Recommendation. World Wide Web Consortium (W3C). https://www.w3.org/TR/2020/REC-wot-thing-description-20200409/

[81] Özgür Kafali, Nirav Ajmeri, and Munindar P. Singh. 2020. DESEN: Specification of Sociotechnical Systems via Patterns of Regulation and Control. *ACM Transactions on Software Engineering and Methodology (TOSEM)* 29, 1, Article 7 (Feb. 2020), 50 pages. https://doi.org/10.1145/3365664

[82] Lalana Kagal, Tim Finin, and Anupam Joshi. 2003. A Policy Language for a Pervasive Computing Environment. In *Proceedings of the 4th IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY '03)*. IEEE Computer Society, USA, 63.

[83] Timotheus Kampik, Andres Gomez, Andrei Ciortea, and Simon Mayer. 2021. *Autonomous Agents on the Edge of Things*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1767–1769.

[84] Michael Kifer, Georg Lausen, and James Wu. 1995. Logical Foundations of Object-Oriented and Frame-Based Languages. *J. ACM* 42, 4 (July 1995), 741–843. https://doi.org/10.1145/210332.210335

[85] Sabrina Kirrane. 2021. Intelligent Software Web Agents: A Gap Analysis. arXiv:2102.06607 [cs.AI]

[86] Sabrina Kirrane and Stefan Decker. 2018. Intelligent Agents: The Vision Revisited. In *Proceedings of the 2nd Workshop on Decentralizing the Semantic Web co-located with the 17th International Semantic Web Conference*. RWTH Aachen University (CEUR-WS Proceedings), Aachen, Germany, 9–17.

[87] Sabrina Kirrane, Alessandra Mileo, and Stefan Decker. 2017. Access control and the resource description framework: A survey. *Semantic Web* 8, 2 (2017), 311–352.

[88] Zoltan Kis, Daniel Peinter, Cristiano Aguzzi, Johannes Hund, and Kazuaki Nimura. 2020. *Web of Things (WoT) Scripting API, W3C Working Group Note 24 November 2020*. W3C Working Group Note. World Wide Web Consortium (W3C). https://www.w3.org/TR/2020/NOTE-wot-scripting-api-20201124/

[89] Pranam Kolari, Li Ding, G Shashidhara, Anupam Joshi, Tim Finin, and Lalana Kagal. 2005. Enhancing web privacy protection through declarative policies. In *Sixth IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY'05)*. IEEE, IEEE, Piscataway, New Jersey, United States, 57–66.

[90] Matthias Kovatsch, Ryuichi Matsukura, Michael Lagally, Toru Kawaguchi, Kunihiko Toumura, and Kazuo Kajimoto. 2020. *Web of Things (WoT) Architecture, W3C Recommendation 9 April 2020*. W3C Recommendation. World Wide Web Consortium (W3C). https://www.w3.org/TR/2020/REC-wot-architecture-20200409/

[91] Kalliopi Kravari and Nick Bassiliades. 2015. A survey of agent platforms. *Journal of Artificial Societies and Social Simulation* 18, 1 (2015), 11.

[92] Timothy Lebo, Satya Sanket Sahoo, and Deborah L. McGuinness. 2013. *PROV-O: The PROV Ontology, W3C Recommendation 30 April 2013*. W3C Recommendation. World Wide Web Consortium (W3C). http://www.w3.org/TR/2013/REC-prov-o-20130430/

[93] Jeehang Lee, Julian Padget, Brian Logan, Daniela Dybalova, and Natasha Alechina. 2014. N-Jason: Run-Time Norm Compliance in AgentSpeak(L). In *Engineering Multi-Agent Systems - Second International Workshop, EMAS 2014, Paris, France, May 5-6, 2014, Revised Selected Papers (Lecture Notes in Computer Science, Vol. 8758)*, Fabiano Dalpiaz, Jürgen Dix, and M. Birna van Riemsdijk (Eds.). Springer, Cham, 367–387. https://doi.org/10.1007/978-3-319-14484-9_19

[94] David Lewis. 1969. *Convention — A Philosophical Study*. Harvard University Press, Cambridge, MA.

[95] Ben Light, Christopher P. Holland, and Karl Wills. 2001. ERP and best of breed: a comparative analysis. *Business Process Management Journal* 7, 3 (2021/05/25 2001), 216–224. https://doi.org/10.1108/14637150110392683

[96] Brian Logan. 2018. An agent programming manifesto. *International Journal of Agent-Oriented Software Engineering* 6, 2 (2018), 187–210.

[97] Ashok U. Mallya and Munindar P. Singh. 2007. An Algebra for Commitment Protocols. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)* 14, 2 (April 2007), 143–163. https://doi.org/10.1007/s10458-006-7232-1

[98] David Martin, Massimo Paolucci, Sheila McIlraith, Mark Burstein, Drew McDermott, Deborah McGuinness, Bijan Parsia, Terry Payne, Marta Sabou, Monika Solanki, Naveen Srinivasan, and Katia Sycara. 2005. Bringing Semantics to Web Services: The OWL-S Approach. In *Semantic Web Services and Web Process Composition*, Jorge Cardoso and Amit Sheth (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 26–42.

[99] Viviana Mascardi, Danny Weyns, Alessandro Ricci, Clara Benac Earle, Arthur Casals, Moharram Challenger, Amit Chopra, Andrei Ciortea, Louise A. Dennis, Álvaro Fernández Díaz, Amal El Fallah-Seghrouchni, Angelo Ferrando, Lars-Åke Fredlund, Eleonora Giunchiglia, Zahia Guessoum, Akin Günay, Koen Hindriks, Carlos A. Iglesias, Brian Logan, Timotheus Kampik, Geylani Kardas, Vincent J. Koeman, John Bruntse Larsen, Simon Mayer, Tasio Méndez, Juan Carlos Nieves, Valeria Seidita, Baris Tekin Teze, László Z. Varga, and Michael Winikoff. 2019. Engineering Multi-Agent Systems: State of Affairs and the Road Ahead. *SIGSOFT Softw. Eng. Notes* 44, 1 (March 2019), 18–28. https://doi.org/10.1145/3310013.3322175

[100] Javier Morales, Maite López-Sánchez, Juan Antonio Rodríguez-Aguilar, Wamberto Weber Vasconcelos, and Michael J. Wooldridge. 2015. Online Automated Synthesis of Compact Normative Systems. *ACM Trans. Auton. Adapt. Syst.* 10, 1 (2015), 2:1–2:33. https://doi.org/10.1145/2720024

[101] Andreasa Morris-Martin, Marina De Vos, and Julian Padget. 2019. Norm emergence in multiagent systems: A viewpoint paper. *Autonomous Agents and Multi-Agent Systems* 33, 6 (2019), 706–749. https://doi.org/10.1007/s10458-019-09422-0

[102] Andreasa Morris-Martin, Marina De Vos, and Julian Padget. 2021. A Norm Emergence Framework for Normative MAS – Position Paper. In *Coordination, Organizations, Institutions, Norms, and Ethics for Governance of Multi-Agent Systems XIII*, Andrea Aler Tubella, Stephen Cranefield, Christopher Frantz, Felipe Meneguzzi, and Wamberto Vasconcelos (Eds.). Springer International Publishing, Cham, 156–174.

[103] Pablo Noriega. 1997. *Agent Mediated Auctions: The Fishmarket Metaphor*. Ph.D. Dissertation. Universitat Autònoma de Barcelona. https://www.iiia.csic.es/research/thesis-details?pastphd_id=85

[104] Nardine Osman, Ronald Chenu-Abente, Qiang Shen, Carles Sierra, and Fausto Giunchiglia. 2021. Empowering Users in Online Open Communities. *SN Comput. Sci.* 2, 4 (2021), 338. https://doi.org/10.1007/s42979-021-00714-5

[105] Sascha Ossowski. 2012. *Agreement Technologies.* Springer Publishing Company, Incorporated, Berlin.

[106] Elinor Ostrom. 1990. *Governing the Commons. The Evolutions of Institutions for Collective Action.* Cambridge University Press, Cambridge.

[107] Julian Padget, Marina De Vos, and Charlie Ann Page. 2018. Deontic Sensors. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence* (Stockholm, Sweden) *(IJCAI'18).* AAAI Press, Palo Alto, 475–481.

[108] Monica Palmirani, Guido Governatori, Antonino Rotolo, Said Tabet, Harold Boley, and Adrian Paschke. 2011. Legal-RuleML: XML-Based Rules and Norms. In *Proceedings of the 5th International Conference on Rule-Based Modeling and Computing on the Semantic Web* (Ft. Lauderdale, FL, USA) *(RuleML'11).* Springer-Verlag, Berlin, Heidelberg, 298–312.

[109] Massimo Paolucci, Takahiro Kawamura, Terry R. Payne, and Katia Sycara. 2002. Semantic Matching of Web Services Capabilities. In *The Semantic Web — ISWC 2002,* Ian Horrocks and James Hendler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 333–347.

[110] Simon Parsons. 2001. *Qualitative methods for reasoning under uncertainty.* MIT Press, Cambridge, Massachusetts, United States.

[111] Peter F. Patel-Schneider, Axel Polleres, and David Martin. 2018. Comparative Preferences in SPARQL. In *Knowledge Engineering and Knowledge Management - 21st International Conference, EKAW 2018, Nancy, France, November 12-16, 2018, Proceedings (Lecture Notes in Computer Science, Vol. 11313),* Catherine Faron-Zucker, Chiara Ghidini, Amedeo Napoli, and Yannick Toussaint (Eds.). Springer, Cham, 289–305. https://doi.org/10.1007/978-3-030-03667-6_19

[112] Olivier Pivert, Olfa Slama, and Virginie Thion. 2016. SPARQL Extensions with Preferences: A Survey. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing* (Pisa, Italy) *(SAC '16).* Association for Computing Machinery, New York, NY, USA, 1015–1020. https://doi.org/10.1145/2851613.2851690

[113] María Poveda-Villalón, Asunción Gómez-Pérez, and Mari Carmen Suárez-Figueroa. 2014. OOPS! (OntOlogy Pitfall Scanner!): An On-line Tool for Ontology Evaluation. *International Journal on Semantic Web and Information Systems (IJSWIS)* 10, 2 (2014), 7–34.

[114] Iyad Rahwan and Chris Reed. 2009. *The Argument Interchange Format.* Springer US, Boston, MA, 383–402. https://doi.org/10.1007/978-0-387-98197-0_19

[115] Anand S. Rao and Michael P. Georgeff. 1995. BDI Agents: From Theory to Practice. In *Proceedings of the First International Conference on Multiagent Systems, June 12-14, 1995, San Francisco, California, USA,* Victor R. Lesser and Les Gasser (Eds.). The MIT Press, Cambridge, 312–319.

[116] Wolfgang Reisig. 1985. *Petri Nets: An Introduction.* Springer-Verlag, Berlin, Heidelberg.

[117] Elena Reshetova and Michael McCool. 2019. *Web of Things (WoT) Security and Privacy Guidelines, W3C Working Group Note 6 November 2019.* W3C Working Group Note. World Wide Web Consortium (W3C). https://www.w3.org/TR/2019/NOTE-wot-security-20191106/

[118] Ahlem Rhayem, Mohamed Ben Ahmed Mhiri, and Faïez Gargouri. 2020. Semantic Web Technologies for the Internet of Things: Systematic Literature Review. *Internet Things* 11 (2020), 100206. https://doi.org/10.1016/j.iot.2020.100206

[119] Dumitru Roman, Uwe Keller, Holger Lausen, Jos de Bruijn, Rubén Lara, Michael Stollberg, Axel Polleres, Cristina Feier, Christoph Bussler, and Dieter Fensel. 2005. Web Service Modeling Ontology. *Appl. Ontology* 1, 1 (2005), 77–106. http://content.iospress.com/articles/applied-ontology/ao000008

[120] Andrei Sambra and Stéphane Corlosquet. 2015. *WebID 1.0 - Web Identity and Discovery, W3C Editor's Draft 28 May 2014.* W3C Editor's draft. World Wide Web Consortium (W3C). https://dvcs.w3.org/hg/WebID/raw-file/tip/spec/identity-respec.html

[121] Murat Sensoy, Timothy J. Norman, Wamberto W. Vasconcelos, and Katia Sycara. 2012. OWL-POLAR: A framework for semantic policy representation and reasoning. *Journal of Web Semantics* 12-13 (2012), 148–160. https://doi.org/10.1016/j.websem.2011.11.005 Reasoning with context in the Semantic Web.

[122] Zohreh Shams, Marina De Vos, Nir Oren, and Julian A. Padget. 2020. Argumentation-Based Reasoning about Plans, Maintenance Goals, and Norms. *ACM Trans. Auton. Adapt. Syst.* 14, 3 (2020), 9:1–9:39. https://doi.org/10.1145/3364220

[123] Munindar P. Singh. 2011. Information-Driven Interaction-Oriented Programming: BSPL, the Blindingly Simple Protocol Language. In *Proceedings of the 10th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS).* IFAAMAS, Taipei, 491–498. https://doi.org/10.5555/2031678.2031687

[124] Munindar P. Singh. 2013. Norms as a Basis for Governing Sociotechnical Systems. *ACM Transactions on Intelligent Systems and Technology (TIST)* 5, 1, Article 21 (Dec. 2013), 23 pages. https://doi.org/10.1145/2542182.2542203

[125] Munindar P. Singh and Amit K. Chopra. 2017. The Internet of Things and Multiagent Systems: Decentralized Intelligence in Distributed Computing. In *37th IEEE International Conference on Distributed Computing Systems, ICDCS 2017, Atlanta, GA, USA, June 5-8, 2017,* Kisung Lee and Ling Liu (Eds.). IEEE Computer Society, Washington, D.C., United States, 1738–1747. https://doi.org/10.1109/ICDCS.2017.304

[126] Munindar P. Singh and Amit K. Chopra. 2020. Clouseau: Generating Communication Protocols from Commitments. In *Proceedings of the 34th Conference on Artificial Intelligence (AAAI)*. AAAI Press, New York, 7244–7252. https://doi.org/10.1609/aaai.v34i05.6215

[127] Munindar P. Singh, Amit K. Chopra, and Nirmit Desai. 2009. Commitment-Based Service-Oriented Architecture. *Computer* 42, 11 (2009), 72–79. https://doi.org/10.1109/MC.2009.347

[128] Munindar P. Singh and Samuel H. Christie V. 2021. Tango: Declarative Semantics for Multiagent Communication Protocols. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*. IJCAI, Online, 391–397. https://doi.org/10.24963/ijcai.2021/55

[129] Munindar P. Singh and Michael N. Huhns. 2005. *Service-Oriented Computing: Semantics, Processes, Agents*. John Wiley & Sons, Chichester, United Kingdom. https://doi.org/10.1002/0470091509

[130] Steve Speicher, John Arwe, and Ashok Malhotra. 2015. *Linked Data Platform 1.0, W3C Recommendation 26 February 2015*. W3C Recommendation. World Wide Web Consortium (W3C). http://www.w3.org/TR/2015/REC-ldp-20150226/

[131] Steffen Staab and Heiner Stuckenschmidt (Eds.). 2006. *Semantic Web and Peer-to-Peer - Decentralized Management and Exchange of Knowledge and Information*. Springer, Berlin. https://doi.org/10.1007/3-540-28347-1

[132] Simon Steyskal and Sabrina Kirrane. 2015. If you can't enforce it, contract it: Enforceability in Policy-Driven (Linked) Data Markets. In *Semantics (posters & demos)*. CEUR, Aachen, 63–66.

[133] Simon Steyskal and Axel Polleres. 2014. Defining Expressive Access Policies for Linked Data Using the ODRL Ontology 2.0. In *Proceedings of the 10th International Conference on Semantic Systems* (Leipzig, Germany) *(SEM '14)*. Association for Computing Machinery, New York, NY, USA, 20–23. https://doi.org/10.1145/2660517.2660530

[134] Christian Straßer and Ofer Arieli. 2014. Sequent-Based Argumentation for Normative Reasoning. In *Deontic Logic and Normative Systems*, Fabrizio Cariani, Davide Grossi, Joke Meheus, and Xavier Parent (Eds.). Springer International Publishing, Cham, 224–240.

[135] Sudeep Tanwar, Karan Parekh, and Richard Evans. 2020. Blockchain-based electronic healthcare record system for healthcare 4.0 applications. *Journal of Information Security and Applications* 50 (2020), 102407. https://doi.org/10.1016/j.jisa.2019.102407

[136] Pankaj R. Telang, Munindar P. Singh, and Neil Yorke-Smith. 2019. A Coupled Operational Semantics for Goals and Commitments. *Journal of Artificial Intelligence Research (JAIR)* 65 (May 2019), 31–85. https://doi.org/10.1613/jair.1.11494

[137] Andreas Theodorou, Robert H. Wortham, and Joanna J. Bryson. 2017. Designing and implementing transparency for real time inspection of autonomous robots. *Connect. Sci.* 29, 3 (2017), 230–241. https://doi.org/10.1080/09540091.2017.1310182

[138] Andrei Tour, Artem Polyvyanyy, and Anna Kalenkova. 2021. Agent System Mining: Vision, Benefits, and Challenges. *IEEE Access* 9 (2021), 99480–99494.

[139] Yathiraj B. Udupi and Munindar P. Singh. 2006. Contract Enactment in Virtual Organizations: A Commitment-Based Approach. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI)*. AAAI Press, Boston, 722–727.

[140] Wil Van Der Aalst. 2012. Process mining. *Commun. ACM* 55, 8 (2012), 76–83.

[141] M. Birna van Riemsdijk, Louise A. Dennis, Michael Fisher, and Koen V. Hindriks. 2013. Agent Reasoning for Norm Compliance: A Semantic Approach. In *Proceedings of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems* (St. Paul, MN, USA) *(AAMAS '13)*. IFAAMAS, Richland, SC, 499–506.

[142] Ruben Verborgh. 2018. Decentralizing the Semantic Web through incentivized collaboration. In *Proceedings of the 17th International Semantic Web Conference: Blue Sky Track (CEUR Workshop Proceedings, Vol. 2189)*, Marieke van Erp, Medha Atre, Vanessa Lopez, Kavitha Srinivas, and Carolina Fortuna (Eds.). CEUR-WS.org, Aachen, DEU, 1–5.

[143] Serena Villata and Fabien Gandon. 2012. Licenses Compatibility and Composition in the Web of Data. In *Proceedings of the Third International Conference on Consuming Linked Data - Volume 905* (Boston, MA) *(COLD'12)*. CEUR-WS.org, Aachen, DEU, 124–135.

[144] D. Walton and E.C.W. Krabbe. 1995. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, New York, NY, USA.

[145] Gerhard Weiss. 1999. *Multiagent systems: a modern approach to distributed artificial intelligence*. MIT press, Cambridge, Massachusetts, United States.

[146] Danny Weyns, Andrea Omicini, and James Odell. 2007. Environment as a First Class Abstraction in Multiagent Systems. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)* 14, 1 (Feb. 2007), 5–30.

[147] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data* 3, 1 (2016), 1–9.

[148] Michael Wooldridge and Nicholas R. Jennings. 1995. Intelligent Agents: Theory and Practice. *Knowledge engineering review* 10, 2 (1995), 115–152.

[149] Michael J. Wooldridge. 2001. *Introduction to Multiagent Systems*. John Wiley & Sons, Inc., USA.