A Parallel World Framework for Scenario Analysis in Knowledge Graphs

Andreas Eibeck¹, Arkadiusz Chadzynski¹, Mei Qi Lim¹, Kevin Aditya³, Laura Ong¹, Aravind Devanand⁴, Gourab Karmakar¹, Sebastian Mosbach², Raymond Lau³, Iftekhar A. Karimi⁴, Eddy Y.S. Foo^{1,5}, Markus Kraft^{1,2,3}

released: 14 February 2020

 ¹ Cambridge Centre for Advanced Research and Education in Singapore, 1 Create Way, CREATE Tower, #05-05, Singapore, 138602. 	 ² Department of Chemical Engineering and Biotechnology, University of Cambridge, New Museums Site, Pembroke Street, Cambridge, CB2 3RA United Kingdom.
 ³ School of Chemical and Biomedical Engineering, Nanyang Technological University, 62 Nanyang Drive, Singapore, 637459. 	 ⁴ Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore, 117585.
⁵ School of Electrical	and

School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, Singapore, 639798.

Preprint No. 249



Keywords: Parallel World, Scenario Analysis, Ontology, Knowledge Graph, Superstructure Versioning, Interoperability, Agent, Modelling, Simulation, Optimisation, Cross-domain, Multi-level

Edited by

Computational Modelling Group Department of Chemical Engineering and Biotechnology University of Cambridge West Cambridge Site Philippa Fawcett Drive Cambridge CB3 0AS United Kingdom



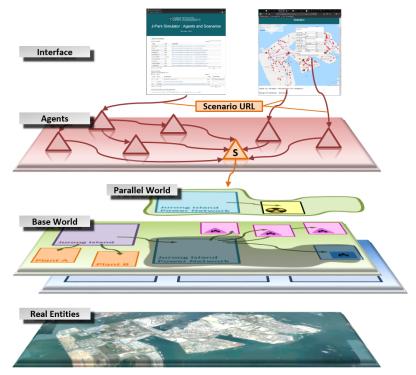
 Fax:
 + 44 (0)1223 334796

 E-Mail:
 c4e@cam.ac.uk

 World Wide Web:
 https://como.ceb.cam.ac.uk/

Abstract

This paper presents Parallel World Framework as a solution for simulations of complex systems within a time varying knowledge graph and its application to the electric grid of Jurong Island in Singapore. The underlying modelling system is based on the Semantic Web Stack. Its linked data layer is described by means of ontologies, which span multiple domains. The framework is designed to allow for conducting multiscale optimisations on complex superstructures within the system. Parallel World Containers, introduced by the framework, ensure data separation and versioning of structures crossing various domain boundaries. Separation of operations, belonging to a particular version of the world, is taken care of by a Scenario Agent. It encapsulates functionality of operations on data, and acts as a parallel world proxy to all of the other agents operating on the knowledge graph. Electric network optimisation for carbon tax is demonstrated as a use case. The framework allows to model and evaluate electrical networks corresponding to set carbon tax values by retrofitting different types of power generators and optimising the grid accordingly. The use case shows possibility of using this solution as a tool for CO₂ reduction modelling and planning at scale due to its distributed architecture.



Highlights

- Architecture of the J-Park Simulator and its Knowledge Graph
- Parallel World Framework based on Semantic Technologies
- Scenario-based Power Grid modelling and optimisation for carbon tax within the J-Park Simulator using the Parallel World Framework

1 Introduction

The fact that simulation plays an important role in the process industry acted as a motivator for design and development of the Parallel World Framework presented in this paper. The time-scale of simulation can vary: from near real-time optimisation (for automated decision making) to long-term policy tests (strategic, support of human decision making, *etc.*). Scenario analysis is a technique used to analyse potential events by examining possible outcomes under given starting conditions. The outcomes are sometimes called 'alternative worlds'. Scenarios can be characterised as processes of varying some initial settings (configuration, conditions, parameters and representations forming parts of the knowledge graph) and then assessing results of simulation and optimisation. The framework was developed within the J-Park Simulator (JPS), an ontology-based system for cross-domain scenarios for the process industry [16], which is part of theworldavatar.com effort which is a general dynamic knowledge graph not restricted to process industry. The main ingredients are: a knowledge graph built on semantic technologies and intelligent agents performing operations on it.

The main challenges addressed by the framework are:

- A need to allow for parallel existence of relevant entities within the knowledge graph forming a semantic representation of a 'parallel world'. At the same time not allowing for interference with other parallel worlds and agent's activities.
- Management of initial settings, intermediate changes and results as well as storing, visibility, finding, reusing (*e.g.* in more complex scenarios or archiving for digital twins life cycle).

There are several aspects, which are strongly related to information persisting, annotating, and versioning. First of all comes persistence. This layer is concerned with reading, querying and writing new or updating existing resources. They come in different shapes of formats as well as storage implementations. Second is an annotation layer. Annotations provide additional information about resources. An example is provenance data, which consists of information about when was a resource created, which agent created or updated a resource *etc.* It can also be used to find resources or to distinguish between different versions of them. The third aspect, which was taken under consideration in this paper is versioning. By setting up an initial configuration or as result of simulation and optimization, some parts of the knowledge graph change. To prevent overwriting the original state by a new state, there is a need to keep track of different versions for these parts.

One can also find a multitude of versioning strategies proposed in the literature. Many of them are built on top of existing solutions for persistence and annotation. None was identified as a solution which would be a perfect fit for JPS. Due to its specific characteristics as a cross-domain integration and modelling tool, it required a new approach. Some of its aspects could be regarded as 'higher' level extensions designed to solve the problem of parallel worlds. There are a few criteria used to differentiate between existing versioning strategies, which will be referred to in the later sections of the paper:

- Granularity level: triple, RDF document (set of triples)
- Reification strategy: e.g. standard reification, named graph
- Storage strategy: independent copies, change-based, timestamp-based
- Change tracking: support for branching and merging or only linear change tracking

As pointed out in [4], distributed version control systems, and Git in particular, could be employed while designing change tracking and versioning solutions for linked data. Operations implemented in Git are transformable to corresponding operations on RDF datasets. It is possible to build systems which support collaboration on distributed resources and heterogeneous data repositories [2] on top of it. However, in case of JPS, distributed nature of Git would make it too complex to make sure that changes in the knowledge graph are visible to all participating agents immediately and without additional push and pull operations. Although agents could still make their changes and read them back on a dedicated URL service via REST API, without extending bare Git, the volume of provenance data is likely to grow beyond the size of the described data [3]. Undoubtedly, amongst its advantages are structural merging capabilities, with which archiving and adding new scenarios on top of the other ones could be easily created. Concerning JPS cross-domain linking requirements and operating under the open world assumption, solutions based on Git have some disadvantages. One can only pull what is in the repository, but can't put the entire world into this repository. The importance of such multi-scale modelling is more extensively elaborated on in [17].

Slightly different approach is taken by the Memento Framework. Instead of minting a new URL for every new version, URL is kept stable and new URLs are minted only for old versions [28]. This approach has become rather widespread. The main idea behind it is to balance quality and scalability by creating traces of interactions with web pages. The created trace also indicates the URL pattern to which the trace applies to as well as provenance information, including the resource on which the trace was created [15]. JPS is designed as a system based on the principle of autonomous intelligent agents operating on a knowledge graph. Although it addresses issues of temporal coherence [1], Memento Framework is not capable of tracing interactions of autonomous agents at the moment.

The purpose of this paper is to present how reduction of CO_2 emissions is possible by the means of the Parallel World Framework, implemented within the JPS. It was envisioned in mind with addressing problems related to versioning and archiving of resources, which belong to representations of complex cross-domain superstructures. Capabilities of the framework allow it to simulate different versions of the world's emission sources and pick the one with the most optimal configuration with regards to the CO_2 emissions. Other ideas, such as multi-dimensional and temporal versioning at the RDF level plus filtering conditions in SPARQL queries [26], are designed to address only relatively narrow problem spaces. To the advantages of the solution implemented within JPS, and presented in this paper, also belongs that it works in a distributed environment. Data is stored on different hosts, while incoming and outgoing links are preserved. A disadvantage of the solution is its prerequisite. It relies on named graphs, each containing only small number of triples. Because it is a very coarse versioning level, operations on named graphs could be very time and memory consuming. Moreover, some of the additional calls to resolve URLs are needed. This approach was chosen in order to pave the way for relatively easy parallel worlds annotations. Annotation is a prerequisite to track modifications (*i.e.* versioning with identifiers) as well as to provide metadata such as time-stamp, creators, scenario *etc.* It is possible to annotate RDF resources at different levels. Parallel worlds originate from the base world and could be regarded as different versions of it. Therefore, within JPS, versioning occurs at the level of named graphs.

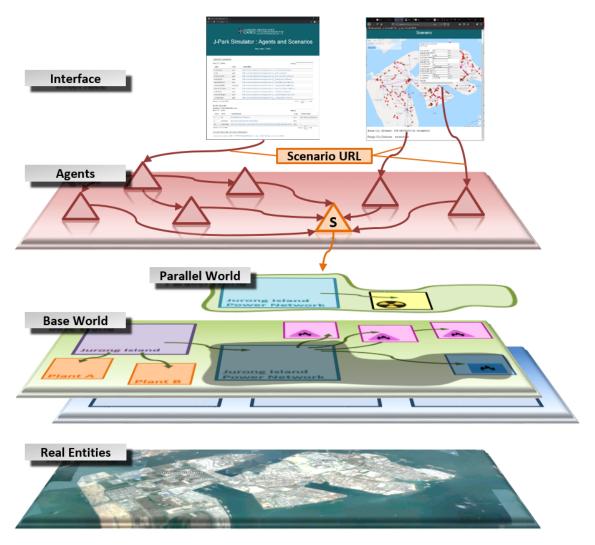


Figure 1: Breakdown of the World representation layers in the JPS system.

JPS and the Parallel World Framework rely heavily on the Semantic Web Stack. Its main architectural building blocks are detailed in [9]. In this paper, section 2 provides an overview of the JPS system as well as its Knowledge Graph Technology. It includes information about relations between high level concepts, such as ontologies and their description language as well as their realisations in the form of named graphs corresponding to real world entities and their representations in the form of Subject-Predicate-Object triples. The use of ontologies and related semantic technologies in JPS for multi-level and cross-domain modelling as well as decentralised management of data and knowledge

is presented in [29, 30]. This paper introduces more specific technological choices made within the system to handle data serialisation, storage and communication protocols. The Parallel World Framework is introduced next. The section which covers it, describes information retrieval and manipulation mechanisms for individual scenarios, identified by unique URLs. It is shown how those operations are realised within the JPS with intelligent autonomous agent driven system architecture on a couple of intuitive examples. The section also describes how those mechanisms allow to organise such information into Parallel World Containers. The fourth section of the paper shows application of the knowledge graph-based Parallel World Framework to the scenario analysis based simulations of the power network, modelled within the JPS system. Results of such an analysis are applied to the power grid optimisation for carbon tax by retrofitting nuclear energy generators into the simulated power grid. It also includes detailed descriptions of interactions between all off the previously introduced components and their concrete implementations.

2 J-Park Simulator and Knowledge Graph

The power network scenario presented later in this paper is based on the OntoPowSys ontology and the model for Jurong Island that was more elaborated on in [8]. The model consists of over 200 bus nodes, 200 electrical lines, five plants with 22 natural gas generators in total, and one plant with six fuel oil generators. JPS uses the Web Ontology Language (OWL) to model these and other real-world entities on Jurong Island, such as chemical plants and Jurong Island itself as an Eco-Industrial Park [21] (EIP). While OWL is a very rich and expressive language, this section summarises only those aspects that are relevant to describing how the parallel world framework takes advantage of semantic technologies. It allows to define classes (types, concepts), individuals (instances of a given class) and properties (relations) and uses URLs as globally unique identifiers for them. It could be illustrated by the means of Subject-Predicate-Object triples. For instance, the triple consisting of the following absolute URLs defines a class for power generators:

- http://www.theworldavatar.com/ontology/ontopowsys/PowSysRealization.owl #PowerGenerator
- http://www.w3.org/1999/02/22-rdf-syntax-ns#type
- http://www.w3.org/2002/07/owl#Class

This triple may as well be expressed in a more compact way with the help of namespaces. However, to keep it more understandable, short and human-readable triple notation is frequently used in this paper. For instance, the triples 'System is a Class' and 'PowerGenerator is a Class' define two classes. Triples 'JurongIslandPowerNetwork is a System' and 'Gen1 is a PowerGenerator' define two corresponding individuals. The triple 'hasSubsystem is an ObjectProperty' defines a property and the triple 'JurongIslandPowerNetwork hasSubsystem Gen1' links the previously defined instances.

Entities described this way form the JPS knowledge graph. Each triple may be represented by two nodes labelled with the URLs of the subject and the object, respectively, and a

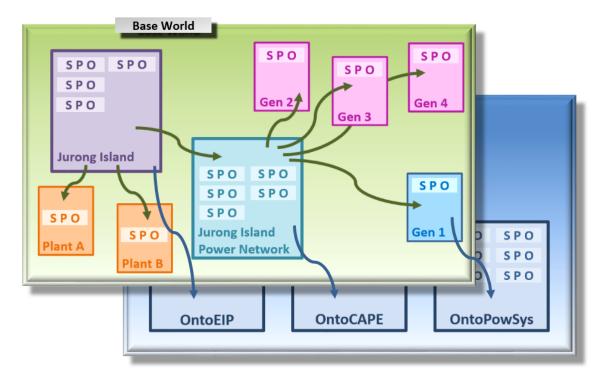


Figure 2: Representation of the Base World fragment by the means of semantic technologies: concrete realisations of domain ontologies as named graphs correspond to real entities and are described by the Subject-Predicate-Object triples.

directed edge in between, labelled with the URL for the predicate. Consequently, a set of triples can be represented by a directed graph which does not necessarily have to be completely connected. A knowledge graph usually denotes a huge graph (*i.e.* a huge set of triples) where the number of instances is much larger than the number of classes [22]. This is the case of the JPS knowledge graph. In principle, the JPS knowledge graph is not restricted to the data generated for JPS but may also integrate triples from other sources. In the same way, other application may make use of the JPS knowledge graph. A plain huge set of triples as a graph without any clear substructures can be hard to navigate through. The question of how to organise these triples and subdivide them into suitable subsets is of high importance for maintainability, performance, distribution, and discovery and also for the Parallel World Framework.

Figure 2 shows a few entities from the power network scenario in a very simplified manner. The box with label 'Gen1' contains only a single triple 'Gen1 is a PowerGenerator'. Meanwhile a power generator in JPS is modelled with the help of over one hundred triples that, among other things, specify appropriate electrical parameters and the bus node it is connected to. Different colours, in Fig. 2, illustrate distinguished classes of individuals: Jurong Island EIP, power plants A & B, power network, two different types of power generators. A triple set may be serialised in RDF/XML syntax¹ and stored as a file. The triple set of the file content may be requested by a specific URL². At the same time, it is important to distinguish the URL that identifies the physical entity as an individual in the

¹https://www.w3.org/TR/rdf-syntax-grammar/

²Requesting the resource http://www.jparksimulator.com/kb/sgp/jurongisland/jurongislandpowernetwork/EGen-

sense of OWL from this specific URL that identifies the triple set describing the model associated with that entity.

Alternatively, the triple set may be stored as a named graph forming a part of a dataset in a triple store. Roughly speaking, a named graph denotes a pair consisting of an URL and a triple set³. Both ways are similar with respect to packaging triple sets, and indeed, the implementation of the parallel world framework is able to switch between them easily. For this reason, we apply the term 'named graph' to both alternatives going forward. However, when querying a huge triple set or a large collection of named graphs, a triple store would provide a much better performance.

JPS separates the class level and the instance level in accordance with the best practices. Classes and properties are defined in modular, reusable domain ontologies. Figure 2 shows three domain ontologies as boxes in the lower blue layer. Together with other domain ontologies, they have been used in previous scenarios in JPS and are also relevant to the power network scenario:

- OntoCAPE [19]: a domain ontology for chemical process engineering based on more general upper ontology for engineering.
- OntoEIP [30]: ontological description for Eco-industrial Parks.
- OntoPowerSys [8]: a domain ontology for power systems.

In fact, the blue box representation can be still regarded as a huge oversimplification: *e.g.* OntoCape contains over 50 submodels, OntoPowerSys around 6, each of which can be represented as a named graph on its own. Therefore, each of the blue boxes could be seen as representing rather a dataset consisting of named graphs. An arrow pointing from one box to another box in Fig. 2 reflects the fact that there is at least one triple in the first box that refers to an individual or class (either as a subject or as an object of that triple) or to a property defined in the second box.

JPS is built upon agent driven architecture. Intelligent and autonomous agents are the components which operate on its knowledge graph. HTTP⁴ is used as a communication protocol between agents, which use GET⁵ method, to exchange data between themselves, as well as POST⁶ and PUT⁷ methods, to update each other about any changes in their states. All data exchanged between agents this way travels serialised in JSON⁸ format as an URL query component⁹ or request message body¹⁰.

^{001.}owl will return the triple set for "Gen1" in RDF/XML syntax. However, the URL of individual Gen1 itself is given by adding #EGen-001 to the previous URL. These URLs with fragments are used frequently.

³https://www.w3.org/TR/rdf11-concepts

⁴https://tools.ietf.org/html/rfc2616

⁵https://tools.ietf.org/html/rfc2616#section-9.3

⁶https://tools.ietf.org/html/rfc2616#section-9.5

⁷https://tools.ietf.org/html/rfc2616#section-9.6 ⁸https://tools.ietf.org/html/rfc8259

^ohups://tools.ietl.org/html/fic82

⁹https://tools.ietf.org/html/rfc3986#section-3.4

¹⁰https://tools.ietf.org/html/rfc2616#section-4.3

3 The Parallel World Framework

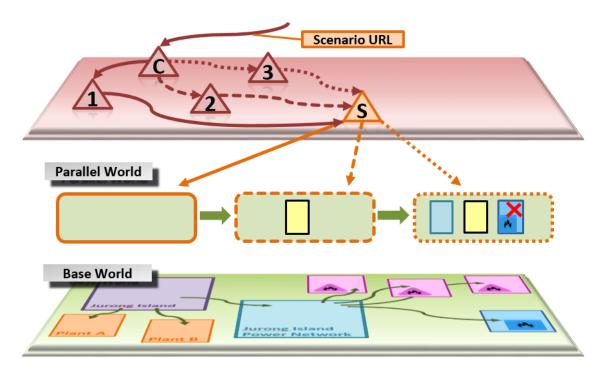


Figure 3: Activities of the JPS agents construct and manipulate entities placed in Parallel Worlds' Containers and accessible by individual URLs to the scenario agent S.

Previous section was concerned with the knowledge graph layer of the power network scenario. This section gives an overview of the agent layer as well as introduces the main ideas and choices behind the parallel world framework.

As described in [9], agents in JPS communicate with each other by the means of the HTTP protocol. They exchange information by updating and querying the knowledge graph with the help of SPARQL¹¹, a semantic query language which is part of the Semantic Web Stack. Those interactions can be realised in one of the following two ways. Agents may request a named graph by its URL directly in order to retrieve information via a SPARQL query or perform an update. They may also send the retrieval or update query to a SPARQL endpoint, a service that performs the query or update on a data set in a triple store. In addition to that, to exchange information directly, agents also provide input and output parameters in the HTTP requests and responses. Input and output parameters are serialized in a form of JSON documents that provide certain URLs which serve as starting points for navigating, querying and updating parts of the knowledge graph.

By default, the agents' queries and updates are performed on the knowledge graph directly, *i.e.* on the main part of the knowledge graph that is labelled in Fig. 3 as the base world. Parallel World Framework changes this behaviour by identifying each parallel world by its own unique scenario URL. Agents that participate in the same scenario propagate the corresponding URL as an additional input parameter while communicating with each other. The framework introduces a separate scenario agent that is associated with

¹¹https://www.w3.org/TR/sparql11-query/

the particular parallel world and works as a mediator, which performs retrieval and update queries on the knowledge graph on behalf of the other participating agents.

Although this approach is not without its own specific drawbacks [26], for the sake of simplicity, current implementation of the parallel world framework in JPS works at the granularity level of named graphs. As noted in [12] this approach is relatively easy to understand as well as allows for reuse of existing data queries. At the same time it also provides a very compact representation for data and queries. In JPS, the scenario agent keeps track of named graphs that have been changed by any agents participating in the same scenario. Any other agents can request the scenario agent's functionality by reading the scenario URL from the input parameters and using it for the HTTP request. If no scenario URL is given as an input parameter, agents fall back to the default mode and operate directly on the base world representation instead.

Because of the above, operations supported by JPS scenario agents can be seen as ones modelled on those described in the W3C recommendations for SPARQL 1.1 Protocol¹² and Graph Store HTTP Protocol¹³. Therefore, GET and PUT operations for entire named graphs are supported with the key 'resource'. A separate DELETE method is not yet implemented as it is in the case of the mentioned protocols. However, it could be easily mimicked using SPARQL updates as, in essence, every update could be broken down into delete and subsequent insert operation on data. The main idea in the parallel world framework is that DELETE in the base world leads to deleting the appropriate named graph physically. In contrast to that, DELETE in a parallel world leads to deleting of the copy in the parallel world and marking of the named graph as deleted with an annotation triple in the meta dataset. After that a scenario agent should respond to requests for it with the same HTTP status code as it would if there was no such resource. Similarly to the SPARQL 1.1 Protocol, GET and POST can be combined with SPARQL retrieval and update queries. In addition to that, GET and PUT can be also used to read from and write to non RDF resources. This fits them very well into Semantic Document Model [20]. The raw scenario URL may also be used to request a semantic description of the parallel world itself by issuing HTTP GET request with <scenario URL> without keys into the JPS.

Figure 3 illustrates some of the agents involved in a sample scenario as red triangles and their calls to each other as red arrows. Their functionality is described in detail in section 4. This section illustrates only some of their interactions with the knowledge graph via the scenario agent. The agent, represented by the orange triangle in Fig. 3, may be deployed to any host. In JPS, it is currently deployed at www.theworldavatar.com. It can be requested by a scenario URL of the form:

http://www.theworldavatar.com/jps/scenario/<scenario name>

The coordination agent C is called with input parameters. It has to be able to identify the part of the knowledge graph representing the part of the base world, which would be altered in the parallel world. In our use case, those parameters consist of 'Jurong Island Power Network' URL and a new scenario URL of the above form (*e.g.* 'pownet1') as a

¹²https://www.w3.org/TR/sparql11-protocol/

¹³https://www.w3.org/TR/sparql11-http-rdf-update/

parallel world name. The agent C calls agents 1, 2 and 3 one by another, providing them with all of the necessary information.

First of all, agent 1 queries the power network to obtain a list of connected entities, which will be subject to alteration in the parallel world. Without the parallel world framework, the default HTTP request

```
GET /kb/sgp/jurongisland/jurongislandpowernetwork
/JurongIslandPowerNetwork.owl HTTP/1.1
Host: www.jparksimulator.com
```

would return the triple set corresponding to the named graph for 'Jurong Island Power Network' as it is stored in the base world. However, since agent 1 is called with a scenario URL, it switches the roles between the named graph URL and the scenario URL, and sends the HTTP request

```
GET /jps/scenario/pownet1?query=
{"resource":"www.jparksimulator.com/kb/sgp/...
/JurongIslandPowerNetwork.owl"} HTTP/1.1
Host: www.theworldavatar.com
```

instead. The original URL becomes the input parameter for the scenario agent¹⁴. Any agent may interact with JPS and use its parallel world framework just by switching the roles between the named graph URL and the scenario URL as described.

Upon the first HTTP request for a new scenario, the scenario agent creates a new container for storing modified named graphs of the particular parallel world. In Fig. 3, the boxes under the parallel world label sketch different states of this container. The empty container on the left side stands for the new container. Upon any of the other consecutive requests, the scenario agent always looks up whether a copy of the named graph already exists in the container. If the appropriate named graph could not be found there, the agent reads the corresponding graph from the base world instead.

Agent 1 performs further queries, *e.g.* to get information about each connected entity. All queries are redirected to the scenario agent in the same way as they are for 'Jurong Island Power Network' representation. Since agent 1 does not update any of these named graphs, the parallel world container is kept empty during such interactions with the knowledge graph.

Next, agent 2 performs several queries and creates – as a result of its optimization – new individuals and their models as triple sets. Because of that it is called with the same scenario URL, it delegates the requests for storing the named graphs to the same scenario agent. The only difference is that it sends a HTTP PUT request for each named graph instead of a HTTP GET request. The request contains the URL of the named graph as an input parameter and the triple set as a message body. Figure 3 shows the state of the

¹⁴Following the software design principle 'separation of concerns', the JPS implementation separates the handling of the scenario URL and the redirection to the scenario agent from the actual functionality of all other agents. This keeps their program code clean.

container (middle box with dashed frame) after storing the named graph for one altered entity.

Finally, agent 3 updates the whole parallel world structure, which is taken under consideration, and replaces some of its original entities by new and altered instances (*i.e.* it deletes the hasSubsystem triples for these entities and adds hasSubsystem triples for the new entities in the named graph of the whole structure). The present implementation of the update operation in JPS does not yet mark named graphs for the replaced entities as deleted in the meta dataset. They are also not deleted 'physically' in the base world. They are only removed in the copy placed in the data container belonging to the particular parallel world. Therefore, comparing parallel world with the base world in order to find out what has changed there, would involve comparison of the whole datasets. This may be time and resource consuming operation, in case of more complex configurations running for certain periods of time. The problem of tracing parallel worlds origins, so to say, will be addressed in the future versions of JPS. This is one of the ways in which the current parallel world framework itself is going to evolve and get enriched in the future. This may be also necessary for regulatory reasons as data traceability issues are looked into by the EU [12].

As it is illustrated in Fig. 1, agents' activities within parallel worlds overlay their activities on the particular parts of the knowledge graph. Modified parts of it get accentuated and gain prominence within contexts of their activities. Meanwhile, their origins within the base world remain unmodified in the background. They may pose as subjects of activities of the same or different agents operating within the base world or independent parallel worlds. At the same time, all corresponding named graphs are independent of each other and evolve in parallel. Knowledge graph parallelisation via copies of named graphs allows JPS agents to operate on a shared world representations without interference. However, the results of their operations could be still eventually tracked back to the base world representation, which always stays untouched in the shadow of their current activities.

Several agents, which collaborate with each other within one parallel world, have to share the same view of the data in the knowledge graph. They shall not interfere with other agents' activities and their operations on the knowledge graph must be separated. This raises a necessity to manage different versions of named graphs (*i.e.* of changed entities). Different versions must be resolved in the same way for all agents participating in the same view. Problems of co-evolution [23] and identification [24] are solved within JPS by associating parallel worlds with unique scenario URLs, which are shared between agents participating in activities related to the particular parallel world. Agents are able to find appropriate parts of the knowledge graph and collaboratively work on them via resolving those unique URLs. Since all of the named graphs for each parallel world are kept in separate containers, all of them can co-evolve independently as well. This idea resembles to large extent the concept of 'context specifiers' from the Multidimensional RDF framework [14].

Entities within the higher level structures and their subsystems are resolvable by their unique URLs. Although they share some of the generic parts with the identifiers of the base world entities, the specific parallel world information is embedded within them as well. Therefore, an agent that aggregates a certain information concerning all entities within such structure is able to resolve every single entity by its unique URL and query

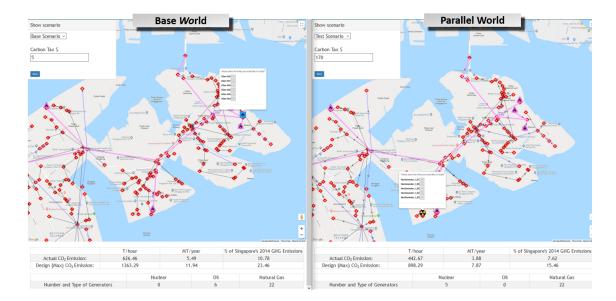
it for the individual property values. Mentioned agent takes for a starting point the string identifying the structure as an input parameter (*i.e.* 'Jurong Island'). Next, it 'navigates along' the knowledge graph via queries to its subsets and their interconnected entities. Bearing in mind that the scenario modifies a particular structure during the parallel world's specific alteration process (*i.e.* optimisation), the aggregation agent needs some functionality that allows it to resolve the 'right' version of the entity. It does so, by traversing the knowledge graph through unique URLs, which point to the resources stored within the particular parallel world container.

As pointed out in [10], minimisation of redundant information and respect of the original modelling as well as provenance information of archives, is one of the main research challenges for structured interlinked data representation systems. Described isolation of JPS agents' operations on the knowledge graph within separate parallel worlds motivated the choice of 'Independent Copies' as a default Archiving Policy. Certainly, from the pure data storage point of view, there are possible more efficient solutions, such as those based on theory of patches [13] and delta calculations [5] or even multi-indexed and compressed delta chains [25]. However, they may not always be a first choice from the version creation and retrieval timing point of view, which is regarded as another key performance aspect of version management systems [27]. This is especially true with regards to complex objects, evolving over time and accumulating many changes building upon one another. Parallel worlds definitely show characteristics of such superstructures. Moreover, storage minimisation oriented solutions, may also need additional annotation of resources [13], which is often mentioned as a more general issue for ontology based knowledge representation systems [18].

Within JPS architecture, agents taking part of a scenario interactions as well as their associated non-RDF resources (such as configuration files), are annotated using RDF and OWL. Therefore a scenario, with all of the corresponding resources, can be viewed as an RDF dataset. This makes scenarios themselves becoming parts of the JPS knowledge graph. The same holds for all of the modified named graphs, which belong to parallel worlds.

All the above does not mean that the problem of storage space was completely disregarded within JPS. Contrary to that, it has been addressed in a form of providing two different copy strategies for parallel world data. Copy-On-Write – a default option – makes sure that only those parts of the knowledge graph, which are modified in a parallel world are copied over to a scenario data container. On the other hand, selecting Copy-On-Read strategy makes agents to copy all data, which they interact with in any way within a parallel world and regardless of that whether they modify it or not. JPS, as a cross domain modelling and simulation system, integrates different types of software which produce heterogeneous data formats. Therefore, its knowledge base manager currently supports Blazegraph, Fuseki and Rdf4j storage back-ends, apart from file based back-end used for RDF/OWL. There are also appropriate abstractions in place, which would allow for potential implementations of cloud based knowledge graph solutions and parallel world data containers in the future.

At the same time, a scenario can be also looked at from beyond a mere persistence layer [6] perspective. It can be viewed as an activity or context, which encapsulates other activities occurring in a particular parallel world. In fact, there is an implementation of a specific JPSContext object, which is used by the ScenarioAgent objects in JPS. The context object holds all of the necessary parallel world information at run-time. This information is passed around between agents, which participate in a particular scenario. This way all agents work within one and the same scenario, where a dedicated scenario agent manages its context and acts on behalf of all the other participating agents.



4 Application to the Power Network Scenario

Figure 4: Result of a scenario based analysis retrofitting altered and optimised power network into a Parallel World.

This section describes an application of the parallel world framework to the power network scenario as well as discusses some of the aspects of the framework itself. More detailed discussion focused on the quantifiable outputs of the power network scenario (carbon tax amount, different configuration *etc.*) will be part of another paper.

Figure 4 presents application of the Parallel World Framework to the simulation of the power network optimisation. The original and unmodified Jurong Island's power grid, which contains six oil generators, is presented on the left side. The right side shows a modified network, where five nuclear generators replaced the oil generators in a parallel world with a simulated carbon tax value of \$170. The total number and type of generators is summarised at the bottom. Due to the level of detail of the map, generators are visualised as grouped in one location. Blue square visualises the six oil generators, whereas the yellow and black circle groups the mentioned five nuclear generators. The bottom section also shows the original and optimised emission details. They are displayed in tonnes per hour and mega-tonnes per year as well as the percentage of the total Singapore's green house gas emissions.

The following capabilities of the parallel world framework, summarised in Fig. 1, are exemplified via the simulation of the alternate power network scenario:

- The lower layer corresponds to a section of the physical real world (in this case Jurong Island)
- The middle layer models the appropriate section semantically (power network, but in addition: carbon tax, *etc.*). It is a part of the knowledge graph persistence layer.
- The upper level contains the agents which operate on the knowledge graph and collaborate with each other. This layer can also be regarded as a 'business logic' layer.

As explained in the previous section, the coordination agent C makes requests to the agents numbered 1 to 3 one after another. The coordination agent can be started by a script providing the required input parameters. It is responsible for the overall service composition and makes sure that all relevant agents are called in a correct order. Intelligent automation of the agent composition process is a work in progress and is considered to be one of the essential elements of the overall JPS system architecture [31, 32].

The coordination agent C calls agent 1 with a carbon tax amount and the URL that identifies any power network as its input arguments. In the presented example the Jurong Island Power Network is taken under consideration. As an instance of the OntoPowSys [8] ontology it forms a part of the JPS knowledge graph. Agent 1 uses this URL to query for generators that are attached to the power network. It combines returned information with data about generating cost, CO_2 emissions and the target carbon tax value in order to make a decision whether replacement of some of the existing generators by Small Modular Reactors (SMR) makes sense from the economic point of view. Agent 1 does not update the knowledge graph in any way. It only reads it and merely returns a serialised, and possibly empty, list of replaceable generators back to the coordination agent, which in turn passes over the list to the agent 2.

Agent 2 optimises the number and capacities of substitutional SMRs of different types and their location on Jurong Island. It takes into account the amount of power delivered to the network as well as costs associated with the whole setup. To do so, it queries the JPS knowledge graph for the list of land lots on Jurong Island and the design capacity of the replaceable generators. Geographical coordinates of load points of the electrical network as well as potential sites are used as input parameters for the optimisation model. Different types of costs and risk as well as the overall project lifespan complete the parameter list. The model aims to balance out corresponding risks, costs and power transition loses for the entire network.¹⁵ Its work results in the creation of new named graphs for the SMRs, that will substitute the replaceable generators in the parallel world. Underlying modelling system makes sure that they are placed in the most optimal places, while taking into account all of the optimisation factors listed above.

Both agents, 1 and 2, make use of the General Algebraic Modeling System (GAMS)¹⁶ to solve the optimisation problems via Mixed Integer Non Linear Programming (MINLP). Due to the complexity of the model required by the agent 2, GAMS computations require around 36 hours to complete. This constraint motivated a development of an Asynchronous Watching Service for JPS. It was designed as an autonomous and reusable part of

¹⁵Please refer to the [7] for more detailed description of the parameters and optimisation goals.

¹⁶https://www.gams.com/

the system, capable of serving agents operating on different use cases and parallel worlds in multiple threads of execution. The service constantly monitors the modelling system's output directories in different parallel worlds and informs instances of the agent 2 whenever the results of GAMS calculations are ready for them to be picked up and processed any further. Mentioned agents' instances themselves are not blocked by waiting for the appropriate models' convergence. The watching service is capable of keeping track of that which agent it has to forward the response back to. Upon their requests, it informs the agents about its own waiting state immediately, so they can carry on with performing any other functions or accept other requests.

Agent 3 uses the list of replaceable generators from the agent 1 and the optimisation result of the agent 2 to retrofit the power network from the base world: it connects the SMRs to the closest proper buses and replaces the hasSubsystem relations for the replaceable generators by new relations between the power network and the SMRs. This is an example of resolving and altering individual instances of entities within a subset of a higher level parallel world structure, described in the previous section. As already mentioned, the retrofitted network is shown as a parallel world on the left side of Fig. 4.

So far, the separation of modelling activities between agents relates to different phases of the overall simulation:

- economic analysis and policy testing
- design phase
- retrofitting and installation

The scenario of the retrofitted power network could be also extended to the operational phase, by adding the Optimal Power Flow (OPF) agent to the base world scenario. The agent was already used in [8]. Its goal is to optimise load supply with the minimum generation cost within the power grid constrains (power line limits, bus voltages, etc.). At this point the network must be already well designed. Otherwise the OPF will not converge. Solving it requires a well designed grid topology, which contains information about the connections between the bus nodes and the attached equipment (power loads, power generators, etc.). The topology needs to be modified every time a change within the grid occurs. Electrical networks are modelled as sets of buses interconnected by branches. The first represent physical points of interconnection among power systems equipment. The second model paths for the flow of electrical current [11]. Inputs to the OPF are matrices for buses, lines and generators. Optimised outputs are line current and power loses, bus voltages and generated powers. Both, inputs and outputs, are read from and written back to the knowledge graph. Some of the inputs are updated along the way of computation and are present in the output value set. More detailed description of the power network ontology and OPF implementation within JPS is presented in [8].

All agents that are called with the same scenario URL have the same view on the knowledge graph, corresponding to the particular parallel world. Figure 1 illustrates this in a way of 'overlaying' the base world by the parallel one and making it go to the background. Figure 5 shows how the optimisation results (both the retrofitted power network and the OPF calculation) are visualised in a browser. The user selects the scenario (step 1) to

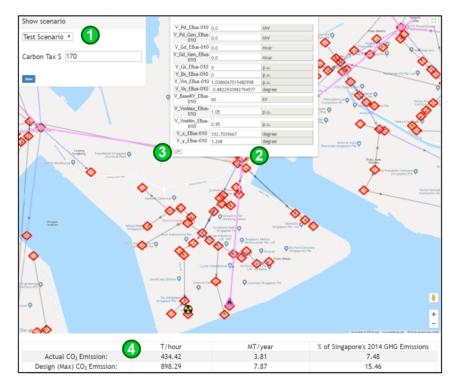


Figure 5: Parallel World user interface of JPS for scenario listing, component alteration and analysis.

visualise the corresponding version of the power network. Upon a click on one of the network entities, a pop-up window presents the current property values of the selected entity (step 2). The user can change the values and call the OPF agent by pressing the OPF button (step 3). After the optimal power flow simulation has finished, another agent aggregates the actual CO_2 emissions of all generators. The aggregated values are displayed at the bottom of Fig. 5 (step 4).

If the user selects the base scenario, the OPF agent is requested without a scenario URL and the browser will visualise the model from the base world, otherwise the model from the corresponding parallel world. Effects of starting simulation this way are isolated solely into visualisation. It operates only on the client side copy of the visualised knowledge graph and is not persistent in any way. This ephemeral simulation could be started from the base world as well as parallel world in order to estimate and visualise any further changes to electrical networks. Adding the OPF calculation on top of the previous optimisation result (retrofitted power network) demonstrates easiness of combining different aspects from different domains within JPS as well as its capabilities of keeping track of complex scenarios by the Parallel World Framework.

5 Conclusions

This paper presents the Parallel World Framework implemented within the JPS. It extends its capabilities by allowing to evaluate multiple versions of modelled complex superstruc-

tures during optimisation process. Electrical network on Jurong Island, represented within the system, is taken as a use case. It is demonstrated that described framework is capable of optimising the network for carbon tax. Therefore, this part of JPS could be used as a CO_2 reduction planning and modelling tool. At the same time, its distributed architecture allows for potential superstructure optimisations at scale.

The framework is designed to keep documents, models and results for the entire life-cycle together as well as to use them for optimisation, prognosis and policy testing without interference with the real world. It has got its parallels to the concept of digital twins. One of the drawbacks of this approach is its demand for data storage. Keeping large volumes of data in always on data warehouses has got its own contribution to the energy consumption and CO_2 emissions. Although it could be looked at as a sacrifice necessary to improve the present state of affairs, there is possible yet another, slightly more sophisticated, approach. Namely, one can think of building small models, which are able to generate data on demand, instead of keeping the original volumes always available, even if rarely used.

The next paper will present deeper analysis of carbon tax price for the power network scenario. It will also explore the use of the framework for hyper parameter optimisation together with annotation for querying time and location of entities subject to atmospheric dispersion modelling in JPS. There is a potential for using the framework in cases where parallel in-memory optimisation for different initial conditions is required. It is possible to use it for spinning off multiple parallel world simulations based on different initial conditions and selecting the world which reaches optimisation goals first as the end result.

Acknowledgements

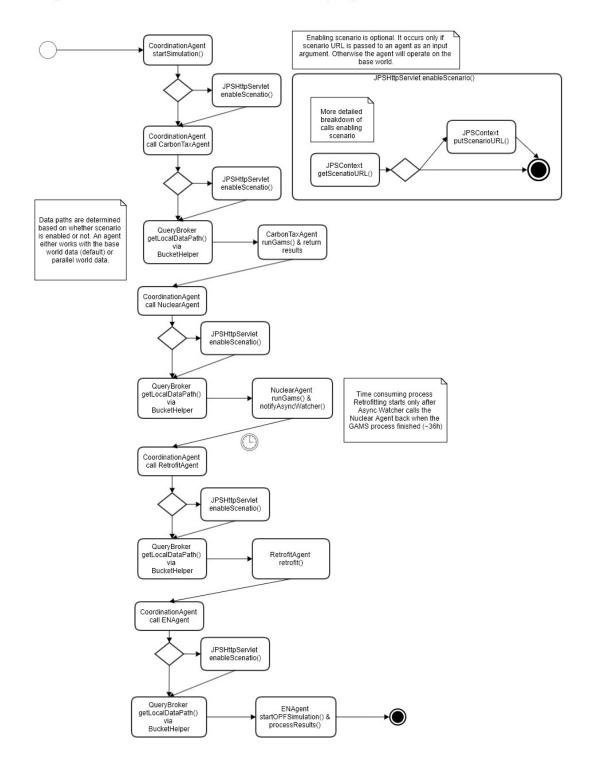
This project is funded by the National Research Foundation (NRF), Prime Ministers Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme. Markus Kraft gratefully acknowledges the support of the Alexander von Humboldt Foundation.

List of abbreviations

EIP	Eco-Industrial Park
HTTP	Hypertext Transfer Protocol
URL	Uniform Resource Locator
JPS	J-Park Simulator
JSON	JavaScript Object Notation
OWL	Web Ontology Language
RDF	Resource Description Framework
SPARQL	SPARQL Protocol and RDF Query Language
W3C	World Wide Web Consortium
XML	Extensible Markup Language
SMR	Small Modular Reactor
OPF	Optimal Power Flow
MINLP	Mixed Integer Non Linear Programming
GAMS	General Algebraic Modeling System

A Parallel world framework – UML diagram

JPS_POWSYS with Parallel World Framework - UML Activity Diagram



B Electrical Network use case – List of Agents

Coordination Agent (Agent C)

- Inputs:
 - 'carbontax' (*e.g.* \$170)
 - 'landlot'
 (*e.g.* 'http://www.jparksimulator.com/kb/sgp/jurongisland/ JurongIslandLandlots.owl')
 - 'electricalnetwork'
 (*e.g.* http://www.jparksimulator.com/kb/sgp/jurongisland/
 jurongislandpowernetwork/JurongIslandPowerNetwork.owl
 #JurongIsland_PowerNetwork)
- Outputs:
 - void

Carbon Tax Agent (Agent 1)

- Inputs:
 - 'carbontax'
- Outputs:
 - 'substitutionalgenerators'

Nuclear Agent (Agent 2)

- Inputs:
 - 'landlot'
 - 'electricalnetwork'
 - 'substitutionalgenerators'
- Outputs:
 - 'plants'

Retrofit Agent (Agent 3)

- Inputs:
 - 'plants'
 - 'electricalnetwork'
 - 'substitutionalgenerators'
- Outputs:
 - void

Electrical Network Agent (OPF Agent)

- Inputs:
 - 'electricalnetwork'
- Outputs:
 - 'folder'

References

- S. G. Ainsworth, M. L. Nelson, and H. Van de Sompel. A framework for evaluation of composite memento temporal coherence, 2014. URL https://arxiv.org/ pdf/1402.0928.
- [2] N. Arndt, N. Radtke, and M. Martin. Distributed collaboration on RDF datasets using git: Towards the quit store. In A. Fensel, A. J. Zaveri, S. Hellmann, and T. Pellegrini, editors, *SEMANTICS 2016: Proceedings of the 12th International Conference* on Semantic Systems, pages 25–32. Association for Computing Machinery, New York, 2016. doi:10.1145/2993318.2993328.
- [3] N. Arndt, P. Naumann, and E. Marx. Exploring the evolution and provenance of git versioned RDF data. In J. Debattista, J. Umbrich, and J. D. Fernández, editors, *Proceedings of the 3rd Workshop on Managing the Evolution and Preservation of the Data Web (MEPDaW)*, volume 1824 of *CEUR Workshop Proceedings*, 2017. URL http://ceur-ws.org/Vol-1824/mepdaw_paper_2.pdf.
- [4] N. Arndt, P. Naumann, N. Radtke, M. Martin, and E. Marx. Decentralized collaborative knowledge management using git. *Journal of Web Semantics*, 54:29–47, 2019. doi:10.1016/j.websem.2018.08.002.
- [5] T. Berners-Lee and D. Connolly. Delta: an ontology for the distribution of differences between RDF graphs, 2004. URL https://www.w3.org/DesignIssues/ Diff.html.
- [6] F. Christ and B. Nagel. A reference architecture for semantic content management systems. In M. Nüttgens, O. Thomas, and B. Weber, editors, *Enterprise Modelling* and Information Systems Architectures: Proceedings of the 4th International Workshop on Enterprise Modelling and Information Systems Architectures, EMISA 2011, Hamburg, Germany, volume P-190 of LNI, pages 135–148. Gesellschaft für Informatik (GI), 2011. URL https://dl.gi.de/handle/20.500.12116/18506.
- [7] A. Devanand, M. Kraft, and I. Karimi. Optimal site selection for modular nuclear power plants. *Computers & Chemical Engineering*, 125:339–350, 2019. doi:10.1016/j.compchemeng.2019.03.024.
- [8] A. Devanand, G. Karmakar, N. Krdzavac, K. Aditya, R. Rigo-Mariani, A. Krishnan, Y. F. Eddy, I. A. Karimi, and M. Kraft. OntoPowSys: A power systems ontology for cross domain interactions in an eco industrial park. Submitted for publication, 2020.
- [9] A. Eibeck, M. Q. Lim, and M. Kraft. J-Park Simulator: An ontology-based platform for cross-domain scenarios in process industry. *Computers & Chemical Engineering*, 131:106586, 2019. doi:10.1016/j.compchemeng.2019.106586.
- [10] J. D. Fernández, A. Polleres, and J. Umbrich. Towards efficient archiving of dynamic linked open data. In J. Debattista, M. d'Aquin, and C. Lange, editors, *Proceedings*

of the First DIACHRON Workshop on Managing the Evolution and Preservation of the Data Web, volume 1377 of CEUR Workshop Proceedings, 2015. URL http://ceur-ws.org/Vol-1377/paper6.pdf.

- [11] S. Frank and S. Rebennack. An introduction to optimal power flow: Theory, formulation, and examples. *IIE Transactions*, 48(12):1172–1197, 2016. doi:10.1080/0740817X.2016.1189626.
- [12] J. Frey, K. Mueller, S. Hellmann, E. Rahm, and M.-E. Vidal. Evaluation of metadata representations in RDF stores. *Semantic Web*, 10(2):205–229, 2018. doi:10.3233/SW-180307.
- [13] M. Frommhold, R. Navarro Piris, N. Arndt, S. Tramp, N. Petersen, and M. Martin. Towards versioning of arbitrary RDF data. In A. Fensel, A. J. Zaveri, S. Hellmann, and T. Pellegrini, editors, *SEMANTiCS 2016: Proceedings of the 12th International Conference on Semantic Systems*, pages 33–40. Association for Computing Machinery, New York, 2016. doi:10.1145/2993318.2993327.
- [14] M. Gergatsoulis and P. Lilis. Multidimensional RDF. In R. Meersman and Z. Tari, editors, On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE, volume 3761 of Lecture Notes in Computer Science, pages 1188–1205. Springer, 2005. doi:10.1007/11575801_17.
- [15] M. Klein, H. Shankar, L. Balakireva, and H. Van de Sompel. The Memento Tracer Framework: Balancing quality and scalability for web archiving. In A. Doucet, A. Isaac, K. Golub, T. Aalberg, and A. Jatowt, editors, *Digital Libraries for Open Knowledge*, volume 11799 of *Lecture Notes in Computer Science*, pages 163–176. Springer International Publishing, 2019. doi:10.1007/978-3-030-30760-8_15.
- [16] M. J. Kleinelanghorst, L. Zhou, J. Sikorski, E. FooYi Shyh, L. K. Aditya, S. Mosbach, I. A. Karimi, R. Lau, and M. Kraft. J-Park Simulator: Roadmap to smart eco-industrial parks. *Proceedings of the Second International Conference* on Internet of Things, Data and Cloud Computing (ICC '17), pages 1–10, 2017. doi:10.1145/3018896.3025155.
- [17] M. Kraft and S. Mosbach. The future of computational modelling in reaction engineering. *Philosophical Transactions*, 368:3633–3644, 2010. doi:10.1098/rsta.2010.0124.
- [18] N. Lopes, A. Zimmermann, A. Hogan, G. Lukácsy, A. Polleres, U. Straccia, and S. Decker. RDF needs annotations. Technical report as part of W3C Workshop - RDF Next Steps, 2010. URL https://www.w3.org/2009/12/rdfws/papers/ws09.
- [19] W. Marquardt, J. Morbach, A. Wiesner, and A. Yang. OntoCAPE: A Re-usable Ontology for Chemical Process Engineering. Springer, 2009. doi:10.1007/978-3-642-04655-1.

- [20] S. Nešić. Semantic document model to enhance data and knowledge interoperability. In V. Devedžić and D. Gaševic, editors, *Web 2.0 & Semantic Web*, volume 6 of *Annals of Information Systems*, pages 135–160. Springer, 2009. doi:10.1007/978-1-4419-1219-0_6.
- [21] M. Pan, J. Sikorski, C. Kastner, J. Akroyd, S. Mosbach, R. Lau, and M. Kraft. Applying Industry 4.0 to the Jurong Island eco-industrial park. *Energy Procedia*, 75:1536–1541, 2015. doi:10.1016/j.egypro.2015.07.313.
- [22] H. Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web*, 8:489–508, 2016. doi:10.3233/SW-160218.
- [23] S. Schlobach and W. Beek. Pragmatic semantics for the web of data. In Semantics for Big Data, AAAI Technical Report FS-13-04, pages 58–62, 2013. URL https:// www.aaai.org/ocs/index.php/FSS/FSS13/paper/viewFile/7607/7554.
- [24] H. Sompel, R. Sanderson, M. Nelson, L. Balakireva, H. Shankar, and S. Ainsworth. An HTTP-based versioning mechanism for linked data. In C. Bizer, T. Heath, T. Berners-Lee, and M. Hausenblas, editors, *Proceedings of the WWW2010 Work-shop on Linked Data on the Web*, volume 628 of *CEUR Workshop Proceedings*, 2010. URL http://ceur-ws.org/Vol-628/ldow2010_paper13.pdf.
- [25] R. Taelman, M. V. Sande, J. Van Herwegen, E. Mannens, and V. Ruben. Triple storage for random-access versioned querying of RDF archives. *Journal of Web Semantics*, 54:4–28, 2019. doi:10.1016/j.websem.2018.08.001.
- [26] J. Tappolet and A. Bernstein. Applied temporal RDF: Efficient temporal querying of RDF data with SPARQL. In L. Aroyo, P. Traverso, F. Ciravegna, P. Cimiano, T. Heath, E. Hyvönen, R. Mizoguchi, E. Oren, M. Sabou, and E. Simperl, editors, *The Semantic Web: Research and Applications*, volume 5554 of *Lecture Notes in Computer Science*, pages 308–322. Springer, 2009. doi:10.1007/978-3-642-02121-3_25.
- [27] Y. Tzitzikas, Y. Theoharis, and D. Andreou. On storage policies for semantic web repositories that support versioning. In S. Bechhofer, M. Hauswirth, J. Hoffmann, and M. Koubarakis, editors, *The Semantic Web: Research and Applications*, volume 5021 of *Lecture Notes in Computer Science*, pages 705–719. Springer, 2008. doi:10.1007/978-3-540-68234-9_51.
- [28] H. Van De Sompel, M. L. Nelson, R. Sanderson, H. Shankar, L. Balakireva, and S. Ainsworth. Memento: Time travel for the web, 2009. URL https://arxiv. org/abs/0911.1112.
- [29] L. Zhou, M. Pan, J. J. Sikorski, S. Garud, L. K. Aditya, M. J. Kleinelanghorst, I. A. Karimi, and M. Kraft. Towards an ontological infrastructure for chemical process simulation and optimization in the context of eco-industrial parks. *Applied Energy*, 204:1284–1298, 2017. doi:10.1016/j.apenergy.2017.05.002.

- [30] L. Zhou, C. Zhang, I. A. Karimi, and M. Kraft. An ontology framework towards decentralized information management for eco-industrial parks. *Computers & Chemical Engineering*, 118:49–63, 2018. doi:10.1016/j.compchemeng.2018.07.010.
- [31] X. Zhou, A. Eibeck, M. Q. Lim, N. B. Krdzavac, and M. Kraft. An agent composition framework for the J-Park Simulator – A knowledge graph for the process industry. *Computers & Chemical Engineering*, 130:106577, 2019. doi:10.1016/j.compchemeng.2019.106577.
- [32] X. Zhou, M. Q. Lim, and M. Kraft. A smart contract-based agent marketplace for the J-Park Simulator a knowledge graph for the process industry. Submitted for publication, 2020.