SiMoNa: A Proof-of-concept Domain-Specific Modeling Language for IoT Infographics

Cleber Matos de Morais  
Universidade Federal da Paraiba  
João Pessoa/PB, Brazil  
cmorais@gmail.com

Judith Kelner, Djamel Sadok  
Universidade Federal de Pernambuco  
Recife/PE, Brazil  
{jk,jamel}@gprt.ufpe.br

Theo Lynn  
Dublin City University  
Dublin, Ireland  
theo.lynn@dcu.ie

Abstract

The Internet of Things (IoT) is a prominent concept in academic and technology business discourse in recent times reflecting a wider trend to connect physical objects to the Internet and to each other. This idea of connecting things started in the beginning of the 2008 through RFID. But after the open hardware initiatives (as Arduino), it gained more visibility and access to experiments with sensors in the environment. The IoT is already generating an unprecedented volume of data in greater varieties and higher velocities. Making sense of such data is an emerging and significant challenge. Infographics are visual representations that provide a visual space for end users to compare and analyze data, information, and knowledge in a more efficient form than traditional forms. The nature of IoT requires a continuum modification in how end users see information to achieve such efficiency gains. Conceptualizing and implementing infographics in an IoT system can thus require significant planning and development for both data scientists, graphic designers and developers resulting in significant costs, both in terms of time and effort. To address this problem, this paper presents SiMoNa, a domain-specific modeling language (DSML) to create, connect, interact, and build interactive infographic presentations for IoT systems efficiently based on the model-driven development (MDD) paradigm. The language and approach are validated using real-world use cases.

1. Introduction

The Internet of Things (IoT) is a prominent concept in academic and technology business discourse in recent times reflecting a wider trend to connect physical objects to the Internet and to each other. This idea of connect things started in the beginning of the 2008 through RFID. But after the open hardware initiatives (as Arduino), it gained more visibility and access to experiments with sensors in the environment. The IoT is already generating an unprecedented volume of data in greater varieties and higher velocities. An IoT system spectrum deals with many variables that can be used to characterize each application. For example, one can characterize an IoT application with just two variables, such as area and data intensity. A Smart Home would be a use case scenario with a small area and a low data intensity whereas a Smart City is illustrative of a use case with a large area and high data intensity. Similarly, a Smart Factory may be characterized in terms of a small area, for example a warehouse, with high data intensity resulting from the use of sensors. While all these applications fall within the Internet of Things, each one not only has a different type of area and data intensity but criticality.

In each of the above scenarios, the data volume is large. The Smart City and Smart Factory also require that data processing and use are near-real time. Despite this, most of the data collected by an IoT system is not processed or often isn’t even stored for future analysis. In fact, it is estimated that less than 1% of IoT data is used for decision making. Using this deluge of data to build information and knowledge for decision making is a significant business challenge. Machine assistance through machine learning techniques are enhancing this ability, generating new information to feed and help decision making but ultimately the human decision maker plays a central role. The human eye is the most data intensive and efficient sense in the human body playing a role facilitating memorization in many cases.

The nature of IoT requires a continuum modification in how end users see information to achieve such efficiency gains. Thus, the computer graphics gained new challenges, especially the data visualization. Infographics are a popular form of diagrammatically representing data. Conceptualizing and implementing Infographics in an IoT system can require significant planning and development for both data scientists, graphic designers and developers resulting in significant costs, both in terms of time and effort. To address this problem, this paper presents...
SiMoNa\textsuperscript{1}, an Infographic Domain-Specific Modeling Language (DSML) to create, connect, interact, and build interactive Infographic presentations for IoT systems efficiently based on the model-driven development (MDD) paradigm.

The language proposed has its roots in prior IoT domain-specific languages, such as \cite{11,12}, but SiMoNa is more focused on Infographic visualization of IoT data rather than the IoT architecture as a whole. From a visual perspective, authors in \cite{13} deal with the representation of Big Data in a geo-spatial context. From a domain modeling language perspective, the works \cite{14,15} are very similar to SiMoNa, but applied to a different domain i.e. automated software engineering tools. As a result, SiMoNa is unique both in terms of the type of data visualization, infographics, and domain being visualized.

This paper is organized as follows. Section \ref{sec:infographics} introduces infographics and data visualization as human interfaces to information. It also introduces Model-driven Development (MDD) and related DSMLs as a key methodology to address the infographic dynamics in IoT systems. Next, the SiMoNa DSML is presented with its metamodel and elements. The language is validated in Section \ref{sec:validation} followed by the Conclusion (section \ref{sec:conclusion}), and Future Works (section \ref{sec:future_works}).

\subsection*{1.1. Contributions}

The main contributions of this work are:

\begin{itemize}
\item A Model-Driven Development approach to development of Infographic interfaces for IoT;
\item The meta-modeling, creation and implementation of a DSML for Infographics in a replicable environment;
\item Validation of the SiMoNa DSML in two IoT use case scenarios.
\end{itemize}

\section*{2. Infographics and Data Visualization}

In the visual display of quantitative information \cite{16}, Tufte encourages designers and statisticians to create accurate visual representations with the basic principle of "graphics reveal data". In the same way, computer scientists are exploring the opportunities emerging from the intersection of digital and interactive graphics and Big Data. Computer science makes a significant contribution to data visualization through reducing the economics of creating the graphic, increasing flexibility to recreate a graphic, and enhancing user interaction with the graphic. In addition to these more general contributions, interaction, in particular, has changed how we experience and understand graphical data.

The simultaneous observation and interaction with a graphic creates a cognitive dual visual experience for the user \cite{17,18}. This interaction experience activates two different parts of the brain at the same time. The visual conscious perception part (vision-for-perception) and the call-for-action (vision-for-action) visual perception are activated. For example, when someone sees a cup of coffee on a table, this is processed in two parts of the brain contemporaneously. The image is separated from the background so that the cup is perceived within the environment stimulus. Then, the call-for-process is instigated to map the physical motor system to trace and pick up the cup. Even if the person does not want to pick up the cup, the brain prepares the human motor system to be ready to do so. In such a way, the user experiences both a visual stimulus and a physical call-for-action when interacting with a touchscreen panel. This dual mental process is especially useful in high-skilled task use case scenarios such as IoT systems. The ability for an IoT system to communicate data visually and interactively is critical for an IoT system, so that end user can perceive a event and act accordingly in a complex environment \cite{18}.

\subsection*{2.1. From Graphics to Infographics}

Data visualization is not a new information technique; examples can be found as far back as 16th century \cite{10}. Notwithstanding this, it has gained momentum in recent years as computing power, computer visual graphics, and data analytics tools have evolved. The goal of visualization is to optimize applications so that they help us perform cognitive work more efficiently \cite{18}. But sometimes there is no preconceived query while analyzing a Big Data scenario. The information is not only presented in one representation, but from multiple different perspectives of the same reality. For such analysis, a more complex visual representation is required.

Infographics are a diagrammatic representations of data \cite{10} that have been popularised more recently, in a basic form, by social media. Infographics are more complex than a series of graphics presented together or a story told with images. At its core, an Infographic represents a purposeful diagramming of each information source, thus each graphic (and even non-graphic information) has a predefined purpose (and associated meaning) in the visual space. There is a narrative in the Infographic scope, with syntax and semantics.

\footnote{SiMoNa is an acronym for Monitoring and Analytics Information System in Portuguese.}
The three main elements of Infographics are (i) data substance, (ii) relevant statistics, and (iii) design [16]. In the IoT context, the data is typically provided by sensors. Despite this, not all sensor data is relevant for a specific use case. For example, even if the power distribution unit (PDU) could offer wattless [19] charge information, this is not necessarily useful for energy efficiency decision making in a Smart Home use case. In contrast, such information might be critical in an industrial or business use case. Useless data represents noise in visualization. The data substance must fit the use case, regarding both the quantity and quality of data to be presented to the user.

Statistics are at the core of data processing. Merely presenting data on a screen does not help the end user in the decision making process. The system must offer information in a clear and thoughtful way to enhance the data-information-knowledge continuum [20]. The capability to process data, compare it, and present those results to the user in a meaningful way is both central and critical to utility of Infographics.

Design is the final presentation of all the relevant information to the user. A narrative bonds the data presentation scope to facilitate the user’s perception of information [21]. This narrative is composed through the aesthetic applied in an effective way to present information. As a language grammar, the visual representation has presentation rules [22].

Based on those these three pillars - data substance, relevant statistics, and design - this work presents some basic principles to define interactive Infographic systems:

1. The data source must fit the user requirements in terms of relevance, quantity, and timeliness;
2. The Infographic must allow the user to compare precisely the data presented in the same context;
3. The design narrative must be consistent and have a meaning for each section of the Infographic;
4. The Infographic should allow the user to query, investigate, explore, mark, create triggers, and compare data and information in the same interface;
5. The Infographic system should react to a data level defined by the user and automatically store new information while feeding back this new information for visualization and analysis.

Conceptualizing and implementing Infographics in an IoT system can thus require significant planning and development for data scientists, graphic designers and software developers teams. Each new element in this complex representation system (new data sources, new graphics, new statistical methods or new narratives) incurs costs in terms of time and effort. To address this problem, this work considers the use of Model-driven Development (MDD) theory as a key strategy to deal with this complexity.

2.2. Model-driven Development (MDD)

Model-driven Development (MDD) is a methodology based on the concepts of software diagrams and software development. According to [23], instead of requiring developers to spell out every detail of a systems implementation using a programming language, creating documentation, and coding, it would be more efficient if developers could just model the system, describing the architecture and functionality.

In this way, by using MDD, developers can deal with high level abstractions to define their system requirements, and then automatically generate the required code [24]. The code samples for the code generator are provided by domain specialists and tested in the unit of production. As a consequence, the software development becomes more resilient to requirement changes (especially in dynamic scenarios, such as IoT systems) and the generated code has higher quality.

To make use of MDD, it is necessary to define a Domain Specific Modeling Language (DSML) to describe the system requirements. DSMLs are easier to specify, understand and maintain than general propose modeling languages. According to [25], DSML promotes productivity of modeling and also contributes to model quality since the DSML concepts should be the result of an especially thorough development process. The integrity of models is maintained because the syntax and semantics of a DSML can prevent nonsensical models. Furthermore, a DSML will often feature a special graphical notation (concrete syntax) that helps to improve the clarity and comprehensibility of models [25].

3. SiMoNa, An Infographic DSML

This work presents SiMoNa, an Infographic Domain-Specific Modeling Language. SiMoNA is an extension of the SiMoN IoT system, developed by [26]. The main requirement of the language is to address a wide range of IoT Infographics in a quick and efficient way through the MDD paradigm.
3.1 Tool and Meta-metalanguage

SiMoNa was built with The MetaEdit+ Workbench 5.5 software and its meta-metamodelling language [27]. MetaEdit+ uses a GOPPRR metamodelling language, widely used in software development and research. GOPPRR is an acronym for the languages base types: Graph, Object, Port, Property, Relationship and Role.

The main reason to select MetaEdit+ relies on the experience, replicability and extensibility of the internal process of its metamodel. Anyone can validate and further extend this work as needed. Also, the main approach to the model is visual so a graphical language has better visual representation of the Infographics displacement and configuration.

3.2 Roles

Although developing an Infographic System might seem to be a task for one person specialising in the use of the DSML, for advanced use cases, such as IOT, a variety of specialised roles are required. At bare minimum, the team is composed of a developer, a modeler, a database administrator, a designer, and a final end user. This team working together can deploy a real solution to an IoT visualization.

The developer instigates the development process by preparing the unit of code to be generated. This is a very important phase, as the unit of code must be error prone both at the component level and system-level. The database administrator (DBA) works with the developer, planning and optimizing the queries to the database in line with use case requirements. This database control is particularly critical in IoT due to the volume, velocity and variety of data generated by IoT scenarios. The designer is responsible for specifying visual layouts of charts in a meaningful way to the modeler. Feedback and validation in the end user space is key to the effectiveness of a given solution. As such end user participation in the process is essential both for setting and validating use cases. Furthermore, as any given data visualisation solution can be iterated numerous times, end users can actively contribute to the final product.

The modeler is responsible for binding all the elements through the language. After discussing the use case, the modeler starts binding the elements of the system. For each data source, the DBA might be asked to specify the most appropriate way to retrieve the data from the database. The new design interfaces should be integrated in to the metamodel so the modeler is constantly updating new visual models, charts, and narratives. Also the model binding should follow the designer’s visual specification for the use case.

![The SiMoNa Metamodel](image)

Figure 1. The SiMoNa Metamodel. The arrows represent possible interconnections between elements. The number next to an arrow represents the multiplicity of the element.
3.3. Metamodel

The metamodel elements are presented in Figure 1.

The main part of the metamodel is the Infographic. In this version, each Infographic screen is composed from up to six charts and a panel for notices and warnings. This standard Infographic configuration emphasizes consistency as required by principle (3) above. The Infographics for a given system are presented as a full screen panel. Each system has at least one Infographic, but may have multiple compositions. An Infographic element can connect to another element of the same type, representing a total screen change with new graphs and actions (see section 6 for limitations in the scope of this work). Also, the Infographic might contain, as a sub-graph, another description of the visual elements represented as a small dark circle in the lower right corner. The main graph could represent the entirety of a system (as presented in Figure 4) and the sub-graph specific use cases (Figures 5 and 6).

The second element is the charts (bar charts, percentage bar charts, pie chart, gauge display charts, 7-segment display charts and line charts). A chart is a representation of one or more variables selected by the designer or modeler in a meaningful way [28]. Each chart has its data range selection to plot and update intervals, allowing the user to generate as many visualizations for each set of data as desirable. These basic types of chart meet principles (1) and (2) above in that graphics must be easily compared and relevant for a use case. Also, charts illustrate extensibility; any new variety of graph offered by the implemented language in the future can be added to the metamodel.

Two elements provide data to the charts - (i) the Data Source and (ii) the Formula. The Data Source is the element that points to the database storing the sensor’s data. In this implementation, the Data Source is a JSON URL to fetch the data from an API/database. The Formula element is a processed Data Source. With Formula, the data plotted in a chart can be parsed with a statistical method (average, median, or mode) or some complex formula made in AsciiMath [29]. In this proof-of-concept, Formula can have up to four elements, including other Formulas. The output of this Formula will be the information plotted in the chart.

The next elements are related to hypothesis, thresholds, and actions. Those elements meet principles (4) and (5) above as the user can explore and set triggers to the system. Hypothesis, thresholds, and actions are sub-elements of each chart for the sake of clarity of the modeling. The relationship between those items and the chart is a decomposition [27], as they are of the same elements family. The Hypothesis element is applied to a Chart to verify a specific threshold of a variable. The Hypothesis can be redefined (or not) by the end user in the graphic user interface, according to the use case. In the case of a non-interactive informational dashboard, this might not be important. However, in an interactive interface, this ability is critical, as the user learns and interacts with the presentation system. With this element, the Action element performs a specific task when a condition is met in the Hypothesis. This action could be, for example, sending an email or an SMS, inserting the value (and its correlated variables) in a database, or a simple warning in the Infographic notification area.

The last element in the metamodel is the Interaction. This element adds an interaction ability to some graphs. There are various use cases so the metamodel language supports both interactive (as in a tablet or computer), non-interactive interfaces (as...
4. Use Cases and Validation

The proposed DSML can be used to design and implement a wide range of IoT Infographic. An IoT energy efficiency use case was selected to illustrate and validate this assumption. This use case is a very common and practical use for Infographics as the end user needs information about the IoT system environment in real time or through historical report analysis.

While a common use case, energy efficiency has multiple end user types and each of which may require different information [30][31]. For example, a home user has different requirements than a business user or a factory user. For the purposes of this paper, six semi-structured interviews were conducted with energy managers to understand business requirements. The sample included two large scale energy users, one energy distribution manager, and three energy quality managers from an energy producer in Brazil. Interviews were recorded, transcribed and coded [32] for analysis. The main output from this data was the guide used to plan the IoT energy efficiency demands for representation in the Infographics.

4.1. Use cases description

The ultimate use cases selected were a home end user and an Energy Efficiency Manager in a university (presented in Figure 2). The home energy efficiency use case is a device-focused energy efficiency scenario. The main concerns are how energy is consumed on each device and the reporting of minor use case incidents such as leaving the lights on during the day. The home energy efficiency use case scenario (Figure 2(a)) is a small apartment with a single bedroom, a living room, a kitchen, and a bathroom. The main concern is managing the appliances’ status and power consumption, spread across 11 sensors in the apartment. The university use case is presented in the Figure 2(b). It is a central power management office with four distinct buildings. As extrapolated from the interviews, the main objective of this IoT system is a high level management system with a low capillary of devices. There is just one main sensor on each building - a power distribution unit (PDU) in a polyphase energy system. This one sensor provides a wide range of variables: active power, reactive power, voltage, and power factor of each building.

At first glance, the data types in both cases can look very similar as both monitor energy efficiency i.e. power in watts (W), energy in kilowatts per hour (kW-h), and tension in Volts (V). But the business power consumption profile has unique characteristics. A business contract has active and reactive power [19] whereas residential contracts do not deal in wattless units. The power factor is also critical in a business IoT system and the business contract frequently has more than one electric power phase (polyphase system), each with its own information profile. Also, IoT systems will have different configurations in each use case. Using the [3] classification, the home use case would have a 2002111 scenario code and the university case would have a 1031211 code. The university case is a complex environment with more data intensity and data density than the home scenario. Also, the university business requirements require more critical call-for-actions in the event of bad energy parameters or failure (see section 5 for more information about the limitations of the home scenario).

4.2. Modeling with SiMoNa

Home energy efficiency is the first scenario modeled. The result is presented in Figure 3. As a simple system, this model has just one Infographic. The main graph is a line chart that plots the last 15 minutes of total energy consumption for four elements - the heating system, living room, bedroom and kitchen. This information, as a whole, aims to present the dynamics of the home to the end user. This chart has an interaction opportunity in that the user can use pan & zoom techniques to enhance visualization. Also, in this graph a hypothesis is constructed to detect if the living room total consumption is over the 10kWh threshold. If so, a trigger would store the values with a time-stamp in the database and send an SMS to the system’s end user.

4.2.1. Home energy efficiency use case

In panel 1, a gauge display chart compares the more voltage variation sensitive devices. The main objective is to monitor the variation of tension to prevent fails or malfunction. The second panel focuses on high energy consumption appliances such as washing machines. Here, the end user may monitor the impact of the appliance at different settings e.g. cold water wash, hot water wash, drying etc. The third panel presents the low energy consumption devices in a percentage bar chart. The 5th panel presents the active power real-time data in a series of 7-segment displays, one for each sensor. This kind of presentation is visualized in a similar way to a real energy monitor. The 6th panel compares the general appliance consumption of each house’s room using a pie chart.

The university model is much more complex.
(see Figure 4). It has five Infographics displayed in a page layout. The main Infographic, in its main line chart on the 4th panel, gathers the consolidated data from all the buildings’ active power. Panels 1 and 2 compare the actual total Active Power and Reactive Power with historic data from the previous month. The third panel presents each buildings’ voltage in a bar chart with a five-minute refresh rate. The 5th panel is a 7-segment display, as per the previous use case, but gathering each building’s active power. Finally, the 6th panel presents a percentage bar chart showing each building’s Active Power and Reactive Power ratio in a percentage bar.

4.2.2. University energy efficiency case The other Infographics were split in to two sub-graphs for better design and clarity. The first model (Figure 5) corresponds to an average energy efficiency scenario and the second to a more specific case (Figure 6). The Business School building, the Art School Building and Health School Building were characterized as the first sub-graph. The Engineering School building was classified as the second type. The Engineering courses demand more Reactive Power compared to other courses due to the use of motors. The energy efficiency in this case requires the management of wattless units as part of the power consumption model.  

In the first model (Figure 5), the main section is a line chart of the Active Power of each phase. This is a characteristic of the scenario as the power lines have different phases in each building. The first panel presents the data from the previous day for comparison with the current usage. The second panel presents the voltage of each phase. As an energy quality measure, this chart has Hypothesis items connected to notify the Energy Efficiency Manager when a voltage problem reaches a specific threshold (in this case, more than 245 volts or less than 215). The third panel shows the current per phase in a bar chart. The last panel presents the Active Power and Reactive power ratio in a pie chart. The Energy Efficiency Manager must check this power ratio periodically but it is not as critical her compared to the second model focussing on Active and Reactive Power as described below.  

The second Infographic model (Figure 6) focuses on Active Power and Reactive Power. This data is plotted in panels 1 to 3, one for each phase. A pie chart represents this ratio with latest data refreshed in real time. Also a Hypothesis is placed in those graphics with a contract level of one-third of the wattless units [19]. Each time that any phase level reaches this threshold, the Infographic will display a warning in the end user interface and log this in the database. This information is useful to recreate contracts [19] or activate compensation devices [33]. As with the first model, the voltage is monitored per phase in a bar chart with a constructed Hypothesis level. The main area has a bar chart with the last 24 hours energy consumption for the building.  

Both use cases are very different by nature. Still,
the SiMoNa language was capable of creating a visual representation of each in an Infographic IoT system. The modeling possibilities are endless to fit a wide variety of requirement. As SiMoNa uses Metaedit+ [27], the modeling software already has a working editor for the language to enable faster adoption and extension to other use cases.

5. Conclusion

The complexity of IoT information systems require a fast and adaptable solution to handle data visualization. This paper proposes SiMoNa, a domain-specific modeling language (DSML) based on model-driven development (MDD) to provide visual information through Infographics to handle data that is generated from IoT systems.

By using SiMoNa, it is possible to model and generate an Infographic system to visualize, compare and analyze data generated by an IoT system. SiMoNa can model a wide range of use cases, from a display screen (non-interactive) to an full interactive touchscreen monitor used by a controller user on an IoT scenario. And even variation in-between of those scenarios. With the evolution of the meta-model, the language can even deal with specific requirements, as a very high resolution screens or a mobile applications.

This paper presented two real-world use cases to validate and evaluate the SiMoNa language - a home and a university. Our results suggest that SiMoNa is flexible, extensible and consistent, as measured by the principles presented in section 2.1. Although the visual representation of each mode might look complex, the code generation is accurate. Thus, the visual documentation of the presentation can fit both human use and machine compiling.

6. Future Works and Limitations

It is expected that some requirements will vary and/or improve during a real world application. The actual version is planned to generate code for Qt 5.11 [33]. Notwithstanding this, it can be expanded to meet alternative use requirements and applications. Future iterations of the model may integrate new Infographic models. These can be easily added to the metamodel as it has been designed to accommodate more than one base Infographic. The selection of Metaedit+ is noteworthy from an adoption perspective. It is a commonly used DSM environment and integrated in to popular modeling, metamodeling, and code generation tools. The MetaEdit+ API supports open SOAP, Web Services and .NET standards and models can be imported and exported via standards-compliant XML files. While this supports adoption by the development community, it should be noted that greater examination of the human, organisational and technological factors that influence infographic system adoption and use is needed. Furthermore, notwithstanding the extensibility and interoperability factors outlined above, further experimentation and testing is required to ensure the scalability of the system given predicted IoT data volumes several orders of magnitude greater than today.

This work considers a home user with basic energy management. While more complex home IoT scenarios were outside of the scope of this work, it should be noted that users can reconfigure criticality settings for different scenarios e.g. e-health applications such as a BiPAp

![Figure 4. The university’s energy efficiency use case modeled in the SiMoNa DSML. Each building Infographic is modeled in a different sub-graph for ease of maintenance and better visualization of the model](image)
Figure 5. The university energy efficiency scenario sub-graph A modeled in the SiMoNa DSML. The main objective of this Infographic is to track energy alteration and react quickly.

Figure 6. The university energy efficiency scenario sub-graph B modeled in the SiMoNa DSML. The reactive energy is the main focus of this Infographic - Technology courses may use motors thus unbalancing the energy load per phase.

machine (Bilevel Positive Airway Pressure breathing machine) and this is worthy of exploration in future works. Furthermore, actuation is an important function of IoT systems. Actuation and how to visually describe it is a complex action that would broaden the scope of this paper too much but again, it is worthy of integration in future iterations of SiMoNa and associated studies.

7. Acknowledgement

This work is partly funded by the Irish Centre for Cloud Computing and Commerce (IC4), an
References


