Towards SWIM Narratives for Sustainable Water Management

Raúl Alejandro Vargas-Acosta[1], Luis Garnica Chavira[2], Natalia Villanueva-Rosales[1], and Deana Pennington[1]

[1] The University of Texas at El Paso, El Paso TX 79968, USA
[2] Git Gud Consulting SAS, 31203 Chihuahua, Mexico
`ravargasaco@miners.utep.edu, luis@gitgudconsulting.com,`
`{nvillanuevarosales,ddpennington}@utep.edu`

Abstract. The creation of scientific models to understand water availability under different scenarios is an important step towards pursuing a sustainable water future. A wide variety of scientific models have been created for understanding the different elements driving water availability in urban, agricultural and ecological settings. The Sustainable Water through Integrated Modeling Framework (SWIM) enables a wide range of stakeholders to run water-sustainability model scenarios through participatory modeling. Although SWIM is a science-driven platform, it was created with input from diverse stakeholders with the goal of improving how water models can be used and shared. SWIM aims to foster a better understanding on the impact that decisions about water usage can have. This paper describes our efforts towards translating the science behind the models generated in SWIM into English and Spanish explanations, also known as narratives. We anticipate that narratives will better communicate the meaning of specific water-economics scenarios under different perspectives, including urban, agriculture and environmental. Thus, assisting stakeholders in decision making.

Keywords: Participatory Analysis, SWIM, Narratives, Stakeholders.

1 Introduction

In 2015, all 193 members of the United Nations agreed on a set of seventeen Sustainable Development Goals to be achieved by 2030, one of which is achieving clean water and sanitation for all [1]. Achieving such a comprehensive goal will require unprecedented collaboration between researchers from many disciplines, professionals in management and decision making, and diverse stakeholders across public sectors [2]. Understanding the range of plausible future water conditions depends on integrating complex hydrologic surface and groundwater models along with models of changing climate, population, land use, economics, technology, management, and policy – each of which has a range of plausible future trajectories [3, 4]. It is becoming increasingly important for researchers who develop such complex models to work directly with stakeholders using "participatory modeling" methods, which engage stakeholders in the process of defining modeling goals, envisioning future scenarios of change, and exploring the results of these potential changes [5, 6]. Participatory modeling of water management may include an array of stakeholders with different roles (e.g., decision makers), from different sectors (e.g., agriculture), representing different scales (local, regional, national, or global) with different levels of power and ability to influence change [7]. Hence, this context incorporates complex models with many plausible configurations due to the large array of stakeholders with different perspectives and interests in the modeling process. This paper describes The Sustainable Water through Integrated Modeling Framework (SWIM) approach for supporting stakeholders to better understand and interpret the underlying water models considering their particular perspective.

Related Work

Communicating hydrologic model's outputs within an understandable context is fundamental when performing water scarcity projections. A literature review shows that a common practice is the use of data visualizations (e.g., graphs, maps). In [8–11] Swain, Demir and Krajewski, and Kulkarni et al. propose the Tethys Platform, the Iowa Flood Information System, and the Integrated Flood Assessment Model respectively to model different environmental events (e.g., flooding) using graphs and maps. Similar to these systems, SWIM provides undirected graphs to communicate provenance (i.e., the origin and processing of data) and an interactive map that shows information about the underlying models. Ongoing work in SWIM includes the dynamic visualization of outputs in a map.

2 Model Narratives in SWIM

SWIM's team identified the need of providing context for inputs and outputs of the system to better communicate the meaning of the model scenarios generated. Further, we recognized the need to target different groups of stakeholders when providing this context, i.e., consider different perspectives. A similar was identified by Gil and Garijo [12], who proposed principles for the automated generation of *data narratives*. Some of these principles align with the SWIM goals, and thus we adopted the term *model narratives* for our efforts.

2.1 SWIM Narrative Components

Following [12] principles, narratives in SWIM are composed by interlinked provenance records, system identifiers for each model run and accounts (i.e., SWIM narrative templates) that provide descriptions of water model elements with a different focus or level of detail. In addition, SWIM enables the customization of narratives from different perspectives, including user roles (e.g., farmers, urban planners), and geographical region (e.g., from both sides of the US-Mexico border). SWIM's stakeholders currently define roles. The core concepts of SWIM model narratives are: i) *Language:* Human language, currently English or Spanish; ii) *User role:* Role of a user that defines the level of detail in model narratives, e.g., farmers, policymakers, water administrators; iii) *Focus area:*

Perspective of the narrative filtering components that the narrative will focus on, e.g., urban, environmental; iv) *Geographical Region*: Perspective of the narrative according to regional interests, currently El Paso (Texas), Las Cruces (New Mexico) and Ciudad Juarez (Mexico); and v) *Model Element:* model inputs (e.g., crop acreage), model outputs (e.g., surface water storage), and scenarios (e.g., big stress climate).

Narratives in SWIM aim to provide context to model elements. Inputs are determined by a predefined scenario from a set of scenarios created by SWIM's modeling team. SWIM enables users to modify selected input parameters within a predefined threshold. Narratives include additional explanations on assumptions applied to each model element and data provenance. Narrative templates were manually created by SWIM's modeling and technical team according to different user perspectives and added to the database collection of narrative components. The structure of SWIM narratives is composed by: i) Text templates, ii) Statistical values, iii) Adjectives, and iv) Responsive explanations. A *text template* is a self-contained unit in a human language that provides a basic structure of a narrative, it includes placeholders that are replaced with statistical values, adjectives, labels and explanations assembled at runtime. *Statistical values*, such as average, are calculated for a specific output value to provide aggregated data used for comparison with other model scenarios or a baseline. Statistical values are also used to complement generated data plots on the SWIM interface. *Adjectives* are used to describe trends and behavior on output values (e.g., increase) and for value comparison with baseline results. Comparative adjectives relate a custom simulation to a baseline scenario. The baseline scenario on SWIM's water model replicates observed data and water management strategies at a historical time period between 1995 and 2015 [13]. With statistical values from a custom simulation, SWIM can establish if an output average was reduced or increased in comparison to the baseline value. *Responsive explanations* describe the context of model elements from a specific perspective. Responsive explanations are determined by provenance, relationships between other model elements, and level of granularity according to the stakeholder role.

2.2 SWIM Narrative Semantics

Our ongoing work includes exposing narratives and provenance as linked data that can be automatically digested by semantically-enhanced applications. SWIM's data and narrative concepts are annotated by extending database documents to JSON-LD[1].
The SWIM vocabulary[2] and SWIM Ontology[3] reuse widely used vocabularies including Dublin Core Metadata Terms[4], The PROV Ontology[5], and Schema.org[6].

---

[1] https://www.w3.org/TR/json-ld/

[2] http://purl.org/swim/vocab

[3] http://purl.org/swim/terms

[4] http://purl.org/dc/terms/

[5] https://www.w3.org/TR/2013/REC-prov-o-20130430/

[6] http://schema.org/

2.3 SWIM Narrative Samples

In this section, we illustrate the generation of SWIM narratives for the execution of the Bucket Model [13]. The Bucket Model is a simple-basin created by SWIM modelers that simulates all major water sources, sinks, usage and economic values as well as institutional constraints governing water supply and use. The Bucket Model covers the geographic area of the Middle Rio Grande between the inflow to Elephant Butte Reservoir and Fort Quitman[7]. This model is categorized as a constrained optimization model targeted to maximize net economic benefits under predefined institutional restrictions.

Structure	~element_label <template> ~adjective_trend <template> ~percent ~adjective_bevaviour <template> ~constant_year <template> ~maxValue ~element_unit <template> ~maxYear <template> ~minValue ~element_unit <template> ~minYear
0-Perspective	General Public
0-Text-EN	Surface Water Storage follows a downward trend with a 75% reduction by the end year 2033 with a peak volume of 2213 KAF in 1997 and lowest volume of 541 KAF in 2033.
0-Text-ES	El Almacenamiento de Agua Superficial sigue una tendencia de caída con una reducción de 75% en las reservas de agua para el año final 2033 con un volumen máximo de 2213 KAF en el año 1997 y volumen mínimo de 541 KAF en 2033.

Table 1. SWIM narrative for the output variable “Surface Water Storage.” Outputs are marked in blue, templates in black, inputs in purple, adjectives in red, and statistical values in orange. The character “~” identifies placeholder fields and ”< >” characters denote template sections.

Consider use case scenarios of SWIM users with different roles:

General Public: A citizen living in the Middle Rio Grande region has just read an online news story about the likelihood that changes in climate that have been observed in recent decades will continue into the future. The citizen is primarily concerned about how much water will be entering Elephant Butte Reservoir. A sample element narrative under this perspective is presented in Table 1.

Scientific: A researcher is studying the effects of climate change on the Middle Rio Grande water supply and interested in reproducing climate simulations used to establish different inflow scenarios at Elephant Butte reservoir.

Water Administrator: An El Paso Water Utilities manager is tasked with long-term strategic planning for water availability for a growing regional population. The manager uses SWIM to project the costs of providing water, assuming a population growth rate equal to the current growth rate, and under a worst-case scenario of Big Stress climate.

2.4 Automated Generation of Narratives

SWIM uses the Natural Language Narrative Generator (NLNG)[8] implemented as a Restful Web Service, it generates model element narratives on request. NLNG retrieves

---

[7] http://purl.org/swim

[8] http://purl.org/swim/services/nlng

narrative components from SWIM's database for processing. SWIM'S database is currently managed with an instance of MongoDB and data interaction is leveraged using Object Document Mapping provided by the Morphia[9] library. The narrative generation workflow is described as following: a user selects his or her role (introduced in section 0) before executing a custom model scenario on the online interface. After submitting the scenario, the backend processes model results along with a provenance trace of the execution (i.e., the origin of the values used to execute the model). SWIM's frontend can then request specific model element narratives on user demand. NLNG performs a GET request to SWIM's data endpoints to retrieve the executed user scenario to be used on the narrative creation, work on the user scenario model is covered in [14]. Fig. 1 shows an excerpt of time series data for the output "water_stocks" in JSON format.

```
{"varLabel": "Surface Water Storage","varName": "water_stocks",
"varDescription": "Reservoir water storage","varUnit": "KAF/yr",
"varValue": [{"res": "Store_res_s","t": "1996","value": 2061.5},{"res":
"Store_res_s","t": "1997","value": 2213.49}...}
```

Fig. 1. Excerpt of an executed scenario serialized as JSON using SWIM's data model. The response contains metadata and result values for the output variable "water stocks."

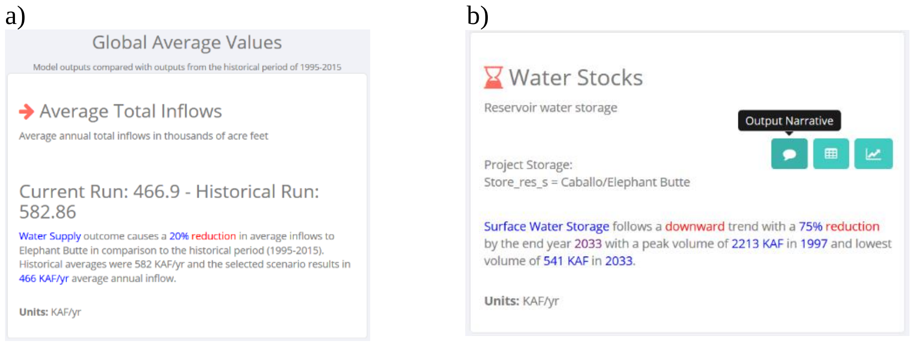
Once the user scenario is retrieved, statistical and aggregation calculations are performed over the retrieved data sets. The resulting statistical data will replace corresponding placeholders on the narrative templates. Narrative templates are also stored as a collection of documents in MongoDB. Fig. 2 illustrates a query for an output narrative template (Q1) returning a template with placeholder fields "~…" (R1). Queries performed internally by the NLNG web service are not available to end users, the query criterion is set as parameters on the service URL. The resulting narrative is displayed in SWIM's frontend as illustrated in Fig. 3.

```
Q1. List<Narrative> outputNarrative = mDataStore.createQuery(Narra-
tive.class).filter("element.name", varName).asList();
R1. ~element_label follows a ~adjective_trend trend with a ~percent ~ad-
jective_behaviour by the end year ~constant_year with a peak volume of
~maxValue ~element_unit in ~maxYear and lowest volume of ~minValue ~ele-
ment_unit in ~minYear.
```

Fig. 2. Query to retrieve narrative template for an output variable and result.

---

[9] https://mongodb.github.io/morphia



a)

Global Average Values

Model outputs compared with outputs from the historical period of 1995-2015

Average Total Inflows

Average annual total inflows in thousands of acre feet

Current Run: 466.9 - Historical Run: 582.86

Water Supply outcome causes a 20% reduction in average inflows to Elephant Butte in comparison to the historical period (1995-2015). Historical averages were 582 KAF/yr and the selected scenario results in 466 KAF/yr average annual inflow.

Units: KAF/yr

b)

Water Stocks

Reservoir water storage

Project Storage:
Store_res_s = Caballo/Elephant Butte

Surface Water Storage follows a downward trend with a 75% reduction by the end year 2033 with a peak volume of 2213 KAF in 1997 and lowest volume of 541 KAF in 2033.

Units: KAF/yr

Fig. 3. Summary narrative (a) and single output narrative (b) corresponding to the result in Fig. 2 - R1.

An example of a call to NLNG is given in Fig. 4 where *varName_value* depicts the model element (input, output or scenario) for which the narrative is to be generated, *userType_value*, *focusArea_value*, *region_value*, and *lang_value* indicates the stakeholder type, water consumer area, region, and language respectively, finally *varValueName_value* and *varValueRegion_value* are keyholders to filter data from a selected model element.

```
http://purl.org/swim/services/nlng?usid=[usid_value]&varName=[var-
Name_value]&userType=[userType_value]&focusArea=[focusArea_value]&re-
gion=[region_value]&lang=[lang_value]&varValueName=[varValue-
Name_value]&varValueRegion=[varValueRegion]
```

Fig. 4. NLNG invocation

3 Discussion and Future Work

The generation of model explanations is a required feature for SWIM - the complexity of the science behind the water models involves domain expertise and context to be transformed into knowledge for decision making. SWIM's current infrastructure supports the generation of model narratives. However, the creation of the narrative template for the presented proof-of-concept was a challenging task – narrative templates required a collaborative effort between SWIM's technical team and modelers to portray the correct message. Our evaluation plan includes holding sessions with: 1) a group of stakeholders (which may play more than one role) from Mexico and the U.S. to analyze the appropriateness and usefulness of the narratives for stakeholders. SWIM's interface will display the wide range of perspectives to all users, highlighting suggested options in real time as the user interacts with the system; 2) the SWIM scientific team to validate the content of the narratives to ensure coverage of Bucket Model scenarios and identify the limitations of our approach. Future work includes the full implementation of auto-generated narratives in SWIM (i.e., beyond proof-of-concept). We aim to expose

SWIM annotated data as knowledge graphs to foster its reuse and document best practices for the generation of narratives in the Water Sustainability context. We will explore the prediction of narratives of interest using the matrix factorization model as a recommender system [15] to suggest the most relevant inputs and outputs of water models for a specific user, based on “ratings” from users with the same or similar role. Ratings in the context of SWIM represent preferences demonstrated explicitly (e.g., scale rating), or implicitly (e.g., the user selecting a specific variable for visualization). This information will also be used to refine user roles based on their interaction with SWIM. Integrating new models into SWIM will provide an opportunity to reuse the NLNG's infrastructure and SWIM's data model to generate narratives across models.

4 Acknowledgments

Authors thank the contributions of SWIM’s modeling team, especially Frank Ward, and Dave Gutzler. This material is based upon work supported by the National Institute of Food and Agriculture, U.S.D.A. Grant# 2015-68007-23130 “Sustainable water resources for irrigated agriculture in a desert river basin facing climate change and competing demands: From characterization to solutions.” This work used resources from Cyber-ShARE Center of Excellence, supported by National Science Foundation Grant #HRD-0734825.

References

1. United Nations: Transforming our world: The 2030 agenda for sustainable development. (2015)
2. Science, I.C. for: Review of the sustainable development goals: The science perspective. Paris: International Council for Science (ICSU). (2015). doi:https://doi.org/10.1002/2017JA024835
3. Hamilton, S.H., ElSawah, S., Guillaume, J.H.A., Jakeman, A.J., Pierce, S.A.: Integrated assessment and modelling: Overview and synthesis of salient dimensions. Environmental Modelling & Software. 64, 215–229 (2015). doi:10.1016/j.envsoft.2014.12.005
4. Laniak, G.F., Olchin, G., Goodall, J., Voinov, A., Hill, M., Glynn, P., Whelan, G., Geller, G., Quinn, N., Blind, M., Peckham, S., Reaney, S., Gaber, N., Kennedy, R., Hughes, A.: Integrated environmental modeling: A vision and roadmap for the future. Environmental Modelling & Software. 39, 3–23 (2013). doi:10.1016/j.envsoft.2012.09.006
5. Voinov, A., Kolagani, N., McCall, M.K., Glynn, P.D., Kragt, M.E., Ostermann, F.O., Pierce, S.A., Ramu, P.: Modelling with stakeholders – Next generation. Environmental Modelling & Software. 77, 196–220 (2016). doi:10.1016/j.envsoft.2015.11.016
6. Gray, S., Voinov, A., Paolisso, M., Jordan, R., BenDor, T., Bommel, P., Glynn, P., Hedelin, B., Hubacek, K., Introne, J., Kolagani, N., Laursen, B., Prell, C., Schmitt Olabisi, L., Singer, A., Sterling, E., Zellner, M.: Purpose, processes, partnerships, and products: four Ps to advance participatory socio-environmental modeling. Ecological Applications. 28, 46–61 (2018). doi:10.1002/eap.1627

7. Halbe, J., Pahl-Wostl, C., Adamowski, J.: A methodological framework to support the initiation, design and institutionalization of participatory modeling processes in water resources management. Journal of Hydrology. 556, 701–716 (2018). doi:10.1016/j.jhydrol.2017.09.024
8. Swain, N.: Tethys Platform: A Development and Hosting Platform for Water Resources Web Apps. All Theses and Dissertations. (2015)
9. Swain, N.R., Christensen, S.D., Snow, A.D., Dolder, H., Espinoza-Dávalos, G., Goharian, E., Jones, N.L., Nelson, E.J., Ames, D.P., Burian, S.J.: A new open source platform for lowering the barrier for environmental web app development. Environmental Modelling & Software. 85, 11–26 (2016). doi:10.1016/j.envsoft.2016.08.003
10. Demir, I., Krajewski, W.F.: Towards an integrated Flood Information System: Centralized data access, analysis, and visualization. Environmental Modelling & Software. 50, 77–84 (2013). doi:10.1016/j.envsoft.2013.08.009
11. Kulkarni, A.T., Mohanty, J., Eldho, T.I., Rao, E.P., Mohan, B.K.: A web GIS based integrated flood assessment modeling tool for coastal urban watersheds. Computers & Geosciences. 64, 7–14 (2014). doi:10.1016/j.cageo.2013.11.002
12. Gil, Y., Garijo, D.: Towards Automating Data Narratives. In: Proceedings of the 22Nd International Conference on Intelligent User Interfaces. pp. 565–576. ACM, New York, NY, USA (2017)
13. Ward, F., Mayer, A., Gutzler, D., Habteyes, B.: Middle Rio Grande Hydroeconomic Model for Policy Analysis: The Bucket Model. Sustainable Water Resources Management in the Rio Grande/Rio Bravo Basin Symposium. , University of Texas at El Paso (2018)
14. Villanueva-Rosales, N., Garnica Chavira, L., Rajkarnikar Tamrakar, S., Pennington, D., Vargas-Acosta, R.A., Ward, F., Mayer, A.S.: Capturing Scientific Knowledge for Water Resources Sustainability in the Rio Grande Area. In: Tiddi, I., Rizzo, G., and Corcho, Ó. (eds.) Proceedings of Workshops and Tutorials of the 9th International Conference on Knowledge Capture (K-CAP2017), Austin, Texas, December 4th, 2017. pp. 1–6. CEUR-WS.org (2017)
15. Koren, Y., Bell, R., Volinsky, C.: Matrix Factorization Techniques for Recommender Systems. Computer. 42, 30–37 (2009). doi:10.1109/MC.2009.263