Collaboration in the Model Web

Sven Schade[1], Patrick Maué[2], and Clodoveu Davis[3]

[1] Joint Research Centre - European Commission, Institute for Environment and Sustainability, Ispra, Italy
[2] Institute for Geoinformatics - University of Münster, Münster, Germany
[3] Computer Science Department - Federal University of Minas Gerais, Belo Horizonte, Brazil

Abstract. Compared to environmental data sets, an open and simple exchange of environmental models on the Web remains a vision. The lack of an architecture facilitating the different aspects of sharing models, in the sense of scientific simulations of environmental phenomena, significantly impairs collaboration between the different user groups interested in the model results. In this paper we discuss current solutions and future challenges which facilitate the collaboration in a web of models. We elaborate on the key issues, and illustrate our vision of a collaborative Model Web by detailing a possible usage scenario. We analyze different aspects of environmental models, investigate the involved actors, and define the different means of collaboration between these actors. This work provides a mid-term goal for the development of the Model Web and thereby for a central part of the Digital Earth.

1 Introduction

The Digital Earth will enable everybody to explore the existing wealth of geographic information (GI) [1]. Visualized on a 3D globe, spatial data is supposed to be seamlessly integrated from a wide range of sources. This includes real-time observations from physical sensors as well as environmental models which assess, for example, the spread of forest fires. In this sense, the Digital Earth can be understood as network of interlinked data sets, models, processes, and tools for visualizing these resources adequately.

Earth observations coming from physical sensors are not continuous over space and time. Environmental models are the means to fill spatial and temporal gaps, or to predict future changes in our environments based on historical or present observations. Coupling sensors with models, for example for the interpolation of temperature values, helps us to better understand spatial and temporal variations of real world processes. Initiatives, such as the global Earth Observation System of Systems[4] or the European Global Monitoring for Environment and Security[5], have established frameworks for the long-term provision of real-time or modeled Earth observations to the public. The idea of the Model

[4] More information available from `http://www.earthobservations.org/`
[5] More information available from `http://www.gmes.info/`

Web as part of the GEOSS work plan envisions interoperable computer models, embedded in a multidisciplinary network of models, data sources, processes, and sensors. It is envisioned that this web of models grows organically and that a sustainable infrastructure will emerge over time [3].

The desired seamless integration into the Digital Earth depends on established means for ensuring interoperability regarding syntax, structure, and semantics [5,2]. The geospatial web service community tries to establish this by introducing standard interfaces and data encodings [4]. However, so far the Model Web remains a vision. We miss a critical mass of content providers as well as consumers. In order to improve the current situations, we need to discuss about the involved users, as well as a description of the relations between the users themselves and between users and environmental models.

In the following paper we discuss existing solutions and future challenges which facilitate the collaboration in the Model Web. We begin with a scenario illustrating the need for new solutions to communicate the results of environmental models. We analyze different aspects of environmental models, identify the involved actors, and define the different means of collaboration between these. We end with a discussion about the participating communities and especially about the role of Digital Earth.

2 Model Web Scenario

In the desired Model Web, a scientist works on identifying the next vectors of deforestation in central Pará state, Brazil. He gathers data on the forest's current situation, and proposes a computer model for land use change that reflects the main driving forces for deforestation. The model is calibrated using simulations on past data, put together from a series of satellite images. After calibration the scientist runs it to indicate areas for probable future deforestation. The resulting scientific publication, the model, maps resulting from it, and the basic data used in the study, are published on the Web, naturally associated to the region of interest and related to the study's time of development. Other scientists who expressed interest in the covered region and subject by subscribing to a notification service get alerts about the scientist's new model results.

The responsible IBAMA (the main Brazilian environmental institute) authorities are also notified. Scientists from another institution are instructed by IBAMA to check the soundness of the model, data, methods and conclusions for the original study. Those scientists access the original raw data and model instructions to replicate the results. Other models with similar purpose are used to compare results with the original study. Meanwhile, a reporter alerted by the notification system plans to publish the study's abstract in a newspaper. The Environmental Ministry refers him to the confirmation study performed through IBAMA, which has been published as GI within the Digital Earth as well. Recent satellite images are used to further illustrate his article. A notified environmental protection NGO based in Brasilia studies the maps to get information on roads that serve the region, and decides to act to convince locals to observe the main roads. They watch out for signs of illegal wood transportation, with the goal to denounce these activities to the authorities. The NGO sends messages to its members and local citizens referring them to legislation included in the Digital Earth site. It asks them to help building a collaborative map of environmental violations in the region in Digital Earth. The result is later covered in a geo-and time-referenced article by the reporter, as a follow-up story on the original study. Top authorities in Brasilia react on the increased number of contributions and overall activity over central Pará within the Digital Earth. They analyze the contributions to understand the extraordinary behavior, and decide to organize a raid in the region with support from the Federal Police and state troopers.

The scenario illustrates how manifold interaction between users in the Digital Earth could be and indicates the variety of involved communities. The Digital Earth is not limited to visualization only; it is a new gateway to GI on the Web. This gateway comprises a broad set of communication channels, including the mentioned 3D globe, but also notification systems running on mobile devices, portals on the Web dedicated to specific thematic communities, or more common digital outlets like blogs or online newspapers. All kinds of GI, including environmental models, are distributed through these channels.

3 Understanding Environmental Models

In essence, environmental models aim to simulate real world processes. Scientists use them to answer questions regarding the spread of forest fires, long-term changes in our climate, human behavior in urban environments, and much more. Computer models implement these abstract views on reality. Stochastic models may rely on simple mathematical functions, numerical models approximate solutions based on empirical data (from physical sensors), while agent-based models aim to simulate interaction between individual entities in complex systems.

Digital libraries let us search and download scientific publications, such as the basic data and maps mentioned in the scenario. Web service standards by the Open Geospatial Consortium (OGC) enable integration of remote spatial data. Environmental models, on the contrary, are scarce on the Web. Although (geospatial) cyberinfrastructures intent to improve this situation in the context of eScience [9], no standards exist for sharing and validating models. Only few search engines and infrastructures to study model results exist, for example AuScope[6] and SEEK[7]. Howerver, those solutions mostly focus on model outputs, i.e. they do not allow to retrieve the underlying model instructions, input parameters, or input data.

In this context, the Model Web is a proposal for a dynamic infrastructure for computer models which serves researchers, managers, policy makers, and the general public. It will be 'composed of loosely coupled models that interact via web services, and are independently developed, managed, and operated' [3]. Here, the model should be understood not as single component, but as a configuration

[6] More information available from `http://www.auscope.org.au/`

[7] More information available from `http://www.seek.ecoinformatics.org/`

of modeling instructions, input data sources, and directions for calibrating input parameters. Unfolding existing models into workflows supports reusability of their individual components, better acceptance due to transparent workflows, and the option to compare models. Using standards to publish models - again as web services - ensures seamless coupling of models (e.g. combining a runoff model with a stream flow model) and the integration into existing scientific workflows [4]. In [7], we propose the use of the OGC standards of the Sensor Observations Service (SOS) and the Sensor Planning Service (SPS) for environmental models. Whereas the first may be used to predefine models (e.g. a temperature forecast model for a certain region using a preconfigured set of weather stations as input), the latter relies on input parameters defined by human users to be executable.

Still, the contemporary realization of the Digital Earth does not include means to adapt content to the clients' needs. Free-flowing knowledge reaching all interested parties, as depicted in the scenario presented in Section 2, is not yet supported. Semantics can help to bridge the different information communities. Formal semantics can capture the meaning of input parameters and how their changes impact the model's outcome. Standard formats for the workflows, data encoding, and metadata (with focus on aspects of data quality including lineage) can at least ensure a minimum level of understanding between the different actors. In [6] we discuss a solution based on rules automatically forwarding notifications to subscribers across spatial and thematic boundaries. The use of different communication channels is out lined in [8]. We yet miss an overview of all involved actors and their interplay. We address this in the following section.

4 Collaboration between Experts, Stakeholders, and Citizens

We distinguish between three kinds of users: (environmental) experts, stakeholders, and citizens. Stakeholders, such as policy-makers or reporters are naturally interested; they have to decide or to report, based on the information they get from the experts. Citizens are interested in result, since they might be potentially affected by the decisions. Experts from other information communities consider the model results to be valuable input for their own studies. The person in the scenario studying deforestation and the scientists hired by IBAMA are the experts. The reporter, the NGO, and the public authorities take the role of stakeholders. One actor may take up multiple roles: by publishing his/her analysis, the reporter himself acts as an expert delivering information to his/her readers (the stakeholders).

Figure 1 illustrates the collaboration between the different roles. Environmental experts on the left side create models predicting changes in our environment, for example the deforestation in the Amazon. It is also an expert's task to draw conclusions from the data and compile it to the information demanded by the stakeholders on the right side. Depending on the stakeholder, the information may be specified in different manners. The scientist publishes the results of a model in a research journal, describing the underlying assumptions and chosen model algorithms. The reporter distributes the information as news article to the readers of his newspaper. Public authorities retrieve short abstracts summarizing the main findings, accompanied by suggestions how to respond. It is in the responsibility of the community of experts to come up with the different realizations of the same information to address all potential stakeholders, at least until we are able to perform such a semantic adjustment automatically.

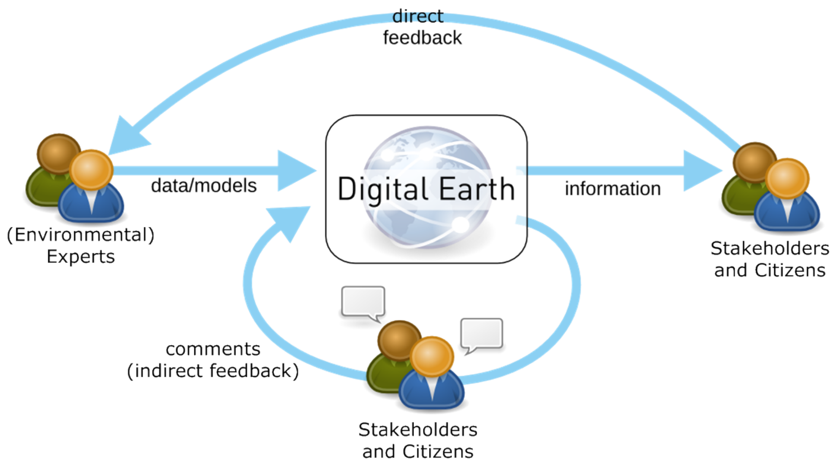

Fig. 1. Collaborating in the context of the Digital Earth

Once the information reaches the stakeholders and citizens, both may provide feedback to experts. Decision makers act to prevent the predicted events; citizens gain interest and demand for more information. We distinguish various types of feedback, with the most evident being the money used to pay the experts. Actions by authorities might have an impact on the model inputs, resulting in the need to re-run the models to update the predictions. Citizens can act as human sensors whose observations confirm (or contradict) the model results. The feedback is then used to calibrate the models, with the ultimate goal to close the gap between environmental models and reality. Users will be informed about the impacts of their feedback in order to motivate further participation.

5 Discussion and Conclusion

In this short paper, we outlined our vision of the Digital Earth using a deforestation scenario. We discussed current solutions and future challenges which facilitate the collaboration in the Digital Earth, especially in respect to the Model Web. We analyzed different aspects of models and discussed the interaction between the involved actors.

This work indicates the large amount of involved (provider and consumer) communities and the need for coordination. Whereas groups working on eScience, cyberinfrastructue, web portals, spatial data infrastructures and many more provide important building blocks, it will be a central task to the Digital Earth community to monitor the developments within these groups and to identify the links between them. Citizen involvement including bi-directional feedback should be one of the central aims. In addition, we will still require a mentality shift in eScience, in which environmental externs become willing to share their data and even more important, their models.

The vision of seamless integration of environmental models requires far more than a sound understanding of their impact on the collaboration between Digital Earth users. We just briefly mentioned model semantics, and we did not even start with issues like uncertainty. Achieving concise model descriptions and providing mechanisms for deploying models as services are central future tasks for the Model Web community.

6 Acknowledgments

The presented ideas evolved in a workshop between IFGI and the Brazilian National Institute for Space Research (INPE), funded by the German DFG (444 BRA 121/2/09). Our research on environmental models is additional funded through the European research project ENVISION (FP7-249120).

References

1. Max Craglia, Michael F. Goodchild, Alessandro Annoni, Gilberto Camara, Michael Gould, Werner Kuhn, David Mark, Ian Masser, David Maguire, Steve Liang, and Ed Parsons, *Next-generation digital earth: A position paper from the vespucci initiative for the advancement of geographic information science*, Internation Journal of Spatial Data Infrastructures Research **3** (2008).
2. F. Fonseca, M. Egenhofer, P. Agouris, and G. Camara, *Using ontologies for integrated geographic information systems*, Transactions in GIS **6** (2002), no. 3, 231–257.
3. Gary N. Geller and Forrest Melton, *Looking forward: Applying an ecological model web to assess impacts of climate change*, Biodiversity **9** (2008), no. 3&4.
4. Carlos Granell, Laura Díaz, and Michael Gould, *Service-oriented applications for environmental models: Reusable geospatial services*, Environmental Modelling & Software **25** (2010), no. 2, 182 – 198.
5. Werner Kuhn, *Geospatial semantics: Why, of what, and how?*, Journal on Data Semantics III **3534** (2005), 1–24.
6. Patrick Maué and Jens Ortmann, *Getting across information communities*, Earth Science Informatics **2** (2009), no. 4, 217–233.
7. Patrick Maué and Christoph Stasch, *Web service standards for environmental models*, Proceedings of the INSPIRE Conference 2010, June 2010.
8. Sven Schade and Max Craglia, *A future sensor web for the environment in europe*, Proceedings of the EnviroInfo Conference 2010, June 2010.
9. Chaowei Yang, Robert Raskin, Michael Goodchild, and Mark Gahegan, *Geospatial cyberinfrastructure: Past, present and future*, Computers, Environment and Urban Systems **34** (2010), no. 4, 264 – 277, Geospatial Cyberinfrastructure.