Co-producing software for complex environmental data visualization

Case Study

XXXXXXXXXXXX[*]
XXXXXXXXXXXX
XXXXXXXXXXXX
XXXXXX@XXXXXX.edu

XXXXXXXXXXXX[†]
XXXXXXXXXXXX
XXXXXXXXXXXX
XXXXXX@XXXXXX.edu

XXXXXXXXXXXX[‡]
XXXXXXXXXXXX
XXXXXXXXXXXX
XXXXXX@XXXXXX.edu

ABSTRACT

Environmental scientists, land managers, and policy actors are increasingly presented with high-stakes high-uncertainty problems stemming from human-ecosystem interactions. These interactions exacerbate already challenging issues associated with environmental policy and natural resource management. To address these problems, scientists and managers frequently use models that produce enormous geospatial and temporal datasets that are constantly modified. To help make sense of this complex and changing data, we are immersed in a co-production effort where software engineers and environmental scientists collaborate on the development of visualization software. We report on this on-going research, and find that visualization is critical not only for communicating science, but integral to many aspects of the science production pipeline and evolving data science field. We also find evidence among our collaborators that this software co-production process helps build legitimacy for the information it produces, with potential implications for generating actionable science for policy and governance.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in visualization**; *Scientific visualization*; *Geographic visualization*; *Visualization design and evaluation methods*; • **Applied computing → Environmental sciences**;

KEYWORDS

Knowledge co-production, visualization, environmental science

ACM Reference Format:
XXXXXXXXXXXX, XXXXXXXXXXXX, and XXXXXXXXXXXX. 2018. Co-producing software for complex environmental data visualization: Case Study. In *Proceedings of 19th Annual International Conference on Digital Government Research (dg.o'18),* Anneke Zuiderwijk and Charles C. Hinnant (Eds.). ACM, New York, NY, USA, Article 4, 5 pages.

[*]xxxxxxxxxxx
[†]xxxxxxxxxxx
[‡]xxxxxxxxxxx

1 INTRODUCTION

As interactions between humans and ecosystems continue into the 21st century, the impacts of these interactions are a critical societal concern [2]. Environmental scientists, ecologists, land managers, and others are increasingly presented with high-stakes and high-uncertainty problems surrounding these human-ecosystem interactions, sometimes referred to as wicked problems [11]. These wicked problems are exacerbated by factors such as a changing climate, human- and natural-caused land use change, and complex system behavior that introduces feedback loops and unintended outcomes from policy decisions [2]. To address these problems, ecologists, environmental scientists, and environmental managers frequently turn to complex models that produce staggering amounts of data across both spatial and temporal scales. Given the societal importance of effectively utilizing these new data sources, new tools and approaches are needed for scientists to explore, understand, and analyze this data, and, perhaps most critically, communicate complex information to non-scientist audiences. Bridging the gap between scientific knowledge and policy action plays a critical role in promoting the incorporation of scientific information into the policy process as this lowers information barriers, promotes interactions and fosters feedback among different communities [6]. To help bridge this information gap, we consider the role of scientific visualization in addressing wicked problems.

Visualizations and data analytics are intrinsically linked, with new visualization approaches taking on important roles in understanding big data, and recent examples including applications in digital governance [7, 14, 15]. For addressing wicked problems with environmental planning and policy components, there is a clear need for scientific visualization tools that efficiently and effectively display large amounts of data across space and time while also meeting the domain-specific needs of scientific users [16]. Designing visualization software to meet these needs, however, presents significant challenges for single discipline experts (e.g. software engineers, computer scientists, ecologists, hydrologists, etc.). An approach for overcoming these challenges is integrating knowledge from multiple discipline- and non-discipline experts in the software development and visualization process, in an iterative and collaboratively-driven effort referred to as co-production of knowledge [8].

In this paper, we explore the application of co-production to the development of software tools for the visualization of environmental and ecological data that span both space and time. The case study reported here considers how emerging methods in co-production can be applied to the development of data visualization software. Additionally, we consider how these resulting visualizations can be used to tackle specific problems associated with understanding large, complex, geo-spatial datasets, with the ultimate goal of informing policy decisions. Rather than presenting technological advancements of our project, we focus on the process by which these technological advancements were made, and continue to be made, through a novel co-production approach involving computer scientists, ecologists, environmental scientists, climate modelers, and social scientists.

1.1 Knowledge co-production

In our project we apply co-production, a framework that involves a collaboration between experts and stakeholders with the intent of generating knowledge [8]. Co-production approaches have been applied across a variety of problem domains and institutional settings, with co-production efforts frequently bringing together domain experts and stakeholders, with actors that include, but are not limited to, scientists, elected officials, bureaucrats, managers, homeowners, and recreational users. Recent examples of co-production efforts include applications to planning for water management projects [3], ecosystem service management and disaster risk reduction [10], coastal management of fisheries [12], conservation planning [9] and usable climate science for decision-making [13].

A co-production approach typically involves multiple iterative interactions between experts and stakeholders where they work together to define research objectives, make decisions about data and methodological design, interpret results, and apply findings. Notably, when stakeholders are involved in the co-production process, the information produced is perceived as more legitimate, and, therefore, is more likely to be applied in decision-making [1]. Additionally, those involved in the co-production process likely have more buy-in related to the created information because they understand how the information was produced. We expect that any software, information, and ultimately action resulting from this co-production process will be more useable, and useful, than information produced under other scientific paradigms [4].

In this paper, we report lessons learned from a case study that focuses on a co-produced visualization tool for environmental scientist and ecologist users. In this project, we consider co-production between computer scientists and software engineers project collaborators who are experts in the field of designing software for visualization, and their potential users, environmental scientists. These scientists, who we also call our collaborators in the co-production process, are often engaged in work that requires understanding and interpreting large scale ecological and/or environmental datasets with both geo-spatial and temporal components, where cause and effect relationships are not necessarily proximate in either space or time.

2 CASE STUDY APPROACH

To understand the effects of software co-production on experts and non-domain experts, a multi-phase approach is employed where the perspectives of project collaborators can be elicited at different stages of the co-production process. In the case study research described below, this elicitation occurs in three phases. In the first phase, the pre-period, a baseline is established where collaborator needs and perspectives can be assessed, and general directions and goals for software production is established. In the second phase, the development period, collaborators interact regularly to provide input and guidance about software design that may include interface features, data structures, and visualization options. In the third and final phase, the post-period, the newly developed software is assessed by those involved in the co-production effort and the overall process of co-production is considered by project collaborators. After this three-stage process has ended, a new round of co-production could potentially occur, allowing for software to be iteratively developed across multiple development phases. This three-phase co-production effort is amenable to a pre/post-test case study design [5] where project collaborators are interviewed prior to their participation in the project, observed during the development period, and interviewed after development has concluded. We formally apply this pre/post design in the assessment of the VISTAS (*Visualization of Terrestrial and Aquatic Systems*) software tool.

2.1 VISTAS

Launched in 2011, VISTAS is a National Science Foundation funded project that seeks to enable scientists to better understand and communicate information about complex environmental problems through visualization. VISTAS is a joint effort among scientists at The Evergreen State College, Oregon State University, and Conservation Biology Institute. Prior work on the VISTAS project established a need for software that was capable of quickly processing and displaying large amounts of data [16]. Additionally, environmental scientists and users identified a need to understand how their data, either modeled or remotely sensed, interacted with impacted topography and other landscape features. This lead to the development of the software platform VISTAS that has the capacity to display a wide range of detailed spatially-explicit landscape data.

VISTAS users are environmental scientists explaining results of their science to interested stakeholder groups. For this case study, we recruited collaborators for the VISTAS project to involve themselves in an effort to co-produce visualization software. In this paper, we briefly review our findings from the co-production pre-period and describe how the outcomes were used iteratively to inform the development period starting in February 2017 in which new features and functionality are added to the VISTAS software platform. This development period is on-going and anticipated to conclude in March 2018.

At the onset of the project there were twelve collaborators recruited in the pre-period for interviews including software developers (n=2), environmental scientist users (n=6), and domain experts from related fields (n=4). Pre-period interviews began in November 2016 and concluded in February 2017 (Figure 1). Each interview followed a semi-structured protocol, with interview length ranging from 25-60 minutes. In the baseline, key interview questions asked collaborators about their current use of visualization in their research, challenges associated with producing visualizations, and presentation of visualizations to other scientists, stakeholders, and the public. Data collection in the development period included observation of stakeholder interactions, tracking of software feature

	Pre-period: Nov - Dec, 2016	**Development period:** Feb 2017 - March, 2018	**Post-period:** March-June, 2018
VISTAS software co-development		Vector flows , visual analytics, Python integration	
Collaborator interviews	Prior to development (N=12)		After development (N=??)
Collaborator observations	Attend weekly development meetings, record interactions, collect documentation		

Figure 1: Timeline of VISTAS research activities

development, and recording of correspondence between our collaborators. Data collection in the post-period will commence after the development period has concluded and will consist primarily of post-project interviews with our collaborator group. All interviews are transcribed and coded for key themes that emerged in analysis.

3 FINDINGS

Findings from the study thus far are based on data collected in the pre-period as well as data collected in the development period, which is currently on-going. In the pre-period, all collaborators identified substantial challenges associated with both managing and understanding data, and the general need for visualization tools (n=12); and each domain scientist described ways that visualization tools could address their particular scientific domain problems and user base. While the current use of visualization tools for data exploration varied among the environmental science participants, all participants agreed that some form of visualization was integral to understanding model behavior and validating model outputs. For environmental scientists who studied hydrology, for example, being able to view data outputs on topographic or 3-D landscapes was essential. For environmental scientists engaged in agent-based modeling at landscape scales, the speed at which visualizations could be displayed was crucial and often identified as a bottleneck in workflow. For both sub-domains, viewing landscape data with complex topography over time was also critical.

A common challenge among VISTAS scientist collaborators was their capacity to understand increasingly large, spatially- and time-explicit datasets. Using the visualization techniques currently available to them, many collaborators felt they were missing important patterns or features simply because potentially important information was not considered due to time, computational, and analytical constraints. Using visual analytics to help identify important or interesting information was one proposed approach to this challenge, and was seen as a way to improve model understanding and the speed and scope at which model outputs could be validated. Four collaborators recommended that machine learning techniques could be leveraged to address some of these challenges.

When presenting model results to other scientists, our VISTAS scientist collaborators perceived visualization to be an important step in the communication process, with some describing situations where their scientist-colleagues, in conferences as well as other academic and agency settings, did not have a firm understanding of the information presented until it was visualized. When our scientist collaborators engaged their own stakeholders, they perceived visualization as critical, especially when the information being relayed was at a landscape scale. Our scientist collaborators spoke of how their stakeholders did not have confidence in model results until those results were visualized, with one study collaborator describing how he uses visual aids like aerial photography and Google Maps to orient stakeholders to a study area. Collaborators described how stakeholder trust in environmental model outputs was crucial in establishing stakeholder trust in scientific results and bringing those to bear in land management decisions, with two collaborators using VISTAS-generated videos of model outputs as a way to build confidence in the information presented to stakeholder groups. From the perspective of VISTAS collaborators, their stakeholders had to first *see* how the model predicted impacts on parcels of land they were familiar with before they accepted these model results as useful for describing ecological processes on the landscape.

Having identified the importance of landscape visualizations for understanding model outputs, communicating information, and build legitimacy among stakeholders, all collaborators acknowledged some level of frustration about their current capacity to visualize information. While visualization was acknowledged as critical to the analysis and communication process, there were still significant barriers to the usefulness, and usability, of existing software for the visualization of complex spatial and time series data. Notwithstanding, three scientist collaborators who had experience using an earlier prototype of the co-produced visualization software described how VISTAS had enabled them to overcome some of these issues. These early VISTAS users noted how the software was tailored to meet their specific needs by consuming gridded time series data, quickly drawing 3-D landscapes, and allowing them to easily export visualizations for presentation purposes. While these VISTAS users had experience with other software platforms capable of generating visualizations, that software did not fit their specific needs, particularly in the case of displaying topographical data and understanding complex processes that evolved on the landscape across time. These VISTAS users described how they had trained colleagues within their government agency to use VISTAS, and how this had led to further adoption of the software among other agency scientists. After further development has been undertaken on the VISTAS software, these current VISTAS users plan on distributing VISTAS software to their non-scientist stakeholders, provide training on how to display and interpret model output visualizations, and hopefully facilitate the use of model visualizations as part of management and policy decision-making.

3.1 Feature development from co-production

After careful consideration of VISTAS collaborator feedback and perspectives, new functionality is being added to the VISTAS software platform in the development phase. See Figure 2 for an example of the current user interface for the VISTAS prototype. At the time of writing, VISTAS features are under development, and will continue to be built out as this co-production process continues. When environmental scientists were attempting to understand the impact of point source contaminants, the ability to track the flow

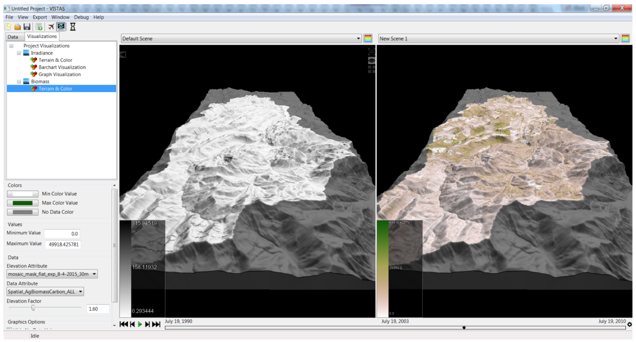

Figure 2: VISTAS software interface example

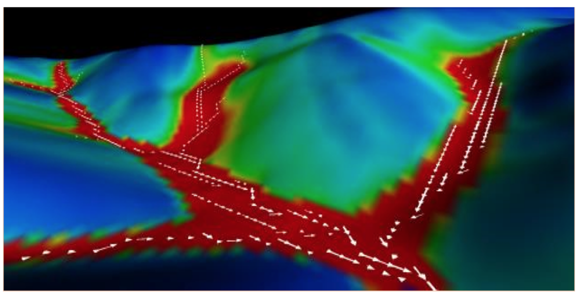

Figure 3: Major watershed flows, visualized in VISTAS

of water on a landscape was identified as an essential feature. The VISTAS development team created a flow vector visualization (Figure 3) to address this request. The impetus for water flow vector visualizations was to better understand model behavior, as well as more effectively communicate how upstream activities can have downstream impacts. This could prove particularly useful when engaging stakeholders and managers who are involved in decision making about land use activities. Another feature identified by VISTAS collaborators was the ability to load, or select, an area of the landscape and summarize cell values across this area. Generating summaries across reporting units and other spatially delineated areas is seen as particularly important for informing the decision-making process because this is the scale at which management decisions are typically made. Finally, current VISTAS collaborators requested the ability to run statistical analyses on the visualized data, and use the output of statistical analysis to help inform the interpretation of the visualized landscape. This feature is currently under development, with a first step in this process including linear regression, with additional statistical analysis to be added at a later time after feedback from users. The request for statistical analysis to accompany the visualization of landscape processes underscores our assessment that visualization is a key component of data analytics effort.

4 DISCUSSION

While these findings reflect data collected in the baseline and development phases of this project, and do not include post-period development, key themes emerge that we feel are significant to the co-production process and of interest to the broader digital government community. These key themes reflect the importance of visualizations for environmental scientists, and articulate substantial challenges associated with understanding and visualizing large datasets in the context of wicked problems. We have robust evidence that among VISTAS collaborators, visualization is an important component of the science production pipeline that spans from problem identification through analysis, results presentation, and legitimization. Visualization, if appropriate to the tasks and practical to use, is critical throughout both analytic and decision processes for data exploration, model validation, and presenting results through multiple formats and media.

It is important to note that while the computer scientists co-design the visualization software with scientist users, it is the users themselves who design visualizations using that software. Since VISTAS collaborators communicated that the scale, scope, and design of each visualization should be particularized to the domain, domain problem, and intended audience, the software is designed to be flexible to account for these differences. VISTAS collaborators were sensitive to such design distinctions and conscious of the needs of users to design visualization to fit the circumstance. This study, however, did not consider how scientists learn to design more useful visualizations using co-produced software, a question that we leave for future inquiry.

There are clear advantages for using a co-production approach for software development, one of which includes a tailoring of visualization software to meet the needs of expert users, thereby improving usefulness and usability for this group. Apart from software technological advancement, there are also impacts on the participants themselves. Scientists and software engineers must communicate on a regular basis, with scientists having to learn, across time, how to communicate their needs to software engineers, and vice versa. In addition to co-production between environmental scientists and software engineers, another consideration with this work is how co-produced software is used to shape interactions between scientists and non-scientist stakeholders. While we did not collect data from these stakeholders for this study, our experience with our scientist collaborators has provided us a window on how co-production activities, like the development of visualization software, can improve efforts that interface directly with other members of the public. For example, one of our VISTAS users frequently interacts with landowner stakeholder groups, providing them information about the environmental impact of land management practices. Data from model outputs, which is complex both spatially and temporally, is displayed using VISTAS for presentation to these stakeholders. Future work will more explicitly consider how the software co-production effort not only produces usable and useful software for environmental scientists and ecologist users, but how VISTAS facilitates interactions of these users with stakeholder groups through creation of visualizations.

When approaching issues involving data sharing and analytics to address societal problems, there are clear advantages to applying a co-production approach, especially in the context of software development. First, if the goal is to provide usable information to the public, and facilitate public-private partnership, then public and private users should be brought together in a process where problem definitions, approaches, and the resultant product can be iteratively co-developed. Second, there is a clearly identified need for better visualization tools, and almost all forms of data, either within or outside the domain of ecology and environmental science, will likely benefit from visualization. While we cannot speak directly to the challenges of visualization of large datasets in other areas, in the case of environmental science these challenges are substantial, and likely become more challenging when both time and space are included dimensions. Third, many of the intractable social problems that benefit from the input of the attentive public are complex in nature, often fitting the mold of wicked problems. Co-production of knowledge is one approach that has been applied in many contexts to address such wicked problems, and, given the experience that we have had working with collaborators confronting these issues, is applicable to data- and information-intensive problem domains. Finally, co-produced software, and potentially the visualizations produced by this software, could be understood as having more legitimacy based on the process by which it was created. For purposes of governance, legitimate information is arguably more actionable, and therefore more likely to be applied in decision-making by policy actors.

This case study laid out a framework for applying a co-production approach to developing software for complex data visualization. We anticipate that this co-production approach, including the pre/post case study research design, can be successfully applied to areas of governance where exploration, understanding, and communication of complex data and information are essential to informing decision-making. While co-production is promoted as way to increase the usability of produced knowledge, an additional outcome of this process may be a wider adoption and dissemination of this produced knowledge, such as data, models, software, and web applications, through the legitimacy and buy-in that co-production can generate. Ultimately, the best designed software and the most information rich datasets will not be applied in the policy process if they are not first adopted by policy actors. We therefore recommend that those in the digital governance community carefully consider whether their governance domains are amenable to collaboration between experts and stakeholders.

ACKNOWLEDGMENTS

This project is support by National Science Foundation grants: 1) XXXXXXXXXXXX and 2) XXXXXXXXXXXXXXX. We also acknowledge our environmental science collaborators and all VISTAS developers at XXXXXXXXX.

REFERENCES

[1] David W Cash, William C Clark, Frank Alcock, Nancy M Dickson, Noelle Eckley, David H Guston, Jill Jäger, and Ronald B Mitchell. 2003. Knowledge systems for sustainable development. *Proceedings of the national academy of sciences* 100, 14 (2003), 8086–8091.

[2] Kathryn Davies, Karen Fisher, Mark Dickson, Simon Thrush, and Richard Le Heron. 2015. Improving ecosystem service frameworks to address wicked problems. *Ecology and Society* 20, 2 (2015).

[3] Jurian Edelenbos, Arwin Van Buuren, Dik Roth, and Madelinde Winnubst. 2017. Stakeholder initiatives in flood risk management: exploring the role and impact of bottom-up initiatives in three âĂŸRoom for the RiverâĂŹprojects in the Netherlands. *Journal of Environmental Planning and Management* 60, 1 (2017), 47–66.

[4] Sheila Jasanoff et al. 2004. *States of knowledge: the co-production of science and the social order.* Routledge.

[5] Jason L Jensen and Robert Rodgers. 2001. Cumulating the intellectual gold of case study research. *Public Administration Review* 61, 2 (2001), 235–246.

[6] Maria Carmen Lemos, Christine J Kirchhoff, and Vijay Ramprasad. 2012. Narrowing the climate information usability gap. *Nature climate change* 2, 11 (2012), 789.

[7] James McCarthy and Jim Thatcher. 2017. Visualizing new political ecologies: a critical data studies analysis of the World BankâĂŹs renewable energy resource mapping initiative. *Geoforum* (2017).

[8] Alison M Meadow, Daniel B Ferguson, Zack Guido, Alexandra Horangic, Gigi Owen, and Tamara Wall. 2015. Moving toward the deliberate coproduction of climate science knowledge. *Weather, Climate, and Society* 7, 2 (2015), 179–191.

[9] Jeanne L Nel, Dirk J Roux, Amanda Driver, Liesl Hill, Ashton C Maherry, Kate Snaddon, Chantel R Petersen, Lindie B Smith-Adao, Heidi Deventer, and Belinda Reyers. 2016. Knowledge co-production and boundary work to promote implementation of conservation plans. *Conservation biology* 30, 1 (2016), 176–188.

[10] Belinda Reyers, Jeanne L Nel, Patrick J OâĂŹFarrell, Nadia Sitas, and Deon C Nel. 2015. Navigating complexity through knowledge coproduction: Mainstreaming ecosystem services into disaster risk reduction. *Proceedings of the National Academy of Sciences* 112, 24 (2015), 7362–7368.

[11] Horst WJ Rittel and Melvin M Webber. 1973. Dilemmas in a general theory of planning. *Policy sciences* 4, 2 (1973), 155–169.

[12] Franke van der Molen, Daniel Puente-Rodríguez, Jac AA Swart, and Henny J van der Windt. 2015. The coproduction of knowledge and policy in coastal governance: integrating mussel fisheries and nature restoration. *Ocean & Coastal Management* 106 (2015), 49–60.

[13] Tamara U Wall, Alison M Meadow, and Alexandra Horganic. 2017. Developing evaluation indicators to improve the process of coproducing usable climate science. *Weather, Climate, and Society* 9, 1 (2017), 95–107.

[14] Lidong Wang, Guanghui Wang, and Cheryl Ann Alexander. 2015. Big data and visualization: methods, challenges and technology progress. *Digital Technologies* 1, 1 (2015), 33–38.

[15] Ben Williamson. 2016. Digital education governance: data visualization, predictive analytics, and âĂŸreal-timeâĂŹpolicy instruments. *Journal of Education Policy* 31, 2 (2016), 123–141.

[16] Kirsten M Winters, Judith B Cushing, and Denise Lach. 2016. Designing visualization software for super-wicked problems. *Information Polity* 21, 4 (2016), 399–409.