The *Smart* Amplified Group Environment

Nurit Kirshenbaum, Dylan Kobayashi, Mahdi Belcaid, Jason Leigh
Laboratory for Advanced Visualization and Applications, University of Hawaii at Manoa, USA
nuritk@hawaii.edu
dylank@hawaii.edu
mahdi@hawaii.edu
leighj@hawaii.edu

Luc Renambot, Andrew Burks, Krishna Bharadwaj, Lance Long, Maxine Brown
Electronic Visualization Laboratory,
University of Illinois at Chicago,
Chicago, IL 60607, USA
renambot@uic.edu
aburks3@uic.edu
krishnaknbharadwaj@gmail.com
llong4@uic.edu
maxine@uic.edu

Jason Haga
National Institute of Advanced Industrial Science and Technology, Japan
jh.haga@aist.go.jp

Chris North
Center for Human-Computer Interaction,
Virginia Tech, USA
north@cs.vt.edu

CHI'20,, April 25–30, 2020, Honolulu, HI, USA
ACM 978-1-4503-6819-3/20/04.
https://doi.org/10.1145/3334480.XXXXXXX

Abstract
SAGE2 is a widely known open-source environment for collaborative immersive analytics on large display walls. This paper briefly outlines our group's approach to evolving the SAGE2 software to support existing and future collaborative scientific exploration workflows through the application of Artificial Intelligence in Immersive Analytics.

Author Keywords
Immersive Analytics, Large Display Walls, SAGE

Introduction
With approximately 4000 users at 800 sites in over 17 countries, SAGE2 - **Scalable** Amplified Group Environment [5] is today's de facto operating system for immersive analytics on large tiled display walls, enabling users to collaborate locally and remotely and access, display, share and manipulate digital media data (images, movies, documents and spreadsheets). SAGE2's ease of use makes it an excellent platform on which to display a variety of related, high-resolution information, which in turn enables collaborators to reach conclusions and make decisions with greater speed, accuracy, comprehensiveness, and confidence.

A 2017 SAGE2 community survey showed that SAGE is popular in many disciplines, including: Archaeology, Architecture, Art, Atmospheric Science, Biology, Chemistry, Civil

Engineering, Communications, Computer Science, Education (preschool through graduate school), Geoscience, Health, Library Science, Medical, Meteorology, Network Engineering, Neuroscience, Physics, Psychology, and Statistics. Moreover, SAGE is used in a variety of ways, including: classroom and distributed education, local and remote presentations, and planned and impromptu meetings.

One of the key advantages to SAGE2 is its JavaScript based implementation, which enables the creation and utilization of web applications from within SAGE. Many existing applications for productivity and research are in use by our community for collaboration (Google Hangout, Slack, Whereby, NoVNC, BlueJeans); for brainstorming, editing, group meetings, research papers and grant writing (Mindmeister, ShareLaTex, Google Docs, Office 365); and, for science (Potree point cloud, VTK.js, ARCGIS, PubMed, and Jupyter). It is therefore a platform particularly suited for integrating current tools with future innovations in the field of immersive analytics.

The Scientific Research Enterprise

We have recently collected usage patterns into a *scientific research, development and training enterprise* [4] that we have observed in SAGE users. Figure 1 illustrates the "kinetics" of such an enterprise, with example snapshots of how SAGE2 is currently used to support the various enterprise phases. Phases include: (C) conceptualization/hypothesis generation, (W) data collection and wrangling, (V) visual analysis, (X) knowledge crystallization, and (P) knowledge presentation. Several popular workflow systems and Gateways (such as Jupyter and Agave) primarily fall under phases W and V. However, research and training frequently transition between all phases, particularly in the early phases as a result of many micro-failures before achieving and reporting success in X and P. We think of our usage patterns as an expansion on Colin Ware's concept of Visualization Thinking Design Patterns (VTDP) [8] to incorporate Design Patterns for collaborative work in large tiled display environments. In the eScience paper we shared several scenarios that use these patterns as a guiding example for other SAGE users who wish to take full advantage of the system in their process.

However, as the enterprise diagram suggests, the activities in each phase have different needs; the icons accompanying the diagram represent group work, individual work, cyber-infrastructure use, visual analysis, interactivity, intensive writing, and intensive coding. Researchers must move seamlessly between the activities without breaking their immersion in the data. Furthermore as the scale and complexity of the data increases it becomes increasingly difficult to continue to manage the content on display walls. Thus far display walls are used largely as passive devices. To help users keep pace with scale and complexity, the display walls need to become active collaborators. This need inspires our concept of **smart end-user services**. With smart services, researchers delegate low-level tasks that are needlessly disruptive to their flow of work to Artificial Intelligence, leaving them to focus on important analysis and sense-making tasks. **The future vision for SAGE (SAGE3 - the *Smart* Amplified Group Environment) is a "human in the loop" coordination platform, with AI-enabled services as their "co-pilot" to autonomously manage and link disparate datasets and orchestrate services, such as connecting to other display walls and computational systems.** The two main ideas of our concept are **SmartBits** (Data building blocks imbued with AI capabilities) and **SmartSpaces** (shareable and customizable AI-enhanced workspaces). This paper introduces these ideas and how they fit within our Human-Centered Scientific Research, Development and Training Enterprise.

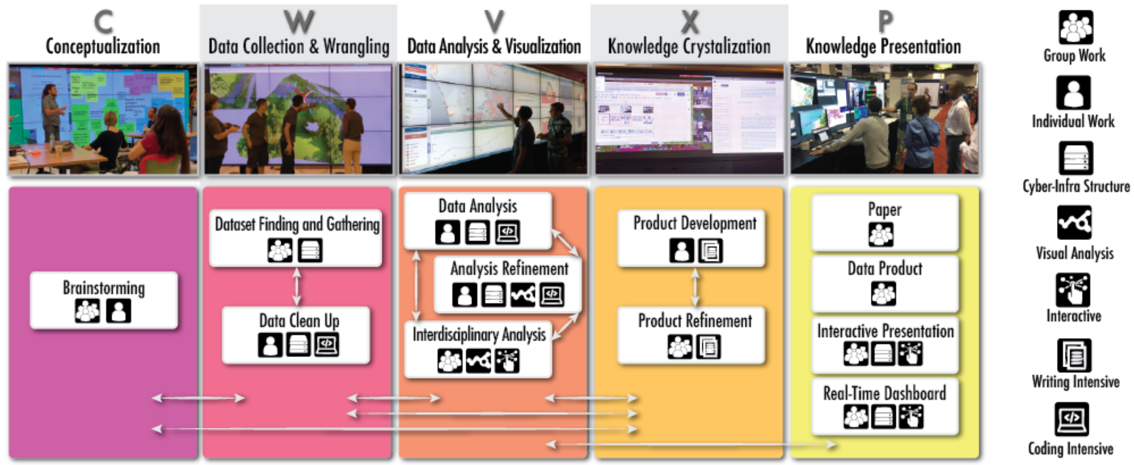

Figure 1: Diagram of the scientific research, development and training enterprise

SmartBits

State-of-the-art work in using AI-based natural language techniques to automatically generate data visualizations showed that when users are relieved of the task of manually operating visualization controls, they are able to produce desired visualizations and derive insights more rapidly [6]. A user may be interested in the identification of trends, seasonality, or anomalous instances in time series data; or detecting communities or nodes of high centrality, etc. SAGE using smart services will have autonomy to analyze datasets in the background and suggest relevant information to users, enhancing their productivity. The concept of autonomous insight is garnering significant interest in the industry, specifically for the analysis of document data, where such frameworks are dubbed "insight engines."

In our vision for a smart SAGE, data will be stored, manipulated and represented through units of information termed SmartBits. SmartBits are first class visualization objects, combining an atomic data-type, rendering information and data-aware code to manipulate the data at various stages of its life cycle. For instance, in the data wrangling stage, a time series SmartBit could fill missing values using different strategies, or resample the data using a different time interval. During the data mining stage, a SmartBit can provide functionality for detecting seasonality of trend, or for mining its data for anomalies. During the preservation phase, a time series dataset can save itself to a database or compress itself to a file. SmartBits lower the barrier to entry into AI for non-expert users, while providing experts with the range of functionality they have come to expect. This will allow scientists to think about solutions at a higher level of abstraction, in terms that are closer to their scientific discipline. SmartBits could leverage a custom data infrastructure using ontology-specific data-indexing strategies and efficient frameworks to query the data; and summarization and thinning strategies to provide statistics on the full datasets or representative subsamples, without compromising the inherent scientific information [2, 7].

Companies are embracing insight engines to interact with users in a more natural fashion, automating generic tasks and learning from context which analytical actions to take. Insight engines, such as Kinetica's, for instance, are currently used in pharmaceutical research, finance, logistics, telecom and healthcare, among other areas. A smart SAGE could adopt a similar paradigm by centralizing its AI in a **Foresight Engine** that governs how SmartBits are synergistically federated to assure utility to the users. The Foresight Engine is a hybrid between AI-based Enterprise Integration Assistants (for example, SnapLogic's Iri) which use machine learning to predict the next step in analytics, and AI-based visualization recommendation systems [1].

By seamlessly interacting with SmartBits to automate repetitive tasks, the Foresight Engine will recommend analyses based on the task's predicted usefulness to the user. Mechanisms to rank usefulness of analysis will be collated based on the idea that steps that were useful with similar datasets, for a similar task, or in a similar context in the past are likely to be useful again. Key to this approach is the human-in-the-loop who will non-intrusively queried for feedback about the utility of the outcomes, creating a mechanism for improving learning.

SmartSpaces

SAGE's user interface must support interactions among many simultaneous users and heterogeneous information. Users and data may be co-located or distributed across multiple remote systems. SmartSpaces will enable user-customizable workspaces that can be packaged and archived for local system recovery or shared with other SAGE walls. SmartSpaces could be shareable across platforms, made portable through containerization, and leverage SmartBits and advanced SAGE backend resources to enable data provenance, recovery, and consistency. Collaborators will be able to seamlessly share and reuse their work between the critical phases to enable a fluid transition between individual and collaborative work. This enables data provenance through automatic workspace checkpointing, allowing recovery and investigation branching from critical data phases. SmartSpace templates (Figure 2) will help teams quickly and intuitively initiate common SAGE usage scenarios in support of the 5 phases of the research, development and training enterprise described earlier.

Groups using the brainstorming template are likely to generate many short, grouped notes. Like the other templates, SmartBits will be utilized for smart service assistance. In this mode, notes will be referenced for content visualizations or related word suggestions among other things. When shifting from brainstorming to dataset finding and gathering, only the key takeaways of the brainstorming session may be important. These key aspects can be summarized by the Foresight Engine which, for example, may be creating a real-time Word Cloud.

The analysis template will make SAGE a first-class collaborator within the domain of collaborative data analysis. From the moment data is deposited onto the wall, the AI provided by SmartBits will perform ongoing analysis to find correlations within and between other data already on the wall. The correlations found may be quickly prototyped into visualizations for rapid hypothesis generation and validation.

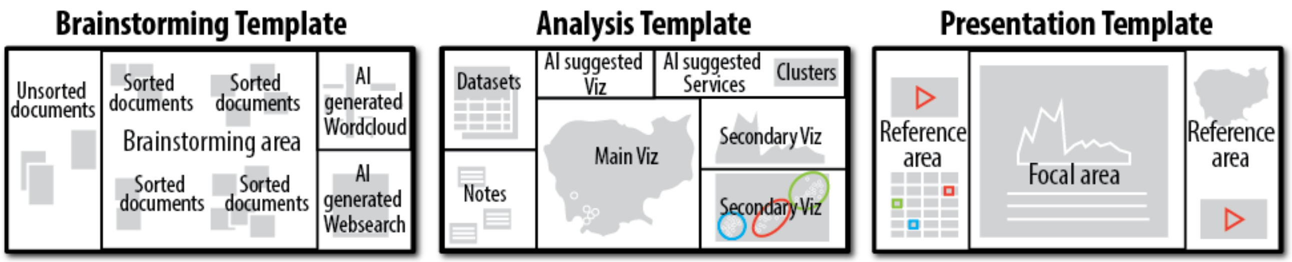

Figure 2: Example SAGE3 SmartSpaces for scenarios such as brainstorming, analysis, and presentation.

The presentation template will leverage the universally familiar conference poster as the paradigm for automatic layout configuration. We recognize the necessity to support a variety of presentation schemes through customization such as live dashboards, interactive posters, and other data products with goals of disseminating research and data exploration results. Throughout the phases of research and development cycle, SmartSpaces will provide a solid foundation of visualizations and data processing building-blocks enabling sharing, reusability, and reproducibility. This foundation will reduce the difficulties involved with summarizing and sharing scientific results while easing the merging of disparate perspectives (human and AI) into one cohesive, data-driven conclusion.

Closing Remarks

By all accounts SAGE would appear to be a success story in the deployment of immersive analytics tools to scientific end users, hence our proposal for SAGE3. However, this success was not achieved easily. Grammel showed that users will typically resort to simple charts in tools like Excel, even though better alternatives exist [3]. We believe the impact of our success could be improved by (1) creating greater trust in the tools and (2) providing more hands-on training with users to apply the tools to their problems.

Trust in the Tools

Trust is not simply about how accurately a tool is able to perform the task. Scientists are busy people so to invest the time to learn a new tool they must trust that the tool will not only provide a substantial benefit but that the tool will continue to be supported. Government funded research rarely provides sufficient resources to sustain efforts for long periods of time, and it is challenging to turn these boutique tools into commercial products at a low cost. Scientists would prefer their software be free but this usually means there is no guarantee the tool will exist into the future.

Hands-on Training

From our experience with SAGE, we realized that there is a tendency for new users to under utilize the capabilities of SAGE, and in many instances users will use it as a presentation device showing one piece of information at a time, whereas so much more can be gained if all the screen space is used. While user studies may be able to prove that users perform better using large immersive displays, users must have a paradigm shift in their approach to using immersive displays, especially when effectively leveraging SAGE across the entire scientific research enterprise. This led us to begin to conduct extensive day-long training programs as part of domain science workshops. Although effective, this is highly labor intensive, hence our desire to come up with automated solutions such as SmartSpaces. Lastly, as part of the training, scientists must also understand that properly crediting the tools they use to create scientific output encourages funding agencies to provide sustained support for these tools.

Acknowledgments

This work is supported by the National Science Foundation award 1441963, 1530873, and the Academy for Creative Media System at the University of Hawaii at Manoa.

REFERENCES

[1] Alex Endert, William Ribarsky, Cagatay Turkay, BL William Wong, Ian Nabney, I Díaz Blanco, and Fabrice Rossi. 2017. The state of the art in integrating machine learning into visual analytics. In *Computer Graphics Forum*, Vol. 36. Wiley Online Library, 458–486.

[2] Dan Feldman, Melanie Schmidt, and Christian Sohler. 2013. Turning big data into tiny data: Constant-size coresets for k-means, pca and projective clustering. In *Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms*. SIAM, 1434–1453.

[3] L. Grammel, M. Tory, and M. Storey. 2010. How Information Visualization Novices Construct Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (Nov 2010), 943–952. DOI: http://dx.doi.org/10.1109/TVCG.2010.164

[4] Jason Leigh, Mahdi Belcaid, Dylan Kobayashi, Nurit Kirshenbaum, Troy Wooton, Alberto Gonzalez, Luc Renambot, Andrew Johnson, Maxine Brown, Andrew Burks, and others. 2019. Usage Patterns of Wideband Display Environments In e-Science Research, Development and Training. (2019).

[5] Thomas Marrinan, Jillian Aurisano, Arthur Nishimoto, Krishna Bharadwaj, Victor Mateevitsi, Luc Renambot, Lance Long, Andrew Johnson, and Jason Leigh. 2014. SAGE2: A new approach for data intensive collaboration using Scalable Resolution Shared Displays. In *10th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing*. IEEE, 177–186.

[6] Yiwen Sun, Jason Leigh, Andrew Johnson, and Dennis Chau. 2009. Articulate: a conversational interface for visual analytics. In *2009 IEEE Symposium on Visual Analytics Science and Technology*. IEEE, 233–234.

[7] Surya T Tokdar and Robert E Kass. 2010. Importance sampling: a review. *Wiley Interdisciplinary Reviews: Computational Statistics* 2, 1 (2010), 54–60.

[8] Colin Ware, William Wright, and Nicholas J. Pioch. 2013. Visual Thinking Design Patterns. In *Proceedings of the 19th International Conference on Distributed Multimedia Systems, DMS 2013, August 8-10, 2013, Holiday Inn, Brighton, UK*. Knowledge Systems Institute, 150–155.