GATEWAYS TO SCIENCE: HARNESSING BIG DATA AND OPEN DATA FOR PRECISION MEDICINE

Sandra Gesing[1], Michael Zentner[2], Maytal Dahan[3], and Katherine Lawrence[4]
[1]*University of Notre Dame, Notre Dame, IN, USA*
[2]*Purdue University, West Lafayette, IN, USA*
[3]*University of Texas, Austin, TX, USA*
[4]*University of Michigan, Ann Harbor, MI, USA*

Sandra Gesing
University of Notre Dame, Notre Dame, IN, USA
sandra.gesing@nd.edu

The importance of Big Data and Open Data to achieve scientific advancements in precision medicine is beyond doubt and evident in many different projects and initiatives such as the Precision Medicine Initiative (All of Us), ICTBioMed, NCIP Hub, 100K Genomics England Project, and PatientsLikeMe.com. Companies from a wide spectrum such as Oracle Health Sciences, Google, and Data4Cure build solutions that help address efficient and secure data sharing with the patient or clinician in mind.

Even with all these advances there are still challenges to address including a recent Precision Medicine World Conference announcement: “We are missing easy-to-use solutions to share patient data.” Science gateways are a solution to fill the gap and help form, by definition, end-to-end solutions that provide intuitive access to advanced resources and allow researchers to focus on tackling today’s challenging science questions instead of becoming acquainted with complex underlying computing and data infrastructures.

Science Gateways have existed for over a decade and a wide variety of frameworks and APIs have been developed to support the efficient creation of science gateways and ease the implementation of connections to distributed data on a large scale. HUBZero®, for example, forms the basis for the NCIP Hub, a resource for collaboration and sharing of data, tools, and standards among the cancer research community. The importance of science gateways has been recognized by NSF by funding the creation of a Science Gateways

Community Institute (SGCI) to serve the community with free resources, services, experts, and ideas for creating and sustaining science gateways.

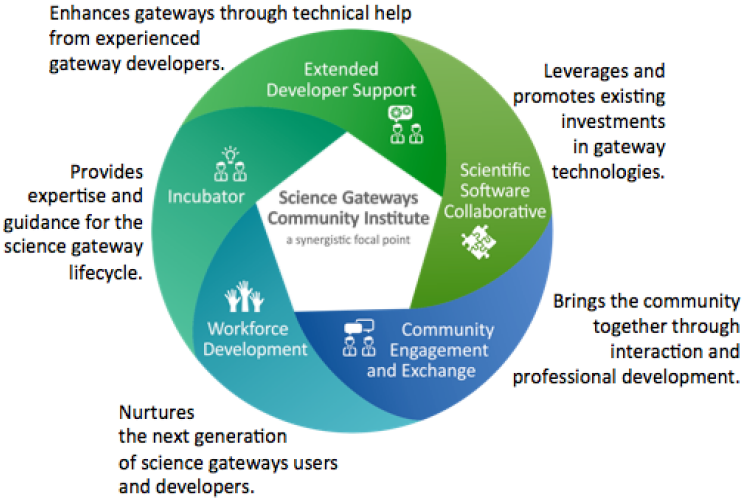

Figure 1: Conceptual overview on the SGCI.

The SGCI is technology-agnostic and serves the community by using technologies and services that are the best fitting solution for the use case. Via this approach, easy-to-use solutions for sharing patient data, analyzing Big Data and/or sharing data for precision medicine can be more efficiently developed or adapted, whenever possible.

References

- Precision Medicine Initiative (All of Us). https://www.nih.gov/AllofUs-research-program/pmi-cohort-program-announces-new-name-all-us-research-program
- Mazurek, C., Pukacki, J., Kosiedowski, M., Trocha, S., Darbari, H., Saxena, A., Joshi, R., Brenner, P., Gesing, S., Nabrzyski, J., Sullivan, M., Dubhashi, D., Thankaswamy, S., and Srivastava, A. Federated Clouds for Biomedical Research: Integrating OpenStack for ICTBioMed. Cloud Networking (CloudNet), 2014 IEEE 3rd International Conference on, pp.294-299, 8-10 Oct. 2014, doi: 10.1109/CloudNet.2014.6969011, 2014.

- NCIP Hub. https://nciphub.org/
- 100K Genomics England Project. https://www.genomicsengland.co.uk/
- PatientsLikeMe. http://patientslikeme.com/
- Precision Medicine World Conference announcement. http://www.pmwcintl.com/big-health-data-sharing-benefits-challenges-and-solutions/
- Gesing, S., Wilkins-Diehr, N., Barker, M. and Pierantoni, G. Science Gateway Workshops 2015 Special Issue Conference Publications. Journal of Grid Computing (2016), doi:10.1007/s10723-016-9389-4
- Michael McLennan (2010), "The Hub Concept for Scientific Collaboration," https://hubzero.org/resources/12.
- Gesing, S., Wilkins-Diehr, N., Dahan, M., Lawrence, K., Zentner, M., Pierce, M., Hayden, L.B., and Marru, S. Science Gateways: The Long Road to the Birth of an Institute. Proc. of HICSS-50 (50th Hawaii International Conference on System Sciences), 4-7 January 2017, Hilton Waikoloa, HI, USA, http://hdl.handle.net/10125/41919