

Roadmap for The integration of biology and earth science data SIG

1. Purpose: The purpose of the Bio/Geo SIG is to foster collaboration and interoperability with researchers and projects outside of the Geosciences directorate with a special emphasis on biological sciences. Special attention will be given to earth sciences research that can be improved with biological data. Outcomes from Bio/Geo SIG efforts include increased communication and collaboration across the earth science and biology domains, new infrastructure that aids access to and integration of data across geo and bio domains, and recommendations for best practices in biology and earth science that facilitate data integration.

2. Communications: The Bio/Geo SIG requires good communication between scientists in both domains as well as social and computer scientists. The SIG currently contains members from all of these communities. The SIG has a presence on the ning site and is open to new members. Use cases are currently being solicited from projects that need biological and earth science data. New members are being solicited. Examples of projects with which the SIG is currently communicating include the Encyclopedia of Life, dbSEABED and iplant. To date, communication has been informal and opportunistic, but there are ongoing efforts to gain funding for workshops and other future engagement activities.

3. Challenges:

The major challenges for the Bio/Geo SIG are

1. Maintaining awareness of the shifting goals and needs of EarthCube and earth science researchers.
2. Gathering requirements and use cases specific to SIG goals
3. Obtaining funds to reach SIG goals
4. Communicating across domains
5. Transferring enthusiasm for the SIG from the initial members to new members and maintaining interest in existing members

4. Requirements: The SIG has already engaged in requirements gathering activities. There was a BOF meeting at the June Charrette, where initial interest was gauged. Conversations have been started with iplant and the Encyclopedia of Life to collect ideas for integration between the biology and earth science domains. The SIG is currently encouraging the development of use cases by SIG members who have specific research questions and by large research projects (iplant) that have biology and earth science data needs.

5. Status: The most common example of integration of biology and earth science data can be found in the form of “this species occurs here” databases. Examples of these are GBIF and OBIS. However, oceanographic databases often handle both physical and biological data. The types of services that GBIF and OBIS can provide include searches by species name and

retrieval of locations and lists of species from a given location. The SIG hopes to expand beyond this type of biology and earth sciences data integration.

6. Solution: The Bio/Geo SIG will identify problems through the submission of use cases. That will inform the SIG on the types of biological data geoscientists need and the types of bio/geo integration questions they are trying to ask. The SIG is targeting existing big data projects in both disciplines and developing a plan or a use case to achieve the desired interoperability. This has been identified as “low hanging fruit”. Additional approaches and their resulting requirements can be investigated as a part of the SIG goal to develop a list of best practices.

7. Process: The SIG will engage in the process of use case and requirements gathering with an emphasis on developing use cases that integrate biology and earth science data.

8. Timeline:

Year 1: Use cases developed and submitted to EarthCube. SIG membership expanded. Face-to-face workshop held. Collaborate with other EarthCube groups.

Year 2: Help members obtain funds to further the data integration mission of the SIG. Continue communication and collaboration mission of the SIG. Work with EarthCube on development and prioritization of the use cases.

Year 3: Develop document on best practices and incentives for data integration between biology and earth science. Continue communication mission of SIG.

9. Management: The Bio/Geo SIG is still a relatively small group that relies on volunteer participation. Most management and governance is handled by the SIG administrator, Anne Thessen, who continues to represent the SIG at EarthCube workshops and alert members of important activities. Moving forward the Bio/Geo SIG needs additional members and an organized communications network. A simple plan is best at this early point, but should consist of:

1. A member network to talk about the SIG at meetings or workshops they may attend and report relevant information back to the group
2. A stakeholder group with members outside EarthCube that have an interest in SIG activities and may become involved in specific development activities
3. An individual responsible for the SIG ning site

The management structure should be reviewed annually and adapted as needed.

10. Risks: The biggest risks to the Bio/Geo SIG are lack of participation and communication. The SIG members are trying to mitigate these risks by obtaining funds for a face-to-face workshop for current and potential SIG members, communicating directly with stakeholders about use case development and talking about the SIG at EarthCube meetings. A unique risk for our SIG is that integration of earth science data with biology data might be seen by the

community as unimportant or at least less important than dealing solely with earth science data. However, many of the science questions discussed at the June Charrette had a strong biology component and the SIG plans to emphasize this point.