

Report 38: Investing in publicly available resources and computational tools for integrating and analyzing environmental health data

Convener: Carolyn Mattingly

Brief History: Across the environmental health community there is a need for centralizing, accessing and analyzing diverse environmental health data through public resources. Currently there are multiple databases being developed by disconnected groups. Data in these resources could be leveraged more effectively through better integration of these resources in combination with feedback from the community regarding research needs.

Discussion Highlights:

- There is a need to integrate existing resources that are relevant to environmental health to better enable searching across data sets more effectively.
- Integration and centralization of environmental data will improve access to relevant information for hypothesis development.
- This integration must be spearheaded by a “leadership” group that combines representatives from technology, biology, the community and NIEHS. This has been done by other scientific areas (e.g., genomics) and we can learn from their example (one example that was discussed was NIF).
- A particular challenge for integration of environmental resources is the diversity of data, variety of endpoints, study structures (e.g., epidemiology, basic science, pathway studies, etc.).

Recommendations:

- Need for information gathering in the form of workshops that address the following aims:
 - We need to gain a better handle on the research priorities that could be enabled by better integration – this requires feedback from the community
 - In parallel, we need to inventory existing resources and associated technologies and vocabularies/ontologies in order to determine how to prioritize integration and streamlining future development based on community feedback (above).
 - Publication of a data resource inventory in a journal relevant to the environmental health community (e.g., NAR database issue)
 - Request for information issued by NIEHS to the community about research questions that require or would be better enabled by databases as well as about resources that they currently use.
- In order to facilitate integration of resources:

- Create a multidisciplinary consortia that may function at a leadership level to drive the implementation of integration of resources, development of needed ontologies, standards for data representation etc.
- NIEHS should invest in providing support for this consortia to be successful (e.g., There is a precedent for this in coordination of NCBI resources and various institutes working together to make genetics and genomic data publicly available).
- A UO1 mechanism might be appropriate for involving members of the community in this integration process.
- NIEHS should also take a lead role in helping communicate the existence of an integrated environmental health resource to the community. Our recommendation is to work with NCBI, which is already the primary resource among biomedical researchers, to integrate EHS resources (there is significant overlap in individual data points such as genes).
- NIEHS should invest in strengthening their internal informatics expertise, which could contribute to the integration of resources and provide insights into needed research initiatives that could benefit from these resources.

Discussion Participants:

Carolyn Mattingly
David Balshaw
Chris Bradfield
Ken Fasman
Julia Gohlke
Heather Henry
Stephanie Holmgren
Robert Kavlock
Antonio Planchart
Kristina Thayer