Data Discovery Index Coordination Consortium (DDICC)

The University of California San Diego

PI: Lucila Ohno-Machado   Grant Number: 1-U24HL126127-01

BioCADDIE: Biomedical and healthCAre Data Discovery and Indexing Engine Center

Biomedical and healthcare data sharing efforts are currently impaired by lack of (1) proper incentives and sharing tools for data producers, (2) practical frameworks for data standardization and indexing of data, and (3) effective data discovery mechanisms. BioCADDIE is a consortium of data producers, curators, publishers, and consumers who will work together to develop practical, sustainable solutions to the problem of biomedical and healthcare data discovery. Through task forces and corresponding pilot projects addressing the barriers enumerated above, we will promote open discussion of why millions of dollars are currently spent in the generation of data that remain captive at their origin or are shared in a sub-optimal way just to comply with mandates from funding agencies and scientific journals. We will promote the development of incentives, policies, and tools for data sharing and data discovery. We will engage researchers, clinicians, patients, and the community in general in an open dialogue focused on pros and cons of biomedical and clinical data sharing. BioCADDIE's specific aims are to: (1) Organize task forces with representatives from communities who have interest in data production, dissemination, and utilization. We will organize an annual symposium, workshops, and internet-based discussions among biomedical and clinical researchers, professional societies, journal publishers, funding agencies, clinicians, patients, and information scientists on best, sustainable practices for making data easily discoverable by different types of users. (2) Promote the development of realistic, minimal, friendly meta-data specifications and annotations for biomedical and healthcare data collections, and corresponding tools for automated indexing so that users will be able to locate data that are relevant to their specific free text searches. (3) Incubate new technologies by funding highly innovative, high-risk pilot research projects that enable the development of novel data discovery and indexing engines and have them tested by our diverse community of stakeholders. We only describe a small number of seed pilot projects in this proposal because BioCADDIE will solicit proposals for new pilot projects every year and select them through a review process involving the various stakeholder communities. PUBLIC HEALTH RELEVANCE: Biomedical research and healthcare data are not fully utilized in part due to lack of incentives and tools to share these data in a way that makes it possible to reproduce results and make new discoveries. We will develop a consortium involving data producers, data disseminators, and data consumers (including patients) to develop tools and processes for easy discovery and access to data.