

[Inspiratron.org - Natural language processing, machine learning and cybersecurity](#)

Open source effort to bring all open biomedical data together

by Nikola Milošević - Saturday, December 12, 2015

<https://inspiratron.org/blog/2015/12/12/open-source-effort-to-bring-all-open-biomedical-data-together/>

Motivation

There is a vast amount of openly available biomedical information on the web in the form of open access journal publications, biomedical, gene or drug databases, drug labels and more. We are talking about millions of full text articles (i.e. indexed in [PMC](#) and possibly some in services such as [Arxiv.org](#)), tens of millions abstracts ([MEDLINE](#)), tens of thousands published drug labels (on [DailyMed](#) and some other drug databases such as [DrugBank](#)), hundreds of thousands clinical trials ([ClinicalTrials.gov](#)) and so on. However, they are all published by certain authority or institution, but they are not integrated. Imagine the data resource where you can get for free all accessible data integrated, text mined, semantically annotated, linked and inferred at one place. Imagine the system where you can get all publically available information about certain drug from drug labels to all clinical trials with extracted drug-drug interactions, adverse events and other things. Also, imagine if it is up to date all the time and all the tools doing mining are open source and available.

I believe this is what health research and health informatics strive towards and what should be an open source community effort. **An integrated data resource built with community created open source tools.**

What needs to be done

There are several things that need to be done. First of all, field experts need to be gathered in single open source organization that can be umbrella organization for the projects that strives towards this goal. It is easy to register organization, however, there are I believe plenty of small bio and health informatics organizations and institutes that are doing some effort, but they are not collaborating enough to make a unique single point of entry data resource and accompanying tools available. Currently we have many organizations, each of them promoting their own standards or schemas, building tools, etc. But centralized effort will be of greater help. Some standards are relatively broad and a lot of researchers apply them, such as publishing resources as [linked data](#) and [OWL](#), but even there, multiple endpoints, schemas and ontologies makes it not as useful as they would be integrated to each other. And here I would not like to talk only about tools and schemas, but to actually build a resource with an endpoint where people can access all biomedical knowledge that is accessible freely.

This definitely is quite a bit task and requires a lot of resources, but these resources will be returned back to the public. In order to host fast service, REST API, website, mail server, and other things needed for such effort there is a need for money as well. Also, people working on integrating data, researching new approaches to extract data from literature, programming and maintaining tools and service needs to live out of something. I would welcome volunteers as well, but in order to be realistic, the most critical people need to be employed and physical infrastructure to be built. Money may come from Grants per projects or by industrial sponsors and supporters. However, they should be aware that all the built resources and tools will be open, free and vendor agnostic.

Projects and local efforts

Since I am part of [OWASP](#) and one of the project and chapter leader there, I would borrow organizational model from them. There has to be a number of different projects inside the organization, which would be led by one or more project leaders. The leaders would need to agree on general resource structure, but different people can work on integrating and maintaining some part of the resource or some tool.

Also, there should be a local chapter or local branches, which would hold a meetings at least 4 times per year and spread the word and awareness about the project, resource or lobby for the good laws and regulations in various countries that could make some anonymized biomedical data accessible through our resource.

What now

I am still doing my PhD at the University and my project is quite related to this topic and could be one of the projects under this umbrella organization. Doing a PhD does not let you to have too much time. However, if there is a critical mass of people who would like to start this type of effort, I would be happy to join or lead the effort. If you are interested, [please let me know](#). If there is no critical mass, maybe it is still early, and this was just one of my shouts and opinion on what needs to be done for better quality of health, research and our community.

All rights reserved and copyrighted by inspiratron.org and Nikola Milosevic