

Something wiki this way comes

Last autumn, Merck & Co. announced the closing of subsidiary Rosetta Inpharmatics in Seattle, Washington (see *Nature* **456**, 26–28; 2008). Last week, Rosetta's founder **Stephen Friend** and scientific director **Eric Schadt** announced a bid to sustain the company's innovative network approach. With logistical support from Merck and US\$5 million in catalyst money from private sources, the duo hope to lay the foundations for a non-profit, open-access research platform called Sage. Its aim: building comprehensive databases that scientists can use to develop more predictive models of disease.

What is the impetus for doing this now?

Stephen Friend: In the past 3–6 months, we have seen examples showing that you can build models of disease that are predictive enough that you want to hang new data on them. Our feeling has been that biology has mostly been archivists building up stacks of data, but not really being able to leverage the data that others have used. The Human Genome Project, for example, has provided linear data that set out the variations, rather than giving us an understanding. We have seen examples generated by Eric and his group, in mice and in humans, that showed it is possible to build frameworks that other people could add data to — and at that point, the scale and scope became very large. And we felt that it was right to go ahead and start that now.

How will Sage be structured?

Eric Schadt: We want this to be open access, and we don't want it to be perceived as owned or dominated by anybody. So I think it is very important that it should be at multiple sites, housed in universities through the incubator phase. We're looking in the Seattle area — it will be the first spot we transition into — where we're still in negotiations with the University of Washington and the Fred Hutchinson Cancer Research Center. And then also our lead potential candidate sites — these haven't been nailed down yet but we are in active discussions — are with Yale University as well as in the Bay Area, at the University of California, San Francisco.

How do you envision the evolution of this open-platform system?

ES: We view the incubation period as being 3–5 years. After that time, this stuff all gets turned loose, in a similar way to Facebook. Facebook started at Harvard University, learned all the necessary rules, developed those tools, and then expanded to other universities, and then to the entire public. That is the kind of thing that, at the end of 3–5 years, we'd be opening up: a truly open-access public platform.



MERCK & CO.

Stephen Friend (left) and Eric Schadt want to launch a predictive open-access medical database.

SF: I think one of the things that has got us excited is that Sage (<http://sagebase.org>) provides us with a mechanism by which important efforts that are already going on in understanding patients — in the National Institutes of Health, in government, in Europe, among companies, among private foundations — all get to come together and aggregate.

What areas of medical research might benefit most from your efforts?

SF: During the incubator phase, we felt as though we have to limit it to two or three disease areas. Because so much of the data generated already is in the area of metabolic diseases, diabetes and obesity, we feel as though that has to be one of the core areas. Another core area that we think is ripe for this type of analysis is oncology. And we're looking to see what other area we might do.

What effect might this have on how scientists approach these disease models?

SF: There is an opportunity to revolutionize how biologists who are in individual labs interface with those who build large data sets. There is a real separation between haves and have-nots that has meant that the people who have intense knowledge of disease biology at a protein–protein, biochemical level, are not

able to interface with those who are building the more genomics-oriented data sets. Our hope is that this can bring together those two groups to work on models in which individual experiments can actually inform the large sets.

ES: This is one of the first efforts to provide these individual researchers with a way to access that scale of information in ways they understand and that affect their research.

Where do you see this effort heading within the next 10 years?

ES: My vision, 5–10 years from now, is of an open-access platform through which research scientists, clinicians and maybe even patients can access petabyte [10^{15} bytes] and maybe even exabyte [10^{18} bytes] scales of data. Where models of disease are actively being used to inform decision-making. And not just where people take, but where they contribute back. So as scientists query their data sets against this platform, they are actively contributing that data to the platform to make it even better. You can think of it as a Wikipedia type of thing where you have this active-contributor network-based approach that just makes the information more and more informative. ■

Interview by Bryn Nelson.