Centers of Excellence for Big Data Computing in Biomedical Sciences (U54) Webinar – Transcript

0:00 Good afternoon and good afternoon all four hundred and eighty-one of you and welcome

0:05 to the meeting. What we're going to do is give a short presentation and then go over

0:13 some of the FAQs and then start taking questions. Remember, please see down on the bottom right

0:18 hand of your screen you'll see the mailbox is BD2KCenterRFA@mail.nih.gov and you can

0:28 also send it right to the chat. Okay so, Big Data has been going for more

0:33 than a year now at NIH and this is the first RFA that came out dealing with centers. This

0:41 is the applicant webinar. We want to remind you that there is a gateway portal at http://bd2k.nih.gov

0:57 So at the beginning, early last year, Dr. Collins had the DIWG Data Informatics working

1:09 group which was advisory to the council. This working group recommended a lot of things that ultimately

1:16 lead to BD2K. The realization of that working group was that we're capable of producing

1:24 large amounts of data, but we're not necessarily capable at NIH of using it. The idea at

1:30 NIH was that we need a sea change. It's got to be a fairly wide sea change, not only

1:40 managing data, but analysis and the management of the programs at NIH.

1:51 Some of..just a minute..i'm just going to hold it for a sec..

2:05 Okay, so some of the main facets are sharing and broad use of data, software and tools that are needed

2:16 to manage and analyze the data, training, and centers. So they're the four core program

2:25 projects that emanated from the DIWG report and have been created under the BD2K. What

2:35 we're talking about today is the fourth bullet point, the centers.

2:41 Going down through a number of key facets. There's a focus on research, near and long

2:46 -term needs. Address big data science challenges. Develop and distribute products. In some sense,

2:58 the products of this FOA are approaches, methods, and tools, and we want to open the door for

3:07 a wide variety of multidisciplinary individuals who can work on this.

3:14 What kind of data types are involved? Imaging, phenotype, we call them the usual suspects.

3:21 Although maybe combining them is certainly something that's become more important recently.
3:27 But also data that's generated for other purposes e.g. social media, mobile is becoming
3:33 important now. And one of the facets that's important to NIH, kind of like our definition
3:41 of Big Data, is combination of data from large coordinated projects but also the accumulation
3:46 of called small data bundles from projects that collectively add value. This last bullet
3:56 here, open science, open data, there's a fairly strong determination from Francis Collins
4:01 down to change the culture to be more collaborative, more open, and we're hoping that BD2K is
4:10 a machine for achieving that. Some of the focus areas. There are four focus
4:17 areas for the FOA: Collaborative environments and technologies, data integration, analysis
4:25 and modeling, and computer science and statistical approaches. Some of the other details of the
4:36 FOA: It's a multi-component U54. This is a relatively new kind of mechanism. I see
4:42 some of the questions I got from a number of you not that familiar with it so we'll
4:46 deal with that in a moment. There's an overall component—component is just like the old
4:52 project within a large center. So there's overall, which is like a gestalt overall,
5:00 an introduction to the program, then there's the data science research which is the main
5:05 body of the research. There's a training component, there's an administration, and
5:10 then there's the center consortium activities. Okay, so first kind of FAQish kind of thing.
5:17 No work necessarily gets described by the overall, but gets performed under the overall.
5:23 It's more or less describing the whole project, so you've got an org chart, you've got
5:27 a philosophy statement about what's going on. But the actual work, the actual people
5:33 goes down in the data science research, training, and center consortium activities.
5:45 Under the component called DSR there are key elements that we really want you to hear.
5:50 And I'm just going to summarize them here: scientific validation; reuse; generalizability;
5:58 the use of modern software development practices that not only help dissemination but helps
6:03 build good software, and I think the community has really made great steps and this is understood
6:09 broadly, and the reviewers will be looking at what you say; user feedback on tool development;
6:16 self-evaluation. Lastly, participation in standards efforts.

6:22 It's a cooperative agreement, which means NIH staff are more substantially involved

6:31 than usual. This round upcoming is 24 million which is going to be about 6 to 8 awards,

6:37 2 million dollars DC for four years. The important dates: letters of intent coming up October

6:45 20th, receipt November 20th, review next spring, and council May, followed by awards in the early summer.

7:02 Now what I'll do is, Jack, can you help me show the FAQs? What we'll do now is just

7:09 go through some of the FAQs that we've posted. Maybe we can head off some questions by briefly

7:14 summarizing the FAQs. We'll run this for about an hour or so with questions coming

7:26 in through the mailbox BD2Kcenter@mail.nih.gov and then if there is enough time, we'll

7:33 open up the line so it will be more of an interactive session. We didn't want to do

7:39 that now with so many people, we may have more than the limit. Yes, we've reached

7:49 the limit.