

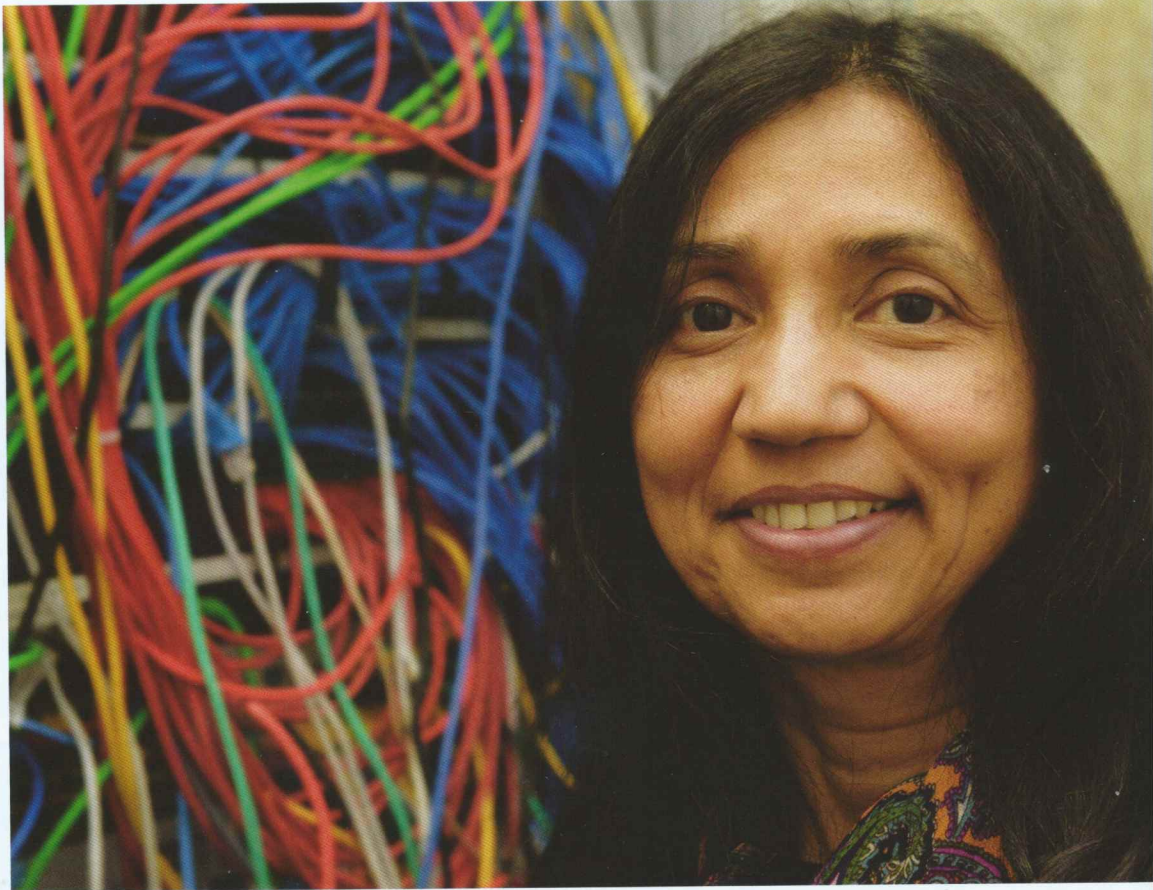
USING BIG DATA TO BREED PERFECT POTATOES

- Laxmi Parida from IBM Research gave a talk on Foundation Day
- Joint plan for Watson Potato computer program

The university celebrated its birthday again on 9 March. This year's Foundation Day talk was given by Laxmi Parida, head of the Computational Genomics Group at IBM Research. She explained how smart methods for linking large numbers of databases can help us treat cancer more effectively or grow better potatoes. Wageningen UR wants to collaborate with her institute in the area of precision agriculture and breeding.

Laxmi Parida works at IBM Research in the US on how to combine large amounts of data to answer specific questions. To do that, they use the Watson computer program. The research institute has developed IBM Cancer, for example. Together with five academic hospitals in the US, the institute has collated lots of information about cancer from hospitals, the scientific literature and social media. IBM is able to use this to make an analysis of the syndromes. Furthermore, doctors can enter their patients' complaints into the application and it then produces a list of the conditions that the patient is most likely to have. That helps the doctors make the right diagnosis more quickly and draw up an appropriate treatment plan.

Richard Visser, head of Wageningen UR Plant Breeding, visited IBM Research last year with his Wageningen colleagues Thomas Been and Richard Finkers. Wageningen UR will be collaborating with the American institute to develop a Watson Potato, a computer program containing all the scientific and practical information on potatoes. Visser: 'We have an aw-



ful lot of information and datasets on potatoes and they know how to link those databases.'

Parida says that using large amounts of data requires in-depth knowledge of data management and integration. 'The computer is unable to see a lot of the data so you have to build computers that function in the same way as the human brain and are able to understand and interpret unstructured data. This is known as cognitive computing. For this, you need language processing and learning methods for the computer. We also use topographical data analysis to detect hidden patterns in the data.'

The idea is that the cooperation with IBM Research will speed up the breeding process for Wageningen's potato researchers and make it more targeted. Visser: 'We

know all the genes in the potato genome but we don't yet know what they do exactly. Let's say we want to develop a potato that's resistant to drought. There could be 500 genes that are correlated with drought resistance but which are the most important ones? By linking datasets, Watson can show which ten genes are most promising for further research and possibly even for trying out in breeding programmes.'

Visser sees many applications for the computer program in Wageningen. 'In biology you have a lot of unstructured datasets that you have to combine to answer complex questions. Think of personal dietary advice and environmental issues. You need a lot of computational power and smart software for that.' Ultimately, the crucial fac-

tor in the use of big data is not so much the amount of data — which is still growing exponentially — as asking the right questions, says Visser.

Parida wants to work with Wageningen UR because it is the global leader in the field of agriculture and environmental sciences. 'To develop computer programs that can reason, understand and learn how to solve complex problems, you need the combined know-how of IT experts and specialists in the subject matter. Those two groups need to constantly work together on software programs such as Watson to 'learn' how it should solve that problem.' That makes Data Sciences a new scientific discipline, says Parida, one that involves researchers and companies from different fields. **AS**