

Noshir Contractor: Welcome to this episode of untangling the web, a podcast of the Web Science Trust. I'm Noshir Contractor and I will be your host today.

On this podcast we bring important leaders to explore how the web is shaping society and how society, in turn, is shaping the web.

Today I'm speaking with Professor James Handler. Jim is the director of the Institute for Data Exploration and Applications, IDEA for short.

He is also the Tetherless World Professor of computer, web, and cognitive sciences at Rensselaer Polytechnic Institute (RPI) in the United States.

He's acting director of the RPI IBM artificial intelligence research collaboration and serves as a member of the board of the UK's charitable Web Science Trust.

Hendler is a man of many accomplishments. He's a fellow of the American Association for artificial intelligence, the British Computer Society, the Institute for Electrical and Electronic Engineers, The American Association for the Advancement of Science, the Association of Computing Machinery, and the National Academy of Public Administration.

Jim might well be the only individual who is a fellow of all of these professional associations. On a lighter note, in 2010, Jim Hendler was named one of the 20 most innovative professors in America by Playboy magazine.

Besides being involved in the start of the web, Jim was one of the pioneers of the interdisciplinary we call web science. Today, I talked with him about the origins of web science, how it has evolved over the years, and its relevance during the COVID 19 era.

Noshir Contractor: Jim, Welcome to this podcast.

Jim Hendler: Thanks. Not sure.

Noshir Contractor: I wanted to start by giving you an opportunity to share with our listeners what is meant by web science. What does that term mean to you and how did it get started?

Jim Hendler: Sure, great question.

So, you know, the web was invented. And there are different dates you could use. 89 is where Tim Berners-Lee actually wrote the proposal for what became known as the web.

By 90 and 91 there was code that was being shared.

Really around 95, 96 he started see the take off of this and people becoming more and more aware that the web was there and that things were happening.

But it really started have much bigger impact by really it was the late 90s where you started to have search engines, you started to have monetization of things on the web. You started to have the social networks growing better now.

So, again, a lot of that was happening all around the same time from the 90s to, you know, the early 2000s and at some point, those of us who had been involved in the web and the web architecture, we're starting to feel like understanding this thing was breaking up into many different pieces - you would go to one kind of conference and hear a lot of discussion about the mathematical underpinnings of networks and networks science, but some of that was really not about the web. The web was just one example of something much bigger.

And then you would go to another meeting and it might be, you know, the social impact of the web or legal aspects, because we were starting to see some of the early days of people beginning to worry about privacy and security on the web, things that are now much bigger issues. And then, there was sort of the engineering of the web. How do we build it? How do we do a better job of knowing that if we deploy a certain technology, you know, how will it get us involved, what impacts might have.

So, some of us started to believe that the web itself had become something we needed to understand. The web sits on top of a larger system known as the Internet, which has a lot of mathematical properties of its networking and, you know, a lot of the routing and things like that that gets talked about happened.

But the web thing was really sitting on top of that, and was its own entity that needed its own understanding.

And there were principles of how you design things that were standards groups, but there really wasn't a lot of research going into the interaction between all these different pieces.

So some of us started to feel like maybe there was kind of a systems science to really understand the web in all its forms.

So around 2005, a group of us and Wendy Hall was one of the organizers. I was one of the organizers. Nigel Shadbolt was there. Danny Weitzner I believe was there and some other people. And a lot of other computer scientists and a few social scientists who were really trying.

It was an invite only workshop about 30 people, held in London, sponsored by the British Computer Society. And the goal was to say, you know, what did we really need to understand to understand the web? So a report was generated by that coming out of that workshop.

The report really got boiled down to a couple pages that got accepted as a perspectives piece by science magazine. So what most people call the start of web science was an article called "Creating a science of the world wide web."

And it didn't use the term web science in that article. It's just that as people started to refer to the thing we had called for, which was something that would put the math, the engineering and the social on the same page, and get those different communities talking and working together. That's when the term web science started to be used.

Noshir Contractor: And that article didn't appear until August of 2006.

Jim Hendler: Yes.

Noshir Contractor: And so really, would you consider that as one of the dates that is most associated with, not the start of the web, but the start of web science?

Jim Hendler: Yeah so 2006 of that article, you know, people tend to try to find some definitive thing. That's the start of something like this. So obviously, a lot of us were talking about stuff that now might be called Web science, but the term itself really grew out of that 2006 article and the first web Science Conference held in Athens was in 2007 based in part by the community started by in part by the community that come together around then.

Noshir Contractor: That you mentioned that as we got started on this. You were focused on a group of people, some of who were invited only at this event, and then subsequently unlike some other scientific interdisciplines, the web science decided to form a Web Science Trust. Can you tell us a little bit about why - what was the thinking behind the creation of the Web Science Trust, as opposed to, say, a learning society or some other division within another context?

Jim Hendler: Yeah, so it's a good question and you know it's a combination of design and accident, but what actually happened was the two e two leading institutions that had really been trying to create web science, sort of institutionalized at that point where MIT and Southampton University.

They created a joint statement to create something called the web Science Research Institute WSRI and very quickly, a few other organizations joined the University of Maryland where I was at that time was so, so I became the third school and there were basically five of us who were kind of leading things at that time.

As it started to grow. We started to create a network of laboratories, including your lab and others, and realized we needed something a more formal way of people to interact because

something that was inherently interdisciplinary has a tendency to coalesce around one of the disciplines, just you know that's historically what's happened.

And it becomes less and less interdisciplinary as it becomes its own discipline and we really felt that was the wrong thing that this needed to be, you know, I used the analogy, sometimes of climate science.

Right when you're studying the climate, you need geologists and you need, you know, people who study the atmosphere and need these people to study the ocean and you need... but not everybody who studies the ocean is looking at climate, not everybody who studies the atmosphere is looking at climate and so it was the same thing with the web. We had people who are studying networks, but only some of them who really cared about the web. We had people who were looking at social impacts of growing communication networks, but some of them in particular we're looking at the World Wide Web.

Now, in the past, you know, 15 years as we've grown the definitions of have slid a little bit about what is exactly the web and where the boundaries and things like that. But the Trust has really tried to be an entity that would help promote web science, help keep this network of labs going, helps make sure there was a conference and eventually a journal. So some of the things I've learners society does, but without really trying to create the way to learner society and be the kind of disciplinarity that tends to come with it

So, again, web science tries to both be an entity that brings people together, but also an entity that doesn't pull people out of the other things they do. So you know, some of them studies network and network science can also be a web scientist without there being any tension between them.

Noshir Contractor: It's fascinating because it takes on the role of both helping to build a community, but also to curate that community intellectually and one of the challenges that I imagine you might have faced is the difference, if any, between those who think that what they are doing is internet science versus web science. Do you have any thoughts on that distinction?

Jim Hendler: You know there's there's been a lot of different terms that have a lot of overlap. So information science in the US was taking off around the same time and a lot of people were arguing that web science belong in information science. Other people were saying no, because information science really is sort of some schools doesn't really include the computational or mathematical side of things. So, so, you know, same thing with internet science. There was a feeling that web might be too limited of a term. And frankly, I would love to see so called internet science and so called web science come further together.

But by in large, the desire to keep the social science piece wedded to the math and the engineering side has been tended to differentiate the websites approach brothers. That's not to

say no one else wants to do it, but I think the dedication to that as a sort of core value that we're trying to bring people together across these different ways of looking at the web. And nowadays you know the mobile web. The, the big companies, the things that look at information and challenges that information, you know, again, they happen at a lot of different places and web science really would like to be an integral place that brings these people together.

Noshir Contractor: And I imagine that some folks also will be questioning whether the study that doesn't happen, specifically on the web, but things are sort of migration into apps and different kinds of ways in which we are navigating this new world should or should not be included as spider web science?

Jim Hendler: You know, there's always a tension in those kinds of directions for any interdisciplinary science, but I think that the goal has been open. The definition at a technical level of what the web is actually very different than what a lot of people think of. So a lot of people when they're opening something on their browser. I'm sorry, opening something on their phone, not on their browser.

And looking at how you know the apps world aren't realizing that it's sitting on top of the web architecturally. So you know some of us like to think of it, as you know, I hate the word ecosystem, but I don't have a better word.

That's evolving and so you have parts of the web that moves one way and part moves the other. And again, part of what web science wants to do is not reject any of those parts and say, really, if we're going to understand this thing we have to understand it as a system or systems as systems. We have to understand the interacting pieces, whether they are considered by a particular practitioner to be "web: or not.

You know, there's a lot of overlap with work that's done at the world wide web conference. But for example, some of the work there, really, is not particularly seen as web science per se because it's technologies to enhance web products more than technologies that are really understanding the interactions happening on this huge network of information.

Noshir Contractor: Great. So 2006 when you wrote that article quarter that article and laid out some priorities. To what extent were those priorities focused more on seeing the web as an opportunity or web science as something that explores new opportunities versus focusing attention on potential concerns as the web became more and more prevalent. And while you think that, are there some areas where you feel that web sciences made the most progress, while others where you would like to see a lot more progress being made?

Jim Hendler: Sure, so you know, in that document and then not long after that we had something we produced we called it a manifesto, which may or may not be the right name.

Something became a book on web science or a publication that you know later you joined for the second edition.

But we were really looking at trying to get thematically what this thing was what was happening, how it worked and and we always wanted to express both the positive and the negative.

Again as a science, you have to be looking at what's happening. But again, part of part of what makes web science somewhat unique is the goal to bring together the people who are building it and can put in mechanisms to try to solve some of these problems with the people who may be studying the problems or the opportunities, trying to understand why does some things on the web, take off and "go viral".

And here I don't necessarily mean just like a video or something. But the whole use of the web for video, the whole how does that change the world when you know people couldn't can video chat, rather than talking, in person or not, you know, again, how does that change from the phone network to the web when you have a web of information.

Not, you know, how to search work. But what a search does, what's the impact of being able to find this kind of information.

As crowdsourcing Wikipedia, things like that grown, that's become one of the things we've been trying to study and understand and that includes misinformation as well as information, right. So if you look at sort of the papers that have been presented at web science. In fact, some of the early best papers. One of them was how trolling on the web, right, influenced an election. This was actually the, Congress person from Boston from Massachusetts and it was interesting that that was one of the earliest studies of trolling and misinformation in an election long before it became part of the national election in the US and Brexit and things like that. So again, web scientists were really looking at these impacts in a very deep way.

Noshir Contractor: Yes because the idea of, for example, seeing the extent to which search became so important in the age of what was happening at the time.

I remember a remark made by someone that said that before search became a thing that the World Wide Web was like a library. With all the books strewn all over the floor and no easy way to do look for the right book and to be able to get to it. And I think, search was an example of something that really helped address an early challenge that was faced by the growth of information on that World Wide Web.

Jim Hendler: Right, to give you the counter example because, of course, it's going to be a science, no matter what someone says, someone's gonna disagree.

But search also at but as search engines took over. It became harder and harder to find the opposing opinion. Right. It's hard to go to some of the major search engines now and say I'm searching for this but show me something very different. Show me... You know somebody who disagrees with this approach that.

So if you say this is the what most people believe. And I actually think, you know, and you've studied some of this that has that's part of what creates information bubbles, because then the people who don't believe some piece of that go off and create their own search structures and their own ways of doing things. So again, a lot of different ways of looking at how this plays out. And, you know, again, it's so we used to say, you know, the sort of the metaphor of surfing the web had a connotation of a little bit of danger and serendipity. You might not end up where you were looking to get to search make some of that different.

And so people have been looking at how do we reimpose creativity and new kinds or search, how do we look at argumentation. So, you know, some of the exciting stuff happening in web science nowadays looks at some of the impact of these technologies becoming centralized and says, can we re-decentralized, can we find a way to put it back into the you know, away from the everything owned by a few big companies and much more back to the hands of the users that remains the tension we look at today and things like privacy privacy preserving technologies and things like that, which I'm sure will be topics for later podcasts.

Noshir Contractor: Absolutely. I think that I have a couple of closing questions and I'm going to wait and ask you a closing question on covid but fast forwarding to 2020 or 2019 domain, what are the areas where you feel that web science has made the most progress and what are areas where you don't see as much progress or you would like to see more progress?

Jim Hendler: So I think where web science has made a lot of progress is helping to focus attention on I say really two different things.

Several things have been really impacted by web science One is transparency and the whole open data movement.

So that was coincidence, with the growth of web science that was in part because a lot of the leading people in web science, including Tim Berners-Lee himself, were very involved in helping to try to get governments to open their data to make it more available to develop some of those technologies.

I think also predicting some of the dangers.

So early web science papers were already saying, you know, let's look at privacy. Let's look at security as companies on the web, grow, they're going to be able to see our information to share it to track it you know as cookies came along.

So, you know, I find when I say things that a web Science Conference that really bother some, you know, if I tell people in a normal setting that when you go on the web and you look at the price of something on a particular website, it may be different than, you know, someone else looking at it.

Because they're using information about you to try to adjust the price, people are surprised web scientists aren't surprised.

We're exploring it we want to understand both what are the algorithms that are being used, but also people who believe that that is problematic, how we might control it, things like that. So I think, you know, we, it's more than we're embracing exploring the problems. But of course, the web is a very fast moving thing.

I think the whole mobile web and app space that you talked about, you know, many of us view it still from a web development platform. Well, other, newer people come into web science are beginning to really look at those apps themselves but then you start getting into these boundary issues, right, if somebody has studied a particular Twitter phenomena is it or isn't it and you know what we try to do is be very embracing right if the work is important and talks to it.

Jim Hendler: That's good. On the other hand, if it's just a pure mathematical analysis of something happening in some different network. Then the paper that shows why that applies the web is going to be much more interesting than the paper that just says all networks have some feature.

So again, the boundaries are very hard to see, but I think that, again, the challenges of the web were something we embraced very early and are still looking at.

I think the opportunities become more apparent to people just as more people, you know, it's just part of our lives.

I think the social impacts or something. So we keep trying harder and harder, bring in more social scientists

Particularly people who really can talk to the qualitative meets the quantitative to really try to, again, look at that triumvirate of the underlying math of what's going on the engineering and building this thing because the web's not a natural phenomena and the social impacts and policy impacts of that.

148

00:23:33.000 --> 00:23:35.340

Noshir Contractor: One last question that we plan to ask everyone was appearing on the podcast during this pandemic time is by what is one thing that you personally believe by which

the web is or was it could have been a real help during covid and or one thing I think the web has hurt society during the covid crisis.

Jim Hendler: Great question. So you know, so you and I are sitting here on opposite ends of the zoom channel and we could be on any of five or six of its competitors talking.

More people are working from home. Imagine the lockdowns that we had around covid and entire countries entire cities without the communication infrastructure.

And the communication infrastructure that provides the bits to move between things is the internet. And so, you know, it's sort of hard to pull that apart. But the thing that lets people really interact at the information level that includes find each other for these videos that includes you know when I clicked on a link to open this zoom chat that clicking on the link and how that happens.

And the protocols that made that happen. That's all web. So the web itself has really been absolutely instrumental in allowing communication to happen.

Jim Hendler: You know, big international conferences that were canceled at the last minute were held online

For this year, again virtually, and that virtual conference is made possible by the very technologies, we've studied

Where the negative comes in as the negatives of the web to get amplified and, you know, cultural differences, things like that impact, but, certainly in the States, we're seeing the astounding growth of misinformation. The weaponization for politics of misinformation about covid. Significant amount of them are, you know, bots and trolls.

In fact, the largest network of pro-covid call it an anti-covid, whatever that means you cover your face or not those sorts of thing both have the same origin from the same trolling point which appears to be mostly trying to push divisiveness rather than push a particular point of view. So again, understanding how that works.

Understanding the math of that and being able to show people that will allow both people to understand what's happening, I hope. But also, you know, engineers to understand what we might do to improve that situation. And what we can't do.

Noshir Contractor: Thank you again, Jim, for taking time to talk with us about the history of web science and how it all got started because you have a very privileged position in order to share those insights with us since you were there at the time, it actually happened. And we're also very involved in making it happen.

Jim Hendler: Thank you.

Noshir Contractor: Untangling the Web is a production of the Web Science Trust. Thanks to Carmen Chan for editing and technical assistance. I am Noshir Contractor. Thanks for listening.