

Video Transcript Lexxon 2.2 Tim Berners Lee: The next web of open linked data

It's actually almost 20 years ago when I wanted to reframe the way we use information, the way we work together. I invented the World Wide Web. Now, 20 years on, Ted, I want to ask your help in a new reframing.

So going back to 1989, I wrote a memo suggesting a global hypertext system. Nobody really did anything with it very much. But 18 months later, this is how innovation happens.

18 months later, my boss said I could do it on the side as a sort of a play project. Kick the tires of a new computer we got. And so he gave me the time to code it up.

So I basically roughed out what HTML should look like, hypertext protocol, HTTP, the idea of URLs, these names for things, which started with HTTP. I wrote the code and put it out there. Why did I do it? Well, it was basically frustration.

I was frustrated. I was working as a software engineer in this huge, very exciting lab. Lots of people coming from all over the world.

They brought all sorts of different computers with them. They had all sorts of different data formats, all sorts of kinds of documentation systems. So in all that diversity, if I wanted to figure out how to build something out of a bit of this and a bit of this, everything I looked into, I had to connect to some new machine.

I had to learn to run some new program. I would find the data maybe, the information I wanted, in some new data format, and these were all incompatible. It was just very frustrating.

The frustration was that there's all this unlocked potential. In fact, on all these disks, there were documents. So if you just imagine them all being part of some big virtual documentation system in the sky, then, say, on the Internet, then life would be so much easier.

Well, once you've had an idea like that, it kind of gets under your skin, and even if people don't read your memo, actually, he did. It was found after he died, his copy, it was found, he'd written vague but exciting in pencil on the corner. But in general, it was really difficult to explain what the web was like.

It's difficult to explain to people now that it was difficult then. But then, okay, when TED started, there was no web. So things like click didn't have the same meaning.

I could show somebody a piece of hypertext, a page which has got links, and we click on a link, and bingo, there will be another hypertext page. Not impressive. We've seen that.

We've got things on hypertext on CD-ROMs. What was difficult was to get them to imagine. So imagine that that link could have gone to virtually any document you could imagine.

That is the leap that was very difficult for people to make. Well, some people did. So yeah, it was difficult to explain, but there was a grassroots movement.

And that is what made it most fun. That was the most exciting thing. Not the technology, not the things people have done with it, but actually the community, the spirit of all these people getting together, sending emails.

That's what it was like then. Do you know what? It's funny, but right now, it's kind of like that again. I asked everybody, more or less, to put their documents, I said, could you put your documents on this web thing? And you did.

Thanks. It's been a blast, hasn't it? I mean, it's been quite interesting because we found out that the things that happened with the web really sort of blow us away. They're much more than we'd originally imagined when we put together the initial website that we started off with.

Now, I want you to put your data on the web. Turns out that there is still huge unlocked potential. There is still a huge frustration that people have because we haven't got data on the web as data.

What do you mean, data? What's the difference? Documents, data? Well, documents you read, okay? More or less, you can read them, you can follow links from them, and that's it. Data, you can do all kinds of stuff with a computer. Who was here or had otherwise seen Hans Rosling's talk? When Hans Rosling was at TED, we had one of the great, yes, a lot of people have seen it because it was one of the great TED talks.

Hans put up this presentation in which he showed for various different countries in various different colors, he showed income levels on one axis and he showed infant mortality and he shot this thing animated through time. So he'd taken this data and made a presentation which just shattered a lot of myths that people had about the economics in the developing world. He put up a slide a little bit like this.

It had underground all the data, okay? Data is brown and boxy and boring and that's how we think of it, isn't it, data? Because data, you can't naturally use by itself, but in fact, data drives a huge amount of what happens in our lives and it happens because somebody takes that data and does something with it. In this case, Hans had put the data together. He found it from all kinds of United Nations websites and things.

He put it together, combined it into something more interesting than the original pieces and then he put it into this software which I think his son developed originally and produced this wonderful presentation. And Hans made a point of saying, look, it's really important to have a lot of data and I was happy to see that at the party last night that he was still saying very forcibly it's really important to have a lot of data. So I want us now to think about not just two pieces of data being connected or six like he did, but I want to think about a world where everybody has put data on the web and so virtually everything you can imagine is on the web and I'm calling that linked data.

The technology is linked data and it's extremely simple. If you want to put something on the web, there are three rules. First thing is that those HTTP names, those things that start with HTTP colon, we're using them not just for documents.

Now we're using them for things that the documents are about. We're using them for people. We're using them for places.

We're using them for your products. We're using them for events. All kinds of conceptual things.

They have names now that start with HTTP. Second rule, if I take one of these HTTP names and I look it up, I go and do the web thing with it and I fetch the data using the HTTP protocol from the web, I will get back some data in a standard format, which is kind of useful data that somebody might like to know about that thing, about that event, who's at the event, whatever it is, about that person, where they were born, things like that. So the second rule is I get important information back.

The third rule is that when I get back that information, it's not just got somebody's height and weight and when they were born, it's got relationships. Data is relationships. It's interesting data is relationships.

It's got this person was born in Berlin, Berlin is in Germany, and when it has relationships, whenever it expresses a relationship, then the other thing that it's related to is given one of those names that starts HTTP. So I can go ahead and look that thing up. So I look up a person.

I can look up then the city where they were born, then I can look up the region it's in and the town it's in and the population of it and so on. So I can browse this stuff. So that's it, really.

That is linked data. I wrote an article entitled linked data a couple of years ago and soon after that, things started to happen. The idea of linked data is that we get lots and lots and lots of these boxes that hands had and we get lots and lots and lots of things sprouting.

It's not just a whole lot of other plants. It's not just a root supplying a plant, but for each of those plants, whatever it is, a presentation, an analysis, somebody's looking for patterns in the data, they get to look at all the data and they get to connect it together and the really important thing about data is the more things you have to connect together, the more powerful it is. So linked data.

The meme went out there and pretty soon Chris Bitzer at the university in Berlin who was one of the first people to put interesting things up, he noticed that Wikipedia, you know Wikipedia, the online encyclopedia with lots and lots of interesting documents in it. Well, in those documents, there are little squares, little boxes and in those information boxes, there's data. So he wrote a program to take the data, extract it from Wikipedia and put it into a blob of linked data on the web which he called DBpedia.

DBpedia is represented by the blue blob in the middle of this slide and if you actually go and look up Berlin, you'll find that there are other blobs of data which also have stuff about

Berlin and they're linked together. So if you pull the data from DBpedia about Berlin, you'll end up pulling up these other things as well and the exciting thing is it's starting to grow. This is just the grassroots stuff again, okay? Now let's think about data for a bit.

Data comes in fact in lots and lots of different forms. Think of the diversity of the web. It's a really important thing that the web allows you to put all kinds of data up there.

So it is with data. I could talk about all kinds of data. We could talk about government data, enterprise data is really important.

There's scientific data, there's personal data, there's weather data, there's data about events, there's data about talks and there's news and there's all kinds of stuff. I'm just going to mention a few of them so that you get the idea of the diversity of it so that you also see how much unlocked potential. Let's start with government data.

Barack Obama said in his speech that he, American government data would be available on the internet in accessible formats and I hope that they will put it up as linked data. That's important. Why is it important? Not just for transparency.

Yeah, transparency of government is important but that data, this is the data from all the government departments. Think about how much of that data is about how life is lived in America. It's actually useful.

It's got value. I can use it in my company. I could use it as a kid to do my homework.

So we're talking about making the place, making the world run better by making this data available. In fact, if you're responsible, if you know about some data in government department, often you find that these people, they're very tempted to keep it, Hans calls it database hugging. You hug your database, you don't want to let it go until you've made a beautiful website for it.

Well, I'd like to suggest that rather before you just make a beautiful website, whom I'd suggest they don't make a beautiful website, make a beautiful website but first give us the unadulterated data. We want the data. We want unadulterated data, okay? We have to ask for raw data now and I'm going to ask you to practice that, okay? Can you say raw? Raw.

Raw. Can you say data? Data. Can you say now? Right.

Raw data now. Raw data now. Practice that.

Practice that. It's important because you have no idea the number of excuses people come up with to hang on to their data and not give it to you even though you've paid for it as a taxpayer and it's not just America. It's all over the world and it's not just governments.

Of course, it's enterprises as well. So I'm just going to mention a few other sorts of data. Here we are at TED and all the time we are very conscious of the huge challenges which the human society has right now.

Curing cancer. Understanding the brain for Alzheimer's. Understanding the economy to make it a little bit more stable.

Understanding how the world works. The people who are going to solve those are scientists. They have hard-formed ideas in their head.

They try to communicate those over the web, but a lot of the state of knowledge of the human race at the moment is on databases, often sitting in their computers and actually currently not shared. In fact, I'll just go into one area. If you're looking at Alzheimer's, for example, drug discovery, there is a whole lot of linked data which is just coming out because scientists in that field realize this is a great way of getting out of those silos because they had their genomics data in one database and in one building and they had their protein data in another.

Now they are sticking it onto linked data and now they can ask a question that you probably wouldn't ask. I wouldn't ask. They would.

What proteins are involved in signal transduction and also related to pyramidal neurons? Well, you take that mouthful and you put it into Google, of course there's no page on the web which has answered that question because nobody has asked that question before. You get 223,000 hits, no results you can use. You ask the linked data which they've now put together, 32 hits, each of which is a protein which has those properties and you can look at.

The power of being able to ask those questions of a scientist, questions which actually bridge across different disciplines is really a complete sea change. It's very, very important. Scientists are totally stymied at the moment.

The power of the data that other scientists have collected is locked up and we need to get it unlocked so we can tackle those huge problems. Now, if I go on like this, you'll think that all the data comes from huge institutions and it's nothing to do with you, but that's not true. In fact, data is about our lives.

You log on to your social networking site, pick your favourite one, you say, this is my friend. Bing! Relationship. Data.

You say, this photograph, oh, it's about, it depicts this person. Bing! That's data. Data, data, data.

Every time you do things in the social networking site, the social networking site is taking data and using it, repurposing it and using it to make other people's lives more interesting on the site, but when you go to another linked data site and you say this is one about travel and you say, I want to send this photo to all the people in that group, you can't get over the walls. The Economist wrote an article about it, lots of people have blogged about it. Tremendous frustration, the way to break down the silos, to get interoperability between social networking sites, we need to do that with linked data.

One last type of data I'll talk about, maybe it's the most exciting, came down here. I looked up on the OpenStreetMap, OpenStreetMap's a map, but it's also a wiki. Zoom in and that square thing is a theatre which we're in right now, the Terrorist Theatre.

I didn't have a name on it, so I could go into edit mode, I could select the theatre, I could add on down the bottom the name and then I could save it back and now if you go back to

the OpenStreetMap.org and you find this place, you will find that the Terrorist Theatre has got a name. I did that! Me! I did that! It's the map! I just did that! I put that up on there. Hey, you know what? The StreetMap is all about everybody doing their bit and it creates incredible results because everybody else does theirs and that is what linked data is all about.

It's about people doing their bit to produce a little bit and it all connecting. That's how linked data works. That you do your bit, everybody else does theirs.

You may not have lots of data which you have yourself to put on there, but you know to demand it and we've practised that. So, linked data is huge. I've only told you a very small number of things.

There are data in every aspect of our lives, every aspect of work and pleasure and it's not just about the number of places where data comes. It's about connecting it together and when you connect data together, you get power in a way that doesn't happen just with the web, with documents, you get this really huge power out of it. So, we're at the stage now where we have to do this, those people who think it's a great idea and all the people, and I think there are a lot of people at TED who do things because even though there's not an immediate return on investment, because it will only really pay off when everybody else has done it, they'll do it because they're the sort of person who just does things which would be good if everybody else did them.

Okay? So, it's called linked data. I want you to make it. I want you to demand it.

And I think it's an idea worth spreading. Thanks.