



Thought Leaders in Learning: Too Big To Know: Scaling Knowledge In A Messy World

Presentation Transcript

December 11, 2012

David Weinberger
Harvard Library Innovation Lab

Sponsor

United States Agency for International Development

Male:

Well, thank you. It's a real pleasure to have you all here today, and those of you joining us remotely. This is another one in our series of thought leaders and learning speaker series. We LER are bringing a group of people into the agency, not only for these presentations, but also for consultations in order to sort of broaden our understanding and our awareness of a number of the interesting ideas and thoughts that are going on and broadly within the world of organizational learning and knowledge management.

And we're really very happy and privileged to have David Weinberger here today. He is with Harvard University's innovation lab, and I think he has recently published a new book that has gotten a lot of attention called – let me be sure I got this right – *Too Big to Know: Scaling Knowledge in a Messy World*. If you want the short, there's some good articles in it in the *Atlantic Magazine* that appeared earlier this year.

But I think the biggest endorsement from him was several of our staff attended the KMWorld, which is a big convention of people who work on organizational learning and knowledge management, and David's presentations that he made there were seen as sort of one of the big highlights there and that he was a really very hot ticket there. So we are really very pleased that he had time out of his schedule to come and speak to us specifically at USAID. So, David, we're very happy to have you here, and we're very anxious to hear what you have to say about knowledge and the relationship to the Internet.

[Applause]

Male:

Oh, thank you. Thank you very much for having me. You do good work, and so it's a pleasure to be here. So I have a presentation that is probably too long. I'm going to talk quickly, if that's okay, if it's okay with the Web. So it seems to me that knowledge is changing. It's changing pretty dramatically, so I want to begin with a very quick reminder – and when I say “knowledge,” only in west. Only talking Western knowledge.

A very quick reminder of what it's been in our western culture. It has some properties. It has a whole bunch of properties, but among them are that it is from very beginning – from Greek origin, 'cause usually we date this back to Athens 2,500 years ago – knowledge was something that was the result of a filtering process where you had lots of opinions. Only a handful of 'em were justified true belief. Plato's formulation, and you had to figure out what that was. So it was reduction. Knowledge was relatively scarce and resulted from filtering.

And second of all, we only call things knowledge in our culture when they have been settled, when the issue is settled, when all reasonable people

have agreed, where reasonable contains within it, a massive politics and culture that we won't even – Pandora's Box we won't even opinion. Nonetheless, if there's serious disagreement, we say, "Well, we don't know yet." Likewise, knowledge has had the property of being very orderly, and, in fact, beautiful.

For about 2,400 years of our 2,500-year-civilization, we assume quite deeply that to know what something was, was to know its single place in the order of the universe. And to deny that there was an order to the universe, a single order, right now in currently modern world that's sort of an archaic idea, but until 100 years ago, it was absolutely taken as axiomatic. You could not have knowledge if there wasn't an order to know, and that order had to be singular.

And so to deny that there was order to the universe, made universe made you quite literally either a heretic 'cause you're denying God's order, a lunatic because you're saying there's nothing to know. "There is no order. Why are we bothering trying to know things?" We've lost that concept, but only very recently, and we haven't lost it entirely. We still have relatively – pardon the expression – dumb arguments about exactly how to define is blogging journalism, is Pluto a planet. These are not interesting discussions. They assume that there is a single order.

And so throughout our history in the west, we've confronted a really basic problem about knowing the world, which is our brains are tiny, little specs in the universe, a kilogram of gray, and we're trying to understand everything from the big bang up to the end of the universe and 5 billion years, or whatever it's going be, and we want to understand human emotion and is there a God. We want to know everything about the world, and we are woefully ill-equipped to do this.

And so we've had to have a strategy, and the strategy has been to try to reduce the world, to break it down into knowable chunks, for example, which is what we do with experts. We allow them to break off a brain-size chunk of the world to know it really thoroughly, and then we can ask them questions, and most important, we can get an answer and move on. We go. We read the book. We ask the expert in person. We get an answer, and we can stop asking the question. And this is, in fact, where the great efficiency of our system of knowledge, which has been incredible. We're the dominant species because of the system of knowledge, but we've constructed it as a series of stopping points. We get our answer. We do not have to repeat the entire history. We don't have to redo the research. We just move on and so we can build on it.

It is not an accident that the properties of knowledge, the ones I pointed to as well as others, are, in fact, also properties of the medium of knowledge,

the traditional medium of knowledge. This is very Marshall McLuhan, but I think he's right about this. These are properties of paper, of writing things down, of books of libraries, and so a library is a filtration of books, books are a filtration of manuscripts submitted to publishers. Most manuscripts don't get published.

Knowledge is settled the way ink settles into paper. You don't publish something until you're really sure, really confident, because you can't take it back. You can't pull the ink out. We have to organize our thoughts into books in very rigorous and constrained ways, the way books are constructed. I don't mean physically. In terms of how they present knowledge is hugely constrained. It's very few ways of doing it in an acceptable form, and we know what those forms are and we live by them and they are all reflections of the fact that books are short, they're disconnected, they are sequential. That's physically sequential. And likewise, we organize our books into a single order because they're physical objects, and in the physical world, everything, just as in the order of the universe of ideas, everything has a single place.

Well in the real world, that's actually true. You have to decide what the principle of organization is going to be for this information, for these books. So you say we'll do it by topic, and within that, by authors' last name. If somebody simultaneously wants to do it, say, simply by chronology, you can't. You can only do one at a time with physical objects, and so we've imported into the world of ideas this horrible limitation of the physical. You only get one way of organizing the physical and so we've thought that there can only one right way of organizing ideas. This is a terrible limitation on our thinking, which we are now able to get past.

And, finally, knowledge consists of a series of stopping points in the way that books, for all of their goodness, et cetera, et cetera. I work in a library. I'm not going to spend time arguing in favor of books, 'cause you already like books, I assume, but they're really limited. They are disconnected from every other book. The ideas are connected, but the book is physically disconnected, and so the construction of books takes that for granted. The construction of expertise takes for granted that a book starts here, it ends here, and it's only about this thick. It's not this thick, 'cause that's too thick.

So books have lots of techniques that we can talk about for dealing with the fact that they are physically disconnected. Footnotes are one. Footnotes, of course, point out. I'm not saying all books – books always come out of a social milieu out of a culture, and the footnotes point back into it, but the footnotes actually are there primarily as stopping points. Once again, you have a question about why the author said this, you can

see why she said it, 'cause there's the footnotes. Very rarely do we actually look up the original, 'cause it's a huge pain in the neck to go back to the library and crawl through the stacks.

And so knowledge has reflected the properties of its medium. But now we have a new medium, and so we can expect that knowledge will start to – and this is a radically new medium. This is as radical as the invention of literacy. The printing press was a very big deal. I would not want to underplay it, but it was taking the fundamental task of writing things down, of ink on paper, or scratches into clay or whatever, and automating that process, huge implications, of course, but it seems to me the Internet in taking away the writing things down part is, in fact, a bigger change than the printing press. It actually is on the scale of the invention of literacy itself. And if I'm wrong about that, it doesn't matter. Nothing follows from it.

So one of the ways of thinking of knowledge taking on properties of the Internet is to say it's becoming a network. We're seeing networks of knowledge. And so the smartest person in the room in a network of knowledge is not the person at the front of the room. It's the room itself. It's the connection among the people. So I want to look at a few knowledge networks in order to give a sense of what they are, and also to take away the impression I've probably already left, that a knowledge network is some special piece of software that you have to buy, install, and maintain. There are, of course, systems that do that, but what I'm talking about is much more common and familiar.

So the first example is what's going on in science. Michael Nielsen has a really good book about this called *Reinventing Discovery*. So if it's 1919 and you've been waiting for an eclipse so that this Einstein fellow's theories can be tested – there was a critical experiment done in 1919 eclipse – you opened up your newspaper and there on the front page of *the New York Times* was an article about it which may have gotten you very excited; very, very interesting. And if you wanted to pursue this idea, this topic, you had some questions or something you want, you couldn't. In 199, you couldn't. This is all that you had.

That was it. You couldn't go to the library and get more information, 'cause there wasn't any. This I the day that was you'd have to wait for weeks or months or years to be articles and books about it, and they'd be difficult to come by. Our curiosity was cut off by the media. This was all that we could know. So in 2011, some data was discovered that seemed possibly to overthrow Einstein. This is the faster -than-light neutrino data that was coming from the Large Hadron Collider. It was really well sourced data, conceivably the most important data since Einstein.

The group that discovered this posted the information, not in a peer-reviewed journal, not in a paper-based journal, but in an open-access journal that has no peer review. Now most open-access journals do have peer review. This one, archive.org, doesn't. It's wildly popular among scientists and mathematicians for whom it's designed, and anybody with any scientific standing, any university affiliation, can post whatever he or she wants there. No peer review at any stage of development. So people post raw data. They post rough drafts. They post finished drafts. Whatever you want. So they could have gotten into any physics journal with their data.

Why'd they pick archive.org, it's for the obvious reasons. They wanted to get the information out quickly, and they wanted to get the information out quickly so that a Web of discussion would form around it, which is exactly what happened. Over the next few months, there were thousands and thousands of articles posts and Tweets and everything else about this data. These came from world-class physicists in named seats, named chairs. It came from complete amateurs that didn't know anything and were idiots. It came from everybody in between. Every type of question was answered.

If you had a question, you could get an answer. If you couldn't, you could pose a question somewhere on the Web. You could follow the chain of links into infinity. There was no natural box around this topic. This was an amazing educational environment, and it is also where the knowledge itself was. It was in this nest of this Web of disagreements, of differences. This network only had value, this sort of subnetwork around this data only had value because of each of these pieces disagreed in one way or another. It was different from the others. So here we have knowledge that is vast, that is non-reduced, that is open to everyone, and that has value because of the differences, not because it has driven out all differences.

So on the one hand, this is wonderful. On the other hand, it's a little scary because it means we now live in a world in which differences do not get resolved. There is some additional good news which is that we seem to be getting better at living with disagreement, rather than insisting those other people are wrong, we're right. We're going to marginalize them. We do better. We can build these networks of knowledge better if we're able to prosper in difference.

So in the 19th Century, there was way too much debate, decades of debate about what this thing was, how it was to be classified. And, in fact, whether it could exist, because it denied the Linnaean classification scheme. Quite literally, a dead one was shipped from Australia to England where scientists dissected it, found eggs inside of it, which is not possible for a mammal, and assumed it must be a hoax because it defied too many

categories.

So in the 19th Century, we wasted a lot of time trying to decide is this a mammal or not. Now we don't have that argument so much. We don't feel the need to resolve it. So at a site like Encyclopedia of Life dot-org, which is a wonderful site, eol.org, they want a page for every species. First of all, they collect every name they can find for it. Any language, slang, vernacular, scientific, whether, they will collect it because they want you to find it no matter what you call it.

So rather than insisting, "Oh, really should call it a platypus. Water mole is just too slangy," they don't care. They just want to help you find it. So they're non-judgmental. It's a scientific site to keep in mind. This is an scientist site. Non-judgmental, any name. And furthermore, if you think that this really has to be classified here, and I think, "No, that's crazy. Your taxonomy's all wrong. Here's how it's really one of these," the site also collects taxonomies.

So I can see it in a taxonomy that I like. You can see in a taxonomy that you like. We don't have to have that argument, 'cause that argument's not helpful. In almost all cases, it's just not helpful. And so we're able to collaborate and build knowledge around this thing without agreeing what to call it or what it is. This is a really fruitful way of managing difference and disagreement, enabling collaboration across fundamental sort of ontological questions about what the thing is.

This is an example of the use of a namespace, which is a really powerful concept which has become all the more useful and powerful in the age of the Web where it's really helpful when you're in a scaled environment, a huge environment. A namespace is a domain of knowledge in which the objects have whatever names they want prefaced by they're within a namespace that is unique. So it's a good way to enable people to talk about things the way that they want to without having to worry about conflicts with other domains, and we can talk about that if you want. That was really bad explanation.

Another way that we're learning to deal with difference fruitfully, you can see at Wikipedia. So this the history of aviation at Wikipedia, in the English version, very American-centered. Aviation flight was, flying machines were invented by Americans, all about the Wright brothers. In the French version of Wikipedia, the same article, not so much about the Americans. It's pretty much turns out the French invented flight. So there's the useful, a real different disagreement. Both of these are reasonable readings of history. They're culturally different. It's always been the case that the American Encyclopedia and the French Encyclopedias, et cetera, have had different histories of aviation and the

rest of it, different cultures. That's what makes cultures different. So that's not new.

What's new is that these, at Wikipedia, because it's an online thing, are linked. So if you're in one, you can click and see the other, and vice-versa. Possibly as a result of this, over the past five years – it's hard to know what it's a result of – but over the past five years, these two articles have actually gotten closer. They're still different. French more MIT to the French; Americans to us. But they've gotten closer in their story, in their histories, possibly because we now have not simply difference, we have linked differences, and that is a huge, huge difference.

So from this, what do we see? Well, from science as a network, in the case of the faster-than-light neutrino data, rather than filtering it beforehand, they put it out and allowed the filtering to occur after publication. This created an ecosystem that every niche was filled. Whatever level of expertise you were at, you could understand what was going on. You could find stuff that was directed to you. This knowledge has value because of the difference, and at our best – and there's a lot of at our worst, but at our best, we're learning how to maintain fruitful differences.

Okay. So a second example of a knowledge network which is an ecosystem. It's the ecosystem that developers have created for themselves. I'm willing to acknowledge this is an overstatement. I'm not sure that it is, but it might be, that software developers have created an insanely efficient ecosystem for learning staggeringly. So if, for example, you need to learn a new programming language, there are tens of thousands of tutorials at various levels of different points of view.

This to me is the most amazing thing. If you have a question, you can go to a site like Stack Overflow. You'll go to Google which will very likely send you to Stack Overflow, which has over 3 million different questions, almost all of which have been answered, which are at an amazing level of specificity because in a scaled environment in a world in which there are a couple billion people on the Web, it becomes less and less likely that any question that you have, you're the first person to have that question, less and less likely.

And so you can go to Stack Overflow, ask it a question that can be surprisingly detailed about something you're having a problem with as you're doing your programming, there will be an answer, very likely, and there will be iteration on that answer. That is somebody else will write in and say, "Oh, that's a good answer, but it doesn't work in IE9, because nothing does, and so here's the code that you need to insert," and somebody will say, "Well, that's good, but actually you could do this

more efficiently this way,” and after six or seven or eight iterations, you have a really solid, robust piece of code that you can just copy and paste. And the next person who has that question and goes there, there’s the code. There’s the explanation. There’s the discussion that explains how they got there. It’s an amazing thing and has made developers crazily productive, crazily efficient.

Likewise in the open source world, sites like GitHub. GitHub is maybe dominant at this point. You post your code there, makes it very easy for people to reuse your code, to fork it, to make it a little bit different, and to make that forked code available as well. So I think we see a bunch of things going on in this environment.

First of all, despite their reputation, this only works because developers are humble enough to be able to admit in public that they don’t know something, and generous enough to spend their time responding. So sometimes it’s not generosity, it’s showing off. We’re human beings. We have a mix of motivations. Nonetheless, that is part of it.

Second is the power of iteration when you’re in a scaled environment. Little differences, little interactions. Wikipedia is also another example of this. Little interactions spread across a large enough set of people result in amazing collaborative work. That’s part of the economics of collaborations. Small iterations spread across a large group.

And maybe for me, the most important thing from this example is the power of learning in public, seeing learning as something that is selfish, but only instructs you. It’s great. You ask somebody a question, you get an answer, say you do it by e-mail, you’ve learned something. That’s wonderful. You have not made the world better, but you do if the next person who has the question can learn from what you learned. Learning in public is, as a public responsibility, I think is a fundamental thing that we’re seeing beginning now. That’s Stack Overflow; that’s what it’s about. You see this in at a bunch of other sites as well.

Another example of something like a knowledge network, so we’re very, very good at organizing things. We do this so we can find them. We do this because we’ve had this investment in the idea that their spot in the organization is what they are. But then you have to look at a case – there are many, many of these, but in this case, this is the Library of Congress about five years ago had a few thousand color photos from World War II, does not have the staff to do all the proper metadata work, all the information about the photos.

And so as an experiment, they decided to put it up on a public site, Flickr, which is owned by Yahoo, so that we could see these photos which are

really beautiful. When you see them with good contrast, this is an even more gorgeous photo. But also so that we could collectively collaboratively fill in the missing pieces of the metadata. And so Flickr has a variety of tools for doing this. You can annotate. You can have conversations, but here's also a tagging system with 75 slots for tags where anybody can come along and label this, basically. And you click on one of the tags and you see all the other photos at Flickr that had been tagged that way by someone, and so you get a wide range of types of tags. Some of them, they can be entirely idiosyncratic.

If I think this looks like Aunt Erma, I might tag it, "Aunt Erma," and it would up. It wouldn't help you at all, but it'd be useful to me, and that's perfect permissible. So you get those sorts of tags. You also get tags of the sort that a professional cataloguer would give, like the year of the photo. You also get tags that are off beat. They're quirky. A professional cataloguer would not have called – she has a really nice hairdo – this hairdo a coif, 'cause it's an archaic term. It is, isn't it? Anybody got a coif? Okay, so it's an archaic term. But it's a perfect term. If you're interested in women's hair from the 1940s, coif is a great tag. You click on this, you'll see a bunch of really interesting coifs. Even though it's archaic, it's exactly right.

And then red is sort of an aesthetic tag. I had not noticed till I saw the tag that the red, which is not that many pixels overall, but it's still really central and dominant. It's the red of her lips and the red of the tip of the probe, but it's an aesthetic tag. It makes a sort of aesthetic criticism. It draws you to a bit of a look of it that otherwise might have been hidden from you.

And then there's this one, which is really why I wanted to wait it. Rosie the Riveter. So this is not Rosie the Riveter. This is Rosie the Riveter. And so this tag is inaccurate. It's clearly inaccurate. That clearly is not Rosie the Riveter. And yet, it's a really useful tag. If you're interested in the role of women in manufacturing in World War II, you think on this tag, you'll see a whole bunch of really interesting photos. So here's a tag that no professional would have applied because it's wrong. It's misleading. Somebody might come along and that that, in fact, is Rosie the Riveter. It's inaccurate. It's a lie. It's a knowing lie. The person who did it very likely knew that wasn't Rosie the Riveter, and applied that tag anyway. The shame of it. But it turns out, this is a really useful tag.

And so my overall point about these photos is that they're messy. They are messy beyond what any professional would do. There's no controlled vocabulary. There's no hierarchy. There's no taxonomy. And the tags range from the idiosyncratic to the quirky, to the outright lies, yet it's really, really useful, and it's really useful because it's the only way you

get meaning to get really, really big to see all of the connections of this photo to everybody's lives, to every topic that we can imagine. It's only by allowing this sort of mess. Messiness has lots of problems, but it scales meaning. It let's meaning get really, really big. And this actually has I think, or arises from a structural change in the nature of metadata, information about information. We're very used to metadata even if we don't call it that. Folders on a label or, in this case, somebody's self-labeling herself. And so we have assumed, as has been the case in our information systems in the past, that metadata and data are separable things. They're two different things. You have data. You give it a nice set of metadata so it can be found. That's much less true now.

So, for example, if you know the name of the author, but you can't remember the book, you type it into Google or whatever, type the author's name in. It's metadata that allows you to find the book. In fact, even to find the content of the book these days if it's online. But likewise, if you don't remember the name of the author, but you do remember a quote – what is that book that had this quote in it? “Call me Ishmael,” might a been the first line. Not sure? You can type the content in and you'll get back the author. You'll get back Herman Melville.

In fact, you'll get back the contents of the book. In fact, you'll get back everything. You'll get back this amazing list. You get his bibliography, his biography, the house in Lenox where he lived, his social network, Nathaniel Hawthorne, the map to Hawthorne's house, the time they went up to Monument Mountain and had a picnic. Nobody knows what they talked about. You'll get information about the whale, the biology, the anatomy, the ecology of the whale. You get back Al Gore. You'll get back everything from that one little piece of metadata. And the point is that metadata is no longer something distinct from data, that we have our data, now we design our metadata. Rather, metadata, there's no difference except in terms of function. It's an operational difference. So metadata is the thing that you know, and you use that to find the thing you don't know. You know the author's name, that's metadata. You know the quote from the book, that's metadata. And you use it to find the thing that you're looking for.

This is a huge change in how we think about information that's occurred thanks to the Web without us doing anything. It just happened. And overall, it's really beneficial because metadata is a lever we use to pry up knowledge. I know this thing. I'm going find out something I didn't know. And if everything now is a lever, we just got much smarter. So the same sort of messy changes, as well as some others, are happening to data as well as to knowledge. And I just want to spend a moment on this because you are rich in data, and my secret aim is to encourage you to make as much of it available in useful form to the public as possible, even though

or, in fact, because you don't know what people will do with it. And the stuff that's dangerous or secure, obviously you're not going to.

So data had been this thing in a cell. It's at the intersection of two pieces of metadata, the row and the column, and we know how to deal with that. That's what it has been. But if you look at data.gov, a site that you may be familiar with, there are many examples. I happen to like this one. Data.gov worked because it accepts raw data. If the government required all the data to be cleaned up before it's presented, then it wouldn't happen. We'd wait 20 years and we'd get clean data that was of no use. So I think the right decision was made that it's better to have raw, unreliable, inaccurate data, than to wait 20 years to get reliable data, perfectly reliable data, that messiness is okay, especially if scale where there's some averaging that occurs.

I know there are times when that's not acceptable, but lots of times it is, and it's only because of that that we're able to get at this sort of data, but it's really messy data. You can't rely on it. It's wildly imperfect. It can even be inconsistent or inaccurate, and yet it can be incredibly useful. And so there's a tradeoff, and in this case, I think personally, we've absolutely made the right tradeoff. Better to get masses of data out in imperfect form, than wait forever to get it perfect.

This move towards releasing data I think illustrates another and related change to the nature of data itself. So the first change is that data is becoming messy in the world of big data as we decide to make the stuff available. The second is that just as with knowledge, it's becoming linked. The links are becoming internal to it. So Tim Berners-Lee, about seven years ago, backed linked data or linked open data, when the data is open. Is linked data familiar? Not a lot of nods. That's fine. There's no reason for it to be. I will do a worse job explaining this than I did namespaces. This is my promise to you.

So linked data is a very popular way of making large amounts of data public. So rather than silos, we think about putting things into clouds, and the big data cloud, quite frequently, as institutions are putting this information out, they're doing so in the linked data format which in its briefest uses links to point at the metadata that explained what the data is. It's actually more extreme than that, but I'll leave it there. And this has the benefit of allowing computers to make sense of vastly different types of data that have not beforehand figured out and agreed upon a shared scheme for expressing the data.

How bad was that? Okay. So the alternative was that when you're going to release data, you figure out what format, which data, what metadata, and you code it all up and that's useful, but it means that a computer has to

know a great deal about how you've structured your data ahead of time. With linked data, it's a technique by which computers don't have to know nearly as much. They're able to traverse and make connections among the pieces far more fluidly with far less work beforehand. How's that?

Okay. So I encourage you to look into this if this is beyond the veil of mysteriousness that I've pull over it because it's actually a preferable way of making data available because it results in the following. It results in the creation of data commons in which all of these clouds are available to anybody who wants to traverse them and make connections between the author of a book and the weather when she was born and the star under which she was born – bad example – the epidemiological conditions of the area in which she was born. Whatever set of connections you want to make you can start to make.

It's a data commons. It's an ever-growing set of clouds, sometimes quite large, of information of various sorts. And the reason why I think this is interesting and important is that it takes us away from an idea that the right way to make data public – if that's what you're inclined to do – is to figure out exactly the perfect structure, or highly structure it, and use it to build careful models of how the world works. And if you get your model right, in a world that is, yeah, I'm sorry to say, too big to know, well, it is gigantic, there's so much complexity that no single way of thinking about it is going to succeed. It just can't. The world outstrips all of our models.

And so I really like what John Wilbanks who was at Science Commons says which is the way that we get smart, the way that science gets smart in particular, is that we need to have your nerds arguing with my nerds. It's not that we all need to collaborate to come up with a single right model of how the world works and we'll feed in the data and we'll turn the crank and now we'll know. What we really need is a marketplace of science, of models, of computer models and scientific ideas so that they can be in contest. And so your nerds can argue with my nerds. And in that argument, not in the victory of either one, but in the argument, we now have an environment in which more accurately represents the world.

The last of the networks is Wikipedia. So if you go to Britannica, the old print Britannica, which is no more, and you opened up randomly, you were right to believe what you read because you knew that it came through a careful editorial process, and so simply being in Britannica gave something credibility.

If you open up Wikipedia randomly, you not as safe believing what you read, 'cause there's a chance you're going hit it at the very moment that some idiot has trashed the page. So simply being in Wikipedia does not confer authority. Now my point is not Wikipedia sucks. Don't believe it.

It's not credible. It's actually the opposite. I find Wikipedia at this point to be generally a highly credible source. My question is why does it have that credibility, to which there are more important answers than the one that I'm going to give you, but I want to point you at one particular phenomenon here.

So one of the reasons is that Wikipedia is very happy to insert warnings about the reliability of its own information. Very happy. There is over 100 of these different warnings, and if you need a new one, you can make one up because we are, after all, all Wikipedia editors. The set of 100 reads like the set of ways in which an encyclopedia can go wrong. So this articles appears to contradict another article, and it appears to contradict itself. It just not worthy of an encyclopedia page. It needs to site sources. And Wikipedia over the past three years has really pushed on this one, and so that's why now you see Wikipedia articles are replete with citations.

It reads like an advertisement. It reads like a sermon. And then this one, which is especially important because the pursuit of neutrality is at the heart of Wikipedia's process and at the heart of its success. So the neutrality of the article is disputed. So it seems to be obvious that Wikipedia's willingness to put up these notices – and there are more and more these days. This has become even more common over time – its willingness to put up notices of its own lack of credibility, of its on fallibility, actually includes Wikipedia's credibility because you just look at it and you know that Wikipedia is not trying to convince us of things because it wants to be the implement authority.

It just wants us to know our world better, and it's very comfortable with this most basic of human facts which is that we get everything wrong. We're fallible creatures. It's at the very heart in our culture of what it means to be human that the human divine distinctions is the divine gets everything right, and we get everything wrong, and then we die. This is the most basic human fact. Authoritative sources, authoritative agencies are very uncomfortable with this. Businesses are very uncomfortable with – look at any business market, any brochure, and it's all about perfection and everything is right. There's not a dent, a scratch. And we all know that it's basically lying. It's a sort of platonic impossible, platonic ideal of what a hotel room looks like. None of 'em look like that.

It's one of the wonderful things about trip TripAdvisor – you know TripAdvisor? Yeah – is that they encourage actual people to upload photos of the hotel rooms. It's really sort of wonderful. Rarely do they match. Anyway, so this is the most basic fact about humans. The denial of this is making sources less credible. In fact, I think the right question to ask is not why does Wikipedia do this, but why we don't see this in more places. Why don't institutions like this do it? And the answer is I think

very obvious, it's that they have an economic and psychological interest in convincing us that they are the authorities and they speak in this tone of voice. That tone of voice is increasingly just going to alienate people and convince people that you are a gas bag and not credible.

So some overall points, and then I want to make a large and tentative conclusion. So what enables a room to be smarter than any of the individuals, or even the cumulative? Part of it is the power of links, links that link differences and disagreements, 'cause the world's very complex and any answer that is simple and universal and unequivocal is very likely wrong.

Learning in public is an amazingly powerful and useful tool I think for any organization. Learning in private, becoming educated by yourself is selfish. It's inefficient and it's a waste of someone's time when that same process could be done in a way that exposes a learning to the public in a findable way. Messiness is how you get things rich. You get things less accurate. You get things less reliable. You get things less universal, but it's how you enable something to be saturated with meaning.

And talking like an institution, there are times when that absolutely is required. You can see this in businesses. We'll go to a business website and we will ask, "What's the wheel base of this car?" when we want to know if it's going fit in our garage and we're going rely upon it, and going to the crowd, going to Twitter and asking, "What's the wheel base –?" that's not a good idea. Much better go to the – so there's absolutely room for that. But if that's the only way you talk, why would anybody listen? Much more interesting things to listen to than that.

So the larger point is we have a set of characteristics of knowledge. These are some of the important ones. In the age of the Internet, knowledge is taking on the properties of the network, and so from paltry, to overwhelming, from settled to unsettled, now it's unresolved rather than being neat. It's messy. It's not disconnected, but rather loosely connected, and held together by interest. The links that are created are created by people who think it's interesting enough to link to this place over there to send you away from my site. I put in a link. I say, "Yeah. Go over here, that's really interesting." And so the Web, which is by and large, a voluntary activity, is an expression of human interest for better and for worse. There's a lot of stuff we would rather not acknowledge, but apparently, we're very interested in it.

And so the properties of knowledge are becoming properties of its medium, it's new medium, the Internet, but it also seems to me that these are also properties of what it means to be a human who's trying to know the world, a world that is overwhelming, that is complex and messy and

unresolved. This is our position s knowers, and I think that's one reason why we've seen such a rapid embrace of this new knowledge environment. It's a better reflection of who we are. And so network knowledge may or may not be harder than – it's a lot easier when you can just look something up and you get the answer, no doubt, and that's still a useful activity.

So network knowledge may be harder, may be more difficult. It may even work out to be not as true about the world. I think it not, but that's possible. But it seems to me to be indisputable that it's a better reflection. It is truer about what it means to be a human trying to know a vastly overwhelming and overwhelmingly complex world. Thank you.

[Applause]

And I will be happy to e-mail each of you a better explanation of namespaces and linked data. I was just trying to go as quickly as I could so we could talk.