MARRAKECH – How it Works: Overview and workings of the World Wide Web Consortium (W3C)
Sunday, March 06, 2016 – 08:30 to 10:00 WET
ICANN55 | Marrakech, Morocco

UNIDENTIFIED MALE: All right, it looks like we have a packed room today. This is the beauty of – I guess the only session worse than being an afternoon right after lunch session is being the first session on a Sunday morning at the ICANN meeting, and being in the basement to boot. We've got all this going for us, but that just means it's going to be a much more personal experience for each and every one of you.

I want to welcome you guys. This is the third installment of our How it Works series. We piloted it two meetings ago, and then ran these types of sessions in Dublin as well. This is our third series of them, and we've got a brand new presentation today. We have Daniel Dardier from the World Wide Web Consortium. He's going to give an overview of the W3C and talk about it in relation to standard settings, so Daniel, I'll pass it on to you.

DANIEL DARDIER: Sure. I'll take this mic. Hello, good morning, bonjour, salaam. My name is Daniel Dardier, I live in France where the W3C has a site. I'll explain that. I'm going to present the consortium and some recent activity as well. You probably heard of W3C. It's a

standard organization specific to the web technology. To give you some feedback, it was created more than 20 years ago. It's a consortia, so it's not an official, de jure standard organization. It's closer to the IETF in that sense. I'll give some detail.

It's funded by membership, that is people joining to participate, to get more advanced information, but most of the working groups are public anyway, so it's more like sponsorship/membership. We have a staff, which is in a sense a model that is closer to the standard organization that are usual, with what's called a secretariat

Our staff is pretty much technical. It's most of the people that have invented the technology that we work on, whether it's the web, like our director, Tim Berners-Lee, or the technology around the web: XML, CSS, metadata, etc. We have people that work in various places on Earth, and we're focused not just on the web technology, but also on some of the environments, the testing.

For instance, we run the validator for HTML, CSS and other markup language on our site, which is a free service that is used by millions of people. We do also a lot of liaising. I'm not sure how familiar you are with the standard landscape just for the Internet and communication technology. There are dozens of

organizations that are specialized in some area of the technology stack.

We have, of course, to make sure that the system works together. We're responsible for one layer, which is the web. Of course we depend on the well-functioning of the layer below us, the Internet, but also we have to pay attention to platforms that are built on top of the web.

As I mentioned, W3C is not incorporated. It may change in the future, or we may decide to become a legal entity, but right now it's a project between different entities that are themselves, of course, legal, and they are usually academic institutions. We have up today four, what we call host sites that are academic institutions that provide a salary for staff of W3C, and also some steering for the consortium administration.

MIT is the original host site. It was created in '94. In '95, we created a European antenna site, which is where I live. I was the director of this antenna for a few years. It's in the south of France, it's Sophia Antipolis. If you know, that's where Etsy is also in terms of technology. We have a site in Japan as well, and more recently we opened a site in China, which is growing, at Beijing University in Beijing.

You might wonder who control the W3C. It's kind of a difficult question to answer. There are different bodies that are in charge

of running W3C whether it's at the technical level, at the administrative level, or at the strategic level as well. One of the most important bodies is called the W3C management, and it's the set of all the technical leaders of the area that W3C deals with, but also people running the system, communication, international relation.

That body decides on the day to day advancement of W3C, but as well on more important and long term initiatives that are then run by an advisory board, which is elected by our membership. We work with the advisory board to come to consensus for all these matters, and usually when it's important enough, we actually run it through our membership. We ask our advisory committee not to vote really, because it's more like a consensus system, but to give their opinion whether or not they would participate in something like that, whether they're opposed to it, whether they don't care, and depending on the resource that we have, the resource that they can provide, we decide to run the new initiative or not.

I'm not sure about this missing piece of the screen, but anyway, we also have a technical architecture group, which is on the other side of things. It looks at the overall architecture of the web platform, and from the 50 or 60 working groups that are generating technology, it makes sure that they fit together. It's also, of course, very much in liaison with other technology

groups that are working on Internet technology, such as at the IETF, etc.

The tag of W3C is pretty much the architectural board. It decides on whether or not something is good for the web, or it may be a danger, a risk for the web architecture. What we produce is called W3C Recommendation. It's the name that we give to our standard, like IETF gives RFC. We have our own term. This is the final product, has been tested, implemented usually.

It's been also agreed by our membership to bear the name of W3C. We might decide after our whole work on the technology that it's not worth becoming a recommendation. It would stay in a different kind of status, and people can use it, but it's not a recommendation. We have a funny example of that, like the web service architecture document. It's never been a recommendation, even though there were a completely entire system built on top of the web service architecture.

That's because people thought at that time that may not be a time to standardize it. Maybe it would evolve. Eventually it stayed like that, and web service became whatever it became. With the cloud coming up, it hasn't really been successful. This is a map that represents where W3C is present on the planet. We see in red, there are the host sites, where we have our staff, like

me. It's probably about 20 people at MIT, 20 in Europe, about ten in Japan and ten in China. That's the staff.

Of course, this staff, they do not produce the technology. They help the members engineer to produce the technology. We manage thousands of engineers in hundreds of working groups, but the staff is there to orient, to explain the process, and also to pay attention to the dependency between different working groups. In blue in this picture, you see the offices of W3C. We have this concept of chapters, which relay people doing outreach for W3C, and also getting input from the local industry, whether or not they want to get involved.

We have one in Morocco, you can see this dot. It's in Rabat. It's hosted by I think it's Mohamaddia Engineering School, in Rabat. They run workshop. They do translation in Arabic for some of our documents. You can contact them if you want more information.

A bit on finance, I mentioned that the membership is what makes most of our revenue. We also get some government grants. Usually, they are for research and development. We try to have them pay for doing standard, but it's not really simple, because standard budget usually is reserved for different kinds of organizations than W3C. Nevertheless, since our standards

EN

are pretty much advance development, we usually get some funding to do some work on that.

For instance, most of the accessibility work of W3C, the web accessibility has been helped by the European Commission from '97-'98 on. We participate in research projects with our institution to help people understand what is web accessibility. I'll give you some detail later on. We get some sponsorship from either our members, or for meetings or for initiatives in particular. For HTML5, we may get a particular sponsorship to pay for an editor because it's a huge spec.

ISOC has also been generous in the past, and has helped us going through a tough time. So overall, we have a budget around 10 million euro, dollars, whatever. Every year, we're basically in the red. We use everything we have, and usually we don't invest because we're a not-for-profit, so we stay at about the same size. We haven't grown much in the past 20 years.

Our members, they come from everywhere. You see that the US has a big part here. It's evident that there's a lot of web technology giants in the US, so you could expect that, but if you add all the European countries together, you'll see that it's about balanced. Of course, we have a deficit in countries like China, India, which are huge in terms of Internet presence and

development, but are not yet fully participating in the occidental consortia or regionally.

So we try to change that. We rotate our meetings. We make sure that everybody can participate and understand the interest to participate in the future web standard. Just a word, if you have been familiar with standardization, it's mostly the same. Ours has some steps that require, whether it's testing, implementation or public review of the specification.

Of course, I'm not going to go into the detail of that, but the standardization world is full of patent problems, so we have to pay attention for all the steps that there is a clear IPR policy for people submitting technology. Usually what happens when we start a new working draft is that there is a set of players that have done already this kind of technology. If you look at HTML5, before we standardized the complete set of markup elements, there were extensions to HTML4 in –whether it was Netscape, Chrome or Firefox.

So we put all those people in a room together and we had them agree on a common set of semantics for all the elements they want to add, and usually they submit their technology, the extension that they have. We have to make sure that the submission is not encumbered with patents because we have made a promise to the community that our standard would be

royalty-free. That is that there's not going to be a problem for implementation for the open source community for instance, or for just new players that cannot afford the price of intellectual property for web technology.

This was done by experience. In around 2000-2001, we had this unfortunate situation where one of the technologies we were working on – it was on privacy for web sessions and things like that – while we were doing the standard, one of the companies that were participating came up and said, "Well, it's fine that you're doing the standard, but we have a patent on that, and so anybody implementing the standard will have to pay something to us."

We realized at that point that we didn't ask systematically to all the players if they had any patent, and they would be willing to let them in the patent pool of W3C in the sense, to have them be used royalty-free. We had to re-engineer our process to make sure that at each step, there was a clear way for people to commit. We don't want to take the patent from people, we just want them to make a pledge that they're not going to use this patent to ask for money to people implementing the web platform. That's our requirement.

Even though we have no enforcement right to chase people that do not do that, we count on the fact that all the industry is

playing the same game. Whether it's Microsoft or Google, they're going to pledge for some of the technology that they have and that they submitted to W3C to improve the platform. In the end, their pledge is such that they promise not to ask for any patent, unless someone is asking them for a patent for the same pool of technology.

So there is a defensive mechanism. If one player starts to ask for a different kind of regime, then all the other big players that have already assets in the patent pool can ask this person, only this person to pay a fee for whatever. XML, link, something that they most certainly use for free today. So there is a system that has been working for more than ten years, that works pretty well.

Why do we do that? I think the industry has recognized very early on that it was much better to have a share of this big pie here than to have the little pie whole for yourself. It's important that the industry understood that the growth of the market is what is going to make them rich. It's not the technology at a particular layer of the stack. They want to make sure that the web platform is open, so that they can add other platforms on top of it.

You can see that Facebook is a platform on top of the web. You have API for Facebook, same way that you have API for the web or API for the TCP/IP level, and so those people, they understand

that the freer and the more flexible and the more open is the web platform, the better it is for developers on top of it. This idea of a level playing field is very important for the web platform.

The advantage of membership, since everything is free in the end and most of the development are public anyway, is that again, it's like a club. They gain early access to where the market is going to be in a few months. That's very important in this area, that they know, even if it's not completely sure, that in six months from now there's going to be a new standard on the web platform that is going to allow this kind of innovation. They can work on the innovation. They already have work on the innovation, in fact.

They just have to re-engineer it on top of the standard, so the important thing for them is that the new standard is as close as possible with what they have done already. The challenge for us is to have two or three big companies that have a lot of things in the web, to agree to change their implementation to be a standard.

What we produce is called the open web platform nowadays. It's HTML, plus the technologies that are necessary for building web applications. The idea is that the web changed from being a web of documents, static documents that you would go navigate

from link to link to document, to just an operating system where you can build applications. It's a network distributed operating system that you can use to write an application with menu bar and user interaction, like on a regular application on any computer.

Of course, the web didn't start as an application platform, so we have to catch up. We have a lot of things that needed to be added to the web. JavaScript provides with a programming language. We didn't pick up JavaScript. It came up as the winner on the market and it's being standardized by another organization called Ecma, which is in Geneva. We haven't really approached the JavaScript standardization. It's good as it is. It's making improvement.

We focused on the HTML5 language, which is really now a very powerful language associated with JavaScript, and also on the fact that on the web, you want to separate the styling of a page from the content of the page. That's very important for a variety of things. If you can separate the form and the style of a document from its content, then you can repurpose. You can change it to adapt to user needs. If you have a mobile, you can change the content presentation to fit better on the mobile screen. If you're blind, you can change the content to be a text audio stream, etc.

This idea of separation is central to the web architecture. HTML5 is the cornerstone of that, and CSS 3 is the language for the style. JavaScript is something else. You have metadata, which is another set of language that helps you describe the content, etc. The open web platform is the biggest programming platform ever. It has billions of computers running, and you can see the result of that. Facebook is an application, Google+, all the Google excel sheets. They are application of the open web platform.

You see the open web platform is sort of a stack, like an operating system. You get a graphic library, a network library, multimedia, and everything has to work, of course, in a secure way with authentication, and also in the network way. That is, the data that you get doesn't come from your disc as a usual application, but they come from a different origin in the network. You do an HTTP request to a different site, and you can present them in your same application, but of course, you have to care about the security aspect of that when you get running code coming from a different part of the web on your system. So it's something that takes a lot of time.

Around the HTML5, we have organized our activity around what we call application foundation. We have a team within W3C and activity within W3C that would be specific, whether it's security or real-time accessibility, internationalization, things that are

normal on any platform. You have different specialty working for different parts of the programming environment. We have lots of documents, you can visit the site that explains what are the various status of our work.

We have this big table that we did a couple of years ago that we're maintaining, where you can see for each of the new or existing technologies present on the web where they are implemented, in which browser, on which server, and how they actively evolve in terms of the standardization process. I mention HTML, CSS. You can see on this slide that there are a lot of technologies that we do not alone, but joint with other organizations as well.

For instance, the HTTP and the URI specification are a standard from the IETF, but they of course are relevant to the web. They are the interface between the web and the Internet, so we work together with IETF on that. The things that are more advanced, like the semantic web technology, usually we have research projects going on which people are doing experiments with semantic technology and link data. I can talk more about link data, which is kind of a new name for the semantic web.

What we see today is that the web is not just a computer scientist technology. We don't get too much of the vertical industry participating in the first ten years, but now we have,

whether it's the automotive industry or the health industry, or people working more on the government side. They all want to use the web to the best possible effect, so they come to us and ask us for extensions. Usually what you find is that people are using applications on a mobile, and they would like to make a website that is a web application that has similar functionality than on the mobile.

But the mobile native operating system will give you access to a lot of the device characteristics that have never been a topic of interest for the web. Whether or not your phone is vibrating, what is the direction of your accelerometer on your phone was not a concern for the HTML of the first ten years of the web because the computer didn't have all these features. Now that we want the web to run on the mobile, we have to catch up with all the API that are available to native applications.

Same for the cars. People doing interface or management, orchestration of device on an automobile, they usually use HTML5 because it's much easier for them to develop in their lab, to emulate the system. They would like to have an HTML version that help them more secure, more packaging because it's very important in a car that you don't give access to the same thing to everybody. So they also come to us and ask us for extensions to HTML.

We have this organization of vertical industry now, where we try to get input, whether directly from the industry or from – usually these industries, they have created their own standards organization or consortia, and they have to pretty much liaise with us and we establish cross-membership sometimes, so that we make sure that we move forward in sync.

The web platform is focused on API, not just on markup definition, because what programmers want is a programming interface. They want to call functions that do something in a program. Most of our specifications now are expressed in terms of APIs. We want to support all the media that you can find on the web in a standard way. Because we want the system to be as free and open as possible, we have to look also at the patent issues that are present in codec and video format, and there's a lot of that.

We want to work on all sorts of screen devices, and that's what the separation of style in CSS is giving us. More and more people want to communicate between users without going through a server, so we've been working with IETF on direct communication between different browsers basically, different web applications. And more and more, of course, we have to care about the privacy, security and all the societal aspects that the web is dealing with on an everyday basis. I'll go down on some of those in the presentation.

One word on the principle of standardization. A few years ago, three, four years ago with IETF and IEEE, and I mentioned IRB and ISOC as well, but to mention the three standard organizations that have brought basically the Internet to the world. If you consider Wi-Fi, TCP and HTTP, you have what you run on your laptop anywhere you are on your planet. Those technologies, they come from organizations that were not official standards organizations, that are governed by their own principle.

So we have our own principle that we tried to present clearly to the rest of the world as the modern way of doing standard. Not just Internet standard, but we think that our ways of doing standard with consensus, with running code, is a good way for various standards nowadays, not just the computer standards. So of course, we recognize that it cannot be the anarchy. We have to have due process that explains how people have to participate. We have to be transparent, so that everybody can participate.

It's obvious, because we are all together the standard organization, let's say there are a few thousand people, and we have to provide technology for a few billion people. So it's just not possible that we do that alone in an opaque way, without showing everybody that we do all the time. It's at the risk of getting it wrong and just do a technology that is not going to be

useful is huge, because there are going to be millions of failure and of people trying this technology and not liking it, and doing something else.

The waste of resources and of time is just unbearable, so we have to make sure that everything we do is transparent, that people can participate and also review and reorient the technology sometime. You can look at the open standard principle and the people that have also vowed to that on the website.

Maybe I'll go now to some specific technology area, but if you have any question on the overall W3C that I presented, that would be fine now. Any question on how we work, how we live, who we are.

DAVID CONRAD: I guess one of the questions I had is when you were talking about the finances, you had indicated that W3C typically operates in the red. One can't operate in the red forever. How is W3C going to deal with that?

DANIEL DARDIER: That's why we have a system of host, in fact. That's the advantage of that we're having a system of host. The members or the government contract, they all come to the host. So in

Europe, ERCIM, which is the host of W3C in Europe, would sign a contract with the commission or sign a contract with France Telecom and get money for them, and they have this amount of money and they pay me and the rest of the staff.

At the end of the year, if we're short because we had to do some important thing, then the host itself – which is a huge organization usually, MIT – they can spare maybe $500,000 just to push us in the green.

DAVID CONRAD: So they're sort of like sponsorships.

DANIEL DARDIER: Well, yeah, it's more like a backup system. The host is a backup. In a year where we have more money, that we're a little bit in the green – which, unfortunately, rarely happens – then they can put some money aside, which is called discretionary fund, and then in the year when we're in the red, they get money back in the system. But it's usually a fraction of that. We try just to be right in the red, in the green. We cannot basically have a capital, so we have to consume everything every year. That's why we're right in the red. So we have this system.

UNIDENTIFIED FEMALE: With open standards, it's clear that you probably have a really good working relationship with IETF and –

DANIEL DARDIER: Could you speak slowly?

UNIDENTIFIED FEMALE: Slowly, yes, that's my biggest problem. So with open stand, it sounds like you have a very good working relationship with standards bodies like the IETF and IEEE. But as you said, the web isn't just limited to the computer anymore. Do you find working with other standards bodies like the ITU in particular, do they see the value in cooperating, or is it more of a challenge working with them, those types of standards bodies?

DANIEL DARDIER: Speaking for W3C, we've had some good relationship with the de jure standards organization. More than IETF, I would say. It's not that they have a bad relationship with ISO, but they don't have of a relationship with ISO. We've became a past submitter a few years ago, which means that once our standards are done within W3C, we can have them stamped by ISO to become international standard. They have to go through a ballot and vote by country, and it happened already three times for the web service, for the

ICANN|55
MARRAKECH
5 – 10 MARCH 2016

accessibility and more recently for MathML, which is a mathematical markup language.

We have a contractual agreement with ISO, by which we submit our specification for stamping. They have no right to change it, and we maintain the specification after what's called the transposition of standard. We do that with ISO JTC1 because they are the computer part of ISO. We don't deal much with ITU because they are a different level. They're more at the network level, so IETF must be dealing more with ITU. We do have liaison we ITU, because they are more and more in the application area. They send us the status of their work because they have W3C as a former liaison.

Usually, if we do something in this area, we would point them at what we do already, explaining that there is duplication or not, and either they should participate or do something that is more of an extension of what we do, or a complement of what we do, so we haven't run into any crisis for standardization, duplication or anything with the de jure world.

We've been more concerned with the fact that our standard or not has recognized the de jure system by a lot of countries. We give it for granted, and not everywhere that people can use the W3C or the IETF standard. In the procurement, for instance, when a government has to buy technology, there are a lot of

rules that forbid them to buy technology that are not public and open to all. Usually, they can only procure technology that are based on standards that have been recognized by the government.

It's difficult for them. it was difficult for them to say, let's say ,"We want to buy 500 workstations that run HTML," because the definition of HTML was not a standard. So anyone can come up and say "Well, I fulfilled your criteria because I use HTML 3.2," or whatever. There's no clear definition, and it was most important for accessibility, where, of course, people with disability want to use the government website because they pay their taxes, so they have the right to do that. Plus, sometimes the government asks to pay your tax by the web, and if you get to an inaccessible website and you're blind or deaf, you cannot pay your tax.

You can say "I'd like to pay my tax, but I cannot." So the government have been paying a lot of attention to making sure the websites were accessible. Now, the question we're getting was, what is the standard that explained what is a website accessible, so that we can buy those authoring tools or user agents? There was no standard at the time. The only standard was the W3C accessibility guideline, so what happened for a few years with that individual government, they would take a specification and sort of make their own version of it.

The accessibility guideline is pretty much a technology that explains how to write content so that it is accessible if you have an accessible browser, but it is at least accessible on the other side of the pipe. And it's in English. We would explain that if you have an image, then you have to provide a textual description for the image. If you have "blah blah blah," and so people, since it's not very technology oriented like a TCP or HTTP protocol, they felt that they could change the sentence in the specification.

And so we ended up, after a few years, with a lot of variants of the web accessibility guideline. It's a nightmare for the industry because the industry has to provide one tool to all the countries. They translate it, of course, but it has to be the same technology behind it. Our members and industry, they push so that we take this accessibility guideline and make them an ISO standard, so that's what we did in 2012. Now, the accessibility guideline for W3C is called Web Content Accessibility Guideline. It's also ISO 4500, so governments can point at that and they don't have to change it anymore. There is less fragmentation, probably no fragmentation at this point in time, so everybody is happy.

There is a real use in transposing some of our standard in ISO, just because ISO is a better brand. It's much older and there are thousands of ISO standards that people are using on a day-to-day basis without problem, so it's less risky for a lot of

governments to also use ISO. So we're good friends with ISO and they accept that we do the web. They don't try to take away our expertise and they're happy to transpose the specs that are very important.

That's how we deal with the de jure system nowadays. Not much with ITU, just the basic. They do things for Internet of Things, and we do things for Web of Things, so we have to look at what they're doing. They have a complementary action on accessibility as well that we look at, because accessibility is not just a technological problem. There's a lot of education and tooling as well.

UNIDENTIFIED MALE: Any other questions from the floor here? Let me do a quick check on our online. In the meantime, I have a question. My simple understanding is that the IETF is currently the caretaker of the HTTP protocol, and the web consortia is everything beyond that, or many things beyond that. What's the line between that? So something like RTC Web, which is I think being co-developed. What determines what lies with the IETF and what determines what lies within W3C?

DANIEL DARDIER: I mentioned HTTP and URI that are really the interface technology. They are done at the IETF. We at W3C have a participation. Historically, you have to understand that the W3C was created as an IETF application area for HTML in probably 1991, something like that. For a couple of years, it stayed as an HTML application area of IETF, and then the industry decided that the web was going to be too big for an application area of the IETF, and they decided to create a consortium on the model of the X Window Consortium.

Not sure how many people remember the X Window Consortium, but it was a consortium, mostly American, dealing with graphical UI on Unix. At that time, the war was Windows-Unix, and Unix didn't have a common platform for doing graphics. It was a problem, so the X Window system was created and it's still there. It's on Linux now, but the X Window consortium has done its job, basically. They had a model of consortia which was different then IETF, giving more flexibility in terms of having a staff that would be completely neutral, not employed by the member of the W3C. it was the X Window Consortium model. I can't state if there are other reasons, but that was the idea.

For that reason only, HTTP and URI, which were at the IETF, they stayed at IETF. We didn't take them, because Tim Berners-Lee thought that it was going fine, and it was not something that

would explode in terms of the technology stack. Whether for the web, we started with HTML and CSS, and now we have 300 specifications that are also on top of HTTP and URI.

Now, when we have a new technology that requires new protocols, that usually requires also an expertise at the TCP/IP UDP level, then what we've done recently for the RTC web – RTC is real time communication web, which is communication between browsers, between hosts – we have split the work between the protocol level work done in IETF group, and the corresponding API in HTML, in the web platform done at W3C, with, of course, the requirement that they are the same.

We work along those lines, and then when we have technologies that are dependent on the network layer, then we go with IETF to do a joint group, usually. We have regular technical coordination with IETF.

UNIDENTFIIED MALE: Any other questions for right now?

UNIDENTIFIED MALE: What time is it?

UNIDENTIFIED MALE: It is 9:20. Do you have more presentation, or was that…

UNIDENTIFIED MALE: Yes.

UNIDENTIFIED MALE: Keep going. You've got time.

DANIEL DARDIER: Specific area. I'll go over the web accessibility because it's the one that I know the best. I started this activity in '97. I always talk about internationalization, because it's of interest to ICANN, I suppose. All the issues, IDN and linguistic security policy, and also I'll talk about some new initiatives that are running at W3C.

What is web accessibility? You see on this picture, that's the basic architecture of the work that we do on web accessibility. The most important thing is the content. The content that you find on a website, that has to be accessible. If that is not accessible, there is not much you can do. Most of our work at the beginning went on defining what is accessible content. What does it mean to be accessible? What does it mean to create accessible content?

It's a difficult topic because accessibility and usability are different things that people usually confound and mix together. Accessibility is really about improving the experience of the user,

regardless of the capacity that they have of accessing a particular media stream. It usually resolves into serializing, being able to serialize the content so that it can be presented in one dimension instead of two or three. If you consider this webpage for instance – it's actually a PowerPoint but I actually had the same thing in HTML. Anyway, you see that there is a logo, a title, there is a bullet list, there is a graphic that is quite semantic loaded. It explains a lot of things.

All that, in the content, it has to be expressed the same way that I just expressed it. There is the title, that's the value of the title. There is a bullet list that explains "Blah blah blah," etc. Once you have described that in HTML, once for instance the navigation structure has been described in HTML, if the graphic here can be summarized into a sentence, you have entered a sentence, then you're saved. Maybe the user agent, the browser that you're going to use is not going to be able to read everything, but you can update or use a proxy to help you.

That's fine, and if you're using an authoring tool, like Dreamweaver or anything to create content, you have to make sure that the tool you're using is going to let you add the data, the accessible data that is missing. The content is the most important thing, and for people that are creating new languages, not just using HTML but at the beginning of the web, it was clear that HTML was not going to be the language for

EN

everything. If you want to express and invoice on the web, you're not going to use HTML. You could, but it would be a poor use of semantics.

If you want to describe an invoice on the web, you create your own language or you use one that exists already that says what is the value of the invoice, the name, the date. You have your own elements that give you the semantics of an invoice. The same thing if you want to describe a plane or anything that can be described electronically. We have a language called XML that you can use to create other language. XML is a metalanguage. It's a grammar that lets you create HTML-like languages.

So for people that are creating new language, then we have to explain to them that if they want to include graphics, multimedia or whatever in their language, they have also to provide the hooks for the author to describe the media in some textual form. Accessibility is for the content in HTML. It's for the content in any language that are using XML, and it's also for the user agent. That is a program that people are going to use to access this data. You have to explain to both sides that if the author has put this kind of information in the content, then your browser should let the user select that information instead of using the image.

So it's a system with an authoring tool, user agent and content that have to work together all the time. Of course, you have added what's called assistive technology that people are using. If you're blind and you're using a web browser, you're going to have either an electronic braille display that converts text into braille, or most of the people now would use audio output. Audio input as well is used for accessibility, but not just for accessibility. The important thing for accessibility is that everything we do for people with disability is used 90% of the time for people without disability.

That's the same deal for the curb cut. You know the sidewalk, when they have a cut on the sidewalk for people with a wheelchair. If you count how many people in a wheelchair pass by this side cut versus the number of people with luggage or the woman carrying a baby, or anything that has wheels, then you'll find that the real use of the side cut is not for the disability people, it's for everyone else. They use it all the time.

It's the same on the web. If you have a page that is accessible, it means that you can redisplay it in a different environment that is better for you, whether or not you're disabled or you have an accessibility problem. If you're in a noisy environment, you really want to see the subtitles of the video that you're looking because otherwise you're not going to understand anything, so

basically you're deaf. In a noisy environment, whether you're outside working or not, you don't have your ear to use.

We try to convey that message, that if website managers take care of accessibility, they're going to help everybody with a better design, simpler design, better navigation, less bandwidth. If your images are described in a way that provide the same functionality, then people can decide not to download the image, period. The image is going to stay on the server. You don't have to download it, you just download the piece of text that is functional for the image.

The idea with accessibility is that is always degradation, of course. We're not going to show a map of France and asking a blind user to select a point on the map to say where they want to book a hotel. They're not going to see the two-dimensioned map, but they're going to see a list of all the big towns in France, which is an alternative version of the map. Basically, the map is what? It's a list of towns which coordinate, and you can provide that with another list. There's going to be a degradation, but the function of the system, which is to have you select a town, is going to be the same in both cases. That's the idea, that it has to be some kind of degradation, but it has to be functional in the end.

I mentioned that the bulk of the work here has been transposed into an ISO standard. There was a real reason for avoiding fragmentation of this particular standard, but for the others, like web service, it was also a fragmentation reason, but it was a different one. So where are we in accessibility nowadays? The main guideline, the WCAG 2, Web Content Accessibility Guideline 2.0, is pretty much stable. It's being now used not just on the web, but for any kind of electronic conversation and communication.

The other two guidelines that explain how to generate pages that are compliant or to read pages that are compliant are less successful because basically, they're not protocol level. They concern either the authoring only or the reading only, so they're good to have, but people can live without them. They're not a network protocol guideline. What else can I say on web accessibility?

I mentioned that this was an area that was heavily funded by government grants, and the reason was clear, that they needed a standard that was stable so that they could use it in their own procurement. Internationalization – we had also this activity forever. When the web started, like the Internet, it was mostly ASCII. That's the way it was in the '80s and the '90s, and then one of the reasons it was ASCII was there was no Unicode at that time. You have to understand that Unicode, which is the

standard that allows us on the web to work in many languages, didn't exist before the web.

It was created basically at the same time as the web. Before Unicode, every country had their own encoding for their own characters. If you wanted to support something like writing Arabic or Chinese, you had to pay attention to the encoding of this country, and countries have different encoding for the same language most of the time. Unicode really gave a level playing field for all the internationalization, but it was not enough.

Of course, it only defined the characters. On the web, you have more than the characters to deal with. You had a lot of things that are related to the presentation of the data. For instance, you look at a date, it's not enough to just translate the month from March to Mars in French. In French, the month and the day are reversed in fact. When you say the date of today is 06 03, or 06 03 or 03 06. You have to know that on the web, so you have a lot of added complexity for internationalization. The direction of writing of course, right to left, left to right, vertical. The date, the checkmark. Some countries have different checkmarks, so everything that we consider on the web has to pay attention to internationalization.

Privacy and security has been a problem forever as well, that we created the web without much security in mind, and now we

have to care for security and privacy everywhere. We work closely with other organizations on that. We don't control the entire stack of the secure network of course, but on the web we can help with some of the aspects. Privacy is mostly important when you dialog with a webserver, that the webserver is taking personal data out of you, either because you're giving it away, you're filling out forms, or just the server knows where you're clicking because you're clicking on its page.

You can track the user one way or the other. Many ways, in fact. So we've decided to create a technology that allows a client, a browser and a server to negotiate the kind of privacy information that can go back and forth between the two. Nowadays, we're working on web authentication to make sure that people try to avoid using passwords but use a more secure technology. We've been working with other consortium that has been working on privacy, security authentication outside of the web. Of course, in the finance industry there is a lot of that when you use a credit card. We have to unify all that using the web.

Let's see, what else. There are several groups that you can look at. Wendy Seltzer was the other contact for [inaudible], and ICANN is in charge of this activity, so you can talk to her at any time. I want to give you a word on policy. Originally, W3C is a technical consortium. We don't deal much with politics, but as it matters, when I first started working at the W3C in '96, the first

thing I had to do was to deal with something called PICS. Platform for Internet Content Selection.

At that time, in '95, the US government wanted to pass legislation that would forbid indecent content on the Internet. That was the term, indecent content. Then the entire industry said "What is indecent? What does it mean?" And so we realized that there is no international legislation on what is indecent, what is nudity or what is sexual abuse, what is pedophilia. If you look at pedophilia, you can see that in some countries, the age of sexual consent is completely different. It would be 13, 16, 21. Also, the display of pedophilia is completely different. In some countries, it's completely forbidden to even draw comics that display children and adults together. In some other country, it's okay to make montage of photos. If nobody was hurt, it's okay.

So the legislation, the social legislation is nonexistent. There is no international legislation for what is child abuse and things like that. So how do we do that on the web? How do we improve the situation? The idea was to create metadata, a metadata language that would let any owner or provider of content describe the content in objective terms. Nudity for instance, you take nudity, frontal nudity is very objective. That you see the front, the breast of a woman or whatnot. That's an objective criteria, and someone can say that you can see front nudity on this photo.

On the receiver end, whether it's an organization, a country or a family, they can decide that everything that has frontal nudity, it's a checkmark, they don't want to see it or the children should not see it. So by using metadata, you had [an indirection]. You let the content provider or someone describe the data along various axes. It can be gambling, nudity, sexual, violence. There is a lot of objective criteria that have been worked on in various societies, so you have to unify with that, and then it's a question of encoding that and letting each party do the filtering that they want.

It's both filtering and censorship, you could say. Once you have the semantic described in the content, you can do filtering for Google. You can say that your employee should not see gambling sites during the work hours, which sounds okay, I would say, as a policy. But you could also say that everything that has frontal nudity should be banned from the country. If you have access to all the ISPs coming into your country and all the content and the web, then you can do that.

Between filtering for harmful content and censorship, there's a fine line. What we do with this kind of semantic web technology, we provide a technology. It's a neutral technology, and if it's used for good, it's good. If it's used for bad, it's bad. That's the only thing we can say. Anyway, that was the first really contact we've had with legislation because we had to explain to

legislators that they could not just forbid indecent content on the Internet, because there is no such thing as a definition of indecent content.

Then I mentioned accessibility, which was also one of the strong reasons for why we've been closer to government policymakers because they really needed an accessibility standard to implement their policy. They have policy that says we have to consider equal treatment for accessibility of disabled user or not disabled user. They want to implement this policy, and they can't. They had to have this standard, and they helped us producing one, which is now an ISO standard, which is good.

Internet governance has been also very active for W3C because we're doing a standard, and standards are part of internal governance. If you have a standard, then you have a way of implementing usually a governance policy. We've been involved in the open standard fight. In 2005, there was a lot of fight around the patent and around what was an open standard. If you remember, you had the OOXML versus the ODL fight for Open Office. All that was not really W3C's problem. It was more on the Office side, but nevertheless, we were involved with the open standard, royalty-free policy matter as well.

Nowadays, there is a lot of debate on privacy, on security, on things that are more high level, like right to be forgotten. Right

to be forgotten is something that popped up a few years ago from Europe, where in France in particular, it's considered to be a right for someone to remove the trace of what they say in the past, providing they're not a public person and there is no public interest in having this data being available.

So if a few years from now I said something that I regret, then in France, I can force the person hosting this data to remove it completely from the web. Well, completely from their site. There may have been copies and everything, but it's a one by one aspect. So we have legislation that impacts on some of the principle that we have on the web, which is that once you put something on the web, it has to stay there forever. We have this very solid notion of permanence on the web. We want to make sure that the links are going to be around in ten years from now, and if we start doing right to be forgotten on a massive scale, then most of the links are going to be 404. It's never going to happen, but that's the idea.

There's a lot of discussion, has been all the time, about deep linking. I can mention a bit that deep linking is something that the policymaker created. There's no such thing as deep linking on the web. Every link is a deep link on the web, but the notion of homepage or portal, which have no technical equivalent. I mean, there's nothing. A homepage is just a link.

It can be w3.org/homepage, or it can be w3.org. It can be anything else. There's no dependency between the syntax of the URL and the fact that it's a homepage or a portal, or something that you want people to go through. Website content providers, they would like people to go through their homepage and then to go through the article below their homepage, but once they provide the article URL, the anyone can make a link to this URL directly. So you have a conflict between a content provider that wants to control the way people navigate, and the web, which does not forbid you to navigate once you have the URL.

Of course, you have access control, that's fine, but hidden by obscurity is bad on the web. It doesn't work. Once you have the URL, everybody has it. You're just a link away from being completely public, so unfortunately there are some countries in Europe where the legislation on deep linking has been successful. People can go to the court and say "This person has made a link to my internal website, although I say on the home page that you cannot do that." Some people are completely afraid by that because you have to read everywhere before you make a link to a page that you know, and that's not good.

What else on policy? Let's see. We're about to start a policy technology working group, interest group to discuss some of these items, because the Internet governance is now very present in our agenda. Our members care about what are we

going to say about such or such topic. Before I conclude, maybe you have questions on this technology or other technology I didn't mention.

UNIDENTIFIED MALE: We have a question in the back.

UNIDENTIFIED FEMALE: Thank you. When you were describing the content filtering –

DANIEL DARDIER: Could you close the door? Because I have a hearing problem. Thanks. [inaudible]

UNIDENTIFIED FEMALE: When you were discussing the content filtering solution, is that the same thing as POWDER, or is that something different? Because I'm aware of POWDER.

DANIEL DARDIER: Yes, POWDER is the new version of PICS. PICS was sort of the initiator project. It was not XML, it was just a LISP syntax that you attach to the page, that would give you a rating for basically four criteria: nudity, violence, sexual content, and I don't remember the other one. So it was very limited with a particular area, and

POWDER is more generic. You can attach any kind of vocabulary, any kind of semantics, and it uses a better format, which is RDF, which is a semantic web format.

A semantic web format, just to give you a short explanation, XML is like HTML. It's a hierarchical format. Everything is in a document. Everything is a child of something. You have a root element, and then you have a section with bullets in it, etc. RDF, which is Resource Description Framework is a semantic web format. It's a graph-oriented format. You don't have hierarchical relation, you have a triplet. You say something is something, or something has something as a particular property. You say Daniel, name is Dardier.

This graph-oriented syntax lets you do much more powerful things. You can say that this page has this rating along this axis, for instance nudity, but you can further say that this statement that this page has this rating along this axis is vowed by this organization. You can say that the French government guaranteed that this page has this value. You can pile up assertions about something that gives a meaning, the quality of the thing.

It's not just about saying that something belongs to this class, but you can really define the semantic of something and actually say something about the statement. POWDER is a language to

describe this thing using the semantic web language, so it's more powerful. It's the next generation of PICS, in that sense.

UNIDENTIFIED FEMALE: Does W3C, do you track and monitor the usage of your standards?

DANIEL DARDIER: We don't really do that. We have a validator, as I said, that would help people providing content to validate the syntax of their content online, or they can download the validator and do all kinds of things, but we don't crawl the web like Google would do and search for valid pages. We don't do that. We don't even check that our members' pages are valid. They may have reason not to be valid, and that's not our problem, basically. We're not a police. It's all a voluntary standard. It's like the root of the Internet. Anyone can point to their own root, but not many people do because they lose the interest of being equal or participating in the same thing.

Same for the web. If people start doing extension to HTML, they have to realize that it may work with one particular browser that maybe asked them to do the extension, but it won't work for most of the people, so it's a challenge. Most of the time, people

would do a programming trick, see what browser is coming in and adapt to them, which is not very good as well.

UNIDENTIFIED MALE: Any other questions? No? Do you have any closing remarks?

DANIEL DARDIER: Yes. Let's see, what's important on this slide? Not much. The risky competition fact. We take for granted that the web is good and it's going to be there, open for all, but as I said, you may have seen that. When you go to a website today, you're asked to choose between downloading an application, a native application on your mobile, or using the website. I would say that there is a tendency of improvement, but if you go to the website, you don't get the same user experience that if you download the application.

There are many reasons for that. The website has to be connected, for one thing. You cannot use it sometimes offline, and since your mobile can do some of the adaptation itself, you can increase the size of the screen pretty much independently of the content, just zooming in. People usually let the desktop website run on the mobile, which is not ideal because the user experience is bad and people would prefer to download the app.

Also, I mentioned the native app on a mobile has access to much more API to do nice things, like direction, location. So we're adding that to the web platform, and more or less, we think it's going to stabilize. There's a limit to the sort of new hardware capability you can add to a mobile. At some point, maybe you'll have a mobile that gets the temperature and things like that, which is not common today, but with the accelerometer, the direction, the geolocation. We think what we're going to converge, and the set of API to do web applications and mobile is going to be similar to what we get on a native platform.

Once we reach that point, we think that most of the programmers will realize the value of doing a web app instead of a native app on a mobile, because you don't have to reprogram it for the iPhone or for the version of Android that comes out, or Blackberry, anything. If you do a webpage and it works well as a mobile application, then that's the only thing you have to do, and usually it would work on all the devices.

That's a lot of work to catch up with the new OS capability, but as I said, we think that we're pretty much close to filling the gap. I mentioned that we have this horizontal activity at W3C, like security, accessibility, internationalization, device independence, that everybody cares about. Everybody wants that their particular technology is accessible, internationalizable, and etc. The work to make all the review of

the specification is something that not a lot of people are ready to pay for. Usually, if you look at a company, they would go into one direction, provide a new technology that is very nice, that they would like to make a standard of it, but in the end, they haven't thought about all these horizontal constraints, such as accessibility, language support, security, privacy.

These are the things that we think should be funded by some kind of organization to do the common good for all these platforms. As much as the automotive industry can pay for the extension to HTML to support HTML in their car system, things that are horizontal in nature are always forgotten, even by the government nowadays. Governments want to fund things that are innovation, and these are not innovation. They are just things that you have to do, so that's one of the messages. Okay, thanks.

UNIDENTIFIED MALE: Great. Well, thank you, Daniel. I appreciate you coming today. Little bit of housekeeping, we passed out a piece of paper. We're constantly looking at how to do this better, so we would appreciate any feedback you have. If you have any ideas of what types of sessions you would be interested in in the future, please write those down. We take your feedback seriously and would like to get as much as we can. We are having repeat of this, too. If

you have colleagues, friends, enemies who might get some value out of learning more about the W3C or any of the other sessions that will be happening today, please tell them because we're doing a repeat of this entire day tomorrow, too. So anyone who might be interested, who didn't hear about us, please tell them and hope that they come tomorrow.

We will also have these slides up on the website in the agenda. They'll be up within the next couple of days, and they will be there in the archives forever and ever, and as long as the Internet lives, so all these slides. With that, Daniel, thank you very much, and we'll see you tomorrow. Our next session here is at 10:30. There is a coffee break right now, and at 10:30 we'll have Alain Durand talking about Internet networking, so we hope to see you guys then as well. Thanks.

[END OF TRANSCRIPTION]