BIBLIOTHECA ALEXANDRINA
مكتبة الإسكندرية

THE WORLD'S LIBRARIES AT OUR FINGERTIPS THROUGH THE NET

Lecture delivered by Vinton G. Cerf

at the Bibliotheca Alexandrina, on 30 May 2004

Alexandria, EGYPT

2005

Bibliotheca Alexandrina Cataloging-in-Publication Data
Cerf, Vinton G., 1943-
The world's libraries at our fingertips through the Net / lecture delivered by Vinton G. Cerf on 30 May 2004. – Alexandria : Bibliotheca Alexandrina, c2005.
p. cm. – (Distinguished lecture series ; 6)
ISBN 977-6163-29-7
1.Digital libraries. 2. Information storage and retrieval systems.
3. Internet -- Computer programs. I. Title. II. Series

004.678--dc21 2005251840

ISBN 977-6163-29-7
Dar El-Kuttub Depository Number 21159/2005

Printed in EGYPT
2000 copies

CONTENTS

INTRODUCTION

An Amazing Legacy

The very name of the Bibliotheca Alexandrina conjures up the image of a glorious past, of a shared heritage for all of humanity. For it was indeed at the Ancient Library of Alexandria that the greatest adventure of the human intellect was to unfold. Launched in 288 BCE by Ptolemy I (Soter) under the guidance of Demetrius of Phaleron, the temple to the muses, or Mouseion (in Greek), or *Museum* (in Latin) was part academy, part research center, and part library. The great thinkers of the age, scientists, mathematicians, poets from all civilizations came to study and exchange ideas.

They and many others were all members of that amazing community of scholars, which mapped the heavens, organized the calendar, established the foundations of science and pushed the boundaries of our knowledge. They opened up the cultures of the world, established a true dialogue of civilizations. For over six centuries the ancient Library of Alexandria epitomized the zenith of learning. The library completely disappeared over 1600 years ago…but it continues to inspire scientists and scholars everywhere. To this day, it symbolizes the noblest aspirations of the human mind, global ecumenism, and the greatest achievements of the intellect.

The Rebirth of the Bibliotheca Alexandrina

Sixteen-hundred years later, under the auspices of President Mohamed Hosni Mubarak, and with the continuous untiring support of Mrs. Suzanne Mubarak, it comes to life again. The Bibliotheca Alexandrina, the new Library of Alexandria, is dedicated to recapture the spirit of the original. It aspires to be:

- The World's window on Egypt;
- Egypt's window on the world;
- A leading institution of the digital age; and, above all
- A center for learning, tolerance, dialogue and understanding.

To fulfill that role, the new complex is much more than a library. It contains:

- A library that can hold millions of books;
- A center for the Internet and its archive;
- Six specialized libraries for (i) audio-visual materials, (ii) the blind and visually impaired, (iii) children, (iv) the young, (v) microforms, and (vi) rare books and special collections;
- Three Museums for (i) antiquities, (ii) manuscripts, and (iii) the history of science;
- A Planetarium;
- An Exploratorium for children's exposure to science;
- Three permanent exhibitions;
- Six art galleries for temporary exhibitions;
- A Conference Center for thousands of persons;
- Seven research institutes covering (i) manuscripts, (ii) documentation of heritage, (iii) calligraphy and writing, (iv) information sciences, (v) Mediterranean and Alexandrian Studies, (vi) arts, and (vii) scientific research; and
- A forum for dialogue and discussion.

Today, this vast complex is a reality, receiving more than 870,000 visitors a year, and holding hundreds of cultural events every year.

The Academia Bibliotheca Alexandrinae (ABA)

The greatness of the Ancient Library resided as much in the remarkable community of scholars that it had helped create as in the vast collection of manuscripts it assembled. They represented the best in the World of their time.

Today, to recapture the spirit of the ancient Museum, we have established the Academia Bibliotheca Alexandrinae (ABA), to include 100 of the greatest minds of the contemporary world. Today, with the magic of the Information and Communication evolution, these eminent men and women can and do reside and work in all parts of the world. The ABA is a "virtual organization" with a small secretariat established at the Bibliotheca Alexandrina in Alexandria, Egypt. The Director of the BA functions as the secretary to the ABA.

The ABA will create and maintain an international network of scientists, artists and scholars dedicated to:

- The promotion of excellence in science and the arts;
- Helping build international goodwill, primarily through collaborations between scientists, scholars and artists;
- Spreading the values of science, and the culture of science in Egypt and the region;
- Fostering openness to the other, through inter-cultural dialogue; and
- Encouraging tolerance, rationality and dialogue.

Beyond the virtual network, a special event of the ABA shall be organized tri-annually. Between the proposed meetings every three years, many activities sponsored by the ABA will take place. Indeed, individual members come and visit the New Library at different times and they and their guests, deliver lectures here.

The Distinguished Lecture Series

In the spirit of spreading the goals and values that the ABA espouses, and the Bibliotheca Alexandrina's commitment to its mission, it was considered appropriate that the Distinguished Lecture Series should be developed to record and make available in an affordable format some of the distinguished lectures delivered at the BA by members of the Academy or their distinguished guests. Thus was the Distinguished Lecture Series born.

There is no specific frequency for the issuing of these publications of the occasional lectures. We would expect no less than three such published lectures to appear every year, and sometimes there will be substantially more. The series is driven by content and quality, not by timing.

In terms of coverage, the scope of the ancient library or the modern BA and its Academy (the ABA), the Distinguished Lecture Series includes science, the arts, politics, and every aspect of the human condition. The only requirement is the rigor of the presentation and the distinction of the lecturer. It is as broad as the human imagination, as varied as the fields of knowledge whose mosaic creates the universal human experience, and as engaging as the talents of the distinguished speakers who bring to life the different topics of their choice. Each lecture stands on its own. It can be appreciated as an experience in its own right. It does not have to be read in relation to any of the others.

It is our hope that by publishing this series, the Bibliotheca Alexandrina is allowing many more individuals to share the lectures than those who have attended the actual event. It safeguards the material for posterity and invites those who are so inclined to view the actual video record of the lecture, which is safeguarded in the Library's multi-media section.

To make the publication more suited for the reader, a special introduction has been included which explains the work of the individual concerned and positions the lecture in relation to that body of work. Each publication also includes a bibliography of selected works and a short biography of the lecturer.

Ismail Serageldin
Librarian of Alexandria
Director of the Bibliotheca Alexandrina
Secretary of the Academia Bibliotheca Alexandrinae

FOREWORD

Often called the “Father of the Internet”, Vinton Cerf has held top positions within such agencies as the US Department of Defense’s Advanced Research Projects Agency (DARPA) and the Jet Propulsion Laboratory. He was the founding president of the Internet Society, and at present chairs the board of the Society. His notable achievements include devising the Transmission Control Protocol (TCP) and the Internet Protocol (IP), and the establishment of the first commercial email service. The recipient of numerous awards and honorary doctorates from universities around the world, Dr. Cerf acts as advisor to various corporations and governmental organizations. Dr. Cerf is also, incidentally, hard of hearing, and is renowned in the deaf community for his work on improving computer facilities for the disabled. He grew up in Los Angeles, California. He is married, with two children.

The Internet entered the public arena in 1994, when Tim Berners–Lee's World Wide Web became accessible through a commercial browser. In the intervening decade, it has become for many an indispensable tool, transforming business and politics, and spawning its own industry of search engines and online shopping. An estimated 60 million computers serve around 200 million users in over 200 countries. However, when Vinton Cerf and his colleague, Robert Kahn, first experimented with connecting computers in 1973, they were struggling to link just three small networks, in California, London, and Norway. The successful demonstration of this linkage was the first step toward the Internet as we know it today.

The notion of connecting systems had first surfaced in 1961, with the development of packet switching technology. This technology was developed under the auspices of DARPA, who had quickly understood the advantages of computers that could exchange information. It was early realized that the most efficient way of passing information was through "packets," or chunks of data, rather than continuous streams. This initial technology relied on the Network Computing Protocol, an inefficient early protocol unable to address a large number of hosts. An effective protocol would need to deal with interference in the transmission channels and be able to translate information to a wide variety of receiving devices. Robert Kahn called upon Vinton Cerf, who had recently finished his doctorate, to create a highly flexible and reliable protocol. In 1973, Cerf developed the TCP, essentially a common language enabling communication between connected computers. In 1978, with the splitting of the TCP into the TCP/IP, Dr. Cerf created a simple protocol employable by gateway computers. While the TCP controlled the flow of data packets, IP addressed and forwarded individual packets. Because the protocol was in the public domain, it was quickly adopted by most systems, and, due to its simplicity and flexibility, remains the primary protocol in use today.

In order to disseminate their knowledge, Cerf and Kahn set up the Internet Advisory Board. They later worked together at the Corporation for National Research Initiatives, developing prototype applications for the Internet, including browsers and digital libraries.

Dr. Cerf subsequently moved into the creation of a globally addressable electronic mail system—email—for MCI. The first email system had been devised by Ray Tomlinson at DARPA, using a program he had developed to pass messages within a single computer and hitching this to a file transfer protocol so messages could be sent to another computer. The first email was sent in 1971, between adjacent computers at the DARPA headquarters. Dr. Cerf was asked by MCI to create a commercial email service. By 1988, the team he led had developed MCI Mail, which Dr. Cerf arranged to be connected to NSFNET, thus allowing it to be used by a wider public. Less than five years later, email services were adopted by major internet service providers such as America Online, and quickly became widespread.

The Bibliotheca Alexandrina has from the outset been concerned not only with printed material, but also with digital information and the complexities of providing access to that information. It houses the only backup copy of the Internet Archive and includes access to the Internet among its services. In areas as varied as archeology and calligraphy, biology and epidemiology, it is involved in preparing CD-ROMs and other digital presentations of information. It was thus an exceptional pleasure for the Bibliotheca Alexandrina to host Dr. Cerf's lecture.

With an admirable cautiousness regarding the staying power of digital technology, and considerable wit, Dr. Cerf speculates in this lecture about the future of the library. Predicting the evolution of the text, he imagines e-books that talk to each other, updating each other as new research augments knowledge. He discusses the efforts now underway to create a Digital Library, which would include not only information presently available in digital form, but also transcribe printed material now in the public domain. Noting that India and China are at the forefront of this latter drive, he envisions the Bibliotheca Alexandrina as a focal point for this initiative.

The Internet has changed the world rapidly, with new applications surfacing almost daily, and perhaps the most exciting portion of Dr. Cerf's lecture explores some of the latest uses of the Net. In anecdotal fashion, he presents technology now in place or being tested, such as Internet-enabled washing machines and refrigerators, and the transformation of highway toll-booths and supermarket checkouts thanks to the RFID—the radio-frequency ID device, which automatically detects credit cards and deducts money from a bank account.

Dr. Cerf concludes with a discussion of his present work, which is concerned with creating an Internet protocol for space. He details the technical difficulties involved and explores the ramifications of a communications system for space. This project will, he hopes, culminate in the launching of a satellite bearing this new Internet protocol.

In the last decade, the Internet has profoundly transformed the world. At times humorous, at times visionary, but always profound, Vinton Cerf in this lecture imagines the ways our lives may continue to be affected by the Internet, and hints at the role the Bibliotheca Alexandrina might play in this ongoing adventure.

PERSONAL
PHOTO

THE WORLD'S LIBRARIES AT OUR FINGERTIPS THROUGH THE NET

Salam! Unfortunately, this is the only Arabic you are going to hear from me today. First, I have to thank Hisham al-Sherif*, as well as Dr. Serageldin, for a wonderful introduction and a wonderful welcome on my first visit to this city and to this Library.

It is a personal point of honor for me to speak at the Bibliotheca Alexandrina. When I think of libraries I always think of the Library of Alexandria, which was a

---

*Dr Hisham al-Sherif, Chairman, CEO, IT Ventures and member of the Pittsburg prize arbitration committee, together with Dr Serageldin, invited Vinton Cerf, who is also member of the same committee, to give a lecture at the Bibliotheca Alexandrina.

focal point of the ancient world's knowledge, and I think we have an opportunity with this new wonderful Library to be the focal point of 21st century knowledge. Therefore, I hope one outcome from today's discussion will be that we can move forward in that direction. A question asked was: "Where would we like to be in the year 2020?" My answer to that is, "Alive!" especially considering how long I have been around and how many years we have put into the Internet. My task this afternoon is to try to convey some ideas to you about what the world can be like, and what it is becoming, as more and more information arrives in digital form. Therefore, the title of today's talk is "All the World's Libraries at our Fingertips through the Net".

The Internet: First Decade

I would like to start by sharing some reactions with you about what I think of when I think of Egypt. Now I am sure that this is colored by my history and background and everything else, but I can tell you that even as a child the history of Egypt fascinated me. As I have grown up, and as my children have grown up, I have noticed that every child seems to be fascinated by the Pyramids, mummies, and the Sphinx, and so we should take advantage of this. Every child everywhere in the world is fascinated by that particular part of your history. The reason this is so important is that once we attract children's attention, then that is an opportunity. We want to teach them to read, to learn more about what they are interested in, and to open up their interests as broadly as possible. I think that because of this special opportunity that today's age affords us, highlighting the opportunity to put information online, this means that we can make the Bibliotheca Alexandrina the focal point of that initiative. The Library is really an icon in history, and it can become an icon for today.

I want to move to a bit of statistical information about the Internet, just to give you some sense for how it is growing. This is a seven-year comparison, so we are going from mid-1997 to early or mid-2004. During this period, the number of servers on the network, the machines that do Web services or email and the like, grew by over a factor of 12. This does not include things like personal digital assistants and the like, things like the blue tooth I am wearing: these devices probably add another 300 to 500 million items to the network. The number of people on the Net is not a known quantity. There is no one place where everyone has to register, so we do not really know who is on the Net. However, the estimates are that there are at least three-quarters of a billion, possibly as many as one billion, users.

While on the one hand that is a big number, on the other hand it is still less than a sixth of the world's population. So suddenly we realize that, as big as the Internet is, it is still small compared to the population that should have access to it, but does not yet have such access. Just for comparison, the world telephone system has grown by over one billion terminations over the last five years: it had about 1.2 billion terminations five years ago, and now it has 2.3 billion, about one billion of which are wireless or mobile devices. Sometimes you probably feel as if all these are right here in Alexandria! However, the point is that these devices have expanded the telephone system dramatically. They are likely to expand the Internet dramatically as well, because many of them either already are, or will later become, Internet-enabled devices. We will revisit that idea a little later.

I want to tell you a story about my personal library. I have a few thousand books; it is not a big library, but I can tell you that when I am looking for something, and I do not know in which book it is, then when I stand in front of my bookshelves I still think I do not have time to look through every book to find what I am looking for. Even if I know which book it is in, I still may not remember what page it is on, and the book might not be indexed. And so I stand there thinking: I love books, but right now I wish they were all online because I would be able to get help searching for the information I want. I really look forward to the time when all the publications that are available in print now are also available online and in searchable format.

There was a time when I was about ten years old when I wanted to do just this. When I was ten, the world was young and dinosaurs still roamed the planet… We did not have the Internet, and we did not really have any computer technology at all, and the only thing that was available was microfilm. I had a smaller book collection then, but I used to be fond of reading books lying down in bed and holding them up above my head. The problem was that these were hardback books, and when I fell asleep, the books would drop and hit my nose, painfully. I wanted to have my books moved over onto microfilm, so I could project the books onto the ceiling. This was in about 1953. I went to the Kodak Company because they made microfilm, and I asked them how much it would cost to have my 500 or so books microfilmed, and they said US$3000. Well, my allowance was 50 cents a week, and so I was not in a position to take them up on this offer. Some 50 years later, I am still eager to find a way of getting my library online, but it will not be with microfilm.

For many people the Internet has always been around. If you are no older than, let us say, 12 years old, there is no such thing as a world in which there is no Internet, because it became visible to the public in 1994. On the other hand, the online searching technology is actually relatively new. I did not get introduced to this in any very profound way until 1992, and I have been using the Net as far back as 1970s. However, even in 1992 the idea of putting information online and having it searchable was still in its early stages.

In 1992, I was president of the Internet Society, and I was going through my email when I found an email from an irate Slovenian complaining bitterly that the suffix ".si" had been taken or reserved by the United

States for something called the Spratly Islands, and he wanted me to fix this so the ".si" suffix could be used for Slovenia. My first thought was: I wonder what the Spratly Islands are, and why is the United States interested in reserving ".si" for them? I did not have the answer to that, so I put the message aside. The next message was an invitation from the University of Minnesota, saying "Why not go to our Web site and try out the Gopher application?" "Gopher" is a double pun, because in English a "gopher" is a little rodent, and this is the mascot of the University of Minnesota. However, gopher also sounds like to "go for" something, and the application essentially allowed you to work your way through a series of menus on the screen.

The Gopher application was not a search engine of the kind that you are familiar with today, such as Google or Alta-Vista. It was just an online information system that you could work your way through in a hierarchical structure. So, I decided to try it out, and I went through and found something called "CIA World Fact Book." I remember panicking, thinking, "Oh my God, they have put classified information on the Net, and we are all going to be in trouble." However, it turns out that this is an unclassified publication, so my next reaction was, well if anyone knows what the Spratly Islands are, the CIA should. I looked them up in the World Fact Book, and discovered that they are unoccupied islands in the middle of the South Pacific, and that they have three airfields and some fortified places, and that nobody lives there. There was no arable land of any kind; there was no land devoted to living space or anything; there was 100% "other" land (whatever that meant), and there were large quantities of guano on the islands. So, I thought that maybe that was why the USA was interested in the islands: they were a source of fertilizer. It turns out that the islands also sit on some important sea links, and the USA also had an interest in those.

My next step was to ask the US State Department why they had reserved ".si" for the Spratly Islands. I never found out the answer to that, because a week later I got a thank you note from the Slovenian saying, "Thank you, the ".si" has been released and is now assigned to Slovenia as its country code on the Internet." I thought at the time that I get blamed for a lot of things that I have not done, but I sent him a note back saying, "You are very welcome, I am glad I could be of some assistance." I still have no idea what happened. But that was my first introduction to an online information system of any significance, and of course not long after that, in 1994, Tim Berners–Lee's World Wide Web burst onto our scenes in commercial form, when a company called Netscape Communications launched a browser that had been built by Marc Anderson and Eric Bina at the National Center for Super Computer Applications in Illinois in the United States. This browser was Netscape Navigator.

The Search for Digital Objects

This history brings me to a general point that I would like to introduce, and which I would like to convey the importance of to you. Today, we are beginning to see an increasing amount of digital information on the network. Some of this is in very well-structured form, but some of it is not. We sometimes speak of this information as "digital objects", which are objects that have a well-defined structure and are understandable to computers as content. The content is actually encapsulated in the middle of software, and you can think of these digital objects almost like viruses. If you are familiar with how a virus works it may be easier to understand the idea of digital objects.

A virus is a bit of DNA wrapped up in a protein coat, and we can think of digital objects as information wrapped up in a coat of software. The external software coat has standard interfaces, through which the object can be requested to do something with its content. For example, you can give the instruction, “Show me your Arabic rendering”, or “Show me your French rendering”, or “Show me your English rendering”. Sometimes you may get the answer back, “We do not have one”; sometimes you might get the answer, “Here is our best estimate of that rendering”: you may have gone through an automatic translation system, of which there are now several, and these produce varying results. But the main idea is that you have a standardized ability to interact with this digital object in order to get its content out in a form that you can use.

In some cases, this content might be rendered as sound. For example, you might have a musical composition stored away, and when you ask it to render itself as sound, what you get back is an orchestral rendering of the music. However, you might equally have asked it to provide you with its print rendering, and in this case you get sheet music from the same object. The idea is that these objects can be represented in a variety of different ways, can be presented and rendered in a variety of ways, and it is the software surrounding them that produces these different presentations. Of course, once we start producing these digital objects, we have the problem of where to store them and how to find them again. In fact, this is the beginning of the reintroduction of the concept of libraries, except in the online environment, we have digital objects not physical ones. However, these digital objects, just like their physical counterparts in the past, have to be catalogued and indexed so we can find them again. Instead of using a card catalog, we use computer programs like the search engine Google to do this. Remember that here we are talking about deliberately-structured information.

I want now to talk about another kind of information, which is the kind we have most of on the Internet. Today, we have a standard representation of content on the Net, this content showing up as Web pages and being typically encoded either in hypertext markup language (HTML) or in extended markup language (XML). This means that the information is not entirely structured, but that it does not have a rigid format either, and it has a lot of tags on it to help us decide how to present the information. In the case especially of extended markup language, these tags help a computer to understand what the content is: it could be an invoice, for example, or a bill of some sort. If so, then the material is tagged in such a way that the computer knows that it is a bill just by looking at it, and it knows that you are supposed to pay it, and it knows it should ask you for your credit card to do so.

The result of having this standardized representation in HTML or XML has led to the creation of things called web crawlers that go through the entire Internet, looking at as many pages as possible. In the case of the popular Google search engine, for example, the crawler will go through some 5 billion pages, which is only a fraction of all the content on the Net. Those of you who are familiar with current physics will know about things called “black holes,” stars of so much mass that their gravity field prevents light escaping from them. I think that there is a lot of “dark information” in an analogous way on the Internet today, sitting in databases that are not represented on the Web pages searched by Google or other engines, or sitting in databases that have to be “quarried” to extract information.

Often, when you are doing some sort of Web search, or interacting with a Web page, what is actually happening is that as you interact, the corresponding Web pages are being generated by information from the relevant database producing the rendered Web page and then presenting that information to you. This is where the “dark information” on the Net is held, because this information is not visible to the search engines. Some estimates suggest that there could be as much as a thousand times more information in that form on the Internet than there are Web pages. So we are only scratching the surface of today's information content on the Internet, even with the immensely useful search engines that we use today.

One possible way of accessing such "dark information"—information that cannot be easily searched—is by associating it with information that can be easily searched. An example of this is what happened in the United States as a result of the Americans with Disabilities Act.

According to that law, television programs had to be captioned so that they could be understood by the hearing impaired. I happen to be hearing impaired, so I was very happy to see that law passed. However, people at the MIT Media Lab then came up with a very interesting idea: they realized that news programs were being captioned as the law said they should, and they recorded the programs on video and separated out the captions as text, using this text as an index for the video. The result was that if you were interested in commentary on a particular subject that might have occurred in a news program, you could search the captioning information, which would then take you to the appropriate frame and allow you to find the point in the program that interested you. Google has a search tool that lets you find images on the same page as a text you are looking for, associating these images with the text. Sometimes, of course, the image has nothing to do with the text you are looking for; it just happens to be on the same page as the text. Often the image is quite relevant. The result here was the development of a kind of media confluence, using indexable material as a way of finding things.

The most recent introduction of unstructured information on the Net has been in the form of what we call "weblogs". These are like diaries, in which people share their daily thoughts with anyone who might be interested. Some of us might wonder why anyone would be interested if you look at some of the weblogs that are currently online. Nonetheless, this kind of information is accessible by carrying out text searches, and this kind of unstructured content is becoming accessible to us in a way that it could not have been in any other environment.

The reason this is important is that in some cases these weblogs may be useful to us, if we are carrying out research, for example. If the weblog is written by someone, a biophysicist or biochemist, for example, who is using it to make notes about his work, then the weblog can lead others interested in the same kind of work to him or her. The Net has thus become an avenue for linking people together who have common interests, which is another important aspect of online information.

Tim Berners-Lee and the Semantic Web

Today, Tim Berners-Lee, the inventor of the World Wide Web, has another very powerful notion. I should say "Sir Tim Berners-Lee", because he was knighted about a year-and-a-half ago. Tim invented the World

Wide Web around 1989, when he was at CERN in Switzerland, and today he is enthusiastic about what he is calling the "Semantic Web".

What he is getting at here is the idea that proper labeling of the information on the Internet can help us search through it more carefully. To give one example: if you are looking for books by a certain author, you could conceivably just type the author's name into the Google search engine or the Alta-Vista search engine and then see what comes up. However, if the information on the Net included the kinds of tags that you can have with extended markup language, then there might be tags associated with an author's name that say: "This is an author's name". This constitutes the meaning of this text, as opposed to just a reference in it to a person's name. By tagging this information as a reference to an author, the search engine will only look for, only respond to, and only tell you about those pages that contain that name with the tag "author" associated with it. This will make Web searches much more effective.

When you search the Web today, you typically receive ten pages of responses, many of which are not relevant. The point here, based on Tim's idea, is that by using tags, and especially by using extended markup language, the codification of searches can be made much more effective. This also has a side effect, which is as important as anything I can think of, because you can also use these same tags to represent information in business documents and in scientific data, or in other kinds of useful exchange. Since these things can be tagged, or recognized, as business documents, for example, a computer can understand that what it is looking at is an order, or an invoice, or a payment confirmation, or a bill, or some other kind of business transaction, and not something else entirely. All the information is tagged, so the software knows what it means semantically.

This system could be taken further, and one can imagine using the same structure of carefully tagged information as a way of controlling scientific instruments through the Net. Once again, the idea is that the information contained in commands given to the instruments is labeled, or tagged, such that the remote computer that is actually controlling the instrument understands what it is you want to do. Of course, when data comes back, let us say, from a microscope or a sensor system out in space, we can use the same labeling system to mark the kind of information that is sent back. As a result, when we analyze this information we have a good tool at our disposal to understand the information received.

Side-effects of Online Information

Clearly, there are already a lot of important side effects of putting information online. One of them is already very, very clear, and that is the accelerating pace of scientific understanding and discovery, particularly when it comes to genomics, the study of human or other genes. We have been able to learn more about genetic information more quickly as a result of online information, because everyone publishing information in paper journals has been required to put the same information into one of the several human genome databases, or other genomic databases, making this information accessible to the entire scientific community. A side effect of this insistence that data be made available in machine-readable form in a timely way was that scientists everywhere had a rapid ability to discover whether anyone else had seen the same gene sequence that they had just discovered, for example, or whether anyone else knew more about what they did, or what it was for, or what it affected.

Today, of course, we have successfully sequenced the human genome, and we have sequenced other genetic sequences from other animals. Now we are at the point of trying to work out what all this genetic information really means. One of the things it means is that for those genes that are expressed, that produce proteins, hundreds of thousands of different proteins, these genes interact with each other in animal physiology. Working out the ways in which they interact is a much more complicated problem than figuring out the gene sequence. In a sense, all that we have done thus far is to climb to the top of the mountain of genetics, but now we need to climb a much higher mountain to understand the resulting proteome, or the collection of proteins that genes express. I expect to see much more rapid evolution of our genetic understanding as a result of climbing this second mountain.

A further development happening today is that universities across the world are starting to share information online. MIT, for example, took an enormous step forward by placing all its technical information and educational materials online, and making these fully accessible to anyone anywhere in the world. Some university presidents thought that MIT was out of its mind to do so—after all, this is valuable intellectual property. MIT had concluded that the fastest way to attract the best students was by saying: "Here's the material we teach. If you understand this already, you don't have to come to MIT, but if you don't understand it, this is the place to come to". They had a very positive response back as a result, and in the course of this advertising initiative, if you like, they have shared wonderfully valuable tutorial material with the rest of the world.

Today, we are increasingly communicating online, using email incessantly—so do the spammers, unfortunately—and this is going to be a never-ending babble. We use instant messaging, and we use other kinds of presence-based communication. Our mobile phones, for example, can now be programmed to say whether certain of our friends are online, so we can know whether it is worth the trouble of calling them. We can even imagine having documents that are malleable, that are active, where you can interact with these documents. I expect to see this technology increase the use of e?documents in business and in government. Many of us interact with government officials, and we would like to do this online, so that we do not have to stand in line physically, or stay on the phone listening to music before someone replies. Online, indexable, machine-readable information should improve our ability to conduct our lives and businesses online.

Standard Identifiers of Online Information

However, as we produce more and more digital objects on the network, we need a reliable method of identifying them so we can find them again. Today, there is a massive initiative by a number of groups aiming to design such standard identifiers for digital objects on the network. Anyone who has ever purchased a book will have noticed something called an International Standard Book Number that is associated with it. Periodicals have similar kinds of identifiers. These can be used as handles, so to speak, for a digital version of these things. My partner Bob Cohn and I have spent quite a bit of time working on a system called the “handle system”, which allows arbitrary strings associated with digital objects to be assigned to them, allowing us to find these objects again wherever they might be.

What is interesting here is that not only do you want to find the object so you can use it again, but you may also want to make reference to any copyright controls there are on that object. Certainly the media, the music community, and the movie community, are very interested in finding ways to make information digitally available, but at the same time to constrain and restrict that information’s reproduction and distribution. These kinds of controls can be put on digital objects, as a way of helping the industry protect its information. I am sure that many of you are familiar with digital signatures, which are a useful way of ensuring that an object has not been altered, as well as a guarantee that it comes from the source it says it does.

Aside from indexing the information, however, another real concern is the increasing amount of digital information we have and our need to put it somewhere. We have a variety of places to put it: disks, tapes and things of that sort; as well as CD-ROMs and DVDs. However, there is an open question about how long these media will actually last. For example, I am very fond of a little new medium that I carry about with me. I have a digital watch, which does not look digital since the face is analogue, but nevertheless has a USB connection and contains 128 megabytes of memory. I carry my Power Point presentations and the spreadsheet containing my wine list and my science fiction bibliography on my watch. However, I do not know how long the information will stay on the watch, so although I am a huge enthusiast for digital information, and always push the idea of putting things online, I have to admit to concern about storage, which I can illustrate by telling a story.

I was once at a library, a specialized library for really old documents, talking to the head librarian about putting things online and making them digitally available. I was saying how wonderful the new technology was, and I was demonstrating some CD-ROMs I had brought with me to show the latest high-density digital technology. The librarian looked at me, excused herself and went into the stacks, and then came back with a vellum manuscript, an illuminated Bible dated 1000 CE, and therefore 1000 years old. She put the book down on the table and said: “Would you tell me again: how long do you think this digital technology is going to last?” I had to tell her I did not know how long it would last, since we certainly had not had a thousand years of experience with a DVD or a CD-ROM. She put me in my place.

This story tells you that if we intend to digitize information we had better get accustomed to moving the digital content from one medium to another and copying it as necessary in order to make sure that we really do retain it. With today's copyright laws it is important that we do retain information, because copyright laws, the Berne Convention in particular, say that copyrights last for 75 years after the death of the author in question, with the result that if you have a youngish author of age 30, let us say, there are around 45 years of life, plus another 75 years, to be copyrighted, which is 120 years that we have to preserve the content to make sure that copyright is observed. So far, no one has demonstrated any digital storage medium that lasts that long, except for paper. I sometimes have the horrible thought that we will go back to punched paper cards in the 21st century as these have been shown to last longest. This would be a terrible outcome, and I hope that some of the newer technologies will be shown to last longer.

Digital Libraries and the Future of the Book

A further project underway today is the so-called "Digital Library", which by now is a generic term. Many universities in the USA are involved in this, as are some of my colleagues on the Petersburg Price Committee. One of the things encouraged in the creation of digital libraries is not only taking information naturally produced in a digital form, using word processors and the like, or material produced as output from some data device, but also transcribing material that is now in the public domain into digital form, making it accessible in that way. Two countries that have been instrumental in producing this material are China and India, because the labor involved in transcribing some of this material is significant and the costs there are reasonable. As a result, we are witnessing an increasing amount of material developed in this way, and of course it would be wonderful if the Library of Alexandria became a focal point for this kind of initiative.

Indeed, I occasionally speculate about the future of the book. Today, we think of books as rather static things: you open them up, and page 25 of a given book still has the same content this week that it had last week, and last year, and ten years ago. However, as we start moving towards digital versions of these things, it would be interesting if books could start to talk to each other, to update each other, so that the next time I publish a new version of something, for example, all references to it can go to the new version, pull the new information out of it and update all previous versions.

Marvin Minsky of MIT has speculated that a hundred years from now we might be saying to each other: "Did you know there were once books that didn't talk to each other?" I sometimes worry about what it would be like to live in an environment where books keep changing, because I am fond of returning to the pages I remember, and I sometimes remember where information is in this way: it is about half way through the book on the right-hand side of the page in the upper half, where the mustard stain is… A dynamic, digital book may not be memorable in the same way as a traditional one, so we will need more tools than we have today to keep track of what books are doing to each other.

The Future of the Internet

I am going to shift gears a bit here, because now I want to give you some examples of what I anticipate is going to happen to the Internet in the future. One thing that has most surprised me over the last 30 years has been the discovery of things on the network that I never thought would ever be a part of it: things like refrigerators with liquid-crystal displays on them, or picture frames that are actually Internet devices and have storage for 20, 30, or 40 images and step through them however fast you instruct them to do so. Even telephones are becoming part of the Internet environment. As a result, I have been speculating about what the world is going to be like, and I would like to share these speculations with you.

Most of the future devices will be programmable and will use high-level languages like Java and Python. Some such devices are already in operation, like Web TVs. For example, I was in New Zealand a year ago, and I discovered that in my hotel room there were four different ways to get onto the Internet: there was a telephone jack that you could use with a modem; there was an RJ48 Ethernet connection you could plug into; there was a Web TV with infra-red keyboard that you could use to surf the net; and if I held my laptop at just the right angle next to the window, I could pick up a hot-spot down the street in an Internet café. I have to admit that I could not type on the laptop very well at that angle, but that was a fourth way of getting onto the Net. We are witnessing more and more of these kinds of devices showing up in a networked environment.

Video games are also becoming Internet-enabled, and this has led me to another interesting speculation. The kids that play these games like to be in the same room together, screaming and yelling as they shoot at each other on the screen. However, sometimes they cannot be together, and they have to be in different locations, or in different cities. What is now happening is that when these games are put on the Net the kids can still play together as a group even though they are no longer in the same physical space. Some of the devices have films as background, and of course they already have speakers so that you can hear the sound effects of the video game and talk to one another while playing. This turns out to be important, because these game devices have about 20 buttons and no keyboard, and there is no time to type or anything, because the player is too busy pushing buttons to shoot, or whatever it may be. Therefore, players use voice information to coordinate their moves.

It occurred to me that with an inexpensive television set, or rather with a small television camera, you could put the camera on top of the television set and pick up an image of the other player, enabling you to see your friends and talk to them while playing the game. That would be interesting because you could have a collection of people physically separate from each other who could nevertheless see each other, hear each other, and shoot at each other, which sounds like a good definition of a video conference! Therefore, I am beginning to think that what will happen in the business world in the future is that these video games, which are inexpensive consumer products, will be brought into the business environment, plugged into high-speed business local networks, and used for video conferencing. Of course, if the video conference gets boring, you can always go back to playing the video game.

Unfortunately, I do not have the time to go through every example of what may happen in the future. However, it is surprising, for example, that there are Internet-enabled washing machines being test-marketed today: college students like them, because you can put your clothes in and go off and have a beer and the machine will send you an email, or page you, to tell you that your clothes are done. You no longer have to wait watching your clothes turning round in the little window. There are also commercial refrigerators that are Internet-enabled, and these have a touch-sensitive liquid-crystal display. I do not know how it is in Egypt, but in many American homes the family communicates with pieces of paper stuck on the fridge with magnets. Now we have been able to improve family communication by using email on the refrigerator.

Another thing happening in our part of the world is the introduction of a new device called an RFID, a radio-frequency ID device. These are now frequently used on automobiles, so that when a driver goes through a toll-booth an automatic payment can be made by the device. These devices are radio-responsive: in other words, when a radio pulse is sent to them, they send back whatever the customer's account number is. There are businesses that want to put RFID slips in the packaging of products in grocery stores: this way, customers will be able to get through the checkout faster. If there is an RFID device in every package, the customer simply goes through the checkout and the machine will read the radio information of everything in his or her basket and then calculate the total amount owed.

What would happen if a refrigerator had the same ability to read RFID packages as supermarket checkouts? If it could read these packages, it would know what it had inside it, and while you were away at school, for example, or at work, the refrigerator would be able to search the Net for recipes using materials that it had inside it. As a result, when you came home the refrigerator's liquid-crystal display would have a list of suggested dinner menus for you. Some people have suggested that this is not going far enough: the refrigerator should also automatically cook the recipe for you!

Imagine also that you are away on holiday. You get an email from your refrigerator saying: "Hello, I don't know how much milk is left, but you put it in three weeks ago, and it's going to crawl out on its own if you do not take it out soon." Or maybe you are shopping at the grocery store, and your mobile goes off. It is your refrigerator again, this time saying: “Don’t forget the marinara sauce. I have everything else I need for the spaghetti dinner tonight”.

On this theme, I am sorry to have to tell you that the Japanese have spoilt things for us by inventing an Internet-enabled pair of bathroom scales. The idea is that when you step on the scales your weight is sent to your doctor, becoming part of your medical record. This is probably OK, but what happens if the refrigerator gets the same information? This might mean that when you come home in the evening you only see diet recipes coming up on the display. Or maybe the machine will just refuse to open, because it knows that you are supposed to be on a diet.

One last thing I want to mention to you as a possibility for the future is Internet-enabled clothing. This is not a joke: when astronauts are sent into space, they wear clothes that have sensors in them to keep track of their vital signs. This same technology could be extended to earthbound medical conditions, where it is important for doctors to be able to monitor the patient. Instead of the patient being tied to his or her hospital bed with wires, that patient could be moving around in the normal way, the information necessary being captured from the clothing he or she is wearing and transmitted, under suitable protection, to the doctor or hospital.

I sometimes think about what the world would be like with Internet-enabled socks. The first thing I would do is to interrogate my sock drawer. I could send a question to the sock drawer saying: "Please report," and it would send back a report saying: "There are 17 matched pairs of socks, except one sock is missing: 144L is not in the drawer." I could then send a radio multicast around the house and get a response from sock 144L: "Hello, this is sock 144L. I'm under the sofa in the living room." We have now solved the problem of the missing sock: an enormous contribution to society.

The Interplanetary Internet

At this point, it is only fair to ask what else will be at our fingertips, now that it is possible to have the contents of the world's libraries online and in searchable form. Here is another piece of speculation that comes from straightforward engineering and has to do with the expansion of the Internet across the solar system. Many people laugh when they hear me introduce this idea, thinking that it is just a joke. However, it is not a joke: it is actually a project I started at the Jet Propulsion Laboratory in the USA six years ago. It is a serious piece of engineering, because we need standards for space communication for the same reasons that we needed standards to create the Internet in the first place. I want first to acknowledge the other engineers who have been working on this project with me, and then to give you an idea of what is currently happening.

You probably know that over the last four years we have been exploring the solar system by sending robotic devices out into orbit around the planets. The deep-space network antennae used to communicate with these devices are 70 meters wide and can communicate with spacecraft that are either flying past the planets or are in orbit around them. There are three such installations of deep-space network antennae: one in California, one in Spain, and one in Australia. As Earth rotates, these antennae look out into the solar system, the three locations being arranged such that at a given time at least one of the three can see the spacecraft we need to communicate with. What you may not know is that in the past every space mission tended to have a communications system tailored to the instrument package on board the spacecraft. These communications packages were not standardized, and when a new mission was launched it frequently could not reuse the communications assets of earlier missions. Our first target, therefore, was how to standardize communication, such that when a new mission was launched any existing communications capabilities out there could be used to help support that mission.

The planet Mars has been a focus of attention for some time now. The Viking system landed on Mars in 1976, and more recently, in 1994, the Pathfinder mission sent back quite a few very interesting pictures of Mars. More recently still, two rovers have landed on Mars, and these are still functioning well beyond their anticipated lifetimes, producing enormous amounts of information. Some of this is stunning, because it has been shown that there was once liquid water on Mars, which is very significant because it means there might at one time have been life on the planet. Perhaps there still is life on Mars in some parts, only it is buried underneath the sand where the water has turned to ice. Perhaps Mars could sustain life, and this would be very exciting if we ever sent astronauts to Mars.

What we are doing in the course of our work today is to design an Interplanetary Internet, really a network of Internets, since we discovered very quickly that the standard TCP/IP protocols that make the earthbound

Internet work do not work in outer space because of the distances involved. The problem is that between Earth and Mars, for example, the delay due to the speed of light ranges from 5 minutes to 20 minutes depending on where Earth and Mars are in their respective orbits. This means that it takes between 10 minutes and 40 minutes for information sent from Mars to reach Earth. TCP is a protocol designed to do what is called "flow control", among other things. This means that when you cannot take any more data, you say: "Stop, I can't take any more data." On Earth, it takes no more than a few hundred milliseconds for a stop request sent out from one machine to reach another, and that is quick enough to ensure that no more data flows until the receiving machine has had time to adjust and can accrue more.

However, if the machines are 20 minutes away at the speed of light from each other, and one says: "Stop, I can't take any more", the machine on Mars will still transmit data for an additional 20 minutes. Where will that data be stored? It is raining down from the sky, and there is nowhere to put it. This means that the basic TCP/IP protocol does not work for interplanetary communication, and we have had to invent a whole new set of protocols. In future, on the surface of planets and in spacecrafts, we will apply standard TCP, but between planets we will apply a deep-space network providing long-haul communication, and we have invented a new set of protocols that will allow the long-haul system to function properly.

To give you a sense of how we have had to rethink our normal intuitions about protocol design for interplanetary communication, think for a moment about what it means to communicate with someone on another planet. One problem is that the two planets are rotating, so if you do not send your information out at the right moment by the time radio signals sent from one planet reach the destination planet the receiver has rotated to the other side and you cannot communicate.

Imagine trying to throw a baseball to someone on a satellite orbiting Mars: if you do not throw it at just the right moment, the ball will fly by before your receiver gets to the point where it is caught.

This problem caused us endless trouble, and we had to invent a system very similar to email. When you send an email message to someone else, you do not know if that person is on the Net at the same time as you are, and you do not really care because your mail is stored and forwarded and then held until the recipient is ready to read it. In the new program, "the system for interplanetary communication", the same idea is used. The full stack of protocols for the interplanetary Internet we hope will be on board a 2010 orbiting satellite designed specifically for telecommunication support on Mars. Thus, by the end of this decade we hope to have an interplanetary Internet in operation.

What does all this have to do with libraries? That is simple: libraries will be the source of enormous amounts of important scientific data we want very much to make available to all the scientists of the world. To be available, it has to be in a digital library in a form that we can find again, and in a form that is sufficiently self-describing; that is to say, it has to be in a form that in 10, 15, 20, or even 100 years from now will be understandable to the scientists of the day in case they want to re-analyze it. More, once we have found information that we did not really understand except 10, 20 or 30 years after its collection, and we needed to go back to the original data source to re-analyze it to discover what it had to tell us.

This is just some of what the next ten years or so holds for us on the Internet. Beyond that, you may have the answer, and I probably do not. Therefore, I am waiting to hear what your ideas are.

BIOGRAPHY

Vinton G. Cerf, one of the world's foremost computer scientists and innovators, grew up in Los Angeles, California. After receiving his doctorate from UCLA, Cerf worked for IBM before joining a program at the US Department of Defense's Advanced Research Projects Agency (ARPA). Cerf worked with Robert Kahn at ARPA develop the protocols used to link computers, a technology which would become the foundation for the Internet. He subsequently joined MCI, where he created the first commercial email service.

Dr. Cerf sits on numerous committees and boards and has received awards and honorary doctorates from institutions around the world. He is currently working in NASA's Jet Propulsion Lab on the Interplanetary Network, essentially developing an Internet protocol for space.

Dr. Cerf is hearing impaired, and has received awards for his work on providing technological services to the disabled. He is married, with two sons.

Honorary Doctorates

Swiss Federal Institute of Technology (ETH), Zurich.

Lulea University of Technology, Sweden.

University of the Balearic Islands.

Palma Capitol College, Maryland.

Gettysburg College, Pennsylvania.

George Mason University, Virginia.

Rovira I Virgili University, Tarragona, Spain.

Rensselaer Polytechnic Institute, Troy, New York.

University of Twente, Enschede, the Netherlands.

Awards

- Marconi Fellowship
- Charles Stark Draper Award
- Prince of Asturias Award for Science and Technology
- Alexander Graham Bell Award
- NEC Computer and Communications Prize

- Silver Medal of the International Telecommunications Union
- IEEE Alexander Graham Bell Medal
- IEEE Koji Kobayashi Award
- ACM Software and Systems Award
- ACM SIGCOMM Award
- Computer and Communications Industries Association Industry Legend Award
- Yuri Rubinsky Award
- Kilby Award
- Yankee Group/Interop/Network World Lifetime Achievement Award
- IEEE Third Millennium Medal
- Computerworld/Smithsonian Leadership Award
- J.D. Edwards Leadership Award for Collaboration
- World Institute on Disability Annual Award
- Library of Congress Bicentennial Living Legend Medal
- US National Medal of Technology

SELECTED BIBLIOGRAPHY

Cerf, Vinton G. 2001. Beyond the post-PC internet. *Commun. ACM* 44(9): 34–37.

Cerf, Vinton G., and Alfred C. Weaver. 1991. Computer networks: past and future. I*EEE Computer.* 24(9) 106–107.

Cerf, Vinton G., and Ed Cain. 1983. The DoD internet architecture model. *Computer Networks* 7: 307–318.

Cerf, Vinton G. 1989. The internet activities board. *Computer Networks.* 17: 337–341.

Cerf, Vinton G., and Gerald Estrin. 1971. Measurement of recursive programs. *IFIP Congress.* 1: 314–319.

Cerf, Vinton G., and Robert E. Lyons. 1983. Military requirements for packet-switched networks and their implications for protocol standardization. *Computer Networks* 7: 293–306.

Cerf, Vinton G. 2003. Requirements for the internet. *RE.* 3–4.

Cerf, Vinton G. 2004. Taking internet's temperature: Prescriptions for the 21st century. JCDL 1.

Leiner, Barry M., V.G. Cerf, D. Clark, R. Kahn, L. Kleinrock, D. Lynch, J. Postel, L. Roberts, S. Wolff. 1997. The past and future history of the internet. *Commun.* ACM 40(2) 102–108.

In the same Series

TIME'S MYSTERIES AND MIRACLES
Consonance with Physical and Life Science

Ahmed Zewail

FROM MATTER TO LIFE: CHEMISTRY?!

Jean-Marie Lehn

THE MANY FACES OF HUMAN CLONING

Lori B. Andrews

ASYMMETRIC CATALYSIS
Roles in Biomedical Science and Technology

Ryoji Noyori

THE CHANGING ATMOSPHERE IN 2004

F. Sherwood Rowland

VEGETAL AND MINERAL MEMORY
The Future of Books

Umberto Eco

more to follow...