**Sixth Meeting of the Board on Research Data and Information**

**U.S. CODATA-World Data System FORUM**

**February 1, 2012**

**The National Academies**

**Washington, D.C.**

**Table of Contents**

# Participants List

# Introductory Comments

# Identification of Issues of Interest and Potential Areas for Cooperation in the United States

# International CODATA and WDS Issues and Plans for CODATA Conference in October 2012

# Concluding Observations

**Participants**  (listed in alphabetical order)

* Tim Ahern, director of Data Services for the Incorporated Research Institutions for Seismology, IRIS, which is an NSF-funded organization and member of the World Data System (application approved last year).
* Laura Bartolo, co-chair of the CODATA Task Group on Materials Data and *ex officio* member of BRDI.
* Dirk Behrend, NASA Goddard Space Flight Center, director of the Coordinating Center of the International VLBI Service for Geodesy and Astrometry, which is a member of the WDS.
* Fran Berman, Vice President for Research at Rensselaer Polytechnic Institute and co-chair of BRDI.
* Phil Bourne, professor of pharmacology at the University of California, San Diego, co-director of the Protein Data Bank, and member of BRDI.
* Mark Brender, Executive Director of the GeoEye Foundation and member of BRDI. GeoEye, the company, owns and operates high-resolution earth imaging satellites, and GeoEye Foundation provides free satellite imagery to universities and NGOs who need it for research.
* Bonnie Carroll, U.S. National Delegate to CODATA, President of Information International Associates, and member of BRDI.
* Robert Chen, international CODATA Secretary General, CIESIN director, Columbia University, and *ex-officio* member of BRDI.
* Sayeed Choudhury, Johns Hopkins University and member of BRDI.
* Dave Clark, member of the WDS Scientific Committee and visiting scientist at the National Geophysical Data Center (participating by telephone).
* Keith Clarke, professor at the University of California, Santa Barbara, Department of Geography, and member of BRDI.
* Dan Cohen, Library of Congress, on detail to National Academies, working with BRDI staff.
* Melissa Cragin, AAAS Science and Technology Policy Fellow at the National Science Foundation, and representing the American Society for Information, Science and Technology.
* Chuck Curran, Data Innovation Initiative, funded by Silicon Valley.
* Paul David, economist at Stanford and in the Stanford Institute for Economic Policy Research, co-chair of the CODATA Task Group for Global Information Commons for Science Initiative, and member of BRDI.
* Howard Diamond, climate scientist with NOAA’s National Climatic Data Center and Director of the World Data Center for Meteorology, which is now a formally approved member of the World Data System.
* Kelvin Droegmeier, University of Oklahoma faculty member in meteorology and vice president for research, and member of BRDI.
* Cliff Duke, Director of the Science Programs Office for the Ecological Society of America and member of BRDI.
* Barbara Entwisle, Vice Chancellor for Research at the University of North Carolina at Chapel Hill and member of BRDI.
* Chris Fox, Director of NOAA’s National Geophysical Data Center and Director of the World Data Center for Geophysics in Boulder, Colorado.
* Tina Gheen, Library of Congress, representing Roberta Shaffer.
* Sara Graves, University of Alabama, Huntsville, U.S. representative to the CODATA Executive Committee, representative of the Global Hydrology Resource Centers (one of the DAACs) which has applied to be a member of WDS, and *ex-officio* member of BRDI.
* Margaret Hedstrom, Associate Dean for Academic Programs and Professor, School of Information, University of Michigan
* Tom Holm, USGS EROS. The EROS Center under the World Data Center System was the World Data Center for Remotely Sensed Land Data and World Data Center for Land Cover.
* Bob Horton, Institute of Museum and Library Services.
* Joe Hourcle, Solar Data Analysis Center at Goddard Space Flight Center.
* Subhash Kuvelker, Board on Research Data and Information.
* Kirsten Lehnert, Columbia University, Director of the Integrated Earth Data Applications, an NSF-funded data facility for earth, ocean, and polar sciences with a pending application to the World Data Systems.
* Cheryl Levey, Board Research Data and Information.
* Sydney Levitus, NOAA, Director of the World Data Center for Oceanography in Silver Spring.
* Jeffrey Love, USGS Geomagnetism Program and chairman of an international organization of national geomagnetism programs.
* Clifford Lynch, Director of the Coalition for Network Information and co-chair of BRDI.
* Chopo Ma, NASA Goddard Space Flight Center, chair of the International Earth Rotation and Reference System Service (IERS), which was part of FAGS and is carrying forward into WDS.
* Alexa McCray, Harvard Medical School and member of BRDI.
* Jean-Bernard Minster, chair of the World Data System Scientific Committee.
* Carey Noll, NASA Goddard Space Flight Center, manager of Coastal Dynamics Data Information System, which is developing an application to WDS, and Secretary of the Central Bureau of the International Laser Ranging Service, which has been accepted into the WDS.
* H. K. “Rama” Ramapriyan,, NASA Goddard Space Flight Center, assistant project manager of the Earth Science Data and Information System Project, which is responsible for 12 data centers known as DAACs, distributed active archive centers, many of whom or all of whom are in various stages of preparing applications for membership under WDS.
* Ryan Scherle, Dryad Data Repository for Data in Bioscience.
* Sylvia Spengler, National Science Foundation.
* Paul Uhlir, director of the Board on Research Data and Information.

**Introductory Comments**

Bonnie Carroll

I am the U.S. National Delegate to CODATA, and am co-chairing this Forum with Bernard Minster, Chair of the World Data System’s Scientific Committee, and Bob Chen, Secretary General of the international CODATA. The purpose of this Forum is to discuss ways the two organizations – the World Data System and the international CODATA – can cooperate on the U.S. implementation of the WDS. To begin, Dr. Minster will give background on the World Data Centers (WDCs) and the World Data System. Dr. Chen will give an overview of CODATA. In addition to his role as Secretary General of CODATA, he was Director of the CIESIN World Data Center for Human Interactions in the Environment, when it was part of the WDCs, and is manager of the NASA Socioeconomic Data and Applications Center, which is one of the data centers that will apply to become part of the WDS. We will then open up the floor for discussion.

**CODATA Overview**

Robert Chen

As you know, CODATA is an interdisciplinary committee of the International Council for Science (ICSU). The World Data System is also under the ICSU umbrella. Bernard Minster will give you a history of the WDC/WDS shortly.

As a result of many ICSU committees and other deliberations, there was a realization that the ICSU bodies, not just these two, need to collaborate more on data management. ICSU recognizes that data management is very scattered across the unions, the global change programs, and so forth. Over the years they have been asking for more collaboration.

ICSU has made structural changes, such as the combination of the old World Data Center system and the former Federation of Astronomical and Geophysical Services into this World Data System. They did not try to combine CODATA or WDS, but asked the two organizations to work in a complementary fashion.

Coming at it from the U.S. perspective, with the dismantling of the old world data centers and services, each service and center in some sense continued, but they were not part of a formal ICSU network until the formation of the World Data System. Within the U.S., we have mostly been in limbo in this regard for a few years.

CODATA, of course, has been moving along on its agenda, but the thinking behind this meeting was that both groups in the U.S. should consider the opportunities for being complementary to each other, how to coordinate and collaborate, and how that fits into the international context.

I was asked to give a presentation and then also co-chair a forum at a meeting in September in Kyoto, Japan, which was the first World Data System Scientific Conference. I co-chaired it with Ruth Neilan from NASA’s Jet Propulsion Laboratory, and today’s presentation reflects that discussion. Bernard Minster will add more about the WDS when he speaks. There is an operational side to the World Data System and Bernard will tell you more about its historical focus on environmental data and observations, and its membership as it has evolved.

CODATA does have a somewhat broader set of disciplines involved, but it is not really intended to be an operational system. It does not oversee data centers and does not have users *per se*. It has a formal existence as a nonprofit in France, members who pay dues, and so forth. CODATA and WDS are somewhat different organizations, but both under the ICSU umbrella.

My initial thoughts on some issues that we could think about here in the U.S. include the question: What is CODATA’s potential to help the WDS? CODATA does have some expertise and experience in data policy and, to some extent, standards. It does have a broader reach in terms of disciplines. It has some links with relevant data initiatives and networks, such as the Group on Earth Observations. It has a number of initiatives on topics like data rescue and data citation.

It can, especially here in the U.S. being a part of BRDI, be an initiator or focal point for consensus studies and interaction when there are key issues that come up, especially that might affect, for example, U.S.-oriented policy in the science data realm. There are strong relations with key federal agencies here. It has some ability to work with a large community of data scientists and managers.

The WDS is equally able to help CODATA in some ways. It is great to have principles and policies that look good on paper. Putting them in practice and making sure that users have the access that theoretically they have is not a trivial problem. There are certainly key challenges that you see when you are managing data and dealing with scientists around the world that you may not see when you are sitting in a committee room. Some of those challenges may need policy attention. I think we will get into a discussion about some of those things, like one of the original roles of the world data centers, which was to provide a channel for scientific cooperation, even during the Cold War.

There are issues related to capacity-building, and certainly CODATA has had an agenda partly encouraged by ICSU, to work on digital divide issues and work with developing countries. The World Data System has had the tradition of doing that, so there is potential room for cooperation there.

Hopefully, the World Data System will establish and push the boundaries on data management, and that work coming out of the WDS can serve as a model in other areas of science, in particular, in the arena of data technology and innovation, which is affecting how everybody operates and how we think we can do things like make data sustainable in the long run.

Several things that came to mind about specific areas where there could be cooperation are:

In the World Data System, a lot of elements have been active in the Global Earth Observing System of Systems (GEOSS), which is under the umbrella of the Group on Earth Observations (GEO). CODATA was very active in the data sharing aspects of the policy issues, and there is a lot of potential work to be done as part of the development of GEOSS.

There is a lot of commonality and interest in the follow-up to the International Polar Year (IPY).

There is already an active CODATA activity on data citation and attribution.

There is a proposal on the table to work on harmonization of data standards, responding to the perception that there are many different unions that are either developing or implementing different standards without a whole lot of collaboration or interaction between different areas.

One of the tricky things is that both CODATA and WDS had limited interaction with some of the key standards bodies, the International Organization for Standardization and Open Geospatial Consortium and so forth. But individuals have a lot of participation, and somehow developing more institutional collaboration is certainly possible and maybe worth pursuing. The World Data System represents actual people working in actual institutions doing data management. They have a lot of skills. How can that be harnessed in a way through networks, through some of the things that CODATA has?

We had a discussion of CODATA and its strategic planning in concert with ICSU. ICSU has a new strategic plan that includes new initiatives in areas like disaster risk, and Earth System Research for Global Sustainability. There is a general issue of getting data even on the agenda of some of these.

As an aside, we had to really fight to get a data session at the big Planet Under Pressure Conference in March 2012. They rejected the first proposals and we had to go to ICSU directly to get us on the agenda. It was done jointly between WDS and CODATA and some other groups.

There is a lot of interest in the general issue of analoguing digital data that are at high risk of loss, and we have a small effort to at least canvas the highest risk things. Many of the world data centers have interest or have data that are at risk. There is also a community of people who do data rescue who are often not entrained in the scientific discussion of what to do and how to do it and how to pay for it. That is another thing that this group is trying to get at.

We obviously have a vested interest in making sure our conference and things are done effectively, and joining forces on meetings, conferences, etc. is important. Both groups have identified the important need for bringing in young scientists and making sure that they are involved in these activities.

That is one agenda, but is by no means the only agenda. Here is an opportunity for general discussion of needs and priorities; the opportunity to learn more about each other’s system.

Thinking of the issues in the U.S. perspective, what are some of the particular ideas we could pursue? The CODATA task groups and the World Data System activities that are in planning, being proposed, or underway is an example of taking action on some ideas. WDS has a secretariat set up in Kyoto hosted by Japan, and they just announced their new executive director who is someone CODATA has worked with closely as well. Additionally, our secretariat has capacity to do quite a bit.

We will talk about the CODATA Conference in Taipei as a coordination mechanism, and then a whole set of initiatives that need a data presence. That includes things like GEO, but also the Intergovernmental Panel on Climate Change, the new Intergovernmental Platform on Biodiversity and Ecosystem Services (which is an intergovernmental effort partly initiated by ICSU in that arena), and the Earth System Sustainability Initiative. I think there is also a set of issues that people have suggested in the pre-survey that was done.

This is a bottoms-up description of CODATA and the activities and possibilities.

**World Data System**

Jean-Bernard Minster

I am a geophysicist, not a data person. I have been asked by ICSU to chair the World Data System Scientific Committee. My experience with data is that I have served for many years as a member and then as chair of the old Committee on Geophysical and Environmental Data of the National Research Council, which has been fundamentally replaced by BRDI.

The WDS is something that was created by a decision of ICSU’s General Assembly in Maputo on October 28, 2008. This came after several years of discussions by advisory committees and people looking at the data problems that ICSU faced. It is an interdisciplinary body to replace the World Data Centers and the Federation of Astronomy and Geophysics Data Analysis Services. We should never forget the services. It is easy to think of data centers, but data services are really critical.

An international scientific committee was appointed immediately and I was asked to chair it. The members are people who are interested in data, and they took to heart the terms of reference. There are six terms of reference: (1) long-term stewardship; (2) quality-assessed data; (3) devise an implementation plan; (4) do not create standards or enforce them, but urge people to live by existing standards; (5) monitor the geographic and disciplinary scope of the system. This directly addresses the so-called digital divide. Also the original WDCs were almost exclusively natural sciences. We are very interested in social sciences and other sciences nowadays; (6) a specific recommendation to develop closer links with CODATA.

We defined our own objectives. We have all heard of full and open access. We are using a new phrase that is going around ICSU, which is universal and equitable access. That is to address specifically the fact that in many countries in the world (members of ICSU), difficult access to data is a serious problem and an impediment to science.

Long-term data stewardship is critical, especially when dealing with things like climate or anything that has a long time constant. In fostering compliance with data standards and conventions, we need to work with CODATA to pay attention to emerging technologies for improving access to data and data products.

Here is a timeline. On the left-hand is October 2008 and the right side is the end of 2012. We opened the system for business in January 2011. We have a fairly large number of institutions worldwide who have expressed an interest. We have defined things that we will talk about, like constitutions, bylaws, and a data policy.



In September 2011, we held our first scientific data conference. That was an idea that was proposed by our Japanese colleagues, and they organized it in Kyoto and Kyoto University. The previous ICSU institutions that we are replacing had not had conferences. We participated in the CODATA General Assembly in Kyiv and again in Stellenbosch, and we will participate in Taipei. That is one of the things to discuss: what is the contribution that the U.S. part of WDS can make to CODATA and to the U.S. part of CODATA?

The red box on the bottom of the timeline is March 5, 2012, which is when the Executive Director of the International Program Office for WDS began his service. We have been waiting 60 years for something like this.

We have a website that is hosted by the Polytechnic Institute in Kviv and operated by colleagues in Kyiv in collaboration with other colleagues in Germany at the PANGAEA Data Center. We will probably expand this initial effort to a major activity in Tokyo. The website contains the documents that you may be interested in if you look at WDS, like the constitution and the data policy, and the list of candidates. It also has a map, on which active members show up in green. It is updated every 10 seconds or so.

The constitution was drafted and approved by the ICSU Executive Board in April 2010. It defines what the system is, what the goal of the Scientific Committee is, and an International Program Office. This is different from the way CODATA is organized. There are some parallels, but the mechanisms are different.

Now that we have members, we should ask the members what bylaws will be most beneficial to them. These bylaws will address the operations of the International Program Office and review of WDS memberships. Our data policy is simple. It is the preamble or the first page of the GEOSS data policy that was developed with enormous leadership from CODATA. We have deliberately stayed away from the 15 pages of exceptions that the GEOSS data policy addresses. We realize that there will be exceptions. We realize that there are privacy concerns, that there are rare species to be protected. There are a number of things that have to be taken into account. We did not work that into our data policy. For now, we are sticking to the concept of full and open access to data, metadata, and data products, with no delays, no sitting on data, and no cost. Most WDS data will be distributed electronically, so the cost should be minimal.

The Program Office is a coordination office and is responsible for implementing the decisions of the Scientific Committee and for day-to-day tasks.

A call for proposals was issued by ICSU to all ICSU members, national members, and to the unions. The National Institute for Communication Technologies (NICT) in Japan offered to host an international program office for WDS with financial support to the tune of approximately $400,000 a year for five years. Given the fact that we have operated the world data centers for 50 years with no money, this changes the ballgame completely.

The proposal was accepted and the office was approved by the ICSU Executive Board in April 2010. After a great deal of negotiations with our colleagues in Tokyo, a letter of agreement was signed in July 2011. A search was made for an executive director.

I chaired the search committee, and we interviewed the top four candidates last December in Tokyo and selected Mustapha Mokrane, who accepted the position. He had worked for ICSU for a long time. He was the ICSU staffer for the WDS effort and for some of the advisory committees that ICSU set up before that. We anticipate that the staffing of the International Program Office will be done by the end of summer 2012. We have looked at the Program Office for the International Research on Disaster Risk (IRDR) that is based in Beijing. Installing an international organization in a country like China or a country like Japan involves interesting negotiations.

As shown below, we have a draft architecture for WDS. The red boxes are basically WDS items like data collection, data archiving, services, and metadata. Publishers are on the left, and libraries are on the right. We realized that a world data system would have to deal with other organizations that are critical in the data world. It is more than just connecting with research facilities, doing education, and outreach. It is dealing with people who have a long history of managing and curating datasets.



One thing we quickly discovered is that we have big centers like NASA Goddard and the MIT Geospatial Data Center, and we have small centers like the International Center for Earth Tides in Tahiti. The point is that one model is not going to fit all possible members. We have to think of WDS as a system of data systems. That is perhaps a mouthful, but this is deliberate.

For example, the International GNSS Service (IGS) has data analysis, data collection, and orbit production systems all over the planet, and it is run out of the NASA Jet Propulsion Laboratory (JPL). This is a big organization and everybody uses the IGS. If you drive your car and have a GPS receiver, very likely you will end up using orbits generated by this service. Even the National Geospatial-Intelligence Agency uses IGS orbits.

So, is that one node in a world system, or are all the places where data get analyzed and stored going to be nodes? As a scientific committee of eight people, we cannot decide that. It is going to be decided by the community, so we need to set up the discussion.

There may be world data centers, services, data analysis services, data activities, and so on and so forth. Those things will have to communicate in what is known to the IT people in the room as a scale-free network. It is basically a system of systems, so no matter on what scale you look at it, you basically have nodes that talk to nodes, and ultimately you get big nodes collecting and doing a service for many other people.

Another thing we discovered is that potential members are not all the same. We have regular members, those who do analysis services, deliver data, do data curation, etc., especially for their own users. Then there are groups of members or umbrella organizations that want to be part of a world data system. The International Oceanographic Data and Information Exchange (IODE) and the International Federation of Seismic Networks are examples. Are those organizations regular members? Often they are just simply organizations. They do not deliver data specifically. They just coordinate what data deliverers do. So we invented network members. Then, International Union of Geodesy and Geophysics (IUGG) wanted to be a member of WDS. So we invented partner members. Then, Elsevier and Wiley said they would like to be members. So, because they are commercial, we called them associate members. We invented those categories and that has worked reasonably well so far, and we will adjust the categories if we need to.

We spent almost two years defining criteria for membership. We purposely did not use ISO standards because they scare people, and in the beginning we did not want to scare potential members.

But we did want to have a transparent and objective base for evaluation, accreditation, and assessment of the members. We want a base for trustworthiness of the data. The data has to be authenticated, and the database has to have integrity for protection, security, and confidentiality, especially for social economic data. We have to balance the security with availability; it does not help to have a beautiful data center if nobody can get data out of it. So, prospective members need to answer a questionnaire, with questions such as: Do you have a data policy? Do you have backups? Do you have a user working group? Do you have scientists on board to do data quality checks?

The questionnaire’s intent was to weed out early applicants who would clearly not be qualified. We now have 37 accredited WDS members. We had 146 expressions of interest. Fifty-eight of them actually went through the whole system and applied for membership. Thirty-three of these applications have been reviewed. We wanted to do peer reviews, but in order to just get started we used the scientific committee members to do the initial reviews. We have one network member, one partner member, and two associate members. Eleven applications have been actively discouraged. That is a question I keep getting: do you ever turn anybody down? We do.

One thing that was insisted upon in 2008 was that we should be able to deal with the IPY data legacy. We participated with a CODATA Polar Commons, the PIC project. We have issued targeted invitations to 80 major IPY data holders. Among the members that we have now, these are centers and services that actually have pooled IPY data. There is more to be done, and, in fact, we will go to the meeting in Montreal and discuss that. It is the last meeting of IPY for science.

We had our first WDS conference, and it was remarkably successful. There were 155 people from 22 countries. Surprisingly, we had participants from natural sciences, social sciences, and IT participants. It was a nice group, very open and willing to discuss issues. One of the things we pursued was a collaboration with the IRDR, holding that in Kyoto after the tsunami and earthquake and nuclear disaster. The day we opened the meeting there was a typhoon going right over Kyoto, so people were exceptionally ready to talk about disasters.

It was an open forum that was run by Bob Chen and Ruth Neilan where members were asked what they would recommend that we should do. One of the things they recommended was to set short-term, medium-term, and long-term goals, which is difficult.

We have a document of “shared understandings” with the Japanese that is very short (about two-thirds of a page) and says, basically:

* ....The WDS should aim at an integration of multidisciplinary scientific data and information to address the needs of ICSU programmes including the Earth System Sustainability Initiative...
* ... data publication and data citation should be adopted and promoted by the World Data System in order to facilitate the release of data, with proper recognition of providers...
* ...Ongoing collaboration between ICSU programmes—WDS, CODATA and IRDR—will be critical for building interoperable data systems that guarantee full and open access to data in support of better disaster prediction, understanding, and mitigation...

In essence, we have four goals: universal and equitable access, long-term data stewardship, compliance with data standards and conventions, and mechanisms to facilitate and improve access to data and data products.

People bring up climate research to justify the need for long-term data curation, but in all natural processes, so much happens that you can never observe again. If you miss Katrina, you can never observe Katrina again, and it changes the way the land looks. There is something about natural processes that requires long-term curation. This is not only climate-based.

DISCUSSION

DR. CHEN: It is important to have a standard out there, but not all elements are sufficiently geared to think about being a long-term trusted digital repository with those functions. There is a gap between what, for example, an active NASA data archive is and what the expectation of a trusted digital repository has expressed in this new ISO standard would have to do. It would be important to work that standard into the system. There may be some category of WDS elements that are encouraged to apply for and receive that certification, and WDS could be an auditor and support the standard, but it is a pretty new standard and that has to be discussed within the community and assessed as to its relevance and how it might be applied in the WDS context.

DR. MINSTER: The social sciences have their own set of criteria that I discussed with them. It is a little bit simpler, but they have other concerns that we do not have.

MS. CARROLL: What we are trying to create is a forum and a community that can discuss that standard and how it really applies to operational data centers.

DR. BOURNE: You mentioned social sciences, but the biomedical sciences do not seem to be represented, nor are open access publishers. I guess the question is: what can this group do to help foster those folks for membership?

DR. MINSTER: We did talk about that and we are interested in talking to organizations or disciplines that are already well organized.

DR. RAMAPRIYAN: For earth sciences, NASA is not mandated to have permanent archiving responsibility. There are other organizations that do this. But as long as there are funded researchers who are using the data, they are going to have active archives. The data is in use, of course, much longer than just the end of missions. Given that and that these data are valuable assets that need to be preserved and transitioned, we have the responsibility of identifying what needs to be preserved and preserve that data until such transition occurs. For all practical purposes, we do have some long-term preservation responsibility.

DR. CHOUDHURY: It is very important to remember that there are different ways people are auditing and certifying right now. TRAC is one way. There are different levels. People do self-audits, they do audits with a partner, and then with some trusted agency, which WDS could be, for example. But a more important point is sometimes we talk about audit and certification like it is a grade and you either pass or you fail. That is not the most useful way to think about it. It is actually about transparency. What are you doing and how are you actually supporting what you are purporting to do? Do you have documentation? Do you have a business plan?

We need to let data producers and the communities decide that according to this audit and certification, this particular entity has this level of expertise. With that spirit of thinking about transparency, the WDS could play a really important role in certification. Are you putting the questionnaire data online? Because that would be very handy just to see how certain data centers are managing their operations.

DR. MINSTER: In the natural sciences, if a colleague has a dataset that you are interested in, you go to him or her to get the dataset. You are not going to go to a NOAA data center or a data service. This means that there is this informal, vague, churning flow of data across the net. There is a danger there. The data can be corrupted either maliciously or by error. What we need is a layer at the top that is guaranteed to be pristine. In other words, this is verified, this is audited. Data can come down, but not back up. This is the role of WDS nodes.

Regarding peer review of datasets, how do you do that? CODATA can certainly help us on the design and mechanism.

For review, we do not want site visits as has been done for the last 40 years. Data centers can be a bunch of computers someplace, and boring to see. But, a “web visit” could work. Two experts in a particular discipline could examine the holdings, make sure that what is available is something that researchers actually use, that the metadata is maintained properly, etc. This could be done pretty much in the same way as an NSF proposal is reviewed with a form.

DR. MCCRAY: Regarding the WDS’s stance on the open and free exchange of data among members in relation to the different categories of members, are the existing members all funded by governments? If that is so, the data in the United States would be open already, in the EU it is more complicated, and around the world it is different. What is the membership right now?

DR. MINSTER: The membership is anybody who wants to apply and is willing to live by the standards that we set. The WDS is not a member organization in the usual sense. If data are stored in a WDS node, access should be full, open, with no restriction and essentially no cost to anybody on the planet.

For instance, if some data gets placed in the repository in the United States and scientists in North Korea wanted to access those data, if that location is a WDS node, we would not place a restriction on access. If the government were to insist on a restriction, then we will have to remove the holdings from the WDS. Anyone has access, not just members.

DR. MCCRAY: You had on your slide today a discussion about publishers being interested in the data that accompany articles. How does that work with regard to publishers being businesses and having business models that would differ from your statement?

DR. MINSTER: That is a very good question. We have started discussions with publishers. Those discussions happen largely in Germany and in the UK. We had a complete session at the CODATA General Assembly in Stellenbosch on this. The initial position of publishers like Wiley and Elsevier, for-profit publishers, is that they will publish the paper and provide access to the data, but the WDS can have the raw data. That is an initial position. We have not talked dollars and cents yet, but there is hope.

MS. CARROLL: Back to the question of who are the members and are they all federal. There are three choices: federally funded, academically funded, or commercially funded, primarily. Everybody has a mixture. How many are federally funded? How many are academically funded? How many are commercially funded? By a show of hands in this room, all are federally funded except one that is commercial. To the point, the data is open. They have aspiring or current members. That is a topic we have on the list to discuss in depth later.

MR. UHLIR: Getting back to the data policy, there are actually three data policies in your presentation. One is that the WDS is fully open to anyone without any reuse restrictions. However, this is not written anywhere. The second one is “universal and equitable access,” which is an ICSU policy for publications and which is not full and open. Universal and equitable means cost recovery. It is a publisher’s policy for journals, and that is what ICSU uses to denote its policy for journals. For data, it is “full and open” and you draw on the GEOSS full and open data principles.

The problem with the GEOSS data principles is that they allow cost recovery as well. There is only a subset of the GEO/GEOSS data principles, which is what you espouse verbally, and that is the Data CORE. Let me read the four principles.

Data CORE stands for “data collection of open resources for everyone.” It is a subset of the larger principles. Within the Data CORE, the data are free of restrictions on reuse. The user registration or login to access the use of data is required, so that is already some restriction.

Attribution of the data provider is a condition of use, which is also a semi-restriction, although not viewed as such by GEO, but it is a condition of use. Marginal cost recovery charges are allowed but marginal cost recovery is very fuzzy. I would say that online it is free, but I have gotten a lot of resistance on having that in writing, so the cost of fulfilling a user request is a general definition of marginal cost. Online you can say it is free.

U.S. federal government data are in the public domain. You can charge, but there are no reuse restrictions. So the U.S. federal data policy of access adheres to the GEOSS Data CORE, but the other countries do not, and even the WDS does not. This is going to be a fundamental point of disagreement between the U.S. and everyone else within the organization, especially if you are going to have two different kinds of data policies, “universal and equitable access,” which is definitely not what you are talking about, and “full and open access,” which requires more precise definition.

MS. CARROLL: We may have some domestic interests that we want to coalesce around as we move to the international discussions.

DR. RAMAPRIYAN: I have concern where you were giving an example of North Korea, for example, being able to get the data if it is in the WDS as opposed to a NASA archive. We do have NASA-imposed restrictions on certain bilateral arrangements. I am a little worried that if we are to apply and become WDS members, we would have to violate those laws that we have to obey.

PARTICIPANT: We have actually gotten a ruling on that. When you are part of a larger organization and you are not talking just specifically to the Chinese, but to a larger group that has a Chinese member, that is okay.

DR. RAMAPRIVYAN: But we also have designated country rules that we need to worry about.

MS. CARROLL: The restrictions on exchanges by nation is on our list of issues.

DR. CHEN: I already alluded to the issue of what are the traditional roles of the world data centers. Bernard did not go back as far in history as I thought he would in terms of the world data centers in the International Geophysical Year and then the subsequent role in promoting data access across the Iron Curtain during the cold war and so forth. That is a whole topic in itself.

I just want to clarify that the GEOSS data principles were negotiated by governments. A friend of mine, Dan Reifsnyder at the State Department actually did the last negotiation, I think, with the Russians in that, so CODATA had no role in those data principles. We came along later trying to help do things like Paul mentioned, which is to answer the question: What does full and open access mean?

In the guidelines that we did help produce and in the white paper, we benchmarked the definition of full and open against the literature, which includes things like the U.S. Global Change Research Program data policy. Also, an Academy report that was done by the old U.S. National Committee for CODATA also defines full and open. We have a role in helping to clarify what these kind of general principles mean in ways that the international community finds acceptable.

Obviously there are still differences and opinions on things like what minimum cost means and what delay means. There is a document in GEOSS that tried to address some reasonable agreement on what some of these terms mean. The WDS is free to use those as they are or develop their own or refine what exists. That is what Bernard meant when he said the community must come together to put forth the effort to come to some kind of agreement.

This tricky thing, for example, is paying attention to members within the ICSU umbrella such as the ICSTI community as well as the publishing community. Developing an understanding varies and is evolving and there is no simple answer. But the WDS will have to grapple with it, and to the extent CODATA can help with that, evolution would be valuable on exactly these differences in interpretation and then in practice when you actually have a set of units that do not fit any nice, simple definition.

In the old days, there were relatively few categories of world data centers and services. Now we have people all through the world data lifecycle and the publication lifecycle who are interested in being part of an international network. One thing that did not totally come out is what is the incentive and motivation for people to participate in the network? You have quality minimum that you apply, but what is the return that one gets from being part of the network?

Previously, the return was that it helped you with sponsors and getting funding because agencies like NASA said they were committed to the World Data Center System, and therefore you could justify people being assigned to be the head or a director and paid for by the agency. There are the implied quality blessings. There are the networking benefits. But it might be good if we at some point also talk about the incentives for participation.

DR. CLARK: Back in the ’70s and ’80s, one of the major advantages that we found as a member of the World Data Center System was that we could interact with our colleagues in, for example, the Soviet Union when there was virtually no contact by the governments, or at least in an adversarial mode. We actually had government agencies that were charged to find out how we could actually do this, and we raised the ICSU flag. They said okay, that seems reasonable. We pointed out the benefits to the U.S. scientists by doing this. That particular aspect was a very important and a critical part of the World Data Center System at the time.

MS. CARROLL: We are well into the discussion of issues, which is the next topic. If there are no specific questions about the World Data System, I want to step back for one minute. We have two organizations here. We have CODATA and the World Data System. Bob Chen did an excellent job of briefing everyone on CODATA. I will just read the mission statement of CODATA. “The mission of CODATA is to strengthen international science for the benefit of society by promoting improved scientific and technical data management and use.” The WDS and CODATA organizations are complementary. They are cousins within the ICSU organization.

There is one very big difference. At this point the World Data System is international in scope. CODATA has national implementations in the form of national committees, so we thought this was a good opportunity for the U.S. National Committee to look at the World Data System and to determine a potential national coalition here where U.S. centers can talk among themselves because we may have some unique issues, and then bring forward the international issues. Of course, you all participate in the international structure as that exists.

That was one of the reasons Paul Uhlir and the co-chairs thought it would be good for BRDI and CODATA to host the data center managers to talk about this kind of thing. First, we want to work together nationally, and second, how do we work with the international system? Is that clear? Is everybody good with the relationships and where and how we want to go? By the end of the day we will have a lot of issues, and as a result of having talked through the issues, we will know whether we have some that we want to separately talk about nationally.

**Identification of Issues of Interest and Potential Areas for Cooperation in the U.S.**

MS. CARROLL: What I will do now is quickly summarize some of the issues that have been raised:

* Certification issues: How our standards meet international standards, and what the World Data System may set as standards.
* Scope and coverage regarding the social sciences: The United States has a very powerful social science component to the sciences.
* Business data: Data that is not scientific but that, in fact, supports the advancement of science. This is a topic that we in the United States have already started talking about.
* Long-term archiving: Who is really responsible for this? Rama was saying it is not necessarily NASA, so maybe the whole relationship between data centers and NARA might be an interesting topic, at least at the national level.
* Return on investment: Is there a return on investment by participating in the WDC internationally, and is there return on investment if we want to get together? Because there is a cost; you spend time. Is there some value in that?
* Access and use policies: How they are domestically, how they are internationally, what the options are, and as data center managers, how do you deal with these policies.

Now let us turn to the identification of issues. There were three groups: policy, good stewardship, and technical, particularly discovery and interoperability. Those are the issues that might be worthy of getting together and talking about.

On the policy issues, we want to ensure free and open data access and communication without suspensions. A related, complicated topic that we can talk about is the non-communication period for NASA with China and with NSF as well.

The ICSU principle of universality of science and the trend toward more open scientific exchanges is not guaranteed. So how do we as a scientific community weigh in on this, particularly as people that are data centers, because of the impact on your data center due to your need to exchange data with other data centers because data is often global in scope?

The WDS and U.S. CODATA can contribute to strengthening multilateral non-government lines of communication. But we need to take into account the current controversies regarding publication, research, and genetic manipulation. Those are clearly topics that a group like this might want to talk about domestically, and we might want to raise some things internationally.

The other policy issues are the intellectual property issues regarding data, their management, dissemination, and use with respect to copyright. We were talking about regulations and implementation, but getting up one step into international intellectual property regimes. Do you as data centers want to have some kind of community of practice discussion about these kinds of issues? You may or may not do anything about them, but a discussion may at least clarify how legislation like the Stop Online Piracy Act (SOPA) and Preventing Real Online Threats to Economic Creativity and Theft of Intellectual Property Act (PIPA) impact you.

We should promote good stewardship and support the international science services. They have to gather and archive as well as generate data for scientific and societal use. They have important components in the United States, but they rely on international cooperation. So, practicing good stewardship in the realm of international cooperation, how do we support international science services?

We need to describe and promote good data stewardship across the entire enterprise. So again, the issue is how do the data centers interact with the publishers? Everybody is interacting with everybody, but are data centers actually in that conversation? Some may be, and it may be you have unique issues as data center managers.

On the issue of technical discovery and interoperability, we should provide discovery services to a wider community and develop interoperable data discovery processes to share members’ catalogs of information. Is this a forum where the data center managers might want to get together and discuss these issues? First, talk about vertical integration within domains nationally and internationally, and then, horizontal and between repositories across domains.

Now that we have gone over all of the issues that were brought up, do we want to add issues to the list? Do we want to talk about any of these issues? We can brainstorm and then list and think about the ones we really want to deal with, and categorize according to domestic and international.

DR. DAVID: Another issue is in regards to the implications of the Research Works Act legislation, which has now been reintroduced, and officiate the NIH policies on deposit of federally funded research publications in open access repositories. This is similar to the SOPA and PIPA issues. But this group has concern for precisely this kind of legislation that would cripple open access policies and which would spike the cost of scientific data and information access.

MS. CARROLL: Having worked in interagency groups who you heard about, federal people cannot engage in the policy aspects of this except through their own organizations. However, what are the implications for technical operations? Some people might not know about the Research Act, so they might be interested in an information exchange, and then a discussion on the technical implications.

DR. LOVE: What about issues like transitioning data from closed access to open access, e.g., data acquired for profit but which now the company might feel willing to share? There is a whole lot of very interesting geophysical data that could be useful for academic research that is not too readily being transitioned to open access.

MS. CARROLL: Actually, as a domestic issue it might be interesting to engage the intelligence community in a conversation. Is that what you are saying? Maybe just by talking, there might be some opportunities for opening up exchange that people do not think about.

MR. BRENDER: The intelligence communities are very concerned about cyber-attacks and cyber-warfare, so under “stewardship” there might be a discussion about security of data from such attacks.

DR. CHEN: The old world data center group worked with, as was mentioned, the Committee on Geophysical Environmental Data, which was under the Board on Earth Sciences and Resources. That had two roles. One was actually kind of a national review of world data centers that included site visits and that whole sort of function. It also had a convening function, which was occasionally to organize world data center direct meetings. Then it helped with the scientific committee and other things.

Since that does not exist, I think one of the issues potentially on the table is, is it valuable for the Academy in some form, like through BRDI, to help in one or more of those functions as the WDS moves forward?

It is valuable to have a group. There are no other groups that include the world data centers and services in the U.S. that would naturally get them together unless an agency decided to do it in an *ad hoc* fashion. Is there a value to having an Academy body provide an umbrella in forums like these? What would be the function of that in terms of having a U.S.-oriented meeting or doing any kind of review or participating in the selection and vetting process or addressing data policy issues, etc.? That is one of the endpoints, just to make it a little clearer. It would be interesting to hear some of the agency people as to whether they think that is actually a useful thing.

DR. FOX: I think it is a very good idea. As we all know, building these federated systems internationally is just that much harder than building them nationally or even within our own agency. At NOAA, it is a challenge to get everyone working together.

In particular, you mentioned the discoverability of datasets. We have recently been making good progress there. It is a convergence of the metadata standards coming online and being adopted, so a lot of our organizations are using ISO and (?) metadata in a common way. At the same time, ESRI has produced this geo-portal software system that allows access to all of these metadata catalogs.

Suddenly, very easily we can produce a portal of portals. You can look at all the collections in my data center. And we have extended that to some of the other agencies like the Corps of Engineers and the USGS. There is actually a core. If we want to find a success, I think that is a place we can do it. If we can help that happen nationally, then I think that forms the example of the core to then extend that internationally to the World Data System.

MS. CARROLL: NASA also has a master directory that does a similar thing, and searching across those by federating those directories might be interesting.

DR. MINSTER: I also think it is a good idea. However, there is an opportunity to get organized communities that do social economic data or biomedical data. There are principles and mechanisms for genomic data, and something that was never addressed in the WCD director meetings, model data. How much do we want to say of the output of those big climate models? That is something to be discussed.

Then to go back to Jeffrey’s question, it is a good idea at some point to have even a small meeting where the board might invite a representative of MEDEA to describe what can be done. This should maybe be a classified meeting with a subset of the board.

MR. UHLIR: There was some discussion of domestic legislation that might affect access through SOPA, PIPA, and the attempt to roll back the NIH deposit requirement, but those are focused primarily on the literature and not on the data centers. There is some connection, but it is marginal. For the data centers, a much more significant congressional action recently was the congressional resolution for NOAA to investigate cost recovery for its data distribution. I think that goes to the heart of the matter.

MR. DIAMOND: I am not sure it was a resolution. I think it has been a discussion, and that was related mostly to the satellite data from the new NPP and JPSS missions where there is somewhat of a ground swell that somehow charging for this data will do a lot of cost recovery, whereas past experience has shown that it just tends to warehouse the data and no one then accesses it. But I think it was related specifically to the satellite data. That is what we have been told and from the articles and things we have seen. There was an article in *Science News* about it and there has been a pretty big push back from at least the climate community about this. People in the community have been very vocal about the problems with this, and this is something we are going to have to pay continued attention to as our goal to is try and provide as much as this data in as free and open manner as possible.

From a World Data Center for Meteorology standpoint, most of our climate data are *in situ* type of data, but we also do serve up quite a bit of satellite data as the National Climatic Data Center also administers the CLASS system, the Comprehensive Large Array Stewardship System. The USGS has done a really nice job over the past few years in making all their Landsat data pretty freely available.

MR. LEVITUS: If we were to charge for data, many foreign countries would no longer send their data to us. Twenty years ago the situation was such that we really had to go out and work for many years to build up relationships with many countries that had been allowed to lapse. If we were to start charging them for their own data, it would end the goodwill that has been built up over the past years.

MS. CARROLL: This is clearly an issue where data center managers have very strong opinions and good technical input to the discussion. How could we as a convening body, assuming that convening was worth doing, voice those kinds of things? What would be mechanisms? What is the return on the investment if we come together and talk about these things? How do we then send them forward?

MR. UHLIR: In the institutional context, we would have to have a formal project that would be approved by the Governing Board, funded by somebody, and would lead to some short study with recommendations, or look at the pros and cons of free access versus cost recovery or research of various kinds. We have done studies here in the past that looked at different data access and reuse policies and privatization. We could certainly do that. It would not be quick and it would not be free, but we are in a position to do that kind of thing. This is possibly a fundamental service or activity of this group.

DR. CHEN: When the previous NAS reports were done, they were influential and helpful to lots of people, but they are old. Having a current assessment of where international cooperation and open access is and can go and what needs to be defended is important. Having an Academy document to back you up is very important when, for example, you are sitting in government and need something to point to in order to justify a position that we are taking for granted because the trends have been positive for things like opening up Landsat.

The point is that none of it is guaranteed. Starting a study now, even though it may not look absolutely necessary now, is worthwhile. If we wait until a crisis, we do not have time to prepare that kind of analysis. So all of the advance work can have an important payoff. We will be ready with the kind of backup that one might need if any of us were trying to deal with, say, an expansion of restrictions on exchange of data well beyond China and NASA, to take one current example.

PARTICIPANT: There have been at least three calls for comments from the federal government about free and open data access in the last two months. There was the NSF, the OSTP, and I cannot remember the third one. They closed right at the beginning of January. So unless you have these requests in advance, you are not going to have time to even find out about them, much less organize a comment.

MR. UHLIR: But the OSTP call was at least for two months. It started in early November, I believe.

PARTICIPANT: On another topic, regarding the creative commons licensing, if someone releases their data with the license stating “no derivative works,” does that mean you are not allowed to write a paper on it?

DR. MINSTER: I bet many people in this room remember the Francis Bretherton Tree for Environmental Data in the NRC Report *Resolving Conflicts Arising from the Privatization of Environmental Data*. The policy that was described in that privatization report was focused on environmental data. I think if you expand the discussion to research data in other fields, then all of a sudden things become more complicated. Engineering data should be treated one way, biomedical data should be treated another way, private data yet another way. That is something that could be a major study.

DR. HOLM: From the USGS perspective, especially when it comes to the pressure in terms of going back to a cost of fulfilling a user request (COFUR) model, we are not seeing anything direct, although since we opened the archives to free and open access in 2008, we are almost annually having to continually defend whether or not we should go back to that model.

When we were selling data, the best year we ever had in terms of daily average was about 53 scenes of Landsat data going out the door per day. Today, it is about 5,700 scenes a day. Scientists are using the data they need, not the data they can afford, and we have got to do everything we can as a community to document and show the benefits of free and open data access to the users on a national and global scale.

A body that would organize world data systems and get groups together would be very useful. At least within the U.S. data centers, there are a lot of lessons learned and a lot of best practices that can be shared. NOAA has a lot of good things. Bring in NARA, the USGS. That would be a good way to utilize the world data system.

DR. BOURNE: That is a very good point, but it is a point at essentially the end of the process, in other words, the scientists using the data after it is deposited in a place and how that use increased. But in terms of coming up with scenarios to enforce the argument, we also need to consider studies that already exist where that occurs by the scientist depositing the data.

There are numerous examples where the scientists are the ones who put it together, for example, in different regions or different societies. Then that leads to all sorts of activities around that data by virtue of the fact that they deposited it. That data will already be interoperable. This has worked pretty well in the biosciences, at least.

MS. CARROLL: The USGS Landsat data went from 500 to 5,000 views. Those are exciting metrics. Have you ever looked at the impacts as opposed to just the numbers? Has there been any work done on that? We were talking about the value of things, and process metrics are one thing, but what are the impacts? That is a question that other communities are looking at very seriously. How do you define the impacts?

DR. HOLM: We do not have a formal effort underway. When the users download data, they give us some metrics of how they are using the data and we document that. We have an understanding of how the data is being used and in what areas they are using it. But we are just this year formalizing a requirements gathering process so we can better understand the requirements of the broader user community. We have done that for decades, but we need to formalize it. When the President issued the National Space Policy, there was a specific statement on understanding the nation’s user requirements, so we are taking steps to gather that. An outcome of that will be impact kinds of statements.

DR. BERMAN: Is there any distinction made between paying for the data and paying for the infrastructure that hosts the data? Because, in some sense, those are two different things.

DR. HOLM: Yes. The more literal way of putting it is data is not free; it is no cost to the user. There is an infrastructure that has to be in place at the data centers in order to maintain that no-cost-to-the-user access to the data. When we were selling data, we were able to retain the sales revenue as part of our operating budget. Then we had to go to a model where we essentially lost all of that. We no longer have that, and now we make up that difference through efficiencies and reducing some of the services. That is what we were forced to do.

DR. CHEN: There are at least two different kinds of metrics efforts ongoing. NASA actually just got a consumer satisfaction award from a federal group. They have been commissioning the American Consumer Satisfaction Index (ACSI) for seven or eight years now. It has been entrained by OMB as a metric for the system. It gets at how users feel they were well served in finding, accessing, downloading, using, and getting help with data. It does not really look at their impact, but it looks at their satisfaction. Centers generally come out above the federal government average, which NASA is happy with. It is a valuable metric that internally has been helpful in responding to OMB and other groups.

There is a very small research community involved in looking at science citation type of studies of impact. Molly Macauley at Resources for the Future (RFF) and I have done some work and a few other people have also, but it is very hard to come up with objective, replicable measures of scientific impact based on citation-type studies. But a lot of people are interested in how one would do that.

As far as I know, there is almost no investment, at least in the fields that we work, in this type of study on the part of agencies, although OSTP did run a workshop looking at those kinds of issues. It is at least on the table, but is not getting the attention that it could, in terms of how do you measure scientific impact of issues like providing open data.

DR. BOURNE: Building on Bob’s point, the notion of bibliometrics as it relates to data and data access is kind of a chicken and egg situation because measuring the impact of data can be hard. Unless you have complete access to the literature that references that data, you cannot actually determine a follow-on citation. You are restricted to the open access literature on a broad scale, and that has some limitations.

But there is no doubt that there is data out there that shows that, first of all, datasets are accessed. That is one metric, but how those datasets are then cited in the literature, which occurs often up to a year or far beyond that, is another. There is a correlation there. H1N1 was an interesting example of that.

MS. CARROLL: This whole topic of citations, e.g., bibliometrics, is an interesting discussion, and it leads to: How do we cite data, Digital Object Identifiers (DOIs) for data, and things like that? This is actually a topic that is going on in communities. Data center managers are always asked to provide input, but maybe there is a conversation to be had among the data managers on that.

DR. BOURNE: It is a conversation that has to be had among the scientists because there are datasets that actually have DOIs assigned to them, but the scientists do not actually use the DOI when they reference that dataset.

MS. CARROLL: CODATA and ICSTI have a citation working group that is looking at a lot of these issues.

DR. MINSTER: There are a couple of other topics to put on the list. One is data architectures in other fields. I have a tendency to think of data systems as a relational database management system, but that does not work for everybody. Virtual observatories work very well for astronomers. Now I hear that other fields like biology are interested in that. This board might want to think about that topic and where the U.S. can lead.

The other one is education. There is a foundation in Sweden called Gapminder that is led by Hans Rosling from the Royal Swedish Academy of Engineering Sciences that focuses on educational systems to teach high school kids about data and fact-based material. That may be another place where we can work from.

DR. RAMPRIYAN: Back to the discussion of impact metrics and DOIs and citation, there are at least two other forums that I have been involved in where such discussions have been occurring. There is the NASA Earth Sciences Data Systems Working Group. There the data centers get together and various other peer-reviewed PIs get together and discuss this issue. Also, there is the Earth Science Information Partners Federation, where a considerable amount of discussion on this has been going on. There is a white paper that Mark Parsons was very involved in developing under the auspices of the Data Stewardship Committee that has been adopted as a baseline at this point that talks about data citation and guidelines for it.

DR. CHEN: Just to clarify, what I was raising is: How do you look ahead and say we are successful with DOIs and we are successful in actually having better means of citations? How then do you create actual objective indicators that measure impact based on that success? In other words, simple things like do you weigh success by the journal impact or how do you make sure that you now have captured the realm of data citation, how do you then construct sampling or measure trends over time and so forth? There is a whole set of issues related to indicator development and testing, and in doing that once you are successful in understanding where your data is being cited.

DR. RAMAPRIYAN: At every point where we come up with an objective measure one can always raise an objection that says that is not quite going to cut it. Not all citations are created equal, so there is a problem there. What is the impact on society, for example? How many lives did you save by providing this data and that sort of thing?

DR. CHEN: There is also the non-scientific impact of the data.

DR. DAVID: This topic about metrics, evaluation, and impact on scientific citations and co-citations of scientific publications in patents, has been extensively studied by economists. These economists are working on innovation processes and interaction between scientific research and the commercialization process, as well as looking at the impacts in public sector information and shaping of policies in Europe and in the United States.

It would be useful for the Board to look at the secondary literature on measures that range across the differences and effects of institutional rules on access to repositories and on how citation impact on continuing downstream applications of scientific publications are affected. There is a lot of interesting work being done, but a barrier exists that people think the only work that economists do is on commercial valuation of citations.

It would be worthwhile to expand some of the work that BRDI is doing on citation impacts to take in that information, and it would be relevant to some of the agencies to make a meta-analysis of these studies in the social sciences available to the people to look at these methodologies to see where it is applicable.

MR. UHLIR: This would be a separate activity or a separately funded kind of thing because the approved citation activity does not contemplate this kind of issue. However, we have been conversing with CENDI about looking at the value proposition of data and getting a better handle on that. We were talking about having a roundtable discussion next fall. Maybe we can look at something to complement the IP and database symposium that looks at the economics of database access.

MS. CARROLL: Many of the people here know what Earth Science Information Partners (ESIP) is or participate in ESIP. Rama talked about the two groups of data center people who are already convening, one being the geosciences-oriented ESIP and the other being more NASA-centric. Is there a role because data center managers are across discipline? Do the disciplines have their own groups of communities of practice and convening organizations? What is the need to convene across those organizations to get into the interdisciplinary areas? If we choose to stay together and have conversations, how do we work with and interact with ESIP and other kinds of individual communities?

PARTICIPANT: The Research Data Access and Preservation Summit have had two meetings so far. The third one will be at the end of March. It is organized by the American Society for Information Science & Technology and the Coalition for Networked Information (CNI) has been a cosponsor.

MS. CARROLL: Can the representative of the one data center that is at least partially commercially funded make an observation about any differences coming from the private sector as opposed to the federal sector?

MR. BRENDER: GeoEye – the company, not the foundation – has about 1 billion square kilometers of high-resolution of the earth. DigitalGlobe, GeoEye’s competitor, has a similar number. That is a lot of data at half-meter of ground resolution, but that is outside of the WDS because it is, in a way, privately held. Is there any connective tissue between all the discussions at WDS and these sort of private databases, which is not technically scientific data, but it certainly could be used, or even with data centers at Procter & Gamble or Facebook or Google?

MS. CARROLL: If you are thinking of joining, that would open up a whole conversation, and back to the issues of paying, charging, and that kind of thing.

DR. MINSTER: Obviously, we need to talk. One thing to realize is that many WDS nodes are hosted by very large data centers, and not all the holdings of data centers get into WDS. It used to be WDC. Sometimes it was a very small fraction. If there is a disaster and GeoEye wanted to make some of their holdings available at some attractive price or no restrictions, that could be discussed. It could be a small fraction of your holdings.

MR. BRENDER: GeoEye has released newly collected imagery over the area that was impacted by the tsunami in Japan, certainly over Haiti. GeoEye has collected new satellite imagery at half-meter ground resolution over the destruction and has put that up on an FTP site or provided it to Google, who has put it up for free and unrestricted access. But it is a company decision made on a case-by-case basis.

DR. MINSTER: Just having some of those data available is a major step towards qualification as a member.

**International CODATA and WDS Issues and Plans for CODATA Conference in October 2012**

DR. MINSTER: The next segment in this morning’s session has to do with international CODATA and WDS issues and as a secondary goal, the discussion of what may be plans for the CODATA Conference coming up in October. WDS has been structured from the beginning as an international organization. However, this meeting focuses on the U.S. aspects. Nonetheless, some of the questions raised in the U.S. context have either international impact or require international input.

Let us open the floor to suggestions of internationally flavored topics to discuss. I will start. Are the problems that involve very large datasets such as OneGeology datasets and the penultimate national and ultimate international project something that the WDS members here in the U.S. and U.S. CODATA would like to raise to an international discussion?

DR. AHERN: There are two points that we are worried about as seismologists. One is that the datasets are large, e.g., terabytes, so it is not easy to move large subsets of the data to research scientists. We are looking at cloud services, putting the data that is accessible to all people next to where the processors are as well. Because that might be something that is generic, can we do that in a coordinated way rather than each of us individually solving the issue of outsourcing to a cloud service?

The second point, specifically within seismology, is we are working towards international federated web data services and data center models where not all the data has to be at our data center; the data can be anywhere in the world. We want it so that for the metadata, you get the time series data from a variety of international data centers using exactly the same query mechanisms, REST-type web service, and you get exactly the same response back.

DR. MINSTER: So you are saying the data holdings themselves are distributed across the world? A specific effort has to be made so that a scientist in Zimbabwe can access a seismology dataset in Europe or in the U.S. and it would look the same to him/her.

DR. AHERN: But they would also have access to the processing, the CPUs, the cores necessary to apply their algorithm to the data. You move the entire dataset, hundreds of terabytes, to a cloud service and you provide processing services. The cost recovery model becomes more important than just distributing the data. We are not worried about trying to recover those costs. But if you start providing computational resources to people globally, how do we pay for that? That is a problem. There were two things: international federation of seismological datasets and data centers and cycles close to large volumes of data.

DR. BOURNE: It seems that the fundamental and biggest problem of all is: How do you actually get the cooperation you need when data is international and funding is national. This is a specific case we face all the time because we actually have a resource that is run out of Asia, Europe, and the U.S. If any one of those partners was to fail in their individual funding, it is not clear that the dataset could be maintained any longer, and yet there is no cooperation between the funding agencies in those three regions to actually ensure the long-term sustainability of that dataset. Certainly any one of these organizations could make a strong statement to the funding agencies internationally to be more cooperative in this regard.

DR. CHEN: The international CODATA viewpoint on U.S. policy issues related to bilateral collaboration might be to deal with the more general issues of bilateral problems or restrictions on access by certain scientists. For example, in seismology, although on paper, there was access, in reality, there have been difficulties actually getting at that data because of the internal Chinese issues.

Getting IRIS involved in a session in Taipei and bringing in some examples of looking at these access issues as the mirror to the argument of restriction, trying to be bilateral, is living up to its obligations under a multilateral framework of making it accessible, and that is part of the counterargument that one potentially needs here in the U.S.

Having a discussion of that in a multilateral framework in Taipei might be valuable to the previous discussion we were having, as well as the more general issue of other countries having problems, whether it is, to take the example, Taipei or mainland China or other rogue countries where the scientists are legitimately working but constrained by the politics.

DR. DAVID: Regarding the problems with porting seismological data for analysis, distributed analysis and a federated database are two separate technical issues. There are models that maybe people working in the United States and federal agencies have not had an opportunity to look at. For example, European Bioinformatics Institute (EBI) works with a very large explorations base, which is created from federated databases. It has analogs in the U.S. database structure, but I think the federation issue is much more extensive. They have a set of search algorithms to search an extended exploration space. This requires a lot of middleware work to render different kinds of database structures compatible and make them searchable.

CERN, in the ATLAS Project, has this enormous volume of data coming in, which then has to be filtered and then distributed. The filter data is enormous and is filtered through high-bandwidth access to distribute it to people who are going to do analysis looking for the Higgs boson or whatever particular signals are of interest in filtered results of the collision particles.

A lot of fundamental tool-building work, which has much wider potential applications, has gone on in these two different fields. People should be looking at this work as models for things that could be generalized. We should think about some mechanisms to inform people about this fundamental tool-building work.

Both of those organizations are open about making this work available to other people, so these are potential sources. The U.S. should not have a not-invented-here attitude, but should be looking globally. Particularly in world data, we should be looking at what is going on in other parts of the world where people are building tools that could be reused.

DR. MINSTER: That goes back the WDS discussions with CODATA that you are not going to build giant silos of data in places like Pretoria or Buenos Aires, by and large. If you want to help scientific collaboration in countries like this, what needs to be done is to build big pipes and make the data available to local scientists.

Three or four years ago a new World Data Center for Biodiversity and Human Health was created in South Africa. The idea was that they could contribute local data and have access to data collected in Europe and in the U.S. The problem was not lack of bandwidth; there is a fair amount of bandwidth between London and Johannesburg. But it is controlled largely by financial institutions and is unbelievably expensive. Neither Europeans nor North Americans nor the South African government is willing to pay for that bandwidth. This is where the international scientific community might be able to put some pressure. We tried to do the same thing in Brazil and we could not find anybody in Brazil who was interested in building a data center, even though they have some of the best biodiversity data available right there. This is part of the mechanics of breaking this divide or bridging it. How do we solve the political problem if we cannot solve the mechanical problem?

DR. DAVID: Maybe someone should begin thinking about a Personal Income Tax (PIT) on these financial pipes to be applied. It would not have to be a very large tax to be applied to support bandwidth for this kind of scientific data streaming.

DR. MINSTER: Who would pay this tax?

DR. DAVID: You tax financial institutions. These are PIT tax proposals, which have been around for a while. The Europeans have moved towards this. The U.S. has said we will never do anything like that, but that has been floated as a general revenue proposition rather than as a very specific purpose, which greatly reduces the amount of revenue you would have to raise in order to support supplementary bandwidth for things like we have been talking about to add bandwidth.

DR. MINSTER: That is an interesting idea. I wonder whether we can invent taxes.

DR. LOVE: I am all in favor of open access to data. Sometimes data acquisition agencies in other countries operate under completely different economic models. The one that I am kind of familiar with is that sometimes a magnetic observatory program will charge a user for data if that user is interested in real-time services. After the data have been acquired in real time because this user needs them in real time, for example, a directional drilling company for oil or something, the data then become open access.

That is a means by which paying for acquiring the data is important. Maybe that gets back to this idea of taxing the most interested user, and then after the data are no longer real time, they are free and open access. Is that relevant?

DR. MINSTER: It is another aspect of a more general discussion that this community or the community that will get together in Taipei could discuss: How do we finance the flux of data across the planet?

DR. LOVE: It is not just digital divide, it is also real-time divide. Many agencies in the U.S. government are interested in real-time data services. They are also interested in getting data from all over the world, sometimes from countries that provide data that are not themselves interested in real-time services. They may have other issues going on, so there is an awareness divide where some of the agencies in economically developed countries just expect the rest of the world to give them their data, in real time even, when the data are being acquired for other reasons in other countries where real time is not an issue.

So, then pay them for it. Sometimes it can be economically efficient to support data acquisition in economically developing countries if that is what you need. You do not have to go out and deploy your own sensors. Sometimes they are there. Just upgrade them to real time.

DR. DAVID: Somebody from NSF could probably comment on this with more accuracy, but NSF has been successful in finding a way to finance oceanographic satellite images for plankton blooms. The satellite is actually picking up many things, but it is also looking at and capturing images in real time of plankton blooms.

They have managed to get an international association of fisheries industries to pay for the satellite to get real-time access. Some blooms last a relatively short period of time, only 10-14 days, and after that they are gone. But the plankton blooms tell you where the fish are, so they will pay for the satellites to take the images.

A month later, nobody in industry is interested in it at all, but the data is available and can be used as an archive for research purposes later on. You can work this both ways. You can have an end user who will finance the overhead, the infrastructure, to capture images that have transient value. Or you can go the other way and capture images and find a public end user that needs this and figure out how you can create a business model around that.

DR. DROEGMEIER: We need to be careful about the phrase “real time.” I am a meteorologist that deals with very fine-scale local weather. Real time is different than streaming instantaneous data.

It is important for WDS to consider – internationally – the real use of real-time data, not just the financial aspects of getting it in real time and making is available later. By real-time data, we really mean streaming data with the latencies on the order of milliseconds to a second or two. That is real time. That is streaming. That has to be considered because a lot of the experimentation that is done cannot be replicated with data after the fact. You can have all of the data, but if you do not have the circumstances of the moment, you have missed out. You have missed where the humans are. You have missed the reactions. That can never be replicated. You can never capture it.

I would like to elevate the notion of real time and streaming to a much higher level of prominence in our conversation because it is critical. A lot of science depends upon that. So far, we have mostly been talking about archives of data, but we cannot forget that real-time component. And, we need to define what we mean by real time.

DR. LOVE: Real time does not necessarily mean within a millisecond. I would say real time depends on the phenomenon that you are monitoring and the timescale over which it evolves. We do not need to get caught into specific numbers.

DR. HEDSTROM: On the developing country angle of this, it is a problem related to how we carve up the world in different areas and different types of “interventions.” But it seems that if you want to engage developing countries in this enterprise, you have to find the connections between what they need and what you need by way of data.

Often that includes things like contributing to the infrastructure, very importantly, the human capital aspect of this, and the fact that their own data needs may well not be your data needs. But how can you take the data that they need for their purposes, let them drive the agenda of what that is, and still be able to integrate that into the larger world system? There is an emerging field called ICT4D, which is Information Communication Technologies for Development, and it might be useful to engage in a conversation with that group, in particular, about tactics.

DR. LYNCH: The National Library of Medicine at the National Institutes of Health has done some very good work with developing countries in terms of setting up and monitoring of emergent diseases. They might be another group that we could exchange models with.

DR. MINSTER: Should WDS have a discussion with them and discuss the data needs and the long-term curation? We could also talk to foundations who are interested in this class of data.

DR. HEDSTROM: Having spent a fair amount of time in South Africa on archive projects and also participating in the first African conference on digital libraries and archives, which was in Addis Ababa in 2008 or 2009, I was really struck by the degree to which the topic of building digital libraries and archives was totally driven by an environmental sustainability and economic development argument. If you listen to what those communities are asking for, it is justification for these kinds of investments. It is not going to be for the sake of data itself or for participation in a world data system.

DR. MINSTER: I had similar discussions, but what struck me is the desire on the part of agencies and local governments to repatriate as many of their own scientists who have been educated in Europe. These are scientists who want to come back home but will not because there is no data. But if data are there, they might come back home.

DR. HEDSTROM: Or if the delivery of data is in a more real-time sense.

PARTICIPANT: One of the earlier comments touched on the technology. At the American Geophysical Union (AGU) conference last year, there was a well-attended session on people who were specifically building architectural components for reuse, trying to lower the barrier of entry to other groups. It is useful to see how many are out there and for other people to see what features other groups have built in. It is useful to scientists so they do not have to develop these components for themselves and for the people who are building these to see how to interoperate. And it reduces the maintenance because then you have a kind of open source project so not every organization has to maintain it themselves.

DR. BOURNE: A different subject that ought to be in this international discussion is competition, which is a good thing, e.g., the idea that there is competition between nations or regions around data. An example raised earlier is what the European Bioinformatics Institute (EBI) does, versus what the National Center for Biotechnology Information (NCBI) and the National Library of Medicine do here. They operate, in large part, from very similar, if not identical, datasets, but they provide different views on that data in a sort of quasi-competitive way. But the end result of all that is good for the scientists. That kind of internationalism and national pride in the data is a good thing.

MR. LEVITUS: At the National Oceanographic Data Center and World Data Center for Oceanography we make all of our data available online, but we also make it available on DVDs specifically because of this problem that less developed countries do not have the bandwidth to download all of these data, and yet they still make important oceanographic observations that they are willing to send to us. We just do not want to lose their goodwill. We do make everything available.

A student, for example, can get access to the most comprehensive global ocean profile database on one or two DVDs and do first-rate scientific work with it. We consider that to be important. Although we have heard many discussions about putting everything online and bandwidth, we continue to work with DVDs and get quite a few requests for those. It is important to keep in mind.

DR. DUKE: In thinking about connections with the developing world, what if we look at that in a human arts context? The American Association for the Advancement of Science (AAAS) has a group called the Science and Human Rights Coalition, which is a coalition of several professional societies including my own , the Ecological Society of America (ESA), which, among other things, is doing a lot of work on the UN Declaration of Human Rights to the Benefits of Scientific Progress.

Some of the foci in there are with regards to indigenous people and people in the developing world, rights to the benefits of information from their cultures, including things like natural history data and potential medicines from plants. Data is an inherent part of that discussion. This context might provide some additional linkages for us in our discussions with them.

DR. CHEN: Since the meeting is in Asia again, there are certain issues that are going to be more interesting. Jane Hunter, who was mentioned as the chair of the CODATA materials task group, is also the chair of the Australian National Committee for CODATA and has many other activities, including working with indigenous peoples and community science. In fact, she came and gave a talk on open community-driven science at Columbia last year.

A topic appropriate to focus on and bring in constituencies in Asia is indigenous knowledge. This would be very appropriate if people were thinking of particular areas where they had interest and wanted to organize something. This is a good opportunity to bring in those kinds of communities in thinking about any activities that might focus on Taipei or on the WDS interaction, especially with the secretariat in Japan. That is another opportunity to think about disaster data.

DR. MINSTER: Natural science data like astronomy data for virtual observatories or local observations of the earth by satellite, whether it be the ocean or topography or global networks like seismology or magnetic observatories, is one class of data where collecting and managing those data requires large, well-heeled facilities with a lot of technology in a country that has the resources, like the U.S. or in a country in Europe. But for things like disease, biodiversity, and human welfare, a lot of work must be done right there on the ground in countries that do not necessarily have large data centers or very large capabilities for that. Is that a general issue that WDS and CODATA should explore? Should we deliberately move more aggressively in areas that are characterized by this different kind of data collection, data management systems?

DR. ENTWISLE: It is important to have social and human data in the mix. It is also important to recognize that, just as it is true with other groups, there are activities already underway, such as specifically a large archive called CESSDA, the Council of European Social Science Data Archives, in the European Union. There is a special part of CESSDA, which is called the European Research Infrastructure Consortium (ERIC). They have been working on a legal framework for sharing, and have been working out a lot of the details for the sharing of certain kinds of survey and administrative data.

The CESSDA-ERIC right now has four to five members who are full members. They are hoping to expand it to the whole European Union, but it provides a structure for other countries also to participate. It raises another issue, which is that in this group we have been talking about individual countries or the whole world. There are regional groups that also need to be acknowledged in thinking about how to go forward.

DR. MINSTER: This could be developed into a session or set of sessions at the CODATA Conference.

To summarize the discussion so far, we have talked about:

* Increasing the bandwidth and using bandwidth more efficiently, including financing the bandwidth by some kind of tax
* Education and human capital and identifying the data needs of countries that are not as well equipped to deal with datasets
* NIH and environment sustainability, which is new for us, so we will have to get the right kind of expertise
* Human rights and a framework for sharing surveys and data

These are ideas that might be discussed in the international context at the CODATA Conference.

DR. LYNCH: An important one, partially related to bandwidth that is not on the list is the one about real-time use of data, because there is a lot more to that than just bandwidth. That is an important emerging new scenario.

MR. DIAMOND: Another important issue is the rescue and digitization of old analog data. At least from a climate standpoint, there is a treasure trove around the world in various back rooms of meteorological services and such. There are activities out there, like an activity called the Atmospheric Circulation Reconstructions over the Earth (ACRE), which is looking at trying to work in things like citizen science activities to rescue these data sources. While some of this can be done by students and other interested persons, there is a need to look at finding resources for this and a better coordination amongst activities.

The best thing we have been able to do, out in the South Pacific, is try to organize the paper and put it in the acid-free boxes and cross our fingers that they do not get wiped out. This may be the case in other fields of study as well. There are lots of data out there, but some of it is getting so old that you walk into a back room and smell a lot of vinegar because the acetate is breaking down the microfiche film and paper. Trying to put sustainable systems in to get new data is important, but there is a lot of data out there that is at risk for never being available for anyone again.

DR. MINSTER: Speaking for WDS, these are areas where we could do something. If the World Bank were to finance someone to do citizen science or a quad sourcing type of experiment, we could conceivably join forces and find a place for the data to be suitably stored, redistributed, reused, and/or made more available. But it would not be part of our purview to actually run these experiments.

DR. CHEN: CODATA has a task group on data rescue. John Faundeen is co-chair. There will definitely be some data rescue sessions in Taipei. I will make sure they touch base with you in terms of organizing that. Also, I work on a DataNet with University of Minnesota, which has data rescue for census tapes. The tapes and paper records are found in some places rotting. This is definitely a cross-disciplinary topic. This will be on the agenda for Taipei and we welcome lots of input. Certainly the task group, which is only trying to survey the problem, could use help in taking the next step to actually getting people organized and more resources involved in actual data rescue.

PARTICIPANT: In that context, an issue that crosses several disciplines is the availability or discoverability of the actual physical specimens and samples that underlie many of the data in natural sciences. That is an aspect that can or should be addressed internationally as well.

In our field it comes down to rocks in quarries that are often generated through international programs such as ocean drilling and continental drilling programs. Issues come up all the time that the actual specimens, also in developing countries, are being discarded in universities where there is no way to save them to maintain repositories and so on. That would be good to add to the agenda.

DR. GRAVES: Along with whether the data is large, small or whatever, there is the whole issue of providence, lineage, capturing whether it is the metadata on the rocks or whether it is the data that we are trying to rescue. Some parts of the world are addressing that more than others. Is that a topic where WDS, in connection with some of the international data centers, can help? Are there tools and techniques that people are using that can be offered? Of course, for providence, it is always better to capture it at the moment the data is created. That is an issue that needs to be discussed nationally and internationally.

DR. BOURNE: In a lot of this discussion, our thought processes relate to specific datasets that we happen to know exist, and then using those as examples to illustrate our points. But the point, generally speaking, seems to be the same, and the point is to make sure there is some form of preservation. Realistically, we can say that, but that is not doing anything about it. Resources to do something about it are going to be very limited. Maybe a function of CODATA or WDS could be to catalog or the most important datasets under threat, and then make recommendations regarding what should be done specifically with those.

DR. MINSTER: We need a mechanism for prioritizing the activities we can get involved with.

DR. SPENGLER: The general repatriation issue of data, samples, and people, is that, in general, if you are considering the developing world, you have to remember that they are not just a source, they are people. There are rights inherent.

DR. HEDSTROM: You may want to look at the UNESCO Memory of the World site where they are doing work in social science, earth science, and culture with regard to collections at risk. This could help with prioritizing.

DR. MINSTER: Thank you very much. We have a number of ideas. We will write a memo to the WDS and send you a copy. Then we will look for volunteers to organize and recruit speakers and/or work with an organization or company to organize something at Taipei in collaboration with CODATA.

 **Concluding Observations**

Bonnie Carroll

As we conclude this forum, we need to discuss action items, or next steps. Bernard took away some ideas for programs, and I am sure he will be calling on individuals to participate. We have had a very interesting discussion, and it is clear there are a lot of issues. Not only did we identify issues, but we could not resist talking about them. We still only skimmed the surface. There are clearly topics on which people would like to continue the conversation.

The closing of this is back to you folks who feel you are going to be participating or would like to participate in the World Data System or in a U.S. group of data center, a community of practice, which all may or may not have to actually participate in the World Data System.

So, what is the next step? Are there next steps for this particular community of practice in relation to U.S. CODATA and what we might do that would be useful to you? Where would we go from here? Here is the range of options: Do nothing and say this has been an interesting conversation. Or minimally, we can plan to do this kind of forum once a year so that we get together and form networks and have one-on-ones. This way we can hear what other problems are too. We would like to have some kind of continuing conversation more often. Then it is a matter of how one does this kind of thing.

MR. LEVITUS: Can the position of the world data centers and other organizations within WDS be made more clear perhaps, or at least the U.S. components of WDS be made more clear? In particular, can it be made clear that we represent non-governmental organizations, because that is so important to me and to other WDC directors in being able to do things for the example I gave with the government of Taiwan. Can there be a formal statement of what our role is under ICSU, for example? A simple statement would be something that would be valuable that we could refer to.

MS. CARROLL: That is an important point. The world data centers were an entity apart from whatever data center you run for the United States government. There was something then that was a world data center. Bernard, is that continuing to be the case, or is it really whatever you are you apply to be, and then it is up to you to create?

DR. MINSTER: There are rules, and the rules continue to be the same as the WDC. If you are the National Oceanographic Data Center (NODC) and you have datasets that you want to make a part of the WDS holdings and you do not want to live with the rules, then we have to tell you no, this is not part of WDS holdings.

MS. CARROLL: This is a somewhat different question. Or are you okay with the answer that if you want to have something that is not your federal data center, like a world data center, how do you create that yourself? There used to be a different mechanism that went away when the World Data Center System when away.

DR. MINSTER: That will depend completely on the location and the country. For instance, we have only one data center in South Africa, but the South Africans have developed a national data policy that is very consistent with the GEOSS data policy, so their data is basically available with no obstructions.

MR. LEVITUS: We had the principles of the WDS, but what I have to do very often – every time we get a new director at NODC – is explain that being director of the WDC is separate from a U.S. government function. I was appointed by an NRC committee and it is not a good idea to be a WDC director and also director of NODC. That is my problem.

A lot of people simply do not understand the separation, and it would be valuable if there were some simple document. I can look back at the old documents that describe the World Data Center System. Perhaps if those were updated and it was reemphasized that these are non-governmental organizations, the WDCs, that would be enough.

MS. CARROLL: Paul, will the NRC have any program to create this kind of thing anymore, or has that gone away to create a non-federal entity that shares the data of perhaps a federal data center?

DR. CHEN: We need to back up. There is a misunderstanding. When I was talking about the Committee on Geophysical and Environmental Data, my perception is that it reviewed and approved candidates for joining the World Data Center System as it existed back then, and that whoever was running it could change. It did not appoint a director of the World Data Center. That was not the Academy’s mandate, from my perception of how it was run under Ferris Webster. The Academy did not take on the role of appointing a director of a World Data Center.

It could be that in the federal government, the interpretation was a little different, but I do not think, moving ahead, that it is likely that this board or any other board would say we are appointing directors of the World Data System. They might give you a peer review process or something that would enable people to qualify. Even if your perception was different from mine historically, it is not clear to me that you should expect that your perception will continue. We can argue about what happened before, but I am not sure that is going to be the model that is feasible.

DR. MINSTER: Chris, I am going to put you on the spot a little bit. The issue is going to be the (?) agreement between ICSU and WDS node and who actually commits to WDS activities.

DR. FOX: The system now is that we were asked if we wanted to apply. We did and we nominated a director. In my case, I had three world data centers within one national data center, and we decided to combine these into one data center. I made myself the director of both the national data center and the world data center with the intention of delegating it down to people like Eric Keane and others that you know.

I differ with Syd. I think that the World Data Center, World Data System title allows us to do a lot of things particularly well by working with our international partners. Sometimes it is better to not be working with a U.S. government agency. To me, our legal responsibility to the federal government trumps anything. I would not do something under the guise of the World Data System that was against the rules of the federal government.

MR. LEVITUS: Not that long ago, I invited someone from Taiwan to visit. I also arranged for them to visit your center. Margarita Conkright, who is director of NODC, could not have done that. She investigated the situation and found out she could do nothing more than send a letter on a blank piece of paper. She could not use her official government title, which is what you are saying.

That is the advantage of having world data center directors separate from a U.S. government data center. This is where I think, from my memory, that I was appointed by the Committee for Geophysical and Environmental Data for this particular reason. This is why I think it would be worthwhile and important that future world data center directors be appointed by an NRC committee.

It is the NRC in the U.S. that oversees, or used to, the World Data Center System in the U.S. because it is non-governmental. When the World Data Center for Oceanography was first formed in the 1950s, it was at Texas A&M University. Then it was eventually moved to be associated with the National Oceanographic Data Center. But it is important to maintain this separation.

I have had some dealings with the State Department, not about the integrity, but about the independence of the World Data Center System. They expressed support. They said this is not governmental and it is of importance to the U.S. I have never had a problem governmentally, and I always felt the State Department would back me up. I forgot the exact details, but this is what happened some years ago. I have invited and worked with people from the Russian Navy and never had a problem with getting either a passport or visa, et cetera. I do it all under the letterhead of the ICSU World Data Center System. This is a good thing to maintain.

DR. MINSTER: Would it solve your problem, if there was a letter of appointment sent by ICSU appointing you as WDS Director and copying the National Academy as the ICSU national member?

MR. LEVITUS: Yes, that would be a good thing to do for all world data center directors. But people have to decide if they want to be directors of world data centers and of U.S. national data centers, because I think it is an inherent conflict of interest. The WDS is a non-governmental organization, and I think they should be separate, the two positions. That is how it was explained to me when I became Director of the World Data Center for Oceanography, and I have used it to my advantage.

DR. FOX: Whether you are the director or another position within the organization, you are still subject to the regulations of the agency. But I would also say that I would never have bothered to check to see if I could have someone from Taiwan visit. I would just say yes. I do think you can do it. I think the State Department does encourage those sorts of things.

I will say that I am also the U.S. representative to GEBCO, the General Bathymetric Chart of the Oceans, which is an international mapping organization. I was nominated and elected by the International Hydrographic Organization into that position, so I have that hat. You can have these different hats and you can use them as leverage to say this is why you need to do that. I do not think it matters whether you are the director or not the director. You still have to play the system you are in. I do not think the committee is going to solve this one. I think this is really an internal government sort of dance.

MS. CARROLL: What I am hearing is: What is the role of the Academy to allow the World Data System to have the U.S. centers? BRDI needs consider this as we talk about BRDI as the home of U.S. CODATA. What is the next step? Do you want to have another conversation and in what timeframe and how to include people or not?

MR. DIAMOND: I can definitely see some benefit to an annual meeting like this where there is just an exchange of information and people get to hear about activities that are dealing with CODATA that cross lines. At a minimum, some kind of loose association of people, such as a mailing list that we establish that U.S. centers could use as a forum for talking to one another, an information sharing and exchange forum of some sort.

DR. AHERN: This forum is where we might raise issues. But we could consider forming a working group when we identify a specific issue that we want to talk in more detail about. We should form the working group with a subset of this group and possibly others from our organizations, and talk thoroughly about the specific issue. Working groups are where work happens. That helps us move forward.

DR. RAMAPRIYAN: This meeting is good in terms of raising issues, but we are only skimming the surface; we have spent about half a day doing this. Also, we have other forums where within our own spheres we have more in-depth meetings to talk about things. The value of this kind of meeting is that we are bringing in people who are not in our own disciplines. For example, my involvement is largely with earth science satellite data. The kinds of problems we talk about are somewhat specific to that.

Here we have people from social sciences, seismology, astrophysics, and others, which we do not discuss in our own spheres. Here, we are looking at things at a somewhat higher level, which means we are not going as deeply into some issues as we do in other forums. We need to define some sort of an agenda that provides the additional value because of these groups coming together from multiple disciplines.

MS. CARROLL: We have a long list of issues. We need to prioritize where the issues might be of real value to have additional conversation on to essentially define the agenda. Then we will figure out if anyone is interested in working on such a thing.

So, we have heard the issues and about where you all are coming from as data center managers. It all goes back to our co-chairs of BRDI, CODATA, and the team and Paul to figure out the resources available and how this is to be done.