



THE NATIONAL ARCHIVES AND RECORDS ADMINISTRATION

GDFR · *Global · Digital · Format · Registry* ·

THE ELECTRONIC RECORDS ARCHIVES (ERA) PROGRAM

HARVARD UNIVERSITY



Proceedings of the Global Digital Format Registry Governance Workshop

14-15 November 2007

Abstract: In order to serve the goal of advancing knowledge and experiences relevant to ensuring the long term sustainability of an international Global Digital Format Registry (GDFR), a governance workshop was held to examine the issues and approaches to the development and implementation of governance models in open, international collaborative based environments. The workshop explored how various consensus driven organizations approach the issue of governance, sharing lessons-learned and best practices.

The workshop was open to participation by all interested parties and stakeholders. The international participation represented many sectors including governments, libraries, archives, research and education communities, and more. The workshop gathered together practitioners in the area of governance from the public and private sectors to discuss approaches and accomplishments in ensuring the sustainability of open, collaboration-based initiatives.

Table of Contents

1	Executive Summary	3
2	Background	5
3	Participants	7
4	Workshop Objectives, Approach, Structure and Deliverables	8
5	Agenda	10
5.1	Published Agenda	10
5.2	Agenda Modifications	10
6	Presentations	11
6.1	"Global Digital Format Registry (GDFR)" Stephen Abrams, Harvard University Library	11
6.2	Workshop Keynote: Peter Brantley Executive Director, Digital Library Federation	11
6.2.1	"GDFR, Roles and Challenges" Peter Brantley, Digital Library Federation.....	11
6.2.2	"Overview of the Digital Library Federation (DLF)" Barrie Howard, Digital Library Federation	12
6.3	"Handle System Governance" Laurence Lannom, Corporation for National Research Initiatives	12
6.4	"Governance as a Socio-Technical System" Kate Ehrlich, IBM Research	13
6.5	"Infrastructure for Preservation" David Giaretta, Director of CASPAR Project and Associate Director, UK Digital Curation Centre	13
6.6	"A Distributed Science-Data-Description Registry/Repository: Implementation and Operational Experience" Donald Sawyer, NASA/GSFC, and John Garrett, SPSystems	14
6.7	"Practical Aspects of Managing a Standard, a Registry, ..." Sally H. McCallum, US Library of Congress	14
6.8	"National Digital Information Infrastructure and Preservation Program (NDIIPP)" Caroline Arms and Carl Fleischhauer, US Library of Congress	14
6.9	"University Perspectives on Governance: MetaArchive Cooperation & DSpace Foundation" Tyler O. Walters, Georgia Tech Library & Information Center	15
6.10	"National Software Reference Library: Computer Forensics Tool Testing" Barbara Guttman, US National Institute of Standards and Technology	15
6.11	"Engineering Formats and GDFR Governance" Josh Lubell, Manufacturing Engineering Laboratory, US National Institute of Standards and Technology	15
6.12	"Information Discovery, Use and Sharing" Mary Brady, Information Technology Laboratory, US National Institute of Standards and Technology	16
6.13	"From Stovepipes to Wind-chimes: Advancing Agility Through Communities of Practice" Susan Turnbull, US General Services Administration	16
7	Discussion	17
7.1	Organizational Beginnings and Evolution	17
7.2	Organizational Characteristics	18
7.3	Stakeholders	19
7.4	Funding	20
7.5	Marketing Strategy	20
7.6	Content Management	21
7.7	Issues	22
8	Going Forward	24

1 Executive Summary

Problem Statement: The wide diversity and rapid pace of adoption and abandonment of digital formats present an ongoing problem for long-term preservation efforts as noted in the October 2002 planning report of the Library of Congress (Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure Preservation Program; <http://digitalpreservation.gov/about/planning.html>), "Longevity of digital data and the ability to read those data in the future depend upon standards for encoding and describing, but standards change over time."

Background: The Harvard University Library (HUL) has been operating under a grant from the Andrew W. Mellon Foundation for a two-year project for the development of a registry of authoritative information about digital formats. Detailed information about the format of digital resources is fundamental to their preservation. The project will result in a new Global Digital Format Registry (GDFR), which will become a key international infrastructure component for the digital preservation programs of libraries, archives and other institutions with the responsibility for keeping digital resources viable over time.

GDFR Governance Workshop: Global registries, such as GDFR, are valuable insofar as they are sustainable. Communities develop sustainability initiatives in many different ways. As a first step to help formalize a structure for the ongoing sustainability of the GDFR, the US National Archives and Records Administration (NARA) Electronic Records Archives Program (ERA), working in partnership with Harvard University Library, hosted this workshop to initiate identification, discussion, and evaluation of options for a potential, sustainable GDFR governance structure.

Workshop Objectives were two-fold: (1) Define the framework for discussion surrounding the mechanisms and organization required for the operation, management and coordination of an International Global Data Format Registry as a global community of collaborating International members; and (2) Identify, discuss, and document issues relative to establishing governance models in order to contribute to future discussions.

Participants: The GDFR governance workshop was open to participation by all interested parties and stakeholders. The diversity and interest in the GDFR project is great, spanning international participation, and representing many sectors including governments, libraries, archives, research and education communities, and more. The workshop gathered together practitioners in the area of governance from the public and private sectors to discuss approaches and accomplishments in ensuring the sustainability of open, collaboration-based initiatives.

Approach: Participants examined the issues and approaches to developing and implementing governance models in open, international collaborative based environments. They explored how various consensus driven organizations approach the issue of governance, and shared lessons-learned and best practices with the central goal of advancing knowledge and experiences relevant to ensuring the long term sustainability of a global digital format registry, such as the Global Digital Format Registry (GDFR) initiative, led by the Harvard University Library (HUL).

Presentations: In order to bring all participants to a common footing on organizational and governance aspects of international format representation and registries, a broad spectrum of presentations was made. The presentations covered relevant diverse topics including: organizational structure, governance and funding models; lessons-learned and issues identified in the establishment, evolution, and maturity of registry, open-source software, and standards organizations; the socio-technical issues of multi-organizational international undertakings; technical aspects of format representation; challenges of format representation from specific domains such as health care, engineering, and bio-molecular science; and the requirements and techniques of e-Discovery and digital forensics.

Discussion: The presentations were followed by a half-day of facilitated, participant-directed discussion which covered topics of GDFR Organizational Beginnings and Evolution, Stakeholders, Organizational Characteristics, Funding, Marketing Strategy, and Content Management. The final topic of discussion was a strawman plan for going forward with an initial Global Data Format Registry organization.

Results: The workshop presentations and discussion was captured in these Proceedings. Based on this material, the following actions were suggested for moving forward

- Establish a GDFR Advisory Council including a light weight governance model for initial decision making and tasking.
 - Write the GDFR Mission Statement
 - Write the GDFR Business Plan
 - Establish the Advisory Group Communication Plan
 - Deploy the infrastructure necessary for the communication plan
- Schedule the next meeting of the Advisory Group (to include a demonstration of the GDFR System work done by Harvard & OCLC. The goal of this meeting is to address the issues raised in the Governance workshop and to work up a strawman organization and governance model for the next phase of the GDFR organization.

The immediate plan for going forward is the vetting and approval of these proceedings by the workshop participants, and thereafter the initial formation of the GDFR Advisory Council.

2 Background

Problem Statement – The wide diversity and rapid pace of adoption and abandonment of digital formats present an ongoing problem for long-term preservation efforts as noted in the October 2002 planning report of the Library of Congress (Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure Preservation Program; <http://digitalpreservation.gov/about/planning.html>), "Longevity of digital data and the ability to read those data in the future depend upon standards for encoding and describing, but standards change over time."

According to Dale Flecker, Associate Director for Planning and Systems of the Harvard University Library, "All digital preservation programs must document the format of the objects they are preserving. Without precise knowledge of format, a digital object is merely a collection of undifferentiated bits. Creating a shared registry of such documentation will save an enormous amount of duplicative effort in acquiring and recording such documentation. It also allows the community to share expertise in formats, so that each institution does not require deep local expertise in every format of data it is preserving."

GDFR System Development To-Date – The Harvard University Library (HUL) has been operating under a grant from the Andrew W. Mellon Foundation for a two year project for the development of a registry of authoritative information about digital formats. Detailed information about the format of digital resources is fundamental to their preservation. The project will result in a new Global Digital Format Registry (GDFR), which will become a key international infrastructure component for the digital preservation programs of libraries, archives and other institutions with the responsibility for keeping digital resources viable over time. Development of the Registry will be informed by the considerable expertise in digital preservation the Harvard libraries have acquired through Harvard's Library Digital Initiative (<http://hul.harvard.edu/ldi/>). An earlier Harvard contribution to the international digital preservation community is JHOVE (<http://hul.harvard.edu/jhove/>), a tool developed in cooperation with JSTOR (<http://www.jstor.org/>) that is widely used to analyze and validate the format of digital objects.

The GDFR system is being established as a distributed service in which participating research libraries, archives, and other organizations with preservation responsibilities can contribute, as well as use, format-typing information. According to Stephen Abrams, digital library program manager in HUL's Office for Information Systems (<http://hul.harvard.edu/ois/>), "GDFR will be a sustainable service available to any preservation institution that chooses to participate. From the outset, we've envisioned the registry as a distributed network of individual 'nodes.' Each node will have a full copy of all the format-typing data in the GDFR. Carefully vetted information and updates will be distributed among the nodes following appropriate technical review. GDFR will also provide a separate track for distributing non-vetted information, so that problems and issues identified in the course of daily work can be quickly shared by participants."

Major American research libraries are supporting Harvard's efforts to develop the GDFR. MacKenzie Smith, associate director of technology for the MIT Libraries, stated, "The establishment of a digital format registry will be a major contribution to our ability to keep digital content viable into the future, and I am grateful that Harvard is willing to take the initiative to build it and coordinate our efforts to use it." In the words of John Ockerbloom, digital library planner and architect for the University of Pennsylvania Library, "Such a system will aid in digital development and preservation not only at my library, but also at many other institutions worldwide. Having open, globally recognized naming, definitions, and documentation of data formats will greatly improve the abilities of libraries and content-management software to use, adapt and share a wide variety of digital content."

Additional information on the GDFR can be found on the project web site at <http://hul.harvard.edu/gdfr/>.

GDFR Governance Workshop – Global registries, such as GDFR, are valuable insofar as they are sustainable. Communities develop sustainability initiatives in many different ways. As a first step to help formalize a structure for the ongoing sustainability of the GDFR, the US National Archives and Records

Administration (NARA) Electronic Records Archives Program (ERA), working in partnership with Harvard University Library, hosted this workshop to initiate identification, discussion, and evaluation of options for a potential sustainable GDFR governance structure.

The Global Digital Format Registry Governance Workshop was held from 14-15 November 2007 in the Jefferson Conference Room of the US National Archives and Records Administration, 700 Pennsylvania Avenue NW, Washington, DC.

The workshop was hosted by NARA which has a continuing interest as a perpetual party in the digital preservation community. NARA is interested in being a partner in an international GDFR undertaking, as opposed to maintaining a GDFR by itself.

The workshop was designed to raise awareness among the participants about current governance practices in consensus-driven organizations, the common threads of successful governance approaches in various sectors, how effective governance examples may provide relevant models for a GDFR governance structure, as well as to consolidate this group awareness in these proceedings.

GDFR Governance Workshop Participants – The GDFR governance workshop was open to participation by all interested parties and stakeholders. The diversity and interest in the GDFR project is great – spanning international participation and representing many sectors including governments, libraries, archives, research and education communities, and more. The workshop gathered together practitioners in the area of governance from the public and private sectors demonstrating approaches and accomplishments to ensure the sustainability of open, collaboration-based initiatives.

Participants examined the issues and approaches to developing and implementing governance models in open international collaborative based environments. They explored how various consensus-driven organizations approach the issue of governance and shared in lessons-learned and best practices with the central goal of advancing knowledge and experiences relevant to ensuring the long term sustainability of a global digital format registry, such as the Global Digital Format Registry (GDFR) initiative, led by the Harvard University Library (HUL).

3 Participants

Caroline Arms,
Library of Congress

Stephen Abrams,
Harvard University Library

Mathew Black,
National Library of New Zealand

Mary C. Brady,
U.S. National Institute of Standards and
Technology

Barrie Howard,
Digital Library Federation

Rita Cacus,
NARA

Robert Chadduck,
NARA

Mark Conrad,
NARA

Kate Ehrlich,
IBM Watson Research Center

Mark Evans, Tesella,
ERA Program

Carl Fleischhauer,
Library of Congress

Magnus Geber,
Sweden Statens Arkiv

David Giaretta,
Digital Curation Centre

Barbara Guttman,
U.S. National Institute of Standards and
Technology

Larry Johnson,
Workshop Facilitator/Editor
TethersEnd Consulting

Chris Jordan,
San Diego Supercomputer Center

Laurence Lannom,
Corporation for National Research
Initiatives

Josh Lubell,
U.S. National Institute of Standards and
Technology

Sally McCallum,
US Library of Congress

Quyen Nguyen,
NARA

John Ockerbloom,
University of Pennsylvania

William Reilly,
MIT Libraries

Jennifer Rigby,
The National Archives of the United
Kingdom

Donald Sawyer,
U.S. National Aeronautics and Space
Administration

Barbara Sierman,
Koninklijke Bibliotheek

Richard Steinbacher,
NARA

Margaret Stewart,
Bibliothèque et Archives Canada

Julie Sweetkind-Singer,
Stanford University

Ken Thibodeau,
NARA

Susan Turnbull,
U.S. General Services Administration

Matthew Walker,
National Library of Australia

Tyler Walters,
Georgia Institute of Technology

Erum Welling,
NARA

Kate Zwaard,
U.S. Government Printing Office

4 Workshop Objectives, Approach, Structure and Deliverables

Mr. Kenneth Thibodeau, Director of the Electronic Records Archives Program and host of the workshop, welcomed the participants. Following the welcome, each participant provided a brief introduction to themselves.

Larry Johnson of the ERA program served as workshop facilitator and introduced the workshop objectives, approach, and deliverables.

Objectives:

- Define the framework for discussion surrounding the mechanisms and organization required for the operation, management and coordination of an International Global Data Format Registry as a global community of collaborating international members.
- Identify, discuss, and document issues relative to establishing governance models contributing to future discussions.

Approach:

- A series of presentations to bring participants to a common footing followed by facilitated discussion of the topics introduced by the presentations and any additional germane topics.
- Examine the issues, options and approaches to developing and implementing governance models in open collaboration-based environments.
- Share in lessons-learned and best practices with the central goal of advancing knowledge and experiences relevant to ensuring the long term sustainability of a global digital format registry
- Explore current governance examples that may provide relevant models for a GDFR governance structure; consider and discuss recommendations for the development of GDFR policies, organizational responsibilities, potential funding models, and oversight of a GDFR editorial process.

Workshop Structure:

The workshop format consisted of two components:

1. One day of presentations and discussion that examined the issues of governance in collaborative consensus-driven organizations. Aimed at gathering existing and potential global digital format registry stakeholders to learn about various governance approaches in collaborative environments, the workshop explored current governance examples that may provide relevant models for a GDFR governance structure; considered and discussed recommendations for the development of GDFR policies, organizational responsibilities, potential funding models, and oversight of a GDFR editorial process.

2. A half-day of facilitated open forum discussion that focused on the following key questions:

- What decisions must be made to ensure effective management and use of a shared global digital format registry?
- Who should make these decisions?
- How will these decisions be made and monitored?
- What are the logical next steps to move this process forward?

Deliverables:

- The "Proceedings" documents workshop results, including
 - Workshop presentations
 - The framework of issues germane to GDFR Governance
 - The initial discussion around those issues

- Potential governance requirements
 - Action items and Recommendations regarding next steps
- Definition of a continuing “Governance Steering Group” to advance, discuss, and resolve the governance issues and provide recommendations for the development of GDFR as an organization.

5 Agenda

5.1 Published Agenda

Day 1: Wednesday, November 14, 2007

GDFR Project Update and Governance Presentations

- 8:30 – 9:00** **Check-in | Breakfast**
- 9:00 – 9:05** **Welcome and Introductions:** Ken Thibodeau, NARA, Director, ERA Program
- 9:05 – 9:15** **Workshop Objectives, Deliverables, and Approach:** Larry Johnson, Workshop Facilitator
- 9:15 – 10:00** **GDFR Project Update:** Stephan Abrams, Harvard University Library
- 10:00 – 10:15** **Break**
- 10:15 – 11:15** **Presentation 1:** Peter Brantley, Digital Library Federation
- 11:15 – 12:00** **Presentation 2:** Larry Lannom, Corporation for National Research Initiatives (CNRI)
- 12:00 – 12:45** **Presentation 3:** Kate Ehrlich, IBM Watson Research Center
- 12:45 – 1:45** **Lunch**
- 1:15 – 1:45** **Optional Presentation: ERA Research Lab Demonstrations,** Mark Conrad, NARA
- 1:45 – 2:30** **Presentation 4:** David Giarretta, Digital Curation Centre
- 2:30 – 3:15** **Presentation 5:** Donald Sawyer, U.S. National Aeronautics and Space Administration
- 3:15 – 3:30** **Break**
- 3:30 – 4:30** **Presentation 6:** Carl Fleischhauer, Caroline Arms, and Sally McCallum, Library of Congress
- 4:30 – 5:00** **Presentation 7:** Tyler O. Walters, Georgia Institute of Technology

Day 2: Thursday, November 15, 2007

Governance Presentations | Summary of Issues | Open Forum Discussion

- 8:30 – 9:00** **Check-in | Breakfast**
- 9:00 – 9:45** **Presentation 8:** Josh Lubell, Mary C. Brady
U.S. National Institute of Standards and Technology
- 9:45 – 10:15** **Presentation 9:** Susan Turnbull, U.S. General Services Administration,
Office of Intergovernmental Solutions
- 10:15 – 10:30** **Break**
- 10:30 – 10:45** **Summary of Issues and Points:** Larry Johnson
- 10:45 – 11:45** **Open Forum Discussion:** facilitated by Larry Johnson
- 11:45 – Noon** **Key Questions:**
- Is there interest in next steps?
 - What do people want to see happen next?

Noon **End of Workshop**

5.2 Agenda Modifications

Due to an illness, Peter Brantley of the Digital Library Federation was unable to deliver the keynote. An abstract of his presentation as well as URL's to his slides and an audio recording are provided below. A presentation by Barabara Gutman of US NIST was added: "National Software Reference Library: Computer Forensics Tool Testing."

6 Presentations

The presentations made at the workshop are listed here with abstracts and URL's to slides and supporting material where available. The presentation abstracts and links have also been posted to the web at URL:

<http://www.archives.gov/era/research/gdfr-presentations.html>

6.1 "Global Digital Format Registry (GDFR)" Stephen Abrams, Harvard University Library

Slides: <http://www.archives.gov/era/research/abrams.ppt>

Mr. Abrams provided an overview of the Global Digital Format Registry project operated by Harvard University Library under a Mellon Foundation grant. The goals and status of GDFR development taking place at the Online Computer Library Center (OCLC) was reviewed. The status of the project was discussed covering accomplishments and work that remains. Mr. Abrams stressed that this project covers technical issues of content and system technology, and does not address the succession of the work to a maintaining organization and attendant issues of content, system, and organizational governance; issues to be addressed in this workshop.

6.2 Workshop Keynote: Peter Brantley Executive Director, Digital Library Federation

Due to illness, the keynote speaker, Mr. Peter Brantley, Executive Director of the Digital Library Federation (DLF) was unable to attend the workshop. He provided his presentation slides, however, as well as an mp3 audio file speaking to them. A summary of his talk is included with links to his slides and the audio file.

Mr. Barrie Howard of the DLF graciously agreed to speak on very short notice in Mr. Brantley's absence. He provided an extemporaneous overview of the Digital Library Federation (no slides are available). Mr. Howard's remarks are summarized below.

6.2.1 "GDFR, Roles and Challenges" Peter Brantley, Digital Library Federation

Slides: <http://www.archives.gov/era/research/brantley.ppt>
Audio: http://sunsite.berkeley.edu/brantley/brantley_gdfr.mp3

Addressing the perspective of organizational form, he recommended a 501(c)3 corporation. Of the two forms, (membership-based and charitable) he pointed out that as a charitable organization, donations are allowed. The Board of Directors should represent diverse perspectives. The organization should be driven primarily by governmental and academic organizations rather than commercial ones.

From the outset the organization needs at least an informal business plan emphasizing sustainability, and it is critical to have funds for at least 4-5 yrs. The funds should be sufficient to cover an adequate staff which should be in place at the beginning, rather than building slowly person by person starting with an executive director. For the long term it is advantageous to start early in soliciting and build an endowment. Other funding approaches such as subscription (for what?), pay for use (difficult to engineer), charitable contributions (from whom?) need to be considered as well. Solicitation for contributions to build an endowment should focus on corporations, particularly large ones and those that actually designed formats that need to be sustained for preservation. These early exemplar corporations will serve as attractors for additional contributions from other corporations.

There is some danger that the GDFR could become an orphan service should it overly focus on the historical perspective of file and document orientation. The world is turning to the web for access, and the concept of file and document is breaking down in this more fluid environment. The web supports formats such as text/html/xml, flash, pdf. However, rich internet applications such as those emerging from Adobe, Microsoft, Google, and Mozilla may drive the future of format development and the concept of how formats are applied to an information stream.

Mr. Brantley closed with a summary of recommendations. The GDFR must stay current with trends in format developments; must emphasize its services rather than assume a role as a static standards organization; must not be file/document dependent; must not get hung up on xml schemas; and must sell the value of preserving the formats of old electronic material as well as preparing for the future as an organization focused on preservation should.

6.2.2 "Overview of the Digital Library Federation (DLF)"

Barrie Howard, Digital Library Federation

[No slides available]

Mr. Howard presented an overview of the DLF which was chartered in 1995 through the efforts of 16 university librarians who realized that the issue of digital libraries had become a recurring theme at conferences. Initially there was no funding for the organization which operated under a charter which was essentially a memorandum of understanding, operating through volunteered time and resources of the university participants. This initial DLF organization defined its goals and mission as well as its future organizational structure and funding model for going forward.

Through a planning grant from IBM, a home was found at the Council of Library Information Resources in 1996, and a single staff member was hired (the Director) in 1997. The DLF was incorporated as a 501(c)3 organization and developed a business plan with the assistance of consultants. The DLF evolved into an incubator organization, fostering projects in digital library technology and application. They see themselves as more of a venture capital firm as opposed to a maintenance organization. Their organizational formula specifies an ideal size of 35-40 members (the latest of which was UCLA, joining last week bringing their membership to 42 organizations).

In their governance model there is the Office of the Executive Director which supports a small staff responsible for setting goals and direction for the Federation, managing day to day activities, and providing oversight of initiative working groups. It is through these working groups that the work is executed in the DLF, drawing on a spirit of volunteerism to the extent possible. Every member has a seat on the Board of Trustees which is the over-arching governing body of the DLF, providing oversight on program activity and initiatives. An Executive Committee acts as a bridge between the Board of Trustees, the DLF Executive Director, and the Council of Library Information Resources and is responsible for governance between Board meetings.

The DLF continues to expand its reach beyond the US, already enjoying the participation of Oxford University, the British Library, and Bibliotheka Alexandrina. The work is done in a distributed fashion, making oceans a small impediment; however, they have found that having occasional face-to-face conferences is an important key to effective operation.

6.3 "Handle System Governance"

Laurence Lannom, Corporation for National Research Initiatives

Slides: <http://www.archives.gov/era/research/lannom.ppt>

Mr. Lannom presented the Handle System which implements a W3C protocol enabling a distributed computer system to store names or handles of digital resources and resolve those handles into the

information necessary to locate, access, and otherwise make use of the resources. He described the client-server architecture including its network topology.

The governance of the Handle System was presented covering topics such as policy making, funding, sustainability, and organizational characteristics. Challenges in carrying the Handle work forward were discussed. Lessons-learned that can be valuable to the GDFR undertaking were presented, stressing the need for simplicity of service profile and focus on organizational incentives.

6.4 "Governance as a Socio-Technical System" **Kate Ehrlich, IBM Research**

Slides: <http://www.archives.gov/era/research/ehrich.ppt>

Ms. Ehrlich defined governance as the activities surrounding decision-making, tracking, compliance, policy formation, and audit; and then introduced the importance of considering social networks in planning governance. Organizations typically form in a hierarchy designed to execute the business and governance of the organization. However, a social organization inevitably evolves that is the "real" decision making network. Understanding the underlying organization is essential to understanding who the key people are (those whose absence would cause the most disruption to effective operation.) The hierarchical structure's evolution needs to be advised by these considerations. It is important to understand that though governance is often designed along the lines of the hierarchy, understanding where governance is really happening requires considering the underlying socio-technical decision network.

Ms. Ehrlich enumerated specific questions that should be addressed in designing an organization and its governance structure:

- Are technical dependencies aligned with social support networks?
- Is decision-making aligned with the goals of the group?
- Are the decision-makers getting the right information?
- Are some people getting overloaded and becoming bottlenecks for decision-making?
- Is there sufficient awareness of others to coordinate work?
- Are there adequate safe-guards against indispensability of individuals?

6.5 "Infrastructure for Preservation" **David Giaretta, Director of CASPAR Project and** **Associate Director, UK Digital Curation Centre**

Slides: <http://www.archives.gov/era/research/giaretta.ppt>

Mr. Giaretta pointed out that there is a gap between information (semantic) and its representation (syntactic). He introduced the Open Archival Information System (OAIS) Reference Model standard (ISO 14721) wherein an information object is a collection of data objects and representation information objects. Both information objects and data objects are described through representation information objects, enabling both the description of format and semantics.

He then discussed the CASPAR project an EU FP6 integrated project). CASPAR is an open-source project based on open source standards intended to manage the gaps between information and data representation. He discussed the nature of "preservation" and the necessary characteristics of systems that purport to support preserved digital information and data. The challenges of registries and repository trust were discussed.

**6.6 "A Distributed Science-Data-Description Registry/Repository:
Implementation and Operational Experience"
Donald Sawyer, NASA/GSFC, and
John Garrett, SPSystems**

Slides: <http://www.archives.gov/era/research/sawyer.ppt>

Mr. Sawyer addressed the organizational characteristics of the Consultative Committee for Space Data Systems (CCSDS) which is a working body for ISO TC20/SC13 responsible for developing science-discipline-independent standards. It contains an Information Interchange Panel to promote the interchange, preservation, and use of space-related information.

For the ingestion and registration of data descriptions, Member Agency Control Authority Offices were established to provide the benefits of a single organization while remaining a loosely federated structure which suits international collaboration. Mr. Sawyer described the operations, policies, and governance structure implemented through the CCSDS Control Authority Office pattern as well as the challenges faced by the organization.

**6.7 "Practical Aspects of Managing a Standard, a Registry, ..."
Sally H. McCallum, US Library of Congress**

Slides: <http://www.archives.gov/era/research/mccallum.ppt>

Ms. McCallum introduced the "Preservation Metadata: Implementation Strategies" (PREMIS), a working group sponsored by the Online Computer Library Center (OCLC). The PREMIS working group aimed to build on the Open Archival Information System (OAIS) reference model (ISO14721) to develop a Data Dictionary of core metadata elements to be applied to archived objects, give guidance on the implementation of that metadata element set in preservation systems, and suggest best practice for populating those elements. Lessons-learned from the experience of maintaining PREMIS were discussed including governance and support facilities. She went on to discuss the governance, organization and funding models of the ISSN (and related) international centers which register international identifiers for such things as books, serials, notated music, etc.

**6.8 "National Digital Information Infrastructure and Preservation Program
(NDIIPP)"
Caroline Arms and Carl Fleischhauer,
US Library of Congress**

Slides: <http://www.archives.gov/era/research/arms.ppt>

The presenters provided an overview of the National Digital Information Infrastructure and Preservation Program (NDIIPP). The evolution and maturation of the International Internet Preservation Consortium was presented in terms of its goals, governance, and organization. Member organizations often maintain format registries specialized by the type of information customarily supported by the format, e.g., scholarly literature, engineering, and social science data. Beyond cataloging information about formats, the program assesses formats in terms of long-term sustainability. The NDIIPP "network of networks" organizational principles were discussed in the context of a "stewardship network." The presentation concluded with the suggestion that GDFR stakeholders should be identified that rely on the GDFR services in order to construct an appropriate oversight program. Lessons-learned from the NDIIPP that are germane to the GDFR were summarized.

6.9 "University Perspectives on Governance: MetaArchive Cooperation & DSpace Foundation"

Tyler O. Walters, Georgia Tech Library & Information Center

Slides: <http://www.archives.gov/era/research/walters.ppt>

Mr. Walters presented the lessons-learned in regards to organizational and governance aspects from two organizations: the MetaArchive Cooperative of the Educopia Institute which began as part of the National Digital Information Infrastructure and Preservation Program (NDIIPP) and the DSpace Foundation which has produced an open-source digital repository system. Modes of sustainability that must be addressed in such organizations were outlined, specifically Organizational, Economic, Technological, and Collections. He elaborated on the difficulties faced by each undertaking in their evolution and the remedies applied. The challenges facing the organizations going forward were summarized, providing a rich source of issues to be addressed in going forward with the GDFR.

6.10 "National Software Reference Library: Computer Forensics Tool Testing"

Barbara Guttman, US National Institute of Standards and Technology

Slides: <http://www.archives.gov/era/research/guttman.ppt>

Ms. Guttman presented an overview of efforts to provide international standard reference data to tool makers that investigators can use in their work and to establish a computer forensic tool testing methodology. She described the National Software Reference Library and the use of its Reference Data Set by law enforcement, investigators, and researchers. The strategies and technologies applied to the system were presented, among them the ability to uniquely "fingerprint" a digital artifact which reliably identifies a file. Testing was emphasized as an essential element in reference efforts stressing well developed specification, peer review of those specifications, and the development of test methodology for quantifying the effectiveness of tools.

6.11 "Engineering Formats and GDFR Governance"

Josh Lubell,

Manufacturing Engineering Laboratory,

US National Institute of Standards and Technology

Slides: <http://www.archives.gov/era/research/lubell.ppt>

Mr. Lubell presented pertinent results from a NIST workshop held in April 2007 on "Long Term Sustainment of Digital Information for Science and Engineering." Participants included implementers of the Open Archival Information System (OAIS) Reference Model standard (ISO 14721), government representatives (NARA, Library of Congress, Navy, US Government Printing Office), and universities. Among the results of the workshop was the identification of the need to collect end-user use-cases and the importance of considering format sustainability factors (disclosure, adoption, transparency, self-documentation, external dependencies, patents, and access restrictions). Though addressed in an engineering context, the sustainability factors are not domain specific. Mr. Lubell noted that format classification schemes will vary by specific domains (e.g., engineering), suggesting that governance and content development will require a coalition of domain-oriented entities. He recommended that the GDFR look at other registry projects that have succeeded, and in particular, at those that have failed to bring that experience to the GDFR organization at the outset.

6.12 "Information Discovery, Use and Sharing"

**Mary Brady, Information Technology Laboratory,
US National Institute of Standards and Technology**

Slides: <http://www.archives.gov/era/research/brady.ppt>

Ms. Brady presented the challenges of information discovery encountered in diverse domains: medical imaging, computational biology, shape searching (from protein structures to 3-D computer-aided design models), and human language processing (including text, video, and speech). Without the capability of finding specific information in a vast sea of data, innovation is impaired. She stressed the importance of testing methodologies and frameworks drawing examples from XML and ebXML.

6.13 "From Stovepipes to Wind-chimes: Advancing Agility Through Communities of Practice"

**Susan Turnbull,
US General Services Administration**

Slides: <http://www.archives.gov/era/research/turnbull.ppt>

Ms. Turnbull presented the effectiveness of light-weight, loosely-coupled, grass-roots, federated Communities of Practice as an effective paradigm for collaboration. Drawing from her experience as co-chair of the Emerging Technology Subcommittee of the US cross-agency Architecture Infrastructure Committee, she cited examples of the use of this paradigm such as the Ontolog Forum, the development of the US Federal Enterprise Architecture's Data Reference Model, and the Communities of Practice established by the Emerging Technology Subcommittee. Key tools enabling these Communities of Interest are wiki's and threaded discussion lists.

7 Discussion

Opening the discussion, the facilitator reviewed the principal objectives of the workshop, stressing that the mission was to identify issues (rather than solve issues) surrounding the governance of an international Global Data Format Registry (GDFR), discuss the issues to the point that they are adequately understood and documented, and then discuss ideas for going forward. The results of this workshop are intended to frame the further discussion of governance as we proceed forward with the GDFR.

The discussion was captured as it flowed by the workshop facilitator. No attempt was made to attribute points, comments, or remarks so as to keep the conversation running freely within the constraints of capturing it. The facilitator/editor re-arranged the workshop notes in a thematic order, editing the language for clarity.

It is important to note that there is a great deal of inter-relationship among these themes. Governance is a complex subject involving many topics that have a bearing on governance and which inter-relate. The discussion focused primarily on the following topics. In listing the topics a sampling of the inter-dependencies are mentioned.

- **Organizational Beginnings and Evolution** – There was discussion on bootstrapping the organization which addressed issues of what is needed to get started in terms of structure, funding, and the organization's evolution.
- **Organizational Characteristics** – Issues discussed included the Legal Form of the organization, the nature of its Board of Directors, internationalism, socio-technical aspects, and aspects of federation. In the governance plan there will need to be procedures and guidelines for the management of the structure of the organization, the assignment of stakeholders to positions in the organization (based, perhaps, on funding considerations), and the day-to-day operations.
- **Stakeholders** – Focusing on roles, who are those that will participate in the GDFR. This has an impact on the design of the Funding Plan as well as the Organizational Characteristic. A governance plan will be required to manage the ingress and egress of participants, whatever their role.
- **Funding** – The issue of funding is intertwined with the Organizational Structure and how the Stakeholders fit into it. Among the discussion points were the importance of the business case, initial funding and various funding models. The funding model selected for a given organization structure will dictate the governance necessary to manage stakeholder contributions and eligibility for membership, for example
- **Marketing Strategy** – In order to build and sustain the organization, awareness must be raised among stakeholders that could be participants. This awareness will be necessary in order to build the core organization and then to sustain it in both consuming and supporting members.
- **Content Management** – The issues involved in the intake, vetting and technical maturation of the GDFR content were discussed. These issues will also affect (or be affected by) how the organization is structured and what its funding model is. Included in the discussion was the "quantification of trust." The consideration and resolution of these issues will lead to the basis for the content governance model at the heart of the GDFR governance plan.

7.1 Organizational Beginnings and Evolution

At the outset it was noted that "the perfect is the enemy of the good" and that we should not try to envision the ultimate organization and governance structure. Regardless of where we start, the organization will evolve based on experience. That being said, it is still important to identify what will minimally work and to craft a viable starting place. If there are pre-planned "phases" that can be clearly anticipated, they should be laid out in the initial business plan; however, we need to start with what we know and work from there, allowing the organization to evolve.

- **Mission Statement** – A mission statement needs to be written that identifies the stakeholders in the enterprise, and what the focus and goals of the organization are both in short and longer term views. The vision/mission statement needs to articulate the benefits by stakeholder.
- **Initial Funding** – From the outset the organization needs at least an informal business plan emphasizing sustainability, and should have funds for at least 4-5 yrs. The funds should be sufficient to cover an adequate staff which should be in place at the beginning, rather than building slowly person by person starting with an executive director.
- **Core Energy** – In order to get the organization off the ground, there will need to be energy at the "center" from a few individuals. The participating organizations that can supply these individuals need to be identified.
- **Maturation** – Eventually, firm rules for resolving differences and procedures for proposing change will be needed, but that will be a sign of success

7.2 Organizational Characteristics

- **Legal Form** – Possible legal forms that the organization might take were suggested and discussed. (This is closely related to the discussion on "Funding".)
 - **Non-Profit** – There was a recommendation that the GDFR organization be organized along the lines of a US 501(c)3 non-profit corporation. Of the two forms (membership-based and charitable). It was pointed out that as a charitable organization, donations are allowed (a form used, for example by fedora).
 - If it is fee-for-membership based, are there classes of members? If so, what value accrues to each class of member? Why do they want to pay their fees?
 - If it is charitable, who are the donors? What is their motivation to endow the organization?
 - **For Profit** – It was also suggested that a separate for profit entity could be created with "shareholders". These must be identified along with their business case. (Even in a non-profit there are ownership issues.)
 - What is it that is being sold? Is there a market for it at a price necessary to sustain the organization?
 - **Internationalism** – Whatever the legal form, it must be compatible with the organization's international nature.
- **The Board of Directors** should represent diverse perspectives. The organization should be driven primarily by governmental and academic institutions rather than commercial ones.
- **Socio-Technical Considerations** – Organizations typically form in a hierarchy designed to execute the business and governance of the organization. However, a social organization inevitably evolves that is the "real" decision making network. Understanding the underlying organization is essential to understanding who the key people are (those whose absence would cause the most disruption to effective operation.) The hierarchical structure's evolution needs to be advised by these considerations. It is important to understand that though governance is often designed along the lines of the hierarchy, understanding where governance is really happening requires considering the underlying socio-technical decision network.

Specific questions that should be addressed:

- Align technical dependencies with social support networks
 - Is decision-making aligned with the goals of the group?
 - Are the decision-makers getting the right information?
 - Are some people getting overloaded and becoming bottlenecks for decision-making?
 - Is there sufficient awareness of others to coordinate work?
 - Are there adequate safe-guards against indispensability of individuals?
- **Federation:** Since we are discussing a global organization, a federated structure should be considered. In putting together a federated organization it is important to consider the politics of institutions and the motivations of individuals. In such an organization some groupings will be tightly coupled whereas others may operate relatively independently.

- **"Inter-GDFR"** – Is the GDFR a single organization or a coordinating body over multiple registries? There are multiple format registries in existence now in various forms. This would be similar to the emergence of the Internet from the multiple networks that existed in the '70's and '80's (Arpanet, Alohanet, etc).
 - Registries will have different policies in place. Bringing them together would require the management of exchange protocols, format representation (and its format), and over-arching concepts of governance so that there is semantic unification among the participating registries (i.e., if "widget" means something in registry "A", it will mean the same thing in registry "B".), etc.
 - These extant organizations need to be identified to consider this option.
 - Registries may wish to cache or replicate representation information from other registries, possibly with different Identifiers. The issue of multiple identifiers will need to be addressed.
- **Domains of Expertise** – There will be diverse format contributors. The "owners" of the digital objects probably know them best. There are representations that require specific skills to understand, for example Computer Aided Design representations will generally require engineering experience.
- **Delegation of Control** – A federated system may not only involve an organization of peers, but there may be tiered subdomains. Each level of the structure must have over-arching principals, guidelines and governance. There needs to be rules for what is in the purview of the subdomain and what is reserved for its superordinate. Some registry entries may require central approval before adding while others may be approved or deprecated after initial addition (the superordinate serving as the final gate of approval for example).
- **Simplicity** – To be successful the organization should maintain simplicity in its structure and the services that it offers, making sure that the service suite is broadly used. Don't try to "boil the ocean."

7.3 Stakeholders

Crucial to the design of the organization, its governance structure, and the funding model is the identification of the GDFR stakeholders. What are the economics for each group? What is their "care abouts" that keep them engaged? Is the stakeholder generating more revenue, lowering cost, increasing quality, or executing some assigned role of community service? The economics and motivations of each stakeholder will have an impact on what is or is not a viable funding model. The vision/mission statement needs to articulate the benefits by stakeholder. A variety of stakeholder categories were identified:

- **Content Custodians** –The archival institutions (traditional or otherwise) that are committed to preserve the documents and material that are in digital formats. These institutions understand that the preservation of format documentation is key to long-term digital preservation and include organizations such as national archives, libraries, etc.
- **Content Developers** – Those who understand the formats and can author their formal descriptions faithfully and in a form that is useful in interpreting a digital artifact that is compliant with it.
- **Service Providers** – The organizations which provide the facilities to operate a node in the GDFR network. It is presumed that there will be multiple nodes that will variously ingest format descriptions and/or mirror other sites providing local availability and load balancing. Lessons-learned can be brought to the table from National Digital Information Infrastructure and Preservation Program with its established experience in soliciting formats from producers (video, geo-spatial, ...)
- **Registry Editor** – Intimately tied with the governance model, the registry editor decides what goes in and what does not. This role could be a scarce resource & therefore could provide difficulties from procedural (bottleneck) and funding standpoints. Deciding who gets to annotate, and who edits the registry is a key issue that needs to be addressed.

- **Format Developers** – The organizations that design formats for their own (often proprietary) purposes. These organizations may have interest in having the format registered imbuing their products with "preservability"
- **Registry Consumers** – What is it that the registry consumer is accomplishing through the use of the registry? What are the economics, i.e., the value accrued to the consumer through the use of the registry? Are there different categories of membership? If so, does each of these categories have different motivations and economic drivers?

The first four categories of stakeholders suggest process, implying that there may be governance applying to each. Whenever we consider governance, we need to consider the degree of formality (heavy weight) or informality (light weight) required. There is a "less is more" rule of thumb that applies here to make sure that the organization is sufficiently governed yet remains agile.

7.4 Funding

- **Business Case** – In forming the organization/funding plan, it is important to rely primarily on organizational incentives (a business case) rather than the enthusiasm of individuals or groups which is more likely to fade over time. In evaluating the funding models it is important to assess who needs this most, who has the most to gain. (See the Stakeholder discussion).
- **Initial Funding** – From the outset the organization needs at least an informal business plan emphasizing sustainability, and should have funds for at least 4-5 yrs. The funds should be sufficient to cover an adequate staff which should be in place at the beginning, rather than building slowly person by person starting with an executive director.
- **Funding Models** – A number of basic funding models were discussed many of which are not necessarily mutually exclusive:
 - **Endowment** – For the long term it is advantageous to start early to solicit and build an endowment. In order to build an endowment, solicitation for contributions should focus on corporations, particularly large ones and those that actually design formats that need to be sustained for preservation. These early exemplar corporations will serve as attractors for additional contributions from other corporations.
 - **Subscription** – What sort of subscription is being offered & for what?
 - **Membership** – Are there different categories of membership? If so, does each of these categories have different motivations and economic drivers? What accrues to the members?
 - **Pay for Use** – It was noted that this can be difficult to engineer.
 - **Charitable Contributions** – From whom? What is the motivation to give?
 - **Registration Fees** – The organizations that design formats for their own (often proprietary) purposes. These organizations may have interest in paying to have the format registered, imbuing their products with "preservability"
 - **Conformance Testing** – Fees might be charged for format conformance testing roughly after the model of the US Underwriters Laboratory.
 - **In-kind Contributions** – for example, donated space, equipment, & personnel to operate a node.
- **Evolution of Funding, going forward globally**
 - As the GDFR develops and matures, industry will become increasingly interested (both producers and consumers). As this happens we can evolve the financial model (and perhaps organization).

7.5 Marketing Strategy

To progress to a comprehensive global format registry organization, we are going to need marketing to increase awareness and solicit participation:

- The US National Archives and Library of Congress have press councils. Can these be leveraged? Many of the participants' organizations have similar resources. Could these resources be brought forward in concert?

- Need elevator speech, brochures, advertisement, articulation of the whys and the benefits. Must be in the major languages that will span the globe most efficiently.
- Before marketing do we need to be further than we are?
 - There should be at least a nascent system capable of ingesting formats.
 - There are two levels to the marketing – marketing to the contributors as to why they should contribute intellectual capital to get the effort going and then another level to the consumer of the services which will require further maturity.

7.6 Content Management

- **Provenance** – The provenance of a registered format must be recorded to make the source clear. In order to avoid hackers and other spurious entries, and to maintain provenance of all entries, the system will require the registration of users.
- **Life Cycle of an Entry** – It is presumed that new format registrations will come into the system from diverse sources and will be assessed through a process for validity and accuracy. The governance process will need to be defined so as to document the "stages" of validation and conformance assessments.
 - **Initial Content Entry** – The information in the GDFR will always come in an initial form that will begin its life cycle and its path of maturation through vetting to a trusted resource item.
 - Initial validation should be simple and light-weight in order to attract content. Don't make the first step too onerous. It is anticipated that a majority of the information will come from 3rd / 4th parties with no financial incentive beyond the contribution.
 - Introduce light-weight community mechanisms (communities of practice, communities of communities) to garner interest as well as initial submission.
 - Introduce maturity attributes based on the degree of vetting. The technical governance model will need to explicitly specify what has been done to vet an item to earn the "higher grade".
 - **Versioning** – Over the life of a specific format registration there will be issues raised pointing out technical problems with the format description. There needs to be a governance plan for updating a format description. Changing a format description should require the documentation of what was changed and why, implying an issue log and issue resolution process.
 - Each version should have its own provenance recorded; including what issues prompted the change.
 - Each version should be kept available, i.e., they are "write-only" and never changed except through the versioning process providing a complete history of provenance and justification.
 - Each registry entry will require a unique identifier plus a version designator.
 - The system should provide reference to "latest version" as well as access to superseded versions.
 - "Standard" formats undergo their own evolution. In addition to documenting versions of refinement of a format, we must track that a given format undergoes its own evolution. It may be important to document the versioning of a specific "named" format.
- **Parameters of Trust** – The GDFR needs to be a trusted repository that national archive organizations and libraries can trust. Having said that, trust cannot be declared, it must be developed.
 - **Quantification of Trust** – The GDFR will need to support metrics for determining what material, if any, can be trusted for what purpose. Information will likely come in from diverse sources and undergo a vetting under a plan of technical governance.
 - Are the contributors to be "rated?" If so here needs to be a procedure in the governance plan by which contributors are certified.
 - The vetting of contributions under the governance plan may have several levels. We need to indicate how carefully the data have been scrutinized, against what

criteria. The "rating" needs to be trusted, even if it indicates that the data are not yet to be trusted.

- **Conformance Testing** – Conformance testing was thought to be key to the ability to make trust assertions concerning the data.
 - There are two levels to conformance.
 - One involves the ability to assert that the registered format representation faithfully documents the format under consideration.
 - The other involves the ability to confirm the assertion that a particular digital artifact is compliant, or not, with the registered format.
 - Lessons-learned from the ebXML experience indicates that conformance testing is crucial to success.
 - We need a Test Framework and Suites used to verify conformance and interoperability. A globally recognized open testing architecture and test suite would promote global GDFR interoperability and utility through a standard set of testing tools.
- **Semantics** – Often the same information can be represented through different formats. A trivial example is that a Microsoft Word document can be transformed to an Adobe Portable Document Format (pdf). The two documents are the same from a content standpoint, but appear as separate documents from a format perspective. Are facilities to be included to indicate the "sameness" of the same information in multiple formats? How far should we go in modeling the semantic content of a given digital artifact? Are these considerations within the purview of a GDFR?
- **Redundancy** – When and how do we recognize the "sameness" of digital artifacts? Emerging techniques such as the "fingerprinting" of digital artifacts can prove useful in this regard. However, the consensus was that though this is an important consideration for a digital preservation system, it's likely out-of-scope for the GDFR.

7.7 Issues

- **Proprietary Format Creators** – These organizations may not be forthcoming in providing intellectual property to a format registry. This sector is populated by organizations which have designed formats for their own tool suites, such as Microsoft, Adobe, Computer-Aided Design companies etc. As a global organization the GDFR will have more clout in soliciting information on formats, but it is not guaranteed. Is there an approach which can protect such IP, yet be useful as a registry?
- **Registry Editor** – Intimately tied with the governance model, the registry editor decides what goes in and what does not. This role could be a scarce resource & therefore could provide difficulties from procedural (bottleneck) and funding standpoints. A key issue that needs to be addressed is: who gets to annotate, who edits?
- **Early Orphan** – There is some danger that the GDFR could become an orphan service should it overly focus on the historical perspective of file and document orientation. The world is turning to the web for access and the concept of file and document is breaking down in this more fluid environment.
- **Lessons-learned – the legacy of success and failure.** We need to examine registry organizations that have succeeded and those which have failed in order to maximize our success.
- **Stakeholders** – We don't know enough about who the end users are and how they will use the data.
- **State of the Art** – We lack the rich documentation required for the archival packaging of data for future use. We need guidelines/requirements for creators to follow while recognizing that they won't be universally followed. We need to think about potential enforcement through contractual mechanisms and peer review, and how to support creators through semi-automated tools.
 - It's hard to find archiving software tools and to figure out how they all fit together
- **Domain Specific Guidance** – We must avoid suffering from a lack of domain-specific guidance (metrics, potential use cases, service level agreements), for example, the special requirements and expertise of "vertical" domains such as manufacturing, engineering, and finance. We need to

understand the needs of involved communities, industry specific requirements analysis, etc. We need to be a community of communities (see the discussion of Federation).

- Different levels of detail will be required in metadata depending on format and context.
- **Scale** – How many formats are there and what is the workload for the governance. Do we know what the size of this organization needs to be?
- **Deadlock** – However the organization is structured, the governance must be designed so as to prevent the deadlock of unresolved contention. There must be a way to make a final decision in the face of controversy.
- **Exit Plan** – Participant interest sometimes changes. If a member loses dedication, funding, or for whatever reason becomes dysfunctional in the organization (or simply wants to withdraw), what is the exit plan? What arrangements are built in to transfer any GDFR assets (real or intellectual) that may be under custodianship by the exiting organization?
- **Rationale for Internationalism** – The question will be asked "Why should the GDFR be global? Why doesn't each country simply do its own thing?" As reflected in their participation in the workshop, each participant felt that the GDFR should be a global undertaking. However, the question will come up. We need to develop a statement of the long term motivation and assure that these resonate with our international audience; otherwise, we will not be able to secure global resources.

8 Going Forward

The last half-hour of discussion was focused on the issues of going forward with an initial organization. The discussion covered what this organization might look like and issues surrounding its formation. The initial organization can (and probably should) be very informal. After the first phase, however, there needs to be more formality.

In a straw pole eleven participants said they would be interested in helping to carry this effort forward through participation in an initial organization.

The following were suggestions and issues raised in the discussion in regard to an initial GDFR organization.

- **Advisory Council** – We should establish an initial advisory council with standing or ad-hoc committees to coordinate the efforts going forward.
 - **Core Drivers** – Going forward is going to take individual and institutional commitment. There was some concern as to where the core of leadership would come from. The importance of a core of individuals/organizations to energize and coordinate the effort came up frequently in the workshop presentations and discussion. There was also concern about the disruption that could be caused by changes of leadership over time in this nascent, informal organization.
 - **Working Groups** – We need to organize working groups around the topics we discussed in this meeting and have the working groups pursue their topics further and make additional preparations for our next meeting. Additional working groups can be established for domain specific issues of the GDFR (e.g., law enforcement, engineering, health, etc.)
- **Mission Statement** – We need to construct a mission/vision statement for the GDFR as an ongoing organization.
- **Business Plan** – Even for this nascent period of the effort, we will need budget estimates for what it will take to operate our startup efforts to flesh out our plans based on the discussion items from this workshop and put them into effect.
- **Community of Practice** – It was suggested that we use a Community of Practice paradigm collaborating through Wiki pages, a threaded discussion list, and anchored in a web site (which may be the wiki site). This could be a light and productive way to start. The US General Services Administration has enjoyed success in this approach for their cross-agency investigations of emerging technology.
 - The US National Geospatial Digital Archive. (NGDA) has a wiki – has it worked? It has been beta-tested but not opened up yet. It's just emerging.
 - GSA is supporting geo-spatial Community of Practice (It was not clear how this cross-fits with the NGDA work). All of their work was done in the openness of a wiki and were not distracted by the noise of the world. They felt it was very effective.
- **Communication** – We need to establish a communication alliance which is inexpensive. All participants should be subscribed to the alliance. The alliance should be open to additional members in that there are many stakeholders out there who have not yet been "captured".
 - We could possibly use a wiki facility as well as a more traditional web site. This will require a DNS name. Harvard has already secured such a name. We need to identify who will host the site: the following hosts were suggested.
 - US National Archives and Records Administration – Though willing to offer whatever support they can to the GDFR effort, NARA in particular (and the participants in general) expressed concern about creating the perception that this is a NARA or US effort. It is important that this undertaking be truly international in perception and fact.
 - US General Services Administration – The GSA operates a number of wiki sites for collaboration of government agencies, industry, and academia on emerging

technologies that are germane across the government. Again there was concern about the perception of this being a US-based effort.

- Harvard/OCLC – The GDFR project itself has web facilities in place. However, the project is scheduled to complete in 2008. Again there was some concern that the effort might be perceived as US based. On the other hand, since the facilities are in place, this could serve as a convenient temporary home to get us started.
- Object Management Group – The OMG is an international standards organization which would likely be willing to provide a wiki/website home for the undertaking. (A member of the OMG's Board of Directors was present.) This option would provide an appropriate international standing for the effort.
- **Timing** – From a timing perspective it was suggested that we proceed with the proof of concept and GDFR demonstration under construction by Harvard and the Online Computer Library Center (OCLC). After that we should convene another meeting and proceed from there.

Strawman Plan for Going Forward

Based on the workshop discussion documented above, the following actions were suggested for moving forward

- Establish a GDFR Advisory Council including a light weight governance model for initial decision making and tasking.
 - Write the GDFR Mission Statement
 - Write the GDFR Business Plan
 - Establish the Advisory Council Communication Plan
 - Deploy the infrastructure necessary for the communication plan
- Schedule the next meeting of the Advisory Council (to include a demonstration of the GDFR System work done by Harvard & OCLC. The goal of this meeting is to address the issues raised in the Governance workshop and to work up a strawman organization and governance model for the next phase of the GDFR organization.

The immediate plan for going forward is the vetting and approval of these proceedings by the workshop participants, and thereafter the initial formation of the GDFR Advisory Council.