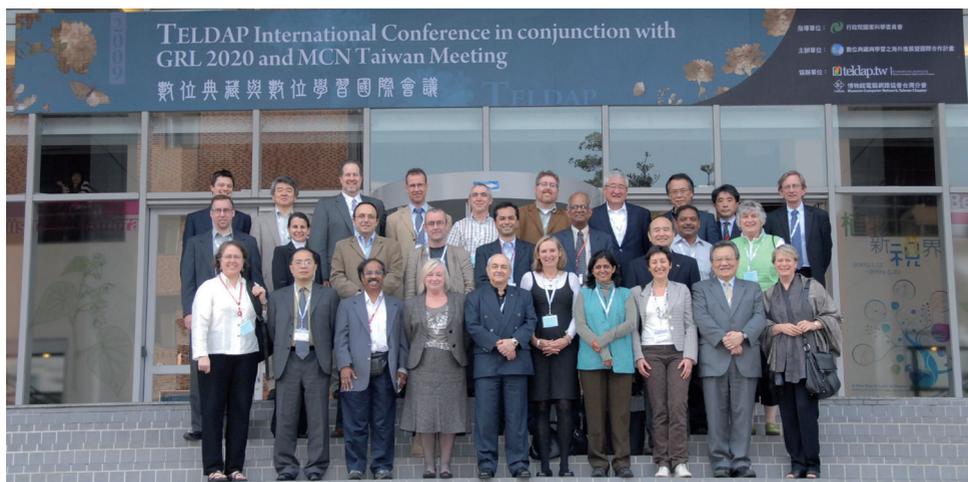


GRL2020 Europe

Call for Action

Outcomes and recommendations of GRL2020 Asia



Participants at the 3rd Global Research Library 2020 Workshop, February 2009, Taipei, Taiwan

23rd -24th February 2009 Taipei, Taiwan

www.grl2020.net





Summary

GRL2020 is a vibrant community that is focusing on top-level challenges facing the global research library of the future with the ultimate goal of creating a knowledge infrastructure that most effectively serves the needs and leverages the competences of domain specialists, librarians and computer scientists. While the challenges are manifold, GRL2020 provides a unique opportunity to harness the expertise that exists on a global level, to define an agenda that prioritises the challenges ahead, to champion the importance of skills development and education, and to enlarge the community through a Call for Action defined during GRL2020 Asia, which was co-hosted by Microsoft Research and Academia Sinica.

The hosts, programme committee and organisers would like to thank all the participants for sharing their expertise and experiences and for enabling effective knowledge exchange. Full details of the participants are provided at the end of this Report. Position Papers and Workshop outcomes are available at www.grl2020.net.

Acknowledgements

GRL2020 wishes to thank Microsoft Research for their support to this publication, to the GRL2020 Programme Committee and to the authors for their valuable insight and contributing work in achieving this informative report.

Disclaimer

The contents of the post-event report have been created by representatives from members of the GRL2020 Programme Committee and with the editorial support from TrustIT Services Ltd. Although content of this report is public, we request that any copying or paraphrasing from the writings or works indicated in the present document should clearly acknowledge the report reference source, its authors, organisation names and dates as indicated here. All rights are retained by the original authors and submitters.

Table of Contents

GRL2020 Mission	7
GRL2020 - An Evolutionary Path	9
Shaping Forces: Data Infrastructure	11
Shaping Forces: Skills and Education - The Library as a Collaborative Environment	13
Shaping Forces: Content & Communities	15
Call for Action	17
Defining a Value Proposition	17
Creating a Think Tank	17
Establishing a Common Glossary of Terms	18
Publishing a White Paper	18
Setting up GRL2020 Work Areas	18
Producing a Roadmap	18
Developing a Case Study	19
Granting an Award	19
Shaping Research Programmes	19
Spotlighting Educational Requirements	19
Ensuring cross-domain, cross-profession and cross-border outreach	20
Web Design & Dissemination	20
Enlarging the GRL2020 Community	20
Expertise & Session Overview at GRL2020 Asia	21
Session Focus	23

GRL2020 Mission

GRL2020 champions the creation of a Global Knowledge Infrastructure that opens up knowledge to research communities worldwide and society at large by responding to the new and specific needs of eResearch. This global infrastructure will advance beyond the current state of the art by creating an environment that provides a broad spectrum of relevant resources, such as publications and datasets, with the research community defining how the material is used. The global knowledge infrastructure of the future will be a more powerful and flexible ecosystem capable of interoperability and collaboration across boundaries and disciplines, underpinned by a commitment towards trusted, long-term services and open access, and enabling the deployment of content resources and technical services on multiple levels.

The creation of a knowledge infrastructure of the future will require considerable technological and organisational effort and will mean a significant rethinking of the role and processes currently performed by research libraries, particularly in terms of the stewardship of digital data. GRL2020 brings together researchers and stakeholders from diverse domains and countries to deliberate on how these challenges can best be addressed. Particular emphasis is placed on both the technical and non-technical hurdles that need clearing as instrumental in defining the steps to advance towards such an infrastructure. Particular emphasis is placed on both the technical and non-technical hurdles that need to be cleared, and on defining the steps that will be instrumental in advancing towards such an infrastructure.

GRL2020 - An Evolutionary Path

GRL2020 Asia was the third in a successful series of Workshops. It built on the headway made at GRL2020 USA in 2007 and GRL2020 Europe in 2008, aimed at defining a research environment that enables collaborative work across disciplines, professions and geographical boundaries. The TELDAP International Conference provided both a backdrop and forum for the Taipei Workshop, which welcomed 35 of some of the world's most highly regarded experts in the area of research libraries and scientific data archives. These experts showcased best practices, case studies and pioneering work that is furthering the innovative approaches fostered by GRL2020. All participants made valuable contributions to the interactive discussions, developing the GRL2020 research agenda and charting a course that clearly defines next steps.

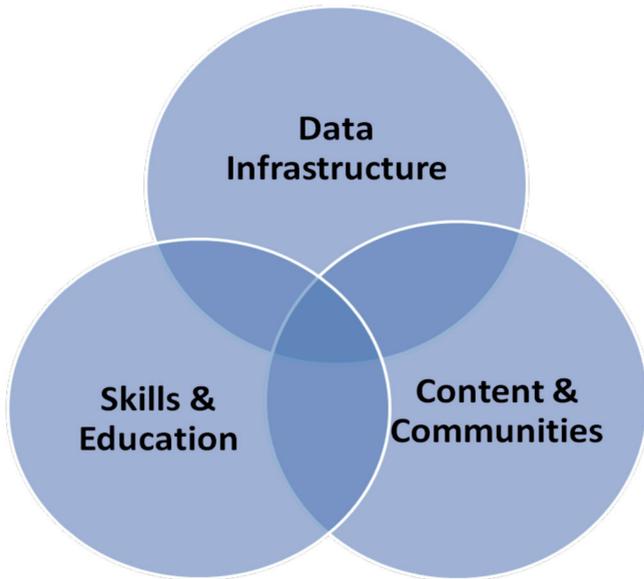
The program allowed participants to gain a unique perspective on ongoing work on digital libraries in the global scene. There were plenty of opportunities to obtain and to provide feedback and also to reflect on the future of digital libraries. Alfredo Sanchez, Universidad de las Americas, Mexico

A distinguished group of experts formed the Programme Committee spearheading thought leadership and driving the workshop that placed much emphasis on a Call for Action to address the challenges and issues surrounding the GRL2020 vision.

GRL2020 Asia Programme Committee

Donatella Castelli	CNR-ISTI	IT
Lee Dirks	Microsoft Research	US
Fabrizio Gagliardi	Microsoft Research	US
Cathrine Harboe-Ree	Monash University	AU
Jessie Hey	University of Southampton	UK
Yannis Ioannidis	University of Athens	GR
Alex Wade	Microsoft Research	US

This Report provides a summary of the main points discussed, with an emphasis on the challenges and potential work areas in which the GRL2020 group can influence the community. These include data exploration and infrastructures, skills development and training, and content and communities. Case studies and best practices presented as an example of how the challenges ahead can be seen as opportunities for innovation beneficial to a broad global community.



GRL2020: Potential Work Areas (identified by Lee Dirks)

Shaping Forces: Data Infrastructure

The term data or information avalanche denotes the huge amounts of data generated by the new wave of data-driven scientific experiments, which are changing the way people conduct research and impacting on the role of librarians in curating and managing scientific information. While it is noted that humanities and social sciences research is also changing dramatically as researchers explore new opportunities made available through technology, this section explores discussion that took place on the technical and non-technical challenges that are inextricably bound up with the data explosion, which is defined as the 4th scientific paradigm. This places new emphasis on data management and requires new technology and data infrastructure to deal with all the stages that data go through to obtain results and identify new discoveries.

It is expected that data-generating sources will continue to improve, that phenomena studied will increase in complexity and that scientific collaborations will become larger, all of which will lead to an exponential growth of scientific data production. Projects like the Large Hadron Collider (LHC) at CERN will be generating 10Petabytes/year of raw data, whereas the International Polar Year brings together over 50000 researchers engaged in hundreds of projects related to the environment in the poles, each project being multidisciplinary by itself and dealing with extremely diverse and heterogeneous data. While capturing, loading, and storing such data is challenging, cleaning and analysing it is significantly more difficult. Current technology cannot deal with the large datasets that will be generated by data-centric science.

There is a real opportunity for libraries to play a pivotal role in working collaboratively with others - research councils, researchers in the humanities and science, university research libraries, computing departments, learning technologists and administrative departments. Jessie Hey, University of Southampton, UK

Specific technical challenges around scientific data comprise size, heterogeneity, semantics, evolution, curation, access, and diversity. Further complexity is added by the difficulty researchers and scientists encounter in understanding the data that needs analysing. A viable approach to this challenge could be the development of simple toolkits that enable the publication of compound objects across disciplines and the extraction of objects from repositories around the world. However, this approach is connected with yet another challenge, that is, the lack of consensus on modelling complex objects, constituents of complex objects and how the constituents link to each other. From a standards perspective, Object Reuse and Exchange (ORE) provides us a descriptive framework but currently falls short of indicating what constitutes an aggregation. This also raises the question as to whether the solution could be to provide tools that deal with the large variety of objects as well as to foster adoption of such tools across domain, with the ultimate goal of making objects and information accessible.

Many of the open issues revolve around how to best steer the evolutionary paths of technology in areas such as data and applications, the need to evaluate how best to capture knowledge generated from the vast amounts of data that are produced, assess use cases on how people are using and scaling data, and combine the different scientific paradigms, such as empirical, theoretical approaches and computational models, in addition to data exploration.

It is also important to explore the impact of new scientific experiments and collaborative approaches from an educational perspective as the boundaries between disciplines are becoming blurred and young scholars are facing new challenges in dealing with the variety and interconnectivity of disciplines. A strategy may be needed to address not only the educational implications springing from new methodologies for modelling and saving data, but also the need to bring into sharp relief that fact that data per se do not constitute a learning tool. In a nutshell, it is the publication and not the data that is significant. Furthermore, it is the meaning and context of the document that adds value and provides an advantage over Google searches that simply helps find a document – hence, the key role of library curators and service providers.

There is a general consensus that the sea-change around data curation and preservation requires particular urgency not only in view of the data deluge but also with regard to data created in the past, which has not always been appropriately tagged and preserved. It is therefore imperative to start addressing these issues now, in order to respond to the new challenges that the brave new world of modern science and research is imposing.

While addressing all of the challenges and issues surrounding data is beyond the scope of GRL2020, it is important for the group to identify areas for further investigation and help steer funding and policy initiatives in the right direction to ensure technology evolves into a knowledge infrastructure. Models highlighted at GRL2020 for further deliberation include the European Commission funded project, DRIVER-II¹, which embodies many of the components that GRL2020 aspires to shape on a broader scale, as well as the Biodiversity Heritage Library, which has created tools for researchers to curate the data collected.

¹ <http://www.driver.community.eu>

Shaping Forces: Skills and Education - The Library as a Collaborative Environment

In the new environment of data-centric science and interdisciplinary research, librarians should ensure the information generated by research is managed for use and re-use. As a consequence, librarians now play a dual role, as creators and curators of content and data and as interfaces with data users. There is an increasing need for librarians and domain scientists to work together closely and to place special emphasis on the priorities and needs of different disciplines, key data, as well as relevant tools and documents. Additionally, librarians need an understanding of the application area. What is more, the emergence of cross-discipline research means that researchers now require access to a broad range of resources. The library needs to adjust to these new demands and create systems capable of encompassing the needs of diverse user communities. Interaction between library professionals and the communities they serve is imperative.

GRL2020 Workshops have convincingly demonstrated why the library must assume a more strategic role in the context of the integration of research repositories into research assessment, leading to a brokering role by supporting a variety of outputs such as theses, data, learning and teaching. Developing the right skill set and keeping it up to date is a crucial element in assuming a more proactive role. Southampton University has been a key driver in this respect.

Case studies at the Workshops have highlighted the shift towards curation as an interactive process encompassing management and leveraging new tools available, enabling the transformation of the library into a collaborative environment. Part of this process involves libraries building applications and providing add-on services that are more oriented towards research as opposed to more conventional library services. This shift could extend across institutional and discipline project co-ordination in the humanities, where large collaborative research is far less common. A good case in point is the innovative approach adopted by the State and University Library of Göttingen in Germany through government funding. Göttingen has spearheaded initiatives that foster institutional and discipline project co-ordination in the humanities. Such initiatives also illustrate the key role played by funding programmes in supporting the co-evolution of libraries and research.

While this sea-change inevitably brings with it many challenges, it is also an important opportunity for libraries to play a partnership role and a key part in transforming the library into a collaborative environment underpinned by value-add services. GRL2020

“Raw technological skills, including the requirement of programming and system integration (to address the concepts of networking, search and retrieval, archiving, and preservation, etc. in the new environment), and other necessary core skills including program management, user interface design” Lee Dirks, Microsoft Research, US

has a pivotal role to play in fostering this transformation and championing radical changes in the way the next generation of librarians are trained and prepared for the work required in the field moving forward. The new challenges call for GRL2020 support and guidance on how to best expand the boundaries of the skill-set of the librarian ensuring that a more tech-savvy workforce emerges. One way to ensure that library staffs develop the right skill set which is then updated. One viable approach could be to enable close collaboration between different inter-library teams, such as IT services and librarians, as well as training subject specialized librarians to conduct data management interviews with faculty heads and to initiate brokering services to manage data, and national and international repository projects more effectively. Such a programme has been developed by Monash University.

Shaping Forces: Content & Communities

In order to address user community requirements and to ensure that the research community is also educated on and sensitive to metadata issues, it is vital that the diversity of the group expands to include more domain scientists who represent the very communities that the global research library will serve. The increased participation of researchers from a variety of disciplines, science, humanities and social sciences, would be hugely beneficial to the GRL2020 group. To ensure productive collaboration and value added results for both the GRL2020 group and the domain scientists, it is imperative that these experts already have knowledge of information management issues such as data management, open access and collaboration.

Demonstrating the benefits that target communities can gain from fully using global research libraries is a crucial step towards this goal. GRL2020 Asia highlighted benefits such as how the sharing of detailed research data can be associated with a higher number of citation rates and in turn an increase in the reputation of scientists – an important driver in fields in all academic disciplines. Furthermore, humanities scholars appreciate that they need to co-operate with librarians in order to have their work available on line.

Another important facet is the variety of practices across disciplines, which calls for careful consideration. Cross-discipline partnership through the GRL2020 forum offers an opportunity for collaborative developments and co-evolution for all domains. It is this type of incentive that should be highlighted to encourage domain scientists to work with the core GRL2020 group. However, such an incentive may not be true of all sectors therefore, the GRL2020 group should seek to identify how best to link the co-operation of domain scientists with the role of reputation in diverse scientific areas.

When thinking of the future, it is also imperative to bear in mind how new colleagues and communities can join the GRL2020 ever growing community. One example is the Library of the World Vegetable Centre, which co-ordinates the work of soil scientists and nutritionists from diverse countries. One of the roles of GRL2020 could be to support newcomers by providing guidance on how an effective partnership between library staff and researchers can be established, how the library can catalogue multilingual, grey literature and pass it on to the user community, in this instance small-scale farmers, as well as how the library can be a catalyst for change.

Additionally, GRL2020 should identify national research libraries in particular disciplines whose collections or repositories could be federated to yield significant benefit through the generation of a global research library. In identifying the training, technologies and content required for this, GRL2020 would provide a focus for attracting funding agencies to offer financial support for training, and for content management and development.

Library systems providers have been conspicuous by their absence in GRL2020. This professional group is developing tools primarily based on the assumptions and pressures of peers, whereas it could leverage the expertise, future vision and best practices that exist within

GRL2020. The current GRL2020 community should define a simple strategy to identify potential contacts in its respective networks, with the aim of ensuring this group actively engages with GRL2020 not only at future events but also on a more regular basis.

All of these examples illustrate how GRL2020 has an important opportunity to engage with a much broader community, while maintaining a compact and focused core group of leaders

Call for Action

GRL2020 Workshops have opened up a window of opportunity to define a series of concrete actions for the group to undertake in order to build on current achievements, foster broader community consensus and effectively convey to funding bodies the need for further research and training through collaborative developments. The section below defines ten actions for the GRL2020 group. These actions range from defining a value proposition for GRL2020, to setting up work areas to ensuring greater visibility for GRL2020.

Defining a Value Proposition

GRL2020 has harnessed a truly global level of expertise to define a far-reaching vision for research libraries and a knowledge infrastructure focussed on long-term sustainability. This base sets the group apart from typical collaborative developments on a national and regional level.

Now there is a need to define a value proposition for the group which can be conveyed to funding and policy bodies and the broader community. This should highlight the expertise that exists on a global level with representatives from funding agencies and policy groups, as well as library directors, librarians, information scientists, domain scientists and members of enterprise, many of whom have spearheaded innovative initiatives on a national, regional and cross-regional level.

Creating a Think Tank

GRL2020 should be seen as a global think tank for research libraries, proactively addressing top-level challenges generated by new approaches to scientific experiments, research in the humanities, and the changing stewardship role of librarians. The forward-thinking approach championed by the GRL2020 group provides a unique opportunity to expand the boundaries and drive forward the vision for 2020. Now it is important for the group to identify specific areas of expertise and clearly define how tangible outcomes can be achieved by individual members in terms of interaction with target groups and contributions to the initiatives outlined in the Call for Action, while ensuring the core group remains compact and focused.

"GRL2020 terminology needs to honed as the group are linguistically forming arguments and still in the process of defining terms"

Peter Young, Library of Congress, US

A practical starting point for identifying problems to be addressed by the group could be the criteria adopted by the National Science Foundation Blue Ribbon Task Force Report, which would help identify of the scale of the issues to be addressed.

Establishing a Common Glossary of Terms

In order to foster an effective, two-way dialogue both within the group and to the wider community, special attention should be paid to the specific meanings, connotations and implications of not only commonly used terms, such as data, objects and documents, but also their usage within the specific context of GRL2020. .

Publishing a White Paper

A White Paper targeting political decision makers on an institutional, national and regional level is highly recommended. The White Paper would provide an opportunity to bring the GRL2020 vision into sharp relief and to the attention of policy makers and peers alike. One of the goals of the White Paper would be to ensure legitimacy when addressing target audiences and a persuasive tool to build consensus around the vision. Additionally, the Paper would enable group members to offer their recommendations on how best to define the vision and the steps needed to make it possible. A suitably qualified author and White paper committee would be set up to produce the paper. An important recommendation is that the White Paper be translated as necessary to make it more accessible to policy groups whose first language is not English.

Setting up GRL2020 Work Areas

Three dedicated work areas and related working groups (WG) are needed to address specific top-level challenges facing the development of future global research libraries. The groups should be multi-skilled, combining domain scientists, librarians, infrastructure experts and application experts. The groups should provide stimulating thought leadership on how to achieve the vision by highlighting success stories, case studies and use cases in the following fields. Findings from the WGs would contribute to the GRL2020 roadmap outlined in the next Call for Action. The following WG topics are proposed:

- Infrastructure – tackling technological issues
- Skills & Training – addressing the education of librarians
- Content & Communities– collaborating with user communities

Producing a Roadmap

The proposed GRL2020 Roadmap would serve as a principal guide for the development of global research libraries. The roadmap, which is aimed at setting up a framework for delivering a global research library that truly meets the needs of the research community, will draw on reports from GRL2020 work areas, inputs from the group's expertise, the outcomes of GRL2020 workshops and additional pertinent meetings, as well as collaboration with librar-

ians, domain specific scientists and representatives from funding agencies, in order to define how technical and non-technical challenges can best be addressed. The Roadmap would be made publicly available and distributed to major stakeholders, researchers, practitioners, and funding agencies across the globe, with the aim of fostering buy-in and potential collaborative developments.

Developing a Case Study

A test bench is now needed to validate the steps defined for the overall vision of GRL2020. An inter- and multi-disciplinary area that is compelling and that constitutes a grand global challenge should be selected. It is paramount that the theme addresses challenges on which governments across the world require support in tackling and that can only be effective through global collaboration. Proposed topics include global warming; clean energy; disaster management; food and water shortages and food and water safety; and inequality between developed and developing countries.

Granting an Award

A high-profile competition focusing on research library and domain science co-operation is recommended as an excellent means for engendering a proactive approach, possibly leading to the creation of further case studies. The competition would be mutually beneficial to participants and the group by fostering engagement with peers and high-profile experts as well as making GRL2020 more visible.

Shaping Research Programmes

Future GRL2020 Workshops should ensure the active participation of representatives from international and national funding agencies, with the aim of shaping future research programmes by placing emphasis on the need for a global knowledge infrastructure. This focus is particularly timely with regards to work between the European Commission Information and Media Society (EC INFOS) directorate and the Cyberinfrastructure Group within the National Science Foundation (NSF) to define a common agenda.

Spotlighting Educational Requirements

A booklet dedicated to educational and training requirements should be produced. This booklet should describe the skills and role of global research librarians, highlighting how library schools should cater for real-world demands by training librarians to be leaders in data preservation, management and electronic publishing. The booklet would be distributed to undergraduate schools, graduate library schools and domain-specific educational institutions. The ultimate goal is to develop a multi-disciplinary programme attracting students from all domains.

Ensuring cross-domain, cross-profession and cross-border outreach

It is essential that the GRL2020 group champions its value proposition to a board audience comprising representatives from national and research libraries, computer science and library technologies, in addition to domain scientists and researchers, new groups and the younger generation. An effective dissemination strategy should be set in place to support this goal and ensure engagement with the target audiences.

Web Design & Dissemination

The GRL2020 website will be re-vamped and designed with the aim of providing dedicated pages on the outcomes of Workshops in the USA, Europe and Asia, showcasing case studies that have been identified, presenting Briefing Papers on the challenges and top priorities deliberated, as well as the Position Papers produced for each Workshop. In order to bring attention to the new website, a special eNewsletter will be circulated to the wider target audience. A special report designed for a broader audience will also be produced and submitted for publication by pertinent journals and magazines. The core GRL2020 group is strongly encouraged to circulate messages about GRL2020, the workshop outcomes and reports and the website.

Enlarging the GRL2020 Community

Dedicated efforts should be made to enlarge the current community by pooling respective networks with special emphasis on domain scientists, library technologists, the younger generation and new communities, in addition to national and research libraries that can benefit from global federation.

Expertise & Session Overview at GRL2020 Asia

International Programme Committee

Donatella	Castelli	National Research Council of Italy (CNR-ISTI)	Italy
Lee	Dirks	Microsoft Research	United States
Fabrizio	Gagliardi	Microsoft Research	United States
Cathrine	Harboe-Ree	Monash University	Australia
Jessie	Hey	University of Southampton	United Kingdom
Yannis	Ioannidis	University of Athens	Greece
Simon	Lin	Academia Sinica Grid Computing Center (ASGC)	Taiwan
Alex	Wade	Microsoft Research	United States

GRL2020 Community at GRL2020 Asia

Jun	Adachi	National Institute of Informatics (NII)	Japan
Subbiah	Arunachalam	Centre for Internet and Society	India
Ya-ning	Chen	Academia Sinica Grid Computing Center (ASGC)	Taiwan
Vic	Elliott	Australian National University	Australia
Nicholas	Ferguson	Trust-IT Services	United Kingdom
Jeremy	Frumkin	University of Arizona	United States
Stefan	Gradmann	Humboldt University Berlin (Humboldt Universität zu Berlin)	Germany
Jieh	Hsiang	National Taiwan University	Taiwan
Huang-Hsiung	Hsu	National Taiwan University	Taiwan
Hong-chu	Huang	Tamkang University	Taiwan
Jane	Hunter	The University of Queensland	Australia
Shuichi	Iwata	University of Tokyo	Japan
Norbert	Lossau	Goettingen State and University Library	Germany
Devika P	Madalli	Indian Statistical Institute	India

Maureen	Mecozzi	AVRDC The World Vegetable Center	Taiwan
Cathy	Norton	Woods Hole Oceanographic Institute Library	United States
Koichi	Ojira	National Institute of Informatics (NII)	Japan
Lisa	Petrides	Institute of Knowledge Management in Education	United States
A.R.D	Prasad	Indian Statistical Institute	India
Arcot	Rajasekar	University of North Carolina	United States
Eloy	Rodrigues	University of Minho (Universidade do Minho)	Portugal
Alfredo	Sanchez	Universidad de las Americas	Mexico
Yoshinori	Sato	Tohoku Gakuin University	Japan
Syun	Tutiya	Chiba University	Japan
Peter	Young	Library of Congress	United States
Xiaolin	Zhang	Chinese Academy of Science (CAS)	China

Sessions Focus

Asian Perspectives on DL development offering in-depth insight into the situation in Asian countries:

- Taiwan Jieh Hsiang, National Taiwan University; Japan Yoshinori Sato, Tohoku Gakuin University; China Xiaolin Zhang, Chinese Academy of Science (CAS); India Subbiah Arunachalam, Centre for Internet and Society.

International Perspectives were also offered:

- Europe Stefan Gradmann, Humboldt University Berlin; USA Peter Young, Library of Congress; Australia Vic Elliott, Australian National University; Latin America Alfredo Sanchez, Universidad de las Americas.

The Workshop also featured focused talks on a variety of domain specific applications and case studies:

- e-heritage Jane Hunter, The University of Queensland;
- Biodiversity Cathy Norton, Woods Hole;
- Climate Change Huang-Hsuing Hsu, National Taiwan University;
- Open Education Lisa Petrides, Institute of Knowledge Management in Education
- Multi-discipline Science Xiaolin Zhang, CAS.

Presentations addressing issues surrounding data were also offered by Shuichi Iwata, University of Tokyo; Ya-Ning Chen, Academia Sinica, Taiwan; Jeremy Frumkin, University of Arizona, US; Eloy Rodriguez, University of Minho; Devika Madalli, Indian Statistical Institute; Maureen Mecozzi, The World Vegetable Centre, Taiwan.

A number of roundtable discussions addressed issues such as challenges facing the digital libraries and infrastructure. The sustainability of the group also came under scrutiny with discussions on a definition of the value add role and identity of the group; the publication of a GRL2020 White paper; the selection of a GRL2020 case study; the establishment of GRL2020 working groups; and collaboration with librarians, domain scientists and funding agencies at future GRL2020 events.

GRL2020 Website www.grl2020.net

- Presentations
- Workshop Outcomes
- Case Studies

