ESI Summit, Rome. 13-14 March 2008

REPORT

Summary

An Earth & Space Science Informatics Summit was convened in Rome 13-14th March 2008 as an Electronic Geophysical Year (eGY) activity. Participants represented the interests of more than 45 leading agencies and initiatives with an interest in geoinformatics. The Summit successfully establishing the basis for better mutual understanding and communication among the leaders of Earth & space science informatics programs worldwide, and confirmed a common resolve to work together cooperatively on data[1] issues that demand a global approach.

Participants noted the extraordinary growth of informatics in the Earth & space sciences, as well as elsewhere, to the extent that informatics is becoming the fourth pillar of the scientific method. At this formative stage, it is inevitable that special interest groups take individual approaches to establishing systems, interoperability protocols, data models, and so forth. Now is a critical time for establishing communication and coordination at the international level to seek uniformity in practices and standards, and reduce replication of effort.

What stood out as the main challenge to be addressed is the lack of infrastructure and governance to (i) cater for the professional needs of scientists and engineers engaged in informatics and (ii) provide an international framework for policy and action. The International Council for Science (ICSU) was recognised as the peak body best positioned to exert the necessary leadership. The Summit applauded the steps already taken by ICSU in this regard, and endorsed enthusiastically the recent recommendations of the ICSU's Strategic Committee for Information and Data (listed in Appendix-A).

Informatics and data stewardship activities are generally a low priority for research scientists. Further, our present reward systems provide little incentive for change. Participants at the Summit regretted this situation as it fails to reflect the growing importance of informatics and the shift in work load from the user to the provider of data. It also compromises the availability and re-use of data. Some simple, achievable steps to rectify this situation are listed in Appendix-C.

In addition to the above broad issues, the Summit dealt with a range of technical, community, marketing, and governance issues. The Summit concluded with a stronger sense of common purpose among the participants and a clearer view of the steps needed to establish a productive international framework for governance and leadership. A series of recommendations were developed under the groupings Governance, Professional Structure and Coordination, Technical and Systems, Marketing, Status and Approaches to take.

[1] The term 'data' is often used generically in this document to cover the data-information-knowledge continuum, spanning both the science and the engineering.

Preamble

The Earth & Space Science Informatics Summit, March 13-14th, 2008 was an Electronic Geophysical Year activity organised by IUGG, IUGS (CGI and OneGeology), and the Earth & space Science Informatics groups in AGU and EGU. The Summit was held in response to an expressed need to establish communication and cooperation among the many geoinformatics and space science informatics entities and initiatives that are emerging rapidly world-wide. Participants at the Summit represented the interests of more than 40 different bodies, agencies, initiatives, and programs (see Appendix-D). The Summit was hosted by the International Geographical Union at the Villa Celimontana (Home of Geography) in the gardens adjacent to the Colosseum in Rome.

Earth & Space Science informatics (geoinformatics in common parlance) is the fastest growing sector of AGU. Schools of informatics are emerging in Universities and elsewhere; new and innovative funding programs are appearing; people are writing use-cases to meet the needs of specialist and general users; standards and their adoption are advancing; community vocabularies/conventions and ontologies are spreading; and working in informatics is gaining respect and becoming popular. Many agencies and bodies are responding.

Data-intensive research requires computer processing power, storage capacity, network bandwidth, and advanced analysis tools. Advances in FLOPS and storage capacity have largely outpaced increases in bandwidth and the development of analysis tools. Because the need to share data, information, and services is so ubiquitous, we all have an obligation to seek unified approaches to common issues in order to reduce replication of effort and a proliferation of different practices.

Origins of the ESI Summit

The need for a summit arose from the several concerns. Science is at the cross-roads. Its relevance and fundability are being re-invented, a gap exists between societal expectations of science and what can be delivered, and informatics is taking on increasing importance in bridging this gap. Interoperability (including common standards and practices) is needed within our science, between sciences, and with users. Geoinformatics groups, meetings, and journals are proliferating, but do they have the same agendas and do they communicate with one another? Greater awareness and coordination among them would be beneficial and major players need to be brought together. Finally, we lack leadership and governance arrangements for geoinformatics at the international level particularly, but also at other levels.

Objectives of the summit

The purpose of the Summit was to learn about each other's informatics activities and plans, identify areas of common needs and overlap where the synergy of cooperation will save effort and produce a better result, promote stronger worldwide coordination, produce a declaration and plan next steps. To this end, participants were asked to address eight questions.

1. What groups of people need geoinformatics capabilities (e.g. scientists, decision makers, educators, private sector) and what do they need them for?
2. What societal, governmental, and commercial imperatives drive investment in geoinformatics?
3. What user-driven needs for geoinformatics are not being met? Are the reasons lack of user motivation, lack of opportunities, or lack of awareness?
4. What current systems of leadership/governance can serve as models for advancing geoinformatics as a discipline, and where is leadership/governance most needed?
5. What principles should be embedded in a declaration of common intent and cooperation?
6. What activities do you undertake or intend to undertake that may be candidates for cooperation and a common approach?
7. What would you like to get out of stronger world-wide coordination in Earth & space science informatics?
8. What are you prepared to contribute to achieving such coordination?

Format of the Summit

The Summit program was organised as a combination of global perspective papers covering the data & information interests of ICSU & SCID, IUGS-CGI, OneGeology, GEO/GEOSS, UN GAID & e-SDDC, the World Data Centres, ESA, EGU, GMES, AGU, GSA-Geoinformatics, NSIDC, and the development of national geoinformatics initiatives in Australia, China, Italy, Russia, and the US. These were interspersed with breakouts and plenary sessions focussing on the questions listed above.

The consensus views of the participants are contained in the statements, comments, and recommendations that appear below.

Informatics – the 4th pillar of the scientific method

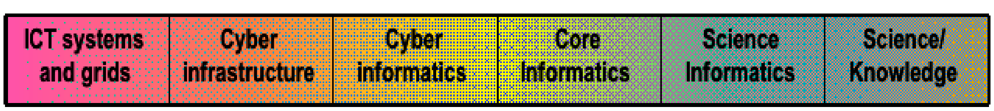
In broad terms, informatics is the science and engineering that fills the gap between the infrastructure of computing systems and networks on one side, and data gathering and knowledge-generation activities for research and decision-making on the other. A definition of informatics, modified from Wikipedia, is:

> *Informatics* - the science of information, the practice of data & information processing, and the engineering of information systems. Informatics encompasses studies the structure, behaviour, and interactions of natural and artificial systems that store, process, and communicate data & information. Informatics has its own conceptual and theoretical foundations, and has computational, cognitive, and social aspects, including the study of the social impact of information technologies.

Informatics is developing rapidly in response to:

- An evolving need for researchers to discover, access, and understand data from outside their own discipline in order to deal with more complex problems.
- Greater urgency - demands for faster response times and rapid access to data & information.
- The need to store, make accessible, and distill an ever-increasing volume of Earth & space observation data; a growing need for processing and user applications to run at the data source.
- The need for software to be available as shared, open source services. (Accessing and analyzing data can be the most time-consuming and costly part of the data-to-information chain.)
- Growing readiness by Governments to support a science information commons[2] as an effective and efficient way to provide societal benefits; wider recognition of the value to Society of open access to and reuse of data & information.
- The power and availability of modern ICT capabilities.

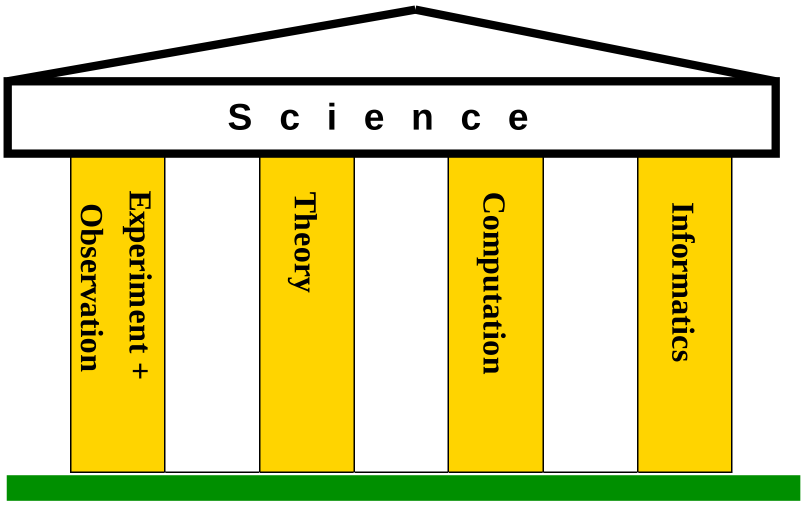
Three sub-divisions of informatics are emerging, forming a sequence from ICT infrastructure systems to science and knowledge generation:



The traditional pillars of the scientific method are observation+experiment, theory, and computation. Modern ICT capabilities now allow us to address a new class of problems and applications – problems that revolve around the organization of data & information leading to knowledge extraction. Examples are recent advances in astronomy, medical science, and hazard risk assessment. Such new capabilities allow science to deliver benefits to society that

[2] An information commons is an environment in which data & information are openly available for everyone to use, either freely or at the marginal cost. The term has arisen by analogy with the traditional agricultural commons – land that was shared by the whole community.

were not possible before, particularly in areas, such as climate change, that involve understanding and modelling the behaviour of complex systems that are not confined to disciplinary boundaries. The information revolution is changing radically the way we conduct our science, and it is little exaggeration to say that informatics has already become the fourth pillar of the scientific method.



Leadership and Governance

To advance the geoinformatics agenda successfully, a balance of leadership (L) and governance (G) are required. These currently occur at many levels and have come about in an ad-hoc fashion, as illustrated in the table below. The biology and the astronomy communities both had strong leadership and each developed a coordinated approach to informatics. As a consequence, they obtained large funding to push themselves forward. The broader geoinformatics community should learn from this.

LEVEL	L & G STATUS	Examples
Working level	L, self-G	Marine Metadata Initiative
National/ regional societies	L, what is role for G?	AGU, EGU, AOGS
'Mission' and 'Production' agencies	G, what is role for L?	BGS, USGS, ESA
Regional programs	some L, no G?	EGSO, INSPIRE US National Geoinformatics System GeosciNet Geochemistry Data Network
Global programs	some L, no G?	GEOSS, GMES, GCOS, OneGeology, IPDA
International Association and Union	some L and some G but not uniform	IAGA, IAU, IUGS(CGI), IUGG(UCDI)
International alliances		IVOA, CEOS, Geoscience Info Consortium
Global, inter-union	G, need L	ICSU, WDC, GEO CODATA, WGISS

- Informatics does not have the ascending structure of other areas of science, which is conceived at lower levels and then geared up to apply at higher levels. Leadership and governance in informatics is evolving in two extreme modes:

  - homogeneous on a working level but heterogeneous at the political/national level (e.g., OGC activities; development of GeoSciML)
  - heterogeneous on a working level but homogeneous at the political/national level (e.g., World Data Centres; GEOSS interoperability).
- In practice, there is a matrix of governance. The two axes are technical and policy - driven by different cultures:
  - the policy culture (in this context) is about ownership and accountability;
  - the technical culture is about delivery of innovative products and capabilities to meet the needs of peers and experts, plus self-Imposed pressure.
- A community consensus approach has produced some notable successes (e.g., the Marine Metadata Initiative, OGC), but has been less successful in other cases. The challenge is how to converge rapidly on a consensus that is lasting.
- The top-down approach has also had variable success. For example, within the World Data Centre System, only 50% of WDCs have achieved their agreed base-line standards of performance, and even within NOAA the three World Data Centres supported have different governing principles.
- Both leadership and governance have a role in how data & information (informatics) efforts respond to user/stakeholder needs, and the increasing need to cross discipline boundaries.
- Areas where leadership and governance are lacking include web services interoperability (OGC is working this area aggressively), IP rights for data and knowledge, and environmental observations (despite the successes of WMO, GCOS, GOOS, GTOS, etc.)
- Two upcoming tests of the effectiveness of community consensus will be the extent of compliance by countries and agencies with (i) the recently-drafted implementation guidelines for the GEOSS data policy, and (ii) the recommendations of ICSU's Strategic Committee for Information and Data (SCID) – see Appendix-D. The recommendations have yet to be ratified by ICSU, but we here assume they will be adopted unchanged.

Common needs

Common needs shared by groups represented at the Summit can be summarized under "seven C's": a need to chart (discover and map what is happening), communicate, coordinate, contract (use the expertise and products of others), collaborate, converge (reduce complexity and proliferation), and seek consensus. Some specific areas of needs that should be targeted are listed below.

Awareness

- Awareness and understanding of the wider scientific and informatics landscape and where the Earth & space science informatics domain fits.
- Better access to information about who is doing what, and awareness of available systems for discovery of data and services.
- Better basis for gap analysis for targeting effort.
- Visibility to scientists and all interested parties of the value of informatics and making data & information readily available - leading to more willingness to partner, preserve and serve data, and actively publish and cite data.
- Support for, and better appreciation of data expertise across the Earth & space sciences

Forum, cooperation, agreements

- A coherent international forum and lobby for addressing informatics issues - raising awareness, promoting better data management practices, coordinating interoperability protocols, rationalising international geospatial standards, and so forth. Such a forum would provide a focus for cooperation on a whole range if issues where there is a risk of proliferation of incompatible standards and approaches, and unnecessary replication of effort.

- Cooperation among bodies at national, regional, and international levels on Earth & space science informatics issues. Greater inter-society cooperation, leveraging of the activities and structures of the Scientific Unions and ICSU bodies
- Wider participation in the GEO Standards and Interoperability Forum

Best practice and policy

- International agreements on best practices and standards, especially for data management and interoperability principles. We need guidelines for developing data models and mark-up languages; services, protocols, and interfaces; models and simulations for forecasting.
- An ICSU data policy, agreed on by all Unions. The data policies of GEO and other bodies provide a starting point. Process of policy developing is important to ensure a widespread sense of ownership.
- Wider adoption of policies for data access, preservation, publication, attribution (citation), and intellectual property rights. Policing of agreed data policy obligations is necessary – for example by funding agencies, which can make future funding conditional on fulfilment of past obligations.
- A system for World Data Centre certification. Can be used for other data centres.
- A system for certification of data and evaluating data quality.
- Contributions & editors for a Best Practices Wiki; coordinate with CODATA?

Systems/Interoperability/Models

- Coordination of data system development and operation; accelerated interoperability and convergence on more specific interoperability principles; avoidance of silos; gradual development of a distributed network of independent but interoperating systems and models with appropriate interfaces to end-users.
- User-friendly, open source processing tools (image processing, GIS, …)
- Better data fusion capabilities and usage, bringing data together for a usable information product.
- Advances in Multidisciplinary Model systems interoperability. This requires a high-level forum to discuss common models and interfaces for multidisciplinary model predictions (see position paper on GEOSS IP3/SIF).
- Advancement of netCDF conventions; establishment of a forum for the advancement of netCDF conventions. CF-NetCDF conventions are truly enhancing interoperability across the geoscience disciplines and the geospatial community (e.g., the OGC GALEON Interoperability Experiment).
- Greater engagement with the development of GEOSS and outreach to all GEO tasks.

Information commons, open access

- Full, open access to metadata, data, and information. Free (marginal cost only) data for education, research and disaster response - available free or for the cost of reproduction and in a timely manner.
- Increased availability of low cost or free remote sensing data.
- Reduction of the growing gap between developed (rich) countries and less economically developed countries (LEDC) in networking conditions, data holdings, and the ability to exploit informatics capabilities.
- Reduction of the Digital Divide - better modern networking capabilities in the LEDCs; promoted and implemented by major global IT companies e.g., Microsoft, Google, Intel, etc. Perhaps also supported by major corporations who have profited recently by world economic conditions. Decrease the gap between developed, developing, and LEDCs in data acquisition, holding, exploration, and research opportunities.
- An internationally agreed process for publishing and citing digital data; advances in making data self-identifying (ownership, lineage). Journal editors can require that authors of data-

based publications cite data sources and processing methodologies, with detail matching some standard, and that data sources are accessible.

Incentives for good data practices

- Higher status for data activities and expertise (e.g., in research and development in informatics, building information systems, providing/serving data & information, publication of data, declaring data policy and achieving compliance), on a par with other scientific disciplines.
- A culture among scientists and engineers that values data management, systems development, and the provision of data, information and services for use and reuse by others. Acceptance of the full life-cycle of managing data, interoperability and access, and so forth as an intrinsic part of the scientific method.
- Personal reward systems that encourage the above (steps to take are listed in Appendix-C).

Common contributions

Groups represented at the Summit are willing to contribute the following to achieve cooperation.

- Raise awareness and encourage scientists in our individual communities and at the regional society level to engage more with informatics.
- Help to establish a peer community - serve on organizational bodies, conference committees, convene sessions, serve on editorial boards, submit and review papers, establish regional nodes, e.g., in Europe; contribute to the establishment and running of proposed forums (particularly by SCID).
- Provide two-way communication between our own domains and the broader international community on issues of informatics, and also science, marketing, and politics. Share access to the networks of people in our own domains – e.g., the extensive network in "ground-truth" geology via the global geological survey network; the WDC coalition. Share information about communities where collaboration is already established.
- Share lessons learned in the course of developing systems and services (e.g., from OneGeology; the European Spatial Data Initiative; ICSU review of WDCs and FAGS).
- Engage with agencies and funding bodies; offer to convene conference/workshop sessions around their programs that will encourage PI participation and broaden dissemination of project information and results.
- Put effort into identifying, developing, and implementing universal and community of practice procedures and standards.
- Identify common issues among peak professional bodies (Scientific Unions, AGU, EGU, and others), promote coordination, and use society meetings and structures to organise joint activities.
- Share the technical expertise, systems, and organisational know-how that are available through our own agencies, bodies, and communities - for example from the WDCs, the OneGeology cookbooks for LEDCs, National Geological Surveys.
- Provide mutual support for each other's programs

Conclusions (the emerging strategic environment)

Information revolution. The modern information era is changing fundamentally the way science will be conducted in the 21st Century and beyond – with a strong focus on data & service access capability, interoperability, large data volumes, breakdown of traditional disciplinary boundaries, real-time response, and so forth. Data & information science has become the 'fourth' pillar of the scientific method (along with experiment & observation, computation, and theory).

Status of informatics. The scientific community under-invests in both time and money in informatics. Data science is widely given inferior status to other branches of science (if, indeed, it is considered a branch of science at all) and research scientists are reluctant to spend time on good data management practices. This can be remedied through a combination of a revision of the reward system for scientists and adoption of data policies by organisations and institutions that dictate good practice. Steps proposed to achieve this are listed in Appendix-C.

Professional structure. There is no formal professional structure for informatics. Such a structure would provide a framework for defining best practices and standards, developing international cooperation, and catering for the professional needs of informatics scientists, engineers, and technicians. A formal structure would also help to place data science/informatics on a par with traditional scientific disciplines.

Coordination. We see a growing convergence between the WDC System, FAGS, CODATA, and GEO. We endorse the SCID recommendation for a new ICSU Strategic Coordinating Committee for scientific data & information, with similar informatics committees in the Scientific Unions and other component bodies of ICSU. (ICSU favours this development – IUGS has CGI, and IUGG is planning a Union Commission for data & information.

WDC System. The network of WDCs is expanding. As many as fifteen new WDCs are coming on line in the coming year, based on existing data centres in developed countries, and a new-paradigm WDC is to be established in a LEDC. A WDC certification scheme (scale of standards/performance) has been discussed as a means of informing users about the capabilities of a particular data centre. Certification might also help to improve the capabilities and performance of data centres and encourage investment from funding bodies. A certification scheme developed for WDCs could be adapted for any data centre - for example, to deal with private-sector data centres that have restricted access to data.

Open access and interoperability. NASA's EOSDIS set new standards for stewardship of data and open access. GEOSS also has the opportunity to lead by example. The view was expressed during the Summit that we should commit to even stronger interoperability principles than are currently contained in the 10-Year Implementation Plan for GEOSS

Mega-projects. Mega-projects have been successful in highlighting advances in data & information science. Examples are One Geology, GEOSS, and GMES. However, some have 'failed' (noted in the presentations). What are the new ones coming along?

Micro Projects. Micro projects are one way we make practical progress, encourage innovation, and create the building blocks of larger systems. There are too many to note here (a draft list called VODISlist can be found under the Working Documents section of www.egy.org). Many are very successful but they seldom have a sustainability path.

We expect that Government funding agencies are increasingly likely to support research initiatives that have a strong professional informatics basis, rather than those that either ignore this area or assume that any informatics requirements can be undertaken peripherally by the scientific investigators.

Next Steps and Recommendations

The needs listed above could and should be fostered via a coherent lobby for improved data management in the Earth & space sciences. There is a clear role here for ICSU, together with the Scientific Unions, CODATA, FAGS, and the World Data Centre System (WDS if the SCID recommendations are implemented). Efforts to foster cooperation among bodies regarding Earth & space science informatics issues should lever off the existing international frameworks of peak professional bodies, alliances, societies, scientific Unions, and ICSU. The Summit agreed strongly with the recommendations of SCID.

After each recommendation below, we suggest [green type in brackets] bodies that should take a leading role in response, but in consultation with the wider community of both informatics experts and users.

1. *Outreach and awareness*

   1.1 Increase awareness and understanding of the wider scientific landscape and where the

Earth & space science informatics domain fits. [Everyone]

1.2 Encourage individual informatics communities to be aware of and plan within the global framework provided by organizations such as ICSU, WDC System, CODATA, and GEO (GEOSS) and find advocates such as: eGY, IPY, societies, and academy panels. [Everyone, ICSU, Scientific Unions]

1.3 Reach out and bring in others: Unions (e.g. soil science); professional societies; cyberinformatics bodies (ACM, IEEE, Computer Society, GIS), mission agencies and commercial interests.

1.4 Publish results, experiences, news, and progress in widely accessible places.

1.5 Market key major projects. [Everyone]

2. *Professional structure and governance*

2.1 Build a formal professional structure and governance for informatics, centred about ICSU. This would provide the international connectivity (coordination) that is mentioned by many programs (e.g. IPY). CODMAC2 would fit well into such a framework as the U.S. voice/inspiration in international deliberations. Start by creating a data & information group within each of the GeoUnions, then spread beyond. IUGS already has such a group (CGI) and IUGG is planning a Union Commission for Data & Information. [ICSU, CODATA, Unions, WDS, FAGS]

2.2 Establish a forum for cooperation and lobbying to deal with specific informatics issues - data access, certification, data interoperability, technical drivers, and so forth. [ICSU, CODATA]

2.3 Explore the feasibility of a distinct informatics society to serve the emerging peer community. (We note that standards bodies such as ISO, W3, OGC do already serve some of the functions of a society.) An informatics society would complement rather than compete with current science-discipline based forums.

2.4 Explore the feasibility of an international institute for informatics. We note the International Space Science Institute (ISSI) in Bern as one example. Another is the Institute in Vienna, IIASA, ICTP, and GBIF.

3. *Best practice & policy*

3.1 Establish international agreements on best practices and standards, especially for data management; interoperability principles; guidelines for developing data models and markup languages; services, protocols, and interfaces; and models and simulations for forecasting. [ICSU, together with SIF, OGC, WGISS, and others. Control should be decided via a Forum of authorities].

3.2 Encourage the relevant groups to press ahead with WDC certification; design a scheme that can be expanded to cover any data centre. [ICSU, WDS, eGY, in consultation]

3.3 Produce a generic ICSU data policy agreed to by all Unions, which can be customised to meet their particular needs. [ICSU]

3.4 Share data policy resources, DFID… [Everyone]

3.5 Participate in the GEO Standards and Interoperability Forum (SIF) and other forums as a means of establishing agreement on best practices and standards. [Representatives of lead bodies and programs worldwide]

3.6 Identify the good ideas that are ready to be shared and bring them together (package them like eGY has done, or as in the INSPIRE synthesis work). Expose these best practices so they are easier to find.

3.7 Continue to press for open and timely access to data for scientific research. [Everyone]

4. *Technical and systems*

4.1 Set guidelines for X-data models and X-markup languages; services, protocols, and interfaces; and models and simulations for forecasting. There is a role here for SIF, OGC, ICSU, WGISS, and others.

4.2 Check out the benefits of even stronger interoperability principles than those currently contained in the 10-Year Implementation Plan for GEOSS [ICSU, Unions, AGU-ESSI, EGU-ESSI, GEO Architecture & Data Committee, SIF, WGISS]

4.3 Establish data citation procedures and conventions. Provide guidance to scientists on how to write a data-based paper, bearing in mind that reproducibility is a cornerstone of science. In data-based research the raw data and a complete history of processing steps is required. Many scientific papers contain an incomplete description of data processing methodology. [ICSU, Unions, CODATA, Societies]

4.4 Establish guidelines, best practices and standards to help make data self-identifying (ownership, lineage). [ICSU, CODATA]

4.5 Work with standards development organizations like ISO and OGC to develop a standard structure for exposing geoprocessing services.

5. *Status*

5.1 Advance data/information and informatics as the 4th leg of society and science progress

5.2 Change the reward system for scientists so that fulfilling data responsibilities becomes a benefit to the scientist (e.g., for career advancement) rather than a hindrance. Change is needed to elevate informatics as a science/service and to motivate scientists to adopt more responsible data management practices.

5.3 Encourage journal editors to require that data-based publications cite data sources and processing methodologies with detail matching some standard.

5.4 Increase the relative funding (%) for data & information management. It was noted that it will be hard to make this case since many assume that data management will be done, and that science activities often takes the share at the expense of data & information.

6. *Approaches to take*

6.1 Keep it simple, network first, then proceed to more structure if warranted. Chart (discover and map what is happening), communicate, coordinate, contract (use other's expertise and products), collaborate, converge (reduce complexity and proliferation), and seek consensus.

6.2 Take a two-prong approach - make contacts with people and work using existing/planned structures

6.3 Communication and coordinate up and down the hierarchy noted earlier in this report, e.g. society to Union. Several gaps are present in this communication flow at the present time. In dealing with gaps, show leadership rather than attempt governance.

Report editors: C.Barton, P.Fox, I. Jackson, S.J.S.Khalsa, M. Messerotti, S. Nativi, L.Wyborn
27 June 2008

APPENDIX-A. ICSU/SCID Key Recommendations

The key recommendations in the Executive Summary of the SCID draft report are that:.

1) ICSU assert a much-needed strategic leadership role on behalf of the global scientific community in relation to the policies, management and stewardship of scientific data & information;

   In order to achieve this, ICSU must reform some of its current interdisciplinary bodies and establish a new committee that will provide overall strategic direction and advice.

2) a new World Data Services system be created (as an ICSU Interdisciplinary Body), incorporating the WDCs and FAGS as well as other 'state of the art' data centres and services;

   This new structure or system must be designed clearly to support ICSU's mission and objectives, ensuring the long-term stewardship and provision of quality-assessed data and data services to the international science community and other stakeholders.

3) CODATA to focus its activities on the three main initiatives identified in its draft strategy and extend its links to other organisations and networks to play a more prominent role within ICSU and within the wider scientific community;

   This will require the close alignment of implementation mechanisms, e.g. working groups and task groups, with the 3 main initiatives identified in the draft CODATA strategic plan. The bi-annual CODATA conference should also be modified to provide closer links to ICSU priorities and the new World Data Services system.

4) a new ad hoc ICSU Strategic Coordinating Committee for Information and Data be established to provide broad expertise and advice to ICSU in this area;

   This Strategic Coordinating Committee will act as an interface between scientists and data & information professionals that can advise on the data needs and possible solutions for existing and new ICSU programmes and other international initiatives. It will enable ICSU to establish visible and effective leadership and ensure proper coordination among ICSU activities.

5) ICSU National Members and Unions be strongly encouraged to establish committees or commissions, where these do not already exist, focussing on informatics issues;

   Where national committees or liaison structures already exist for CODATA and/or the WDCs, consideration should be given to amalgamating and expanding these to integrate data policy, management, stewardship and other informatics issues. Professional data services must be recognised and supported at the national level as part of the long-term infrastructure of science.

APPENDIX-B. ANSWERS TO QUESTIONS

This appendix summarises the responses from participants to some key questions and supplements the information in the main part of the report.

What groups need informatics capabilities and what do they need them for?

Answer: people involved in the data-information-knowledge continuum who use modern ICT in order to accomplish their objectives. Nowadays, this includes most professionals and most sectors of society, including the general public. The broad categories are:

- Governments, mission agencies. Needed to do their jobs.
- Industry, commerce, and agriculture (wealth generators).
- Decision-makers and policy-makers - health care, natural resource management, land and urban planning/management, resource exploration and assessment, sustainable development, natural hazard mitigation, civil protection, emergency management. Needed for data & information discovery; decision support and expert systems to access comprehensive, cross-disciplinary, distributed data, tools, and services.
- Scientific researchers (specialist and multidisciplinary), field scientists, information & computer scientists, librarians, and students. Needed as a source of fundamental tools and to guide application research.
- Educators. Needed to access teaching resources, computer-aided learning, and management tasks.
- Developing countries (LEDCs). Needed for the same purposes as any other country, but also as a low cost and efficient means of accessing data & information resources, international expertise, the ability to participate in regional and worldwide initiatives and activities.
- The public. Needed for information gathering, synthesis, visualisation, planning, news, and alerts.

What do users need in general?

- Information models: data & metadata semantic and content models; encoding model & markup languages
- Services (systems) interfaces: service models, API models, etc.
- Systems for non-specialists as well as experts
- Observational data, simulation models, real-time response, data management facilities and tools, processing capabilities, …
- Formalization, standardization, harmonization, usability, simplicity, transparency, extensibility, flexibility, interoperability, accessibility, security, quality of service …

	Useful services	**Main requirements**	**Present availablity**
Scientists	Discovery and access Data access Dataset archiving	Harmonisation Interoperability	Few
Decision makers	Value-added processing Knowledge extraction Transparency	Harmonisation Security Interoperability	Few
Civil protection	Discovery and access Value-added processing Scenario planning Knowledge extraction	Harmonisation Near-real time Security Quality of Service Interoperability	Few
Educators	Discovery and access Resource sharing Computer-aided learning	Harmonisation Interoperability Open access	Poor

	Dataset archiving		
Private sector	Service delivery Marketing	Propriety rights	OK

What societal, governmental, and commercial imperatives drive investment in geoinformatics?

- Effective management of the natural world to provide the ecosystem services needed to sustain humans. Natural resource monitoring, assessment, and prediction; sustainable use of the environment.
- Health care; security and hazard mitigate
- Wealth generation & recreation – industry, commerce, finance, tourism, ….
- The realisation by funding agencies that infrastructure for data & information is necessary but expensive, and that significant redundancy and ineffectiveness still exists. We note that commercial support of information repositories is flourishing,
- Access to research data & information by non-traditional stakeholder (e.g., decision-makers, users and developers of GEOSS); a need to integrate diverse information sources, to assess suitability, reliability, etc. Barriers are a lack of resources and political/technical issues.
- Access to research data & information for education and training, bringing relevant and timely data into the view of students.
- Appreciation by commercial companies that re-package data & information as services (e.g., weather) that they must invest in informatics in order to utilize available data and support systems.

What user-driven needs for geoinformatics are not being met?

- Quality assured shareable data content – gaps and lack of standards (i.e., good data management).
- Interdisciplinary data and service discovery.
- Shared data & information infrastructure.
- Interoperability of services, higher-level products.
- Data integration, data fusion.
- Access to model predictions that can answer user's questions.
- Appropriate tools for applications end users.
- Data management expertise in LEDCs to improve their contribution and ability to participate.
- Ability to find out who is doing what - successes, failures, and challenges.
- Lost opportunities: many application areas where there is currently under-utilisation of the products of research and observing programs, and where infrastructure in support of new programs/initiatives is lacking.

Are the reasons lack of user motivation, lack of opportunities, or lack of awareness?

- All are factors that hinder the development and usage of informatics. The main factor is motivation. Science demands are high and research scientists are rewarded for publication of new scientific results only. So scientists are not likely to devote time to data management, serving data, or developing open community systems. The ability to work with informatics professionals is often a low priority, and there is little or no reward for going beyond inventing it yourself.
- As research scientists we typically rely on chance identification of opportunities and connections. In informatics, it is more productive to put the emphasis on forward-planning of facility capabilities and models, etc. to meet current and anticipated needs – needs of not only traditional science, but also of the new science and applications that are now possible.

- Collaboration is not rewarded, recognized, nor sought. Collaboration is seldom facilitated by societies, which are based on discipline boundaries.

Where has leadership/governance been successful (e.g., for established best-practices)?

As in all endeavours, leadership and governance originate from individuals and bodies with a clear, simple, realistic vision and determination. A commercial imperative is a strong driver.

- Climate and Forecast conventions
- CODATA – which has a strong governance structure but is weak on leadership in modern informatics.
- INASP – strong in both leadership and governance for open access to publications.
- CEOS-WGISS – successful in both governance and leadership for satellite missions and satellite data management.
- The WDC System: long established and successful, but being outpaced by informatics developments. (ICSU has recognized this and new guidelines are on the way.)
- Web services standardization - W3C, OGC, ISO
- e-Business standardization and interoperability - OASIS
- Geospatial Information Standardization efforts - ISO TC211, OGC
- Spatial Data Interoperability initiatives - INSPIRE for Europe
- Global Earth Observation interoperability - GEO & GEOSS.

Where is leadership/governance for data & information most needed?

- The greatest need is in the senior leadership of peak professional bodies (particularly in the ICSU family) and in national science organizations, where data management is a low priority for investment.
- Within the Earth & space science community, particularly for achieving interoperability across disciplinary boundaries and catering for non-specialist users.
- Interoperability between the Earth & space science community and the spatial information and remote sensing community.
- Environmental observatories.
- WDC and FAGS - see ICSU-SCID recommendations for a World Data Services system.
- Millennium Ecosystem Assessment (MEA).

Geographic information (GI) Standardisation framework		
ICT	**GI**	**E&SS**
Semi-structured models Science Markup Languages WS-I Grid services MDA SOA ROA …..	ISO 19100 series CEN TC 287 spec.s OGC OWS, GML, GeoRM, SWE, .. ESA application profiles, ….	GEOSS IP3 GEOSS AIP GMES services INSPIRE pilot tests OGC inter. tests OGC GALEON IE CEOS …. …

What principles should be embedded in a declaration of common intent and cooperation?

- Full open access to metadata, data & information within the constraints of legitimate propriety or privacy restrictions. Removal or reduction of commercial & legal restrictions on the availability of data. Open access to the discovery of data/information/knowledge by all (not just scientists).
- Timely availability of data, including real-time when appropriate.

- Minimal cost. Free (or cost of supply only) data for education, research, and emergency management.
- Long-term view, sustainable by design in all elements.
- Open exchange of expertise, practices, and technologies.
- Scientists and stakeholders should be expected to participate together in a community of practice, with governance roles.
- Capitalisation on the evolving discipline-agnostic technical infrastructure.
- Enablement of informatics skills and knowledge development for the next generation of scientists and engineerings.
- Funding for research should carry with it a requirement to manage, preserve, and serve project data.
- In a time of rapid change and interdisciplinary use, publications should not only require data citation, but also mandate that the data are made available.
- Interoperability requirements and opportunities should be catered for (following accepted best practices, adopt existing standards, etc.)
- Existing entities for standards for metadata and interfaces, especially service-oriented architectures, for data & information frameworks should be used to facilitate interoperability

What activities do you undertake or intent to undertake that may be candidates for cooperation and a common approach?

- Participation in Assemblies, meetings, and workshops of the ICSU Scientific Unions and Associations, peak professional societies, and the CODATA bi-annual conference. Joint sessions at conferences to advance shared understanding and expertise
- INASP and CNDI
- The Model Web - the model component of the Semantic Web.
- The TerraLook project - http://terralook.cr.usgs.gov
- Expansion of the WDC system, including a new data centre for LEDCs.
- GeoSciML (a markup language for geoscientific, mainly geological data). OneGeology
- SCCID forum
- eGY and following legacy activities (CODATA Working group and IUGG Union Commission for Data & Information)
- eGY-Africa: to get better internet connectivity for African scientists.

Disciplinary working groups that contribute to specifications

- Discipline-specific WDCs
- Scientific Unions (IAU, …)
- SPASE
- IAU commission 12 Data WG
- CEDAR
- Climate and Forecast conventions
- IVOA
- IOOS/DMAC

Interdisciplinary working groups that contribute to specifications

- CODATA
- INASP
- CDSI
- CEOS - WGISS

APPENDIX-C. Steps for rewarding (promoting) good data practices

The reluctance of many research scientists and other professionals to spend time on preserving and serving their data can be attributed to a lack of rewards. Scientists get promoted primarily on the basis of their research publications, not on responsible data stewardship. The following steps are proposed to change the reward system so as to make it worthwhile for scientists to invest time in ensuring that data are properly managed, preserved, and made available for re-use.

(1) Establish a ‘data’ attribution system – a system for acknowledging in publications and metadata the source provider of data, systems, or services that are used. This can be modelled on our traditional method of citing sources in research papers. For both individuals and organizations (data centres, for example), a ‘data’ citation index would be a metric of performance and relevance.

(2) Require that all projects and project proposals have a data policy. Funding agencies can insist on this as a prerequisite for funding. Data policies could even have a certification rating that would help to establish what type of data policy is expected by all parties concerned. Data policy certification could be included as an addendum to the proposed World Data Centre certification scheme.

(3) Police project data policies to ensure compliance. Funding agencies are well-positioned to do this by making future funding for investigators conditional on meeting past commitments. Some funding agencies are already doing this. Such a system has long applied to help ensure that investigators publish the results of their scientific work.

(4) Include ‘data’ contributions as a routine consideration for career advancement. People should maintain a list of their main ‘data’ contributions and might wish to give citation figures as a measure of impact. A candidate’s ‘data’ performance should be considered on equal terms with his/her scientific performance for career advancement purposes.

(5) Raise awareness. Publish articles in newsletters, EOS, and elsewhere; include material on the importance of data stewardship in conference presentations; engage with CODATA, ICSU, and others who share the same objectives.

(6) Challenge the view that data acquired using public funds is the personal property of the responsible scientist(s). Rather, portray the scientists as a custodian with an obligation to ensure that data are managed to maximise their value to the whole community today and tomorrow.

APPENDIX-D. Bodies and initiatives represented

1. AGU-ESSI – AGU's Earth & Space Science Informatics group
2. Australian Academy of Science - National Cmte for Data for Science
3. Chinese Academy of Sciences
4. CNR – Italian National Research Council
5. CODATA - ICSU's Committee on Data for Science & Technology
6. EGU – ESSI and Hydroinformatics
7. eGY - the Electronic Geophysical Year, 2007-2008
8. eGY-China
9. eGY-Europe
10. eGY-Germany
11. eGY-Italy
12. eGY-Russia
13. EGU – Earth & Space Sciences Informatics programme group
14. ESA ESRIN, ESA Interoperability Program (SSE)
15. GEO SIF (Standards & Interoperability Forum) and ADC
16. GEOSS AIP,
17. GMES-Global Monitoring for Environment & Security,
18. GSA Geoinformatics Division
19. IAGA - International Asssociation of Geomagnetism & Aeronomy
20. IASA, Russian Academy of Science
21. IASPEI - International Association of Seismology & Physics of the Earth's Interior
22. ICSU - International Council for Science
23. ICSU-SCID Strategic Committee for Information & Data
24. IEEE Geoscience & Remote Sensing System; Committee on Earth Observations (ICEO) Standards Working Group
25. iGeoInfo
26. IGU - International Geographical Union
27. INAF – National Institute for Astrophysics (Italy)
28. INGV - Istituto Nazionale di Geofisica e Vulcanologia (Italy)
29. IUGG - Union Commission on Data & Information
30. IUGS-CGI – Commission for the Management and Application of Geoscience Information
31. IZMIRAN – Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation
32. LASP – Laboratory for Atmospheric and Space Physics
33. National Geoinformatics System (US)
34. NCAR – National Center for Atmospheric Research
35. NOAA - National Geophysical Data Center
36. NSIDC - US National Snow & Ice Data Center
37. OneGeology
38. RAS Geophysical Centre
39. Russian Academy of Science (Geoinformatics)
40. Russian Institute of Physics of the Earth (Geoinformatics),
41. SEEGrid – the Solid Earth & Environment Grid (Australia)
42. UN-GAID e-SDDC
43. Virtual Observatory initiatives
44. WDC System and Panel
45. WGISS – WG on Information Services & Systems, Committee on Earth Observation Satellites

APPENDIX-E. Acronyms

AGU – American Geophysics Union www.agu.org
AOGS – Asia Oceania Geosciences Association www.asiaoceania.org
BGS – British Geological Survey www.bgs.ac.uk
CEDAR – Coupling, Energetics, and Dynamics of Atmospheric Regions cedarweb.hao.ucar.edu
CEOS – Committee on Earth Observation Satellites www.ceos.org
CGI – IUGS Commission for the Management and Application of Geoscience Information www.cgi-iugs.org
CODATA – ICSU's Committee on Data for Science & Technology www.codata.org
CODMAC – (US) Committee on Data Management, Archiving, and Computing
DFID – Department for International Development
EGSO – European Grid for Solar Observatories www.egso.org
EGU – European Geosciences Union www.copernicus.org/EGU
eGY – The Electronic Geophysical Year, 2007-2008 www.egy.org
ESA – European Space Agency www.esa.int
eSDDC – UN-GAID Global Alliance for Enhancing Access to and Application of Scientific Data in Developing Countries www.un-gaid.org/en/node/165
FAGS – Federation of Astronomical and Geophysical Data Analysis Services www.icsu-fags.org
GBIF – Global Biodiversity Information Facility www.gbif.org
GEO – Group on Earth Observations earthobservations.org
GeoSciML – Geoscience Markup Language
GEOSS – Global Earth Observing System of Systems www.earthobservations.org www.epa.gov/geoss/index.html
GCOS - Global Climate Observing System www.wmo.ch/web/gcos/gcoshome.html
GICSI – CODATA'S Global information Commons for Science Initiative
GIS – geographical information system
GMES – Global Monitoring for Environment and Security (European) www.gmes.info
GOOS - Global Ocean Observing System ioc.unesco.org/goos
GSA - Geoinformatics Division geoinformatics.sdsc.edu/gsa_division
GSA – The Geological Society of America www.geosociety.org
IAU - International Astronomical Union www.iau.org
ICSU – International Council for Science www.icsu.org
ICTP – International Centre for Theoretical Physics, Trieste www.ictp.it
IEEE – Institute of Electrical and Electronic Engineers www.ieee.org/portal/site
IIASA - International Institute for Applied Systems Analysis www.iiasa.ac.at
INAF – Italian National Institute for Astrophysics www.inaf.it
INASP - International Network for the Availability of Scientific Publications www.inasp.info
INGV – Italian National Institute of Geophysics and Volcanology www.ingv.it
INSPIRE – Infrastructure for Spatial Information in Europe inspire.jrc.it
IOOS/DMAC – Integrated Oceans Observing System / Data Management, Archiving, and Computation
IP3 – GEOSS Interoperability Process Pilot Project
IPDA – International Planetary Data Alliance www.planetarydata.org
IPY – International Polar Year www.ipy.org
ISSI – International Space Science Institute, Bern www.issibern.ch
IUGG – International Union of Geodesy & Geophysics www.iugg.org
IUGS – International Union of Geological Sciences www.iugs.org
IVOA - International Virtual Observatory Alliance www.ivoa.net
LEDC – Less economically developed country
NSIDC – (US) National Snow & Ice Data Center, Boulder, Colorado
OGC – Open Geospatial Consortium www.opengeospatial.org
SCID – ICSU's Strategic Committee on Information and Data /www.icsu.org/5_abouticsu/STRUCT_Comm_Adhoc_scid.html
SCCID - ICSU's Strategic Coordiinating Committee on Information and Data
SIF – GEO Standards and Interoperability Forum
SPASE - Space Physics Archive Search and Extract www.igpp.ucla.edu/spase
SCID – ICSU's Strategic Coordinating Committee for Scientific Information & Data
SCCID – proposed ICSU Strategic Coordinating Committee for Scientific Information & Data
UCDI – proposed IUGG Union Commission for Data & Information
UN-GAID – UN Global Alliance for ICT and Development. www.un-gaid.org
VO – Virtual Observatory
WDC – World Data Centre (System) www.ngdc.noaa.gov/wdc/wdcmain.html
WDS – World Data Services