PERMAFROST - Seventh International Conference (Proceedings),
Yellowknife (Canada), Collection Nordicana No 55, 1998

PERMAFROST DATA AND INFORMATION: ADVANCES SINCE THE FIFTH INTERNATIONAL CONFERENCE ON PERMAFROST

M.J. Clark[1], R.G. Barry[2]

1. GeoData Institute, University of Southampton, Southampton S017 lBJ, UK.
e-mail: mjc@soton.ac.uk

2. World Data CenterA for Glaciology, University of Colorado, Boulder, CO 803900449 USA.
e-mail: rbarry@kryos.colorado.edu

Abstract

In the decade since 1988, the International Permafrost Association Working Group on Data and Information has overseen a remarkable transformation of permafrost data strategy, and a substantial acquisition of permafrost data, metadata and information. The international commitment to the principles of data preservation and rescue is now strong in the scientific community, but funding remains problematic and individual transfer of data is often slow. The technical problems of data model and structure, data management and data dissemination are now well on the way to solution, but the "political" hurdles of data commodification and personal reluctance to share data remain significant. The value-adding capabilities of temporal and spatial comparison, identification of the drivers of variation, hypothesis testing and support for global typology definition offer powerful incentives to continue the data preservation and access objectives of the IPA Global Geocryological Database (GGD).

Introduction

Data increasingly fuel the scientifically lifesustaining processes of description, analysis and modelling. It is a truism to say that without data there can be no cold regions science or engineering, but against this background it is astonishing that the preservation and sharing of such a missioncritical resource comes so low on the priorities of the average polar scientist or engineer that they ignore it altogether. In terms of data, as in so many contexts, we are a wasteproducing society, and repeatedly ignore (or worse, reject) opportunities to preserve, recycle and enhance the "energy" which drives our work. There are many missed opportunities in the annals of the far North and South, but none so great as this. The global technical community has long recognised the inherent and lasting value of data, and scientists planning the International Geophysical Year 1957/58 responded with the establishment of the World Data Center network including WDCA, B and C for Glaciology, but the polar community, and permafrost scientists and engineers in particular, have responded rather more slowly. Other structural changes are reflected in the attention now given by U.S. and European funding agencies to the delivery of data sets by funded projects as a condition of the funding support.

Since the Fifth International Conference in 1988 in Trondheim, some significant progress has been made, but the initiative is vulnerable and complacency would be ill-placed. While the Global Geocryological Database (GGD) project of the International Permafrost Association (IPA) now begins to show the way, it is the culmination of a decade of effort and will require significant continuing investment if the impetus is to be maintained. The underlying question is whether or not such commitment provides a meaningful return on this investment. The answer to this question lies partly in an understanding of the evolution of IPA data strategy, and partly in a creative recasting of that strategy to meet key needs of the profession in the future.

Background to the IPA Data Strategy

The groundwork of the IPA strategy for data and information management was laid during the 1988 IPA Conference in Trondheim, Norway. An informal workshop on requirements for permafrost data was held, with over 100 attendees, and a related paper was presented at the conference (Barry, 1988). The interest expressed by the workshop attendees led to the establishment of the IPA Data and Information Working Group (DIWG) and to the holding of a formal workshop at the Beijing IPA Conference in 1993. The DIWG has subsequently organised workshops in Southampton and Oslo (1994), Potsdam (1995) and Boulder (1996), which have resulted in progressive steps being made in the prioritisation and assessment of

permafrost data requirements (Table 1), the identification of critical data sets for scientific and engineering purposes including global change, monitoring, modelling, impact assessments (Table 2) , and efforts to acquire the necessary resources to assemble some of these data.

The next crucial item was the provision of a funding base to support the work since there is no agency with responsibility for data on frozen ground conditions, at least in the United States, and Canada has effectively abandoned this area within Environment Canada as a result of agency retrenchment (Barry, 1995). Support from the IPA for the Working Group planning meetings was bolstered by a NSF grant for developing a pilot Global Geocryological Database (GGD). This supported the digitisation and assembly of selected data and information from Russia and the inventory and organisation of other priority data sets identified by the Working Group. Subsequently, a CD-ROM version of parts of the GGD has been published in conjunction with the IPA Yellowknife Conference (Table 3).

Data futures for the IPA

Background to the Current Debate

The period 1988-1998 has built for the IPA both a data strategy and an embryo data archive. In so doing it has laid a foundation for honouring that most basic of scientific traditions respect for data as the validation of present work and the illumination of future trends. But such ideals are not selffulfilling: they require careful nurturing if they are to achieve their potential. Figure 1 shows the primary underlying functions of the GGD as being:

- Process understanding (basic science)
- Change detection
- Model validation (for science and engineering)
- Engineering design
- Impact evaluation (and management response)

These objectives range from pure research to management application, and from the site scale to global processes. In so doing, they provide a set of worthy aims, but beg the question as to how these targets might be met. To this end, a number of possibilities have been considered including:

- Comparison in time and space: In practice, most engineering and science investigations have been at the site scale across just a very few field seasons. A fundamental purpose of the GGD is to offer supplementary data (and metadata) through which to assess and incorporate spatial and temporal trends.

- To identify the drivers of variation: Clearly, an immediately beneficial spinoff from the ability to use aggregate data sets to isolate trends is that it becomes feasible to consider identifying the factors or processes which drive and constrain these trends. This is a powerful added facility in its own right, but is rendered doubly effective if (as seems likely) some of the systems are scale-dependent. Under such circumstances, sitescale analyses may provide a poor indication of the functioning of regional and global systems, and the use of data aggregation offers completely new investigative possibilities.

- Hypothesis testing: Since many potentially fruitful hypotheses are generated only through the analyses which follow the end of data collection in a study, immediate hypothesis testing and refinement are often rendered speculative. It is common practice to provide first order validation or refutation through reference to the published findings of other authors, but this is a poor substitute for reworking primary data to assess the robustness of the hypothesis. Carefully documented data sets made available (through GGD) for this purpose can bring new rigour to cold regions science, where project extensions and followup visits may be prohibitively expensive.

- Global typology: The outcome is a set of more carefully validated descriptive and explanatory statements which can be used to build a prototype global typology of attributes or processes. Such a perspective is not common in cold regions science at the present, and where genuinely global typologies do exist they are often deductively generated from speculative overviews (perhaps driven by coarse resolution remotelysensed data) rather than inductively driven from fieldscale studies. There is a very real role for both approaches, but the muchneeded fusion cannot take place without data set aggregation.

Problems and Resistance

It is apparent that even rudimentary success in building structures for global cooperation in data assembly and management is highly challenging. The technical hurdles are significant, but the political and economic barriers are still higher. Economics provides a useful metric for the process in offering the notion of Benefit-Cost Analysis.

This reminds the cold regions science and technology community that while it is readily possible to articulate a series of claimed benefits for the GGD, it is equally important to acknowledge the breadth and scale of the tangible and intangible costs. The analogy of a BenefitCost Analysis is also useful in focusing attention on the payback period. Many of the "costs" have to be carried during the years (or even decades) of the setup period, while the "benefits" may be reaped much later. The stark conclusion is that those who carry the costs may not individually be those who attract the greatest

Table 1. Evolution of the IPA Data Strategies

The Trondheim Objectives (1988)	The Oslo and Potsdam Workshop Imperatives (1994, 1995)	The Berlin Resolutions (1995)
Designate agreed indices of permafrost	To advance scientific understanding of permafrost	To encourage active and open participation in GGD
Increase awareness of data variability	To improve the basis of engineering design	To encourage interaction with WCRP-GCOS and GTOS, IPCC, IGBP and ITEX
Create guidelines for data documentation	To aid global/regional climate change studies	To prevent data loss, and encourage data protection and rescue
Establish preferred data model and structure	To detect environmental change at many scales	
Build inventories of data sets	To enhance environmental planning and protection	
Create open system-independent access to data		

benefits. Manifestly, it is a measure of the maturity of a professional community that it should confront these issues, and respond to them. Commitment to the GGD is a symbol of the IPA's commitment to its ongoing professional role, and to the continuing intellectual and practical importance of permafrost, especially its contribution to global change issues. Anything less would be an individual and corporate abrogation of duty.

Technical Hurdles

Nevertheless, when the problems of building a longterm global data strategy for permafrost are honestly confronted, it is clear why many individuals chose to turn aside. The first 10 years of the IPA Working Group on Data and Information have established the scale of the technical challenge. In the first place, it has been necessary to converge on an agreed data model (the consensus represented by Figure 1) and data structure (a working version of which is incorporated in the interim data protocols now in use for reporting and managing data sites within IPA) (see Barry et al., 1995, Appendix 2 and Appendix 4)).

Table 2. Priority assessment of key permafrost variables (Original Version from Frozen Ground, 18, 1995)

	APPLICATIONS				
PARAMETER	Process Understanding	Engineering Design	GCM Model Validation	Change Detection	Impact Evaluation
			GEOMETRY		
Permafrost extent	M	H	H	H	M
Permafrost thickness (shallow[1])	H	H	H	H	H
Permafrost thickness (deep[1])	M	M	L	H	L
Active layer thickness	H	H	M	H	H
Ground ice extent	H	H	M	M	H
Lateral & vertical displacement	H	H	L	H	H
			THERMAL STATE		
Temperature (shallow[1])	H	H	H	H	H
Temperature (deep[1])	M	M	M	H	L
			PROPERTIES (SOIL/ROCK)		
Moisture content (water - ice)	H	H	M	H	H
Bulk density	H	H	H	L	L
Texture	H	H	L	M	M
Chemistry (water - ice)	M	M	L	M	M
Trace gases	H	L	H	M	M

H M L = High, Moderate and Low Priority
1The demarcation between shallow and deep is taken as the local depth of zero annual amplitude.
Site descriptions include location, geology, geotechnical properties, vegetation etc. Metadata include techniques, equipment, precision, post-processing, data ownership etc.

Table 3 Current contents of GGD and CAPS (Circum-Arctic Permafrost System) CD-ROM

Major components
• The digital version of the IPA Circum-Arctic Permafrost Map
• A bibliography of permafrost literature spanning 1978 - 1997
• A glossary of frozen ground-related terms in twelve languages
• An index of more than 730 Russian permafrost maps
• Selected data sets of active layer and borehole temperature profiles, representative of important detailed time series from many of the member countries of the IPA, including soil temperature data from Russia and data developed under the Circumpolar Active Layer Monitoring (CALM) program.

Data quality is an even greater challenge. Internal consistency checks are much improved as data are drawn into the GGD, in practice through the professional services of World Data CenterA, Boulder, Colorado. But data collection quality remains the domain of the individual scientist or engineer. At first sight, this appears reasonable, but on reflection the data collection norms established by other professional groups that are as yet only just beginning to emerge within the permafrost community offer genuine attractions. For example, it is difficult to imagine regional or global climatic description and modelling being achievable without the comprehensive data collection standards assembled and promulgated through the World Meteorological Organisation and World Climate Research Programme's Global Climate Observing System (GCOS) (WMO, 1997). This is not an imposition on the freedom of expression of the individual researcher, but a means of linking those individuals into an effective professional community. The CALM network of permafrost monitoring sites supported by IPA is, perhaps, a vital first step in this direction, and moves are underway to identify this network as a component of the GCOS.

Data model, data structure and data quality all have the strong incentive that they serve the individual scientist or engineer as well as supporting the corporate activity of data aggregation. This is clearly not the case when one turns to the problems inherent in storing and disseminating data and information. Here, both the challenge and the solution have to be communal and a degree of compromise is essential. The established procedures and protocols of organisations such as the World Data Centers offer both a default standard and a practical infrastructure for achieving the desired functions. They are less obviously effective on their own as a mechanism for disseminating awareness, and it is not surprising that the IPA has sought to supplement this networking. In the first instance, its own World Wide Web site (http://www.geodata.soton.ac.uk/ipa) offers a potential gateway to the contents of the GGD for permafrost practitioners. For those many professionals in related disciplines and organisations for whom IPA would not be a selfevident source of information, gateways such as links through the UNEPGRID International Arctic Environmental Data Directory (ADD) may be more effective. It is fortunate that rapid development of Internet hypermedia linkages (see "cold links" on http://wwwnsidc.colorado.edu) makes it likely that the problems of identifying and acquiring suitable data from wellmanaged repositories will be reduced in the future. The overriding problem then will be the political one of getting data into the repositories in the first place.

Political Hurdles

Politics embodies the assumptions, values and infrastructures through which societies seek to organise themselves. On this basis, the GGD is an act of massive political significance. It is the manifestation of a set of objectives of increasing overt political significance as the permafrost community moved from Trondheim (1988), through Oslo (1994) and Potsdam (1995) to

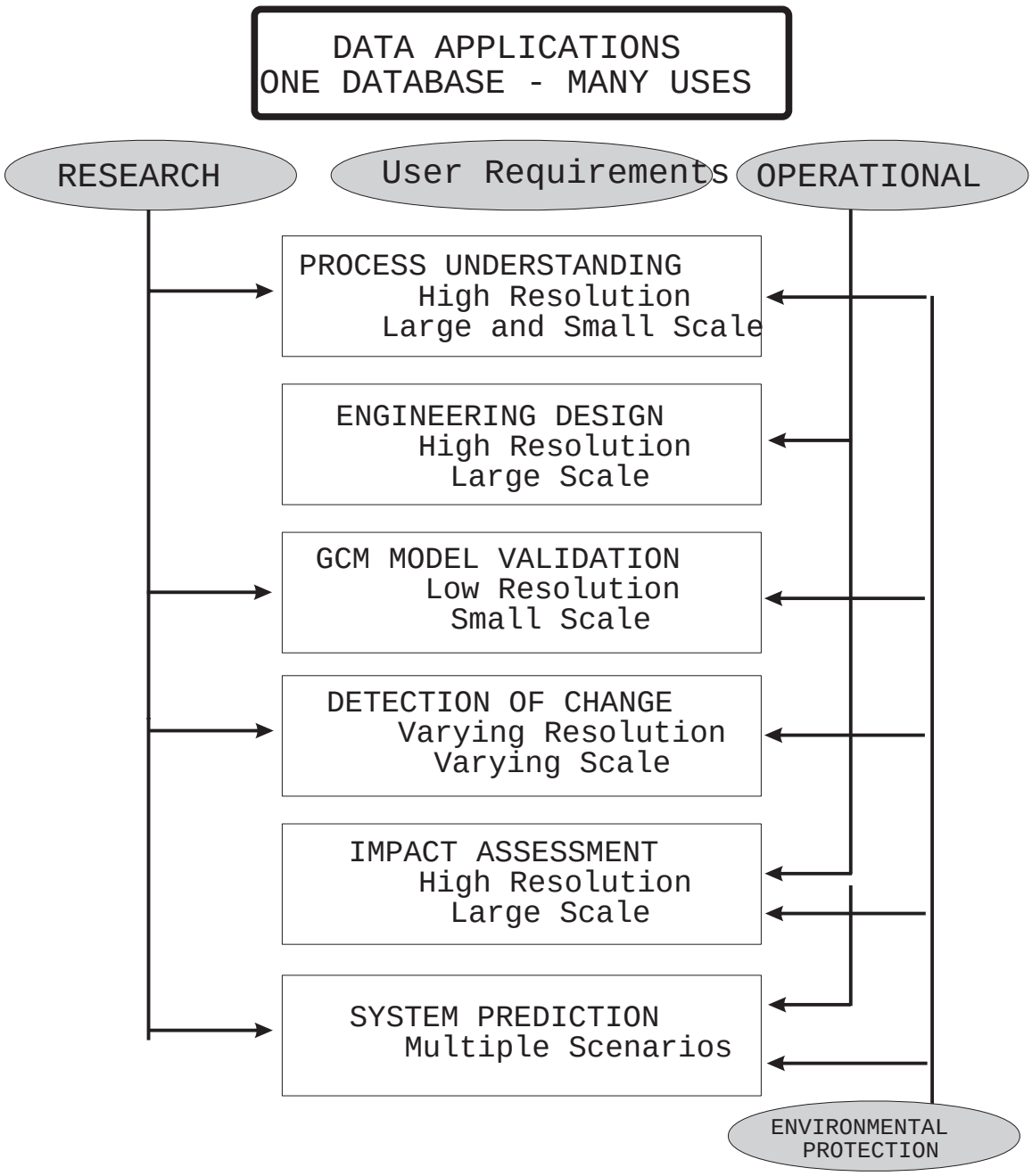


Figure 1. Possible application focuses for different GGD user sectors.

Berlin (1995). Along the way, its development has fostered substantial collaboration between permafrost data specialists in the USA, Russia, U.K. and China, for example, with National Geocryological Databases in Moscow, and at WDCA for Glaciology, Boulder, WDCD for Glaciology, Lanzhou and at the GeoData Institute, University of Southampton, U.K. A decision by the IPA Executive at Yellowknife (1998) to reorder the management of its datal and information role would constitute political recognition and a further advance in this process.

Most basic amongst the "political/professional" debates is the ongoing consideration of the rights and obligations of the individual against those of the professional community. Such issues lie at the heart of the future of GGD in particular, and of IPA data strategy in general. While every effort is made to ensure proper

accreditation to the author of a data set every time it is disseminated and subsequently reworked or republished, the act of depositing data centrally is nevertheless a personal sacrifice by the individual. The GGD is increasingly working to reduce this cost by assisting in the investment needed to prepare data for GGD, and to increase the benefit through accreditation and through returning to the individual the valueadding capability of access to related data.

Separate but equally difficult issues are those of data commodification and pending new intellectual property laws (Webster, 1997). Whether the tangible cost of data acquisition is carried by the individual, industry (commerciallydriven contracts) or society (research grants), the ensuing data do have value which renders them a commodity to be traded. Increasingly, this produces a conflict of interest. Understandable resourcing pressures ensure that the present commodity value of data sets may make it unattractive to consider gifting the same data to the GGD so as to reap the much greater longterm and communitywide benefits. Ultimately, the only immediate response is to use GGD as an awareness gateway which carries metadata through which to direct the "customer" to the data supplier. Much more robustly, GGD can serve as a longterm archive for the data, while dissemination rights rest with the data originator.

The very real progress made in acquiring data sets for GGD (Table 3) suggests that in at least some circumstances, the perceived professional Benefit-Cost ratio is now seen as being positive. This brings one step closer the ideal of creating a data democracy in which access to information drives an enhanced and accelerated development of cold regions science and engineering based on permafrost understanding. At the same time, it recognises the needs and rights of the individual data donor and attempts to meet them. There is a long way to go in turning these first encouraging signs into a new professional expectation for datasharing, but the foundation has been laid for a future data strategy for permafrost.

Future Scenarios

Who could have predicted the rise of the Internet and the World Wide Web at the inaugurating meeting of the IPA Data Working Group at Trondheim in 1988? Technoforecasting is notoriously unreliable, and the evaluation of professions is equally complex and unpredictable. Nevertheless, three trends appear sufficiently wellestablished to form part of the future scenario for IPA data strategy:

- Widearea longterm monitoring holds the key to the creation of standardised data now desperately required to build and validate spatial and temporal models and assist in efforts such as those of the IPCC and IGBP in studying cold region responses to global change. CALM is an excellent example, and may become a key permafrost contribution to programmes such as GTOS and GCOS.

- Communally organised exercises in data aggregation demonstrate the valueadding potential of the GGD and related strategies. An early and perhaps trendsetting example is the IPA Circumpolar Permafrost Map of the Northern Hemisphere (Brown et al.,1997). Despite all the inherent problems of merging nonstandardised data models and data structures, this map represents a major international success in establishing a baseline status report on globalscale permafrost. Future aggregation initiatives might focus on mapping individual permafrost indicators or synthetic indices, or might be directed towards making a wider and more robust contribution to models and to status reports such as that of the IPCC.

- Ultimately the key to IPA data strategy lies at a simpler but perhaps more far reaching level. It lies in turning data into information. "Information" implies an active process: that of informing. This in turn means that there is an identified user community with questions to be answered and operations or decisions to be underpinned. These potential users include resource managers, the educational community, northern peoples, and interested general public. The IPA has to find these people and highlight their needs. It means that what we transmit to these users has to be truly "informative". Key indicators (often individual variables such as active layer depth) and indices (often synthetic variables such as frost susceptibility) need to be specified and created. It means that we have to be proactive in building and publicising infrastructures for information and data dissemination. Problems of technology, rights, and commodity values will need to be overcome. If the will is there, they can be overcome.

Perspectives

All of this is a long way from the technology and sciencedriven thrust of the original 1988 Permafrost Data Workshop in Trondheim. The data strategy has evolved from a science focus to a user focus. Its priorities have evolved from being introvert (identifying and exploiting the needs and capabilities of the IPA membership) to being extrovert (seeking a formal and ongoing contribution to IUGS, IPCC ADD and WCRPGCOS/GTOS). Partly through its data and information activities, IPA has established a professional role and infrastructure, and has put permafrost back into the centre of the coldregions debate.

Acknowledgments

The following collaborators have been actively involved in the development of many of the above activities: J. Branson, J. Brown, D. Gilinchinsky, R. van Everdingen, C. Hanson, W. Haeberli, J.A. Heginbottom, M. Leibman, F.E. Nelson, N.N. Romanovskii, and A. Taylor, and their contributions are gratefully acknowledged.

Support of the GGID has been variously provided by the International Permafrost Association, U.S. Natural Resources Conservation Service, the National Science Foundation through NSFATM 9528007 and OPP9614557, as well as Russian agencies for the Moscow node, and the GeoData Institute, University of Southampton, UK.

References

Barry R.G. (1988). Permafrost data and information: status and needs. In Senneset, K.(ed.), *Proceedings 5th International Permafrost Conference, Trondheim*, Vol. 1, Tapir Publishers, Trondheim, pp 119-122.

Barry, R.G. (1995). Observing systems and data sets related to the cryosphere in Canada: A contribution to planning for the Global Climate Observing System. *AtmosphereOcean*, **33**, 771807.

Barry, R.G., Heginbottom, J.A. and Brown J. (1995). Workshop on Permafrost Data Rescue and Access. *Glaciological Data Report GD28*, WDCA for Glaciology, Boulder, CO. (133 pp).

Brown, J. Ferrians, O.L, Jr., Heginbottom, LA. and Melinikov, E. (1997*). Circumarctic Map of Permafrost and Ground Ice Conditions* (1:10 million). U.S. Geological Survey, CircumPacific Map Series, CP 45, Reston, Virginia.

Webster, F. (1997). Threat to full and open access to data. *Science International, Newsletter,* **65**, ICSU, Paris, 11 12.

World Meteorological Organization (1997). *GCOS/GTOS Plan for Terrestrial ClimateRelated Observations*. Version 2.0. GCOS32. WMO/TDNo. **796**, Geneva.