Data Rescue: experiences from the Alberta Hail Project

by B. Kochtubajda [1, 2]
C. Humphrey [2]
M. Johnson [3]

Abstract
A valuable meteorological data archive collected by the Alberta Research Council over the course of the Hail Studies Project in central Alberta is in jeopardy of becoming unusable as the digital data stored on magnetic tape degrade over time, and expertise in the data collection, calibration, and interpretation becomes scarce. The overall goal of this project was to preserve the digital radar, aircraft, upper air and surface precipitation data along with supporting calibrations and documentation; to transfer this archive to the University of Alberta; and to make the archive available to the scientific community.

There were three distinct operations carried out to ensure the long-term preservation of the archive; retreival of the digital data and all supporting (secondary) datasources; transfer of digital data from magnetic tape to compact disk; and the collection and preparation of relevant documentation describing the data. The archive will provide researchers with a documented dataset to support further research in radar meteorology, climate change, hydrology, cloud physics, mesoscale meteorology and severe weather phenomena.

1. Introduction
The acquisition of atmospheric data is an expensive endeavour, and the data are usually irreplaceable. The subsequent research uses of good data are often not contemplated by the original data collectors. For example, data can be re analyzed to test new hypotheses, or can be used for comparative analyses with other geographic areas. Data may become unusable when supporting documentation is lost or destroyed, or when the physical media on which these data are stored become no longer readable through degradation over time, or through the lack of equipment capable of reading the physical medium due to its obsolete format.

The Alberta Hail Studies Project (1956-1985) was established to study hailstorm physics and dynamics and to design and test means for suppressing hail. Central to these activities was the Alberta Research Council's (ARC) radar facility located at the Red Deer Industrial Airport in central Alberta (Figure 1). A vast amount of data was collected from several platforms to conduct research into precipitation mechanisms, severe storm development, hail suppression, hydrology and microwave propagation.

Since the termination of the Alberta Hail project in 1986, numerous research projects have demonstrated the value of using the Alberta data archive. During the period 1990-1994, 23 archive-based publications have appeared in refereed journals and conference proceedings and 4 scientific reports have been prepared. There have also been nine graduate theses (2 Ph.D., 7 MSc) awarded at 3 universities during this period. The areas of study have included radar meteorology, cloud physics, hydrology/hydrometeorology, computer science, instrumentation, and synoptic, dynamic and mesoscale meteorology. Scientific research and collaborations continue to this day.

Recognition that this valuable meteorological data archive was in jeopardy of becoming unusable as the digital data stored on magnetic tape was degrading over time, and expertise familiar with the data collection and calibration procedures, and their interpretation became scarce, prompted an effort to save this unique dataset.

2. Objectives
The specific objectives of this project were:

1. to transfer the computer-readable radar, aircraft, upper air and surface data from the existing short-term storage medium (magnetic tape) to an archival medium (CD ROM).

2. to collect all available supporting (secondary) data sources and develop the necessary documentation to describe the computer-readable data files.

3. to coordinate these efforts with the University Data Library and develop appropriate mechanisms to make the archive available to the scientific community.

3. Approach and Work Plan
There were three distinct operations carried out to ensure the long-term preservation of the archive; retrieval of the digital data and all supporting (secondary) data sources; transfer of digital data from magnetic tape to compact disk; and the collection and preparation of relevant documentation describing the data.

3.1 The Data Archive
The radar facility near Red Deer consists of a unique polarization-diversity S-band (10 cm) weather radar, a standard C-band (5 cm) weather radar, and an X-band (3 cm) radar used to track aircraft through a transponder

system. With the addition of computer interfaces in 1974, a systematic archive of radar data was initiated. This archive now includes close to 200 Gb of data, representing approximately 12, 000 hours of multi-parameter radar data. In addition to the radar archive, an extensive archive of aircraft, surface precipitation, and upper air data has also been collected. Approximately 18 Gb of data were recorded between 1983 and 1985 aboard an instrumented research aircraft flying through convective storms and cumulus clouds. Also, quantitative precipitation reports (hail and rain) were obtained from approximately 500 ground stations within the radar coverage, between 1974 and 1985 and in 1989. These data exist on several media, including 800, 1600, 6250 bpi magnetic tape and 8 mm cassettes. Comprehensive radar, aircraft and upper-air software packages are also available for data analysis and display.

3.2 Retrieval of the digital data and all supporting data sources

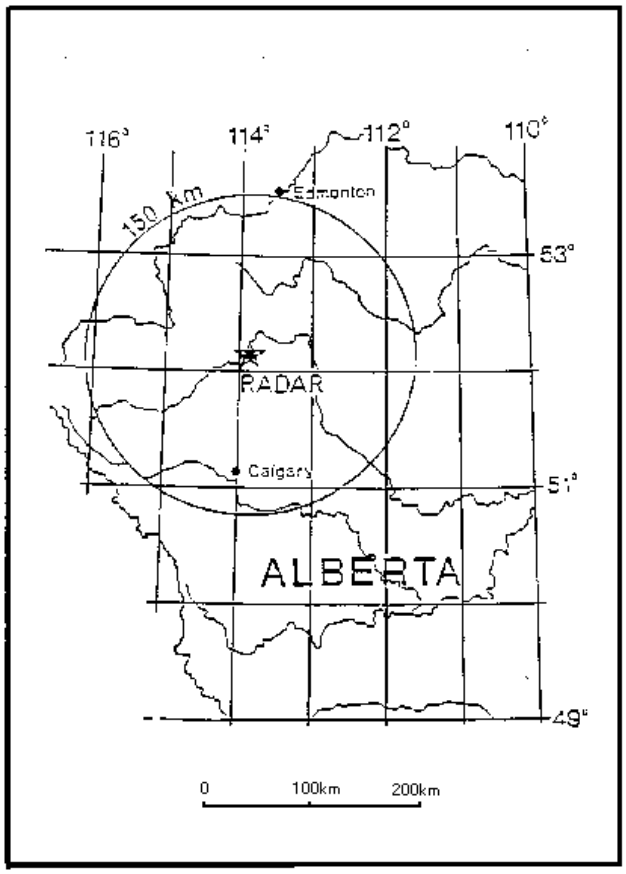
A number of assumptions were made before the digital data were recovered. To provide the broadest range of research potential of the data, unprocessed aircraft and radar data would be provided. This would yield a quicker retrieval of the data and (given the time and budgetary constraints of the project) would result in more of the data being recovered. To maximize the research potential, we would archive all data types and work backward by year, from 1985 towards 1974, thus ensuring a complete multi-platform data set.

3.2.1 Aircraft data

Physical experiments designed to explore the potential of hailfall suppression, and rain augmentation through airborne glaciogenic seeding on convective cells were conducted in central Alberta between 1983-1985, as described in Humphries et al (1986). These studies emphasized in-situ aircraft measurements to investigate natural and artificially modified precipitation processes. The primary observational platform used in these studies was the Intera/ Alberta Research Council cloud physics instrumented research aircraft, a Cessna 441 Conquest, pressurized twin-engine turboprop aircraft. Data from the instruments were managed by a computer based data system which provided data acquisition, recording, and real-time calculations and display (Johnson et al, 1987).

Approximately 300 magnetic tapes were processed for all the research flights conducted from 1983 to 1985. Data were segmented into unique files for each hour of the day. The file names follow the ISO 9660 level 1 standard, and are of the form YYMMDDHH.ADB where YY - year; MM - month (numeric to help in sorting); DD - day; HH - hour of the day; and ADB represents the file extension identifier for Aircraft Data Block. Aircraft data were recorded with Coordinated Universal Time. Yearly index files summarizing the amount of information collected for each hour of the research flight were prepared. The first line provides a brief description of the purpose of each research flight, including the date, start and end time of data collection (hh:mm UTC) and the type of study carried out. The subsequent lines describe the filename, file size (in KBytes and MBytes) as well as the number of records collected (including the 2-D imagery).

figure 1



Location at the Alberta Research Council's weather radar facility and coverage area.

3.2.2 Polarization radar data

The S-band polarization-diversity radar, installed in 1967, operates at 2.88 GHz and has a parabolic reflector antenna with a 6.67 m diameter dish that produces a 1.15_ beamwidth in both azimuth and elevation. The radar sweeps out a helical volume scan rotating at 48_s^{-1} (7.5 s per revolution) and rising 1_ in elevation for every 360_ in azimuth, up to a maximum elevation of 8 or 20_ selected depending on the proximity of the storms to the radar. The radar records data with an approximate azimuthal resolution of 1_, for 147 range gates, from 3 km to a distance of 157 km from the radar, with a range resolution of 1.05 km per range gate.

The radar can transmit any polarization, but has usually transmitted left-hand circular (LHC) polarization with 450 kW peak power. The receiver circuitry digitally records four measurements from the

Table 1: Summary of supporting data sources for 1983 and 1984.

Year	Photo/Maps	Shelf Loc.	Data Timelines	Summaries	Shelf Loc.	Operational Logs	Shelf Loc	Checks	Shelf Loc	Calibrations Data	Shelf Loc.		
1983	Maxmaps		F6 G6	Comp. tape	A6B	Vehicle logs	A5B	Daily oper	A6B	Computer calibs			E5B
	Trans maintenance chart		E5B	Daily timeline	A6B	Daily journals	A5B	Midnight checklist	A6B	Major calibs			A6B
	Computer calib plots		A6B	Radar summaries	E5B F6 G6 A7B B7B	Daily appointments	A5B						
	Cloud photos		D3B	Wx trend sheets	F6 G6	Radar logs	A6B						
				R/A data	C2B D2B E2B	Tape logs	E5B						
						C/S band comparisons	D2A						
1984	Maxmaps		C5B F6 G6	Daily timeline	C5B	Daily appointments	A5B	Transponder checks	B5B	Antenna calibs			B5B
	AnVI disc		C5B	Wx trend sheets	F6 G6	Computer Radar	B5B	Midnight checklist	B5B	Receiver calibs G2A			B5B
	Cloud photos		D3B	Video flight	F6A	Tape logs	C5B	Receiver checks	B5B	Computer calibs			B5B
	Snow flight photos		E3B	Radar const.	B5B	Radar logs	C5B	Daily oper.	B5B	Major calibs			E5B
				R/A data summaries	F2B G2B A3B			Transmitter checks	C5B				
1985	Computer plots		D5B	Daily timeline	C5B	Appointment diary	A5B	Transmitter checks	C5B	Major calibs			D5B
	CAPPI 'S	F6 G6	Video flight sums	F6B	Daily journal	A5B	Midnight checklist	D5B					
	Major con graphs		D5B	R/A data summaries	B3B	Vehicle logs	A5B	Transponder checks	D5B				
	Cloud photos		D3B E3B			Tape logs	C5B	Receiver checks	D5B				
	CSU photos		E3B			Radar logs	C5B F2A	Daily Oper checks	D5B				
	Snow trip photos		E3B			Radar calib oper.	D5B						

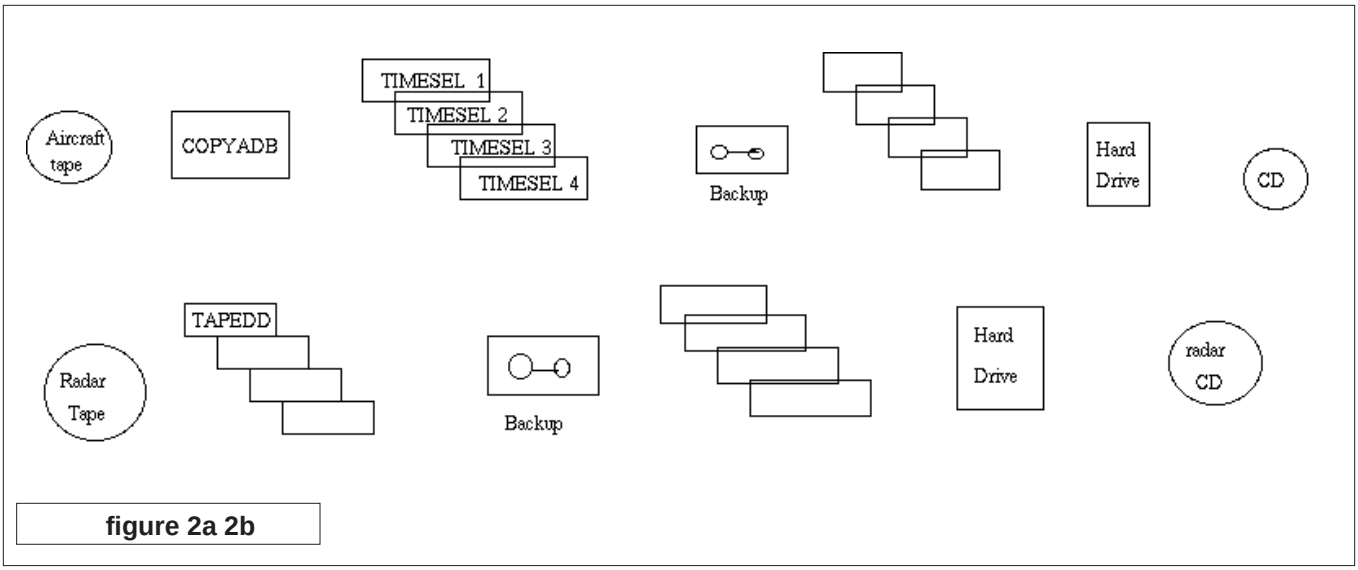


figure 2a 2b

LHC and RHC components from each range bin. These are the RHC co-polar signal power, the LHC cross-polar signal power, and the correlation and phase between the LHC and RHC signals.

Approximately 240 S-band radar tapes (previously copied to 6250bpi) have been examined for the period 1980 to 1985. Radar data were extracted into files containing one complete 3D volume scan, representing either 1.5 minutes for the 8_ scan, or 3 minutes for the 20_ scan. A quality control report was produced with each data file containing information about the azimuth and elevation time histories of the volume scan. The file naming convention adopted for the radar data uses all 11 characters (YYMMDDHH.MMR) where YY - year ; MM - month (numeric to help in sorting); DD - day; HH - hour of day (24 hour clock); MM - minute of first data in file; and R - represents the data type (S for S-band, C for C-band, Q for quality control reports). Radar data were recorded with local (Mountain Daylight) time.

Table 2: Summary of Data Compact Disks

Data Type	Year	No. Files	No. CDs	Total MB
Aircraft				
	1983	349	5	2947.5
	1984	226	3	1706.8
	1985	166	2	1284.8
Radar				
	1980	13	3	1433.1
	1981	9	1	467.9
	1982	20	3	1657.8
	1983	48	8	3681.5
	1984	56	9	4775.2
	1985	51	11	5093.3

3.2.3 Surface hailfall and rainfall data

Volunteer observations of precipitation events were used to supplement quantitative surface measurements obtained by specially equipped vehicles, which were directed beneath thunderstorms. After each storm, telephone surveys were conducted to collect hail and rain reports. Report cards were also received by mail from volunteer farmers. This information was useful to develop hail climatologies and correlate hailstone characteristics with crop damage.

The surface data collection includes the digital hail and rain report files (YYHAIL.DAT, YYRAIN.DAT), from the telephone surveys for the period 1957 to 1985 (except those files missing from 70-73); selected time-resolved hail and rain truck observations (YYMOBILE.DAT); and daily precipitation measurements collected by approximately 500 volunteer farmers during the months of June, July and August for the period 1975 to 1983

(YYPRECIP.DAT).

3.2.4 Upper air data (LIMEX)

A mesoscale upper-air study, *Li*mestone *Mo*untain *Ex*periment (LIMEX-85) was carried out over the foothills and mountains of southwestern Alberta during July, 1985 (Strong, 1989). The objectives of the field experiment were focused on mesoscale convective processes, orographic effects, and interactions with synoptic processes, with particular emphasis on severe storm forecasting applications.

The archive data includes two-hour soundings from nine upper-air sites with an average spacing of 50 km, continuous SODAR profiles, research aircraft soundings at 20-km intervals, surface data from eight automated systems, and an extensive cloud photo set. The compressed upper air data and accompanying analysis software are currently archived on 2 high density diskettes.

3.2.5 Supporting data sources

An equally important component of the retrieval process was the collection of the secondary (supporting) data sources including aircraft mission scientist notes, radar plot summaries, various operational log books, checklists, calibration notes, cloud photographs and video tapes. The data were boxed and transferred from the Alberta Research Council to the meteorology division at the University of Alberta. Subsequently, the boxes were itemized and given a location identifier. The contents of each box were stratified into one of five categories (photos and maps; data summaries and timelines; operational logs, checklists; and calibrations and data. A listing has been prepared which stratifies the data according to the type of information and the year of collection. A subset of the listing (for 1983-1984) reproduced in Table 1, illustrates the variety and richness of the materials available.

3.3 Transfer of digital data from magnetic tape to compact disk

The procedures used to produce the aircraft and radar data compact disks are depicted in Figures 2a and 2b. A series of programs were used to copy unprocessed aircraft data from tape to file (COPYADB), and to generate hourly flight files (TIMESEL). These hourly files were backed up on a series of 8mm EXABYTE data tapes and high capacity SONY compact tapes. Complete 3D volume scan data files and associated quality control reports were generated from the radar tapes (TAPEDD) and also backed up on EXABYTE and SONY tapes.

Compact disks were produced using a Pinnacle Micro RCD-1000 recordable compact disk recorder with Macintosh authoring software and a Quantum 1 GB Fast Scuzzi 3 hard drive for preparing a CD image. The requirements on the hard drive included an average seek time of 12 milliseconds or faster, a transfer rate of 1.2 MB per second or better, and an intelligent calibration feature (ie: thermal recalibration not performed during a continuous read) to avoid a write interruption which would render the CD invalid. The ISO 9660 level 1 standard was selected as the format for the CD-ROMs. This standard allows the same CD to be read and interpreted on Mac, MS-DOS, UNIX, VAX/VMS, and other computer platforms. This includes the restriction of file names to 8 characters, with a 3 character “extension”.

An inventory of the compact disks produced is summarized in Table 2. There are 46 CDs in total, including 10 research aircraft data CDs ; 35 CDs containing the S-band polarization data from 197 days between 1980 and 1985; and 1 CD containing the surface precipitation files, dataset documentations, and the radar and aircraft software source codes. Each aircraft CD contains a series of sequential hourly data files and 3 text folders (MAC, DOS, and UNIX) containing the file summaries and a disclaimer. The directory structure of a radar CD is described in Figure 3. A radar CD contains a series of daily files (YYMMDD). Each file contains 4 sub-directories (DATA, QUALITY, CALIB, LOGS). The DATA sub-directory contains the sequence of radar volume scans. The quality control reports for each radar file are located in the QUALITY sub-directory. The calibration text files and the daily radar and transmitter log files are found in the CALIB and LOGS sub-directories, respectively.

3.4 Documentation describing the data

A series of documents have been gathered and/or prepared to describe the various datasets. Aircraft experiment descriptions including study objectives; flight procedures; aircraft description and instrumentation list; 2D image processing; aircraft tag descriptions; daily flight assessments (instrument evaluations); and sensor calibration files accompany the digital aircraft files.

Digital radar and transmitter logs for the period 1977-1985; as well as descriptions of the radar characteristics and scan protocols; data structures; and calibration files accompany the digital radar files. The primary documentation for the surface hail and rain dataset is a coding sheet describing the file format. The daily farmer precipitation files from 1975-1983 has an accompanying text file.

An extensive bibliography of Hail Project related papers has been compiled. The original hard copies are currently stored in the meteorology division at the University of Alberta. Software packages to analyze and display radar, and upper air data, developed at University of Essex and at AES in Saskatoon, have been obtained and can be shared by users. A summary of the archive as it is currently configured is presented in Appendix 1.

4. Conclusions

Potential users of the archive have indicated that the radar and ground measurements dataset would be used to continue severe hailstorm and rainstorm studies; to provide input to

distributed hydrologic models; to carry out radar-based precipitation climatology studies; and to validate numerical models being developed during MAGS, BOREAS, GCIP, or BASE. The aircraft archive would be used to improve our understanding of the chemical composition of cloud water and the processes which affect it, and for icing research. The LIMEX upper air dataset would be used for the atmospheric correction of NOAA-AVHRR data in estimating regional evaporation, as well as in moisture budget estimates and evaluation of evapotranspiration studies.

A documented archive of radar, aircraft, surface and upper air data has been provided from which further research in these areas can be carried out. The retention and preservation of the archives through the University of Alberta will ensure the continued accessibility and long-term survivability of these datasets.

Acknowledgements
This project was financially sponsored by the University of Alberta, Alberta Research Council, and the Atmospheric Environment Service. Mr. S. Kozak and Mr. F. Bergwall assisted in the data retrieval.

The authors also acknowledge the contributions and support of Drs. EP Lozowski, GW Reuter, and T Gan (UoA), Dr. A. Holt (UoEssex), Dr. DR Rogers (Colorado State Univ), Drs. GA Isaac, P. Joe, GS Strong, (AES), and Dr. BL Barge, and Mr. CF Richmond (ARC).

References

Humphries, R.G., M. English, and J.H. Renick, 1986: Weather modification research in Alberta Canada. 10th Conf. Planned and inadvertent weather modification, Arlington, AMS, 357-361.

Johnson, M.R., L.E. Lilie, and B. Kochtubajda, 1987: A data structure for acquisition, analysis, and display of meteorological data. 6th AMS Symp. on Met. Obs. and Instr., New Orleans, AMS, 397-400.

Strong, G.S., 1989: LIMEX-85: 1. Processing of data sets from an Alberta Mesoscale Upper-air Experiment. Climatological Bulletin, 23, 98-118.

1 Paper presented at the IASSIST confernce

2 University of Alberta Edmonton, Alberta 3 Alberta Research Council Edmonton, Alberta

APPENDIX 1: ARCHIVED DATABASE SUMMARY

Data type	Period	Archive filename	Calibration	Documentation
Digital S-band radar	1980-1985	YYMMDDHH.MMR	(n = 1-4)	radar data structure
		YYMMDDnA.TXT		radar characteristics
		RLYYMMDD.TXT		calibration procedure
		TLYYMMDD.TXT	(n = 1-2)	
		YYMMDDnX.TXT		
Digital aircraft data	1983-1985	YYMMDDHH.ADB		Expt desc data index aircraft + instruments daily scores Video logs
surface reports: 1957-1985 hail and rain		YYHAIL.DAT YYRAIN.DAT		hailcard coding form
daily rainfall reports: 500 stations	1975-1983 (June 1 - Sept 1)	YYPRECIP.DAT		file description
mobile reports hail and rain		YYMOBILE.DAT		file description
Upper air data LIMEX-85 9 upper-air stns. 8 auto surface stations	(July 4 - 23, 1985)			
Analysis Software: radar / aircraft / upper air				

Data Rescue: experiences from the Alberta Hail Project

ADDENDUM

The Earth and Atmospheric Sciences Department in collaboration with the University of Alberta Data Library, and the Information Systems Department of the Alberta Research Council have just completed a 15 month effort to rescue the Alberta Hail Studies Project dataset. The project included the organization, retrieval, and formatting of the digital data and all supporting (secondary) data sources; the transfer of digital data from magnetic tape to compact disk; and the collection and preparation of relevant documentation describing the data.

There are 62 CDs in total, including 10 CDs of research aircraft data collected between 1983-1985; 47 CDs containing the S-band polarization data from 287 days from 1979 to 1985, and in 1989 and 1991; 4 CDs of coincident C-band radar data collected on those days when both radars were operating simultaneously (44 days) between 1979 and 1991; and 1 CD containing the surface precipitation files, aircraft transponder files, dataset documentation, and the radar and aircraft software source code. The archive also includes the collection of supporting data (such as; operational log books, manuals, photographs, slides and videos) and 2 diskettes of upper air data and accompanying analysis software.

Example access software (in the C programming language) and documentation has been developed for quick inspection of the original unprocessed aircraft and radar files from the CDs, and as a demonstration of data access.

A set of World Wide Web pages has been developed and is now available on theInternet via browsers such as Netscape and Mosaic. The “Alberta Hail Project Meteorological and Barge-Humphries Radar Archive” can be accessed through network services provided by the Data Library at the University of Alberta, by opening the URL: http:// datalib.library.ualberta.ca/AHParchive/

Data can be accessed in one of three ways. Researchers can obtain hail and rain data files directly from an anonymous FTP site: (ftp://datalib.library.ualberta.ca/AHParchive).

To obtain aircraft and/or radar data from the compact disks, click on the ORDER FORM and submit a specific request. For small amounts of aircraft and/or radar data (e.g.: a single case study), the set of hourly aircraft files or the daily radar directory will be transferred from the CD library and placed in the anonymous FTP site for subsequent retrieval. Requests for larger amounts of data will result in the production of customized CDs and shipment to the researcher for minimal cost.

Use of the Archive, is subject to the following conditions:

1. These data are to be made freely available only to the scientific research community, whether national or international.

2. These data are provided for the exclusive purposes of teaching, academic research and publishing, and/or planning of educational services and may not be used for any other purposes without the explicit written approval, in advance, of the Data Library at the University of Alberta.

3. The Alberta Research Council, the Atmospheric Environment Service and the University of Alberta will be acknowledged in any anticipated presentations and papers associated with the ARCHIVE.

4. The citation to be used for the ARCHIVE is: Alberta Research Council. The Alberta Hail Project Meteorological and Barge-Humphries Radar Archive: [computer files], Edmonton, Alberta, CANADA. Alberta Research Council [producer], University of Alberta Data Library [distributor]. August 1995.

<URL: http://datalib.library.ualberta.ca/AHParchive> <URL: ftp://datalib.library.ualberta.ca/AHParchive

This project was financially sponsored by the University of Alberta, Alberta Research Council, and the Atmospheric Environment Service.Thank you for your support of this project.

Bob Kochtubajda Ed Lozowski Chuck Humphrey Steve Kozak Mark Johnson Ford Bergwall