Data Management Plan

Data generated by ICECHIP will consist of four main groups: (1) field observations, (2) derived products synthesizing multiple observations, (3) outcome products, and (4) all other research products. Data management for each is described below.

1. Field observations. ICECHIP will use the EOL (Earth Observing Laboratory) Field Catalog customized to meet project needs. The EOL Field Catalog is a web-based central repository of project planning documents, IOP reports, facility status updates, field data images, satellite and model products and other information that are invaluable for field operations and as a post-project reference. The catalog will help the project document activities in near real-time, provide a single point for updating field campaign status, and provide a repository for preliminary in-field research data products. The Field Catalog will be further used following the field phase to assist in data analysis, as well as for providing a long-term archive of the project and an online portal for interested members of the community. ICECHIP participants will also work with EOL to prepare and test web-based forms that will provide the basis of in-field documentation. Examples include daily operations summaries, daily facility status reports, expendable resources status reports, and daily weather forecasts.

All field observations to be collected during EOPs 1 and 2 are listed as datasets O1 (airborne), O2 (remote sensing), O3 (ground-based), O4 (ground-based precipitation), and O5 (laboratory) in the Scientific Traceability Matrix (Table 1 below). All instrument PIs (including non-funded PIs from IBHS, ABoM, and NHP) will contribute their observations to the EOL Field Catalog following EOL guidelines for quality control and time allotted for processing (typically 6 months). In addition, ICECHIP PIs will work with EOL to survey instrument PIs and prepare an explicit set of format, storage, and metadata requirements for all generated data prior to field operations. Data external to the campaign (e.g., operational weather satellites, O2.5-6; radar, O2.7; NWS soundings; surface observations; and model data) will also be integrated. Level 2 data management support is requested from EOL given the large amount (100s of TB) and diverse types of data that will be collected. The entire ICECHIP dataset will be available only to ICECHIP PIs for 1 year after the conclusion of each EOP. After that year, each EOP's ICECHIP dataset will be open to the general scientific community.

2. Derived products synthesizing observations. Three derived products, listed as datasets O6 in Table 1, will be uploaded to either the EOP Field Catalog or the Geoscience Data Exchange for NSF (GDEX), in consultation with NSF program managers and EOL staff, with appropriate documentation in scientifically accepted formats (e.g., netCDF) no later than 1 year after completion of the project. These include 4-D wind and moisture retrieved from radars (O6.1), geolocated and temporally defined hail swath characteristics incorporating all surface hail observations (O6.2), and a spatially targeted EnKF analysis of environmental conditions during each IOP (O6.3). Generation of a digitized, synthesized hail swath dataset is of particular importance for distribution to the scientific community at large, given the necessary but very basic nature of some of the hail observations (e.g., hailpads, hand measurements or pictures of hailstones). These datasets will be used frequently among ICECHIP PIs and the full scientific community upon release. By providing already synthesized data, ICECHIP proposes to reduce the time other researchers need to produce significant new science with the observations. The software used to make the O6 datasets will be archived as described in Section 4.

3. Outcome products. ICECHIP PIs have identified 7 specific outcome products that will be produced by the proposal. These are listed as products O7 in Table 1. Products O7.1-7.6 are all software-based and will be uploaded to a public ICECHIP Github repository, with appropriate documentation and use cases, no later than 1 year following the conclusion of the project. Three of the products have immediate operational use (O7.2, O7.3, O7.6) and will be provided directly to NOAA via existing PI contacts. Product O7.7 is a public web portal that will include interactive online mapping tools, graphics and statistics, and future hail disaster scenario narratives, and will not be subject to the 1-year data embargo. It is designed to provide hail impact information gleaned from the ICECHIP project (H4.5, 4.6) to stakeholders and end-users, laying the groundwork for efficient policy recommendations and changes.

Table 1: Science Trace	ability Matr	Image Theme 1 Theme 3 Theme 4 Theme 5	
PI(s)	Institution	Observations 00 htt 1 ht 2 ht 3 ht 4 ht 5 ht 5 ht 2 ht 2 ht 2 ht 2 ht 2 ht 3 ht 2 ht 3 ht 2 ht 3 ht 3	
		Airborne assets	Color Legend
Bailev	LAOF/UM UMich	King Atrinsitue prostatorica conceptional out.1 x d d d x d x x r d d d x d x x x y x d d d d x d x x x x	CMU
French	LAOF/UW	KingAirfordsondes 013 x x x x x x x x x x x x x x x x x x x	Purdue
French	WN MOL	King Air Ka-Band Profiling Radar (KR-2) 014 x x x d d d x x x x d d x x x x x x x	AER
Argrow		CILBAANINUIATCIAITERESE ULJ. X Y Y Y Y Y A A Y X X Y Y Y Y Y Y Y Y Y	DSLI
Argrow	5 3	CURANTIVISTIVIED DUTY / 2 X X X X X X X X X X X X X X X X X X	2 2
Soderholm, Giammanco	ABoM, IBHS	Aerialimagevy forces 018 x	NIN
Soderholm	ABON I AOF/UW	Kina Surdenta Trajector Protesystem 01.9 x x d d d x x x x x x x x x x x x x x	UNL IBHS
		Remote sensing assets	Villanova
LAOF	NSF	NCAR5Pol (Year 1) 02:1 × × × × × × × × × ×	CSU
Chandrasekar	CSU	CHILLS-andX-band (Year 1) Q2.2 x x x x x x x d d x x	O
Kosiba, Wurman	DIID	2 mobile X-band, 1 mobile C-band DOWs 023 x x x x x x x x x d d d x x x x x x x	UIUC
Kosiba, Wurman		Như cũ cho transmission cuy hành cuy transmission cuy	UMich
NASA	NASA	OCOS-MAIL TANK POST-STOTI POST-STOTI TANK POST-STOTI POS	Symbol Legend
NOAA	NOAA	NWS WSR-88D5 in MESOSAILS mode upon request 02.7 X X X X X X X X X X X X X X X X X X X	x: Observation needed
	5	montectant anticome	to test hypothesis
Gancini Allan	NILL CMIT		v*: Only denloyed in
Gensini Allen Keeler	NILL CML	2) Windsonteevoluming yettiin yoosa ya	 Computed in the second sec second second sec
Friedrich	no	Upstreamsourcements are an array of a constrained and array of a constrained array of a constrained array of array of a constrained array	d: Hypothesis
Friedrich	ŋ	12-chamel microwaveradiometer 03.4 A x x x x x x x x x x x x x x x x x x	dependent upon
Keeler	CML	Mobilemesonet and UAS command center 03.5 x x m m m x m m d m	another that uses
Kosiba, Wurman	NIN	Cound hered acrete infractional acrete and acrete acre	these observations
Allen	CMI		broduct
Adams-Selin	AER, NHP	4 custom hall collection/video systems 042 x x x x x x x x d d x x x d x x x d x x x d x	m: UAS management
Friedrich, Giammanco	CU, IBHS	5(2.CU.31BHS)mobile-optical distrometers 04.3 X x x x x x x x x x x x x x x x x x x	
Dawson	Purdue	45 Portable In-Situ Precipitation Stations [PIPS] 04.4 x x x x x x x x x x x x x x x x x x x	
Kosiba, Wurman		MKK.CEMCIATAMENEIOROSIGISTOMERES 04.5 X X X X X X X X X X X X X X X X X X X	
Hevmsfield	NCAR	www.efvingerationic constraints with a start of a start	
Giammanco	IBHS	24 impact distormeters with GoProcements O438 x x x x x x x x d d x x x x x x x x x	
Schumachar		CoCoRAHS hallpad network; solicited hallpad netw	
Giammanco	IBHS	3.3.1.3extramest Sutting Variants of the state of the sta	
Giammanco	IBHS	3 3 portable compressive strength terminary systems (2001.11) × × × × × × × × × × × × × × × × × ×	
Adams-Selin	AER	Post-storm collection team, coolers G4.12 x x x x x x x x x d d x x x x d x x x d x x x x d x	
		Laboratory-based assets	
McMechan	INN	Hall machine, Cropt test plots 05.1 x d d	
Heymsfield	NCAN	Coll BackGroning 05.31 x x x x x x x x x x x x x x x x x x x	
Dodd	NIN	Water isotopic laboratory 05.4 X X A A A A A A A A X X X A A Y X X X X	
Heymsfield	NCAN	Unit Mainzeventical Wind Market and Market	
	10	Devind Accountional and the second seco	
Adams-Selin	AER	Determine observationant products. After multi-Donobler wind field analysis with DLA-estimated [06,1] × × × × × × × × × ×	
Allen	CMU	Geolocated synthesized hail swath characteristic database 66.3 x x x x x x x x x x	
Zhang	PSU	Spatially targeted EnKF analyses [06.3] x x x x x x x x x x x x x x x x x x x	
		Output products H11 H12 H13 H13 H14 H15 H16 H21 H22 H23 H24 H31 H32 H33 H34 H35 H41 H42 H43 H44 H45 H45 H51 H52 H53 H54	
Strader, Ashley	Villanova, NIU	Improved HailMC Monte Carlo hail damage impact model 07.1 0 0	
Mansell, Dawson	OU, Purdue	Improved NSSL (M20) bulk and bin microphysics 07.2 0 o o o o o o o o o o o o o o o o o o	
Adams-Selin	AER	Improved 3D-HALLOCKTH Inforcestingsystem 07.3 0 0 0 0 0 0 0 0 0 0 0 0	
Kumijan. Lombardo	PSU	Improved ASII hait rajectory model 07.4 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	
Homeyer, Kumjian	OU, PSU	Radar-based hall size concertration agorthmics of a base	
Strader, Ashley	Villanova, NIU	Online (CECHP portal for public and private policy makers 07.7 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	
Gensini	NIN	Suite of stochastic microphysical parameter perturbations 07.8 o o o o o o o	

4. Other research outputs and datasets. Finally, all scientific research planned as part of ICECHIP will produce many observational analyses, model configurations, simulation data, and analysis codes that do not fit into the first three categories discussed above. During the project, all data will be stored on individual institution computing resources as detailed in each institution's Facilities and Equipment Statement, or on NSF-provided supercomputing resources such as Cheyenne or Derecho. Following the completion of the project, all model code, initial conditions, selected model outputs, and analysis software not included as part of the first three categories will be archived following the best practices outlined in Mullendore et al. (2021) to ensure scientific reproducibility. All ICECHIP publications will inform the community of the data's availability and provide access instructions. Archiving practices will vary by institution and are briefly detailed below.

AER personnel will retain the described products in Amazon Web Services (AWS) S3 Glacier Deep Archive storage for at least three years after the project. Funding for the cost of this storage is supported through AER's indirect computer overhead costs. AWS S3 access to interested researchers can easily be provided via creation of a customized URL specific to the requested data. NCAR processing software SODA (github.com/abansemer/soda) is already publicly available. NIU model simulations and analyses will be conducted via NIU's Campus Center for Research Computing and Data (CRCD). CRCD will archive all simulation results in RAID-5 redundancy in netCDF or GRIB2 format. Analysis code will be developed in Python and deployed in an open-source GitHub workflow. CMU analyses will be conducted using the local server cluster, the computers purchased for graduate students and an allocation through Michigan State University's Institute for Cyber-Enabled Research. Project data for long-term storage will be stored on archival grade storage through the CMU tape system. Output formats will be JPG, csv or ascii text for the raw observational data, KMZ for the raw analyses, and gridded netCDF for all synthesized observational data. Software in python/fortran/R will be shared through Github. UI will make two archive copies of data which will be maintained in the UI FARM permanent data archive, and data will be provided to NCAR following the CIF guidelines. UI FARM facility data will be available via FTP from the UI FARM permanent data archive. CU Boulder will archive data and products at the CU PetaLibrary, which is a CU Research Computing service that supports the storage, archival, and sharing of research data. Funding for data storage is listed in the budget. Purdue will archive model simulation configurations, simulation output, and associated analyses on the Research Data Depot (spinning disk) and Fortress (tape) archival systems which are supported by Purdue's Rosen Center for Advanced Computing. Funding for storage on the Research Data Depot is listed in the budget and details of these storage services can be found in Purdue's Facilities and Equipment document. Analysis software will be maintained at Purdue's Github Enterprise server (github.itap.purdue.edu). Villanova University will archive all HailMC model simulation results via external SSD (Raid 1) and back up using PI-Strader's NAS. Data such as NOAA hail events, exposure cost surfaces (i.e., land use-land cover, land value, housing units, building footprint, etc.) are publicly available and devoid of any information that can be linked to individual persons. Analysis software utilized in this portion of the project includes python, ArcGIS Pro, SPS. PSU will archive relevant data and analyses performed by Penn State science PIs. Two university-sponsored publicly available archival systems will be used: DataCommons, which houses the scientific digital research data, and ScholarSphere, which archives finished research including publications, presentations, figures, etc. Public access to CSU's data/analysis will be through secure servers using software development platforms like GitHub and DOE servers. To preserve IP rights of project software, supporting tools will be provided with terms of use and disclaimers consistent with pre-existing software provisions. All gridded NEXRAD radar data, model output, and related computing code generated at OU will be made publicly available through OU & Regional Research Store (OURRstore), a large-scale tape archive maintained by the OU Supercomputing Center for Education & Research (OSCER). No usage charges are required. All data products produced will be stored in netCDF. Diagnostic tools and code will be made available in IDL/Python. Data collected from the simulated hail machine field experiments will be stored on the UNL OneDrive with a DUO secured server. This data will include but is not limited to field operations, crop characteristics, various plant injury assessments, UAS data, time-lapse images, and metrics from the hail machine not already provided to the Field Catalog. Standard space for users starts at 1TB per user, which can be increased to 25TB per user upon request. The

Deep Blue Data repository offered by the **University of Michigan** Library provides access and preservation services for digital research data developed or used in support of research activities. The U-M Library will provide stewardship and curation of the data in the repository and consult with U-M project participants to facilitate the preparation of data for archiving. Data sets will be assigned Digital Object Identifiers (DOIs), which will serve as persistent identifiers.