

Canadian Ice Island Drift, Deterioration and Detection (CI2D3) Database

Database Documentation v1.0

Created 28 May 2015

Last Revised 09 September 2017

Collated by Anna Crawford, Luc Desjardins, Ron Saper and Derek Mueller

Water and Ice Research Lab  
Department of Geography and Environmental Studies  
Carleton University  
1125 Colonel By Drive  
Ottawa, ON K1S 5B6

Recommended citation:

Crawford, A., Desjardins, L., Saper, R., Mueller, D. 2017. Canadian Ice Island Drift, Deterioration and Detection (CI2D3) Database Database Documentation V1.0. Water and Ice Research Laboratory, Carleton University, Ottawa. 18 pp.

## Contents

1. Introduction .....	5
1.1 Initiation of the Database .....	5
1.2 Funding .....	5
1.3 Personnel .....	5
1.4 Initiation and requirements .....	6
1.5 Overview of workflow .....	6
2. Database contents .....	6
3. Data sources.....	8
4. Software and tools .....	9
4.1 SigLib .....	9
4.2 ArcMap.....	10
4.2.1 – Image retrieve and load .....	10
4.2.2 – Attribute filling tools. ....	10
3.3 PostgreSQL/PostGIS .....	10
4. Workflow.....	11
4.1 Data discovery (or some other term).....	11
4.2 Image manipulation .....	11
4.3 Some of the pre-digitizing steps .....	11
4.4 Digitizing and lineage establishment .....	12
4.5 Lineage .....	13
4.6 Quality Control.....	13
4.7 Database consolidation and field generation .....	14
5.0 Error Analysis .....	15
6.0 Conclusions and Further Steps.....	15
7.0 Other documentation and references .....	17
Appendix 1: Database attributes .....	19

## Glossary

**“Best fitting” digitization:** The physical process of moving and rotating an ice island polygon to its newly observed position.

**Calving:** The initial break-off event of ice from an ice shelf or glacier terminus (Cogley et al., 2011).

**Daughter:** A new ice island created after a mother ice island fractures, or the succeeding observation of the mother ice island.

**Fracturing:** A breakage event occurring to an ice island; will likely create multiple daughter ice islands.

**“From scratch” digitization:** Original digitization of an ice island.

**Ice island:** A term used traditionally in the Arctic to refer to a large, tabular iceberg which calved from an ice shelf or floating ice tongue. They have a thickness of 30 to 180+ m and surface area extents can range from a few thousand square metres to > 500 sq. km. Their Antarctic counterparts, referred to as ‘tabular icebergs’, are known to have even greater dimensional extents. Depending on their origin, ice islands can be characterized by a regularly undulating surface giving a ribbed appearance from the air (CIS, 2005).

**Ice island fragment:** An ice mass that has fractured from a mother ice island (CIS, 2005). We refer to both the originally calved pieces as well as those generated through subsequent fractures as ‘ice islands’ due to their overlapping sizes.

**Ice shelf:** A thick, floating ice mass attached to a coast. They are often characterized by large areal extents and level surfaces. An undulating ridge-and-trough surface pattern may also be apparent. Their growth occurs through snow accumulation and/or the flow of land glaciers which form (floating) ice tongues, which may merge with ice shelves or be solitary (see below). (CIS, 2005; Cogley et al., 2011).

**Ice tongue:** A type of narrow, floating ice shelf that is an extension of an outlet glacier or ice stream. They can also be referred to as a glacier tongue (Dowdeswell and Jeffries, 1997).

**Lineage:** The hereditary tree of ice islands. This is represented in the database through mother-daughter ice island linkages whose relationship information is stored through the use of unique identifiers. These linkages can be made between two consecutive observations of the same ice island or between a single ice island and the 2+ fragments generated due to its fracture.

**Mother:** The parent of a daughter ice island; ultimately, the “original mother” is the glacier.

**Operator:** A person who contributed to the analysis and interpretation of satellite images and the digitization of the various ice islands’ shapes contained as entries in the CI2D3 Database.

**Orphan:** An ice island whose mother cannot be ascertained with certainty. Such fragments usually drifted for some time after a calving or fracturing event before showing any recognizable characteristics on radar images.

**PostgreSQL/PostGIS:** an open-source database program (PostgreSQL) that is geospatially enabled with the PostGIS extender and used for housing the CI2D3 Database. 'SQL' or 'structured query language' is used to quality control the database contents.

**"Trimming" digitization:** Digitization process resulting in the diminution in size of an ice island polygon. The trimming is accomplished after copying the polygon representing the ice island's mother to the SAR scene in which its daughter is identified and "best fitting" the polygon to its new position and then removing small portion(s) along its perimeter in order to match what is being observed in the latest SAR scene.

**Unique identifier:** A character string assigned to each ice island polygon for identification and lineage tracking purposes.

# 1. Introduction

## 1.1 Initiation of the Database

Ice islands are massive, tabular icebergs which calve from ice shelves and floating ice tongues. The ability to identify, monitor and predict the drift and deterioration of these immense ice hazards is crucial for mitigating the associated risks to marine navigation and offshore infrastructure in their vicinity. The Canadian Ice Island Drift, Deterioration and Detection (CI2D3) Database is a joint initiative between the Water and Ice Research Lab (WIRL; Carleton University, Ottawa, Canada) and the Canadian Ice Service (CIS; Environment and Climate Change Canada) to digitize and catalogue major ice island calving events in Canadian waters for research purposes.

The CI2D3 Database project seeks to extract pertinent information from available satellite imagery and build a geospatial database for ice island drift and deterioration analyses, remote-sensing signature analysis, and model calibration and validation. The CI2D3 Database currently contains entries associated with the influx of ice islands through eastern Canadian waters after massive calving events at the Petermann Glacier in 2008, 2010 and 2012. Ice islands from other northern Greenland ice tongues (i.e., Ryder, Steensby and C.H. Ostenfeld) are also included, though the monitoring of these ice islands was less systematic than those originating at the Petermann Glacier. Two small, previously undocumented 2011 calving events that produced three small ice islands from the Petermann Glacier are also included in the database.

Thousands of archived RADARSAT-1 and -2 (Canadian Space Agency/MacDonald, Dettweiler and Associates) and Envisat (European Space Agency) synthetic aperture radar (SAR) images were exploited to track ice islands until they were too small to delineate ( $< 0.25 \text{ km}^2$ ). As of the date of this document, more than 14,700 ice island polygons have been delineated in ArcMap (ESRI) by operators using custom productivity tools. The lineage of each ice island were captured to permit spatio-temporal studies of fracturing, deterioration and drift.

## 1.2 Funding

This project was undertaken with the financial support of the Government of Canada through the federal Department of Environment. Funding was also provided through Polar Knowledge Canada's 'Safe Passage' project.

## 1.3 Personnel

The CI2D3 Database project is led by Derek Mueller, a professor in the Department of Geography and Environmental Studies at Carleton University and co-director of the Water and Ice Research Lab (WIRL). You may contact him at [derek.mueller@carleton.ca](mailto:derek.mueller@carleton.ca). The project was initiated by individuals associated with WIRL and the CIS who collectively realized the potential of mining the CIS SAR archives. Tom Carrieres, Matt Arkett, Angela Cheng and Ron Saper were CIS employees involved in the CI2D3 Database planning and funding sourcing.

A team of WIRL members developed the workflow and project management structure for the CI2D3 Database generation. This included: Derek Mueller, Anna Crawford, Ron Saper, and Luc Desjardins. Desjardins led the ice island digitizing, and other digitizers included Jeff Shephard and Correy Schaad. Two groups of undergraduate students from Carleton University's GEOM 4008 and 4009 courses

developed the customized ArcGIS tools that were used by the operators. Individuals associated with the tool development include Dan Patterson, Adam Garbo, Sonya Haskell, Sara Mannseichner, Marisa Ramey, Cameron Samson, Diandra Buttazoni, Ryder Burt, Erik Duncan, Malek Singer and Emma Zerr. Finally, the PostgreSQL database was setup by Gregory Lewis-Paley and Sougal Bouh Ali.

## 1.4 Initiation and requirements

The CI2D3 Database project was initiated with the overall objective of capturing the ice island flux through the eastern Canadian Arctic and sub-Arctic after recent, large Petermann Glacier calving events that occurred in 2008, 2010 and 2012. Imagery was to be searched based on the spatiotemporal distribution of the ice islands, ranging from the Lincoln Sea to the Grand Banks of Newfoundland. Monitored ice islands were to be identified and their 2D planimetric outline recorded in the database at a minimum of once every two weeks. The lineage of the ice islands was to be also recorded and is a unique characteristic of the CI2D3 Database. See the Lineage section below for more details. These monitored ice islands were followed until they had deteriorated to 0.25 km<sup>2</sup>, though many were followed below this threshold. 'Orphan' ice islands (i.e., those for which a prior lineage could not be established) were monitored if they were > 0.25 km<sup>2</sup> when first observed. This increased to > 0.5 km<sup>2</sup> for digitizations representing orphan ice islands present in the region after July 2011. For a complete set of requirements please consult the dedicated project requirements document (Saper and Crawford, 2016).

## 1.5 Overview of workflow

Once the project objectives and requirements were developed, an iterative workflow was followed to generate the CI2D3 Database. Footprints of the SAR scenes within the CIS archive were used to select which scenes would likely be useful to the project based on their likelihood of capturing ice island based on the operators' knowledge of where these ice islands were located over time. The relevant SAR scenes were transferred from the CIS archives to a secure WIRL server and converted into geotiff files (LCC projection) with customized Python scripts. Operators made monthly MXD files (ArcMap projects), imported all scenes associated with the relevant month, searched all scenes and identified those which captured ice islands. Some scenes were known to contain an ice island, based on their previously observed locations and possible drift speeds, but image quality hindered their identification. Those which did not capture an ice island, or which an ice island could not be identified within, were omitted from the project. The operators followed the steps included in the 'Digitization' section below. Additional SAR imagery was searched for, if necessary, to meet the project requirements. See the 'Data sources' section below for more info. Shapefiles were uploaded to a PostgreSQL/PostGIS database and quality controlled with a customized SQL script that identified problematic entries. These entries were flagged for the operators who corrected or commented on the entries before they were re-checked and ultimately added to the final database table.

## 2. Database contents

The geospatial CI2D3 Database contains entries pertaining to ice islands originating from three Petermann Glacier calving events. Entries representing ice islands sourced from the Ryder, Steensby and C.H. Ostenfeld ice tongues are also contained in the database, though they are not monitored as

frequently as those from the Petermann Glacier, especially during the winter months when the ice consolidates in these fjords and prevents any motion. A small number of ice islands from northern Greenland, whose glacier origin is unknown, are also included and referred to as ‘NGIIs’. Table 1 contains the number of entries associated with each calving event, the size of the original calving event, the number of SAR scenes from which ice island polygons were digitized (though more were searched) and the period which ice islands were monitored after each event.

**Table 1: Representation of ice island calving events in the Canadian Ice Island Drift, Deterioration and Detection (CI2D3) Database.** The date field refers to the date of the original calving event. Polygon and image counts are reported as of September 2017. PII = Petermann Ice Island. Originally published in Crawford et al. (forthcoming).

Name	Date	Total calved ice (km <sup>2</sup> )	Polygons	Images <sup>‡</sup>	Monitoring period
PII-2008	Jul 10	36.4 ±4.0	266	171	07/2008 - 07/2009
PII-2010	Aug 5	302.4 ±33.0	8025	2244	09/2010 - 12/2013 <sup>‡</sup>
PII-2011	Aug 16	4.3 ±0.42	502	489	08/2011 - 10/2012
PII-2012	Jul 17	144.6 15.8	1823	286	07/2012 - 12/2013 <sup>‡</sup>
Ryder*			2345	185	08/2011 - 12/2013 <sup>‡</sup>
Steensby*			697	228	09/2011 - 12/2013 <sup>‡</sup>
C.H. Ostenfeld*			568	27	07/2012 - 12/2013 <sup>‡</sup>
North Greenland*			488	184	07/2012 1 12/2013 <sup>‡</sup>

<sup>‡</sup>This only constitutes a subset of the images that were investigated; many images were scrutinized without any ice islands being identified

\*Ice islands originating from these glaciers were not always monitored from the time of initial calving.

<sup>‡</sup> Anticipated monitoring period, as digitization is ongoing.

Each entry contains a number of attribute fields (Appendix A). These pertain to information specific to the concerned ice island (‘ice island info’; e.g., surface area), the SAR scene from which it was digitized (‘image info’; e.g., sensor, beam mode), and lineage (Sections 4.2.2 and 4.5). A number of more subjective fields (‘Operator filled’ in Appendix A) were also filled with a choice from a selection of available values. Finally, a comment field was also filled in manually by the operators.

The lineage is captured through the use of unique identifiers. A unique identifier (“INST”) was assigned to each polygon and a custom GIS tool (Section 4.2.2) was used to capture the unique identifier of the ice island’s “mother” (“motherINST”; the previous INSTance which the same ice island was identified or the parent from which a daughter ice island fragmented). This lineage capture is a unique aspect of the CI2D3 Database and allows the family histories of ice islands to be queried. The ‘mother’/‘daughter’ terminology came naturally to the project. Gendered terminology has been utilized in Antarctic ice island literature, and is well accepted in the field of cell biology with its use in the description of mitosis.

### 3. Data sources

The CIS Radarsat-1 and -2 (R1 and R2) archive was the main source of SAR scenes used to identify, monitor and digitize ice islands. If additional scenes were necessary to meet the project’s requirements (Saper and Crawford, 2016) or to aid in monitoring of a particular ice island, additional data sources were searched. Additional R1 and R2 scenes were acquired from the National Earth Observation Data Framework (NEODF) Catalogue. Envisat Advanced SAR (ASAR) scenes were acquired, when needed, from the European Space Agency (ESA) eoPortal. Table 2 provides information regarding the satellites and sensors that acquired the SAR scenes. Specifications of each sensor and beam mode are provided in Table 3, along with the number of each used in the generation of the CI2D3 Database. The ScanSAR Narrow and ScanSAR Wide beam modes, with maximum resolutions of 50 and 100 m, respectively, are the most frequently used by CIS and therefore comprise the majority of SAR scenes that were utilized during database generation.

Misalignment was apparent in some SAR scenes’ georeferencing. The misalignment was not always in a consistent direction and the magnitude of misalignment varied, though greater misalignment was experienced at higher latitudes. An estimation of the magnitude was captured by the digitizers and is contained in the ‘georef’ attribute field (Appendix A) of any polygon that was digitized from the image.

**Table 2: Satellite and sensor specifications** (Kult, 2009; CSA, 2015; RADARSAT International, 2004; USGS, 2011). SAR = synthetic aperture radar, ASAR = advanced SAR, EO = Earth Observation, ALI = advanced land imager, CSA = Canadian Space Agency, ESA = European Space Agency, NASA = National Aeronautics and Space Administration, UV = ultraviolet, NIR = near infrared, SWIR = short wave infrared.

Satellite	Sensor	Agency	Frequency band/class	Time period	Polarization
<b>RADARSAT-1</b>	SAR	CSA	C-band radar	2008-2012	HH
<b>RADARSAT-2</b>	SAR	MDA	C-band radar	2008-2013	HH, HH+HV
<b>Envisat</b>	ASAR	ESA	C-band radar	2008-2011	HH
<b>EO-1</b>	ALI	NASA	Visible – UV, NIR, SWIR	2010	



**Table 3: Specifications regarding the beam modes of the synthetic aperture radar and visual data used in the construction of the Canadian Ice Island Drift, Deterioration and Detection (CI2D3) Database** (Kult, 2009; CSA, 2015; USGS, 2011). SAR = synthetic aperture radar, ASAR = advanced SAR, EO = earth observation, ALI = advanced land imager. Originally published in Crawford et al. (forthcoming).

Satellite Sensor	Beam mode	Swath width (km)	Resolution (m)	Images used <sup>‡</sup>
R1 SAR	ScanSAR Wide	500	100	1205
	ScanSAR Narrow	300	50	69
	Wide	150	30	13
	Fine	45	8	10
R2 SAR	ScanSAR Wide	500	100	12311
	ScanSAR Narrow	300	50	306
	Extended low	170	60	6
	Standard	100	25	5
	Wide	150	25	28
	Fine-quad	25	12	25
	Multi-look fine	90	8	30
	Fine	50	8	3
Ultra-fine	20	3	2	
Envisat ASAR	Wide-Swath	100	150	414
EO-1 ALI	ALI	37	10	1

<sup>‡</sup>This only constitutes a subset of the images that were investigated; many images were scrutinized without any ice islands being identified.

## 4. Software and tools

Several software packages and custom tools were brought together to assemble and quality-control the CI2D3 Database. Many of these were open-source tools or derived from them. However, we relied heavily on ArcMap and customized scripts to display the imagery, digitize polygons and generate attribute data. These software tools are described below, followed by the workflow that was performed to generate the final result, the CI2D3 Database.

### 4.1 SigLib

For data discovery and manipulation to geotiff we used a custom Python program called SigLib that was developed in-house. SigLib stands for 'Signature Library' and is a suite of tools to query, manipulate and process SAR remote sensing imagery (primarily RADARSAT-1 and -2) and store the signature and metadata in a geodatabase. It uses open source libraries and can be run on Windows or Linux. This software itself will be made open source at a later date. It runs on Python version 2.7 with dependencies

to several scientific and geospatial computing libraries, notably gdal/ogr, numpy, pandas, psycopg2. It is set up to interface with a PostgreSQL/PostGIS database.

## 4.2 ArcMap

All digitization was conducted in ArcMap, Versions 10.2– 10.5. Three customized ArcGIS toolbars were developed for this project.

### 4.2.1 - Image retrieve and load

Convenient subsets of the CIS archive index shapefile were created based on timespan and geographic region. The operator then selected outlines of images desired for examination, and a custom ArcPy script was invoked to bulk load all the images from the local store of projected geotiff images. Those image frames that were not in the local store were identified by this procedure and a catalogue number was reported so that the images could be requested from CIS.

### 4.2.2 - Attribute filling tools.

Two customized toolbars were developed to help operators efficiently fill the polygons attribute fields.

1) Ice Island: This toolbar populates the ‘ice island info’ and ‘image info’ attribute fields and also contains a number of dropdown menus that provided operators with value options for more subjective attribute fields (Appendix A).

The toolbar contains two main ‘buttons’. The first (Populate Info) populated all ‘Image info’ attribute fields while the second (Populate Geometry) computed all attributes regarding the geometry of the polygons contained within the file (i.e., surface area, perimeter, and centroid values).

The image info fields were filled by parsing the SAR scene’s file name (‘imgref1’), which was also inserted as an attribute field (‘imgref1’). The ice island’s unique identifier (‘INST’) was also established at this time. This consists of a character string, “YYYYMMDD\_HHMMSS\_SENSOR\_FID\_XXX”, where XXX is a random 3-letter string. All other characters were parsed from the imgref1 field and the shapefile’s feature identification (assigned when digitized in ArcMap).

2) Lineage: This toolbar populates an ice island polygon’s motherINST field through a consistent ‘selection’ order of the concerned polygons (mother and daughter). A modification was added to help insert a new identification of an ice island in between two already linked entries. This evolved toolbar will be useful if future effort is put towards densifying observations.

## 3.3 PostgreSQL/PostGIS

The database is housed in a PostgreSQL database. A script was written in SQL for quality controlling the data and run through a Python software with database connection.

## 4. Workflow

### 4.1 Data discovery (or some other term)

Several Regions Of Interest (ROIs) were defined based on existing knowledge of where ice islands from the Petermann Glacier are expected to drift. Five ROIs were defined as large, quasi-rectangular, mutually exclusive polygons from Kennedy Channel to Newfoundland. They were clipped using the [Global Self-Consistent Hierarchical, High Resolution Geography Database](#) coastline (Wessel and Smith, 1996), which served as a reference vector layer in our workflow. Using SigLib, an index shapefile containing the footprint of all the scenes in the CIS archive was added to a PostgreSQL/PostGIS database and were queried against the ROIs over certain date ranges. This identified images of interest which were obtained as Level 1 products from the CIS archive (or NEODF).

### 4.2 Image manipulation

SigLib was employed to convert images stored in zip files into projected geotiff files as follows. Each image was unzipped, converted to an amplitude image, projected to a Lambert Conformal Conic (LCC) projection (see below) using bilinear resampling and each band underwent a separate 2 standard deviation stretch, to maximize contrast and were byte-scaled and LZW-compressed to reduce file size. Image pyramids were created to increase the response time in ArcMap during panning and zooming operations. The LCC projection used in this project is identical to the one used by CIS (standard parallels at 49°N and 77°N and an origin at 40°N, 100°W) except that we use the WGS84 ellipsoid instead of the Clark 1866 ellipsoid. Our LCC projection in various formats can be retrieved here: <http://spatialreference.org/ref/sr-org/6718/>.

### 4.3 Some of the pre-digitizing steps

The initial phase of the project consisted of collating all pertinent information with respect to each calving event. Through time, many people/agencies contributed images (screen captures of radar and visual satellite images) showing the location of the larger ice islands as well as products (e.g. beacon drift track, ship chart, visual images from shipborne or aerial perspective) related to the various ice islands during their lifespan. The purpose of collating all this information was to provide a timeline for the individual calving events and to help select appropriate images over pre-defined region of interest in order to efficiently sift through thousands of potential R1 and R2 scenes in the CIS archive or Envisat scenes from the ESA's eoPortal so as to only retrieve those believed to be relevant.

The digitization process started once the appropriate imagery was retrieved and converted to projected geotiff files. An 'Ice Island Template' shapefile was developed to ensure that a standard set of metadata fields were recorded for each polygon (see toolbar). A final version of this template was settled on after many iterations and some fields defined in the final version were missing in the original template. These missing fields were later filled in by going back to the various shapefiles and manually editing their contents when needed.

A number of options for the digitization process were considered. One possibility was to have one operator concentrate on a single calving event at a time; another possibility was to have multiple

operators working on multiple simultaneous events at the same time. A third possibility was to have various operators concentrating on the detection and analysis of ice islands over a prescribed region of interest (ROI). This methodology was eventually abandoned as it would generate discontinuity at the boundary of each region if one operator followed a fragment which was never detected by the operator responsible for the adjacent ROI.

In the end, it was deemed better to have one operator per event; this way, each operator would be self-aware of all the fragments associated with their own calving event. Starting from the initial calving event and scrutinizing every possible source of images in a chronological manner ensured the best possible final product.

#### 4.4 Digitizing and lineage establishment

Digitizing occurred in monthly batches. Operators identified SAR scenes that potentially covered ice islands from the CIS archive. This was done by overlapping footprint shapefiles of the archived scenes with the area and time period in which ice islands may be present. An ArcMap project was created for each month and all images selected as potentially capturing ice islands were loaded in a chronological order.

Each image containing one or more fragments would serve as a naming convention for a shapefile. Each fragment digitized within a shapefile was also given an alias by the operator. Successive occurrence of the same alias was necessary to ensure a proper lineage. Since the geographical coverage of an image changed from one day to the next, different fragments moving in separate or parallel directions would not necessarily be re-sighted on the next image. A new shapefile could easily contain several fragments comprised from several previous shapefiles.

The following layers were used to help with image archive searching, digitizing, or were created by the digitizers:

- image index file (used to search the image archive),
- images
- coastline files
- area of interest boundary, and
- polygon shapefiles (one associated with each image).

Digitization was done from scratch only for the first instance that an ice island was identified. For the successive sightings of the same ice islands, the representative polygon was copied and pasted forward to the new scene, best-fitted to its new location and only trimmed if necessary. This ensured that changes in surface area were genuine. Due to various image parameters (e.g., incident angle), the apparent surface area of a fragments can change from one image to the next. Reporting every possible change would result in a constant fluctuation (trim/growth) of the perimeter of a fragment. This is not realistic nor desirable and is why the trimming workflow was established.

The customized toolbars were then used to fill all attributes except for the lineage aspect ('mothercert' field).

## 4.5 Lineage

The ice island lineage, introduced in Section 2, captures the life history of the ice islands represented in the database. The lineage was typically not established until after the subsequent month's digitizing was complete. This provided the operators with more information regarding the history of each ice island and decreased the number of modifications needed.

Operators attempted to identify each monitored Petermann ice island at least once every two weeks. The observation repeat period is, however, "optimal" but is subject to image availability, image quality and/or the presence of sea ice (which can sometimes obscure the ice islands or at times reveal its presence through an open-water lead developing as the a result of the differential speed between the ice and the ice island). The once per 2 week observation requirement has been achieved for the vast majority of ice islands. Those that are not represented every two weeks can be identified with a query based on ice islands' aliases and timestamps. This may be necessary for certain analyses or future database densification efforts.

The ice islands sourced from other northern Greenland ice tongues were only re-sighted once per month during the winter months (when ice trapped the pieces in their respective fjords). Fragments outside these fjords were re-sighted as often as imagery permitted.

Entries with 'na' values within the motherINST field are either orphans or calved directly from an ice shelf or ice tongue.

## 4.6 Quality Control

All database entries were queried through SQL code written to alert the operators if a 'rule' was broken or a 'flag' was raised. Entries that broke a rule were investigated and fixed. Flagged entries were also investigated. However, following a check, these could be allowed to pass through to the database upon judgement of the operator. Rarely, an exception was also given to a broken rule after investigation by a operator.

### 'Hard and fast' rules:

**r1:** The imgref1 needs to be of a certain formula YYYYMMDD\_HHMMSS\_ss\_bm\_\_\_pz\_a\_prj where ss=sensor, bm = beammode, pz=polarization, prj = projection. This has consequences for INST, motherINST and scenedate fields

**r1a:** checks for correct INST format. If this is flagged makes sure to check lineage and that connections are all made correctly.

**r1b:** checks for correct motherINST format. If this is flagged makes sure to check lineage and that connections are all made correctly.

**r2:** There should be no spaces before or after the Alias entry

**r3:** Make sure the operator field is correctly filled with 'Luc', 'Jeff' or 'Correy'

**r4:** The operator field must be filled.

**r5:** The imgref1 and datepoly fields must be filled.

**r6:** The INST field must be filled.

**r7:** The scene\_date, sensor, beam\_mode and polarization fields must be filled.

**r8:** The poly\_area, perimeter and centroid\_lon, centroid\_lat fields must be filled.

**r9:** The shpcert field must be filled.

**r10:** shpcert must equal 1, 2, or 3.

**r11:** An INST cannot be repeated.

**r12:** The date of the INST must be after that of the MotherINST

**r13:** If two ice islands have the same motherINST, their poly\_area's should not be exactly the same

**r14:** A mothercert value should not exist if motherINST is not assigned.

**r15:** Flag all times when the summed surface area of daughter ice islands is > 125% of the immediate mother.

**r16:** For the same alias, there should not be a change in mothercert values between two successive digitizations of that ice island.

### **Flags:**

**1:** Flag instances where 'traceback' is checked. This may be used in-house for potential densification of observations at a later date.

**f2:** Flag when the area of a polygon calculated within postgres is different than poly\_area

**f3:** Flag all times when drift speeds between two ice islands connected via the lineage tool (motherINST to INST match) are greater than 24 km d<sup>-1</sup> but are less than 50 km d<sup>-1</sup>

**f3a:** Flag all times which drift speeds between two ice islands connected via the lineage tool (motherINST to INST match) are greater than 50 km d<sup>-1</sup>.

**\*f4:** Flag all times when the summed surface area of daughter ice islands is > 110% but less than 125% of the immediate mother.

**f5:** Flag instances where IICERT value changes from one instance to the next (lineage must be set). IICERT was an attribute used in-house for digitizing.

\*This % value was used to capture fracturing events which occurred in viewing conditions different than the initial occurrence of the mother.

## **4.7 Database consolidation and field generation**

Polygons were uploaded to a final PostgreSQL database after the quality control process. The database was assigned the LCC projection defined above.

Three new fields were generated at the time of upload: length, calving event and calving location (Appendix A). The latter two refer to the year and location of the calving event from which the concerned ice island originated. These were filled by information contained in the scene date and alias attribute fields.

Length was determined by finding, via SQL query, the length of the longest line between any two vertices of a polygon.

Logic rules were also followed to establish consistent default values for the following attribute fields:

- If debrisfld is null, set to 'N'
- If surf\_feat is null, set to 'N'
- If dinfo is null, set to 'drifting'
- Set mothercert to 'n' if motherINST is null
- If mothercert is null and motherINST is not null, set mothercert to 1

## 5.0 Error Analysis

The uncertainty of both digitization methods ('from scratch' vs. 'best fit' and 'trimming') were assessed by conducting two error analysis experiments. Five operators were involved in each experiment, including those employed as 'operators' by the CI2D3 Database project, plus others involved in the project but not involved in digitizing. The 5 operators either digitized 4 individual ice island fragments 'from scratch' on 5 different occasions (experiment 1) or they followed the 'best fit'/'trimming' method for 4 pairs of scenes and modified the polygon for the second scene if the shape was judged to have changed (experiment 2). The latter was also repeated five times to approximate independent trials. Operators took at least 1 day in between digitization rounds. Two types of errors were derived from the relative standard deviations in the underlying data. The first represents the internal consistency of each CI2D3 operator's digitization and the second represents the agreement between all operators (CI2D3 and others) on the size of a given ice island from their first digitization trial. On average, the CI2D3 operators internally consistent within 5.3% (range 3.4 to 8.6%) for the 'from scratch' workflow and were within 2.5% (range 1.6 to 3.4%) repeatability when conducting 'best fit'/'trimming'. All operators agreed on the surface area within 8.4% for 'from scratch' digitization and were within 6.2% of each other, on average, for the 'best fit'/'trimming' method.

## 6.0 Conclusions and Further Steps

The CI2D3 Database provides the most comprehensive monitoring of the ice islands produced after major calving events from the Petermann Glacier. The database has two unique features. First, the CI2D3 Database tracks Petermann ice islands until they reach a surface area of < 0.25 km<sup>2</sup>. Second, the lineage of these ice islands is established so that the life history can be queried.

The objective of the project was to document the 2008, 2010 and 2012 Petermann ice islands. The CI2D3 Database covers the entire lifespan of ice islands originating from the 2008 event. The database monitored the ice islands that stemmed from the 2010 event for 3.5 years, and those that originated from the 2012 event for 1.5 years. Elongating these monitoring periods (possibly up to 'real time') is a potential next step for the database.

While the database does contain entries representing ice islands that originated from other northern Greenland glaciers, these ice islands were not monitored with the same dedication as they were already afloat and trapped within their respective fjords when initially digitized (note: the requirement to

digitize all fragments  $> 0.5 \text{ km}^2$  was not adhered to for these ice islands that were sourced from glaciers other than Petermann, as this would have yielded even more fragments). Densifying the observations of these ice islands is a potential future step for the database project, as is representing ice islands generated after ice shelf calving events along the northern coast of Ellesmere Island.



## 7.0 Other documentation and references

### Publications:

Crawford, A.J., Crocker, G., Mueller, D., Desjardins, L., Saper, R., and Carrieres, T. Forthcoming. An introduction to the Canadian Ice Island Drift, Deterioration and Detection (CI2D3) Database. *Journal of Glaciology (submitted)*.

Crawford, A.J., Wadhams, P., Wagner, T.J.W., Stern, A., Abrahamsen, E.P., Church, I., Bates, R., and Nicholls, K.W. 2016. Journey of an Arctic ice island. *Oceanography*. 29(2).

### Database requirements:

Saper, R. and Crawford, A. 2016. Canadian Ice Island Drift, Deterioration and Detection (CI2D3) Database Requirements Specification Document V1.0. Water and Ice Research Laboratory, Carleton University, Ottawa. 11 pp. [https://wirl.carleton.ca/wp-content/uploads/2016/04/CI2D3\\_Requirements\\_v1.0.pdf](https://wirl.carleton.ca/wp-content/uploads/2016/04/CI2D3_Requirements_v1.0.pdf)

### CI2D3 Database data:

Desjardins, L., Crawford, A., Saper, R., Shepherd, J., Schaad, C., and Mueller, D. (2016) Canadian Ice Island Drift, Deterioration and Detection (CI2D3) database. Water and Ice Research Laboratory, Carleton University. <https://wirl.carleton.ca/ci2d3/>. DOI: 10.21963/12678

### References:

CIS (Canadian Ice Service). 2005. MANICE: Manual of standard procedures for observing and reporting ice conditions. Marine and Ice Services Division, Environment Canada, Ottawa <http://www.ec.gc.ca/Glaces-Ice/default.asp?lang=En&n=2CE448E2-1>

Cogley, J.G., Hock, R., Rasmussen, L.A., Arendt, A.A., Bauder, A., Braithwaite, R.J., Jansson, P., Kaser, G., Möller, M., Nicholson, L., and Zemp, M. 2011. Glossary of Glacier Mass Balance and Related Terms, IHP-VII Technical Documents in Hydrology No. 86, IACS Contribution No. 2, UNESCO-IHP, Paris.

CSA (Canadian Space Agency). 2015. Satellite Characteristics. Date accessed: 13 June 2017. <http://www.asc-csa.gc.ca/eng/satellites/radarsat/radarsat-tableau.asp>

Dowdeswell, J.A., and Jeffries, M.O. 2017. Arctic ice shelves: an introduction. In: Arctic Ice Shelves and Ice Islands, Copland, L. and Mueller, D. (eds). Springer Polar Sciences. pp 3-21. DOI: 10.1007/978-94-024-1101-0\_1

Kult, A. 2009. Volume 8: ASAR products specifications. Envisat payload data segment. European Space Agency and MacDonald Dettweiler and Associates.

MDA (MacDonald, Dettwiler and Associates, Ltd.). 2008. RADARSAT-2 product format definition.

RADARSAT International. 2000. RADARSAT data product specifications. RSI-GS-026 v. 3.0. 125 pp.

USGS (United States Geological Survey). 2011. Sensors – Advanced Land Imager (ALI). United States Government of the Interior. Date accessed: 13 June 2017. <https://eo1.usgs.gov/sensors/ali>

Wessel, P., and W. H. F. Smith, A Global Self-consistent, Hierarchical, High-resolution Shoreline Database, *J. Geophys. Res.*, *101*, 8741-8743, 1996

## Appendix 1: Database attributes

A set of customized ArcGIS tools were created for efficient filling of geometry and SAR scene information in the respective attribute fields. **Operator filled** **Ice island info** **Image info** **Lineage**

All ice island entries have the following attributes (Part 1 of 2):

FIELD NAME	DESCRIPTION	FILL OPTIONS, FURTHER EXPLANATION & NOTES
<b>alias</b>	Name chosen by an operator to identify a fragment over time	
<b>operator</b>	Name of person who digitized the polygon	Luc, Correy or Jeff
<b>shpcert</b>	Represents how confident an operator is in their delineation of the ice island's true edge	1 (very certain), 2 (fairly certain) or 3 (uncertain)
<b>imgref1</b>	Populated from the file name of the image from which the polygon was digitized.	YYYYMMDD_HHMMSS_ss_bm__pz_a_prj where ss = sensor, bm = beammode, pz = polarization, prj= projection
<b>surface area</b>	Surface area of the polygon (km <sup>2</sup> )	Precision = 4 (Significance of last figures are not necessarily known.)
<b>perimeter</b>	Perimeter of the polygon (km)	Precision = 4 (Significance of last figures are not necessarily known.)
<b>INST</b>	A unique identifier for each polygon	YYYYMMDD_HHMMSS_SENSOR_FID_XXX, where XXX is a random 3-letter string
<b>motherinst</b>	The INST of the polygon which represents the 'mother' of each ice island. This is set to 'n' for Orphan ice islands.	YYYYMMDD_HHMMSS_SENSOR_FID_XXX
<b>ddinfo</b>	Information regarding the drift and deterioration of each ice island	Drifting, grounded, grounded? (unsure), trapped (within consolidated sea ice), trapped? (looks to be within consolidated sea ice, but unsure)
<b>mothercert</b>	Indicates how confident an operator is in the lineage connection (i.e., that a daughter fragment truly came from a certain mother)	1 (very certain) or 2 (fairly certain), n = none

Attribute fields (Part 2 of 2):

FIELD NAME	DESCRIPTION	FILL OPTIONS, FURTHER EXPLANATION & NOTES
georef	How far off an image is misaligned relative to the coastline vector	0-100m, 100-200m, 200-400m, > 400m
scenedate	Parsed out of imgref1	YYYY-MM-DD HH:MM:SS
sensor	Parsed out of imgref1	r1 (Radarsat-1), r2 (Radarsat-2), es (Envisat), and al (Advanced Land Imager)
beam_mode	Parsed out of imgref1	Numerous*
polarization	Parsed out of imgref1	hh, hx, vv, vh, qp, vi where h = horizontal, v = vertical and vi = visible, qp = quad-pol, x = dual cross-pol
surf_feat	Is there a visible surface feature?	'Y' or 'N'
centroid_x	Longitude of polygon centroid in decimal degrees	Precision = 4 (Significance of last figures are not necessarily known.)
centroid_y	Latitude of polygon centroid in decimal degrees	Precision = 4 (Significance of last figures are not necessarily known.)
calving year	Year of initial calving event	This is unknown, and therefore not filled, for RG (Ryder Glacier), SG (Steensby Glacier), CG (C.H. Ostenfeld Glacier) or NG (North Greenland).
calving location	Ice tongue where the ice island originated	Options include PG (Petermann Glacier), RG (Ryder Glacier), SG (Steensby Glacier), CG (C.H. Ostenfeld Glacier) or NG (North Greenland).
length	Maximum distance between all two vertex combinations in the polygon	
geometry	All vertices of an ice island polygon. Will be contained as part of the ice island shapefile.	

**\*Beam modes. The numbers that follow indicate frame # and represent different incidence angle ranges.** (CSA, 2015; MDA, 2008; RADARSAT International, 2000)

f\* = Fine

fq\* = Fine-Quad

el\* = Extended Coverage Extended Low beam

mf\* = Multi-Look Fine Resolution

s\* = Standard

scn\*, sn\* = ScanSAR Narrow

scw\*, sw\* = ScanSAR Wide

uf\* = Ultra-fine Resolution

w\* = Wide

wsm = Wide-Swath Medium Resolution

xxxx = ALI beam mode