Title

Towards Interoperability of ICEPMAG, PINT, and MagIC Databases

Author

Carrasco, Trinity Lauren

Publication Date

2023

UNIVERSITY OF CALIFORNIA SAN DIEGO

Towards Interoperability of ICEPMAG, PINT, and MagIC Databases

A Thesis submitted in partial satisfaction of the requirements
for the degree Master of Science

in

Earth Sciences

by

Trinity Carrasco

Committee in charge:

Professor Catherine Constable, Chair
Professor Jeffrey Gee, Co-Chair
Professor Emily Van Allen

2023

The Thesis of Trinity Carrasco is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

To my family

To mom

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ICEPMAG	Iceland Paleomagnetic Database
PINT	Paleointensity Database
MagIC	Magnetics Information Consortium
IGSN	International Generic Sample Number
GAD	Geocentric Axial Dipole
VADM	Virtual Axial Dipole Moment
VDM	Virtual Dipole Moment
VGP	Virtual Geomagnetic Pole
AF	Alternating Field
GEOMAGIA50	Paleomagnetic and chronological data for the past 50 thousand years
KARAR	Geochronology database for ^{40}Ar/^{39}Ar and K/Ar ages
PSVM	Paleosecular variation of the Miocene dataset
PSV10	Paleosecular variation of the past 10 million years dataset
FAIR	Findability, Accessibility, Interoperability, and Reusability Guiding Principles
DOI	Digital Object Identifier
ID	Identifier
UID	Unique Identifier (in ICEPMAG and PINT)
External Database ID	External database name (to MagIC) paired with the record ID where data are found
GPS	Global Positioning System
Metadata	Contextual information about data

LIST OF SUPPLEMENTAL FILES

carrasco_MagIC_method_code_uses.xlsx

carrasco_MagIC_contribution_corrections.xlsx

carrasco_jpytrnb_how_to_search.pdf

ACKNOWLEDGEMENTS

I would like to acknowledge Professors Cathy Constable and Jeffrey Gee as the chairs of my committee. Thank you for welcoming me into the field of geophysics and paleomagnetism and for fostering my ability to learn from and contribute to the field of Earth Science. Your combined guidance and mentorship have been invaluable to my research endeavors.

I would also like to thank and acknowledge Professor Emily Van Allen for being part of my committee. I want to thank the MagIC group, including Lisa Tauxe, Ruppert Minnett, and Nick Jarboe for your patience and willingness to answer my many questions regarding the MagIC interface and EarthRef API. Thank you to the ICEPMAG researchers, Justin Tonti-Filippini and Max Brown, for acknowledging my findings and making the necessary updates to ICEPMAG. Thank you to the PINT database researcher Richard Bono. The use of the MagIC interface and the ICEPMAG and PINT databases was essential to this research.

This work was supported by the Earthcube ICER (2126427) and the National Science Foundation (NSFGEO-NERC, EAR 1953778).

ABSTRACT OF THE THESIS

Towards Interoperability of ICEPMAG, PINT, and MagIC Databases

by

Trinity Carrasco

Master of Science in Earth Sciences

University of California San Diego, 2023

Professor Catherine Constable, Chair
Professor Jeffrey Gee, Co-Chair

Data archiving and accessibility is necessary for paleomagnetic research as researchers repurpose published data to resample, reinterpret, and update records to determine information about the past geomagnetic field. MagIC is the most comprehensive database for magnetic measurements. Here, we examine the overlap and consistency across specialized databases specifically PINT and ICEPMAG which focus on geomagnetic paleointensity and regional Icelandic data respectively. We provide an overview of some current database issues and offer suggestions for improved interoperability. We highlight the need for straightforward unique data identifiers to enable database interoperability, and address inconsistencies and inaccuracies in some records that should be identical across MagIC, PINT, and ICEPMAG. Method codes are used to describe field and lab procedures that provide important quality information to users in each of the three databases but differ in their definitions. Only a fifth of the MagIC method codes are actively used and many codes appear redundant or are inconsistently applied. This limits their effectiveness for evaluating data and linking methods across databases. We recommend the use of field mapping applications to improve the accuracy and precision of location data. We suggest that MagIC should remove redundant method codes, implement a hierarchy of method codes for common experiment names, support a tool to help users select method codes, and implement validations for calculated paleomagnetic data. We consider it vital that the paleomagnetic community establishes a unique identifier for each data record (whether it be IGSNs or another identifier) to improve traceability as new databases are developed.

Chapter 1

Paleomagnetism and Database Interoperability

1.1 Paleomagnetism

The geomagnetic field is an essential feature of the Earth that is studied across geoscience disciplines. Earth's magnetic field protects the planet and our atmosphere from cosmogenic radiation and high energy particles produced by solar wind. It also helps us (as well as birds, bees, and a few other animals) with navigation and is a large part of modern technology (Tauxe et al, 2018, Butler, 1992). With its importance to daily life on the planet, it is essential to study the geomagnetic field and how it changes with time.

The geomagnetic field can be modeled as a vector field around the Earth with mostly inward directed flux field lines in the northern hemisphere and outward in the southern hemisphere. These field lines vary in direction over time along with the strength of the field. One way to imagine the field is in its first order approximation as a dipolar source, such as that produced by a bar magnet placed at the center of Earth with magnetic field lines pointing inwards in the northern hemisphere and outwards in the southern hemisphere (Figure 1.1, Tauxe et al., 2018).

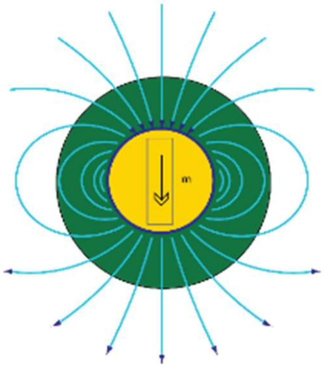


Figure 1.1: Magnetic flux lines produced by a magnet centered in a sphere (Tauxe et al., 2018)

The vector field that describes the geomagnetic field is modeled by three dimensional vectors at the surface of the Earth. At any point on the surface a vector will have two angles (declination and inclination) and a length (strength of the field, B), shown in Figure 1.2. At a point on Earth the inclination, I, is the angle from the horizon to the magnetic field direction and is positive downward as it ranges from +90º to -90º. At this same point the declination, D, is the angular difference between geographic and magnetic north, it is positive eastward and ranges from 0º to 360º (Tauxe et al., 2018). The field strength, often referred to as field intensity, is measured in microtesla, μT, with typical values ranging from about 30 μT at the equator to 60 μT at the poles for the modern field. In this thesis, the declination and inclination angles determined from the magnetization of rocks are referred to as directions, or paleodirections, and the field strength will be referred to as paleointensity.

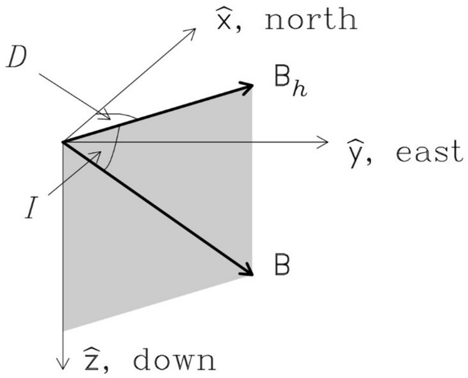


Figure 1.2: Geomagnetic field vector depicting geomagnetic directions: Declination (D), Inclination (I), and strength (B).

The geomagnetic field is generated and sustained by the dynamo action in Earth's outer core (Merrill, 1995). The chaotic movements of very conductive liquid iron and other lighter materials in the outer core cause changes over time that are strong enough to reverse the polarity of Earth's magnetic field such that the dipolar field has flipped hemispheres (Tauxe et al., 2018). The paleomagnetic polarity that matches today's geomagnetic polarity is considered "normal" polarity, if it is opposite to today's polarity then it is considered "reverse". In the field of paleomagnetism, we can use the magnetic properties of rocks that have preserved a record of the ancient magnetic field to deepen our understanding of the geomagnetic field behavior over time.

The time-averaged geomagnetic field is assumed to correspond to that of a geocentric axial dipole (GAD). The geomagnetic field is often characterized by deviations from the GAD field. There are a few calculations that will be used to evaluate paleomagnetic field behavior in this thesis. Directional variations of the geomagnetic field are often characterized by the inclination anomaly relative to the GAD inclination, given by $tan^{-1}(2tan(\lambda_s))$, where λ_s is the site latitude. Paleointensity measurements of the geomagnetic field are used to calculate the virtual dipole moment (VDM), which is the magnetic moment (measured in units of ampere meters squared, Am^2) of the equivalent geocentric axial dipole that would have generated an observed intensity at a particular site. The VDM is defined in (Equation 1.1) and uses the magnetic co-latitude from (Equation 1.2); these appear as equations 2.16 and 2.12, respectively, in (Tauxe et al., 2018). Using the site co-latitude, defined in (Equation 1.3), instead of the magnetic co-latitude in the VDM equation allows us to calculate a virtual axial dipole moment (VADM) (Tauxe et al., 2018).

(Equation 1.1) $$VDM = \frac{4\pi r^3}{\mu_o} \frac{B}{\sqrt{1+3\,co\ {}^{2}(\theta_m)}}$$

(Equation 1.2) $$\theta_m = cot^{-1}(\frac{1}{2}\tan(I))$$

(Equation 1.3) $$\theta_s = 90° - \lambda_s$$

To account for how the geomagnetic pole position has changed over time, we can transform an observed direction into its equivalent geomagnetic pole by calculating the virtual geomagnetic pole (VGP) position (given by the VGP latitude, λ_p, and longitude, ϕ_p) (Figure 1.3, Tauxe et al., 2018). The VGP coordinates also indicate the polarity of the paleomagnetic field such that VGP latitudes (λ_p) between 45º to 90º indicate a normal polarity, -90º to -45º indicate reverse polarity, and -45º to 45º indicate a transitional polarity. The inclination anomaly, VGP, and VADM allow us to describe how the field has changed over geologic time scales, on the scale of thousands to billions of years.

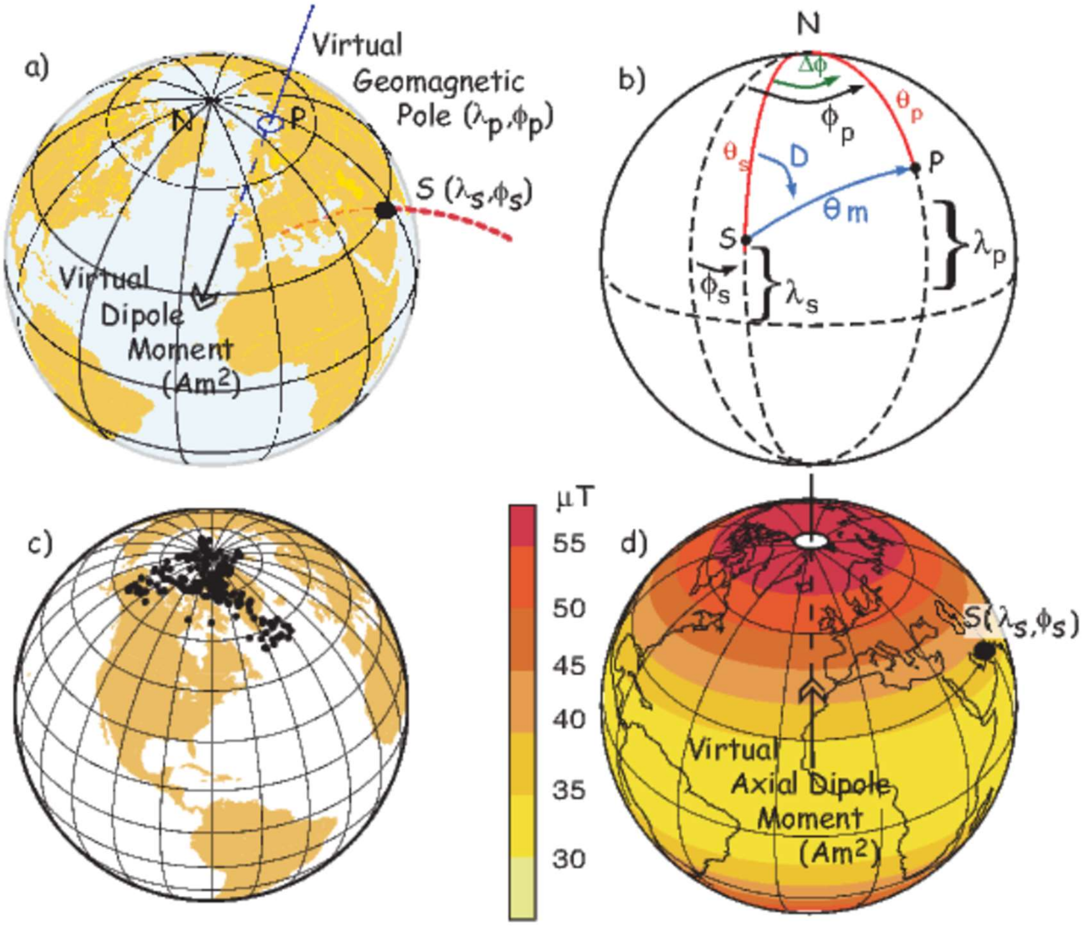


Figure 1.3: Global view of a vector measured at S (site latitude λ_s, site longitude ϕ_s) transformed into a VGP position P (pole latitude λ_p, pole longitude ϕ_p). The VADM models a geomagnetic field aligned with the geographic axis. a) The dashed red line going through S is the magnetic field line observed at this point. The dashed red field line represents the magnetic field line produced by the VDM at the center of the Earth. The VGP position, P, is the point in which the axis of the VDM pierces the surface of the Earth. b) Transformation of a vector measured at site S into its VGP position P using the observed inclination and declination. c) Example of transforming directions into their VGP positions. d) The VADM gives rise to the observed intensity at position S (Tauxe et al., 2018).

1.2 How Data are Collected and Measured

The key to a material becoming magnetized is its preservation of a magnetic remanence acquired when a field was applied to it. In paleomagnetism, we can measure the acquired remanent magnetization in materials containing ferromagnetic minerals including sediments, man-made artifacts such as pottery, and lava flows to obtain information about the ancient field (Tauxe et al., 2018). The remanent magnetization depends on the direction and strength of the geomagnetic field at the time of acquisition, but each material has varying recording abilities and complications. For example, sediments provide us with a time series of geomagnetic field directional variations but only relative changes in the intensity of the field can be estimated from sedimentary sequences. Both igneous rocks and man-made artifacts (e.g., clay pots that have been heated) can acquire a thermoremanent magnetization that can provide us with direction and intensity in the form of "spot" readings from the time when they cooled (Tauxe et al., 2018, Butler, 1992). This thesis will deal only with data collected from lava flows that have acquired a thermoremanent magnetization during cooling.

When molten lava cools, the magnetic moments within the ferromagnetic grains in the rock statistically align with Earth's magnetic field at the time, preserving the field's magnetic strength, polarity, and direction. Individual lava flows should provide a spot reading of the geomagnetic field and this cooling unit forms a paleomagnetic sampling site (Butler, 1992). Several (typically 5-10) oriented samples are taken from each flow in order to allow averaging of uncertainties related to sample orientation (Figure 1.4). Additional considerations include selecting samples with minimal alteration and sampling from portions of the flow that are unlikely to have been affected by reheating from subsequent lava flows. In the lab, oriented rock samples are sub-sampled into specimens whose characteristic remanent magnetization is isolated (by stepwise thermal or alternating field demagnetization) to ultimately recover information about the paleomagnetic field at that location.

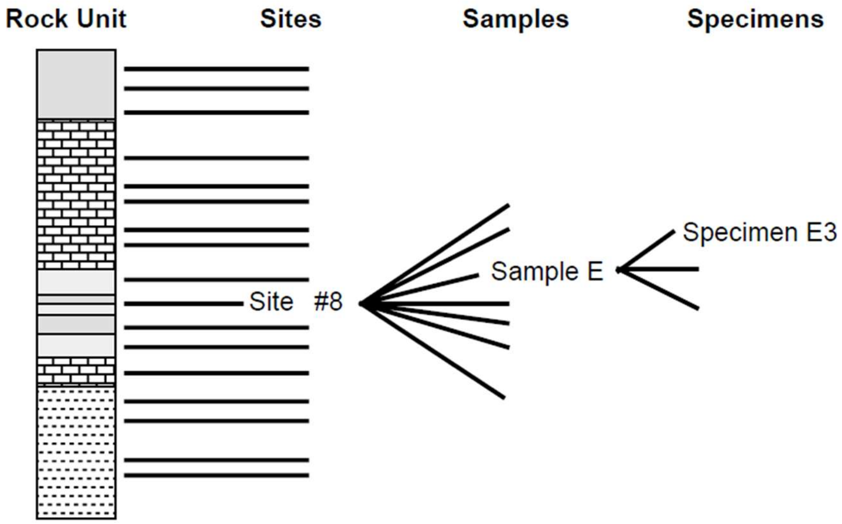


Figure 1.4: Paleomagnetic sampling site hierarchy (Butler, 1992). A set of lava flows layered on top of each other make up a rock unit. Several oriented samples are taken from each flow, which corresponds to a paleomagnetic site. To prepare for measurements, the sample is broken down into specimens. In the context of databases, measurements are made on specimens, but that data may be reported at the measurement, specimen, sample, or site level.

In many cases, it is possible to estimate the geomagnetic intensity recorded in lava flows by reproducing the process of remanence acquisition during cooling. Measuring magnetic intensity typically involves a stepwise double heating process where the natural remanent magnetization is removed (by heating and then cooling in zero field) and then replaced by a thermal remanent magnetization (acquired by cooling in a known laboratory field) that is measured after each step. The ancient field intensity is determined by calculating the ratio between magnetization lost at the zero-field step and the magnetization gained at the in-field step (Cych et al., 2023). Such measurement level data are beneficial to include in data records because there are a variety of experimental approaches (e.g., AF-based methods, multispecimen, Thellier) and, additionally, because there is no consensus on the best way to interpret paleointensity data. Various approaches of intensity estimation (e.g., selection criteria, best fitting, Bayesian statistics) are used and the availability of measurement level data is important as it allows reinterpretation of existing data as newer, and potentially more reliable, techniques are developed (Tauxe and Yamazaki, 2015).

Inferences about the geomagnetic field drawn from paleomagnetic data require that the age of the magnetization be known. Although many radiometric dating techniques are available, the most common method used for volcanic rocks is the $^{40}Ar/^{39}Ar$ technique, which has largely replaced the older K/Ar technique (Dickin, 2005). $^{40}Ar/^{39}Ar$ ages are reported relative to standards whose accepted ages continue to be refined and updated, so it is important to manage the paleomagnetic age records as they get updated. As age data are usually published alongside paleomagnetic records, the age may be updated but won't be reflected in the already published data. This issue could be fixed with interoperable databases that would provide the necessary data links from the database to the original records.

With every paper that is published, funding agencies and publishers require that the supporting data be accessible. This does not necessarily mean that the archived data will be complete or compatible with other published data. Researchers often publish the site averaged data instead of the many individual sample, specimen, or measurement data. The number of site level records associated with each publication typically ranges from 20 to 200 and can include the sample's site coordinates, paleointensity, paleodirections, age, standard deviations, pole coordinates, and other descriptive information about the sample and how it was collected. Many paleomagnetic data have been published over the past decades, but the level of detail varies considerably. Despite increases in current expectations about making data available, many researchers choose to archive only higher-level data such as site averages, rather than measurement level data. A substantial number of published data have been collected into a database interface, as discussed below, but improvements would maximize the usability of the data.

1.3 Databases and their Uses

Scientific research often builds on previous findings for the purposes of reproducing, merging, or reinterpreting data. This can only be done if the data are easily accessible to the greater community, accurate, and consistent across sources. Modern databases are large containers of data publicly hosted online with a basic model that allows for flexibility for available data but overall uses consistent vocabularies, presents the data in a standard format, and can be queried.

The Magnetics Information Consortium (MagIC) is an interactive online database that allows the paleomagnetic and rock magnetic community to freely upload, search, and download data with potential for reinterpretation (Tauxe et al., 2018, https://www2.earthref.org/MagIC). The comprehensive, versatile, and versioned interface allows users to upload published data and provides the paleomagnetic community the opportunity to produce an accessible archive of paleomagnetic and rock magnetic data at all levels. Other online databases focus on specific selection criteria such as age, location, or paleomagnetic data type, are not versioned, and contain limited metadata. One of the databases devoted to paleointensity data is the Paleointensity database, PINT (Bono et al., 2022, http://www.pintdb.org/). The Iceland Paleomagnetic Database, ICEPMAG, combines Icelandic paleomagnetic directional and intensity data (Tonti-Filippini and Brown, 2019, http://icepmag.org/). The GEOMAGIA50 database contains archeomagnetic/volcanic and sediment paleomagnetic and chronological data for the past 50 thousand years (Brown et al., 2015, https://geomagia.gfz-potsdam.de/index.php). Recently, a database for ^{40}Ar/^{39}Ar and K/Ar ages has been developed, called KARAR, which would archive age data and aims to work alongside other data repositories (Jarboe et al., 2021). Updates to these databases are made by the database owners and designated as new versions but often only the most recent version is available to the public.

Other data containers are essentially tabulations of datasets such as the Paleosecular Variation of the Miocene dataset, PSVM (Engbers, 2022), and the Paleosecular Variation of the past 10 million years dataset, PSV10 (Cromwell et al., 2018), which are made public as supplemental files to published studies but are solely accessible as downloadable files. Many of the data used in ICEPMAG and PINT and other databases are sourced from MagIC. The MagIC database has over 4,400 contributions (each corresponding to a publication with associated data) that contain over 229,800 individual site records. ICEPMAG contains 9,491 individual site records from 79 studies and PINT contains 4,353 individual site records collected from 413 studies (Figure 1.5). The MagIC database acts as a parent database to specialized databases such as PINT and ICEPMAG because it is capable of hosting all levels of paleomagnetic data from the measurement level (specimen and sample data) to the site level data.

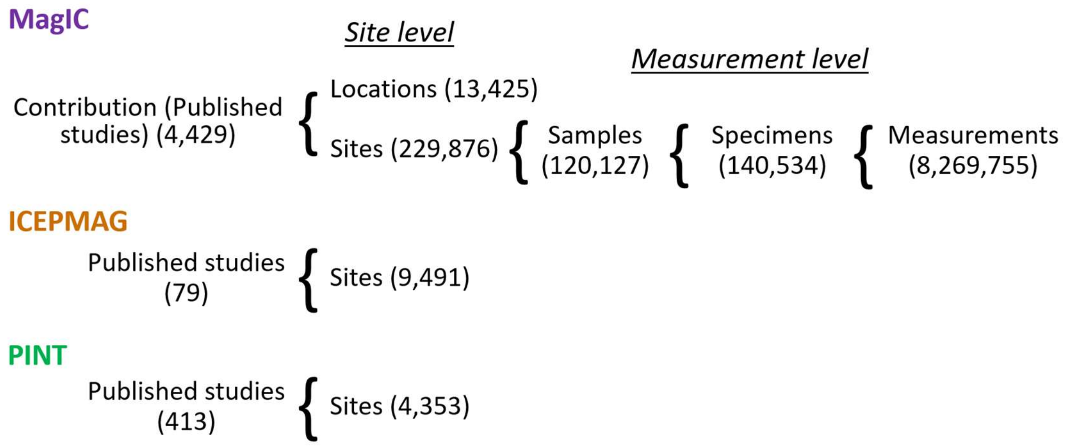


Figure 1.5: Statistics of the data presented across MagIC, ICEPMAG, PINT. In MagIC, data can be put into Locations, Sites, Samples, Specimens, and Measurements tables to indicate the level of data. In terms of the MagIC database structure, the ICEPMAG and PINT databases are essentially large tables of site level data.

In order to access the data in PINT, one could download the entire database or use the online search query to pre-select and view data before downloading (Figure 1.6). When selecting data from the online query, the data are only available to download as a text file. The entire database can only be downloaded as an .xlsx file. Additional resources for the database include files of all references, method codes, deprecated records, and other notes along with a PDF file detailing the descriptions of column headings.

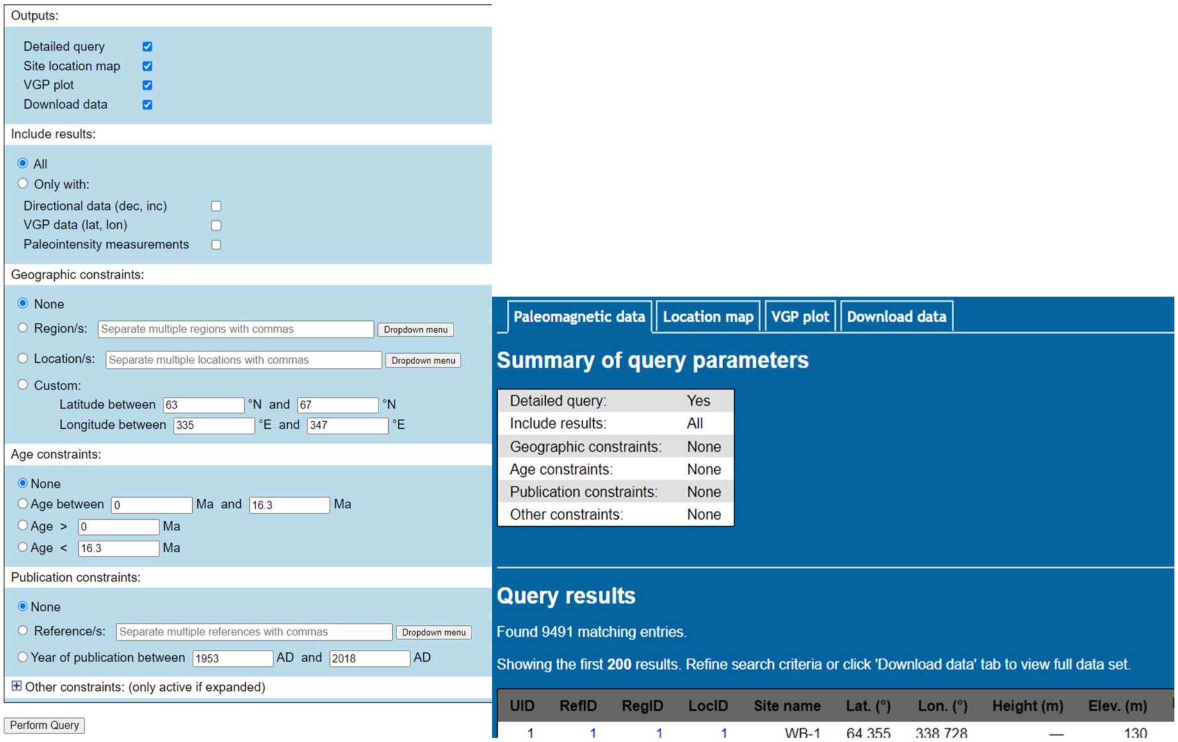
Interactive database query search engine

Download PINT v8.1.1 as Excel spreadsheet (XLSX).

☐ Min. Intensity Specimens 0
☐ Max. dB (muT) 0
☐ Max. dB % 0
☐ Min. Directional Specimens 0
☐ Max. a95 0
☐ Min. k 0

Figure 1.6: Database download options (left) and some query selections (right) from http://www.pintdb.org/.

The ICEPMAG database allows users to download data only through online query. The user may select data based on geographic, age, or publication criteria before performing the query, which will prompt the option of downloading the results from the query as a .csv file (Figure 1.7). In navigating the http://icepmag.org/index.php site, users will find additional information about included studies, credits, and external links, along with options to plot the data.



Outputs:

- Detailed query ☑
- Site location map ☑
- VGP plot ☑
- Download data ☑

Include results:

- ● All
- ○ Only with:
- Directional data (dec, inc) ☐
- VGP data (lat, lon) ☐
- Paleointensity measurements ☐

Geographic constraints:

- ● None
- ○ Region/s: Separate multiple regions with commas
- ○ Location/s: Separate multiple locations with commas
- ○ Custom:
  - Latitude between 63 °N and 67 °N
  - Longitude between 335 °E and 347 °E

Age constraints:

- ● None
- ○ Age between 0 Ma and 16.3 Ma
- ○ Age > 0 Ma
- ○ Age < 16.3 Ma

Publication constraints:

- ● None
- ○ Reference/s: Separate multiple references with commas
- ○ Year of publication between 1953 AD and 2018 AD

⊞ Other constraints: (only active if expanded)

Summary of query parameters

Detailed query:	Yes
Include results:	All
Geographic constraints:	None
Age constraints:	None
Publication constraints:	None
Other constraints:	None

Query results

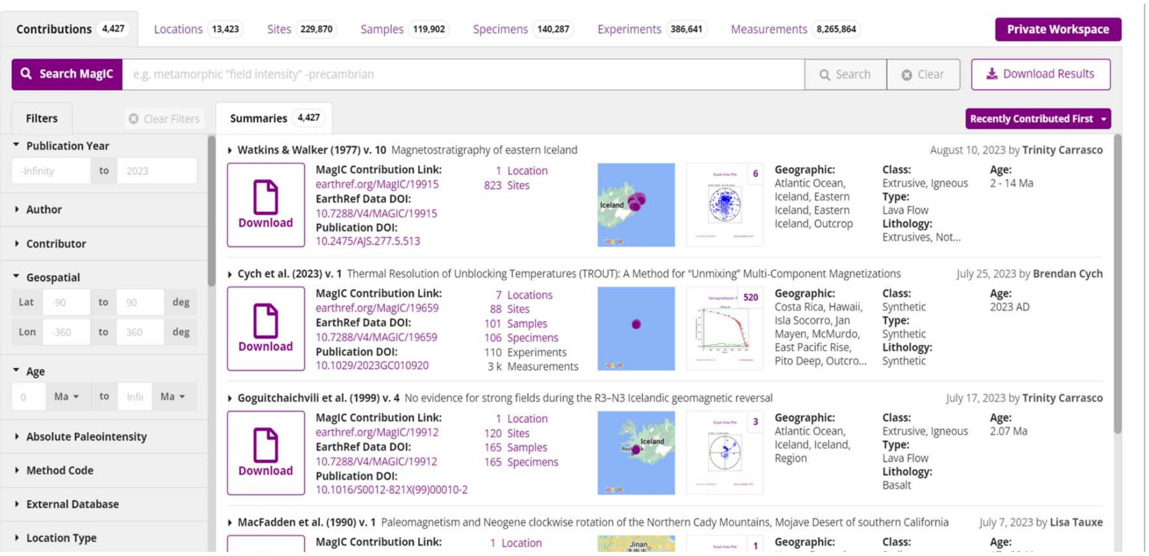
Found 9491 matching entries.

Showing the first **200** results. Refine search criteria or click 'Download data' tab to view full data set.

UID	RefID	RegID	LocID	Site name	Lat. (°)	Lon. (°)	Height (m)	Elev. (m)
1	1	1	1	WB-1	64.355	338.728	—	130

Figure 1.7: ICEPMAG query selection (left) and query results (right) from http://icepmag.org/query.php.

The MagIC interface allows the user to search, upload, or manage their contributions with resources explaining the data model, vocabularies, and external resources. Users can search for data based on an extensive list of search criteria ranging from publication details to measurement data to metadata. Users can download individual data directly using the "Download" icons or all the results from a given search using the "Download results" feature. Users can also upload data using their "Private Workspace[s]" (Figure 1.8), which allows data to be archived and be available for reviews prior to publication.



Watkins & Walker (1977) v. 10 Magnetostratigraphy of eastern Iceland — August 10, 2023 by **Trinity Carrasco**

MagIC Contribution Link: earthref.org/MagIC/19915
EarthRef Data DOI: 10.7288/V4/MAGIC/19915
Publication DOI: 10.2475/AJS.277.5.513
1 Location
823 Sites
6
Geographic: Atlantic Ocean, Iceland, Eastern Iceland, Eastern Iceland, Outcrop
Class: Extrusive, Igneous
Type: Lava Flow
Lithology: Extrusives, Not...
Age: 2 - 14 Ma

Cych et al. (2023) v. 1 Thermal Resolution of Unblocking Temperatures (TROUT): A Method for "Unmixing" Multi-Component Magnetizations — July 25, 2023 by **Brendan Cych**

MagIC Contribution Link: earthref.org/MagIC/19659
EarthRef Data DOI: 10.7288/V4/MAGIC/19659
Publication DOI: 10.1029/2023GC010920
7 Locations
88 Sites
101 Samples
106 Specimens
110 Experiments
3 k Measurements
520
Geographic: Costa Rica, Hawaii, Isla Socorro, Jan Mayen, McMurdo, East Pacific Rise, Pito Deep, Outcro...
Class: Synthetic
Type: Synthetic
Lithology: Synthetic
Age: 2023 AD

Goguitchaichvili et al. (1999) v. 4 No evidence for strong fields during the R3–N3 Icelandic geomagnetic reversal — July 17, 2023 by **Trinity Carrasco**

MagIC Contribution Link: earthref.org/MagIC/19912
EarthRef Data DOI: 10.7288/V4/MAGIC/19912
Publication DOI: 10.1016/S0012-821X(99)00010-2
1 Location
120 Sites
165 Samples
165 Specimens
3
Geographic: Atlantic Ocean, Iceland, Iceland, Region
Class: Extrusive, Igneous
Type: Lava Flow
Lithology: Basalt
Age: 2.07 Ma

MacFadden et al. (1990) v. 1 Paleomagnetism and Neogene clockwise rotation of the Northern Cady Mountains, Mojave Desert of southern California — July 7, 2023 by **Lisa Tauxe**

MagIC Contribution Link:
1 Location
1
Geographic:
Class:
Age:

Figure 1.8: MagIC search interface at https://www2.earthref.org/MagIC/search.

The PINT, ICEPMAG, and MagIC databases allow users to study specific time periods and locations with their own desired constraints on the data. It is important that, along with being able to access the data within the more specialized databases, each data record is consistent and accurate when compared to the same record in either the original study or other databases and is included in the larger parent database, MagIC.

1.4 Database Interoperability in Earth Science

Data accessibility is crucial to scientific discovery and novelty. The design of public geoscience and paleomagnetic databases is intended to organize data so that it is accessible to individual users and separate databases. Interoperability is the ability for different databases to connect and communicate in an effective way, without much effort from the user (Wilkinson et al., 2016). In order to make paleomagnetic data more useful and cross-organizational collaboration more user friendly, we need to make our databases with interoperable components such as translatable vocabularies and traceable and consistent data, as outlined by the FAIR (Findability, Accessibility, Interoperability, and Reusability) Guiding Principles (Wilkinson et al., 2016). When data are accurate, consistent, and can be identified across databases, the quality of data is improved as more sources can be integrated, the time needed to process data is minimized, and interpretations can be traced, thus increasing the community's organizational efficiency. The complications of data exchange are some of the oldest problems in data science (Kolaitis, 2005). At present the ICEPMAG, PINT, and MagIC databases share data and some metadata but lack consistency, accuracy, individual unique data record identifiers, and translatable vocabularies.

Data is structured differently between MagIC, PINT, and ICEPMAG. This requires that we define the terms used to identify the data within each database. A data record in MagIC can refer to a single line of data in any of the data tables within a contribution, but in this context, unless otherwise stated, the term "MagIC data record" will refer to a site level record. Since PINT and ICEPMAG are essentially sites tables, a "data record" within these two databases will refer to an individual site record. In PINT and ICEPMAG, a UID is assigned for every site record and is intended as a data record identifier. MagIC does not yet use a public facing individual data record identifier which hinders a user's ability to search on the basis of individual records.

Ideally, all paleomagnetic data would be correct and represent consistent results across databases. Among ICEPMAG, PINT, and MagIC, errors and mistakes in data records have been found to propagate outwards from MagIC to ICEPMAG and PINT. This indicates that researchers need to adopt methods to ensure that data are correctly recorded and that the MagIC database needs to integrate internal validations for calculated data, especially as the number of users grows.

Data identifiers are a part of the metadata for data records. An essential part of digital data next to the data itself is metadata. Metadata provides contextual information about that data and is essential to understanding paleomagnetic data records because it describes where the data came from, what was done to it, how it was dated, and other key attributes. Metadata should explicitly include a unique data identifier so that the data are findable and traceable across databases (Wilkinson et al., 2016). Unique data identifiers can be some sequence of numbers used to identify an individual data record as it is reproduced beyond its original database. There is not yet a uniform protocol to identify individual paleomagnetic records in this way. Though, in PINT and ICEPMAG, something called a unique data identifier (UID) is assigned to each site record, but only in that respective database and is not widespread. A similar type of identifier that is accessible by the user does not yet exist in MagIC. Moreover, the data within MagIC are derived from physical samples and it may be desirable to establish connections between physical samples and digital data.

Two types of existing identifiers are IGSNs and data DOIs. IGSNs, International Generic Sample Numbers (https://www.igsn.org/), provides a unique, persistent identifier for physical samples. DOIs, Digital Object Identifiers (https://www.doi.org/) are codes used for the permanent identification of digital objects are found in data repositories, but researchers may be familiar with their use to identify published studies (Wilkinson et al., 2016). The IGSN identifiers are known for tracing physical objects while data DOIs are used to identify digital data. Since paleomagnetic digital data are usually acquired from physical samples and may need to be linked to other information such as age data, something similar to both IGSN and data DOIs would be needed to consistently identify data records.

In order to support consistent metadata, database vocabularies should be able to be mapped from one database to another. The metadata for paleointensity experiments will be used as an example of where controlled vocabularies are needed. There are inconsistent uses of paleomagnetic experimental metadata in MagIC and between ICEPMAG and PINT. There needs to be a more concise and consistent use of metadata in MagIC that can map into the more specialized databases.

As new databases are developed, there is an increasing need to identify data records digitally. It may seem implied, but we must also ensure that data are correct and consistent within databases. Improving database validations, metadata, supporting a common data record identifier, and ensuring data are correct and consistent will greatly enhance current databases. These issues among current databases will be further discussed in the following chapters.

Chapter 2

The State of Current Databases

In this chapter, we will examine some of the inconsistencies across ICEPMAG, PINT, and MagIC databases. First, we will outline the MagIC data model, and the database structure of the PINT and ICEPMAG databases. We will compare the location and VGP data for individual data records between the PINT and ICEPMAG databases and investigate their inconsistencies. Lastly, we will illustrate the various uses of method codes to describe experimental protocols across MagIC, ICEPMAG, and PINT and suggest improvements to the MagIC method codes.

2.1 Database Structures

Some data and metadata are labeled differently across paleomagnetic databases. The ICEPMAG and PINT databases present all their data in one table while the MagIC database may have up to 9 tables of data for each contribution. In this section, the structure of MagIC, ICEPMAG, and PINT will be described.

2.1.1 The MagIC Data Model

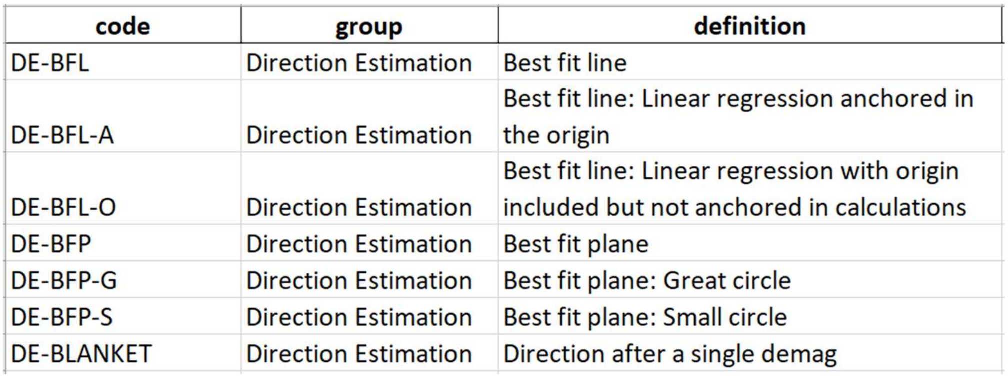
The MagIC database archives data using data files referred to as contributions. Each contribution is related to a published paper via DOIs. All contribution data files follow a data model that requires a minimum amount of data and metadata to be included in the file based on a hierarchy of data and measurements information most often gathered in paleomagnetic and rock magnetic studies. The current version of the MagIC Data Model is version 3.0 with the capability of maintaining up to 9 different data tables: Contribution (contribution metadata created after upload), Location (groups of paleomagnetic sites), Sites (rock units with common age and magnetization), Samples (samples from a unique site), Specimens (sub-samples measured), Measurements (measurements used during the analysis), Criteria (list of passing criteria), Ages (measured ages), and Images (images and plots) (Tauxe et al., 2016). Data within MagIC contributions follow this hierarchical flow of data from the measurements table up to the sites table (as shown above in Figure 1.5).

There are several pathways to searching for data in MagIC that depend on the specific needs of the user (e.g., specific age intervals, location, rock magnetic measurements, stratigraphic information, etc.). For the purposes of this study, we are interested in paleointensity and directional data which are mostly archived in the Location, Sites, and Samples tables. The Location data includes columns to describe attributes of the paleomagnetic site locations. In the Sites table, the key data columns are site name, latitude (λ_s) and longitude (ϕ_s), age estimation, units of age estimation, directional declination, directional inclination, virtual geomagnetic pole (VGP) latitude (λ_p) and longitude (ϕ_p), measured field intensity, virtual axial dipole moment (VADM), method codes, and external database identification. The Samples table may include similar data to the Sites table but for data collected from samples from individual sites.

At the time of uploading a contribution, the data model allows the contributor to decide how much of their recorded data they will upload, which in turn will dictate which of the 9 tables will be populated. The data uploaded may be from the site, sample, specimen, or measurement level. If paleointensity data records are included, they may be put into the paleointensity group of columns (absolute paleointensity, paleointensity sigma or standard deviation, the calculated VADM, and VDM) in either Site or Samples tables. The external database identification (ID) column provides the identification number of the data record in separately published databases. The external database ID may refer to additional external databases such as PINT, ICEPMAG, GEOMAGIA50, and others listed here: https://www2.earthref.org/vocabularies/controlled. Each table also includes important metadata, called method codes, that are intended to describe any methods applied to the samples during the collection, measurements, and analysis phases. Although each table requires method codes to be listed, there are not unique sets of method codes for each table; instead, there is only a large set of method codes that can be used in any table. Also, it is important to note that some, but not all, method codes follow the hierarchy of data within MagIC.

The format of MagIC method codes are short codes that describe the various methods associated with a record (Tauxe, et al., 2016). Each method code starts with the few letters that represent the main category followed by another set of letters that describe the method to varying degrees of detail (see the supplemental file "carrasco_MagIC_method_code_uses.xlsx", for the full lists of method codes and their uses). As an example, a few of the 28 method codes within the Direction Estimation (DE) main category are shown in Table 2.1. The upper method code "DE-BFL" (direction estimation using a best fit line) is more general, while "DE-BFL-A" provides more detail within that group of codes. The full list of Method Codes with their definitions and descriptions is here: https://www2.earthref.org/MagIC/method-codes.

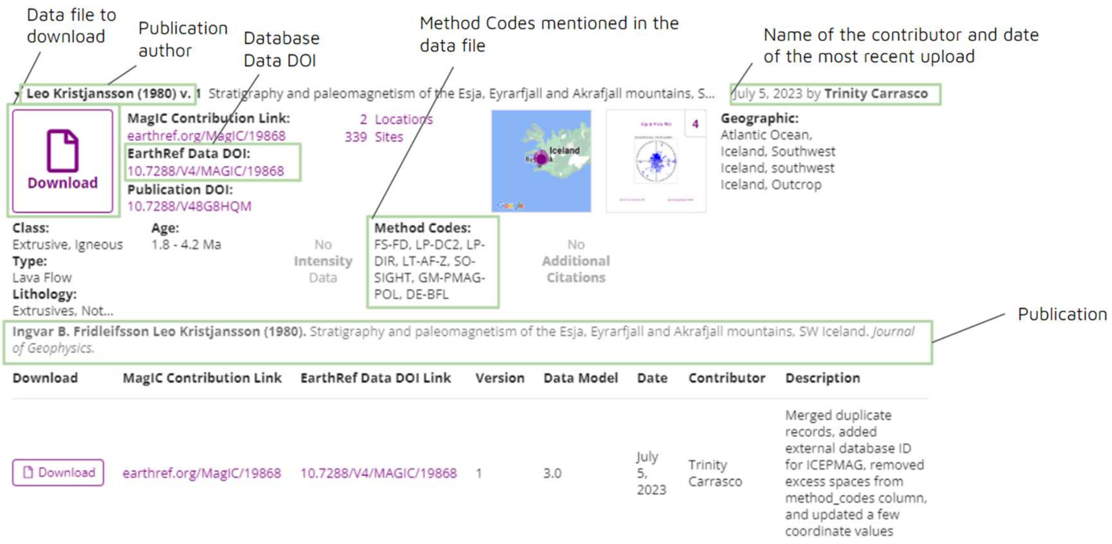
Table 2.1: A few Direction Estimation MagIC method codes and their descriptions.



code	group	definition
DE-BFL	Direction Estimation	Best fit line
DE-BFL-A	Direction Estimation	Best fit line: Linear regression anchored in the origin
DE-BFL-O	Direction Estimation	Best fit line: Linear regression with origin included but not anchored in calculations
DE-BFP	Direction Estimation	Best fit plane
DE-BFP-G	Direction Estimation	Best fit plane: Great circle
DE-BFP-S	Direction Estimation	Best fit plane: Small circle
DE-BLANKET	Direction Estimation	Direction after a single demag

Effective use of method codes throughout the entire MagIC interface is an important goal, but in many instances their usage is unclear, with either too few, too many, or inappropriate codes used that, in turn, results in failure to adequately document how the data were acquired. Along with this issue, we find that other paleomagnetic databases use different method code names and formats to describe similar laboratory methods applied to the specimens.

All published data contributions are publicly accessible for any MagIC users which allows them to search through all contributions based on publication year, title, author, contributor, age of data, study location, paleointensity values, method codes, defined external databases, and various other metadata groups. Most contributions are structured as in Figure 2.1, with publication author, year, and title at the top of the contribution page and metadata labeled below. The EarthRef Data DOI link is a digital data identifier for the data uploaded to MagIC and is also linked to the MagIC contribution. Using the download icon, users can download this data file to their personal machines. Community members can upload published data as well as updates to published data that then appear as new versions of a contribution.



Leo Kristjansson (1980) v. 1 Stratigraphy and paleomagnetism of the Esja, Eyrarfjall and Akrafjall mountains, S... July 5, 2023 by **Trinity Carrasco**

MagIC Contribution Link:
earthref.org/MagIC/19868
EarthRef Data DOI:
10.7288/V4/MAGIC/19868
Publication DOI:
10.7288/V48G8HQM

2 Locations
339 Sites

Geographic:
Atlantic Ocean,
Iceland, Southwest
Iceland, southwest
Iceland, Outcrop

Class:
Extrusive, Igneous
Type:
Lava Flow
Lithology:
Extrusives, Not...

Age:
1.8 - 4.2 Ma

No Intensity Data

Method Codes:
FS-FD, LP-DC2, LP-DIR, LT-AF-Z, SO-SIGHT, GM-PMAG-POL, DE-BFL

No Additional Citations

Ingvar B. Fridleifsson Leo Kristjansson (1980). Stratigraphy and paleomagnetism of the Esja, Eyrarfjall and Akrafjall mountains, SW Iceland. *Journal of Geophysics.*

Download	MagIC Contribution Link	EarthRef Data DOI Link	Version	Data Model	Date	Contributor	Description
Download	earthref.org/MagIC/19868	10.7288/V4/MAGIC/19868	1	3.0	July 5, 2023	Trinity Carrasco	Merged duplicate records, added external database ID for ICEPMAG, removed excess spaces from method_codes column, and updated a few coordinate values

Figure 2.1: A typical MagIC contribution with important aspects emphasized from https://www2.earthref.org/MagIC/search.

2.1.2 ICEPMAG Database Structure

The ICEPMAG database archives data into a single table with 70 data columns. The data columns we are concerned with here are the UID for each data record, reference study ID, the site name, latitude (λ_s) and longitude (ϕ_s), age estimation in millions of years (Ma), bounds of the age estimate, declination, inclination, VGP latitude (λ_p) and longitude (ϕ_p), sample ID type, alteration check ID, demagnetization type ID, paleointensity method ID, absolute intensity in μT, and intensity standard deviation. Most of the data columns here have similar requirements as in the MagIC database table columns, although UID is the ICEPMAG specific data record identifier and the metadata describing data acquisition may be quite different from the method codes in MagIC. Sample type ID, alteration check (ALT) ID, demagnetization (DEMAG) type ID, and paleointensity (PI) method ID columns relate to the data acquisition (Table 2.2). These metadata describe the field sampling, field alteration checks, demagnetization method, and paleointensity experiments applied to a sample, respectively. These aim to describe individual method codes listed in the study's MagIC contribution, but they do not exactly correspond to MagIC method codes (again, the full list of MagIC method codes is here https://www2.earthref.org/MagIC/method-codes).

Table 2.2: ICEPMAG data acquisition definitions.

SAMPLE TYPE ID

ID	NAME
1	Mini-cores (diameter 1cm)
2	Portable drill cores (diameter 2.54cm)
3	Block or hand samples
4	Existing samples (remeasurement)
5	Conventional drill cores (diameter 4.4cm)

ALT ID

ID	NAME	DESCRIPTION
-1	Unspecified	No alteration check or not specified
1	PTRM	pTRM check
2	AFARM	AF demagnetization of ARM
3	PMRM	Partial microwave induced TRM check
4	SUSC	Susceptibility does not change after successive heatings

DEMAG TYPE ID

ID	NAME	SHORT_NAME
0	None	None
1	Alternating field	AF
2	Microwave	MW
3	AF and thermal	AF-TH
4	Thermal and microwave	TH-MW
5	Thermal	TH

PI METHOD ID

ID	NAME	DESCRIPTION
-1	Unspecified	No palaeointensity measurement
1	Thellier	See Thellier and Thellier (1959)
2	ZI (Coe)	Thellier modified by Coe et al (1967)
3	Shaw	See Shaw (1974)
4	Microwave	See Walton (1993)
5	MT4	Modified Thellier type four (Leonhardt et al 2004)
6	Wilson	See Wilson (1961)
7	Van Zijl	See Van Zijl et al (1962)
8	IZZI	Modified Thellier (Tauxe and Staudigel 2004)
9	MSP-DB	Multispecimen parallel differential pTRM (Dekkers and Boehnel 2006)

2.1.3 PINT Database Structure

The PINT database archives data into a single table with 32 data columns. The columns we are interested in for this study are the UID for each data record, the old data record UID, latitude (λ_s) and longitude (ϕ_s), location metadata (location name, sampling site name from the study source, continent, and country), age estimate, age uncertainty, intensity method, inclination, declination, mean paleointensity, and paleointensity standard deviation. The older data record UID is the UID reference number from the previous version of PINT. Of course, this is not a unique and persistent identifier if it changes between versions of the database. The other columns have similar definitions to the ICEPMAG and MagIC definitions. The list of paleointensity methods is more extensive than the paleointensity experiment method list in ICEPMAG, but, again, does not translate exactly from the MagIC method codes.

2.2 An Icelandic Data Comparison

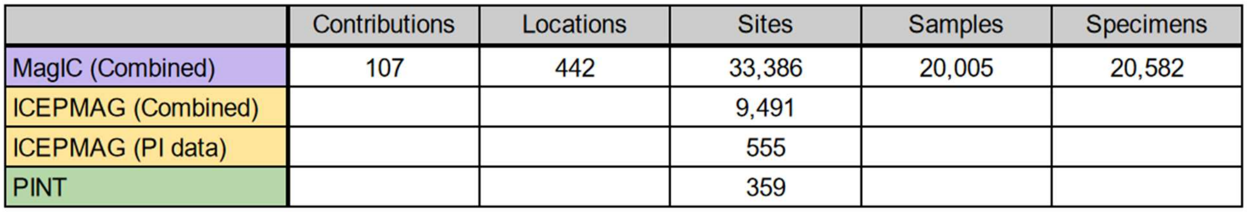
Across ICEPMAG, PINT, and MagIC there is an overlap of data since ICEPMAG and PINT source data from MagIC, but some of the data that should be the same is not. Here we will investigate inconsistencies in location data and VGP data.

2.2.1 Location Data Analysis

Using ICEPMAG and PINT databases, we compiled Icelandic paleointensity and directional data that span the past 16 million years (the complete age range for lavas exposed on Iceland). All of the data used in ICEPMAG and PINT are found in MagIC. Because of this, there is an overlap of Icelandic paleointensity site records in ICEPMAG and PINT. Only the statistics of the total Icelandic data available in MagIC were collected because there was uncertainty in the accuracy of MagIC results produced by searching for Icelandic paleointensity or directional data. For example, searching for “Iceland” in the search query results in 107 contributions while searching for Iceland using the bounding latitude and longitude results in 93 contributions. Both of these searches are essentially the same and are common examples of how researchers might want to use the function so they should yield the same results. On top of this, searching for “Iceland” while selecting the “LP-PI” filter for paleointensity experimental data results in 8 contributions which cannot be correct because PINT and ICEPMAG contain data from over 14 contributions that they sourced from MagIC. The variability of results for searches that should be equivalent undermines the present nature of the MagIC search and must be improved or, at the very least, should have a guide to using the search query.

Given the considerations with the MagIC search query, the amount of Icelandic data available in each database is shown in Table 2.3. The way that specialized databases source data from MagIC is one way data are carried across databases. In this section, we will depict the need for proper data tracking in order to maintain consistency across databases by comparing the location differences that arise for the same studies.

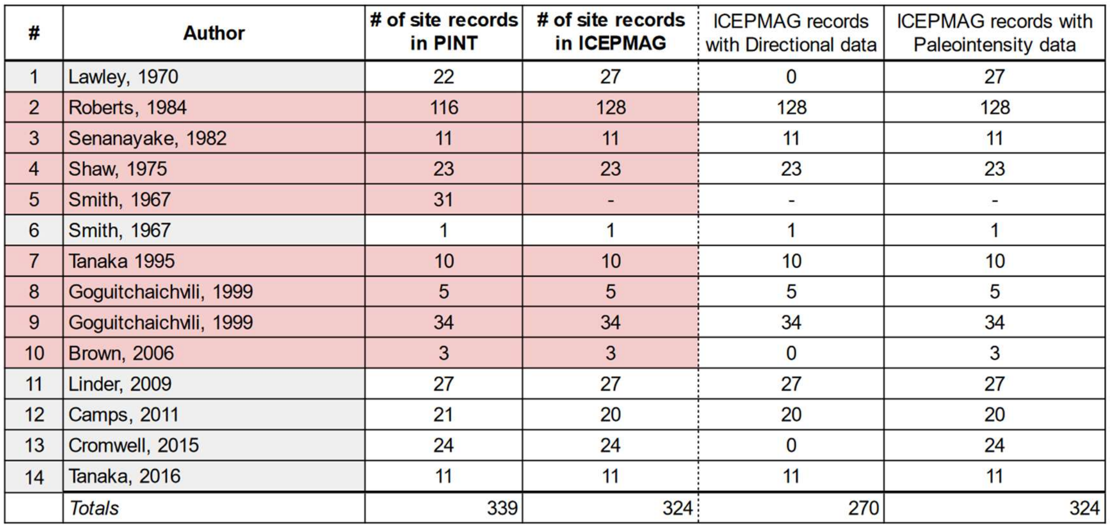
Table 2.3: Icelandic data records available in MagIC, ICEPMAG, and PINT. Note that the Icelandic data records from MagIC were produced by searching "Iceland" in search query. The table shows the number of combined paleointensity and directional data in MagIC and ICEPMAG as well as the paleointensity records in ICEPMAG and PINT. As discussed earlier, ICEPMAG and PINT only record site level data.



	Contributions	Locations	Sites	Samples	Specimens
MagIC (Combined)	107	442	33,386	20,005	20,582
ICEPMAG (Combined)			9,491		
ICEPMAG (PI data)			555		
PINT			359		

A first order check of the data consistency is the site location. There are 14 studies with paleomagnetic site data that are included in both PINT and ICEPMAG and therefore should share identical data. We find that there are multiple site location inconsistencies in 8 out of 14 shared studies, these studies are highlighted in Table 2.3. We will point out that the Smith (1967) study is plotted below (see row 5 of Table 2.4) using coordinates from PINT that appear to be chosen simply to indicate that the site was in Iceland, despite the original study not including the latitude (λ_s) and longitude (ϕ_s) coordinates. ICEPMAG, however, does not include location coordinates for this study. When line 5 is excluded from Table 2.4, the total number of records still differ between PINT and ICEPMAG for these 14 studies. The small discrepancies in the number of records across a few studies are due to the PINT database consolidating site records (e.g., several individual site records K31A, K31C, K31F were combined into one record with the site name "K31A/C/F") (Table 2.4).

Table 2.4: Shared studies between PINT and ICEPMAG with their shared amounts of site location data. Rows highlighted in pink are the studies containing errors.



#	Author	# of site records in PINT	# of site records in ICEPMAG	ICEPMAG records with Directional data	ICEPMAG records with Paleointensity data
1	Lawley, 1970	22	27	0	27
2	Roberts, 1984	116	128	128	128
3	Senanayake, 1982	11	11	11	11
4	Shaw, 1975	23	23	23	23
5	Smith, 1967	31	-	-	-
6	Smith, 1967	1	1	1	1
7	Tanaka 1995	10	10	10	10
8	Goguitchaichvili, 1999	5	5	5	5
9	Goguitchaichvili, 1999	34	34	34	34
10	Brown, 2006	3	3	0	3
11	Linder, 2009	27	27	27	27
12	Camps, 2011	21	20	20	20
13	Cromwell, 2015	24	24	0	24
14	Tanaka, 2016	11	11	11	11
	Totals	339	324	270	324

We compared each of the site location coordinate data between ICEPMAG (orange) and PINT (green) for each referenced study (Figure 2.2). If all the location data for these same studies were the same between databases, then all the orange and green circles should overlap in Figure 2.2, but they do not. Most inconsistencies are in the Northern and Western Icelandic data.

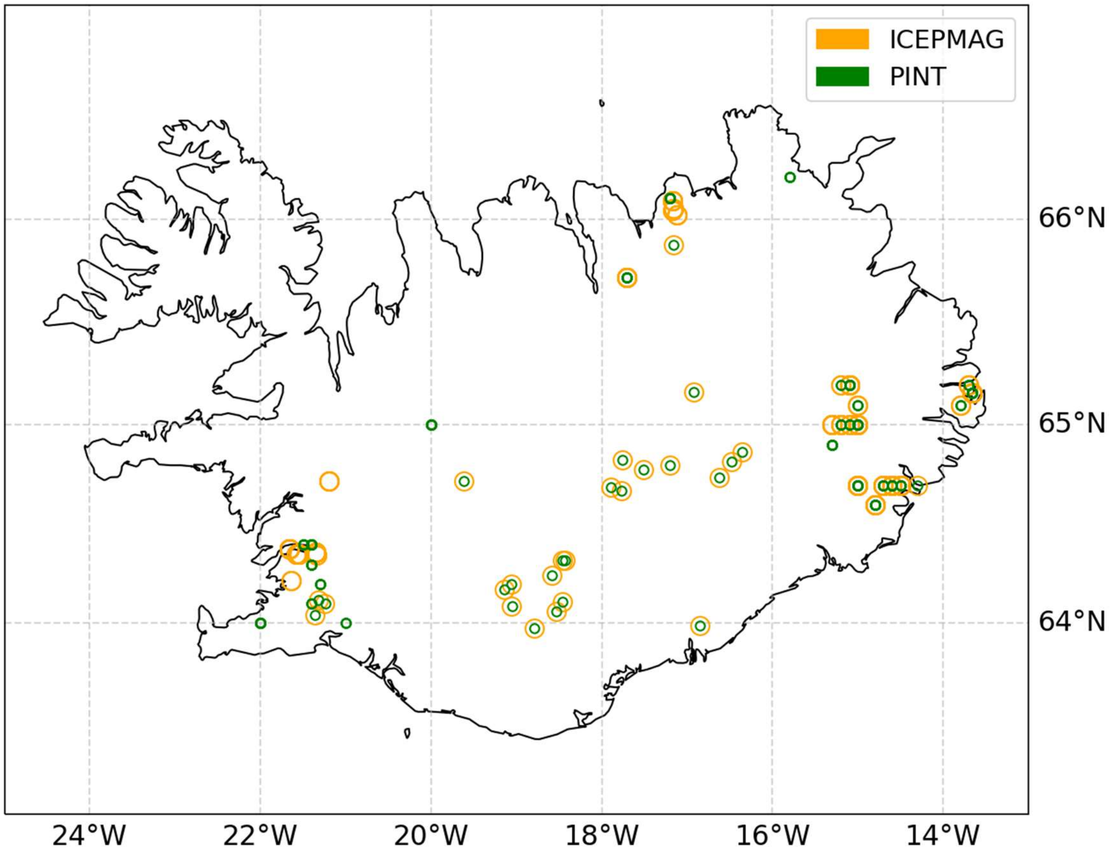


Figure 2.2: Locations of 320 sample sites included in both PINT (green) and ICEPMAG (orange) databases. A few individual site locations were not included in this plot because they were not shared between the two databases. Every green and orange circle should have a coincident pair of the other color.

Although many data overlap, we found some inconsistencies that vary by 0.1° to 0.5° (Figure 2.3), with the greatest difference being over 4° in longitude (Figure 2.3 (a)). These differences are significant because the actual site location may vary by 10-50 or over 100 km, making it difficult and almost impossible to find the original site if one were to try to resample a previous study's sample site. The issue of relocating sampling sites is not limited to Iceland, see for example, Tauxe et al. (2003). When studying any site, as well as in good science, an accurate location and sufficient precision allows the site to be located. These location errors across ICEPMAG and PINT are because the original studies did not include the exact site coordinates, or the data did not get transcribed into the databases correctly. These errors depict the importance of accurately and precisely recording the original data in our studies so that there is no need for interpretation by the reader.

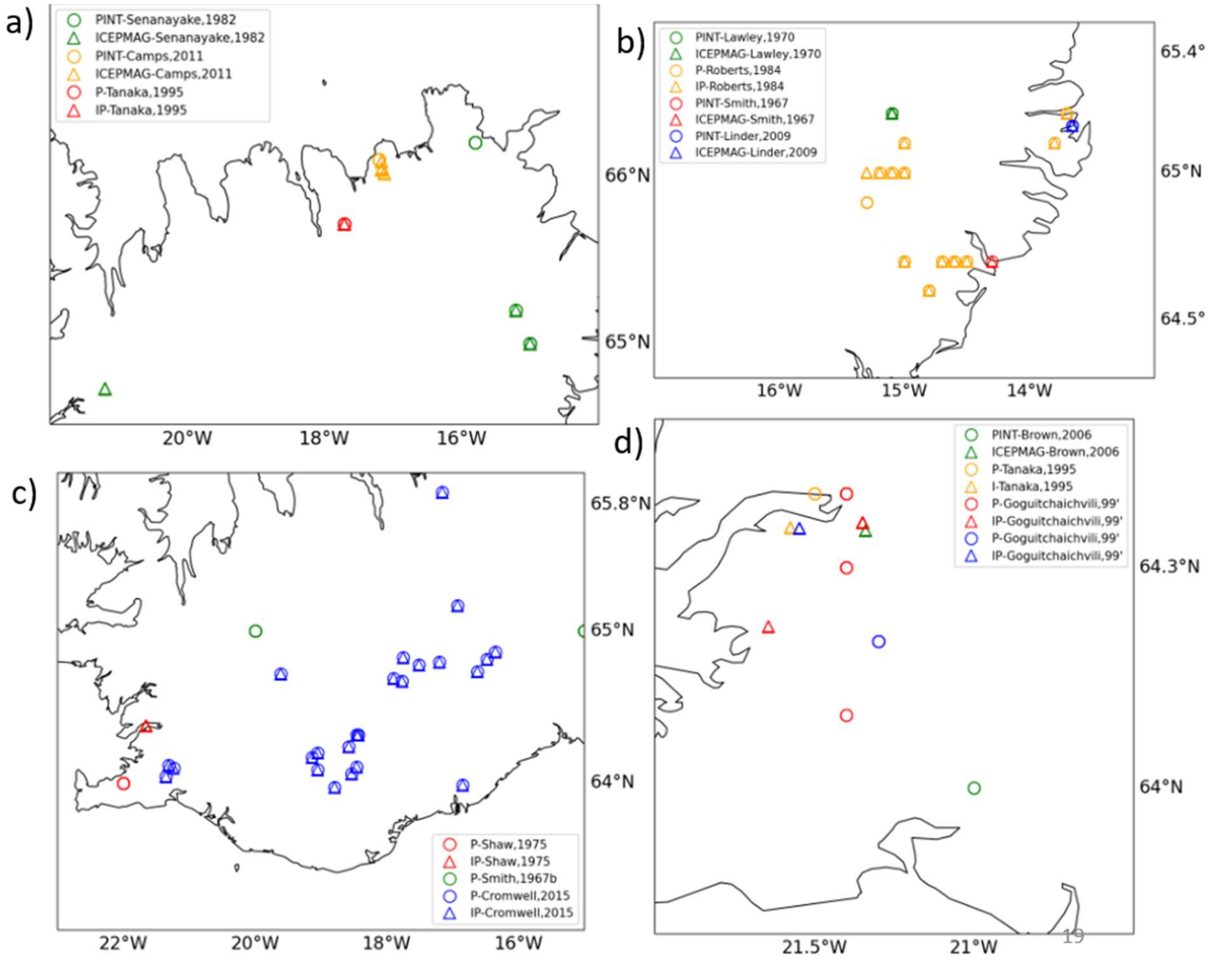


Figure 2.3: PINT (P) and ICEPMAG (IP) locations colored by reference study (Author, year). A single color indicates the study, open circles indicate the coordinates in PINT, and open triangles indicate the coordinates in ICEPMAG. In panel (c), all of the blue symbols overlap indicating the site locations for the Cromwell (2015) study are consistent while the red symbols do not, indicating an inconsistent site location between ICEPMAG and PINT for the Shaw (1975) study. In panels (a), (b), and (d), the same color symbols that do not overlap mean there were inconsistencies and those same color symbols that overlap are consistent.

Published paleomagnetic datasets need to include precise and accurate information about the paleomagnetic sampling site. When studies don't include site locations at all, readers and database curators are left to infer site locations and we risk having conflicting interpretations as shown in Figure 2.3. These interpretations can propagate to other databases or future studies that make use of either database.

One way we can ensure that we have precisely recorded data is by improving the quality of our data collection by using field mapping applications (nowadays GPS allows for improved accuracy in location). FieldMOVE Clino (https://www.petex.com/products/move-suite/digital-field-mapping) and StraboSpot2 (https://www.strabospot.org/overview) are field geology applications that can be downloaded directly to personal devices that allow real-time plotting and verification of site locations. FieldMOVE Clino acts as a compass and digital compass-clinometer to measure the orientation of geological features and most importantly, allows the user to record data measured by the application along with varying field notes associated with the given location determined from GPS. StraboSpot2 allows the user to record geographic data automatically by defining a field site as a "spot", where all notes and measurements relating to this site can be recorded. There may not be internet service at sampling sites, so each application supports uploads of reference maps. Both applications allow users to export the data recorded.

The use of these applications in the field should improve site location accuracy. In addition, there are intentions of improving the interoperability of StraboSpot with MagIC to share data (Nelson et al., 2023). As StraboSpot works with MagIC to incorporate the MagIC vocabulary and workflow to become interoperable, the two systems will help increase data accuracy and the efficiency of entering data into MagIC. It is clear that in the past there may not always have been accurate means of recording data, but going forward there is no excuse.

2.2.2 Directional Data Analysis

For our Icelandic paleodirectional analysis, we will remind the reader of the geomagnetic field approximations, coordinates, and calculations used. The geocentric axial dipole (GAD) is a time-average approximation of a dipolar field that approximates Earth's magnetic field. The calculation for the inclination of the GAD magnetic field is shown in (Equation 2.1). Inclination anomalies are directional variations from GAD and are calculated here by subtracting the GAD inclination from the observed inclination (Equation 2.2). Virtual geomagnetic pole (VGP) positions are the transformation of geomagnetic directions, declination, and inclination, into their equivalent geomagnetic pole as VGP latitude (λ_p) and longitude (ϕ_p) (see Figure 1.3).

(Equation 2.1) $$I_{GAD} = tan^{-1}(2tan(\lambda_s))$$

(Equation 2.2) $$\Delta I = I - I_{GAD}$$

Here, λ_s represents the paleomagnetic site latitude in degrees.

Along with paleointensity data, paleodirectional data can be used to characterize past geomagnetic field behavior. Iceland has played an important role in documenting past geomagnetic field variations as several thousand lava flows spanning the past 16 million years are readily accessible. Data from these lava flows can reveal the directional behavior of the field at high latitudes (λ_s) over the past 16 million years, similar to how data are modeled in Figure 1.3 (c). Indeed, paleomagnetic data from these lava flows played an important role in identifying the pattern of polarity reversals during the Neogene. The directional data gathered from ICEPMAG includes declinations, inclinations, VGP latitudes (λ_p) and longitudes (ϕ_p). Our initial step had been to look at normal polarity data, during which we encountered issues within the directional data. A dominant dipolar normal polarity field has field vectors that are similar and close to today's field. It is conventional to select VGP latitudes (λ_p) greater than 45° so that variable transitional polarities are not included. To gain an understanding of normal polarity data in ICEPMAG we selected all VGP latitudes (λ_p) greater than 0º and calculated the GAD inclination as well as the inclination anomalies relative to GAD, ΔI. The inclination anomalies relative to GAD should not be much less than -45°, but we found many points to be much more negative than expected (Figure 2.4).

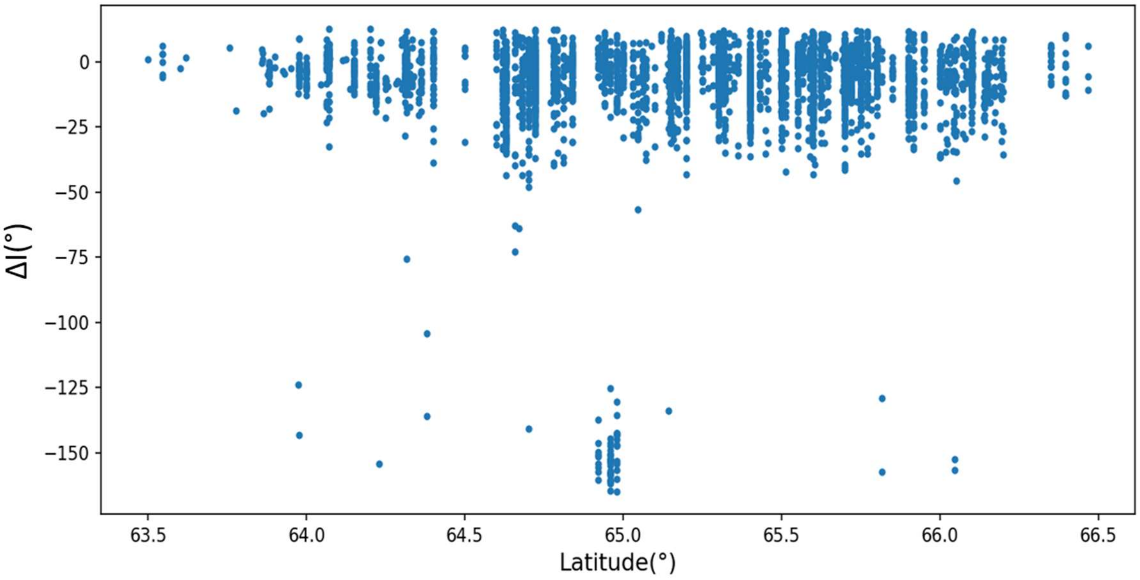


Figure 2.4: Initial ICEPMAG inclination anomalies. Inclination anomalies from GAD less than -45° are unexpected and likely of a transitional polarity. Those less than -90° are in the southern hemisphere and are from a reverse polarity.

Upon further investigation and recalculating all VGP latitudes (λ_p) from ICEPMAG we found that the VGP latitude values were inconsistent with the given declination and inclinations and that some records were assigned the wrong polarity due to the VGP latitude missing a negative sign. By selecting VGP latitudes that were labeled as normal but were not actually of normal polarity, we were seeing mislabeled polarity data by having incorrect VGP latitudes. There were 275 (of 8562) miscalculated VGP latitudes spanning 12 different studies. Before any other geomagnetic field behavior could be investigated these issues needed to be corrected.

The sources of error were determined to mostly originate from their original MagIC contributions. We were able to download each contribution and make the respective VGP latitude edits in the sites table according to the original studies. Most edits were on the order of 10 records and could be made in the MagIC contribution file and re-uploaded as an updated version of the contribution. Along with editing any incorrect data we found, we also added in the UIDs for the individual data records from the ICEPMAG database into the external database IDs column. In MagIC, an external database ID is a pair of the ID assigned to the individual data record in another database with that database name (in the form "DATABASENAME[Database data ID #]"). Including the ICEPMAG UIDs in the external database column in the MagIC contribution, makes the data in ICEPMAG traceable to the corresponding records in the MagIC parent database.

One data file in particular accounted for most of the errors, with 177 (out of 275) VGP latitude (λ_p) mistakes in ICEPMAG. This was Watkins and Walker's 1977 study which in total contains 1058 paleodirectional records in Appendix I and 830 in Appendix II; these appendices represent two different interpretations of the same data (Watkins and Walker, 1977). Data from these two appendices were inconsistently used when entered into the MagIC and PINT databases. In the MagIC contribution for this study, there were notably many missing VGPs. This is just one example of how incomplete records in MagIC propagated to a specialized database. As with the other errors, we made corrections accordingly before uploading the data to the MagIC contribution and sharing them with the ICEPMAG author (see supplemental file "carrasco_MagIC_contribution_corrections.xlsx" for the detailed list of corrections made).

After all the edits were made in the MagIC contributions and after the updates were made in ICEPMAG, we were able to recalculate the inclination residuals that look more like what was expected with most inclination anomaly values from GAD greater than -45° (Figure 2.5). Since the data from ICEPMAG were sourced from MagIC and the reference file included most EarthRef Data DOIs, we were able to trace data from MagIC to ICEPMAG and find these inconsistencies. This exercise highlights the need for accurate data archiving and data consistency at both the source of the data and as it is exported.

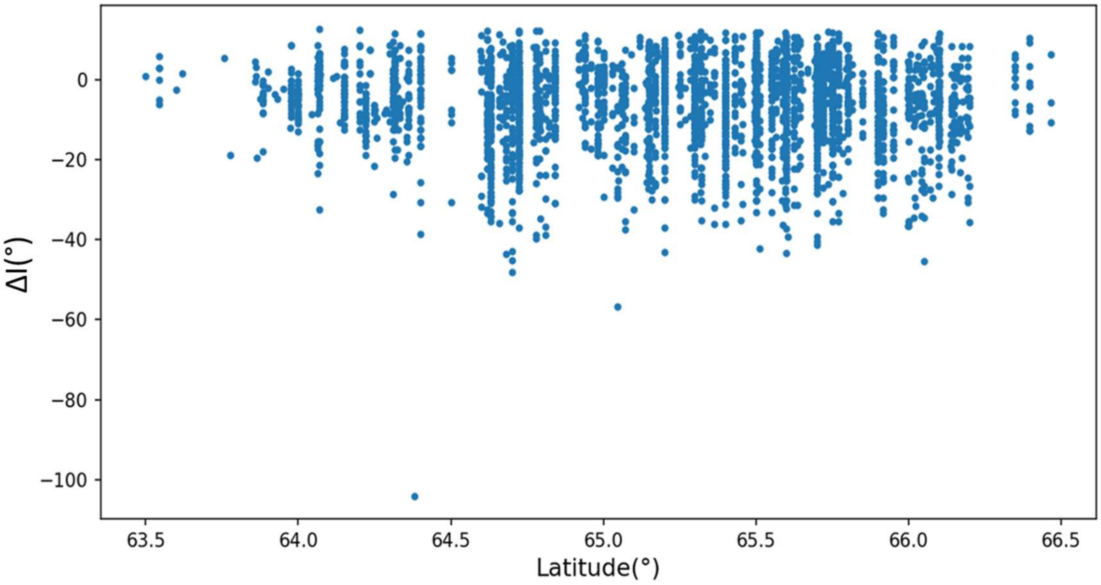


Figure 2.5: Updated ICEPMAG inclination anomalies relative to GAD. Inclination anomalies from GAD greater than -45° are expected of a normal polarity.

2.3 Analyzing Method Codes across MagIC, ICEPMAG, and PINT

The use of method codes describing paleomagnetic data acquisition is not consistent across databases. In this section we will illustrate the inconsistent use of method codes across MagIC and between MagIC and ICEPMAG and PINT. Then, we will introduce solutions to improve method code usage in MagIC and translatability of method codes to external databases.

2.3.1 MagIC Method Code Distributions

The number of method codes, their definitions, and how they are used varies across databases. In the MagIC database, there is inconsistent usage of method codes across the database and within the method code categories. We find that in the MagIC database there are many method codes that are not actively used and represent very specific details about the data that may not be broad enough to be useful. There are a total of 672 method codes with 14 categories of rock and paleomagnetic methods that describe procedures applied to the samples during data collection, measurement, and analysis processes. Of these 672 method codes, 325 have been used and only 122 of these method codes are used in at least 10 of the 4,427 contributions. The categories of method codes are displayed within Figure 2.6, representing the method codes used in at least 10 contributions. For example, none of the method codes in the "Stability Tests" category are used in more than 10 contributions so this category is not shown.

We have compiled a full list of the method codes and their current amount of uses across contributions as of August 2023 in the supplemental file "carrasco_MagIC_method_code_uses.xlsx". The present list of method codes uses in contributions can be viewed under the "Method Code" filter using the MagIC search interface, https://www2.earthref.org/MagIC/search. The most used method codes are within the Geochronology Method, Lab Treatment, Lab Protocol, and Data Estimation categories (Figure 2.6).

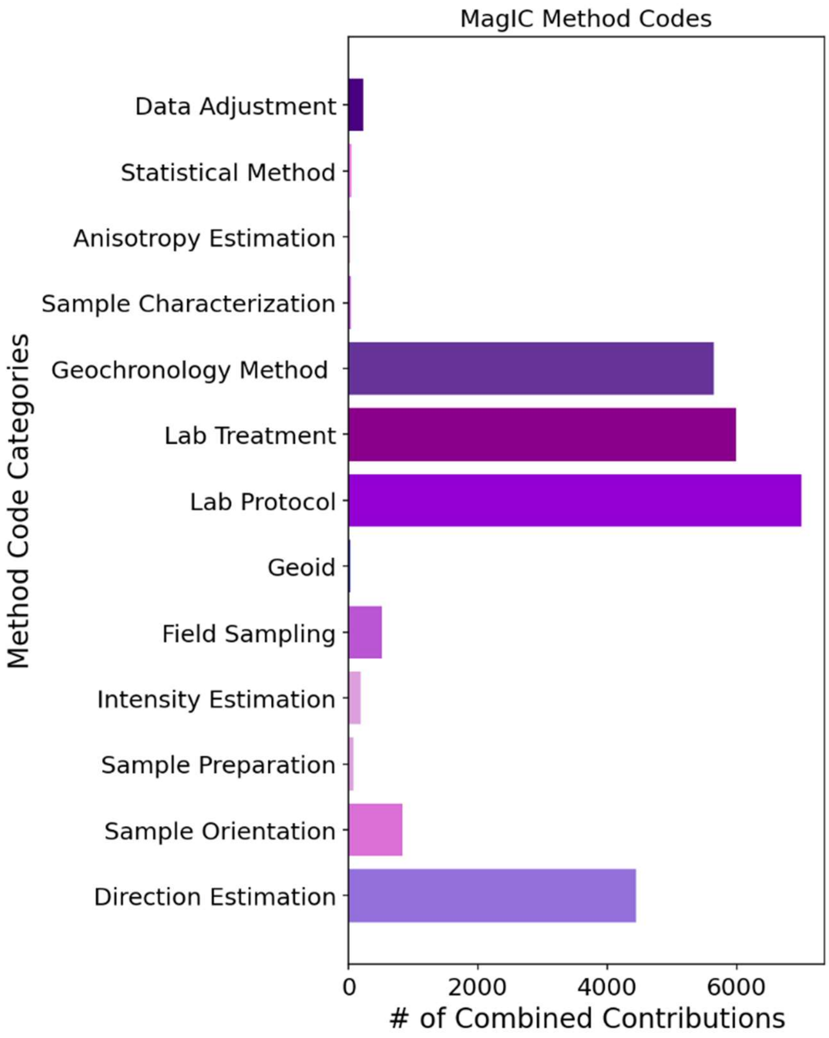


Figure 2.6: Bar chart of MagIC Method Code usage by category. Each category shown in this figure sums the number of contributions using each method code from that category. Some categories (e.g., Lab Protocol) surpass the total number of contributions in MagIC (4,429) because there may be more than one method code from that category per contribution – see Figure 2.7 for this breakdown.

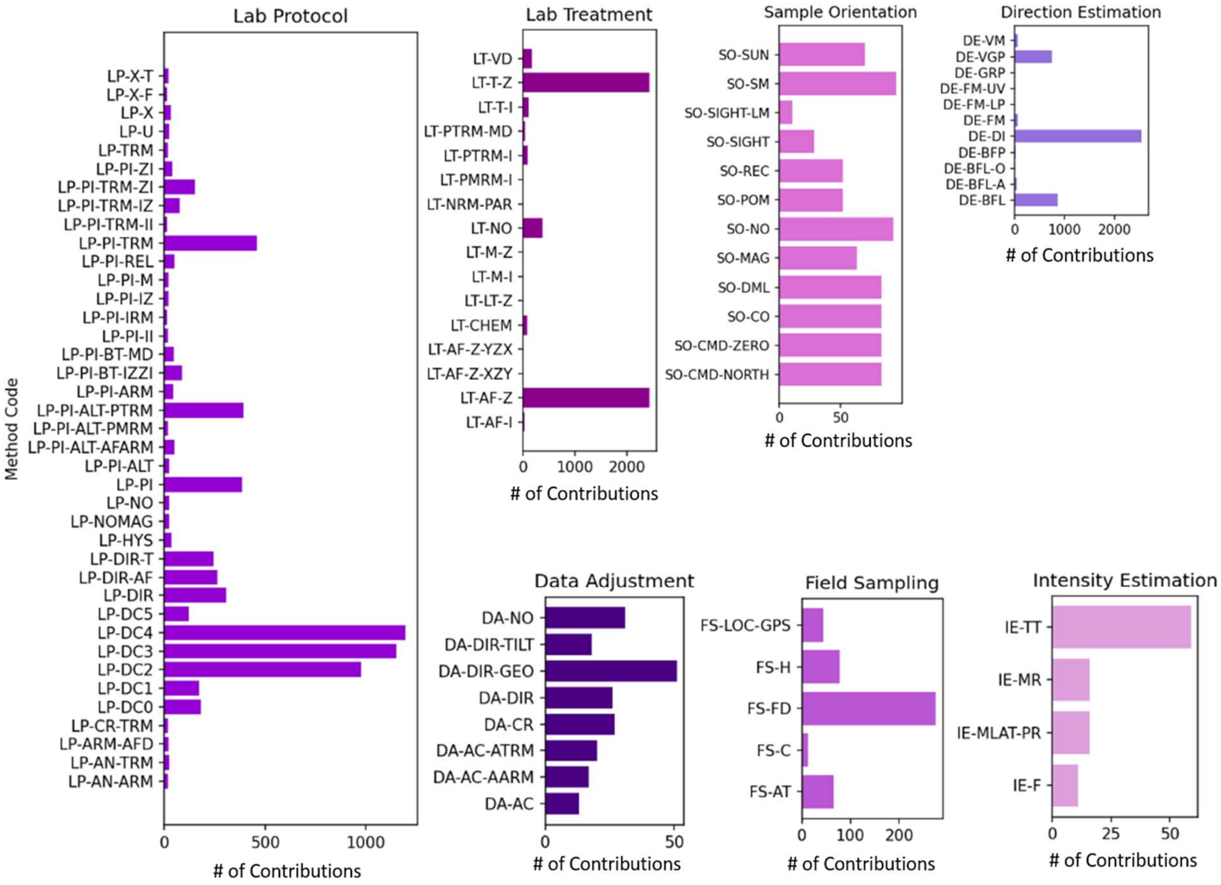


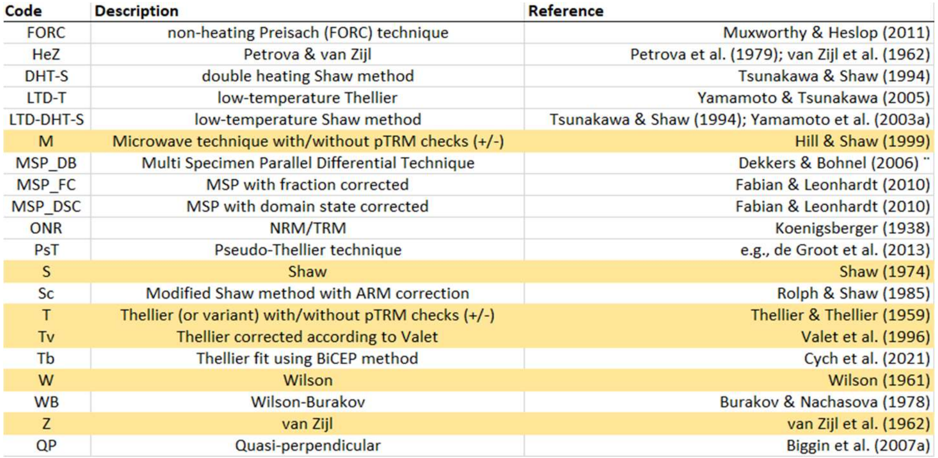
Figure 2.7: Bar charts of the most commonly used MagIC Method Codes. Each bar chart color corresponds to their Method Code sub-category in Figure 2.6. The “Lab Protocol” bar chart shows that the LP-PI method code is not very useful because using it to search does not produce the detailed codes stemming from LP-PI. With the current use of these method codes, it is only possible to search for something with very specific method codes.

The method codes shown in Figure 2.7 are the ones relevant to directional and paleointensity studies and are unevenly used, even within their subcategories. We would prefer to group together method codes in MagIC so that we have a manageable number of method codes for the most used paleomagnetic experiments. The categories that would give us insight about paleointensity experiments would be Field Sampling, Sample Orientation, Lab Protocol, and Lab Treatment. The approach for directional data would be to collect method codes that describe demagnetization (under the Lab Protocol category), Field Sampling, Sample Orientation, and Direction Estimation. Among these categories, we don’t find enough usage to effectively collect groups of method codes in this way to define paleointensity experimental methods or directional data collection methods. Due to this uneven distribution of method code usage, we propose incorporating the implied hierarchy of MagIC method codes such that all higher-level categories would be inherited (LP-PI-ALT as a child of LP-PI) and creating MagIC method codes for general experiment types. This would improve the MagIC search query and allow contributors to indicate their experiment type by the first order when uploading a contribution. In the next sections, 2.3.2 and 2.3.3, we will show how the different paleointensity experiment codes in PINT and ICEPMAG could be grouped together in a way that could be applied to MagIC to reduce the amount of variation in labeling paleointensity experiments. Then in section 2.3.4, we will discuss how data discoverability in MagIC can be improved by using the implied hierarchy of method codes, creating new method codes for common paleointensity experiments, and providing a tool to help contributors select the appropriate method codes for their contribution.

2.3.2 Method Codes in ICEPMAG and PINT

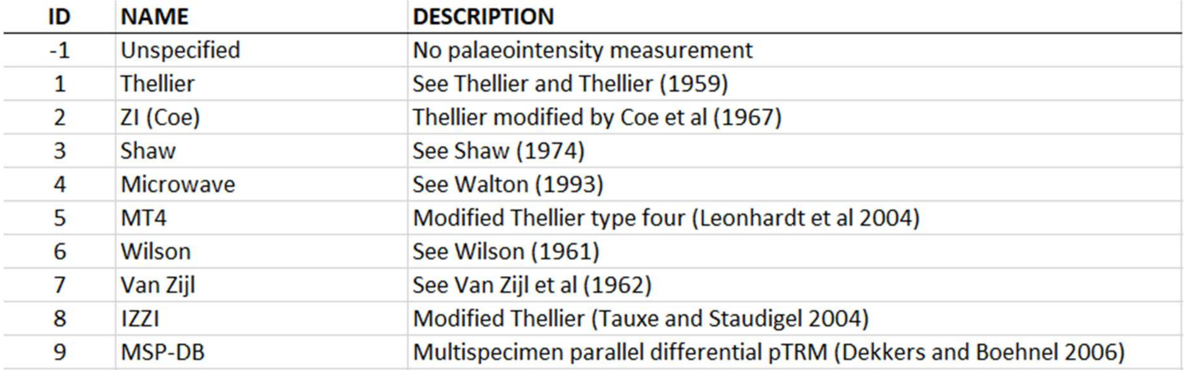
There are method codes in ICEPMAG and PINT intended to specify details of the paleointensity experiment because there are several common methods to measure paleointensity (double heating methods, Alternating Field-based methods etc.). Searching for specific kinds of experiments is useful in gaining additional information that may be used to evaluate reliability. Between ICEPMAG and PINT, the shared 14 studies (as in Section 2.2) are assigned different codes representing paleointensity experiments according to the variation in the two database's paleointensity method code vocabularies (Table 2.5, Table 2.6).

Table 2.5: PINT Paleointensity method codes with their given description and corresponding study reference with author and publishing year (Bono et al., 2022). The highlighted rows here are the method codes used in the Icelandic selection of PINT data and can be compared to the ICEPMAG paleointensity method codes – see Table 2.6.



Code	Description	Reference
FORC	non-heating Preisach (FORC) technique	Muxworthy & Heslop (2011)
HeZ	Petrova & van Zijl	Petrova et al. (1979); van Zijl et al. (1962)
DHT-S	double heating Shaw method	Tsunakawa & Shaw (1994)
LTD-T	low-temperature Thellier	Yamamoto & Tsunakawa (2005)
LTD-DHT-S	low-temperature Shaw method	Tsunakawa & Shaw (1994); Yamamoto et al. (2003a)
M	Microwave technique with/without pTRM checks (+/-)	Hill & Shaw (1999)
MSP_DB	Multi Specimen Parallel Differential Technique	Dekkers & Bohnel (2006) ¨
MSP_FC	MSP with fraction corrected	Fabian & Leonhardt (2010)
MSP_DSC	MSP with domain state corrected	Fabian & Leonhardt (2010)
ONR	NRM/TRM	Koenigsberger (1938)
PsT	Pseudo-Thellier technique	e.g., de Groot et al. (2013)
S	Shaw	Shaw (1974)
Sc	Modified Shaw method with ARM correction	Rolph & Shaw (1985)
T	Thellier (or variant) with/without pTRM checks (+/-)	Thellier & Thellier (1959)
Tv	Thellier corrected according to Valet	Valet et al. (1996)
Tb	Thellier fit using BiCEP method	Cych et al. (2021)
W	Wilson	Wilson (1961)
WB	Wilson-Burakov	Burakov & Nachasova (1978)
Z	van Zijl	van Zijl et al. (1962)
QP	Quasi-perpendicular	Biggin et al. (2007a)

Table 2.6: ICEPMAG Paleointensity method code IDs with a description that includes the reference author and publishing year (Tonti-Filippini and Brown, 2019).



ID	NAME	DESCRIPTION
-1	Unspecified	No palaeointensity measurement
1	Thellier	See Thellier and Thellier (1959)
2	ZI (Coe)	Thellier modified by Coe et al (1967)
3	Shaw	See Shaw (1974)
4	Microwave	See Walton (1993)
5	MT4	Modified Thellier type four (Leonhardt et al 2004)
6	Wilson	See Wilson (1961)
7	Van Zijl	See Van Zijl et al (1962)
8	IZZI	Modified Thellier (Tauxe and Staudigel 2004)
9	MSP-DB	Multispecimen parallel differential pTRM (Dekkers and Boehnel 2006)

The PINT and ICEPMAG databases have similar numbers of records for a few paleointensity methods for the Icelandic region in Figure 2.8. For example, the alternating field (AF) based methods such as the Shaw technique (labeled "Sh" in ICEPMAG and "S" in PINT in Figure 2.8) are assigned to a similar number of records in each database (about 140 records). In contrast, there are a variety of modifications of double heating Thellier methods ("ZI","MTh", "Th", and "IZZI" in ICEPMAG and "S;T-", "T+", "T+;Tv", and "T+;S" in PINT in Figure 2.8), leading to a more heterogeneous use of method codes.

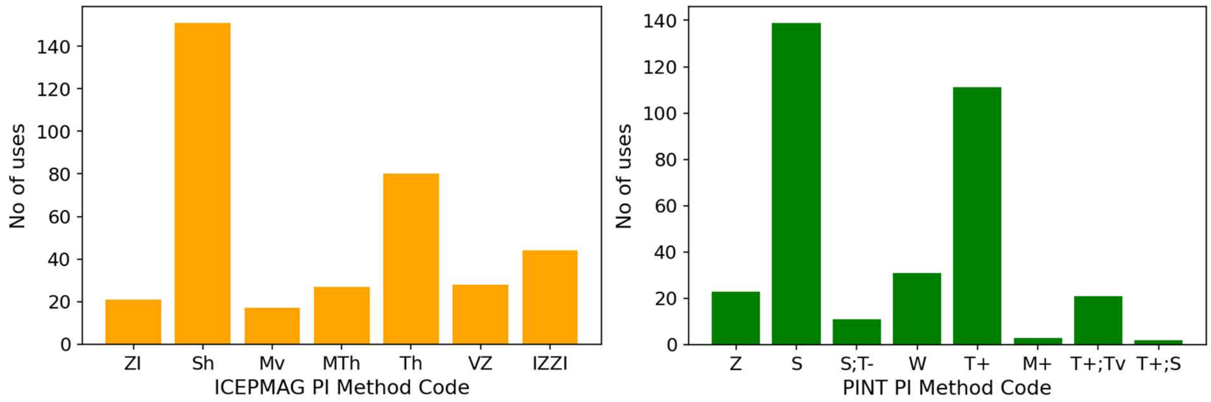


Figure 2.8: Histogram of ICEPMAG and PINT database method code uses for paleointensity data.

The method codes from ICEPMAG and PINT can be combined (along similar experiment types and experiment reliability) into a common set of method codes (referred to as "updated method codes") while still fairly representing their unique paleointensity experiment protocols (Figure 2.9, Table 2.7). For example, many researchers view double heating experiments with pTRM checks as more reliable. The T+ group collects T+ (Thellier with pTRM checks) and IZZI experiments, which also incorporates pTRM checks along with additional information on reciprocity. This means that the code "T+" in Figure 2.9 represents the sum of records in PINT with "T+" and "T+;S" method codes and in ICEPMAG's with "Th" and "IZZI" method codes from Figure 2.8. This shows that it is possible and effective to consolidate paleointensity method codes to a more general level, which could be applied in MagIC. Consolidating the method codes more globally allows database users to quickly parse through data while still understanding a broad description of the experiments used.

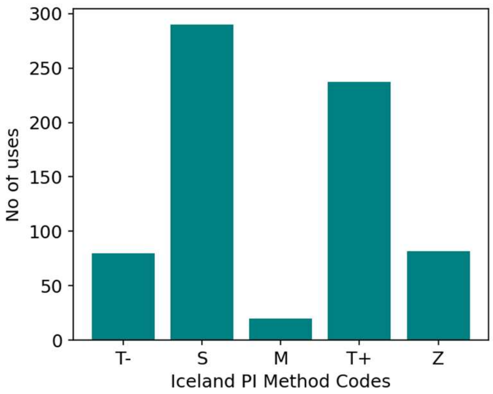


Figure 2.9: Histogram of ICEMAG and PINT paleointensity method codes with my common list of method codes.

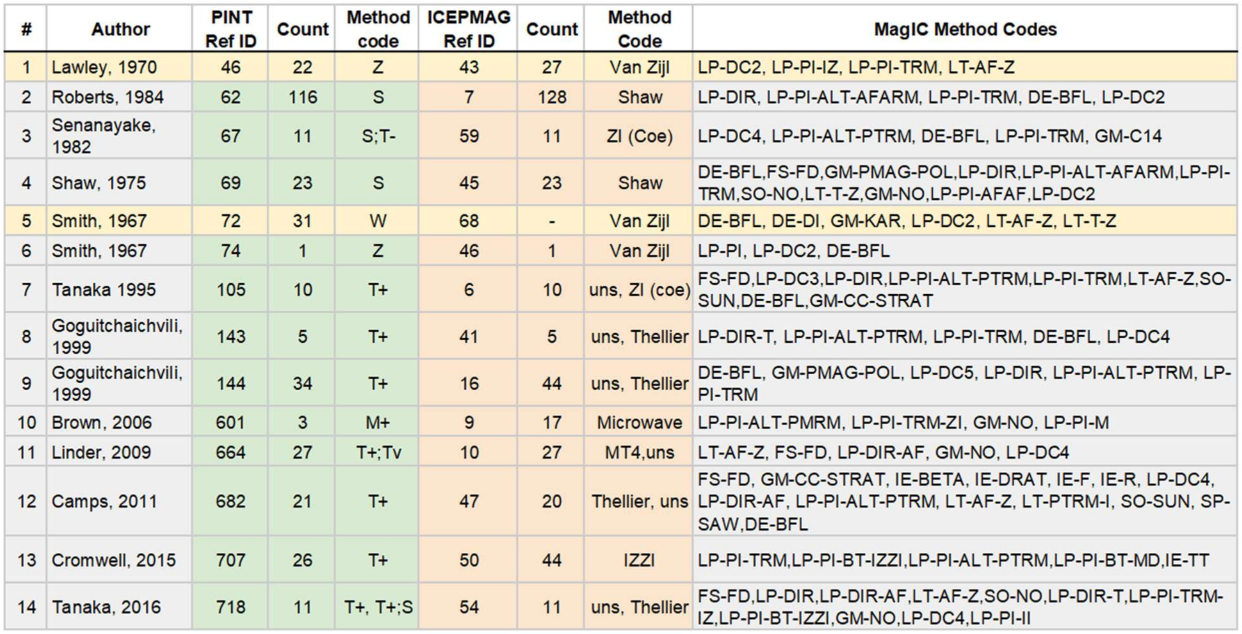
Table 2.7: My combined updated version of method codes and which database method codes they correspond to.

PINT	ICEPMAG	Updated Method Code
T+, T+;S	Thellier, IZZI	T
S;T-,T+;Tv,	ZI, MT4	T-
S	Shaw	S
M+	Microwave	M
Z,W	Van Zijl	Z

2.3.3. Comparing the Method Codes from PINT and ICEPMAG to those in MagIC

Using the overlap of 14 studies between PINT and ICEPMAG, we can compare how different method codes for paleointensity experiments are assigned in the two databases compared to how their MagIC method codes are assigned. PINT and ICEPMAG use different sets of method codes with different definitions and abbreviations (“Method code” columns in Table 2.8). In order to make a direct comparison between the databases, we compared the assigned MagIC method codes (“MagIC Method Codes" column in Table 2.8). When attempting to compare methods based on the assigned MagIC method codes, a consensus could not be made about which group of codes were associated with a given paleointensity method because there is not a consistent use of method codes. There are multiple cases where the same type of paleointensity experiment is given a different set of method codes. For example, in lines 1 and 5 in Table 2.8, the MagIC method codes are very different despite both sharing the same paleointensity method code in ICEPMAG. Presumably database developers for ICEPMAG and PINT created their own set of method codes for grouping different types of experiments that were more practical for their search criteria. The amount of variation among labeling paleointensity methods not just in PINT and ICEPMAG, but also in MagIC, shows the need for a commonly defined set of paleointensity experiment vocabulary. This can be accomplished with improvements made to the MagIC method codes.

Table 2.8: Study references used in both PINT and ICEPMAG database. Column definitions: Ref ID– the study's reference ID number in that database, Count– the number of records the method code was assigned to, Method Code– method code assigned in the respective database. See tables 2.3 and 2.4 for Method code descriptions for PINT and ICEPMAG.



#	Author	PINT Ref ID	Count	Method code	ICEPMAG Ref ID	Count	Method Code	MagIC Method Codes
1	Lawley, 1970	46	22	Z	43	27	Van Zijl	LP-DC2, LP-PI-IZ, LP-PI-TRM, LT-AF-Z
2	Roberts, 1984	62	116	S	7	128	Shaw	LP-DIR, LP-PI-ALT-AFARM, LP-PI-TRM, DE-BFL, LP-DC2
3	Senanayake, 1982	67	11	S;T-	59	11	ZI (Coe)	LP-DC4, LP-PI-ALT-PTRM, DE-BFL, LP-PI-TRM, GM-C14
4	Shaw, 1975	69	23	S	45	23	Shaw	DE-BFL,FS-FD,GM-PMAG-POL,LP-DIR,LP-PI-ALT-AFARM,LP-PI-TRM,SO-NO,LT-T-Z,GM-NO,LP-PI-AFAF,LP-DC2
5	Smith, 1967	72	31	W	68	-	Van Zijl	DE-BFL, DE-DI, GM-KAR, LP-DC2, LT-AF-Z, LT-T-Z
6	Smith, 1967	74	1	Z	46	1	Van Zijl	LP-PI, LP-DC2, DE-BFL
7	Tanaka 1995	105	10	T+	6	10	uns, ZI (coe)	FS-FD,LP-DC3,LP-DIR,LP-PI-ALT-PTRM,LP-PI-TRM,LT-AF-Z,SO-SUN,DE-BFL,GM-CC-STRAT
8	Goguitchaichvili, 1999	143	5	T+	41	5	uns, Thellier	LP-DIR-T, LP-PI-ALT-PTRM, LP-PI-TRM, DE-BFL, LP-DC4
9	Goguitchaichvili, 1999	144	34	T+	16	44	uns, Thellier	DE-BFL, GM-PMAG-POL, LP-DC5, LP-DIR, LP-PI-ALT-PTRM, LP-PI-TRM
10	Brown, 2006	601	3	M+	9	17	Microwave	LP-PI-ALT-PMRM, LP-PI-TRM-ZI, GM-NO, LP-PI-M
11	Linder, 2009	664	27	T+;Tv	10	27	MT4,uns	LT-AF-Z, FS-FD, LP-DIR-AF, GM-NO, LP-DC4
12	Camps, 2011	682	21	T+	47	20	Thellier, uns	FS-FD, GM-CC-STRAT, IE-BETA, IE-DRAT, IE-F, IE-R, LP-DC4, LP-DIR-AF, LP-PI-ALT-PTRM, LT-AF-Z, LT-PTRM-I, SO-SUN, SP-SAW,DE-BFL
13	Cromwell, 2015	707	26	T+	50	44	IZZI	LP-PI-TRM,LP-PI-BT-IZZI,LP-PI-ALT-PTRM,LP-PI-BT-MD,IE-TT
14	Tanaka, 2016	718	11	T+, T+;S	54	11	uns, Thellier	FS-FD,LP-DIR,LP-DIR-AF,LT-AF-Z,SO-NO,LP-DIR-T,LP-PI-TRM-IZ,LP-PI-BT-IZZI,GM-NO,LP-DC4,LP-PI-II

2.3.4 MagIC Method Code Suggestions

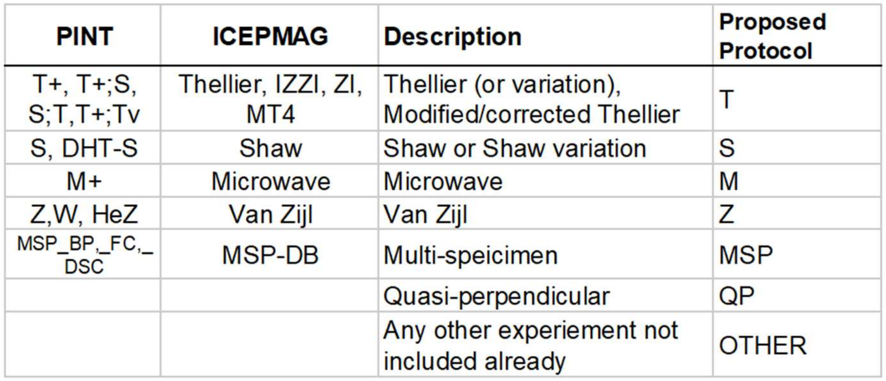
The inconsistent use of MagIC method codes across PINT and ICEPMAG suggests that method codes should be updated to improve the utility of the codes and their translatability to other databases. We expect that the community would make better use of the method codes if we (1) leverage the existing method code categories and subcategories to establish a true hierarchy in which all levels would be available in the search query and when contributors label data records in contributions, (2) implement new method codes that represent general experiments and relate them to the detailed method codes that characterize each experiment, and (3) create a tool to help researchers determine the best method codes to list in their data file contributions.

Out of the 672 MagIC method codes, 325 of them are not used in any contributions so far, making some appear to be unnecessary but efforts should be made to incorporate them by making use of the implied hierarchal order of method codes. For example, method codes could be related to their “parent” method codes (which do not formally exist) as a part of a true hierarchy. As an example, for the method code LP-DIR-AF (directional data collected from single stepwise alternating field demagnetization), the corresponding parent method code would be LP-DIR (directional data). Including both in a contribution data file would help users search for general directional data while providing the option to search for more details on the acquisition of directional data. Labeling data records in contributions with their parent method code and their more detailed method code improves the discoverability of paleointensity, directional, and all other data available in MagIC.

Presently, many contributions in MagIC do not regularly make use of the implied hierarchical system of method codes that would require higher level (parent) method codes to accompany the more detailed ones. In many contributions, the child method code (LP-DIR-AF) is often listed without its parent method code (LP-DIR) or vice versa. As an example, the parent method code LP-PI (lab protocol for a paleointensity experiment) has children method codes that describe detailed aspects of paleointensity experiments, including LP-PI-AFAF (double AF demagnetization paleointensity experiments), LP-PI-ALT (paleointensity experiment alteration check), and many others. Contributors often select method codes at these more detailed levels, but the parent method code (LP-PI) is not automatically included, leading to ambiguity in search results. On the other hand, if a contributor only uses the LP-PI code, there is room to include more detail about the paleointensity experiment with other method codes, so would it not then be better to include both the more detailed code and parent method code? This would allow to search for all data in the parent category in which a search for "LP-PI" returns all paleointensity data and searching for "LP-PI, LP-PI-ALT" returns all paleointensity data that used alteration checks. This would account for many of the unused method codes. The danger with the current system is that users will never know if they have found all the LP-PI studies when using this search. The ambiguity of including parent method codes in contributions should be addressed and with this, there is the possibility to create a function within MagIC that automatically includes a method code's parent method code if it is not already there.

Researchers often refer to paleointensity experimental methods to determine the credibility of the data. In the MagIC search interface, it is currently not possible to search for specific experiments by name such as a Thellier or Shaw experiment. One way we may be able to search for data based on the paleointensity experiment would be to implement a hierarchical method code group that defines general paleointensity experiments (Table 2.9).

Table 2.9: Proposed hierarchical MagIC method codes to be added under the Lab Protocol paleointensity experiment (LP-PI) category and how they would correspond with other database codes.

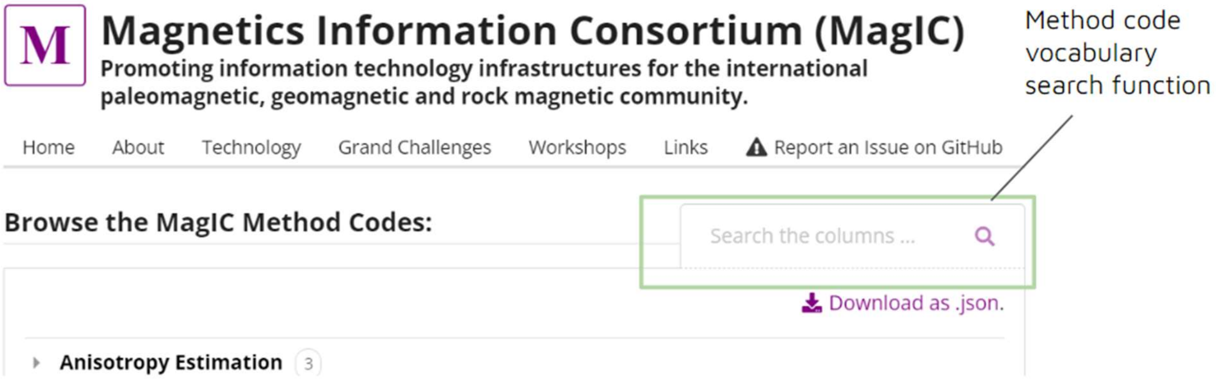


PINT	ICEPMAG	Description	Proposed Protocol
T+, T+;S, S;T,T+;Tv	Thellier, IZZI, ZI, MT4	Thellier (or variation), Modified/corrected Thellier	T
S, DHT-S	Shaw	Shaw or Shaw variation	S
M+	Microwave	Microwave	M
Z,W, HeZ	Van Zijl	Van Zijl	Z
MSP_BP,_FC,_DSC	MSP-DB	Multi-speicimen	MSP
		Quasi-perpendicular	QP
		Any other experiement not included already	OTHER

These method codes would allow users to easily recognize the general paleointensity method (e.g., Thellier type of experiments) that was applied to the data and have the option of investigating the other method codes listed beyond the "LP-PI-Th-…" (e.g., IZZI experiment). This would be more useful than attempting to find a general experiment by searching for a group of several method codes that still might not represent the desired method if a given method code was not listed in the contribution. This still allows for the user to search through more detailed codes for data if they choose to but would make it easier for them to search on a surface level for general paleointensity experiments.

The MagIC developers should consider the necessity of method codes with minimal usage, make use of the implied hierarchy of method codes, and remove truly unnecessary codes that do not have potential for future usage. The decision on lesser used method codes is essential before developing a tool to help users select method codes for their data. A method codes tool guide will improve the researcher's experience of selecting method codes for their paleomagnetic data upload as well as the user experience of searching for paleomagnetic data in MagIC by filtering with method codes. There is potential for all method codes to be useful if users know how to correctly implement them.

An essential tool that is missing from the MagIC interface is an aid for users to determine the best method codes to label their data that goes beyond the search function that is available at present (Figure 2.10). Part of the reason that half of the current MagIC method codes are being underused, (Figure 2.6), could be attributed to users being unaware of all the options they have for method codes. We propose adding a tool to the Method Codes page on the MagIC interface that would help users determine which method codes they should include in their contribution. This form would ask a series of questions (on the scale of 5-15 questions), relating to the type of paleomagnetic or rock magnetic experimentation, collection, and data processing applied to the data they wish to upload and would return the list of method codes that should be mentioned in the data file. The key to the success of this form would be making it brief enough that scientists would realistically spend time using it while still being all encompassing and informative.



Magnetics Information Consortium (MagIC)

Promoting information technology infrastructures for the international paleomagnetic, geomagnetic and rock magnetic community.

Browse the MagIC Method Codes:

Method code vocabulary search function

Figure 2.10: MagIC interface Method codes page as of 8/14/2023 with search functionality highlighted from https://www2.earthref.org/MagIC/method-codes.

Chapter 3
Interoperability in Action

Along with being accurate and consistent, it is crucial that data records be reliable and findable across databases. In this chapter, we will discuss methods for improving data reliability with additional validations and more frequent database updates. We will discuss a set of functions that were created to support data traceability and we will also address the need for a common unique data identifier across databases. At the end of this chapter, a list of bullet points summarizes the suggestions to make paleomagnetic databases more interoperable.

3. 1 Suggestions for Interoperability

3.1.1 Validations for Calculated Data

The design of the MagIC database is to archive and allow discovery of data. In cases where mistakes or issues are found in uploaded MagIC contributions, it is efficient that edits can be made with transparency as contribution versions. There is only so much we can do to prevent natural human error, but MagIC should implement validations for calculated data to reduce the amount of error in VGP coordinates and other calculated data.

In our analysis of location data across the same studies in PINT and ICEPMAG in section 2.2.1, we found inconsistencies in data that should be the same. This emphasizes the importance of precise and accurate data.

In our directional analysis from Section 2.2.2, we found that errors had propagated from MagIC into ICEPMAG causing an incorrect selection of VGP coordinates. Upon making corrections to the data records in the MagIC contributions, we were able to make the correct VGP latitude (λ_p) selection. With the most negative lone point accounted for from Figure 2.5 as a VGP latitude value that was just slightly greater than 0°, we were able to select the VGP latitude values to be greater than 45° (the traditional benchmark between transitional and normal polarity data). At this selection, we find inclination anomalies from GAD that are more useful because they are correct (Figure 3.1).

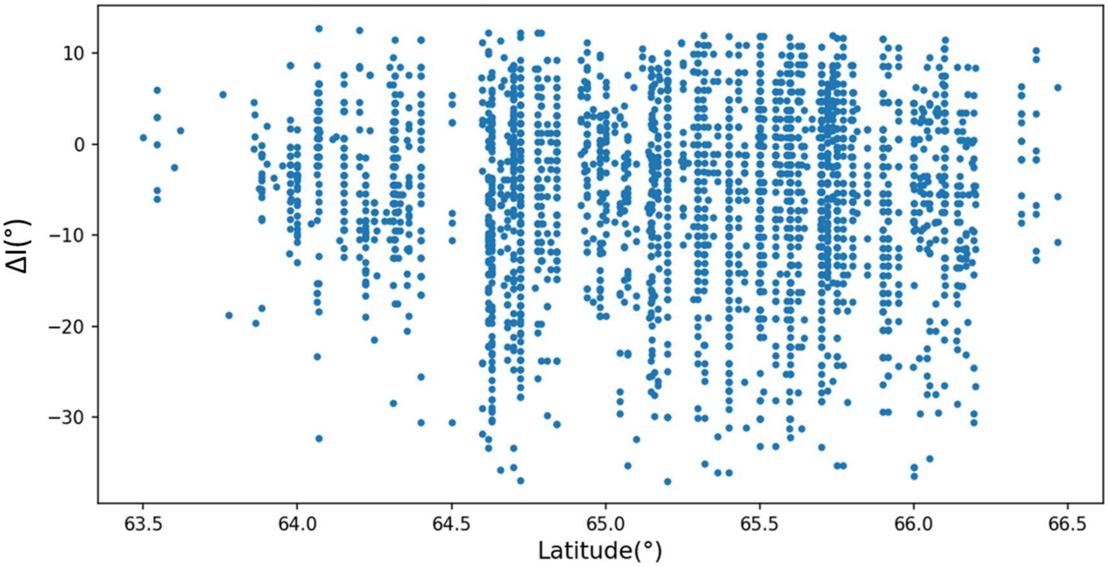


Figure 3.1: Updated ICEPMAG inclination anomalies relative to GAD with VGP latitude values greater than 45° selected, representing records with normal polarity from Iceland for the past 16 million years.

The location and directional data analyses showed how data propagates across databases and the importance of accurate data archiving and data consistency. Validations for calculated data in MagIC would help to support accurate data archiving.

The MagIC interface would benefit from an additional validation feature that reviews data file contributions for internal consistency. MagIC would accept the contribution upload, given that it successfully passes the primary validations, then make a duplicate of the file with VGP latitude (λ_p) and longitude (ϕ_p) coordinate calculations and would compare them to the original values that were uploaded. If a significant difference was found, say 1-2°, between those in the uploaded file and those calculated by MagIC, then in the online MagIC contribution (somewhere near the various versions at the bottom of Figure 2.1) there would be the option to download the original file or the file with calculated VGP coordinates with a note detailing how many VGP values were corrected. This could also be applied to other paleomagnetic calculations in the contribution data file such as paleointensity VADM or VDM. Another option would be to return the file to the contributor during the validation process. This internal consistency inspection upon uploading into the MagIC database would benefit the reliability of data. This would be one strategy to implement on the widely accessible database scale but there may also be tactics that specialized database authors might consider.

3.1.2 Specialized Database Updates

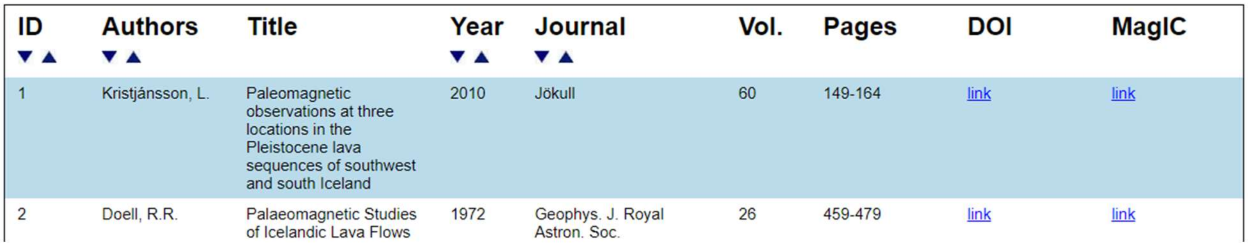
Specialized paleomagnetic databases, including ICEPMAG and PINT as discussed in this thesis, are useful in that they focus on certain selections of data and often source some or all of their data from MagIC. As we strive to enhance our current databases, database authors may consider assessing their strategies for implementing database updates. Having correct and complete external database links would be a simple solution. One suggestion would be to implement an automatic updating feature that would notify database authors when changes are made to the sources used to compile their database or potentially update on their own by directly sourcing the data. Another way to ensure specialized databases are up to date would be to schedule periodic review to check if any of the database sources have been updated. These strategies would help to ensure database accuracy, longevity, and integrity.

3.1.3 Data Record Traceability with DOIs and Data DOIs

Enhancing the consistency and accuracy of paleomagnetic databases improves data. We can further improve databases by assigning unique database identifiers that trace individual data records back to their source and across databases. As discussed in Section 1.4, unique data identifiers can be some sequence of numbers used to identify an individual data record as it is reproduced beyond its original database. Currently, there is not a unique data identifier to trace individual paleomagnetic site records across databases (an issue that will be further discussed in the next section), so other unique characteristics of individual data need to be used as identifiers.

The reference lists in ICEPMAG and PINT outline the studies included in databases by including reference details such as authors, paper title, year of publication, journal, and journal location where the study was published, DOI link, and MagIC EarthRef Data DOI link (Table 3.1).

Table 3.1: The first two rows of the ICEPMAG reference table.



ID ▼▲	Authors ▼▲	Title	Year ▼▲	Journal ▼▲	Vol.	Pages	DOI	MagIC
1	Kristjánsson, L.	Paleomagnetic observations at three locations in the Pleistocene lava sequences of southwest and south Iceland	2010	Jökull	60	149-164	link	link
2	Doell, R.R.	Palaeomagnetic Studies of Icelandic Lava Flows	1972	Geophys. J. Royal Astron. Soc.	26	459-479	link	link

Although ICEPMAG and PINT include supplemental reference lists of the studies compiled within the database, they do not include all the study DOIs, particularly in PINT. The DOI links are used to identify published studies used in these databases, but they only identify collections of data (not the individual data records), and their use is not consistent. This present method of identifying individual data records needs to be supported by database owners until a common individual data identifier is found.

The DOI is typically used to identify published papers while the MagIC EarthRef Data DOI identifies a dataset in the MagIC database. One issue that can arise in these reference files is that some or most of the DOI and/or the MagIC EarthRef Data DOI links are not provided or not viable, such as the case with PINT that only had 36 DOIs and no functional EarthRef Data DOIs out of the 413 included studies, which makes it difficult for the user to find the original source.

Examples of why researchers would need to trace and be sure of the original source of data within a database could be to certify the quality of the data, determine if it is up to date, or perhaps corroborate method codes. With the large number of references in many databases, finding all the DOI and Data DOI links is daunting but is necessary for the completeness of the database.

To assist paleomagnetic database authors and researchers with collecting study DOIs and MagIC EarthRef Data DOI links we have written functions in Python using the Crossref (https://github.com/sckott/habanero) and EarthRef/FIESTA (https://api.earthref.org/v1#tag/Public-Data/operation/v1PublicDownloadFiles) APIs, Application Programming Interfaces, to collect these published paper DOIs and data links.

Instructions on how to use these functions are in a Jupyter notebook, made public on the GitHub page: https://github.com/tcarrasc/DOI_MagIC_link_search (see the supplemental file "carrasco_jpytrnb_how_to_search.pdf" for a PDF of this Jupyter notebook). This notebook walks through an example on how to automate the collection of DOI links and MagIC EarthRef Data DOI links. The PINT reference file is used as the main example to show how one can start with a file of references that is mostly barren of study DOIs and data DOI links and proceed to filling them out by about 80% and 50%, respectively. Populating the reference file with all the DOIs and Data DOIs was not achievable with these functions because there were mistakes in the study name in the PINT references file, the published study may not have been uploaded into Crossref, or the MagIC EarthRef Data DOI link did not exist (or was still queued for publishing). Upon downloading the repository (the folder containing the example notebook and necessary functions from GitHub), researchers can populate their own list of references quickly and robustly while achieving a degree of interoperability. Using APIs, we were able to populate the

PINT database reference file with study DOIs and data DOI links. This exercise re-establishes the importance of including publication DOI and MagIC EarthRef data DOI links as a step towards data traceability.

3.1.4 The Future of Data Record Identifiers

We have shown the importance of data consistency, accuracy, traceability, and the need for simplification of method code metadata. As data moves across paleomagnetic databases we would hope that these data are accurate and consistent. To ensure traceability and that a given record is the same as cited by another database and to the original source of data, we should be able to trace it using a unique data identifier. There is a need for a unique identifier according to this definition, despite PINT and ICEPMAG each having something called a unique data identifier (UID) assigned to each site record. This UID poses as a hinderance since it is not widespread across all databases so more information is still needed to connect the record to its original study. In both databases, the site name where paleomagnetic samples were taken and the publication information about the study can be used to trace the individual record to its original study. This way of identifying individual data requires two pieces of metadata and is particularly inefficient when study identifiers are not included in the database, as in PINT. While a similar type of data record identifier is yet to exist in MagIC. Publication DOIs refer to a published paper but it is also possible to get a DOI for a dataset (for example, the MagIC EarthRef Data DOIs attached to contributions). Overall, the community needs a common data record identifier for individual paleomagnetic records (that may refer to both digital and physical data) and consistent usage of DOIs as study identifiers.

There are two types of existing data identifiers that can be assigned to physical data samples or digital data. To remind the reader of the definitions mentioned in Section 1.4: IGSNs, International Generic Sample Numbers (https://www.igsn.org/), provides a unique, persistent identifier for physical samples and associated data while DOIs, Digital Object Identifiers (https://www.doi.org/) are codes used for the permanent identification of digital objects such as data repositories. Since paleomagnetic data are often a combination of physical and digital data, something similar to both IGSN and data DOIs would be needed to consistently identify data records.

There is more depth to this issue in MagIC as each contribution can contain multiple tables of data. A logical approach to data identifiers within a contribution would ensure that the data identifier follows the hierarchy of the MagIC data model tables. For example, a UID assigned to a record in the Measurements table would need to propagate up to the Sites table in a sensible way. Every sample that is taken would need a unique identifier and should relate to the site record identifier, perhaps in a parent-child relation such that the site would be given a primary unique identifier and the sample would be a derivative of that primary site identifier. A unique identifier could be supported by sampling hierarchies for different materials. For igneous rocks, researchers would need a unique identifier to relate everything from the same cooling unit while with sediments it would need to relate everything to a sediment horizon. If a unique identifier for paleomagnetic data was defined by MagIC, other databases could directly link an individual data record back to, not only the MagIC contribution, but to individual site or measurement level records in MagIC. This is a concept that should be further considered by MagIC so that data are able to be traced with confidence as they expand interoperable abilities to other databases, including age databases.

An upcoming data repository named KARAR was announced for the ^{40}Ar/^{39}Ar and K/Ar geochronology community with the intention of being interoperable with MagIC (Jarboe et al., 2021). The idea would be that the data in MagIC with age information would be cited to the KARAR data repository. In order for this idea to work effectively we would need to consider how it will be traced across MagIC, KARAR, and any other sources that utilize the data. It is important that the data in MagIC have an effective method of being traced, that is well implemented so that contributors have a clear definition of how they should be used, to avoid the disruption to workflows between databases.

3.2 Future Work and Conclusions

We will briefly remind the reader of the suggestions given in this thesis for making the ICEPMAG, MagIC, and PINT databases interoperable:

- To ensure researchers record precise site locations for paleomagnetic field work, they are encouraged to make use of field mapping applications such as FieldMOVE Clino or StraboSpot2 (https://www.petex.com/products/move-suite/digital-field-mapping/, https://www.strabospot.org/overview).
- To ensure users receive consistent results from the search query, MagIC needs to improve the search function so that regional names and regional selections in the bounding box produce the same results and/or provide a guide to using the search query.
- To make better use of the MagIC search query for paleomagnetic data, we suggest implementing the implied hierarchy of method codes into practice, such that all higher-level method code categories would be inherited when they are used in contributions and made available for user searches. Some method codes (e.g., the 347 method codes that are currently not used in any MagIC contributions) seem unnecessary and should be implemented or removed to improve the researcher's experience of selecting appropriate method codes for their data. Upper-level method codes on the hierarchy of MagIC

method codes should be implemented to allow for the ability to search for common paleointensity experiments without using several method codes (Table 2.9).

- Adding an interactive tool on the MagIC Method Codes page would help users determine which method codes they should include in their contribution based on a brief form about rock magnetic experimentation, collection, and data processing.
- To aid in the accuracy of information in data files, MagIC should consider including validations for calculated data, such as VGP coordinates, VADM, and VDMs.
- Specialized database owners should implement automatic updates or timely reminders to check for updates to source data.
- To aid in the traceability of records and usage in APIs, we need to ensure that data studies are correctly linked using publication DOIs and Data DOIs links, such as the EarthRef Data DOIs.
- The community should decide on a unique individual record identification method in an effort to effectively trace individual records across databases.

To make the most use of data repositories, efforts are required to ensure that data are accessible, accurate and consistent, and described by useful amounts of metadata that are translatable to other databases. Efforts are required from users to upload the most correct data and use metadata such as method codes as defined by the MagIC database. Efforts from database owners include implementing validation efforts to ensure accurate data are being uploaded, simplified method codes, and providing methods for tracing data as changes are made within the database or as it is reinterpreted by others. As we adopt more interoperable methods and utilize interoperable tools, we can maximize the utility of our paleomagnetic data.

BIBLIOGRAPHY

Bono, R. K., Paterson, G. A., van der Boon, A., Engbers, Y. A., Michael Grappone, J., Handford, B., Hawkins, L. M. A., Lloyd, S. J., Sprain, C. J., Thallner, D., & Biggin, A. J. (2022). The PINT database: A definitive compilation of absolute palaeomagnetic intensity determinations since 4 billion years ago. *Geophysical Journal International*, 229(1), 522–545. https://doi.org/10.1093/gji/ggab490

Brown, M. C., Donadini, F., Korte, M., Nilsson, A., Korhonen, K., Lodge, A., Lengyel, S. N., & Constable, C. G. (2015). GEOMAGIA50.v3: 1. general structure and modifications to the archeological and volcanic database. *Earth, Planets and Space*, 67(1), 83. https://doi.org/10.1186/s40623-015-0232-0

Butler, R. F. (1992) *Magnetic Domains to Geologic Terranes.* Blackwell Scientific Publications. https://www.geo.arizona.edu/Paleomag/

Cromwell, G., Tauxe, L., and Halldórsson, S. A. (2015). New paleointensity results from rapidly cooled Icelandic lavas: Implications for Arctic geomagnetic field strength. *Journal of Geophysical Research: Solid Earth*, 120, 2913–2934. https://doi.org/10.1002/2014JB011828

Cromwell, G., Johnson, C. L., Tauxe, L., Constable, C. G., & Jarboe, N. A. (2018). PSV10: A global data set for 0–10 Ma time-averaged field and paleosecular variation studies. *Geochemistry, Geophysics, Geosystems*, 19, 1533–1558. https://doi.org/10.1002/2017GC007318

Cych, B., Tauxe, L., Cromwell, G., Sinton, J., & Koppers, A. A. P. (2023). Changes in non-dipolar field structure over the plio-pleistocene: New paleointensity results from Hawai'i compared to global data sets. *Journal of Geophysical Research: Solid Earth*, 128, e2023JB026492. https://doi.org/10.1029/2023JB026492

Dickin, A. P. (2005) *Radiogenic Isotope Geology, Second Edition*. Cambridge University Press. https://doi.org/10.1017/CBO9781139165150

Engbers, Y. A., Bono, R. K., & Biggin, A. J. (2022). PSVM: A global database for the Miocene indicating elevated paleosecular variation relative to the last 10 Myrs. *Geochemistry, Geophysics, Geosystems*, 23, e2022GC010480. https://doi.org/10.1029/2022GC010480

Jarboe, N., Minnett, R., Koppers, A., Sprain, C., & Renne, P. (2021). *Introducing KARAR: A FAIR (Findable, Accessible, Interoperable, and Reproducible) ^{40}Ar/^{39}Ar and K/Ar Geochronology Data Repository Utilizing the FIESTA Domain Repository System*. Paper presented at AGU Fall Meeting, New Orleans, LA. id. V25B-0111. 2021AGUFM.V25B0111J

Koymans, M. R., vanHinsbergen, D. J., Pastor-Galán, D., Vaes, B., & Langereis, C. G. (2020). Towards FAIR paleomagnetic data management through Paleomagnetism.org 2.0. *Geochemistry, Geophysics, Geosystems*, 21, e2019GC008838. https://doi.org/10.1029/2019GC008838

Kolaitis, P. G. (2005). *Schema Mappings, Data Exchange, and Metadata Management*. Paper presented at PODS '05: Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems, New York, NY, 61-75. https://doi.org/10.1145/1065167.1065176.

Merrill, R. T., and McFadden, P. L. (1995). Dynamo theory and paleomagnetism, *Journal of Geophysical Research: Solid Earth*, 100(B1), 317–326, https://doi.org/10.1029/94JB02361.

Nelson, E.M., Tikoff, B., Newman, J., Walker, J.D., (2023). *Incorporating the StraboSpot Ecosystem into Rock and Paleomagnetic Data Collection Practices*, Poster presented at MagIC Workshop, La Jolla, CA, https://earthref.org/events/MAGIC/2023/2023MagICVolume.pdf

Shaw, J. (1975). Strong Geomagnetic Fields during a single Icelandic Polarity Transition. *Geophysical Journal International*, 40(3), 345–350. https://doi.org/10.1111/j.1365-246X.1975.tb04136.x

Smith, P. J. (1967). The Intensity of the Tertiary Geomagnetic Field. *Geophysical Journal International*, 12(3), 239–258. https://doi.org/10.1111/j.1365-246X.1967.tb03120.x

Tauxe, L., Banerjee, S.K., Butler, R.F., & van der Voo, R. (2018) *Essentials of Paleomagnetism, 5th Web Edition*, La Jolla, CA: University of California Press. https://earthref.org/MagIC/books/Tauxe/Essentials/#x1-140002

Tauxe, L., Shaar, R., Jonestrask, L., Swanson-Hysell, N. L., Minnett, R., Koppers, A. a. P., Constable, C. G., Jarboe, N., Gaastra, K., & Fairchild, L. (2016). PmagPy: Software package for paleomagnetic data analysis and a bridge to the Magnetics Information Consortium (MagIC) Database. *Geochemistry, Geophysics, Geosystems*, 17(6), 2450–2463. https://doi.org/10.1002/2016GC006307

Tauxe, L., & Yamazaki, T. (2015). Paleointensities. In *Treatise on Geophysics* (2nd ed., Vol. 5, pp. 461–509). https://doi.org/10.1016/B978-0-444-53802-4.00107-X

Tauxe, L., Constable, C. G., Johnson, C. L., Koppers, A. A. P., Miller, W. R., & Staudigel, H. (2003). Paleomagnetism of the Southwestern U.S.A. Recorded by 0-5 Ma Igneous Rocks. *Geochemistry, Geophysics, Geosystems*, 4(4). https://doi.org/10.1029/2002GC000343

Tonti-Filippini, J. A. D., & Brown, M. C. (2019). The Iceland Palaeomagnetism Database (ICEPMAG). *Earth, Planets and Space*, 71(1), 83. https://doi.org/10.1186/s40623-019-1060-4.

Walker, M. (2005). *Quaternary Dating Methods*, New York: Wiley.

Watkins, N. D., & Walker, G. P. L. (1977). Magnetostratigraphy of Eastern Iceland. *American Journal of Science*, 277(5), 513–584. https://doi.org/10.2475/ajs.277.5.513

Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J., Growth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. https://doi.org/10.1038/sdata.2016.18