

Integrating Data Rescue into the Classroom

CIARA RYAN, CATRIONA DUFFY, CIARAN BRODERICK, PETER W. THORNE, MARY CURLEY,
SÉAMUS WALSH, CONOR DALY, MAIRÉAD TREANOR, AND CONOR MURPHY

Historical observations are fundamental for advancing understanding of past changes in climate. Fundamentally, what you do not observe, you cannot understand. Globally, however, the temporal extent of digitized records to this day remains grossly incomplete, with many records available in image or hardcopy formats only (Allan et al. 2011; Brunet and Jones 2011). The task of digitizing these data is daunting, both logistically and in terms of the sheer volume of records that still require rescue.

The capacity to extend current observational data holdings is largely dependent on the resources available to carry out the digitization and transcription process. Some records have been, and continue to be, rescued via professional keying, but this practice is far from sufficient to key all known records. Engagement of nonexperts or “citizen scientists” on a voluntary basis has become increasingly integral to the rescue and refinement of observational data across multiple scientific disciplines (Bonney et al. 2014). The success of ongoing citizen science applications—for example,

OldWeather.org (www.oldweather.org/) and Data Rescue@Home (www.data-rescue-at-home.org)—underscores the potential of crowdsourcing as a data rescue strategy.

To date, however, the potential of university students to engage in data rescue activities for credit has not been systematically explored. Students represent a pool of interested, qualified, and instructor-guided talent that could, if harnessed, significantly accelerate progress in historical data rescue. Student–scientist partnerships such as The Global Learning and Observations to Benefit the Environment (GLOBE) program are collaborations that engage students, teachers, and scientists in authentic research projects delivering context-rich, hands-on approaches to science (Mitchell et al. 2017; Allan et al. 2011). Students gain important insights into processes underpinning research, while scientists gain access to otherwise unobtainable information (Vitone et al. 2016; Harnik and Ross 2003).

The present study describes a novel research-led accredited assignment as part of the geography program at Maynooth University in Ireland, in which final-year undergraduate students successfully transcribed more than 1,300 station years of daily precipitation data and associated metadata for the period 1860–1939. The student-rescued data relate to pre-1940 (post-1940 is already digitized) daily rainfall observations from 43 stations across Ireland. Most records commence in the early 1900s; however, a few extend into the 1800s. The earliest record transcribed commences in 1866, with the shortest in 1927. In this article we explore i) the potential for integrating data-rescue activities into the classroom, ii) the ability of students to produce reliable transcriptions, and iii) the learning outcomes for students. We provide an overview of the assignment, its management, and the student aids implemented. Details of the workflow employed to evaluate student performance and facilitate the creation of a “corrected” data series are provided. In the final section, we discuss the student’s perceptions of

AFFILIATIONS: RYAN—Irish Climate Analysis and Research Units, Department of Geography, Maynooth University, Maynooth, Co. Kildare, and Climatology and Observations Division, Met Éireann, Dublin, Ireland; DUFFY, BRODERICK, THORNE, AND MURPHY—Irish Climate Analysis and Research Units, Department of Geography, Maynooth University, Maynooth, Co. Kildare, Ireland; CURLEY, WALSH, AND DALY—Climatology and Observations Division, Met Éireann, Dublin, Ireland; TREANOR—Library, Met Éireann, and Oireachtas Library, Houses of the Oireachtas, Dublin, Ireland

CORRESPONDING AUTHOR: Ciara Ryan, ciarapryan@gmail.com

The abstract for this article can be found in this issue, following the table of contents.

DOI:10.1175/BAMS-D-17-0147.1

A supplement to this article is available online (10.1175/BAMS-D-17-0147.2).

©2018 American Meteorological Society

For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#).

the project and reflect upon learning outcomes. Our experience demonstrates the substantial potential to extend such approaches to other universities, thus providing an enriched learning experience for the students and a lasting legacy to the scientific community. Project resources, including transcription templates, MATLAB code, and student aids, are provided as online supplemental material.

OVERVIEW OF ASSIGNMENT AND STUDENT AIDS. The assignment was set in the historical climatology component of the final-year climate change course that focuses on the importance of long-term records for understanding past and contemporary climate change. The course takes place in semester 1 (September–December) and is a large class with 142 undergraduate students enrolled during the 2016–17

academic year. The assignment, which accounted for 33% of the course marks, was designed as a research-led teaching experiment to reflect on the importance of historical climatology, to provide insights into the process of data rescue, and to explore the power of the crowd in rescuing and transcribing meteorological data. As a lead-in to the assignment, students were given a guest lecture by staff from Met Éireann to convey the scientific and cultural importance of the data they would be working with. To coincide with the transcription component of the coursework, students were required to submit a written reflection on the importance of historical climatology, utilizing material and publications presented in the module. Learning outcomes were designed to provide students with the following:

- firsthand experience in working with historical climate observations;
- a critical appreciation of the processes involved in data rescue, digitization, and quality-assurance procedures that are essential to understanding past climate variability and change; and
- firsthand experience with the powerful contribution that citizen science can make to the study of climatology, geography, and other disciplines.

Scans of more than 1,300 pre-1940 annual rainfall sheets (e.g., Fig. 1) containing daily precipitation values and associated station metadata, together with templates used in transcribing the data (Fig. 2), were provided by Met Éireann. Of the sheets provided, 274 had been previously transcribed (keyed once) by Met Éireann. These data were used as a benchmark for student performance.

Upon receipt of the digital images, the file format was converted from a Portable Network Graphics (PNG) file type to a Joint Photographic Experts Group (JPEG) format, reducing file size from 17 Mb to ~1.4Mb, without any apparent loss of resolution. This facilitated the distribution of images to students and avoided potential logistical issues concerning students accessing large files. Each student was assigned 18 annual rainfall sheets to transcribe, and each sheet was assigned twice (i.e., to two different students). Such double-key data entry is a widely used method of quality control to detect incorrectly keyed information. For the purpose of this project, double-key entry was necessary for the allocation of student grades, as a component of the grade was based on transcription accuracy, which could only

Date	Jan.	Feb.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.	Date
1	—	.01	.44	.07	.40	—	.52	.04	.15	.15	—	.14	1
2	.24	—	—	.09	.02	—	.07	.08	.03	.24	—	.04	2
3	.18	.28	.01	.07	.08	.02	.01	.37	.01	.73	.14	.06	3
4	—	—	—	.01	.01	—	—	.42	.02	.24	.01	—	4
5	—	.02	.01	—	.09	—	.04	.05	.07	.53	—	—	5
6	.02	—	.16	.08	.07	—	.11	—	—	.77	—	.05	6
7	.02	.24	.01	.01	.05	—	.57	.34	.01	—	.01	—	7
8	.18	.04	—	.12	.21	—	.59	.04	—	.01	.01	.01	8
9	.81	.25	.01	.43	—	—	.27	.03	.22	.01	—	—	9
10	.19	.26	—	.14	—	.19	.06	—	—	—	—	.01	10
11	.86	—	.22	.30	.32	.05	.10	.11	.02	—	—	.17	11
12	.20	.01	.15	.13	.02	.56	.02	.01	—	.01	.15	.21	12
13	.13	.07	.04	.06	.01	.01	—	—	.12	.11	.13	.03	13
14	.02	.01	.05	.26	—	—	.03	.01	.14	.37	.08	.07	14
15	—	.16	.07	.06	.07	.31	—	.01	.15	.17	.08	—	15
16	—	.01	.07	.08	.03	—	.30	—	.11	.14	.02	—	16
17	.02	—	.05	—	.07	—	.02	.19	—	.11	—	—	17
18	.26	.03	.01	.06	.50	—	.15	.01	.01	—	—	—	18
19	.01	.15	—	.07	.03	.05	—	.02	—	.78	—	.01	19
20	.07	.01	—	.17	—	.04	.02	—	.04	.06	—	.02	20
21	—	—	.01	.05	.02	.01	.10	—	—	.01	—	.07	21
22	.21	—	—	.02	—	.02	.33	.15	—	—	—	—	22
23	.42	—	.03	.11	—	—	.12	.02	—	—	—	.15	23
24	.05	—	.14	.05	—	.01	.02	—	.01	—	.12	.19	24
25	.26	.12	.09	.04	—	.08	1.35	—	—	—	.04	.02	25
26	.02	.08	.03	.21	—	.10	.22	—	.01	.06	.12	—	26
27	.22	.07	.26	.03	.41	.22	.16	—	—	.10	—	—	27
28	—	.01	—	.05	—	.15	—	—	.01	.04	.24	—	28
29	.16	—	.02	.05	.44	—	.15	—	—	.06	.50	—	29
30	.03	—	.01	.08	.04	.08	—	—	.03	.59	.35	.06	30
31	.08	—	.24	—	.02	—	.03	—	—	.02	—	—	31
TOTAL	4.66	1.23	2.13	2.90	2.91	1.90	5.36	1.90	1.14	5.07	2.31	2.17	34.28
Days with 1/8 in. or more	24	19	23	28	21	16	25	17	16	21	16	20	246
Days with 1/16 in. or more	17	11	13	24	13	11	18	10	8	14	12	14	165

FIG. 1. Example of an annual rainfall sheet distributed to students.

Metadata	Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Month	M_total	M_rainday	M_wetday	QCmonth	
image_no	1163	1		0.01	0.44	0.07	0.4	0.52	0.04	0.15	0.15		0.14	1	4.66	24	17	4.66	
station	Royal Botanic Ga	2	0.24			0.09	0.02	0.07	0.08	0.03	0.24		0.04	2	1.83	19	11	1.83	
county	Dublin	3	0.18	0.28	0.01	0.07	0.08	0.02	0.01	0.37	0.01	0.73	0.14	0.06	3	2.13	23	13	2.13
observer	Sir F.W.Moore	4				0.01	0.01			0.42	0.02	0.24	0.01		4	2.9	28	24	2.9
lat_degree	53	5		0.02	0.01		0.09	0.04	0.05	0.07	0.53			5	2.91	21	13	2.91	
lat_min	23	6	0.02		0.16	0.08	0.07		0.11				0.05	6	1.9	16	11	1.71	
lon_degree	6	7	0.02	0.24	0.01	0.01	0.05		0.57	0.34	0.01		0.01	7	5.36	25	18	5.49	
lon_min	16	8	0.18	0.04		0.12	0.21		0.59	0.04		0.01	0.01	8	1.9	17	10	1.96	
year	1920	9	0.81	0.25	0.01	0.43			0.27	0.03	0.22	0.01		9	1.14	16	8	1.14	
start_month	1	10	0.19	0.26		0.14			0.19	0.06			0.01	10	5.07	21	14	5.07	
start_day	1	11	0.86		0.22	0.3	0.32	0.05	0.1	0.11	0.02		0.17	11	2.31	16	12	2.31	
funnel_diameter	8	12	0.2	0.01	0.15	0.13	0.02	0.56	0.02	0.01		0.01	0.15	0.21	12	2.17	20	14	2.17
gauge_height_ft	1	13	0.13	0.07	0.04	0.06	0.01				0.12	0.11	0.13	0.03					
gauge_height_incl	0	14	0.02	0.01	0.05	0.26			0.03	0.01	0.14	0.37	0.08	0.07					
gauge_pattern		15		0.16	0.07	0.06	0.07	0.31		0.01	0.15	0.17	0.08						
railway	Glasnevin	16	0.02	0.01	0.07	0.08	0.03		0.3		0.11	0.14	0.02						
railway_distance		17	0.26		0.05		0.07		0.02	0.19			0.11						
elevation_msl	55ft	18	0.01	0.03	0.01	0.06	0.5		0.15	0.01	0.01		0.01						
jan_1_following	0.41	19	0.07	0.15		0.07	0.03	0.05		0.02			0.78	0.02					
year_total	34.28	20		0.01		0.17		0.04	0.02		0.04	0.06	0.07						
year_rainday	246	21			0.01	0.05	0.02	0.01	0.1			0.01							
year_wetday	165	22	0.21			0.02		0.02	0.33	0.15				0.15					
time_of_observat	9:00 AM	23	0.42		0.03	0.11			0.12	0.02				0.19					
comments_sheet		24	0.05		0.14	0.05		0.01	0.02		0.01		0.12	0.02					
observer_notes		25	0.26	0.12	0.09	0.04		0.08	1.35				0.04	0.12					
		26	0.02	0.08	0.03	0.21		0.1	0.22			0.01	0.06						
		27	0.22	0.07	0.26	0.03	0.41	0.22	0.16				1.01	0.24					
		28		0.01		0.05		0.15				0.01	0.04	0.5					
		29	0.16		0.02	0.05	0.44		0.15				0.06	0.06					
		30	0.03		0.01	0.08	0.04	0.08			0.03	0.59	0.35						
		31	0.08		0.24		0.02		0.03			0.02							

FIG. 2. Template distributed to students for transcribing the data. Columns D–O contain the 31 × 12 array of values extracted for comparison against the corresponding file. Differences detected by the MATLAB code between the two corresponding files are highlighted in yellow. Columns Q–S contain respective values for monthly totals, total number of rain days per month, and total number of wet days per month as recorded by the observer. Column T (red font) shows the quality check used by the students to compare automatically generated monthly totals to the monthly totals recorded by the observer and input to columns D–O.

be ascertained through comparison. The assignment was set in an hour-long lecture that outlined both the assignment and the rationale. Individual directories containing 18 randomly selected annual rainfall sheets were created and distributed to each student via Dropbox along with a Microsoft Excel template for keying the data. Students were provided a link from which they downloaded their personalized directory and performed the transcription component of the coursework. Students were given five weeks in which to complete and submit their transcribed data.

Several additional student aids were implemented:

- A simple video tutorial was produced to describe the different sections of the annual rainfall sheets and to demonstrate the transcription process, and posted to Moodle, the university’s online learning platform.
- An automated quality-assurance check was integrated into the Excel template to generate a monthly total based on the daily values transcribed. Comparison against the monthly totals recorded on the original sheet provided students a

basic quality check to ensure transcribed monthly totals were consistent with the originals (Fig. 2).

- An online discussion forum was set up on the Moodle course page through which teaching staff could address queries raised by students. Students were invited to post questions relating to, for example, differences identified in monthly totals and difficulties interpreting metadata, or to clarify illegible values.
- An in-class check-in clinic was organized at the midway point of the assignment to highlight frequently asked questions from the online forum and to allow students the opportunity to raise questions in class.

Once the assignment was complete, students compiled the transcribed files into a compressed zipped directory, maintaining a consistent file-naming convention, and then uploaded the files to the online course portal. Developing a file-naming convention that uniquely identifies the image and student (e.g., image number_student number.xlsx) was essential to data postprocessing. This allowed for a comparison between equivalent sheets based on image number

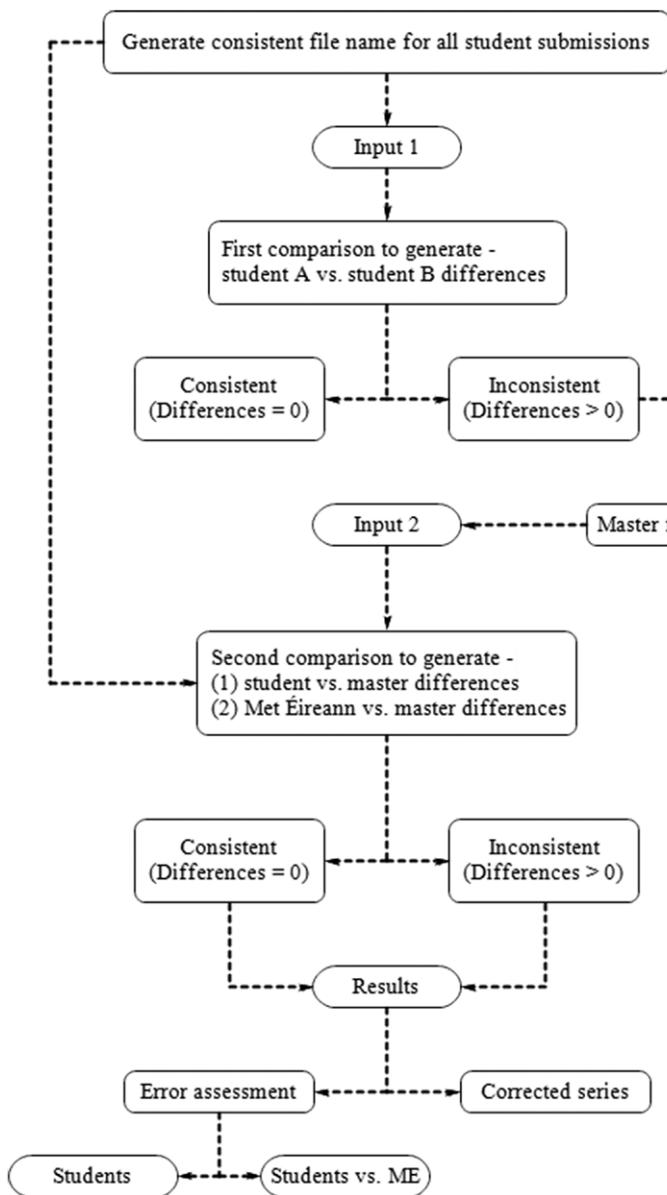


FIG. 3. Flowchart outlining the workflow from receiving student submissions to final output and assessment of errors.

and assignment of grades to students based on an evaluation of their performance.

EVALUATING STUDENT PERFORMANCE.

Figure 3 provides an overview of the workflow implemented once student transcriptions were received. Having verified that consistent naming conventions existed for all student files, an initial comparison of the double-keyed data was performed. Student files corresponding to the 274 previously transcribed files

provided by Met Éireann were also included. MATLAB code was developed to compare double-keyed data by extracting the 31×12 array representing daily values for one year and to highlight cells where differences existed (Fig. 2). From this, two directories were generated: i) a consistent directory comprising student-transcribed files indicating zero differences between double-keyed sheets and ii) an inconsistent directory of student-transcribed files containing differences between double-keyed sheets.

Utilizing results from the first comparison, a “correct” version of each file in the inconsistent directory was created by teaching staff. These corrected or “master” files were manually created by examining highlighted differences against the original scanned images and then adjusting transcribed values accordingly.

Because this required consideration of solely a small subset of the total values transcribed, it was achieved in a matter of hours by two members of staff.

With master files now available for all original annual rainfall sheets (i.e., those showing no differences, together with the manually corrected files), a second comparison was performed to evaluate i) all student transcriptions against the corresponding master files to assess student errors and ii) student transcriptions against Met Éireann transcriptions for common sheets to benchmark student performance relative to Met Éireann. This second comparison generated a new set of consistent and inconsistent directories containing files with zero differences and files with greater than zero differences, respectively. Results were derived from these two output directories.

Student grades were assigned based on the number of differences within any transcribed file. A file containing zero differences obtained full marks; for files showing differences, marks were deducted for each incorrect entry. Figure 4a displays the frequency (%) of student submissions being less than or equal to x , where x is the number of errors per transcribed file (bounded by 0 and 365 for all years, except leap years when 366 are possible), together with a bar graph (Fig. 4b) categorizing the total number of incorrect files by actual number of errors per transcribed sheet. When compared against the corrected master files, 62% of student-transcribed files showed no errors. In 96% of student-transcribed files, fewer than 5% of data entries were incorrect (i.e., 96% of files had

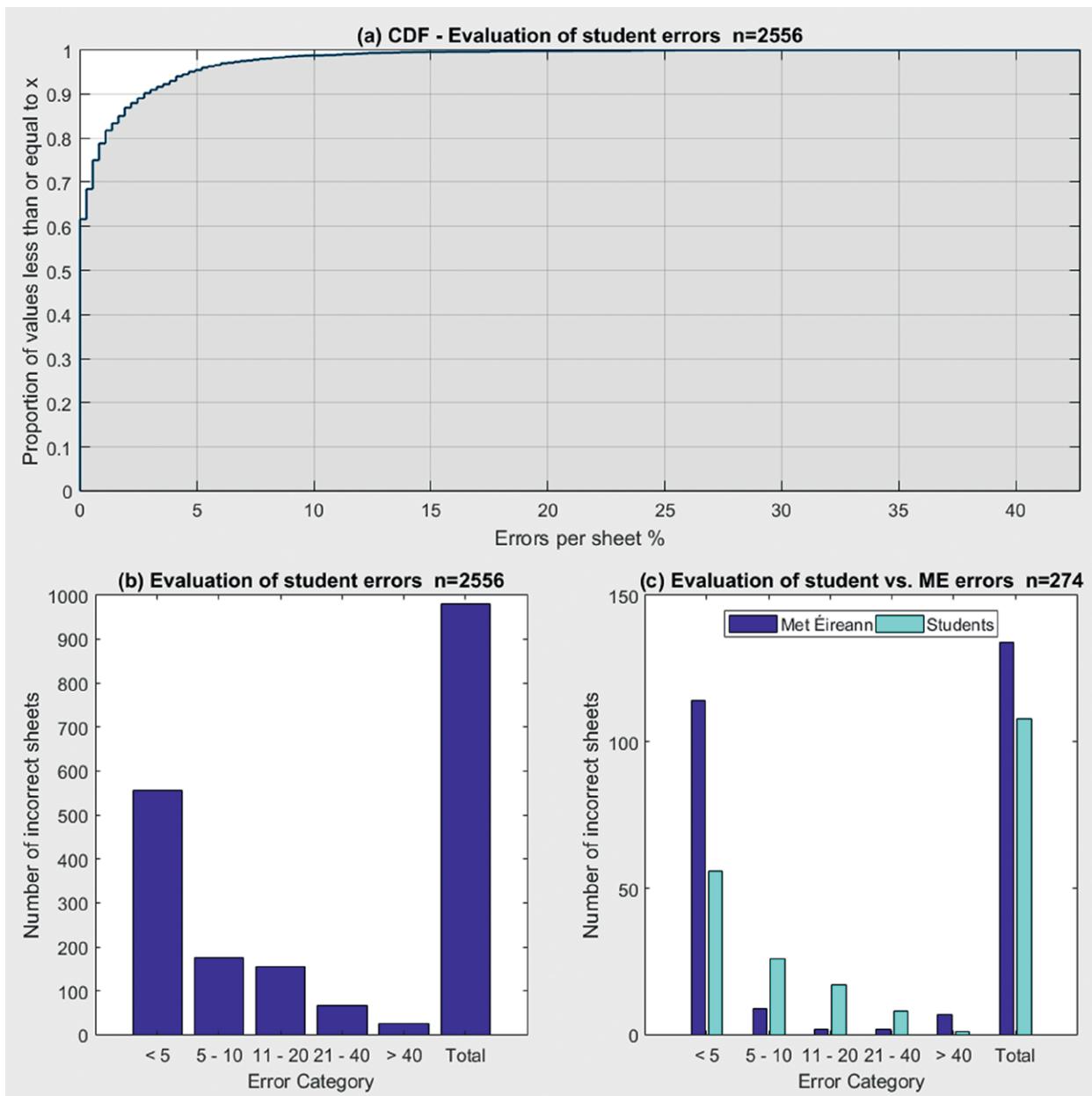


FIG. 4. (a) Empirical cumulative distribution function (CDF) showing the frequency (%) of student submissions ($n = 2556$) being less than or equal to x , where x is the percentage of errors per sheet. (b) Total number of incorrect student sheets by number of errors per sheet. (c) Comparison of errors by error category for sheets transcribed by both students and Met Éireann ($n = 274$).

fewer than 20 errors). A review of all incorrect files reveals that 57% of the transcribed files containing errors had fewer than 5 errors, 90% had fewer than 20 errors, and only 3% had greater than 40 errors. Cumulative error across all 2,556 files transcribed by students reveals a percentage error of less than 1%. In the unlikely event that both students were wrong (i.e., that they produced identical errors),

the errors would not be detected in the comparison. Such errors would, however, be identified during the subsequent application of more comprehensive quality-assurance techniques.

BENCHMARKING STUDENTS AGAINST PROFESSIONALS. A final assessment was carried out utilizing the 274 sheets transcribed by both Met

Éireann and the students. This provided a benchmark against which student performance could be evaluated. Figure 4c displays the proportion of incorrect files by error category. While the students have a smaller number of incorrect files overall (39% for students compared with 49% for Met Éireann), the majority of Met Éireann's incorrect files lie in the lowest error-propensity category. Of the incorrect files from students, 52% contained fewer than 5 errors and 47% contained 6–40 errors, with the remaining 1% of incorrect files showing more than 40 errors. For Met Éireann, 85% of incorrect files contained fewer than 5 errors; however, Met Éireann also had a greater number of incorrect files with more than 40 errors.

Upon further investigation, it was noted that different approaches to the transcription process (i.e.,

row based vs column based) had an impact on the number of errors produced within individual files. Specifically, the row-based approach employed by Met Éireann resulted in the majority of incorrect files falling into the lowest error category (i.e., ≤5 errors). Alternatively, the column-based approach adopted by the students produced a greater number of files containing errors in the intermediate categories (i.e., 6–10, 11–20, and 21–40). Utilizing a column-based approach, the incorrect placement of a data value occasionally propagated down the entire monthly column. While the individual values were correct, differences were flagged due to the data values being input to the incorrect date/cell. An example of errors arising from the two different approaches is highlighted in Fig. 2.

TABLE 1. List of statements provided to students to assess learning outcomes and the extent to which proportions of the class (%) agreed or disagreed with these statements (n = 62).

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree	Total
The assignment provided insights into the process of data rescue.	58	40	0	2	0	100
I now have an appreciation of the role of historical data in climate research.	55	39	5	2	0	100
Supports provided were sufficient to aid completion of the assignment.	66	29	2	3	0	100
The process of working with the data/images was clear and easy to follow.	41	41	10	7	2	100
The assignment provided insights into the power of citizen science.	44	35	16	5	0	100
I would prefer to participate in CA like this over more traditional assignments.	60	20	10	7	3	100
The workload was appropriate to the level of credit given.	45	32	11	8	3	100
Overall, I found the assignment to be a valuable learning experience.	44	44	6	6	0	100
I was more motivated than is usual for me in doing this assignment.	35	23	39	3	0	100
I feel that I have contributed to understanding Irish rainfall trends.	44	47	10	0	0	100
I think that future students would be pleased to participate in a similar assignment.	52	37	5	5	2	100
I think that working with data is a useful skill to have gained.	52	44	3	2	0	100
I did not gain any valuable skills from this assignment.	0	2	11	34	53	100

EVALUATING THE STUDENT EXPERIENCE.

On completion of the project, feedback was provided to the students. First, we explained how the evaluation was performed and what the results told us about the accuracy with which the students had transcribed the data. Second, data transcribed by the students for a local station were collated and presented to the class to exemplify their contribution to understanding changes in historical rainfall. We discussed trends and notable events present in the early record and how it commensurately nuanced our understanding of long-term local climate. To investigate the students' perspectives of the project and reflect upon their perceptions of learning outcomes, a formal assessment was conducted via anonymous questionnaire completed in class in the final week. Table 1 presents the list of statements provided to the students and the extent to which proportions of the class who responded agreed or disagreed with each of these statements.

The response was positive across all aspects of the assignment. The majority of students (>90%) gained insights into the process of data rescue and an appreciation of the role of historical data in climate research. The assignment afforded students a firsthand experience in working with raw historical climate data and dataset development—in particular, processes behind the cataloguing and imaging of historical rainfall sheets, the importance of double entry to avoid gross errors, and the value of metadata recorded by the original observers. More than 90% of participants stated that they could see value in their contribution (extending available digital records) to understanding Irish rainfall trends. Presentation of the long-term time series at the end of the project supported this consensus. Notably, 80% of respondents stated that they would prefer to participate in class assignments (CA) like this over other, more traditional, assignments. Students were given the opportunity to continue working with the data, but in smaller groups, in the second semester as part of a research methods workshop whereby they could conduct statistical analysis on the transcribed data and present findings in individual research reports.

Students were asked to identify which student aids they found most useful in completing the assignment. The most popular were the online discussion forum and the video tutorial, at 44% and 36%, respectively. The management of these resources demanded a significant investment of time

by teaching staff, particularly the online discussion forum, which received more than 500 queries from students. Nevertheless, effective management of the project and student aids facilitated the development of the corrected dataset by notably reducing the propensity for errors and the amount of time required to carry out postprocessing of the data. While questions have been raised over the accuracy and reliability of citizen science-produced datasets, a number of projects have demonstrated the potential for enhancing data quality through practical management (van der Velde et al. 2017; Kosmala et al. 2016). Despite the success of this initial experiment, a number of issues arose that require consideration in future iterations:

- Given that the misplacement of daily values, rather than incorrect entry of data, was the main source of error, simple additions to the Excel template used to transcribe the data (e.g., delineating the margins of the columns and rows) could significantly reduce the propensity for errors. Additionally, adopting the row-based approach used by Met Éireann to transcribe the data could further reduce errors associated with the incorrect placement of data values.
- Reducing the number of sheets allocated to the students could potentially increase student motivation and further reduce the number of errors.
- Having successfully integrated the transcription component into the coursework, the next iteration of the assignment will build in basic quality assurance and statistical analysis of the resulting dataset. Development of such skills will further deepen the research experience offered to students.

SUMMARY. We have outlined an initial successful attempt to integrate data rescue into the classroom using a research-led approach to teaching and learning that complies with the prerequisites of pedagogy within the university curriculum. Additionally, the project motivated students by engaging them in a practical exercise whereby their contribution adds considerable value to research. Such initiatives promote the development of mutualistic collaborations between national meteorological services and higher-level institutions with the paired objectives of accelerating science. At the same time, students gain a firsthand understanding of the processes that underpin data rescue and research.

Having established confidence in the transcription, the next steps in developing a long-term daily rainfall network for Ireland will involve i) repeated iterations of this assignment with subsequent classes, ii) the application of comprehensive quality-assurance and homogeneity techniques, and iii) analysis of the derived long-term record to assess changes in the characteristics of extreme rainfall events. Metadata transcribed by the students will be systematically extracted and catalogued to facilitate the process. A final objective is to make the data widely available to national and international researchers. To this end, the data shall be shared with the recently awarded Copernicus Climate Change Service Global Land and Marine Observations Database service led by coauthor Peter W. Thorne.

It is hoped that the framework outlined in this paper may be integrated into teaching programs across other universities and used to highlight the importance of historical meteorological sources to students and to encourage their involvement in data-rescue efforts. We will provide the developed student aids to help realize this objective. These materials are available via both online supplemental material and at www.maynoothuniversity.ie/icarus/data-rescue-classroom. The latter shall be updated as we make modifications in subsequent years.

ACKNOWLEDGMENTS. The authors thank and commend Maynooth University (GY313: Climate Change) students who participated in the project. Ciara Ryan gratefully acknowledges funding provided by the Maynooth University John and Pat Hume Doctoral Scholarship and the Irish Research Council Employment Based Scholarship. Conor Murphy acknowledges support from the Irish Environmental Protection Agency project 2014-CCRP-MS.16.

FOR FURTHER READING

- Allan, R., P. Brohan, G. P. Compo, R. Stone, J. Luterbacher, and S. Brönnimann, 2011: The International Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative. *Bull. Amer. Meteor. Soc.*, **92**, 1421–1425, <https://doi.org/10.1175/2011BAMS3218.1>.
- Bonney, R., J. L. Shirk, T. B. Phillips, A. Wiggins, H. L. Ballard, A. J. Miller-Rushing, and J. K. Parrish, 2014: Next steps for citizen science. *Science*, **343**, 1436–1437, <https://doi.org/10.1126/science.1251554>.
- Brunet, M., and P. Jones, 2011: Data rescue initiatives: Bringing historical climate data into the 21st century. *Climate Res.*, **47**, 29–40, <https://doi.org/10.3354/cr00960>.
- Harnik, P. G., and R. M. Ross, 2003: Assessing data accuracy when involving students in authentic paleontological research. *J. Geosci. Educ.*, **51**, 76–84, <https://doi.org/10.5408/1089-9995-51.1.76>.
- Kosmala, M., A. Wiggins, A. Swanson, and B. Simmons, 2016: Assessing data quality in citizen science. *Front. Ecol. Environ.*, **14**, 551–560, <https://doi.org/10.1002/fee.1436>.
- Mitchell, N., M. Triska, A. Liberatore, L. Ashcroft, R. Weatherill, and N. Longnecker, 2017: Benefits and challenges of incorporating citizen science into university education. *PLOS ONE*, **12**, e0186285, <https://doi.org/10.1371/journal.pone.0186285>.
- van der Velde, T., D. A. Milton, T. J. Lawson, C. Wilcox, M. Lansdell, G. Davis, G. Perkins, and B. D. Hardesty, 2017: Comparison of marine debris data collected by researchers and citizen scientists: Is citizen science data worth the effort? *Biol. Conserv.*, **208**, 127–138, <https://doi.org/10.1016/j.biocon.2016.05.025>.
- Vitone, T., K. A. Stofer, M. S. Steininger, J. Hulcr, R. Dunn, and A. Lucky, 2016: School of ants goes to college: Integrating citizen science into the general education classroom increases engagement with science. *J. Sci. Commun.*, **15**, 1–24.