Transfer of old Nature Calgary (CFNS) bird sighting data to eBird

By Andrew Hart, revised July 2015, posted to Albertabird July 26, 2015.

From about 1986 to 2001 several thousand (about 22,000) sets of birding field trip records were submitted to the (then) CFNS by members and others. These records were submitted in hard copy using a standardized checklist, and entered in to a Paradox database by Olga Droppo, Marie Baibey and others. I believe the intent was to create a database that CFNS members and others could access. For a variety of reasons, including the technology of the day, that did not happen. It seems that the effort languished in about 2001. There are records of numerous discussions about the fate of the data, but it essentially languished and presumably slowly became forgotten.

As is often the case a series of isolated events combined to bring these data to light. After the June 2013 floods Jamie Noakes asked me take the old CFNS computer and several boxes of data off his hands to free up space in his basement for flood damaged items he needed to store for someone else. Cathy Noakes was an enthusiastic supporter of this request. The material included an incomplete set of the hard copy records (probably about 50 percent complete). It also included one of the old CFNS computers (fortunately the most recent one). From the hard copy records it was clear that at one time CFNS and FAN (now Nature Alberta) had been co-operating. So I asked Petra Rowell at Nature Alberta if they knew anything about it or had any of the missing data, I told her I was looking to upload the data to eBird. The one thing that came out of this is that Petra passed my interest on to Greg Wagner who has been combing records for Frank Lake sightings and uploading those to eBird. Greg put me in touch with Mike Burrell at Bird Studies Canada who made several helpful suggestions. While this was going on Greg visited the Inglewood Bird Sanctuary to borrow records pertinent to Frank Lake, and came across nearly all of the missing hard copy binders. So I now had an almost complete set. How the original binders came to be divided between a basement and the IBS is an interesting question.

The data binders include thousands of hard copy bird checklists generally submitted with three checklists per double sided sheet. These lists were numbered 860001-1, 860001-2 etc. by Olga and her team. 860001-1 represents the first column of data (first field trip record) of the first sheet submitted in 1986. Confirmation of data entry for each sheet is stamped on the sheet by both the original data entry person, and by a second person who provided the quality check that the data entry had been done correctly. This is the same quality control procedure that I am familiar with in the EPC industry and it is very effective. It is unlikely that there is a significant incidence of transcription errors in the electronic database. To date I have only come across one instance of a transcription error which looks like it can be attributed to poor handwriting on the original submission sheet. I understand from anecdotal conversations with current Nature Calgary members who were somewhat familiar with the process, and from notes in the binders, that "questionable" sightings were sometimes challenged, or further back-up obtained. But it seems that in general the submitters were given the benefit of the doubt.

However, none of this would help much if the data had to be manually re-entered from the hard copies (the 22,000 field trips contain about 225,000 sightings records). While these dialogues

had been proceeding I was able (with a surfeit of help from my workplace IT department and my IT savvy son) to first recover the data from the CFNS computer hard drive, and then convert the recovered Paradox data to Excel and Access formats.

That brought me to the beginning of the task of conditioning the data for eBird upload. The biggest challenge was the location descriptions for each sighting. These were entered in free format, making them very difficult to search (or specifically upload into eBird) in any systematic manner. I think this was a problem that the earlier compilers recognized and were addressing when the project stopped about 16 years ago. I decided to address this by entering latitude and longitude data for each of the 22,000 locations, less about 1400 where the trip record covered a distance significantly bigger than that recommended by eBird. Those 1400 or so sightings have not been uploaded to eBird. That process took me several months. I then went through the data again to substitute some of the latitude and longitude data with eBird "hotspot" locations to make it easier for eBird data miners to explore. This means that the several thousand separate location descriptions in the original database have been reduced to just over 3000. In most cases this reduction is simply the reconciliation of different ways of spelling the same location, converting the spelling to be consistent with current hotspot locations, or combining locations in the same local area.

While preparing the data I became familiar with the names of some of the more prolific submitters to the database. I compared these with the “top 100 Alberta eBirders” in the period covered to see if any of the old CFNS data may have been submitted already. While eBird already has many instances where duplicate sightings are entered, e.g. if several people on the same field trip enter data independently rather than sharing it, BSC advises that duplication should be avoided if possible. Based on this I contacted Andrew Slater and Mike Harrison who both confirmed that all of their sightings that had been submitted to CFNS had already been submitted to eBird. So I excluded their sightings from the Nature Calgary account that I used for the upload.

The data shows up in eBird with "Nature Calgary”, as the observer name. The Nature Calgary eBird account is also being used to track data for current Nature Calgary field trips, when such data are available and shared with the account.

The original observers are noted in the comment section of each uploaded field trip record and the original field trip record number (assigned by Olga and her team) is also included so that the original data can always be traced back. As a basic security precaution I have changed location domestic addresses such as "1234 Main Street" to "Main Street" in the location description in eBird.

Some of the sightings have additional back-up information in the form of photographs and/or transcriptions of field notes, sometimes including sketches, In some instances I have included scans of this additional back-up as part of the uploaded eBird record.

As you can imagine a task like this is not without its challenges. As a simple example, several of the bird species in the old CFNS records, such as "Northern Oriole" have since been split into two species. In instances like that I have taken guidance from BSC. In other cases the names of

the species have changed (e.g. from Oldsqaw to Long-tailed duck), and in others just the spelling has changed. On the advice of Bird Studies Canada I matched all the species names to the current eBird master list.

When uploading current sightings to eBird it is normal to enter sightings as either "stationary", all sightings from a fixed location, or "traveling", where a specific (known) distance was traveled. For archive data these parameters were not always recorded so data have traditionally been entered as "casual" or "incidental" (eBird allows both descriptions). The old hard copy sheets had space to enter start time and stop time (from which duration can easily be calculated electronically), but not distance. So I scanned through all the records and where it seemed clear that the observations were "stationary" I changed the protocol from "casual" to "stationary", subject to both start time and duration being available. I did this for sightings that were obviously for a street address, or where the location description was along the lines of "at the community hall" or "at the firehall feeder".

Before getting on to the systematic upload of all the sightings to eBird I tried a small test file. All seemed to go well and I was poised for the balance of the uploads. At that moment, for some reason, I noticed that eBird has added a new protocol "historical" for files just like these. This protocol applies to sightings where birding was the main purpose of the trip, but not all of the data necessary to classify the trip as "stationary" or "traveling" (start time, duration, and distance – for traveling) are available. So I deleted the few files I had uploaded, changed "casual" to "historical" and started again. When I finished, in early July, ~19,000 field trip records had been submitted. For the curious I split the data into 42 separate CSV files in "eBird Record Format (Extended)" for the uploads. So on average each upload file contained 440 field trip records and took a few minutes to upload.

Since I started the upload process I have had some feedback from the regional eBird moderators. It appears that two of the sightings were challenged at the time by the provincial rare bird sightings committee, so I removed them.

It struck me as ironic that the eBird database was launched in 2002 and started to gain traction just barely after the CFNS effort languished. A small change in timing and there may have been a more timely transition.

I have given all of the hard copy binders to the Inglewood Bird Sanctuary for permanent storage. I have retained scanned images of all of those binders, producing those was a mini challenge of its own. The scanned images, or a sample of them, may be made available on the birding resources page of the Nature Calgary website at a future date.

If you have any questions or concerns please contact me at birdstudy*@*naturecalgary.com .

Andrew Hart

Chairman, Calgary Bird Study Group
President, Nature Calgary