

## Accessing and using weather-related data in OCaml

Hezekiah Carty, [hezekiah.carty@mdaus.com](mailto:hezekiah.carty@mdaus.com), MDA Information Systems LLC

This talk will cover uses and applicability of OCaml for wrangling weather data at MDA and during my time as a graduate student at the Department of Atmospheric and Oceanic Science at the University of Maryland. I will focus on the library bindings required for this work as well as the challenges and benefits of using OCaml in a field where the most common environments are Fortran, Matlab, IDL or (recently) Python.

### Bindings

GRIB<sup>1</sup> API (`ocaml-grib`): An API and suite of command-line tools for working with GRIB1 and GRIB2 file formats. The software is developed by the ECMWF<sup>2</sup>. GRIB formats are commonly used for distributing output from meteorological models. Most global and regional weather forecasts put out by government and private agencies are provided as GRIB1 or GRIB2.

HDF4<sup>3</sup> (`ocaml-hdf`): A file format and API for working with HDF4 and NetCDF data. These formats are commonly used for data collected from Earth-observing satellites.

### Library binding successes:

- `ocaml-grib` and `ocaml-hdf` have been used successfully in production for years, processing hundreds of gigabytes of compressed data each day.
- The resulting interfaces are (hopefully!) safer and easier to use than their C/Fortran counterparts.
- OCaml's C FFI is fairly easy to use in simple cases.

### Library binding challenges:

- `ocaml-hdf` is a mix of `camldl`-generated and hand-written bindings.
- The first iteration of `ocaml-hdf` was one of my first projects while learning OCaml so there are some ugly pieces of code left around.
- Passing large arrays between C and OCaml requires balancing ease of data access with copying/duplicating large amounts of data between the two environments. This is not always an easy balance to achieve.

### Future binding work and potential improvements:

- Adding blocking section indicators could allow for some level of parallelism when encoding and decoding data, potentially at the cost of extra data copying.
- `ocaml-ctypes` could make the bindings simpler to update and maintain.
- As always - more examples and more tests! As of this writing, most of the current tests are external to the released bindings.

---

<sup>1</sup> GRIBbed Binary. Bindings available at <http://github.com/hcarty/ocaml-hdf/>

<sup>2</sup> European Center for Medium-Range Weather Forecasts - <http://www.ecmwf.int/>

<sup>3</sup> Hierarchical Data Format version 4. Bindings available at <http://github.com/hcarty/ocaml-grib/>