

# The Open Dictionary Project

**Tyler Nickerson**

Linguistic Inc.

`tyler@linguistic.io`

The majority of dictionary formats in use today are built either upon antiquated or proprietary technologies, making offline, programmatic access to lexical data needlessly complex. In addition, most dictionary formats store entries as unstructured HTML, making the extraction of hierarchical or tagged data difficult and varying across formats.

The Open Dictionary Project (ODict) aims to create a new standard for dictionary formats that is open, extensible, and easy to use. Introduced in 2017, ODict is a file format specification, compiler, and API for creating and accessing dictionaries. Files are compiled to tiny, standalone binary files from easy-to-read XML and can retrieve definition entries in under one second. Along with its command-line interface, ODict ships with bindings for Node, Python, and the JVM.

In this presentation, the rationale and motivation behind ODict's development will be discussed, along with a demonstration of its full features and capabilities.