From remote files to running API

Builds and maintains production-ready API (high availability, scalable, standard, secured)

Includes BioThings Hub:
- keeps data up-to-date
- produces data releases
- uses parallelization

Ready-to-use
Pre-configured system dependencies, including:

- elastic
- mongoDB

Easy-to-install
- docker image

Find us on GitHub:
Core SDK Application:
https://github.com/biothings/biothings_api
https://github.com/biothings/biothings_studio

From remote files to running API

BioThings API workflow

1. Write a data plugin
   - write a parser
   - write a manifest:
     - data location
     - parsing function
   - register the plugin!

2. Inspect data
   Automatically generate an Elasticsearch mapping (rules used to specify which fields should be indexed and how)

3. Define a build configuration
   - select which sources to merge
   - choose merge strategy

4. Create a build
   - merge multiple data sources
   - provide basic statistics & logs

5. Create and manage data releases
   - full: create new index on Elasticsearch
   - incremental: compute differences between previous and current builds (json-diff) and patch index data. Reduce release size and time to deploy.

6. Create new API instance
   - use query interpreter to search and serve Elasticsearch data
   - default queries can be customized as needed

7. Automatically deploy API as high availability, scalable clusters
   - on the cloud
   - or hosted on biothings.io

Full tutorial available

This project is supported by the NIH through the Big Data to Knowledge initiative, award number: U01 HG008473, U54 GM114833 and R01 GM083924.