

scientific data, javascript, and you

presented by @maxogden



why I like node:

streaming

cross platform (os and web)

excellent dependency management



DAT

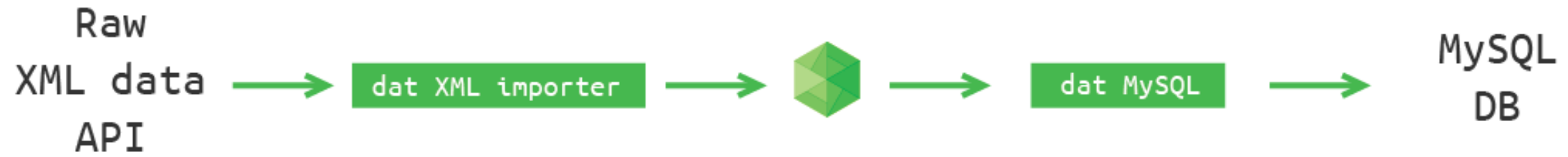
versioned data

**github.com/maxogden/dat
(1 year old, but still pre-alpha)**

What Is Dat

- Combination Tabular store and Blob store (Mutable/Immutable)
- Row level or file level versioning
- Push and pull replication
- 100% streaming + real-time
- CLI, JS, HTTP and RPC APIs
- Small core philosophy
- Everything is pluggable

Dat Pipeline



**Ecosystem of shared components
and workflows**

Protocol Buffers
stored in
LevelDB
(thanks, Google!)

Ultimate Goal of Dat:

**Enable a vibrant
ecosystem of reusable
data modules**

grunt gulp **gasket**

- **composable pipe chains**
- **like `npm run` but better**
- **100% unix pipes & streams**
- **cross platform**
- **reproducibility through npm**

npm install data-plumber -g

```
{
  "name": "my-module",
  "version": "0.0.1",
  "gasket": [
    "xml-json",
    "jsonfilter body.vehicle.*.$id"
  ],
  "dependencies": {
    "xml-json": "^1.0.0",
    "jsonfilter": "^1.0.1"
  }
}
```

```
ncbi.search('sra', 'txid7400[Organism] AND strategy wgs[Properties]')
    .pipe(fork1)
    .pipe(ncbi.parseXMLProperties(['expxml', 'runs']))
    .pipe(storeSRA())

fork1.pipe(collectUIDs()).pipe(fork2)

fork2
    .pipe(ncbi.link('sra', 'biosample'))
    .pipe(indexLink('sra-biosample'))
    .pipe(fork3)
    .pipe(collectBiosampleUIDsNotYetSearched())
    .pipe(ncbi.search('biosample'))
    .pipe(ncbi.parseXMLProperties(['sampledata']))
    .pipe(storeBiosampleMetadata())

fork3
    .pipe(collectSRAUIDs())
    .pipe(ncbi.link('sra', 'pubmed'))
    .pipe(indexLinkPubmed())
    .pipe(collectPubmedUIDsNotYetSearched())
    .pipe(ncbi.search('pubmed'))
    .pipe(storePubmed())
```



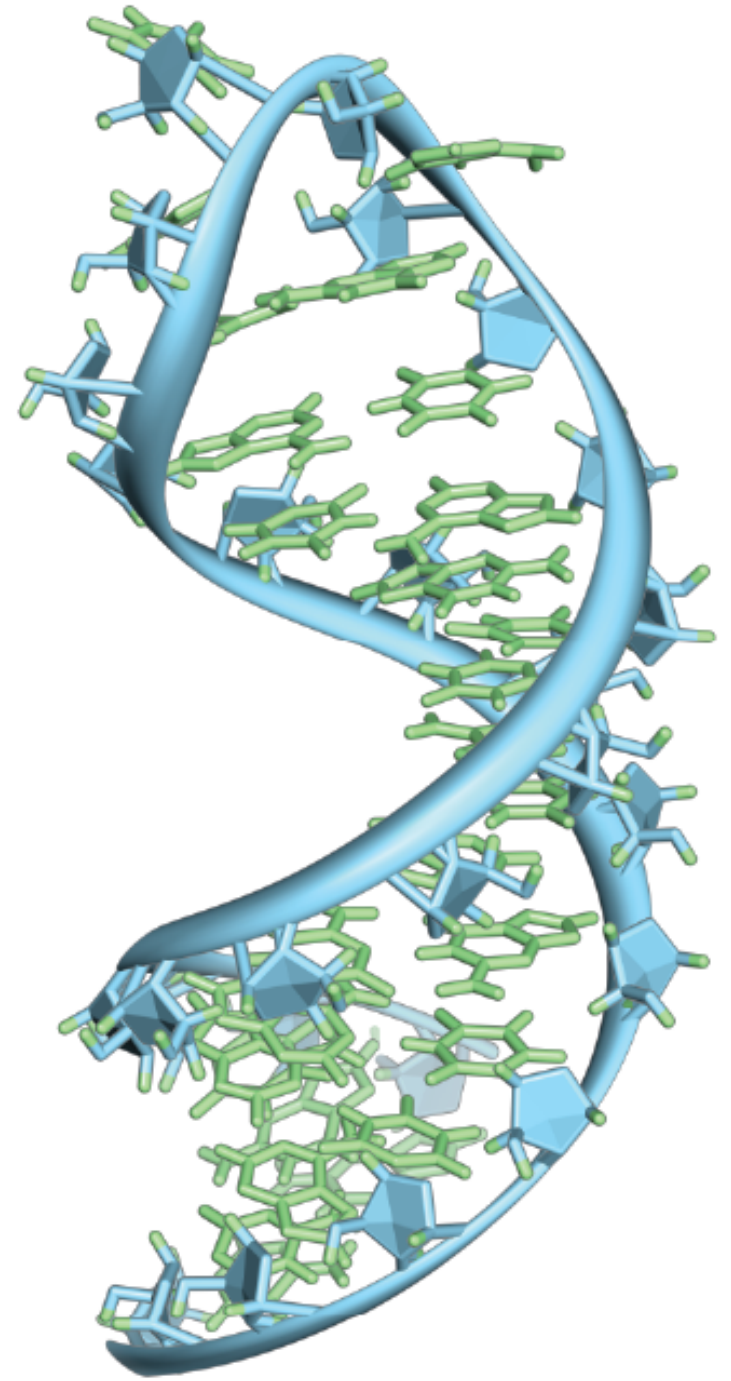
Pilot Use Case SDSS

sdss.org

trillian.scicoder.org

Pilot Use Case

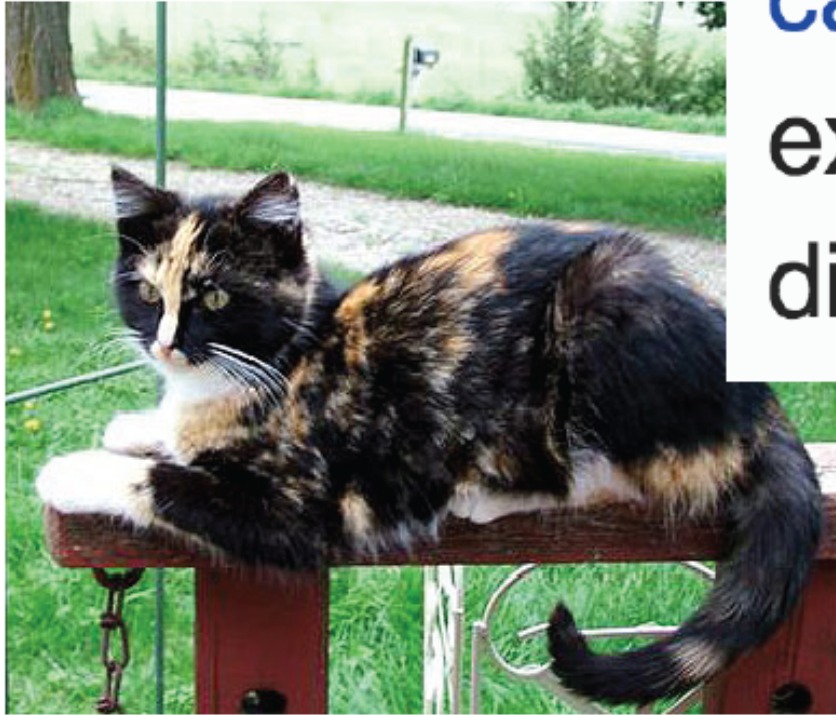
RNA-Seq



Bionode github.com/bionode

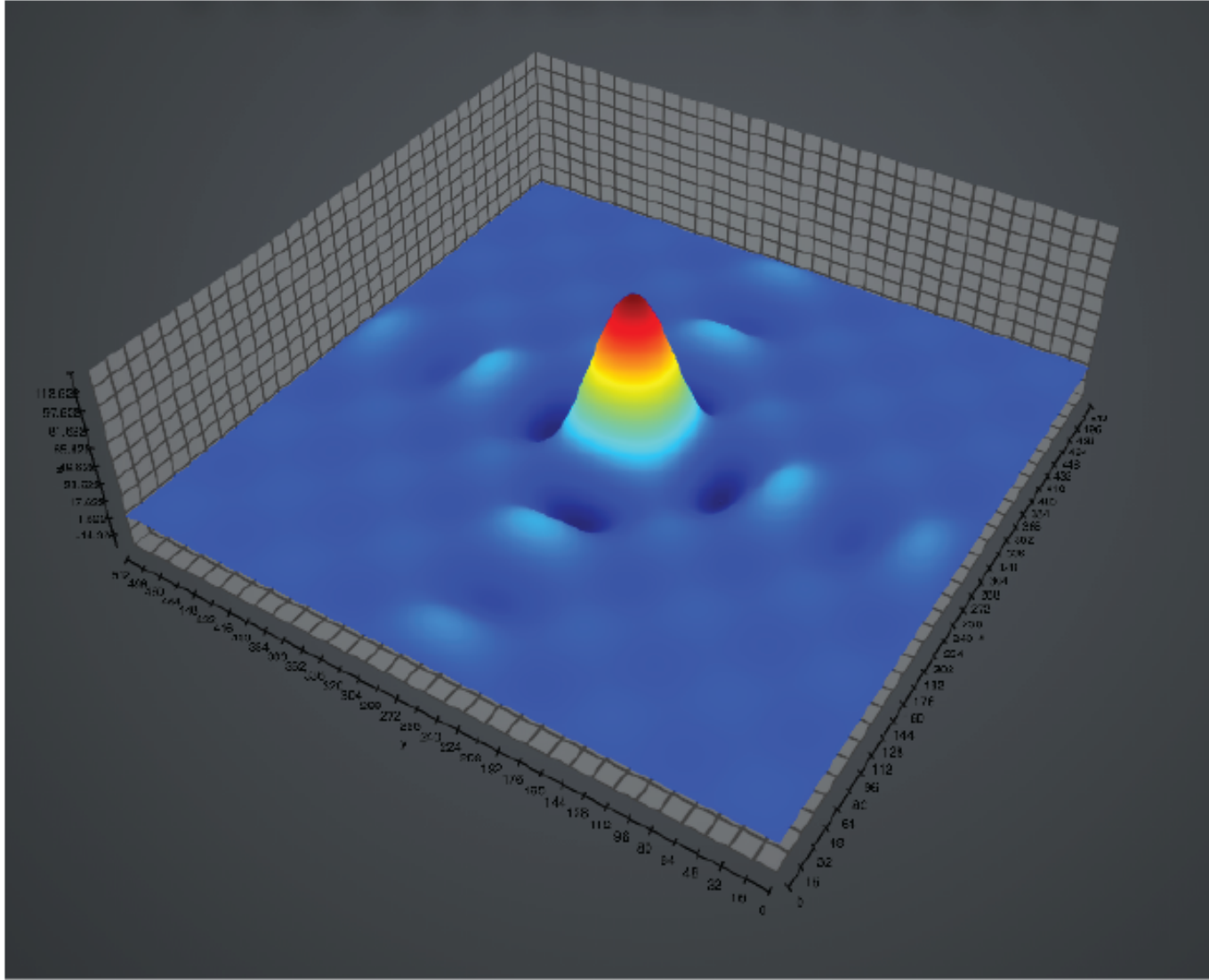
Genome.js github.com/genomejs

Bio.js github.com/biojs



The patchy colours of a **tortoiseshell cat** are the result of different levels of expression of **pigmentation** genes in different areas of the **skin**.

source: https://en.wikipedia.org/wiki/Gene_expression



ndarray
github.com/mikolajsenko/ndarray

FUTURE

Data Package Manager

BitTorrent

IPFS/FileCoin

Distributed Computation

First stable release soon!

dat-data.com

Developed in the open

github.com/maxogden/dat

**Contribute by
sharing use cases,
ideas, test data**

github.com/maxogden/dat/issues