

the State of the Union 2012



We built a lot of stuff

We built a lot of
reusable
stuff

We put it on CPAN

And we're all using
parts of Catmandu in
production

Hurray!

THE GOOD

We built a lot of Importers

- Atom
- CSV
- JSON
- YAML
- Z3950
- MARC
- SRU
- PubMed
- PLoS
- ArXiv
- LDAP
- CrossRef
- AWS

We abstract our data storage

- MongoDB
- Solr
- ElasticSearch
- DBI
- SBCatDB

(And we abstract querying the data with CQL)

We get the data out

- Plain Perl data
- JSON
- YAML
- CSV
- BibTeX
- RIS
- Template
- XLS
- (OpenOffice)

We're building components to expose the data

- `Dancer::Plugin::Catmandu::SRU`
- `Dancer::Plugin::Catmandu::OAI`
- `Catmandu::Plack::REST`

We built lots of utility functions, a web app generator, command line tools

Catmandu/Util.pm

Catmandu.pm

Catmandu/Cmd/

dancat

We're building applications

- Repository search
- Digital collections
- Discovery
- Scanning workflows
- Lots of CSL, embedding related stuff
- Digital archiving
- Project data
- Researcher data
- Analytics

THE BAD

We need more
documentation

We need more
tests

We need more
examples

We need a **file** store abstraction

- FS
- S3
- FedoraCommons
- DSpace, EPrints
- HTTP
- ...

We need a data
validation
module

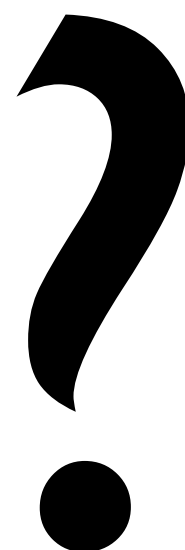
We haven't done a lot to reuse
components on the **client**
side
(widgets)

**There is no common data
model
(should we have one?)**

Catmandu core needs
consistent **error handling**

Catmandu core needs
consistent **logging**

still too many **dependencies**
(split off more functionality)



We only model relations as data attributes

$\{\text{part_of} \Rightarrow \{ _id \Rightarrow 6, _path \Rightarrow [6,4,2] \}\}$

+ scales well

+ simple

is there a need for more?

is there a need for transactions?

hot off the press: **components**