ENKI

Refuge

Data in the cloud

2012 - Benoît Chesneau @ Erlang Factory

- I’m benoitc
- web craftsman since 1995
- Apache Couchdb developer
- doing some stuff in Python too (gunicorn, circus, couchdb*)

about me

ENKI

WHAT?

- refuge at CCC: sensors data platform https://vimeo.com/27662399

initial concept

- bigdata?
- What about you? Do you really have 1 Petabyte of useful data?
- Distributing

bigdata?

ENKI

- big data center aren’t a solution.
- can’t scale indefinitely
- Flooding, tornado, ...
- Security
- expensive

bigdata?

PIONEN

OmniView MATRIX 2X8
Sun
2
0
3
1

- big data center aren't a solution.
- can't scale indefinitely
- Flooding, tornado, ...
- Security
- expensive

bigdata?

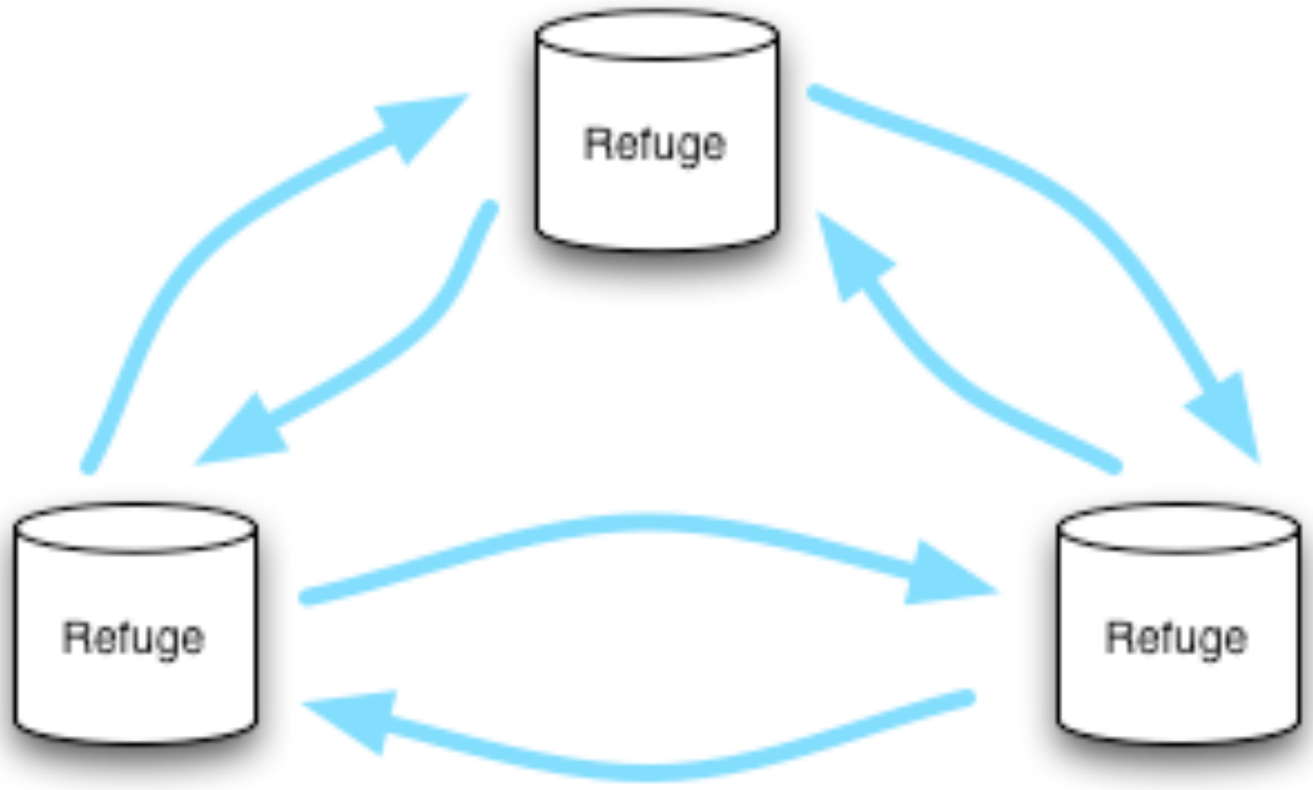
- Do you really need to centralize the data storage?
- Publishers aren't everywhere,
- Multiple data silos
- Query them directly

decentralize

- data should be yours
- keep privacy
- but be efficient

`data should be yours`

- decentralized data platform
- opensource
- based on Apache CouchDB
- Erlang powered
- multiplatform



refuge project

ENKI

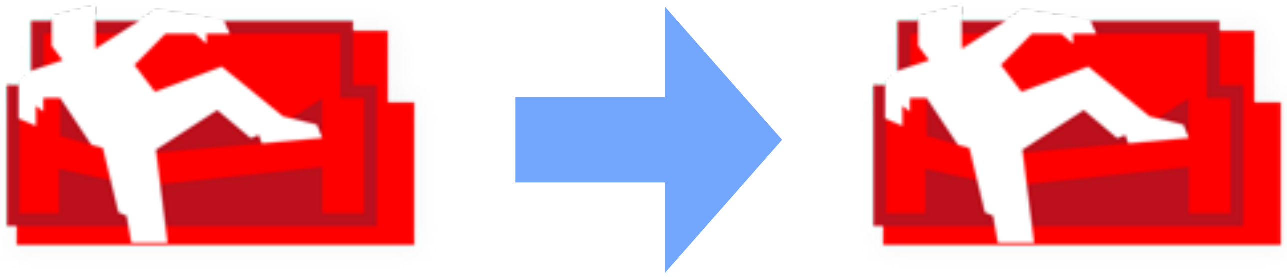
- Collect data, Store Data
- Query Data
- Render & Transform
- everything you can do but decentralized

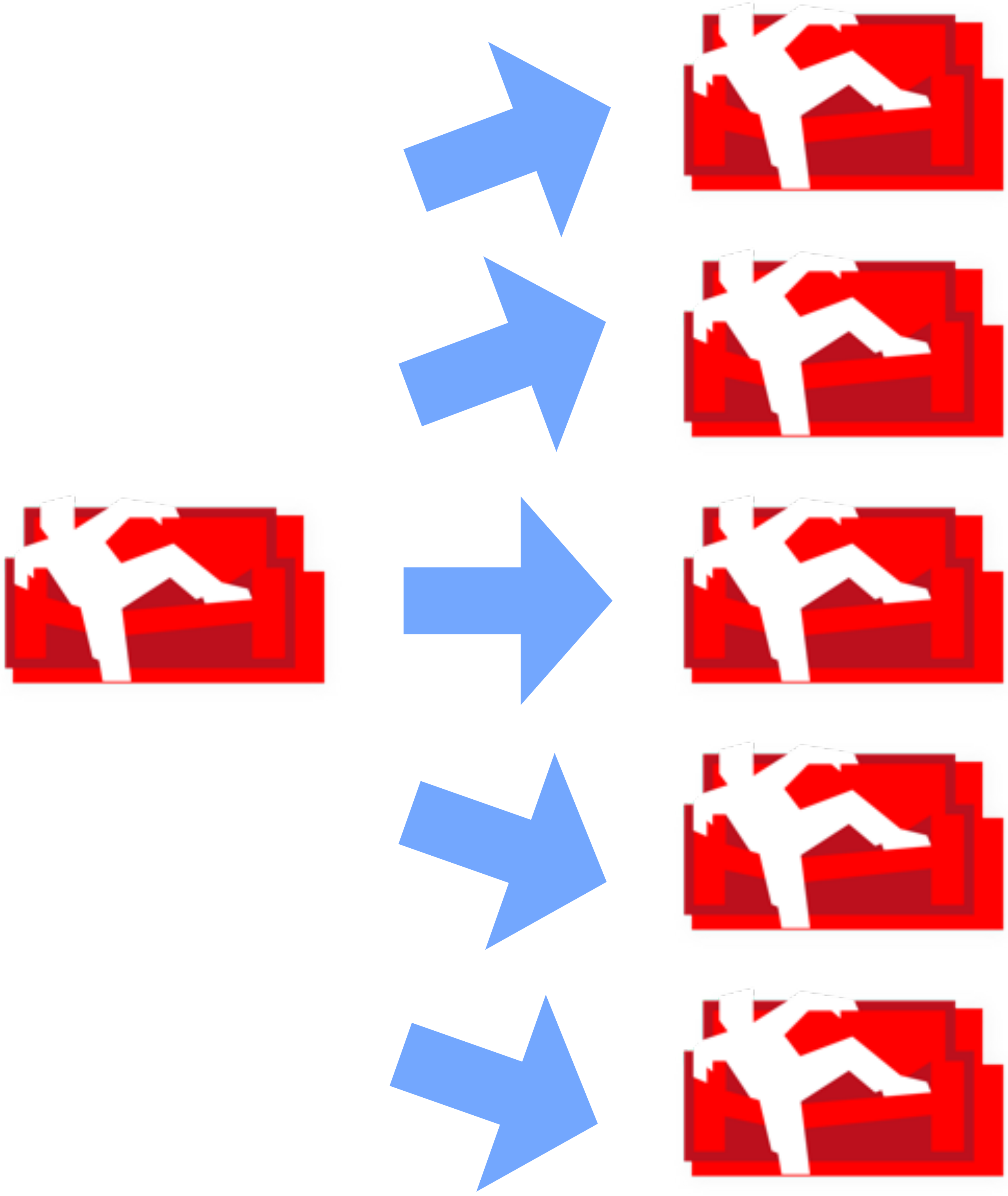
data platform?

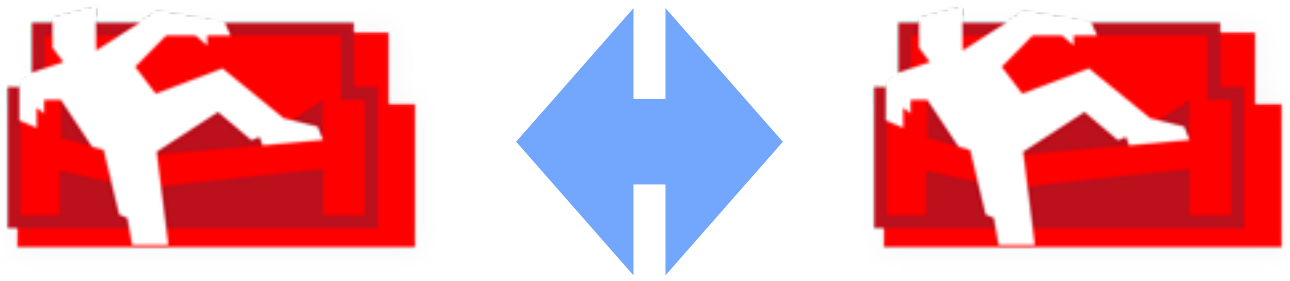
- because I know it....
- robust
- easily hackable
- replication & `_changes`
- couchapps

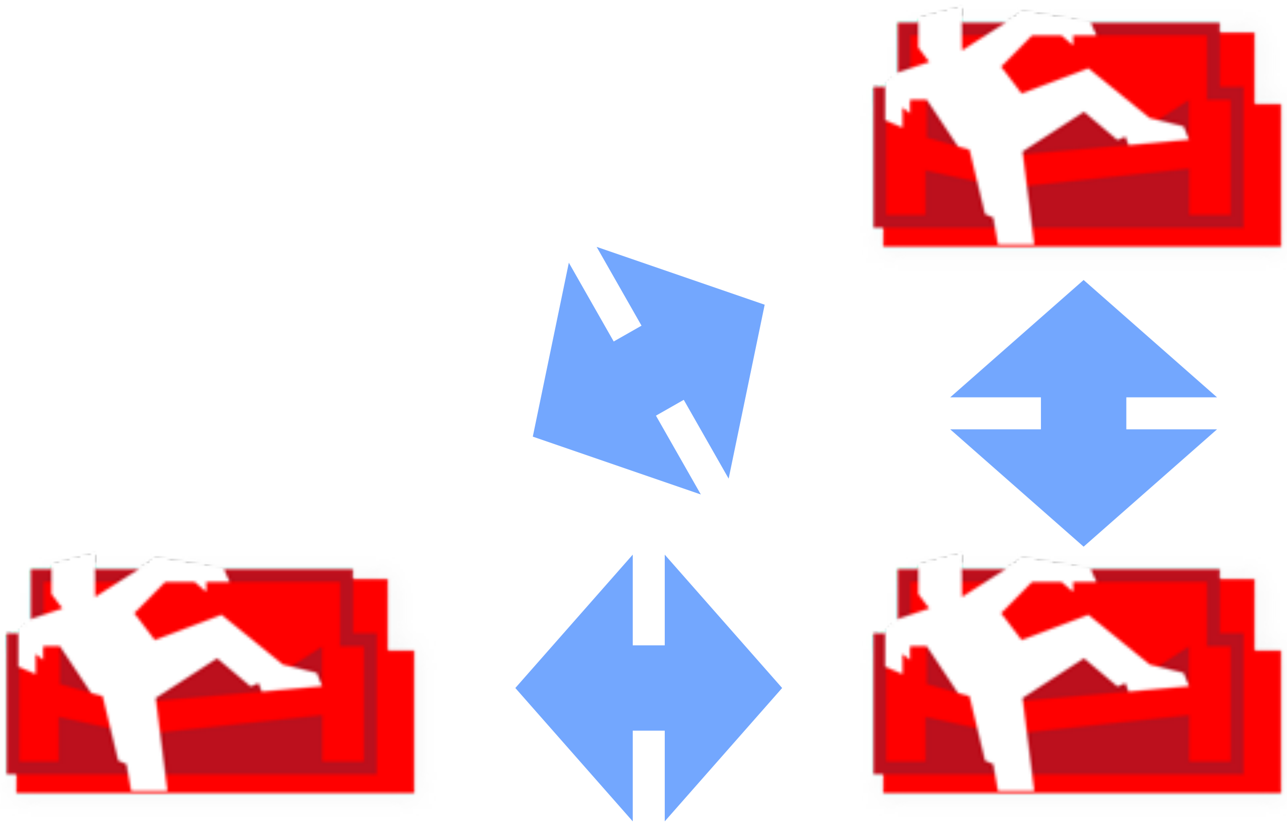
Why Apache CouchDB?

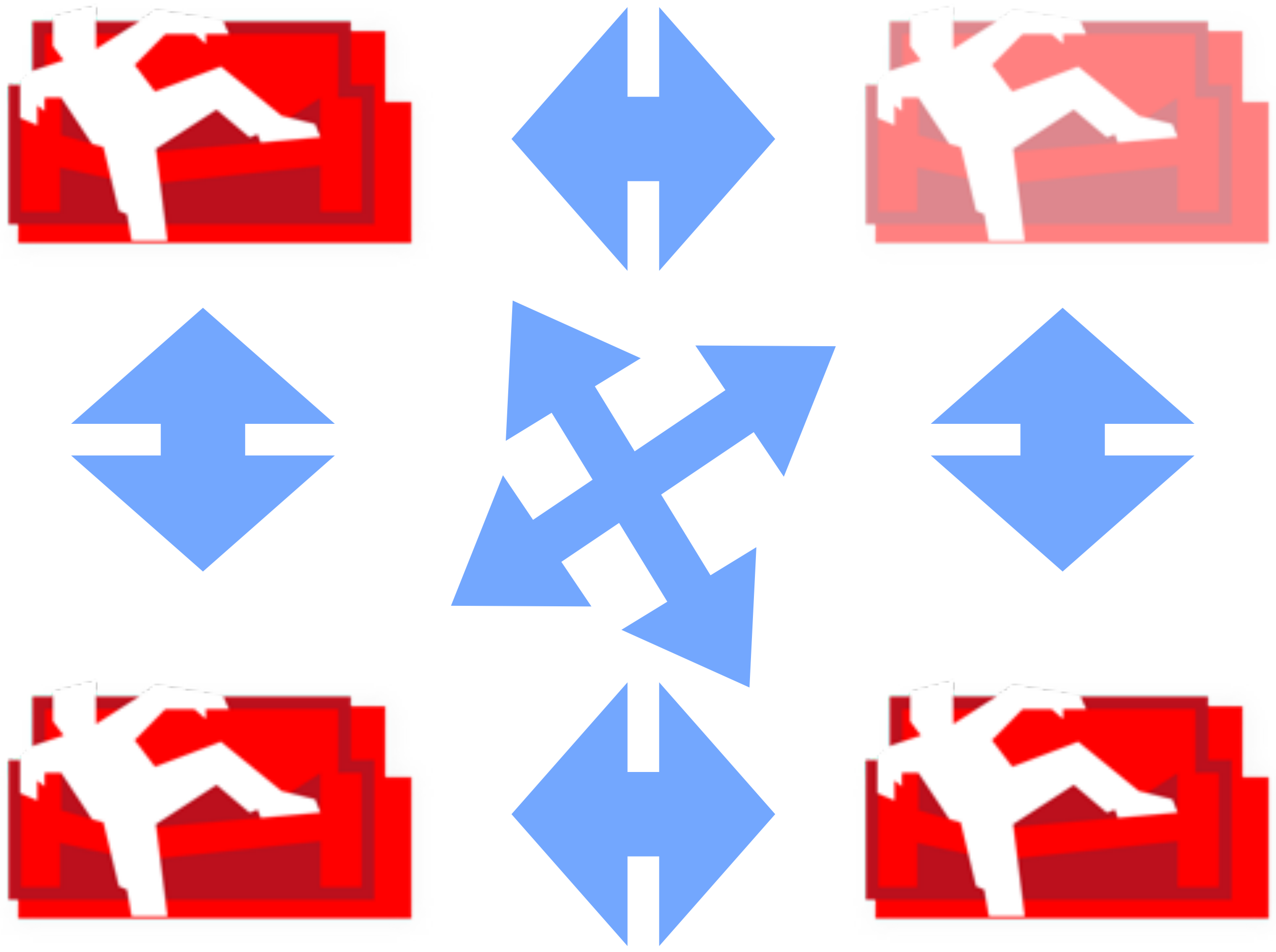
INCREMENTAL REPLICATION

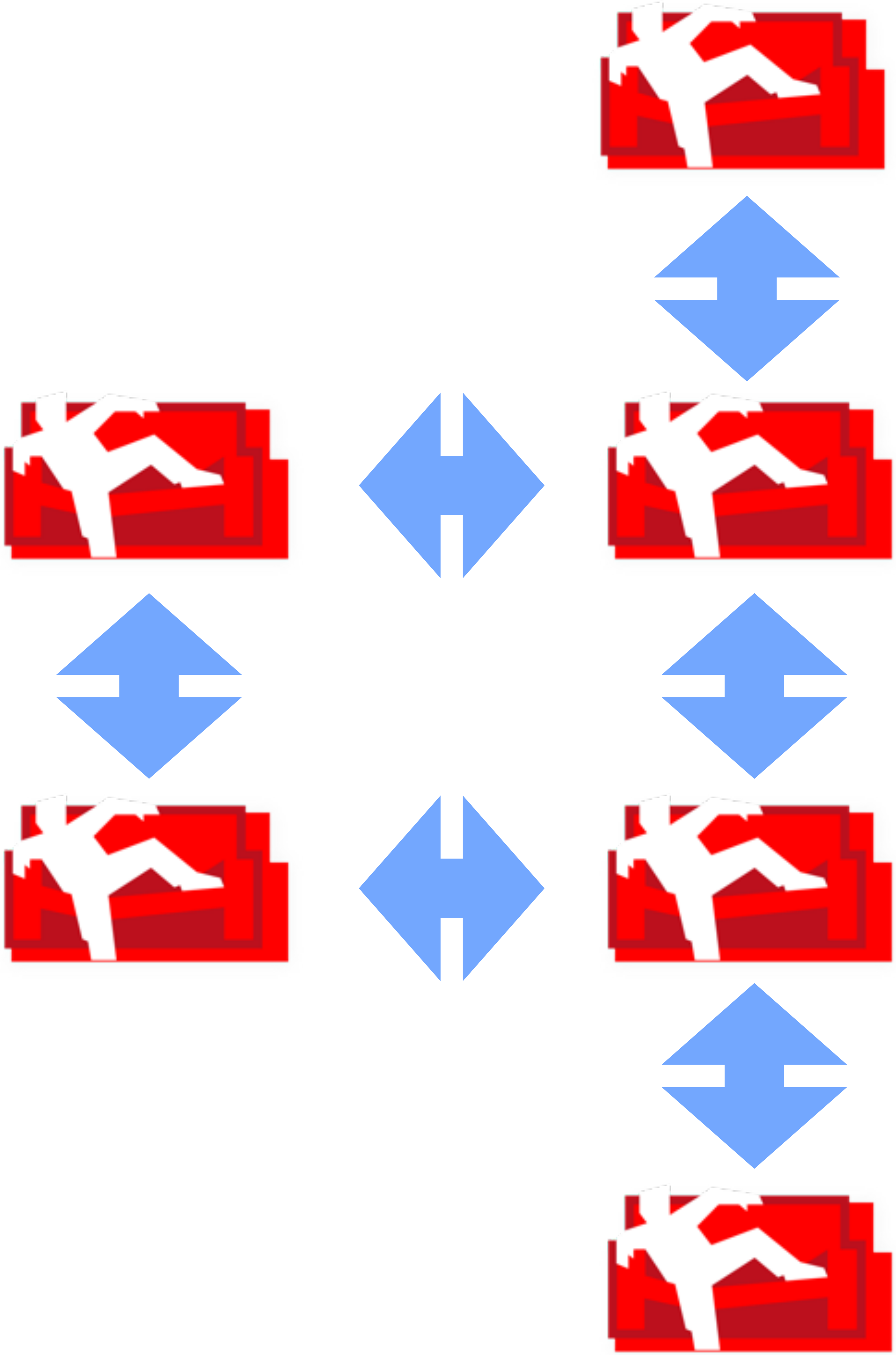


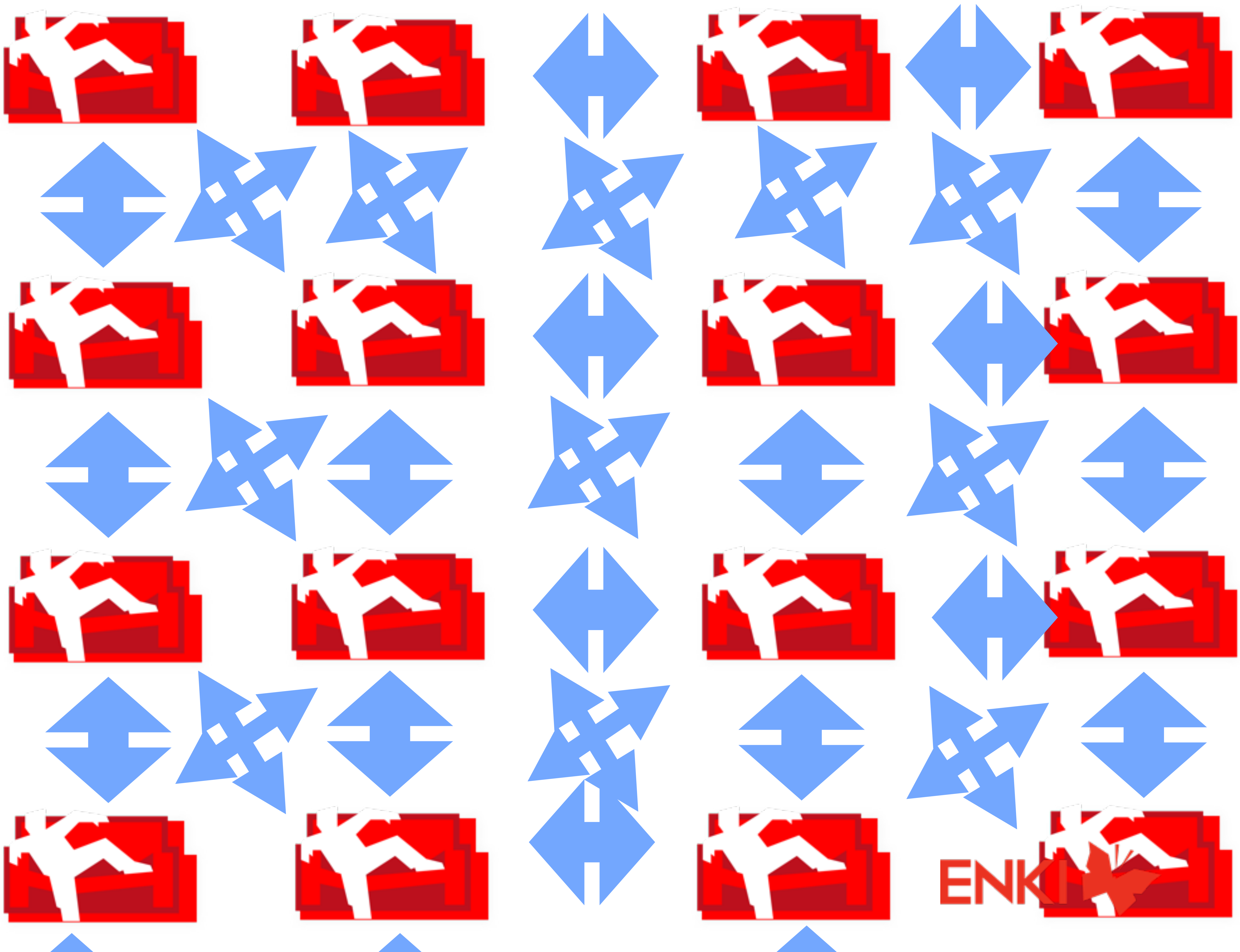











ENKI

- changes feed
- Collect a group of Document/Revisions ID pairs
- send them to the target database
- GET each revisions from the source Database

replication in a nutshell

- Most of the Apache CouchDB code in Erlang
- functional language
- adapted to the product
- we are building a network

why Erlang?

packaging & fork

- distributing is about packaging well
- building an Erlang release....
- ...with rebar
- and make Apache CouchDB even more hackable
- improve some parts

`packaging & fork`

- Autotools & gnu frivolities
- Spidermonkey
- Not embeddable
- Don’t provide Erlang releases

`state of packaging in couchdb`

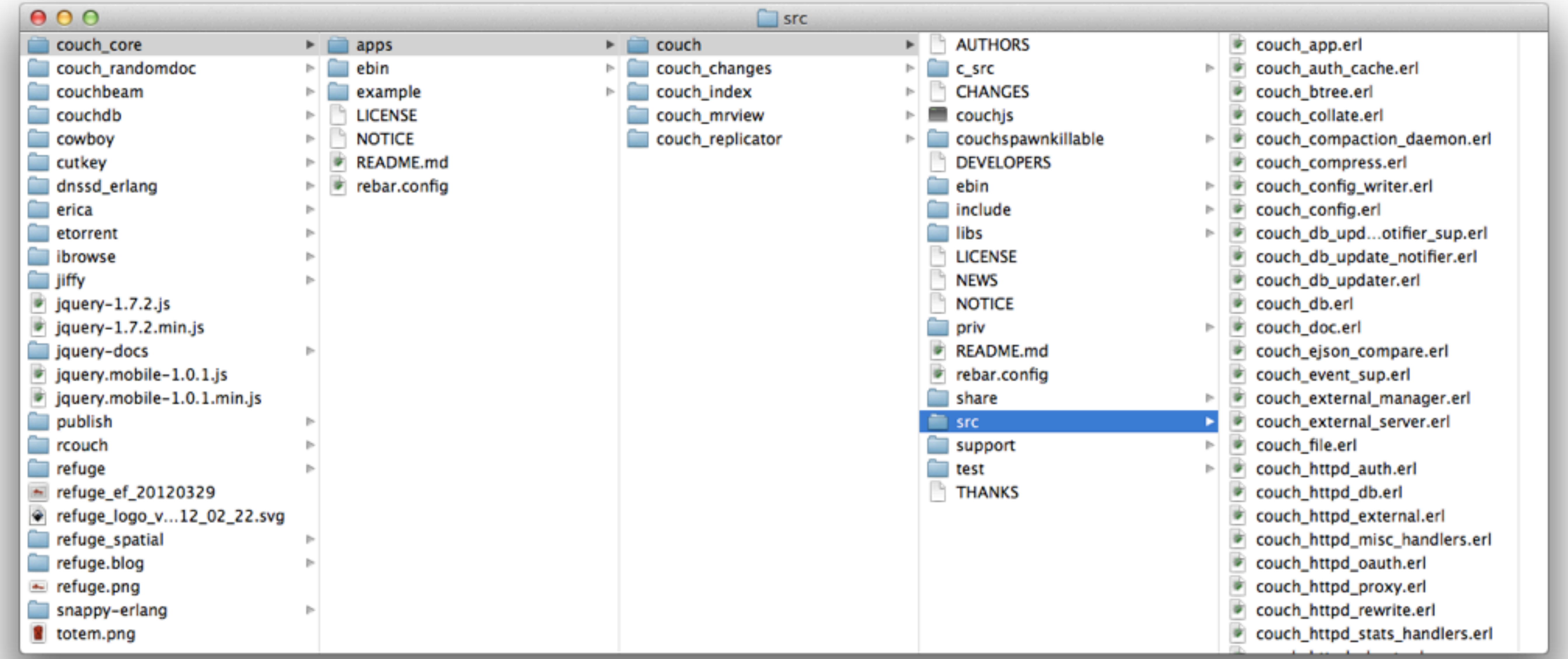
- nifs & driver: difficulties with multiplatform
- windows support: should be more opiniated (which compiler to use...)
- Still a question: how to manage IOS, Android

rebar challenges

- split in small apps
- improve supervision
- new mochiweb, jiffy, lager, ...

rework the code

ENKI



src

couch_core
couch_randomdoc
couchbeam
couchdb
cowboy
cutkey
dnssd_erlang
erica
etorrent
ibrowse
jiffy
jquery-1.7.2.js
jquery-1.7.2.min.js
jquery-docs
jquery.mobile-1.0.1.js
jquery.mobile-1.0.1.min.js
publish
rcouch
refuge
refuge_ef_20120329
refuge_logo_v...12_02_22.svg
refuge_spatial
refuge.blog
refuge.png
snappy-erlang
totem.png

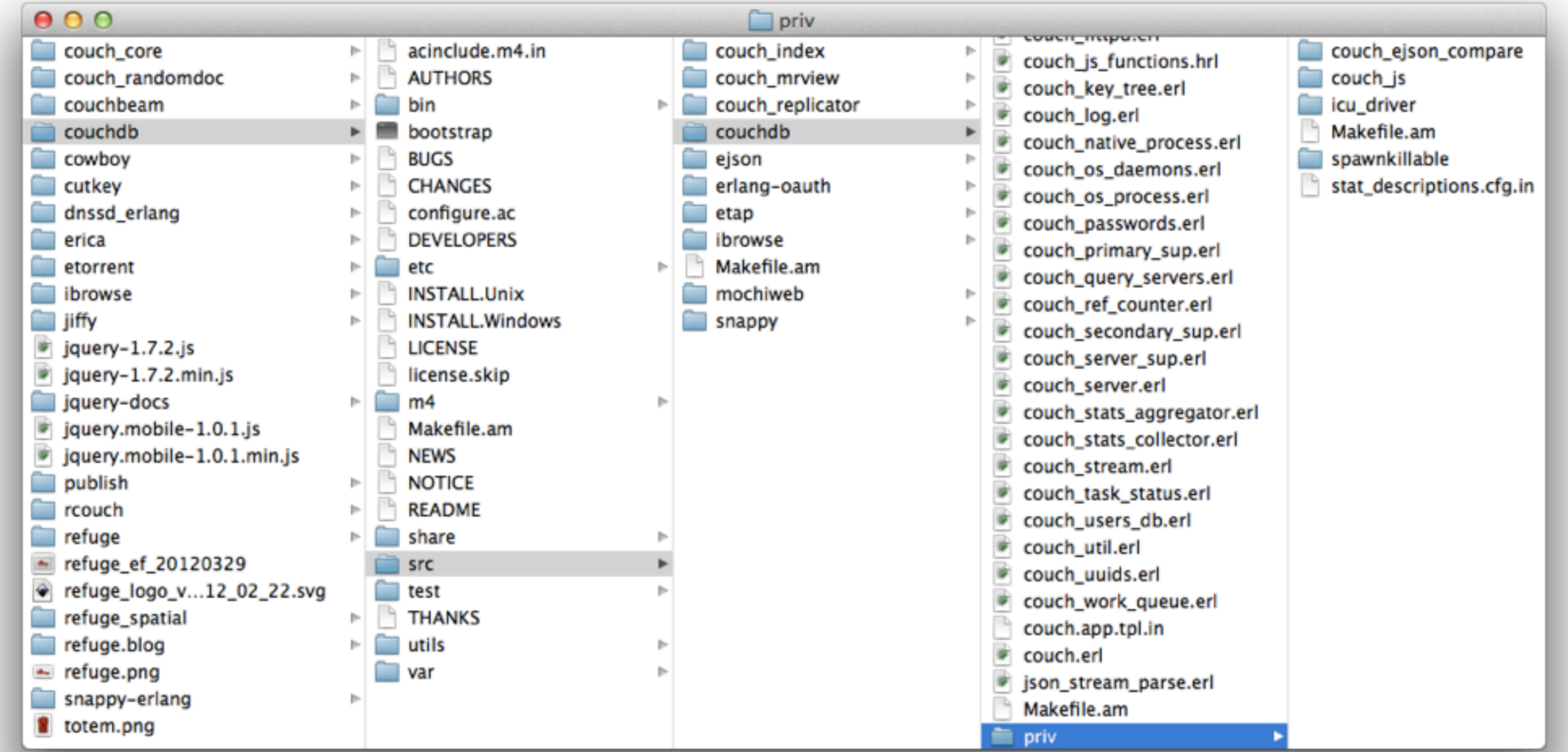
apps
ebin
example
LICENSE
NOTICE
README.md
rebar.config

couch
couch_changes
couch_index
couch_mrview
couch_replicator

AUTHORS
c_src
CHANGES
couchjs
couchspawnkillable
DEVELOPERS
ebin
include
libs
LICENSE
NEWS
NOTICE
priv
README.md
rebar.config
share
src
support
test
THANKS

couch_app.erl
couch_auth_cache.erl
couch_btree.erl
couch_collate.erl
couch_compaction_daemon.erl
couch_compress.erl
couch_config_writer.erl
couch_config.erl
couch_db_upd...otifier_sup.erl
couch_db_update_notifier.erl
couch_db_updater.erl
couch_db.erl
couch_doc.erl
couch_ejson_compare.erl
couch_event_sup.erl
couch_external_manager.erl
couch_external_server.erl
couch_file.erl
couch_httpd_auth.erl
couch_httpd_db.erl
couch_httpd_external.erl
couch_httpd_misc_handlers.erl
couch_httpd_oauth.erl
couch_httpd_proxy.erl
couch_httpd_rewrite.erl
couch_httpd_stats_handlers.erl

ENKI



priv

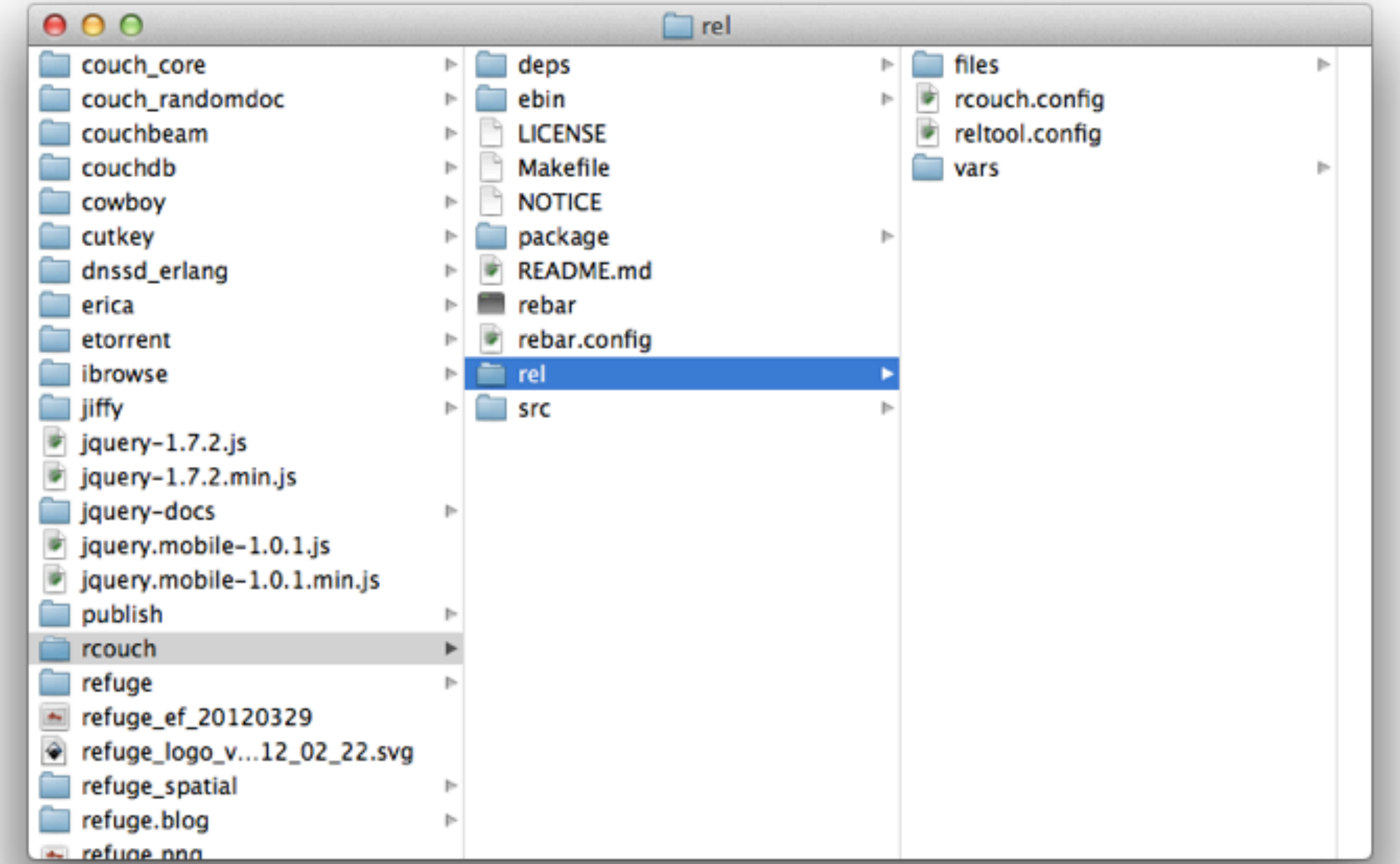
couch_core
couch_randomdoc
couchbeam
couchdb
cowboy
cutkey
dnssd_erlang
erica
etorrent
ibrowse
jiffy
jquery-1.7.2.js
jquery-1.7.2.min.js
jquery-docs
jquery.mobile-1.0.1.js
jquery.mobile-1.0.1.min.js
publish
rcouch
refuge
refuge_ef_20120329
refuge_logo_v...12_02_22.svg
refuge_spatial
refuge.blog
refuge.png
snappy-erlang
totem.png

acinclude.m4.in
AUTHORS
bin
bootstrap
BUGS
CHANGES
configure.ac
DEVELOPERS
etc
INSTALL.Unix
INSTALL.Windows
LICENSE
license.skip
m4
Makefile.am
NEWS
NOTICE
README
share
src
test
THANKS
utils
var

couch_index
couch_mrview
couch_replicator
couchdb
ejson
erlang-oauth
etap
ibrowse
Makefile.am
mochiweb
snappy

couch_js_functions.hrl
couch_key_tree.erl
couch_log.erl
couch_native_process.erl
couch_os_daemons.erl
couch_os_process.erl
couch_passwords.erl
couch_primary_sup.erl
couch_query_servers.erl
couch_ref_counter.erl
couch_secondary_sup.erl
couch_server_sup.erl
couch_server.erl
couch_stats_aggregator.erl
couch_stats_collector.erl
couch_stream.erl
couch_task_status.erl
couch_users_db.erl
couch_util.erl
couch_uuids.erl
couch_work_queue.erl
couch.app.tpl.in
couch.erl
json_stream_parse.erl
Makefile.am
priv

couch_ejson_compare
couch_js
icu_driver
Makefile.am
spawnkillable
stat_descriptions.cfg.in



rel

couch_core
couch_randomdoc
couchbeam
couchdb
cowboy
cutkey
dnssd_erlang
erica
etorrent
ibrowse
jiffy
jquery-1.7.2.js
jquery-1.7.2.min.js
jquery-docs
jquery.mobile-1.0.1.js
jquery.mobile-1.0.1.min.js
publish
rcouch
refuge
refuge_ef_20120329
refuge_logo_v...12_02_22.svg
refuge_spatial
refuge.blog
refuge.png

deps
ebin
LICENSE
Makefile
NOTICE
package
README.md
rebar
rebar.config
rel
src

files
rcouch.config
reltool.config
vars

refuge protocol - the mesh

- Peers are linked with trust relationships.
- Each refuge node user generate an 1024 RSA key pair .
- The pubkey is used as the identity among peers.
- Keys are exchanged manually or using the social graph

identity

- hubs: maintain a list of registered users
- public hubs
- untrusted users used to forward lookups
- untrusted users can access to the public data

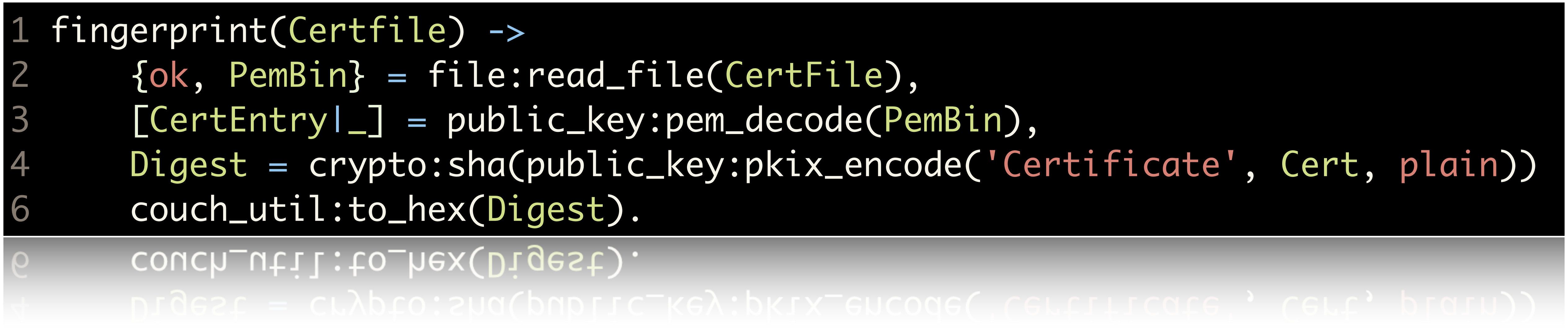
groups & untrused users

- don’t try to generate the RSA key (*erl_make_cert.erl* is just for testing)
- use openssl to generate it
- use a port
- or an erlang driver: **cutkey** http://github.com/andrewjt/cutkey

generating RSA keys in Erlang

- create/read a certificate
- use & parse keys
- use it to return public & private key used by the ***crypto*** module.
- calculate a fingerprint

publickey - handy functions



```erlang
fingerprint(Certfile) ->
    {ok, PemBin} = file:read_file(CertFile),
    [CertEntry|_] = public_key:pem_decode(PemBin),
    Digest = crypto:sha(public_key:pkix_encode('Certificate', Cert, plain))
    couch_util:to_hex(Digest).
```

public_key - handy functions

- validate_fun: if you only want to check and validate a certificate (notary mode)
- ssl:peercert/1
- couch_auth_ssl & https

ssl - authentication and more

- don’t try to generate the RSA key
- use openssl to generate it
- use a port
- or an erlang driver: **cutkey**
  http://github.com/andrewjt/cutkey

generating RSA keys in Erlang

- manual key exchange/ link
- dns-sd to get them in the lan
- hubs
- DHT

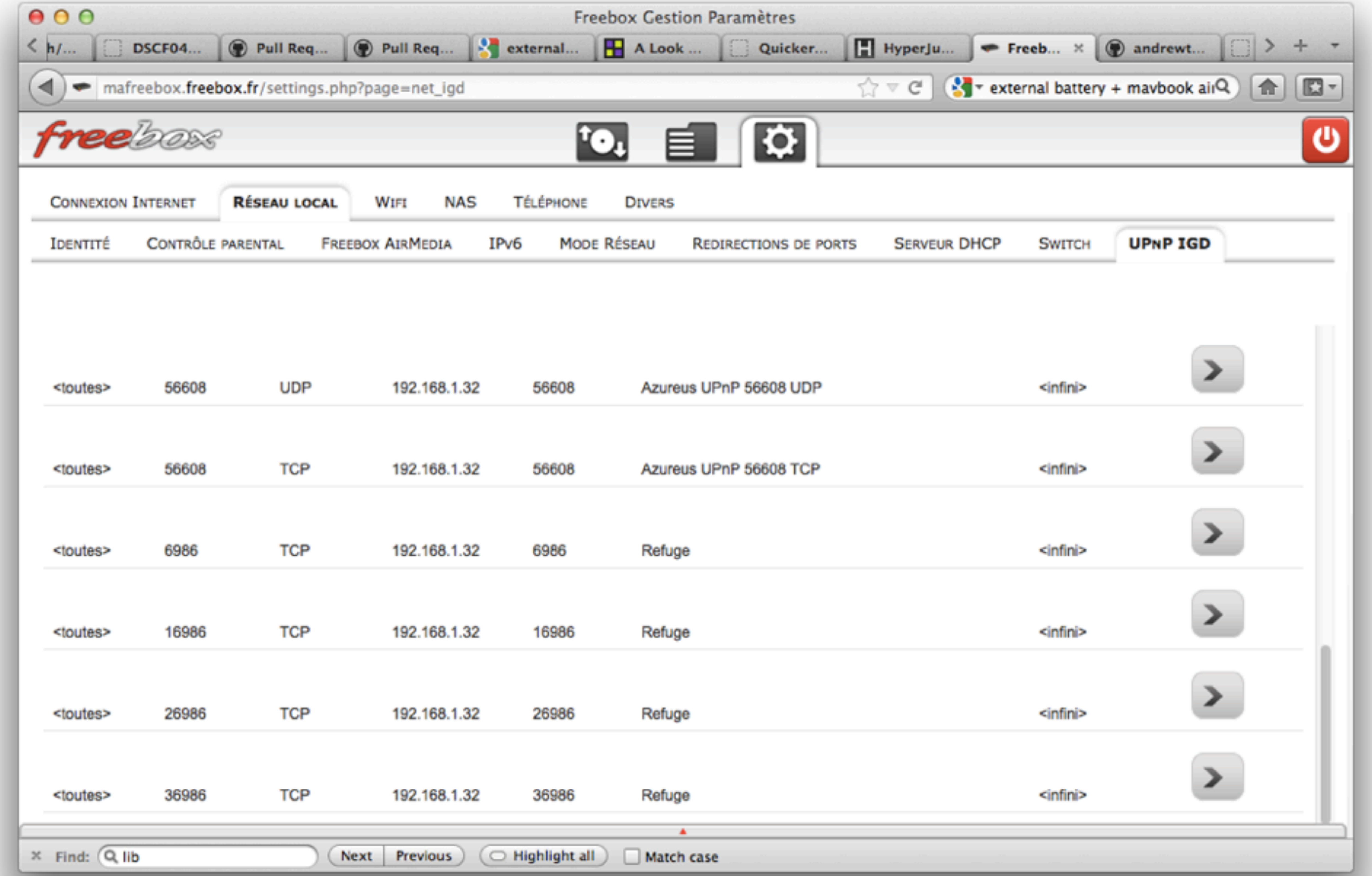
discover nodes

- bonjour protocol
- dns-sd: http://github.com/andrewtj/dns-sd
- upnp: make them available to the world

```
dns-sd
```

- refuge 0.5 - DNSSD support https://vimeo.com/39416245

DEMO



<toutes>	56608	UDP	192.168.1.32	56608	Azureus UPnP 56608 UDP	<infini>
<toutes>	56608	TCP	192.168.1.32	56608	Azureus UPnP 56608 TCP	<infini>
<toutes>	6986	TCP	192.168.1.32	6986	Refuge	<infini>
<toutes>	16986	TCP	192.168.1.32	16986	Refuge	<infini>
<toutes>	26986	TCP	192.168.1.32	26986	Refuge	<infini>
<toutes>	36986	TCP	192.168.1.32	36986	Refuge	<infini>

- gen_event
- nodeup / nodedown event
- monitor nodes

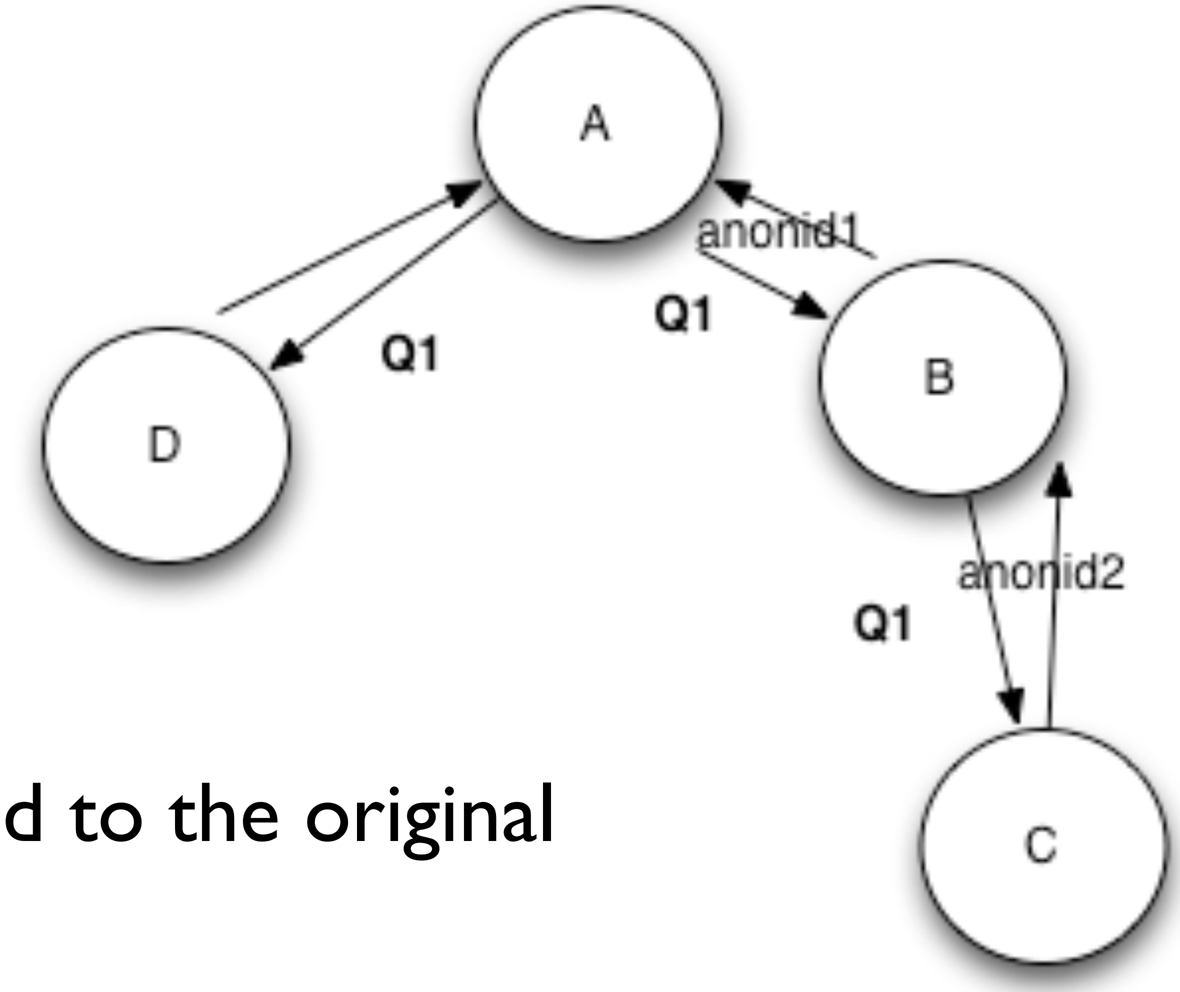
```
nodes generate events
```

- opportunistic replication
- doc creation on demand
- application interatcons

opportunistic actions

- flooding nodes with the query
- content id: crypted with pubkey

query the nodes



- query are flooded
- result are forwarded to the original requester
- follow the path

query the nodes

other things we do

- view changes (branch view_changes)
- forked geocouch, support knn
- chained m/r (wip)
- distribute couchapps & jobs
- ...

discover nodes

The refuge project:
http://refuge.io

@benoitc
benoitc@refuge.io