

Sharing Data, Hiding Complexity

with RDF and Clojure

Malcolm Sparks
James Henderson

TechMesh 2012

Data: Setting the scene

Financial, regulatory pressures in banking are presenting IT with new challenges.

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Data: Setting the scene

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Financial, regulatory pressures in banking are presenting IT with new challenges.

How can we process data :-

Data: Setting the scene

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Financial, regulatory pressures in banking are presenting IT with new challenges.

How can we process data :-

- ▶ Faster (as always)

Data: Setting the scene

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Financial, regulatory pressures in banking are presenting IT with new challenges.

How can we process data :-

- ▶ Faster (as always)
- ▶ At greater volumes (risk sensitivities - HistSim)

Data: Setting the scene

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Financial, regulatory pressures in banking are presenting IT with new challenges.

How can we process data :-

- ▶ Faster (as always)
- ▶ At greater volumes (risk sensitivities - HistSim)
- ▶ More precisely (finer granularity - Dodd Frank initial margin)

Conway's Law

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

“Organizations which design systems ... are constrained to produce designs which are copies of the communication structures of these organizations”

Mel Conway (1968)

Conway in action

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

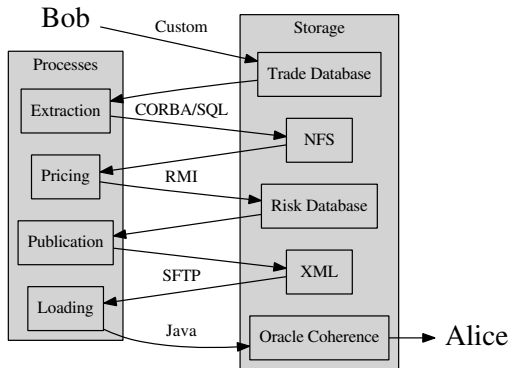


Figure : An example communication structure

Q: Do Bob and Alice get the same view of the data?

What does it all mean?

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

CommonReference Narrative NumericId LegalEntity
ManagementArea GLRegion LocalBankingTrading
AggregationLevel SettlementEngineSource ISIN CUSIP
PrivatePlacement OtherSecurityId FinancialProductType
ReportingCurve InternalExternalFlag BuySell Fas133Flag
SpvId MtnId LinkageReference LinkageReference2 NodeRef
ReasonCode YNFlag

Observation: Data doesn't travel well

What went wrong?

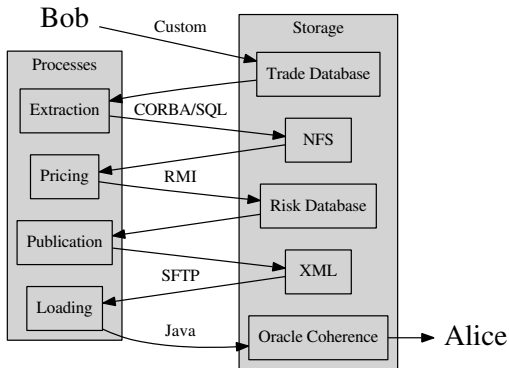
Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs



Conclusion: We need our data models to span across our systems.

Does modelling scale?

As we scale up :-

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Does modelling scale?

As we scale up :-

- Complexity increases

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Does modelling scale?

As we scale up :-

- ▶ Complexity increases
- ▶ Deviations increase

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Does modelling scale?

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

As we scale up :-

- ▶ Complexity increases
- ▶ Deviations increase
- ▶ Concepts become more abstract

Does modelling scale?

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

As we scale up :-

- ▶ Complexity increases
- ▶ Deviations increase
- ▶ Concepts become more abstract
- ▶ Understanding decreases

Does modelling scale?

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

As we scale up :-

- ▶ Complexity increases
- ▶ Deviations increase
- ▶ Concepts become more abstract
- ▶ Understanding decreases

As data escapes its domain model :-

Does modelling scale?

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

As we scale up :-

- ▶ Complexity increases
- ▶ Deviations increase
- ▶ Concepts become more abstract
- ▶ Understanding decreases

As data escapes its domain model :-

- ▶ It loses assumptions (meaning)

Does modelling scale?

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

As we scale up :-

- ▶ Complexity increases
- ▶ Deviations increase
- ▶ Concepts become more abstract
- ▶ Understanding decreases

As data escapes its domain model :-

- ▶ It loses assumptions (meaning)
- ▶ Local identifiers fail to become global ones.

Implications

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

- ▶ Our primary tool for flexible data processing (data modelling) does not scale!

Implications

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

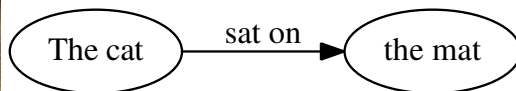
Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

- ▶ Our primary tool for flexible data processing (data modelling) does not scale!
- ▶ We are fire-fighting this problem with a code hosepipe!

Introduction to RDF



```
http://animals.com/ginger
  http://actions.com/satOn http://tech-mesh-hotel.co.uk/mat ;
  rdf:type http://animals.com/Cat ;
.
```

```
http://actions.com/satOn
  rdfs:label "sat on" ;
  rdfs:comment "The action of sitting on a given object." ;
.
```

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

RDF features

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

RDF features

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

- ▶ URIs - globally unique identification and transparency

RDF features

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

- ▶ URIs - globally unique identification and transparency
- ▶ Namespaces (for federated governance)

RDF features

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

- ▶ URIs - globally unique identification and transparency
- ▶ Namespaces (for federated governance)
- ▶ Storage and transfer neutrality

RDF features

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

- ▶ URIs - globally unique identification and transparency
- ▶ Namespaces (for federated governance)
- ▶ Storage and transfer neutrality
- ▶ Metadata

Observations

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

RDF has the power to break silos

RDF is a non-disruptive way of applying federated data governance

The RAP

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

The RAP

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

- ▶ Exposing the data in the 'ODS'

The RAP

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

- ▶ Exposing the data in the 'ODS'
- ▶ The tables in the ODS are built from a meta-model

The RAP

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

- ▶ Exposing the data in the 'ODS'
- ▶ The tables in the ODS are built from a meta-model
- ▶ We get the meta-model in XML format

So why do we transform it to RDF?

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

So why do we transform it to RDF?

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

- ▶ Disambiguation of terms

So why do we transform it to RDF?

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

- ▶ Disambiguation of terms
- ▶ Merging different data sources

Exposing a data dictionary

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs



Forming a View

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Forming a View

Sharing Data,
Hiding Complexity

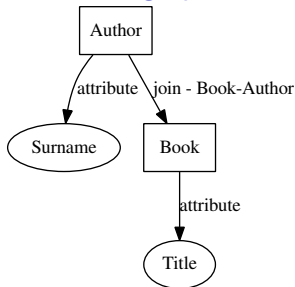
Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

A View in graph form:



Forming a View

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

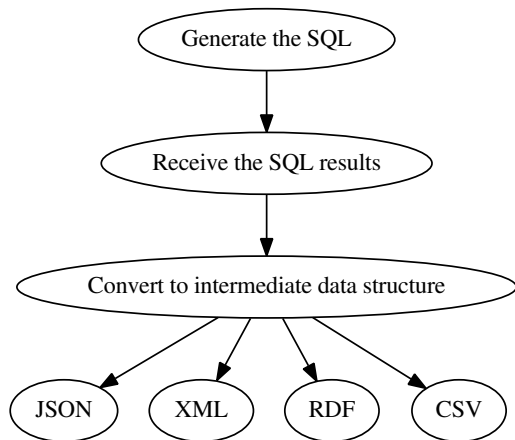
RDF and the RAP

The Power of
Graphs

RDF storage of a view:

```
views:demo-view
  <select> [ <node> rap:Author ;
            <selectedAttribute> [ <attribute> rap:Author-Name ; ] ;
            <childNodes> [ <node> rap:Book ;
                          <relationship> palace:Book-Author ;
                          <selectedAttribute>
                            [ <attribute> rap:Book-Title ; ] ; ] ; ]
```

Delivering the View



Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

The Power of Graphs

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Storing config in RDF

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

```
<hosts/p21> <instance> <instances/PROD1> .
```

```
<hosts/u11> <instance> <instances/UAT2> .
```

```
<hosts/d30> <instance> <instances/SIT1> .
```

Storing config in RDF

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

```
<hosts/p21> cmdb:hostName "p21.db.com" .
```

```
<instances/PROD1> <service> <services/PROD1/http> .
```

```
<services/PROD1/http> cmdb:port 8080 .
```

Why RDF for config?

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Why RDF for config?

- ▶ Disambiguation
- ▶ Meta-data

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Why RDF for config?

- ▶ Disambiguation
- ▶ Meta-data
- ▶ Querying the config

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Why RDF for config?

- ▶ Disambiguation
- ▶ Meta-data
- ▶ Querying the config
- ▶ Multiple config sources

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Querying the config

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Graph-Zip:

<https://github.com/james-henderson/graph-zip>

Querying the config

Sharing Data,
Hiding Complexity

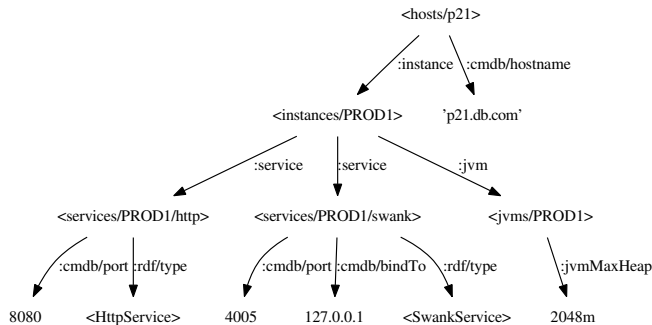
Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

An example config graph:



Querying the config

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Importing the graph:

```
(def config-map
  (build-in-memory-graph
    [{:subject "<hosts/p21>" :property :hostname :object "p21.db.com"}
     {:subject "<hosts/p21>" :property :instance :object "<instances/PROD1>"}

     {:subject "<instances/PROD1>" :property :service :object "<services/PROD1/swank>"}
     {:subject "<services/PROD1/swank>" :property :port :object "4005"}
     {:subject "<services/PROD1/swank>" :property :rdf/type :object "<SwankService>"}

     {:subject "<instances/PROD1>" :property :service :object "<services/PROD1/http>"}
     {:subject "<services/PROD1/http>" :property :port :object "8080"}
     {:subject "<services/PROD1/http>" :property :binds-to :object "127.0.0.1"}
     {:subject "<services/PROD1/http>" :property :rdf/type :object "<HttpService>"}

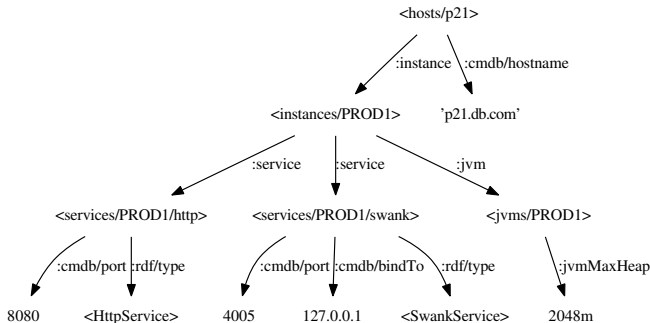
     {:subject "<instances/PROD1>" :property :jvm :object "<jvms/PROD1>"}
     {:subject "<jvms/PROD1>" :property :jvmMaxHeap :object "2048m"}]))
```

Querying the config

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

An example config graph:



Data Loses Fidelity
As You Move It

RDF and the RAP

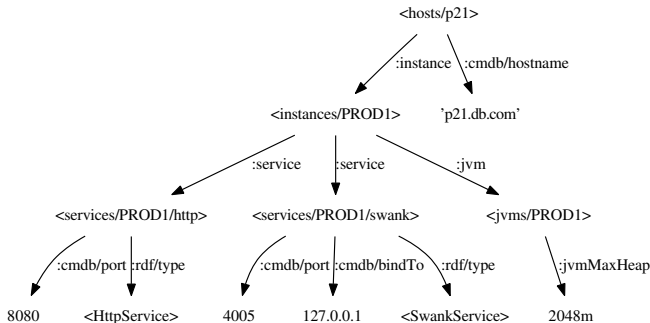
The Power of
Graphs

Building the zipper:

```
(def prod-host-zipper (graph-zip config-map "<hosts/p21>"))
```

Querying the config

An example config graph:



Navigating the graph:

```
(zip-> prod-host-zipper
:instance
:service
node)

;; -> ("<services/PROD1/swank>" "<services/PROD1/http>")
```

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

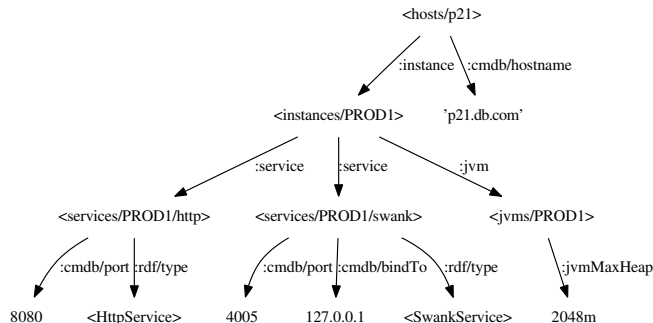
Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Querying the config

An example config graph:



Finding out the Swank port:

```
(zip1-> prod-host-zipper
:instance
:service
[(prop= :rdf/type "<SwankService>")]
:cldb/port
node)
```

```
:: -> "4005"
```

Querying the config

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Graph-Zip:

<https://github.com/james-henderson/graph-zip>

Neo-Zip:

<https://github.com/james-henderson/neo-zip>

Combining Zippers with RAP Views

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Combining Zippers with RAP Views

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

In XML:

```
<hosts>
  <host>
    <hostname>p21.db.com</hostname>
    <instances>
      <instance>
        <label>PROD1</label>
        <services>
          <service>
            <type>swank</type>
            <port>4005</port>
            <bindsTo>127.0.0.1</bindsTo>
          </service>
          <service>
            <type>http</type>
            <port>8080</port>
          </service>
        </services>
      </instance>
    </instances>
  </host>

  ...

</hosts>
```

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Combining Zippers with RAP Views

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Using traditional zip syntax:

```
(for [host (xml-> doc :host)]
  {:hostname (xml1-> host :hostname text)
   :instances (for [instance (xml-> host :instances :instance)]
     {:label (xml1-> instance :label text)
      :services (for [service (xml-> instance :services :service)]
        {:type (xml1-> service :type text)
         :port (xml1-> service :port text)
         :bindsTo (xml1-> service :bindsTo text)}}))})
```


Combining Zippers with RAP Views

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Using a datalog syntax:

```
(let [db (d/db conn)]
  (for [[host hostname] (d/q '(:find [?h ?name]
                                :where [[?h :type :host]
                                         [?h :hostname ?name]])]
    db])

{:hostname hostname
 :instances (for [[instance label]
                  (d/q '(:find [?inst ?label]
                            :in [$ ?host]
                            :where [[?host :instance ?inst]
                                   [?inst :label ?label]])]
    db host)]
 :label label
 :services (for [[service type port binds-to]
                  (d/q '(:find [?service ?type ?port ?binds-to]
                            :in [$ ?instance]
                            :where [[?instance :service ?service]
                                   [?service :port ?port]
                                   [?service :type ?type]
                                   [?service :binds-to ?binds-to]])]
    db instance)]
 :type type
 :port port
 :binds-to binds-to}})))))
```

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

Combining Zippers with RAP Views

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

An enhanced zipper query:

```
(xml-> doc
  :host
  {:hostname ^:1 [:hostname text]
   :instances [:instances
               :instance
               {:instance-name ^:1 [:label text]
                :services [:services
                           :service
                           {:type ^:1 [:type text]
                            :port ^:1 [:port text]
                            :interface ^:1 [:binds-to text]]}}]])
```

Combining Zippers with RAP Views

An enhanced zipper query:

```
(xml-> doc
  :host
  {:hostname ^:1 [:hostname text]
   :instances [:instances
                :instance
                {:instance-name ^:1 [:label text]
                 :services [:services
                             :service
                             {:type ^:1 [:type text]
                              :port ^:1 [:port text]
                              :interface ^:1 [:binds-to text]]}]]}})
```

Resulting data structure:

```
[{:hostname "p21.db.com"
  :instances [{:instance-name "PROD1"
                :services [{:type "swank"
                             :port "4005"
                             :interface "127.0.0.1"
                             {:type "http"
                              :port "8080"}]}]}]
{:hostname "u11.db.com"
  :instances [{:instance-name "UAT1" :services ...}
               {:instance-name "UAT2" :services ...}]}
{:hostname "d30.db.com"
  :instances [{:instance-name "SIT1" :services ...}]}
```

Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs



Sharing Data,
Hiding Complexity

Malcolm Sparks
James Henderson

Data Loses Fidelity
As You Move It

RDF and the RAP

The Power of
Graphs

James Henderson

@jhenderson_89

<https://github.com/james-henderson>

Malcolm Sparks

@malcolmsparks

<https://github.com/malcolmsparks>