Nigel Roberts

Director

Channel Islands Registry (.GG and .JE)

Consulting Engineer

TLD Registry Solutions LLC
Technical DB consultancy

nigel@roberts.gg
http://www.roberts.gg
Databases

- Oracle
- MySQL
- PostgresSQL
Databases

ORACLE

- enterprise-scale
- fast
- functional
- reliable
Databases

ORACLE

- expensive
Databases

MySQL

- free
Databases

MySQL

- free
- extremely limited
Databases

MySQL

- Free
- Extremely Limited.
Databases

PostgreSQL

- scalable
- reliable
- functional
- fast
- (... and FREE)
Databases

CoCCA Registry
is built on PostgreSQL
.GG, .JE & CoCCA Registry

- formal evaluation in 2009
- deployment in March 2010
Data Cleansing and Migration

- **extract** the data
  - from several legacy systems
- **clean** the data
- **format** the data
- **load** the data
Island Networks

- 'thick' registry model
- *resellers* not registrars
- bills (direct) registrants
  - by post

*However* ... is moving to registry-registrar model
Core Concept #1

“Views”
Core Concept #1

“Views”

(a nice view of Belgium)
Core Concept #1

“Views”

A virtual representation of data (e.g. a subset) presented in a particular way.

A read-only interface created from query and programming logic for the Billing Process.
Legacy System

- connect to and integrate with new backend
  - seamlessly
  - invisibly
    (i.e. no one should notice!)
- “core interface views”
- upload to new registry
Backup and Availability
Backup and Availability

- *PG_DUMP*
Backup and Availability

- `PG_DUMP`
- filesystem copy
Backup and Availability

- `PG_DUMP`
- filesystem copy
- “frozen filesystem”
Backup and Availability

- \textit{PG\_DUMP}
- filesystem copy
- “frozen filesystem”
- Point of Failure
Backup and Availability

- `PG_DUMP`
- filesystem copy
- “frozen filesystem”
- Point of Failure
- Point In Time Recovery (PITR, continuous archiving)
Backup and Availability

Point In Time Recovery

• “Recover right up to the point it all went wrong!”
Backup and Availability

Point In Time Recovery
  - like having your own time machine...
Backup and Availability

Point In Time Recovery

...or even an 'UNF**KIT' Button.
Backup and Availability

- `PG_DUMP`
- filesystem copy
- “frozen filesystem”
- Point of Failure
- Point In Time Recovery

and more ...
Backup and Availability

- **PG_DUMP**
- Filesystem copy
- Frozen filesystem
- Point of Failure
- Point In Time Recovery
- High Availability Warm Standby
Backup and Availability

- *PG_DUMP*
- Filesystem copy
- Frozen filesystem
- Point of Failure
- Point In Time Recovery
- High Availability Warm Standby
- . . . even more . . . . .
Nigel Roberts

Director
Channel Islands Registry (.GG and .JE)

Consulting Engineer
TLD Registry Solutions LLC
Technical DB consultancy

nigel@roberts.gg
http://www.roberts.gg
Data
Registry Data

- Data Quality  
  - matters

- Poor Quality Data  
  - has real costs  
    - to the registry  
    - to registrars  
    - the public
In an ideal world

- Registry contact data should be
  - accurate
  - up-to-date
  - consistent
  - presented in an easy to use fashion
Who cares?

- “The registrars sort it out.”
- “We don't ever send postal mail.”
- “It's all electronic.”
- “It doesn't matter anyway.”
Costs

- undeliverable or misdirected (postal) mail.
- untraceable registrants.
- unpaid renewals (for direct registrations).
- legal hassles
  - Bad actors → bad data
- staff time wasted = $$$
Costs
Improvement plan
LET ME PROVE IN 7 DAYS that I can make YOU a NEW MAN!

By CHARLES ATLAS

"The World's Most Perfectly Developed MAN"

ALL I want is one week to prove that I can give you a powerful body of muscle and muscle like mine—with my quick dynamic

FREE BOOK

You know that most books on physical training are full of trash and mumbo jumbo. I can fit you into two weeks of shape and tone you up in no time. And you can have the results in just one week! The savings in time and money is amazing. I'll prove it with my FREE BOOK.

Accept my FREE BOOK

Why be just 6' of the time you can be 6' 5000! If you accept my FREE BOOK, I will show you how to develop your body to its full potential in just 14 days. It's easy! Just follow the instructions and you'll see results in no time. And you don't need any special equipment or tools. Just follow my easy-to-follow plan and you'll be on your way to a stronger, healthier body in just one week.
Data migration

Three disparate databases

- Technical
- Billing
- Admin/Ticketing
Technical:

- No issues: domain names & NS
Billing

• Addresses in postal format
  • One field for city, state, postcode
    - Alderney, (C.I.), GY9 3JZ
    - 23563 Lübeck
    - Miami Beach FL 33139-4612
    - Iqaluit, NU X0A 0H0
Admin and Ticketing

- Admin Contact details
  - similar format issues
  - even more accuracy issues
    - not self correcting
    - a postal letter once a year keeps address data warm!
Consolidation

- A number of magic processes fixed the legacy data
Now ...

- nicely separated
- right data in right place
- but formats still differs from record to record
  - Due to the different ideas of original inputter
- registrars/registrants
  - “rebugging”
- can it be kept improved and improved further?
BUT!!!

- EPP!

Neither

- registries; nor
- registrar
have much control over registrants' delusions
ISO

United Kingdom
  • GB: \{GG, JE, IM, GB\}

Finland
  • FI: \{FI, AX\}

France
  • FR: \{FR, MC\}

Italy
  • IT: \{IT, VA, SM\}
Postcodes

• UK & CI
  GY1 3JZ
  GY 13 JZ
  GY13JZ
  GY1
  ENG
  Fabio Cappello

• Sweden
  234 50
  23450
  SE 234 50
  SE-234 50
Informal variants

US states
- Cal
- Cal.
- California
- Calif
- Calif.
- CA
- CA.
State

Island-as-state
- Guernsey
- Bailiwick of Guernsey
- Channel Islands

(this is otiose – in the ISO-code field)
State

German/Dutch/French states and provinces

- must never part of official address at all
- causes routing confusion if used
• Island Parishes
  - St Pierre-du-Bois, St Saviour
    (variants St Peter, St Peter's, St Saviour, St-Saveur, St Saviour's)
  - important locally.
  - found in Street, Town and occasionally in State field
Town/City

- UK Villages
  - e.g. “Lawford, Manningtree”
  - Village or sub-division is otiose
  - but often perceived by many people in UK as necessary or desirable.
British Islands

“Standard Address”

- PAF
  - 'Post town' not village
  - County is otiose and deprecated
  - Every address in GB, GG, JE and IM in a database
    Min unique expression
    e.g “GY9 3JZ-13”
Street Address

- 3 lines of Street address
- Separated well in data migration
- But what of the future?
Another key motivator everyone has their own standard!!

(... and they all thing thinks they are right!)

“Two men say they are Jesus, one of them must be wrong”, M. Knopfler
If you have consistent data you can do more magic

- Map registration hotspots
- Break down registrations by country, state, town, telephone area or other keys
- Normalise!
  - Minor typos in different versions of the same contact prevent normalisation.
The amazing thing

- Data still contains
  - Inherent semantic meaning
  - which can improve life for users
    - by reformatting data
    - and enhancing its presentation
- Just need to bring it out
Core Concept # 2

“Trigger”
Trigger
**Trigger**

- “A trigger is a piece of code which is executed whenever a certain database operation takes place

- **ON UPDATE**
- **ON INSERT**
- **ON DELETE**
What can it do

- make presentation of data human friendly
- restore apparently lost semantic content
  - algorithmically; or
  - by data lookup
Addresses

Correcting “wrong” ISO codes

- where country or territory in neighbours postcode system
  - VA, AX, MC, SM, GG , JE, IM

removing otiose data

- states where the country has none (in the official postal addresses)
- 'postcodes' when the country has none
Addresses

- removing duplication
  - e.g. identical name as street address AND city
- automatically creating “IDN” format local addresses.
Addresses (2)

- Standardising identifiers
  - for states/provinces
    - Canada
    - Mexico
    - Australia
    - USA

- Ensuring standard city/town names from postcode.
Addresseses (3)

- Misc format corrections
  - CEDEX
    - Capitalisation from 'Cedex', 'cedex' &c.
  - MLO
    - similarly
  - P O Box
    - 'PO Box',
    - 'po box',
    - 'Po Box' etc etc.
Telephone numbers

E.123

National formats
  • (01481) 822800
  • (06) 6982

International format
  • +44 1481 822800
  • +39 06 6982
RFC 4933, section 2.5 claims to be based on E.164

- But E.164 is the numbering plan, not the recommendation for the notating standard!!

- +44.1481822800
- +39.066982

removal of spaces is removal of semantic content – not mandatory per RFC.
Telephone numbers
(contd 2)

As-you-dial-it

(0044) 1481 822800
(01144) 148 182-2800

Local (national) standards

02/234-5678
(Belgium)

2234 5678
(Denmark)

- closed dialing plan
Common “informal standards”

“Printers international”
+44 (0) 1481 822800
0044.1481.82.28.00
Nigel Roberts

Director
Channel Islands Registry (.GG and .JE)

Consulting Engineer
TLD Registry Solutions LLC
Technical DB consultancy

nigel@roberts.gg
http://www.roberts.gg