Domestic Load Research Seminar
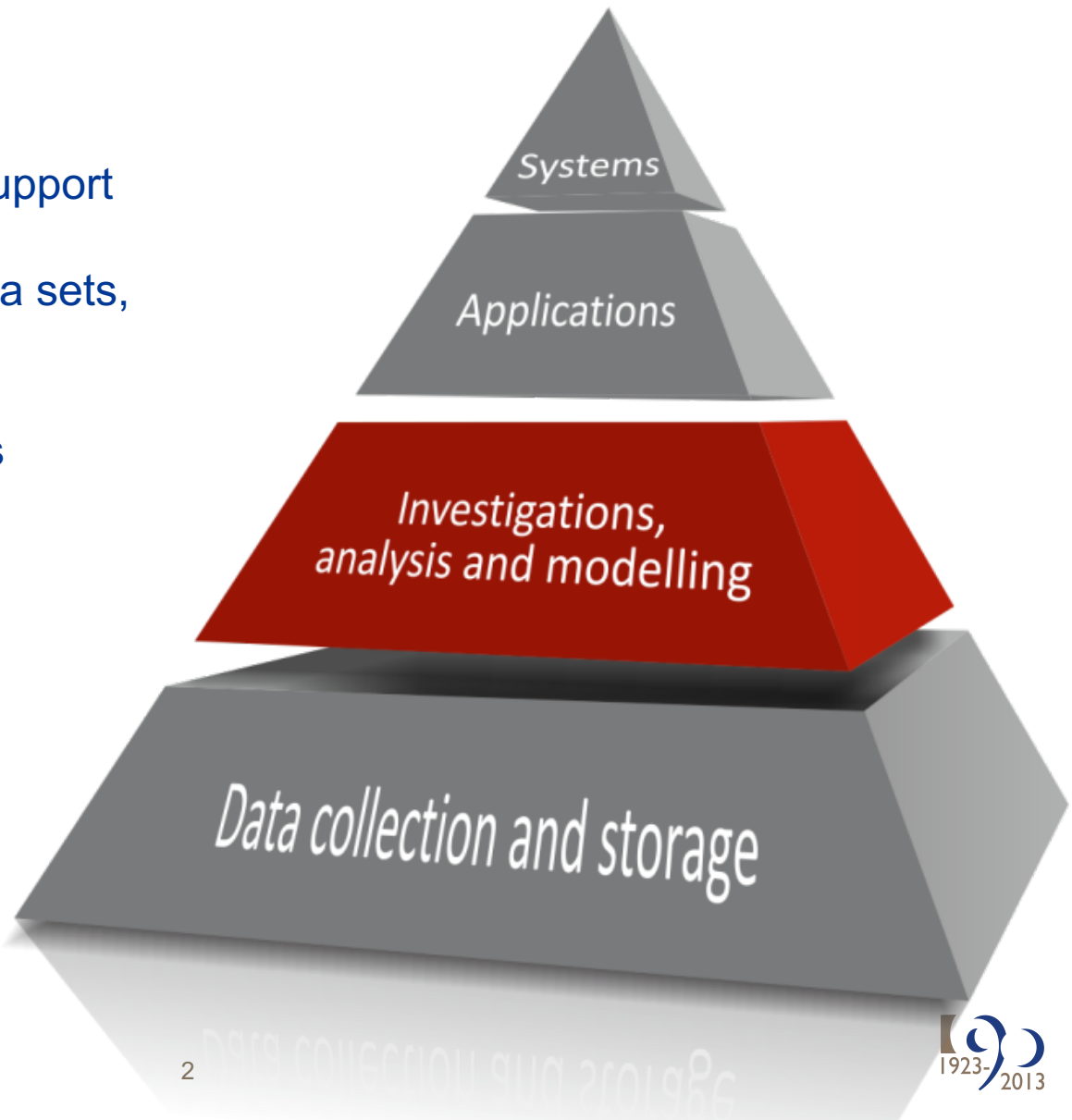
The Source Data Store

Presenter Marcus Dekenah

12 November 2014

Powering your world

1923-2013

Objective

- Making data for decision support

- We want to build "tight" data sets, with long shelf life.

- Consistency of approach is important (Standards).

- Manage the "keys" actively

- Validate each input.

Systems

Applications

Investigations, analysis and modelling
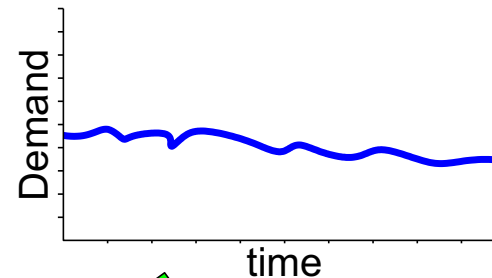
Data collection and storage

Contents

- Description of the source data assembled during course of domestic LR project

- Load data

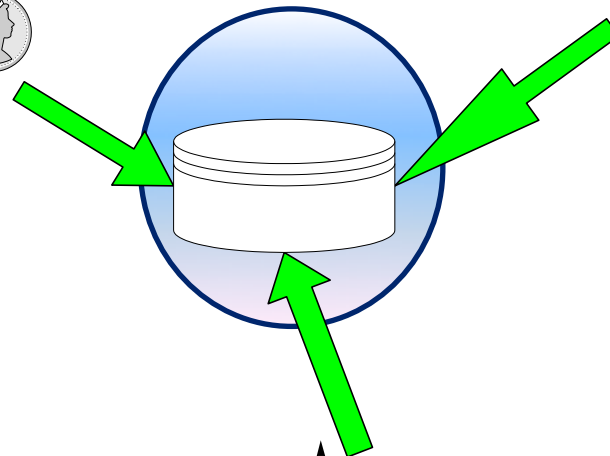- Sociodemographic data

- Link data

- External data sources

Household demographic information
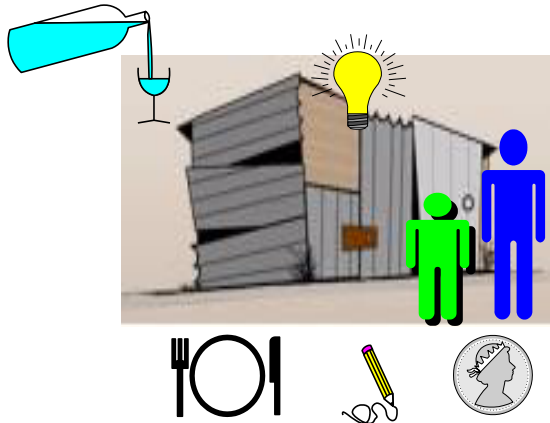
Household electricity consumption

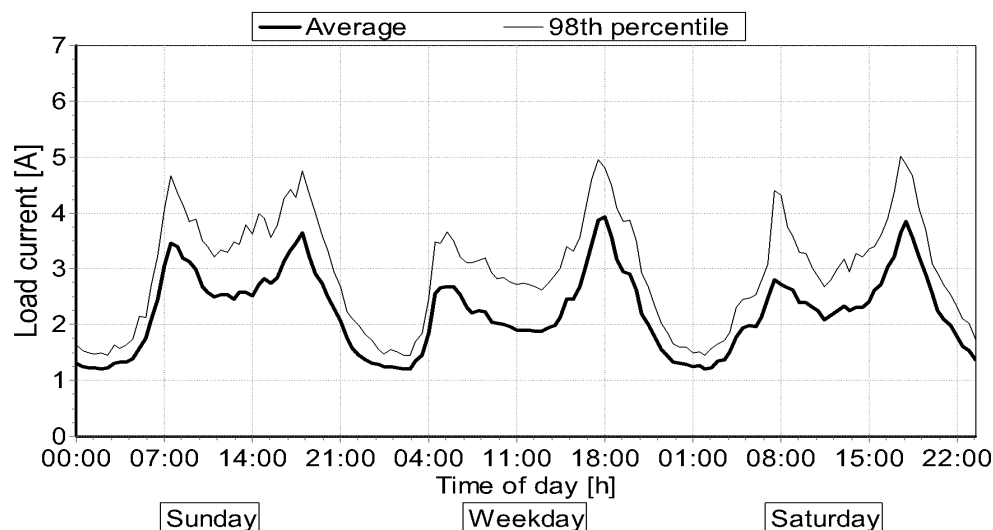Utility sales & billing information

- Five minute averaged load data at the house connection level.

- 1995-2008 : multi-channel loggers used current + voltage (1 phase only).

- 2009-2014: 3-channel energy meter kit. V, I, KW, KVA, Hz.

There are **more than 1.3 Bn load readings** on database now.

Includes data from domestic monitoring in Namibia, and limited shop/school/clinic data in South Africa.

Validation of load data

- All load data is marked with validation result, as blocks per day.\

- Validation is used to exclude effect of technical metering failure.

- Validation rules cover missing data, inactive data, bounds-checking, and model-based checks.

- Since 2008, additional status flags from MV90 (interval and measurand) have been used to add a further layer of validation.

- In 2013 data-marking rules were augmented (rules to handle voltage and the additional measurands)

- Readings are stored as a row per measurand per time-stamp.

Sociodemographic data

- There are 6 distinct survey forms on the database. (2 SA*household, shop/school/clinic, and Namibian household).

- All Socio data is captured on database via electronic forms, which are themselves a form of data.

- The captured responses are encoded and distributed by data type, stored as either Text, Numbers or Binary Large Object (BLOB), (a strategy to deal with technology at that stage) and indexed by the AnswerID. The storage density is sparse.

Socio demographic data

- A socio demographic survey form is collected per household per year.

- Typically 80% of sample is surveyed per year.

- Socio data collected:

- Inhabitants (age, income, sex, education)

- Dwelling construction and size

- Appliance ownership

- Cooking habits

- Perceptions about power supply

Last year 800 socio survey forms were collected.

Sociodemographic data validation

- Most Socio data capture forms have an associated set of validation rules covering:

- Domain tests (control nulls, zero's and excessive values)

- Redundancy tests (force consistency where sub-totals can be made)

- Model-based tests (seldom used in reality)

Examples : Income data must be within a reasonable upper limit, and non-null unless the householder declines. Need to control number of ways to say "nothing".

All hand-captured data is validated. All validation results are written to auxilliary tables, and each flag is verified on capture to ensure.

Sociodemographic source archive

- All socio demographics captured are labelled with their database index.

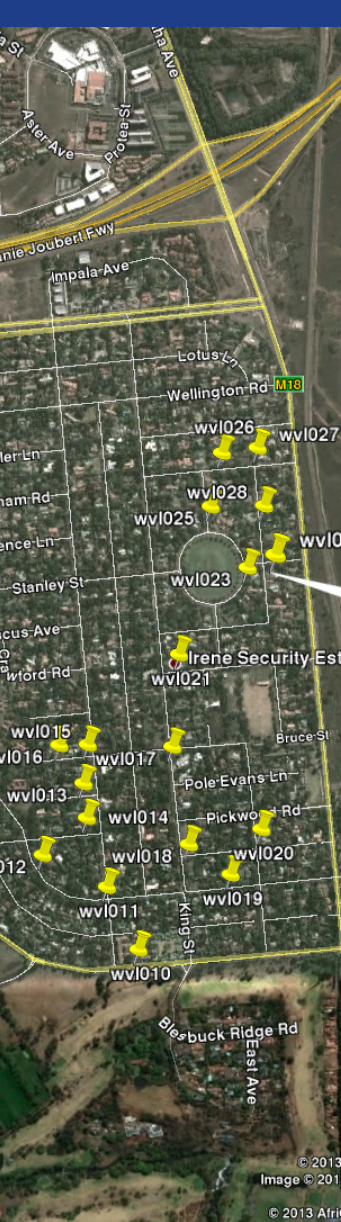- Paper forms are then scanned to PDF and the paper version destroyed.

Link data (connectivity between logger and Socio)

- Link data associates Load data profiles with Socio responses

- Used to locate the customer sample with low uncertainty

- Each year since Y2000 the customers connected to each logger are traced and connectivity is documented.

- Field operations can "break" the connectivity over time, so the reference is reviewed annually.

- During the documentation process, customer Name, address , GPS, Meter Number, logger position and logger channel are all captured in an auxiliary database.

- Galvanic testing is preferred method to resolve connectivity unambiguously.

Link data collection <u>now paperless.</u> Includes photographs of dwellings.

Results of "paperless" link-data collection

Export format can be "mobile-friendly" (ie KMZ on google earth)

wvl029

start	Fri Oct 25 12:57:57 UTC 2013
end	Fri Oct 25 13:05:26 UTC 2013
deviceId	353988050224061
village_name	wattville
recorder:logger_id	wvl029
recorder:pole_no	Jugliugli6glig6
pole_location:Latitude	-25.8683683689
pole_location:Longitude	28.2215624192
pole_location:Altitude	1504.6057088425
pole_location:Accuracy	8.0000000000
recorder:channel_count	3
household_a:channel_no	channel1
household_a:photo	View
household_a:owner_name	Ttyyuujui jjjkoiypp
household_a:meter_no	44378966436
house_location:Latitude	-25.8685168897
house_location:Longitude	28.2213936639
house_location:Altitude	1508.1358294805
house_location:Accuracy	4.0000000000
household_a:eskom_no	Kuhiuhoih
household_a:other_no_a	
household_a:other_no_b	
household_a:channel_test	load_at_house
household_a:comments	
household_b:channel_no	2
household_b:photo	View
household_b:owner_name	ljyvjyviyuh hiihih
household_b:meter_no	55777534880
house_location:Latitude	-25.8690309574
house_location:Longitude	28.2214028881
house_location:Altitude	1522.9112210407
house_location:Accuracy	8.0000000000
household_b:eskom_no	Vyghyvjyvj893343
household_b:other_no_a	5565
household_b:other_no_b	
household_b:channel_test	load_at_house
household_b:comments	

WVL029-channel2

Channel No	channel2
Owner Name	B James
House location	-25.8690309574 28.2214028881 1522.9112210407 8.0000000000
Meter no	55777534880
Eskom no	E893343
Channel test	load_at_house

Eskom

Link data validation

- Link data represent a snapshot of customer data, connectivity and metering year-on-year.

- Practical issues: full data set are not always possible (not at homes).

- Validation tests are typically applied within each year and between years.

- Types of tests:

    - Check meter vs control code

    - Spatial testing of GPS coordinates (relative & absolute)

    - Time-based consistency

External data

- Hourly climatic data from from SA weather service.

- Limited prepaid sales data (much more required)

- SA population data (SA Census's)

- AMPS survey results & LSM data (Eighty20 web-service)

- CPI data (SA reserve bank)

Key Success Factors

- Persistence

- A "Grand design" to approach the problem.

- Rapidly analyse existing data to understand what is not known.

- Long term view

- Relatively generic database design

- Very good team support

THANK YOU!!!