MANCHESTER
CITY COUNCIL

Local Datavores Workshop LGA – July 13

Sarah Henry – GM Connect and Head of Intelligence and Performance – Manchester City Council

s.henry@manchester.gov.uk

A families database – from an individual to a holistic view

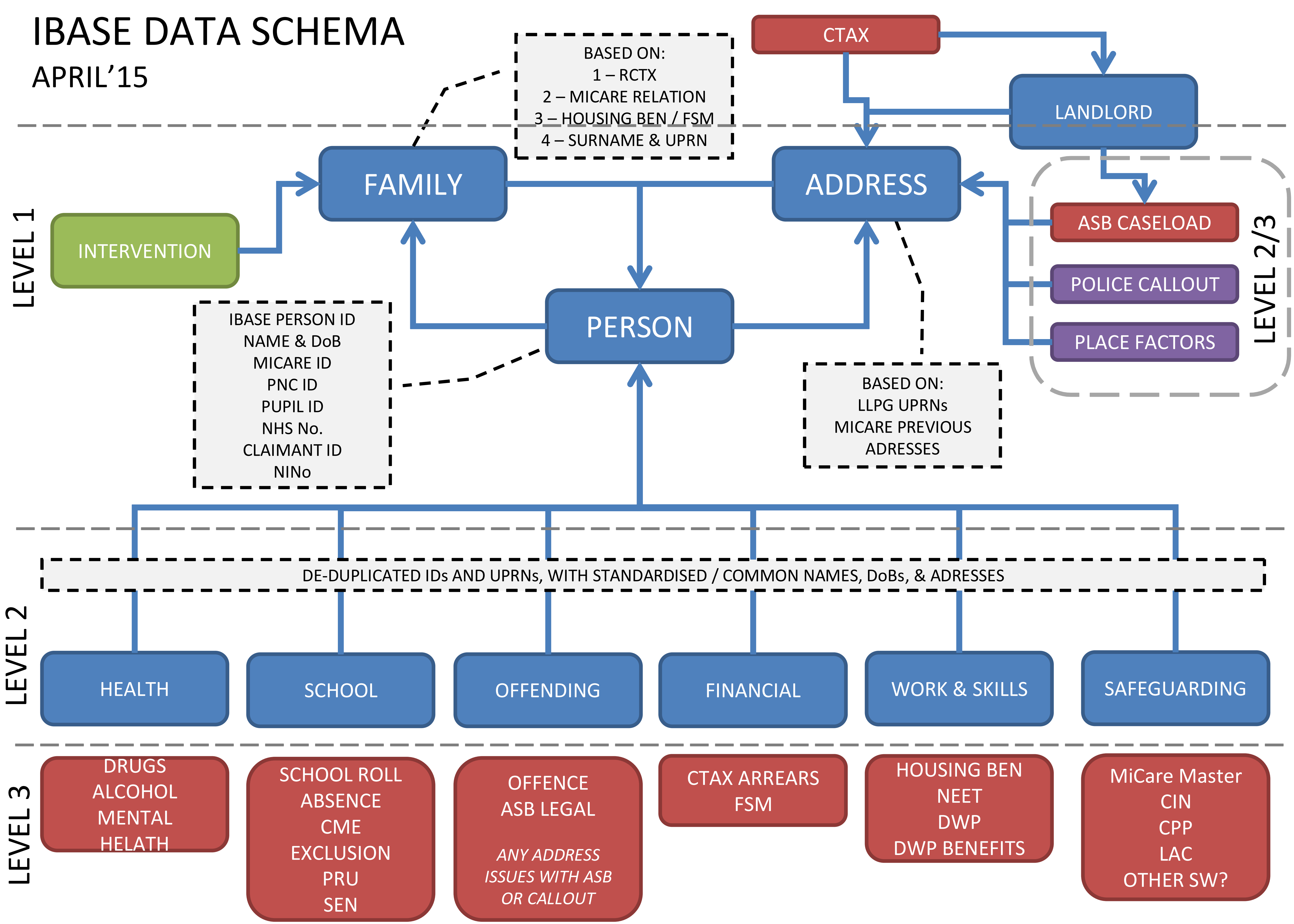
People, and especially children, are affected – positively, and sometimes negatively - by their family circumstances and the people around them

But what we collectively know – the data - sits in systems about individuals so we need:

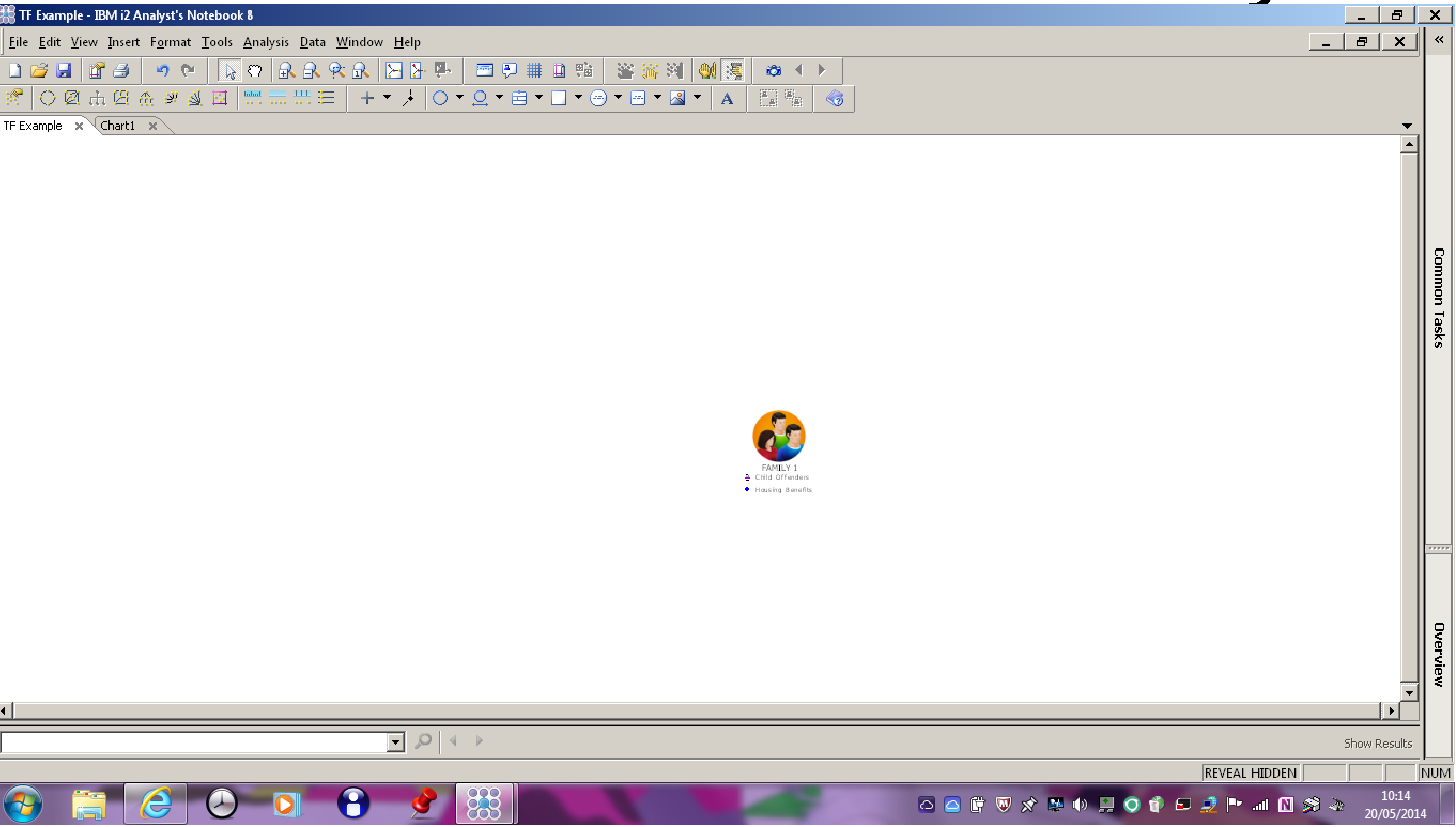
- To link individuals to ‘families’ and ‘households’
- Intelligent linkages between individuals e.g. family relationship, mention on benefit claims
- Linkages to organisations such as schools
- Linkages to and tracking of interventions, services and outcomes, recognising the engagement with the family is not static

Our system -

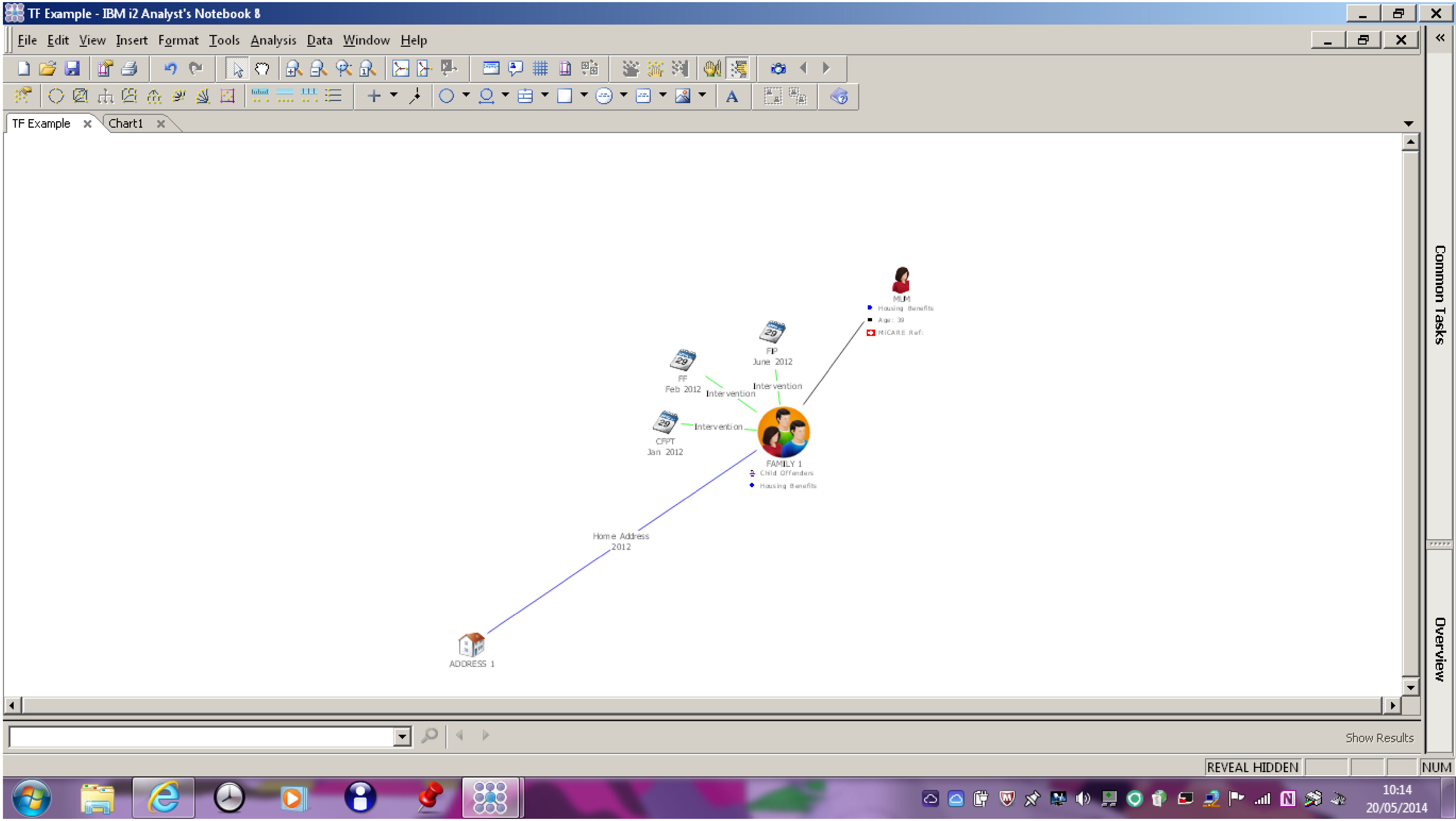
- Replaces and complements manual processes to find out what others know about the family – MASH, Panels and conferences, desk-top work etc
- Visualises the family and the issues
- Creates data that can be analysed over time
- Brings together 18 datasets including offending, ASB, absence from school, low attendance at school, various benefits (including out of work), debt, social care, children missing education, NEET, poverty indicators, substance misuse, mental ill health etc.


IBASE DATA SCHEMA
APRIL'15
BASED ON:
1 – RCTX
2 – MICARE RELATION
3 – HOUSING BEN / FSM
4 – SURNAME & UPRN
CTAX
LANDLORD
LEVEL 1
INTERVENTION
FAMILY
ADDRESS
ASB CASELOAD
POLICE CALLOUT
PLACE FACTORS
LEVEL 2/3
PERSON
IBASE PERSON ID
NAME & DoB
MICARE ID
PNC ID
PUPIL ID
NHS No.
CLAIMANT ID
NINo
BASED ON:
LLPG UPRNs
MICARE PREVIOUS
ADRESSES
DE-DUPLICATED IDs AND UPRNs, WITH STANDARDISED / COMMON NAMES, DoBs, & ADRESSES
LEVEL 2
HEALTH
SCHOOL
OFFENDING
FINANCIAL
WORK & SKILLS
SAFEGUARDING
LEVEL 3
DRUGS
ALCOHOL
MENTAL
HELATH
SCHOOL ROLL
ABSENCE
CME
EXCLUSION
PRU
SEN
OFFENCE
ASB LEGAL
ANY ADDRESS
ISSUES WITH ASB
OR CALLOUT
CTAX ARREARS
FSM
HOUSING BEN
NEET
DWP
DWP BENEFITS
MiCare Master
CIN
CPP
LAC
OTHER SW?

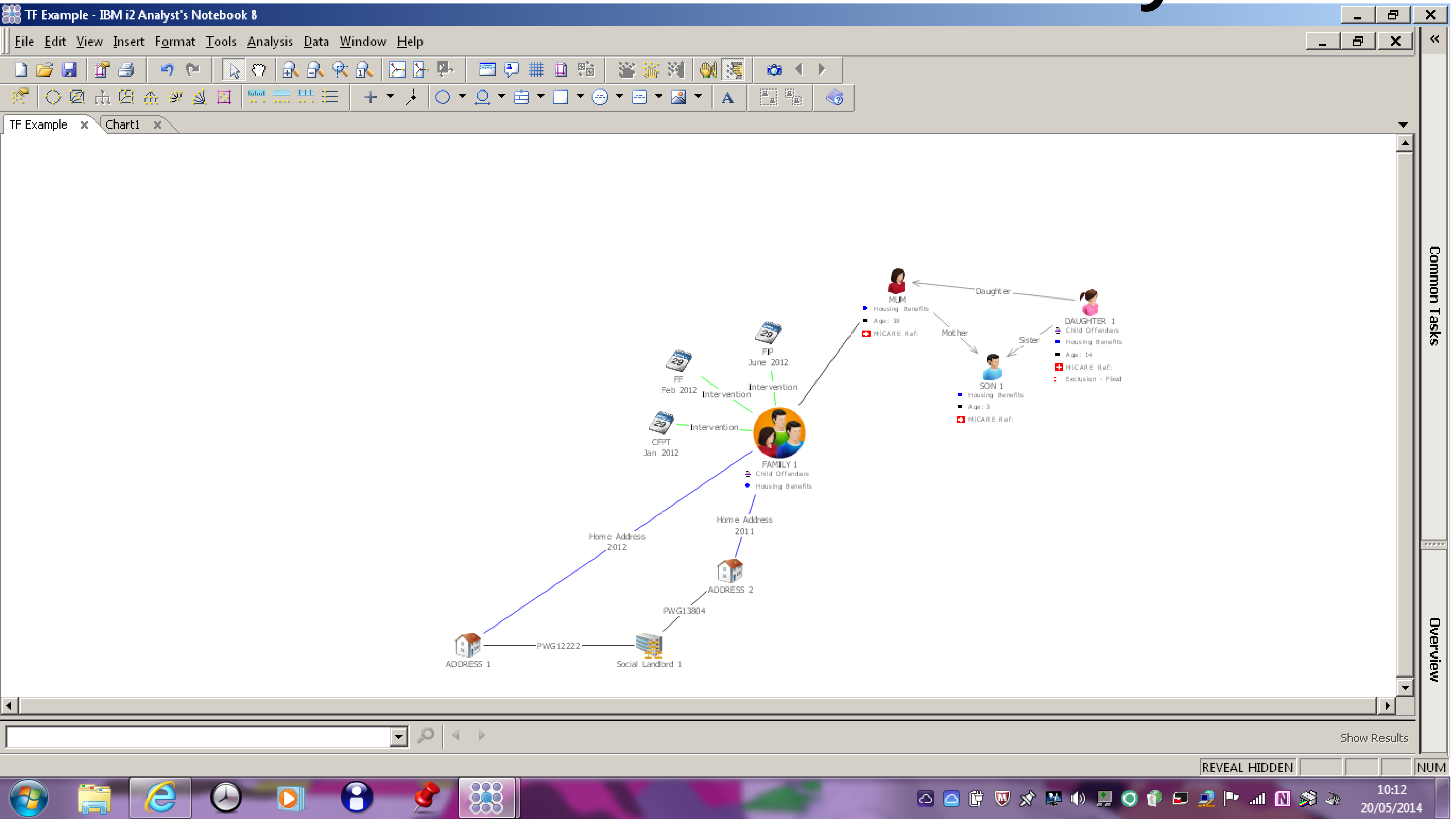
1. Find the Case / Family



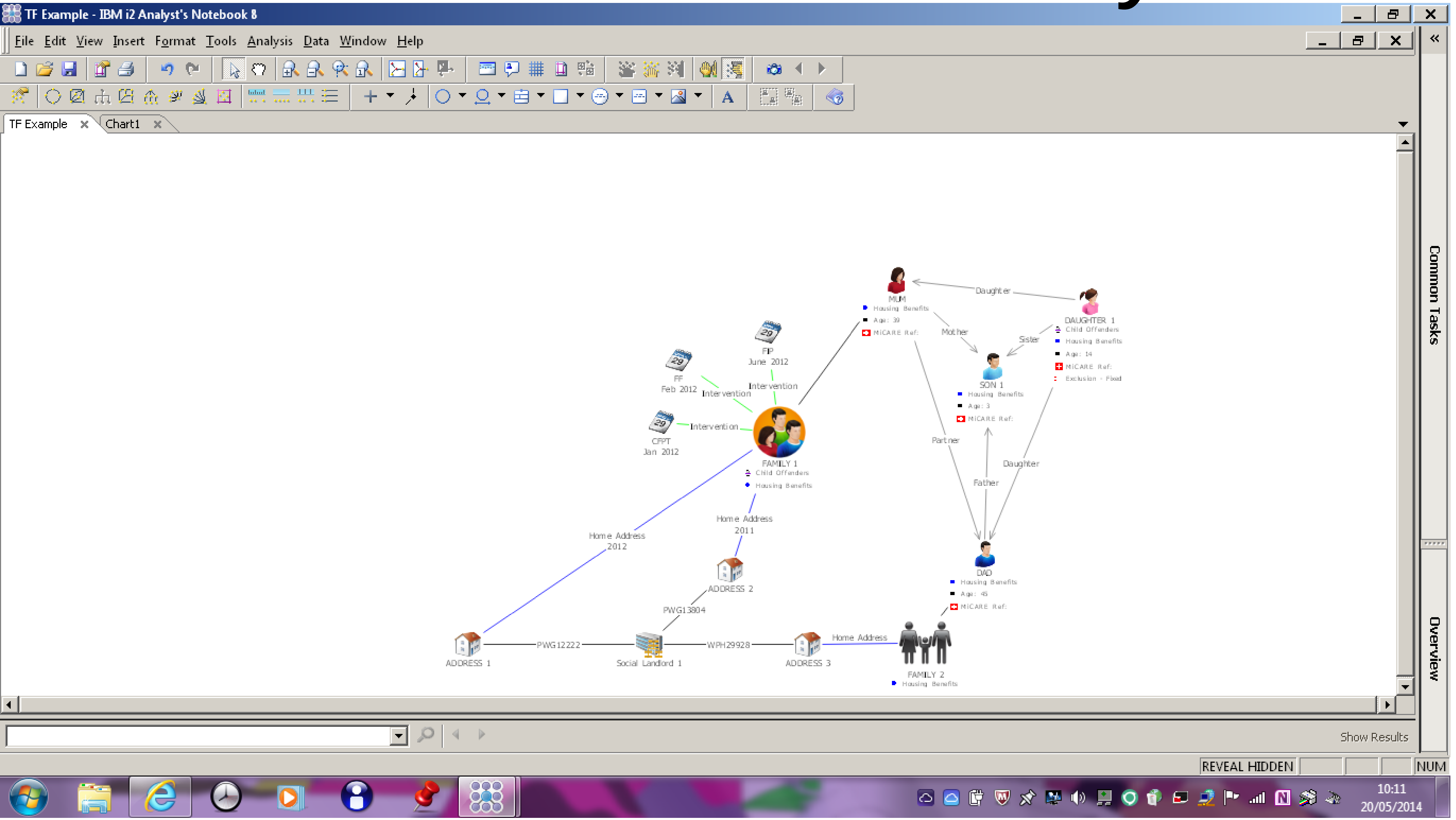
2. Direct Links to Case



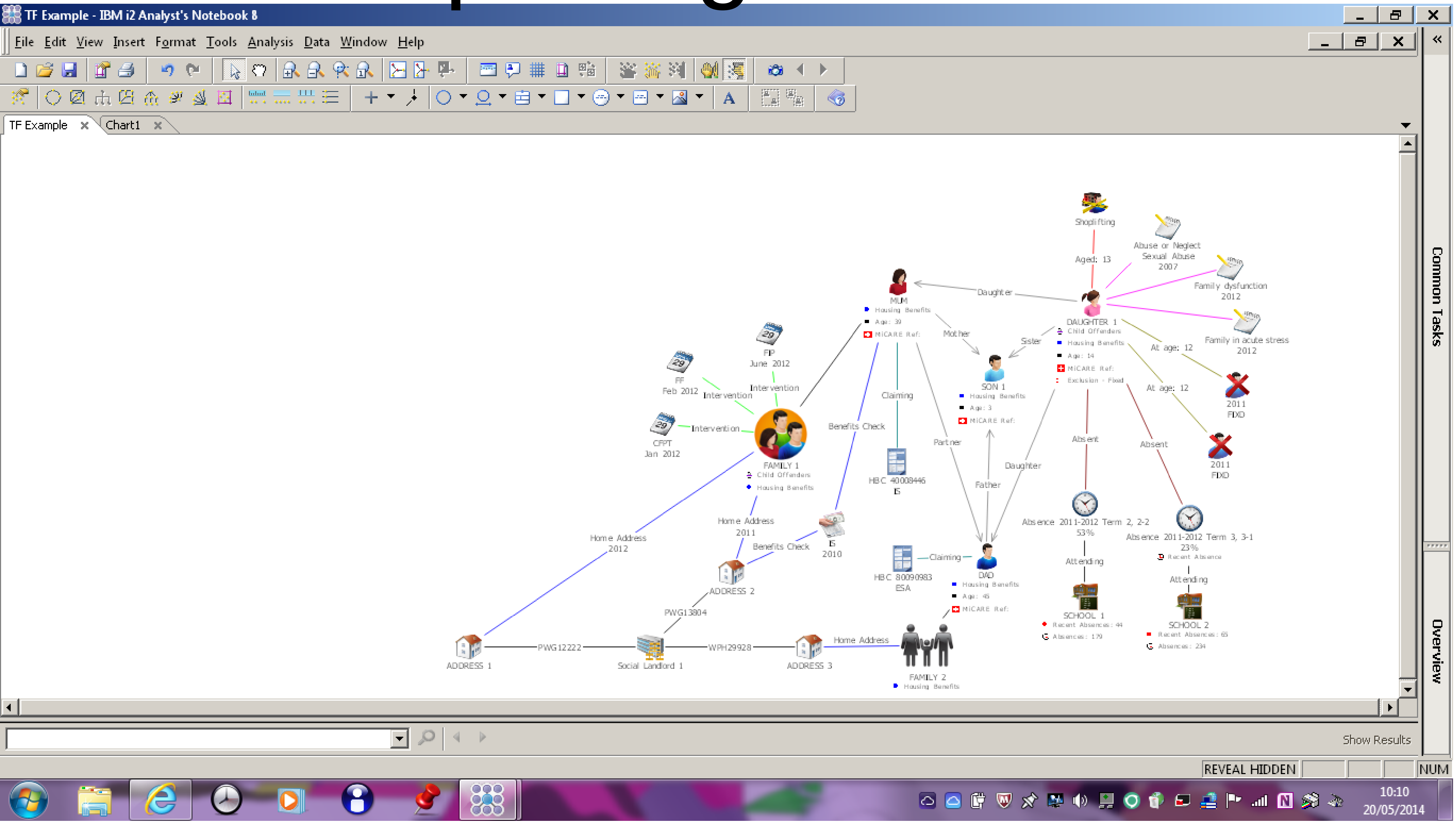
3. Immediate Family



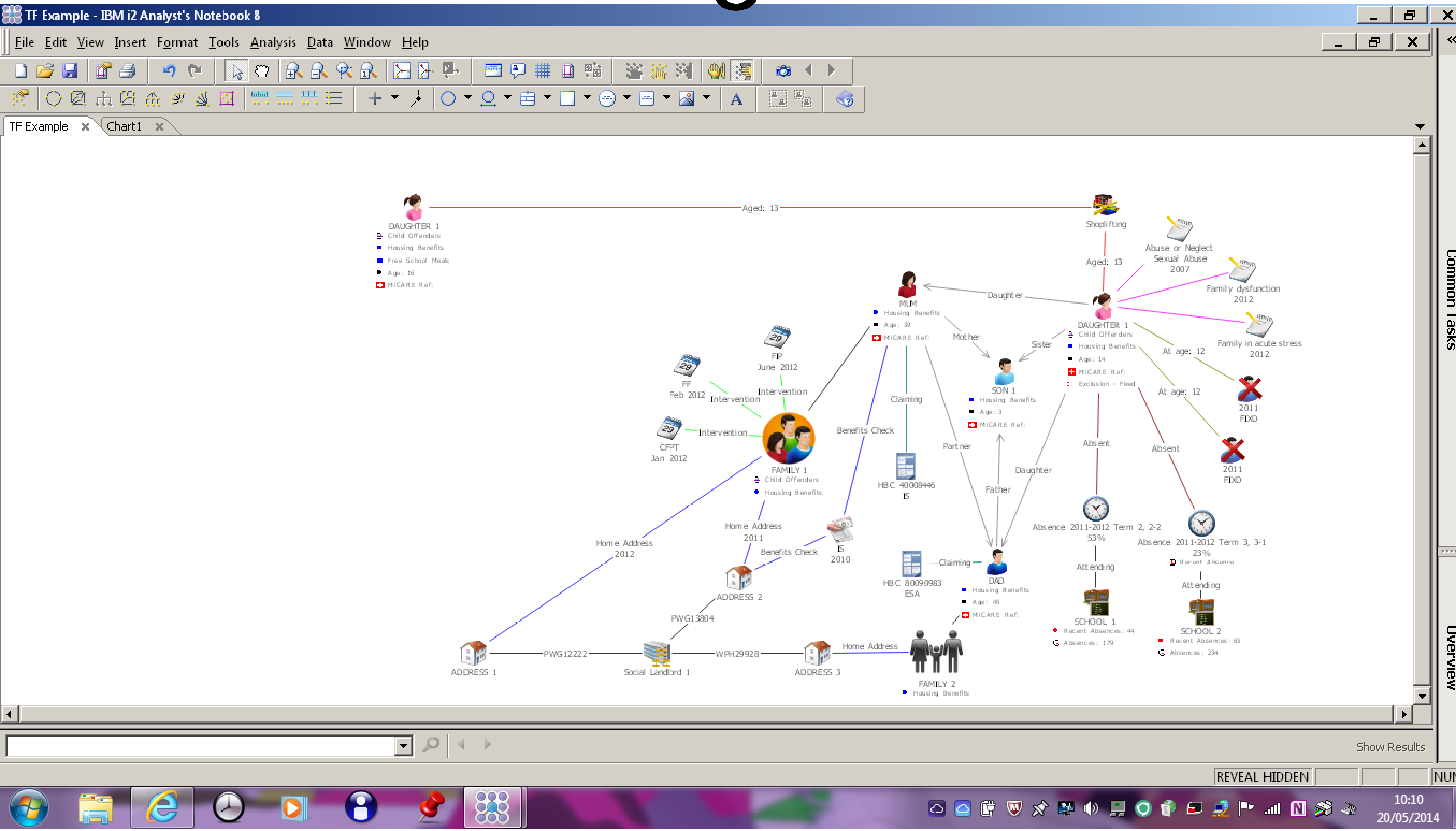
4. Extended Family



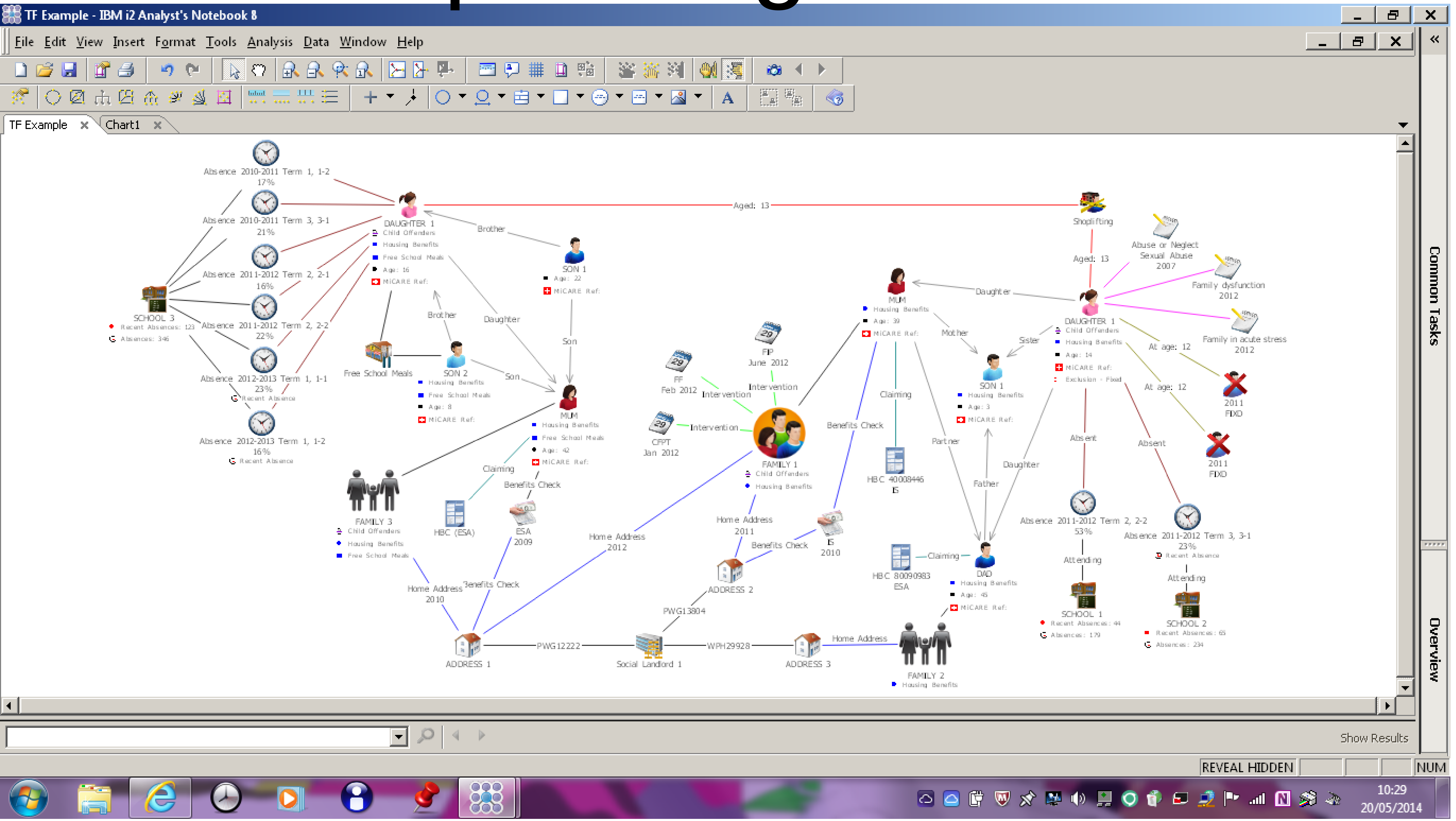
5. Exploring the Issues



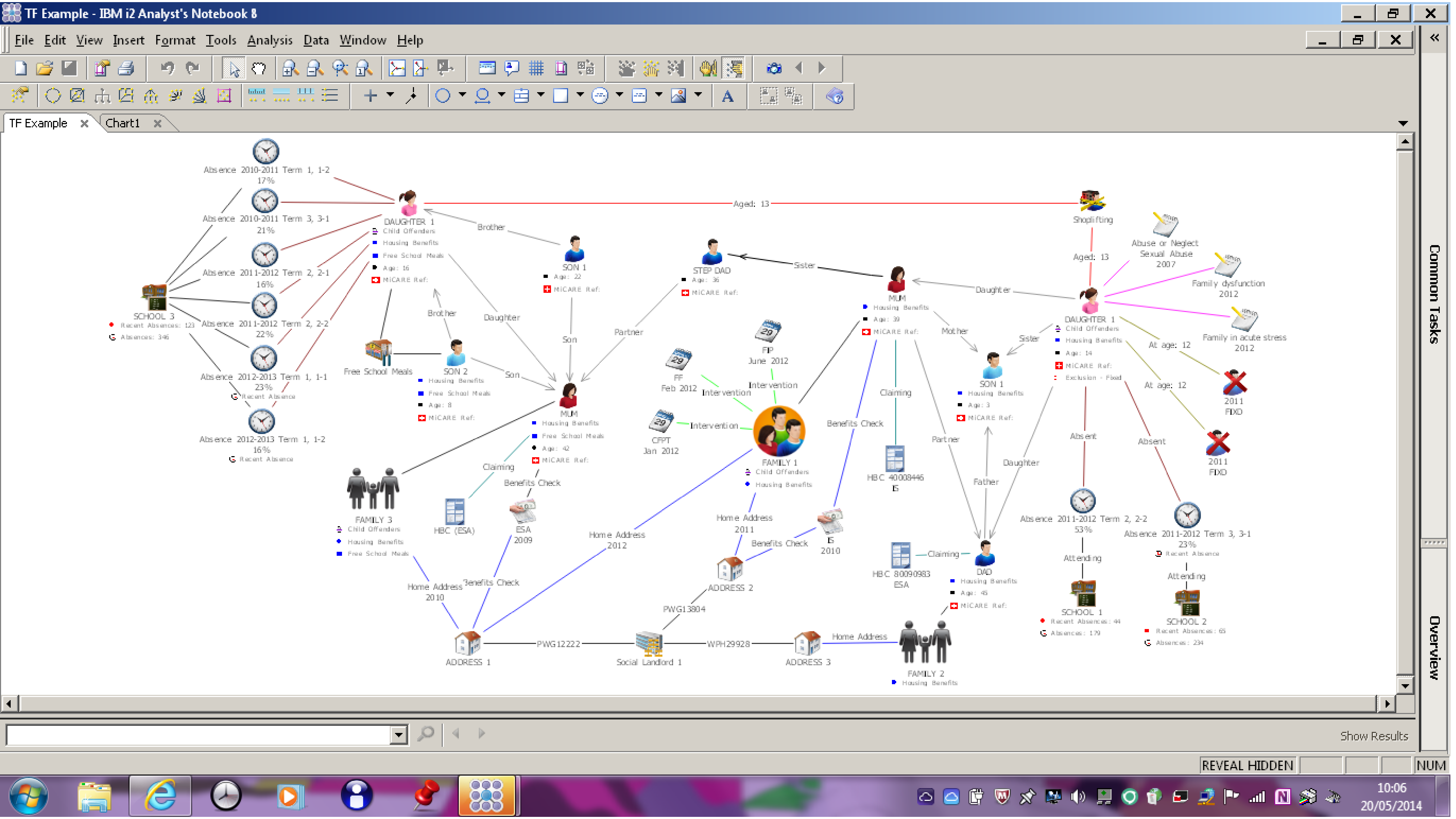
6. Looking for Links



7. Expanding the Links



8. Finding the Connection



Challenges

- Information Governance
- Technical – selection of a system
- Information Governance
- Developing and agreeing rules and processes – data architecture
- Information Governance

Data Protection and Security

- Underpinned by data sharing agreements and a Privacy Impact Assessment
- Consent, where needed and appropriate
- Removal of the decision to share or not to share away from individual workers – consistency and corporate accountability
- Who should access? – Restrictions, training and access control
- Policies and procedures – including security, retention, destruction
- Footprints

Legal gateways

- Human Rights Act – right to life is absolute
- Right to privacy – conditional
- Duty to safeguard vulnerable people
- Local Government Act – general power of competence
- Duty to prevent crime (and fraud)
- Regulations designed specifically for programmes like Troubled Families
- And more…

Critical Success factors

- Reasonable budget (but it didn’t break the bank!)
- Senior strategic sponsor
- Clarity of purpose – but agile enough to build in new things
- It wasn’t an IT project!

Research and insight

- Understanding the impact on the system (so it can be planned)
- Understanding the causal factors behind the outcomes (so we provide the right service)
- Predictive modelling
- Ultimately leading to decision making tools

Data Mining & Testing

There are numerous data mining techniques that could be used in an analysis of this large dataset

Five techniques have been used in this project:

- Decision Tree
- Cluster Analysis
- Regression Analysis
- Sequence Analysis
- Spatial Analysis

The analysis has been based on 326,486 records, representing 21,584 individuals over a 5-year period

Data included:

- Child in Need episodes
- Domestic Abuse incidents
- School Absence issues
- School Exclusions
- Anti-Social Behaviour incidents & legal actions
- Offences

Decision Trees

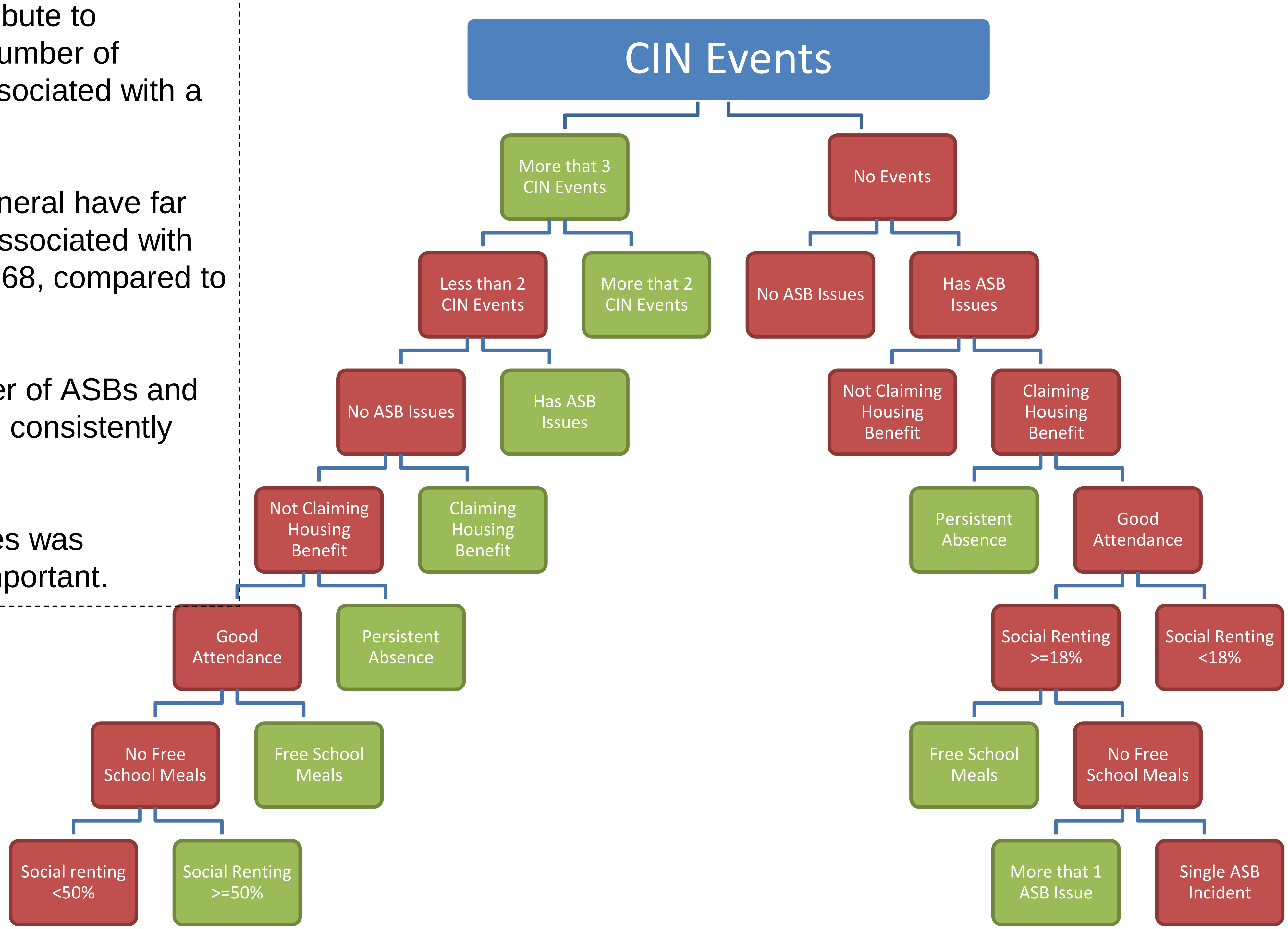
Reviewing recent history and using the common decisions to predict a likely future decision

The most important attribute to predicting TF was the number of MiCARE CIN events associated with a family.

Troubled Families in general have far more MiCARE events associated with them – an average of 5.68, compared to 0.31 for the non-TF

School absence, number of ASBs and PRU also were deemed consistently important.

Number of adult offences was consistently the least important.

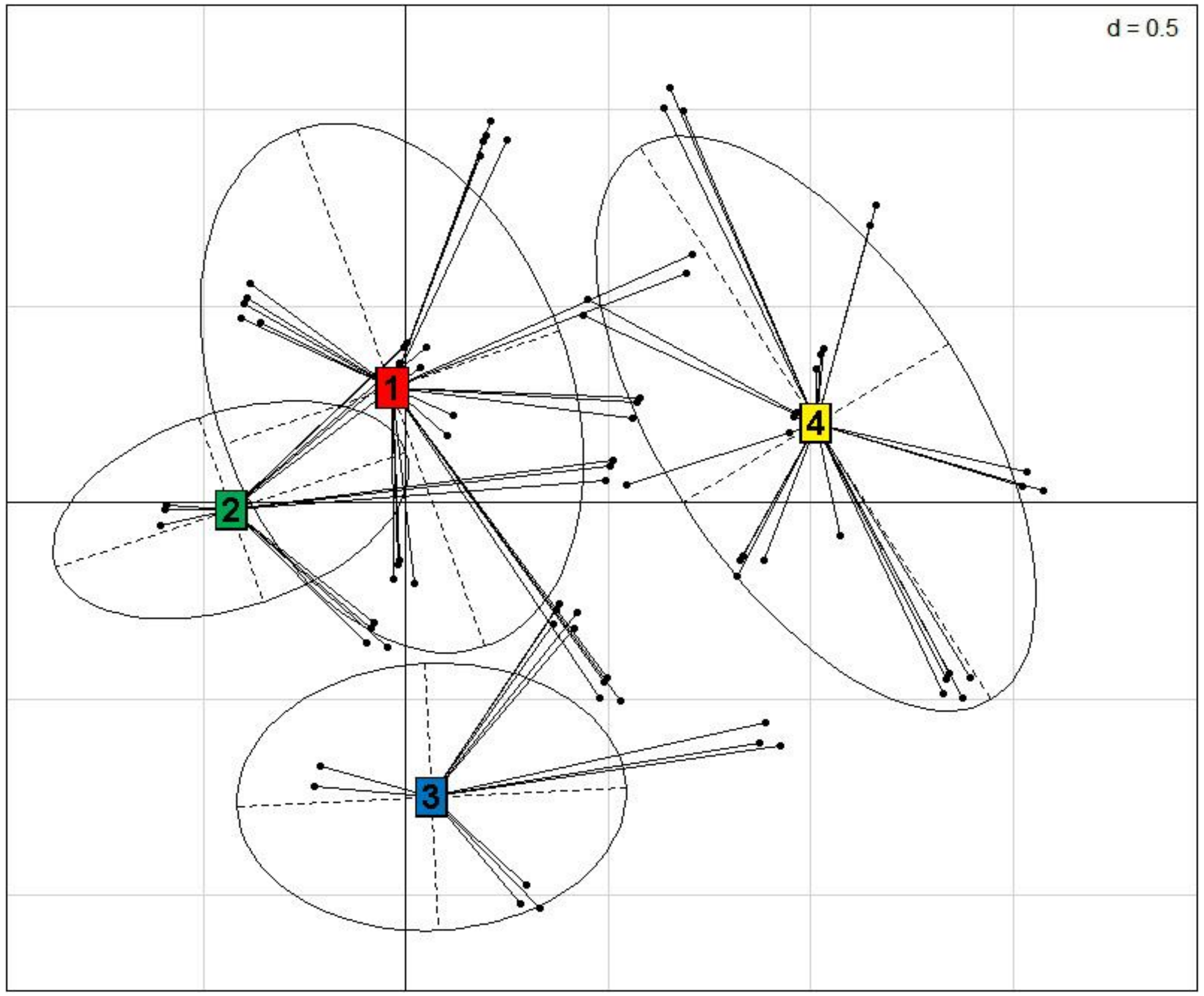


Cluster Analysis

Reviewing the current distribution to identify common groups / characteristics

Cluster Analysis was performed on 2,028 families, where there had been a referral made to the FRS and where there was at least 2-years worth of event data

Four clusters were identified. No completely distinct groups but there were still patterns in the data



Cluster 1 – 689 families - No MiCARE CIN events but a quarter of all DA events in this cluster. Just under a third of families had absence, a quarter had exclusions, and 22% had offences but none involving minors.

Cluster 2 – 724 families - All had MiCARE CIN events. No school absence, but 13% had exclusions. 18% had offences, none involving minors.

Cluster 3 – 308 families - All had MiCARE CIN events. All had absence, 37% have exclusions. 17% had offences but none involving minors.

Cluster 4 – 307 families - Half had MiCARE CIN events. Half had absence, **62% have exclusions. All families had offences, 73% of these offences involving minors**.

Regression Analysis

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	direction
Step 1[a]	Previous CIN referral	1.418	.071	401.379	1	.000	4.127	+
	Age	-.054	.009	33.884	1	.000	.948	-
	Absence periods (0,1,2,3)	.227	.076	8.959	1	.003	1.255	+
	Previous absence	-.096	.033	8.245	1	.004	.909	-
	Exclusions periods (0,1,2,3)	.084	.093	.814	1	.367	1.087	
	Previous exclusions	.012	.035	.110	1	.740	1.012	
	Person Offending (0,1,2,3)	-.133	.185	.516	1	.473	.876	
	Previous Person Offending	.072	.071	1.038	1	.308	1.075	
	Family Offending (0,1,2,3)	.292	.064	20.586	1	.000	1.339	+
	Previous Family Offending	-.163	.030	30.670	1	.000	.849	-
	Constant	-3.454	.115	903.863	1	.000	.032	

a. Variable(s) entered on step 1: Pre_CIN_R, Age, x_periods, total_abs, Exc, Exc_Pre, P_Off, P_Off_Pre, F_Off, F_Off_Per.

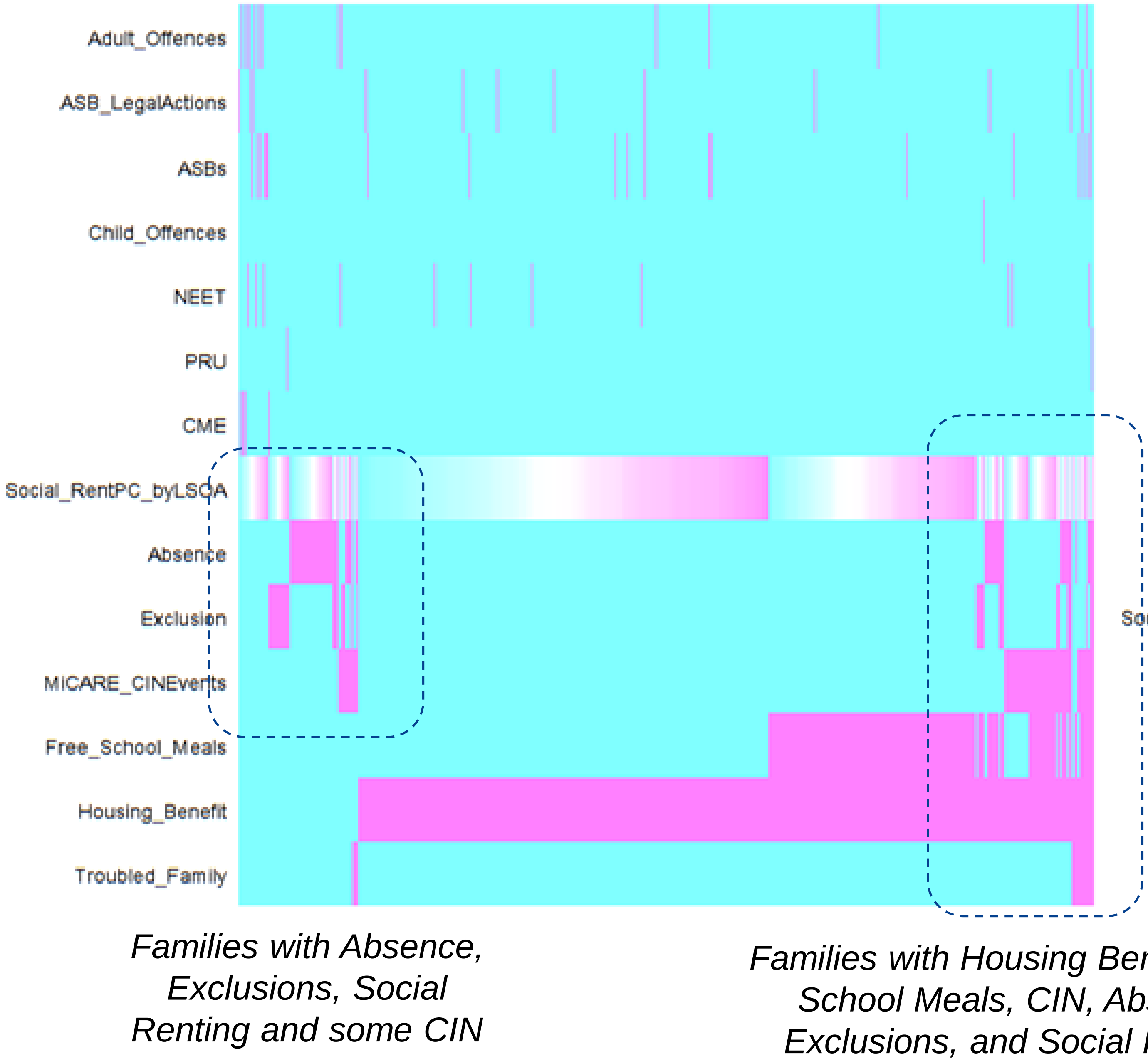
Provides an insight into a child's short-term life history

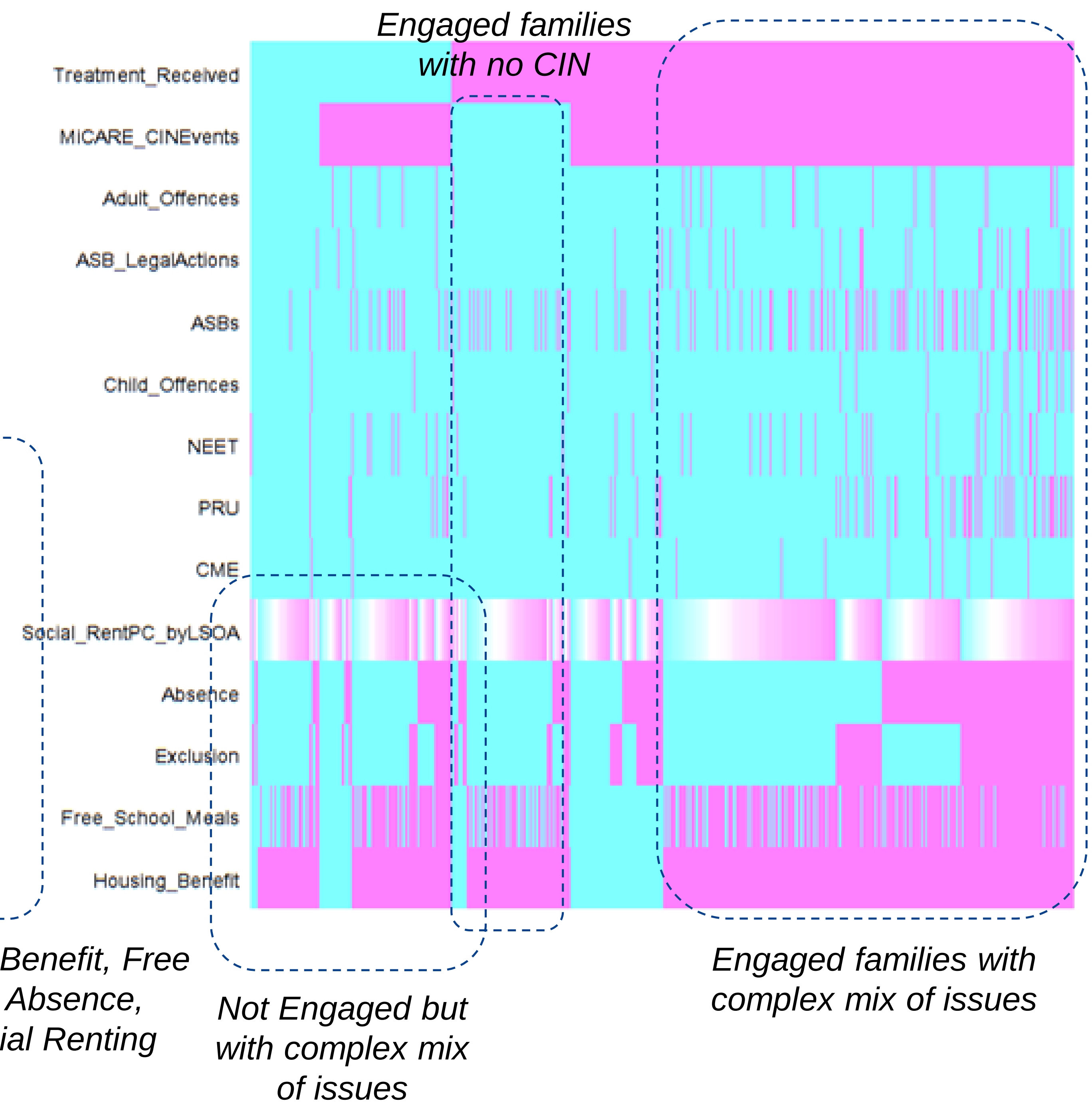
- Having a **previous Child in Need (CIN) referral** is a strong predictor of a new Child in Need referral
- An **increase in the *frequency* of Unauthorised Absence** (in previous 3 half-terms) is a predictor of CIN
- Exclusions and Personal offending are *not* significant
- An **increase in the frequency of Family offending** (in previous 3 half-terms) is a predictor of CIN
- The likelihood of a **CIN referral is less when the young person gets older**

Sequence Analysis

Figure 1 – this represents a subset of all family records, a purple colour indicates the presence of an event (e.g. receiving housing benefit, or having school absence events) whereas a turquoise colour indicates the absence of such events. c.80,000 records

Figure 2 represents just the TF data (with the extra attribute of whether they have received any intervention treatment), this shows far more events occur for these families. c.3,000 records



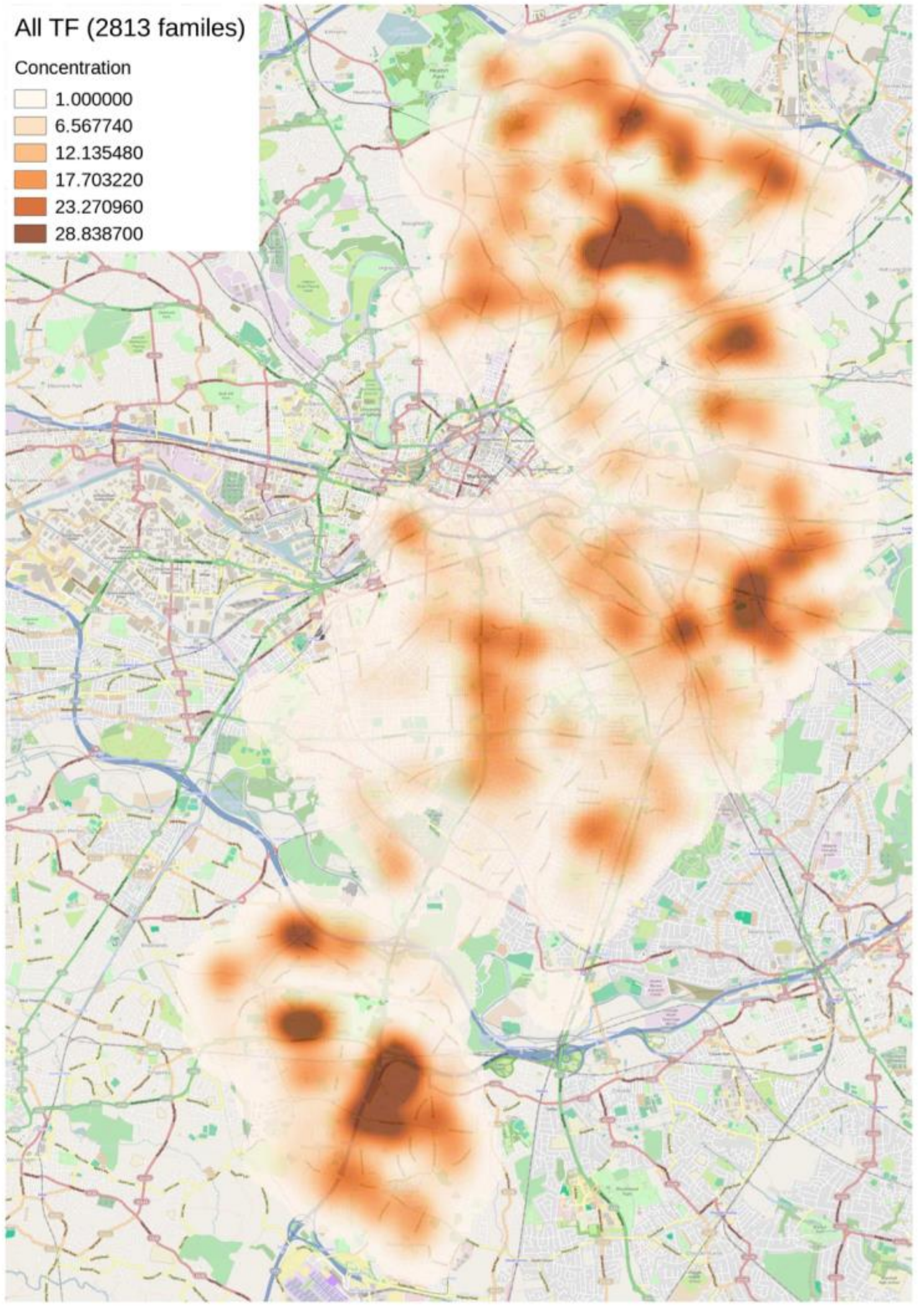


Spatial Analysis

Location of Troubled Families

These areas are similar with regards of Median Income, Types of problems the TFUs have, Deprivation indices

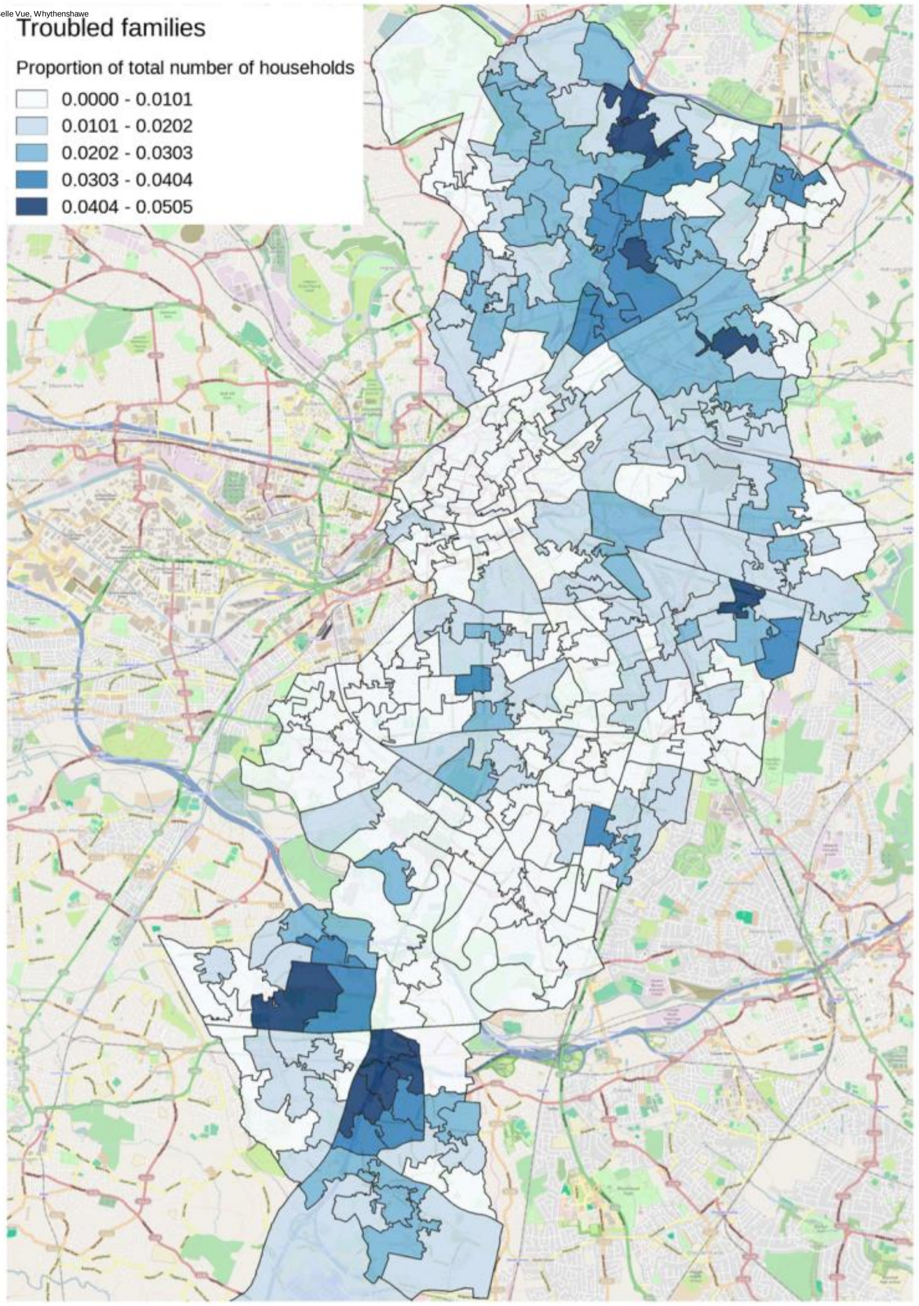
But they differ in: Tran



Density of Troubled Families

TF scattered across all the city. Very few areas with no TFUs (Didsbury). Number of TFU never exceeds 5% of all households in any LSOA

Substantial clusters: Harpurhey, Gorton / Belle Vue, Whythenshawe



Predictive Modelling

The techniques used in this project culminate to aid with Predictive Modelling, this would include:

- Using data mining to identify your clusters / cohorts
- Testing cluster / cohorts characteristics to identify the significant factors
- Apply the significant factors to the whole population to identify scale
- Use decision trees and sequence analysis type tools to test likely impact of decision for this broader cohort
- Use this to inform the business planning