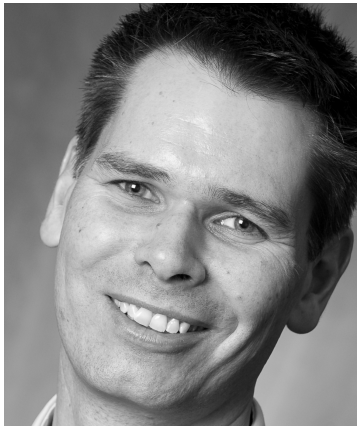
Debunked

The Top 10 Myths About Data Quality

The best source of
knowledge is experience

EQUILLIAN
Information Strategy

Your Presenter



- Jon Evans
- Information Strategist
  - 20 years experience
- Self-confessed Data Quality geek

- Founder of Equillian
- Experts in Enterprise Information Management
- Three founding values
  - Independence
  - Passion
  - Knowledge

EQUILLIAN
Information Strategy

Why Should I Care About Data Quality?

EQUILLIAN
Information Strategy

Straight from the Horse's Mouth

Differences in how data is captured, passed from one system to another, and later presented results in reluctance to trust the final output…

Incorrect or minimal data can be entered resulting in remedial work further down the line…

So what is the industry saying about data quality?

$6m has been spent on projects to clean up our data warehouse in the past two years…

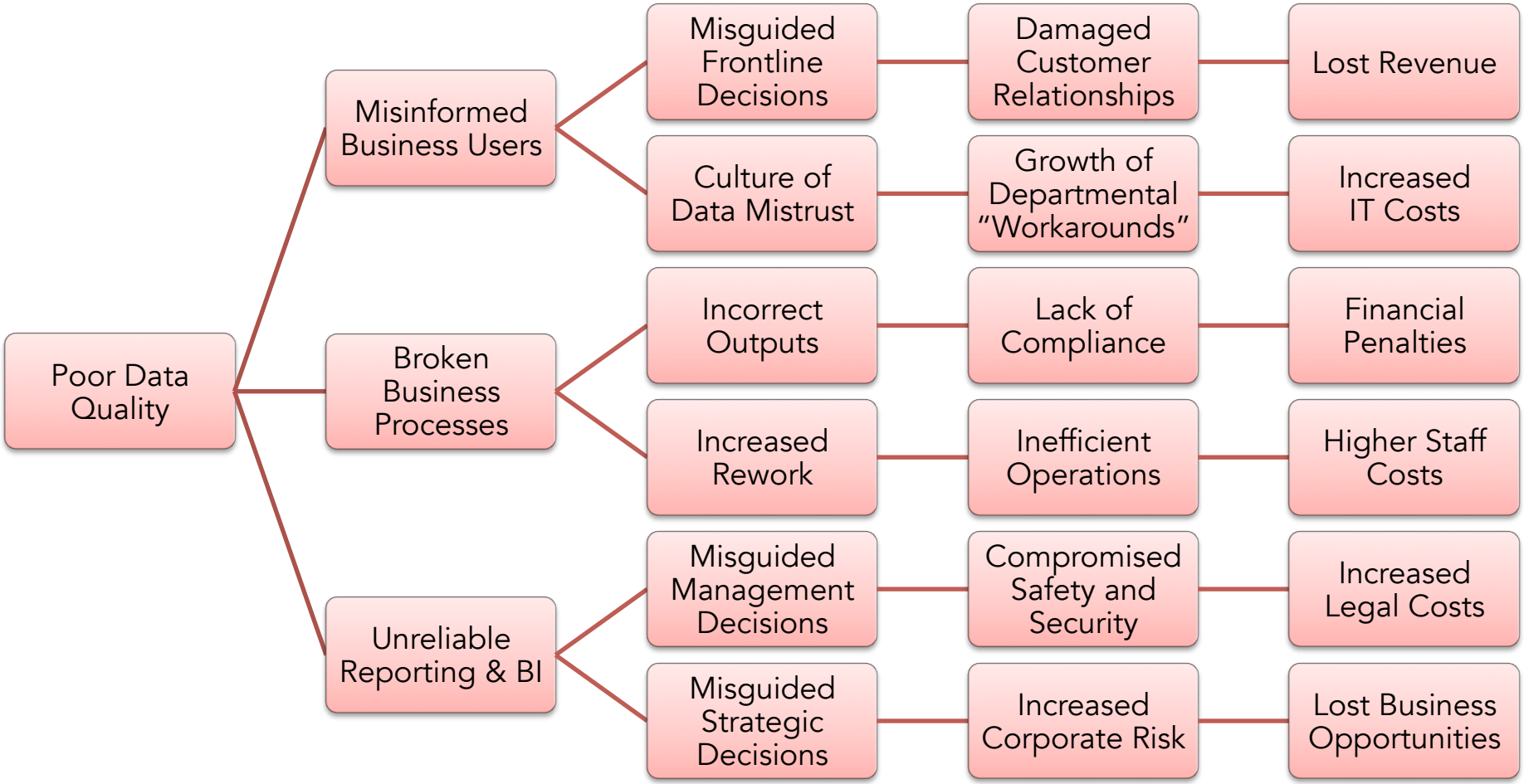
Key subject matter experts are relied upon to review detailed data from various systems to ensure accuracy…

You can't run our sort of business without good data and we're crying out for it…

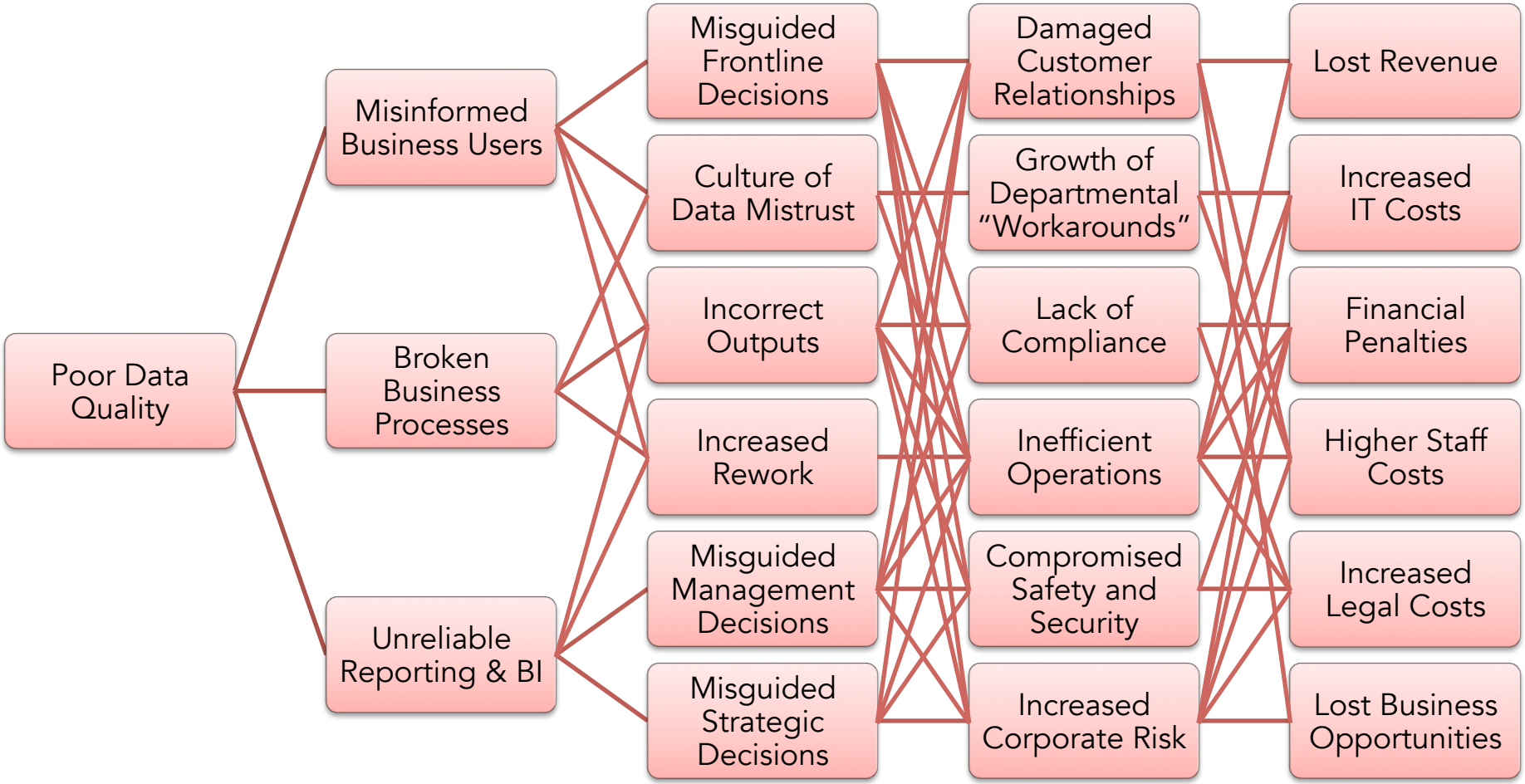
The financial impact of poor data quality is clear

- Actually, most of the time it's pretty subtle – many of the consequences of poor data quality only have an **indirect** link to financial impact
- When you talk about data quality, be prepared for the "**so what?**" question from the sceptics
- So create a compelling response
  - Use data profiling to quantify existing data quality levels
  - Collect qualitative (anecdotal) evidence from the business
  - Explore all the possible impacts (direct and indirect)
  - Develop your rationale and state your assumptions
- Just don't expect building a business case for DQ to be easy!

Financial Impact is Often Indirect

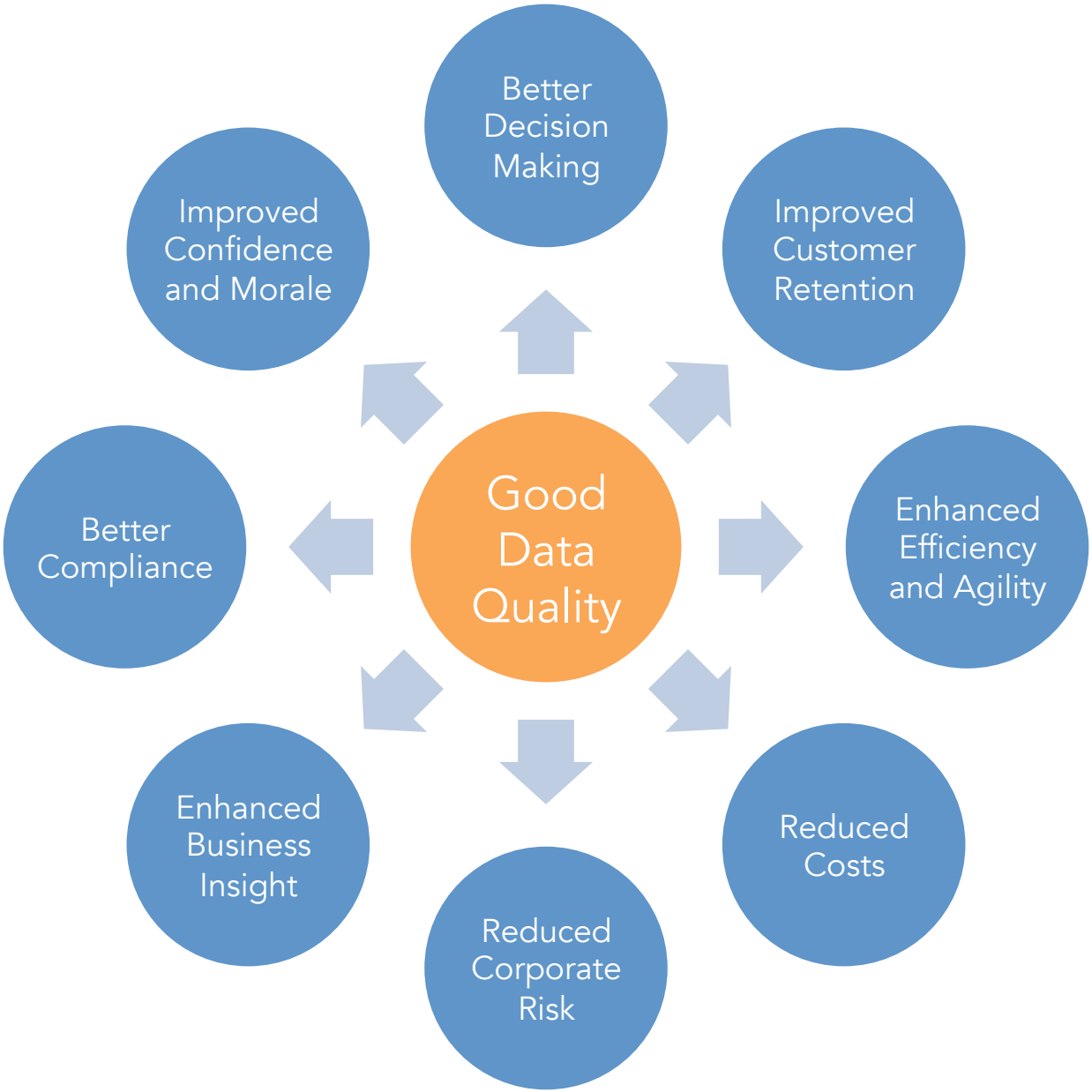


…But if Only it Were that Simple



Organisations with poor data quality will never be the most successful

- Really? I've worked with some extremely successful organisations, many of which were suffering widespread data quality problems
- Success in business is influenced by many factors – let's not pretend that data quality is top of the list
- However, good data quality **does** bring many benefits
  - Better decision making
  - Enhanced operational efficiency
  - Improved customer relationships
- So just think how much **more** successful your organisation could be armed with excellent data quality


Better Decision Making
Improved Customer Retention
Enhanced Efficiency and Agility
Reduced Costs
Reduced Corporate Risk
Enhanced Business Insight
Better Compliance
Improved Confidence and Morale
Good Data Quality

Poor Data Quality Management means poor data quality

- Not necessarily – some organisations manage to maintain DQ without having **any** formal Data Quality Management
- However, these organisations are undoubtedly…
  - Relying on luck, the heroics of their staff and significant manual effort
  - Completely oblivious to the current level of data quality and the direction of travel
  - Carrying a much higher level of risk
- Eventually, their luck will run out and the risks will materialise
- Far better to **ensure** data quality through a coherent set of policies, standards, processes and supporting technology

"Ultimately, poor data quality is like dirt on the windshield. You may be able to drive for a long time with slowly degrading vision, but at some point you either have to stop and clear the windshield or risk everything."

Ken Orr, the Cutter Consortium

Trying to tackle data quality before Data Governance is pointless

- It depends how long you want to wait – formal Data Governance takes time to become established and the business case is even harder to sell
- Fortunately, the **monitoring** side of Data Quality Management can start with very little Data Governance
- This non-intrusive "bottom-up" approach
  - Requires less business change
  - Establishes baseline data quality levels
  - Raises awareness of DQ issues
  - Helps to build the case for wider initiatives
- So don't delay tackling data quality – or you may never start

Data Warehouses provide the ideal starting point for data quality

- Sometimes – but only if the data is coming from **outside** your organisation and there's no means of accessing it at source
- By definition, your organisation's **key business information** is created and managed by your very own staff, processes and systems
- Therefore, tackle data quality at the point of entry in order to
  - Provide benefit to your operational functions
  - Ensure all consumers of the data see the very best version
  - Prevent DQ problems from being propagated downstream
  - Reduce inconsistency and aid reconciliation
- Why should BI be the only beneficiary of good quality data?

Full validation at the point of entry will ensure accurate data

- Oh, if only it were that simple! Ever moved house? Got married? Changed your credit card?
- Even the most comprehensive validation routines can't prevent data from degrading over time
- Just accept it and put in measures to try and reduce its effect
  - Devise processes that proactively check those data items most susceptible to change (e.g. condition rating of a piece of equipment, the email address of a customer)
  - Make use of third-party data sources to scrutinise your own data
- Finally, don't confuse validity with accuracy – your data might be 100% valid, but that doesn't mean it's 100% accurate

Organisations should aim for 100% accuracy in their data

- Only if you're prepared to continually report failure against a stationary target – not the best start to your DQ programme
- Setting an absolute target for any measure of data quality (not just accuracy) is often a bad idea for two reasons
  - There's generally no logical justification for the chosen figure (is 95% any less arbitrary than 93%?)
  - It doesn't reward continual improvement – one of the key characteristics of a good Data Quality Programme
- Especially in the early days, it's better to measure direction of travel against the baseline and reward progress
- Data must be **fit for purpose** – perfection is rarely essential

Quality vs Fitness for Purpose

Quality: *"the standard of something as measured against other things of a similar kind"*

Fit (for purpose): *"well equipped or well suited for its designated role or purpose"*

Source: Oxford English Dictionary

So when we talk about data quality, do we really mean data fitness? Let's go camping and find out!

I need a new tent for my next holiday

Sleeping Capacity	2 Adults
Dimensions (Folded)	10cm x 10cm x 30cm
Weight	3kg
Max Wind Force	Force 10
Assembly Time	15 mins (1 person)
Guarantee	2 Years

- What's more important…
  - the standard of my chosen tent as measured against other tents (i.e. its generic **quality**)?
  - or that my chosen tent is well suited for the type of holiday I'm planning (i.e. that it's **fit for purpose**)?

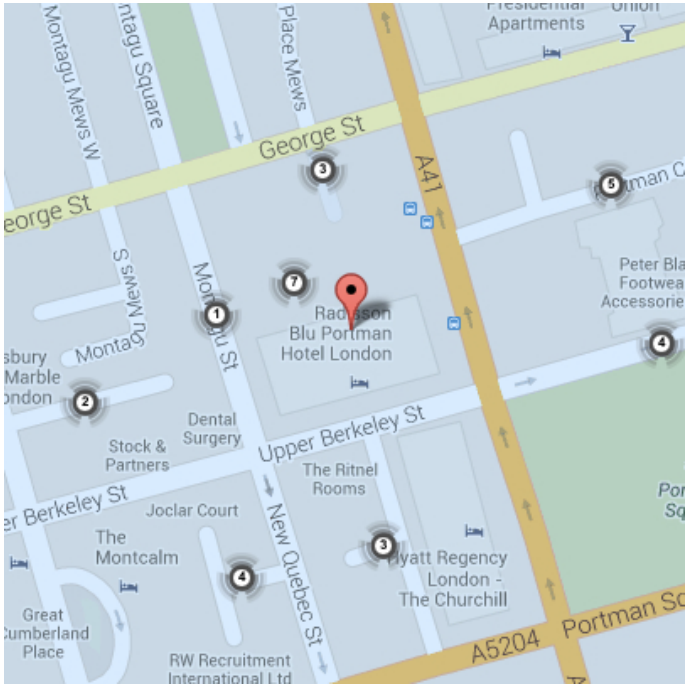
Generic quality doesn't really matter provided data's "fit for purpose"

- Well, yes…and no. It's true that data needs to be fit for its intended use, but what about all the unintended uses?
- Organisations are now starting to recognise the true value of information can only be realised through re-use, including
  - Extensive downstream analysis of operational data
  - Data mashups involving previously unconnected data sources
- New uses will constantly emerge so measuring "fitness for purpose" across every potential use-case is impossible
- Data quality isn't an exact science, so don't disregard **generic** measures of DQ as a useful yardstick when "fit for purpose" today doesn't necessarily mean "fit for purpose" tomorrow

The problem is, circumstances change

Let's Look at a Real Example

In February 2011, the UK government launched a crime-mapping website for England and Wales (www.police.uk)



Unfortunately, for a number of reasons, the postcode allocated to a specific police incident didn't always correspond to the precise location of the crime.

The net result was that poor accuracy in the recording of geographical information led many quiet residential streets to be incorrectly identified as crime hotspots.

NEWS LANCASHIRE

Home | World | UK | England | N. Ireland | Scotland | Wales | Business | Politics | Health | Education | Sci/Environment

1 February 2011 Last updated at 14:44

Share

Preston street branded most crime-ridden

A look round 'crime-ridden' Glovers Court, in Preston

A street in Preston has been branded the most crime-ridden in England and Wales by a government website.

Glovers Court - a quiet street in Preston city centre - has been the location for 150 offences in December, according to the online map.

But police say the crimes were actually committed across the whole of Preston city centre.

Ch Supt James Lee said only three crimes had been reported at Glovers Court.

"I don't accept these figures. The postcode relates to the whole of the city centre."

He said the people of Preston should be proud to live there. "All crime was actually down 4.5% during the month of December in the city centre."

A spokesman from the Home Office said the figures on **www.police.uk** were supplied by Lancashire Police.

The spokesman said: "The information is provided by local forces and we simply input the raw data."

Glovers Court, which is off the busy city centre thoroughfare of Fishergate, does not have any flats or houses - but it does boast two pubs - the Wellington Inn and Glovers - as well as Brown's Cafe Bar.

'Cheesed off'

Simon Nash, who lives on nearby St Austins Place, is dismayed the figures put Glovers Court at the top of national crime statistics.

Mr Nash said: "I have lived here for eight years and have never ever seen one crime.

"As a resident it's cheesed me off as it isn't a violent place to live; it's a great place to live."

NEWS LANCASHIRE

Home | World | UK | England | N. Ireland | Scotland | Wales | Business | Politics | Health | Education | Sci/Environment

1 February 2011 Last updated at 14:44

Share

Preston street branded most crime-ridden

A look round 'crime-ridden' Glovers Court, in Preston

A street in Preston has been branded the most crime-ridden in England and Wales by a government website.

Glovers Court - a quiet street in Preston city centre - has been the location for 150 offences in December, according to the online map.

But police say the crimes were actually committed across the whole of Preston city centre.

Ch Supt James Lee said only three crimes had been reported at Glovers Court.

But does it really matter?

In the context of creating aggregated statistics to assess relative crime rates between counties, the data quality is perfectly acceptable.

✓ Data fit for purpose

However, if the same data is used by an insurance company, there is an issue for the homeowners who receive inflated home insurance premiums.

✗ Data not fit for purpose

The business own the data, so data quality is their problem

- At long last, we recognise that data doesn't belong to IT – but is it really fair to shift **all** responsibility to the business?
- **People** in general are the primary cause of poor data quality
  - Business users not taking enough care when entering data
  - Technical staff not taking enough care when designing systems
  - Senior management not taking data quality seriously
  - Everyone thinking it's someone else's problem
- The business must acknowledge that **data is a business asset** and they need to drive the data quality agenda from the front
- But **everyone** must play their part in making it happen – including IT, end-users and senior management

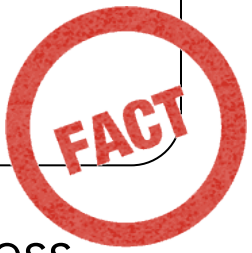
Data Quality Management requires significant investment in technology

- Hmmm – let's put things in context
- It's true that certain data quality activities are far simpler if you streamline them using clever technology, but don't forget
  - Options range from traditional licenses to subscription-based models to no-fee open source software (community editions)
  - It's possible to get started with a subset of the available functionality (e.g. monitoring and reporting)
- Also, regardless of which route you take, there will be many data quality challenges that no technology can overcome
- If you're worried about the technology costs, here's a reality check – the total cost of your DQ programme will be **far** more

My Advice – Ignore the Myths

1. Building a business case for DQ is necessary, but don't expect it to be easy
2. Poor data quality isn't a barrier to success, but tackle it and you'll reap the rewards
3. Don't rely on the heroics of your staff – take a formal approach to data quality
4. Start your data quality journey now – don't wait for Data Governance
5. Fight the data quality battle at the frontline, not in the data warehouse
6. Don't expect good data to remain good forever, if you do nothing to maintain it
7. Data must be fit for purpose – don't strive for perfection or you will always fail…
8. …but also don't assume that "good enough" today means "good enough" tomorrow
9. The business needs to lead the way, but everyone has a role to play in data quality
10. Ensuring DQ requires much more than technology, so budget accordingly

The best weapon in the fight against poor data quality is education



- Absolutely! Nothing can compete with people's awareness and behaviour when it comes to tackling data quality
- Educating your staff can start with no changes to processes, no investment in technology and only limited budget
  - Schedule **high-level briefings** with senior management, to highlight the importance of data quality to your organisation
  - Set up informal **"lunch and learn" sessions** to help data quality awareness across your user community
  - Arrange **specialist training** for IT to introduce data quality best practice and ensure systems are designed with DQ in mind
- Your staff may well be the most common cause of poor DQ, but they're also the most effective cure – educate them

Summary

- Despite the myths, there are two genuine reasons you should care about data quality
  - poor data quality can have subtle consequences, but will eventually hit your bottom line
  - good data quality can bring many benefits, and will help to make you even more successful
- People are the root cause of most data quality problems
- But if you educate them they'll also be your most effective weapon

Thank You for Listening

For more information contact
jon.evans@equillian.com