

how stories become data and data makes stories

Tobit Emmens
Research & Innovation, Devon Partnership NHS Trust

@tobite / tobit.emmens@nhs.net

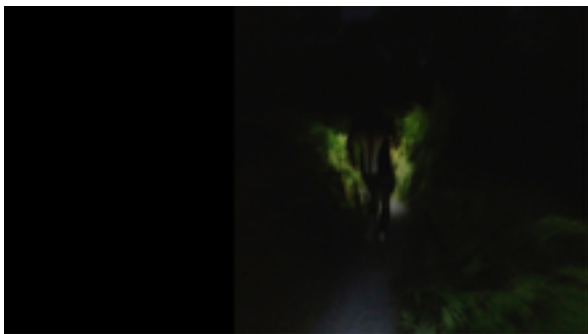
12/11/2014 - BBC Data Day

its 4am, I'm cold, exhausted - mentally, emotionally, physically. My right knee has just given out. My one man, skeleton crew suggests, no tells me, that I should should sleep. But if I sleep...

I've been up for the previous 22 hours, and I started running 16 hours ago, and somehow, despite my best plans and preparations I have only managed to cover 65 miles. I spent so long prepping for this



I give in, and as I crouched in my tent, putting on pretty much all my clothes, shivering uncontrollably, I lay down, and with in seconds passed out. An hour later, Andy comes back and wakes me. 5:15AM and I eat some crisps, a bacon sandwich, take some layers off, find my head torch, my cold, damp shoes and stagger out of my tent. I let the event team know I am back on the course, and start out again. At this point, Andy, who has been by my side for most of the last 7 hours tells me he needs to go home to bed. What can I say but thank you.



My legs are stiff, unwieldy, and they hurt, but I set out at a slow trot, into the darkness. It is true, that the hour before dawn is often the darkest. And this particular dawn, also felt like one of the loneliest. I stumble around a 5 mile trail course, I had already done 13 laps so i know it pretty well, but in this pre-dawn period, nothing makes sense, the landscape seems almost alien, its really hard to work out what is, and what is not real. The greyness of dawn arrives, where light and dark merge, things seem to emerge,

and for a short while, as the sun climbs, everything makes sense. Until I am jolted as I jar my knee.

My ultra running experiences remind me very much of my data experiences. At heart, I'm more of a qualitative kind of person, an ethnographer. I see, hear and watch the stories of people who use NHS services unfold around me. In a world of key performance indicators, metrics, huge datasets, performance management data. What I notice is that the more we sanitise and process the data, the more abstract it becomes.



So this is Devon

and to work within data protections, etc, what you are about to see is somewhat fictionalised.



These dots shape how we as an organisation provide aspects of care, together they tell us where to focus attention, where to place interventions.

BUT what do you think these dots actually mean?

each dot represents where someone died by suicide during 2011-2013 - around 500 people, or around 200 people each year.

We add this data to our databases, that have multiple streams of data flowing in to it, we process, analyse it, bases decisions on it, argue over it.



But what about this, somehow we end up representing John, who having recently been made redundant is on school run duties. After dropping his children off at school, smiled at his friends, texted his wife about picking up some apples, drove to a railway station, bought a ticket, stood on the platform for an hour, before stepping into the track in the path of an oncoming fast train.

Somehow we have to navigate the tension of respecting John, his family and friends and the people who worked with him, and the ever

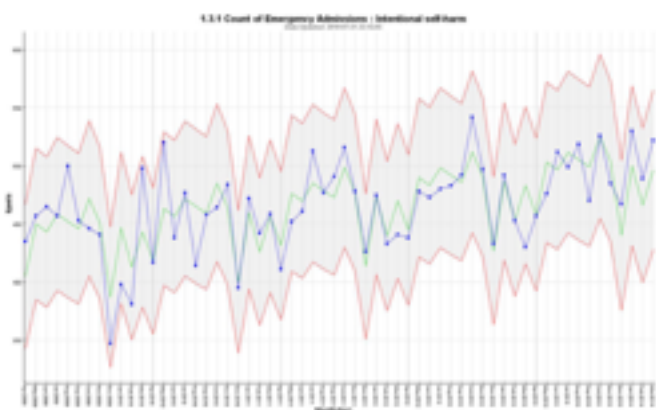
pressing need to have and use data.

As John's story gets turned into data, along with another 200 or so other suicides each year, strategy is made and policy is written, and reports generated, action is taken in an attempt to improve this situation.

and we commit to doing more, but the charts don't get better, they get worse.

This is the big picture for people being admitted to hospital for Self Harm for the period 2009 - 2014 (the last 5 years) a big picture

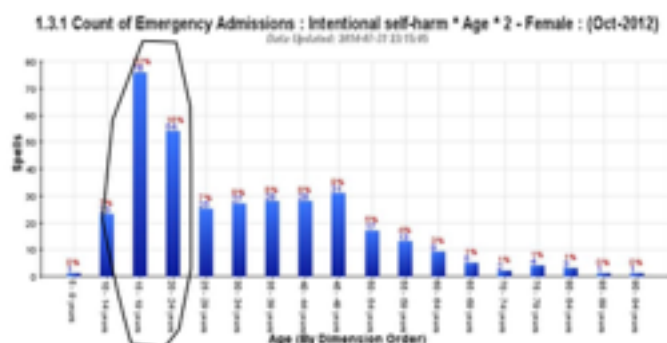
The rise continues and continues. Each one of those 500 data points each month is a person. A person dealing with some of life's most difficult traumas. so what is behind these data



this shows us what we often think about when we talk about self harm

By far the largest peak in self-harm is young women - a one month snapshot shows over 70 teenager between 15-19 attending A&E for self-harm.

(as an aside, we know that the self-harm that presents at A&E is just the very tip of the iceberg)



This group takes up a huge amount of media attention



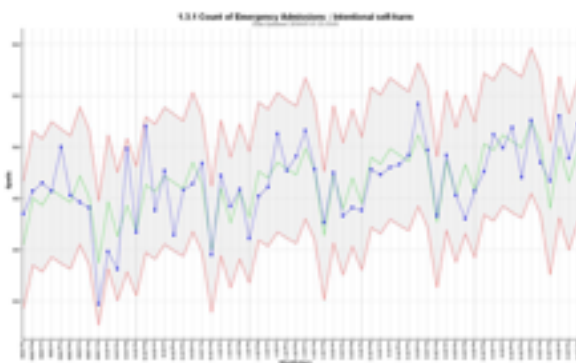
But what about the others? aside from young men, a group that is declining, The fastest climbing group for self harm and for suicide are middle aged men... and researcher have largely attributed this to the impact the financial crisis,



the biggest increase in self harm and suicide is men between 40 and 55



We have 'suicide generation, suicide risk men, unemployment, debt and breakdown... and researchers, and the media have largely attributed this to the impact the financial crisis,



lets go back to this chart

with all the access to data (and when I say data, I mean the myriad of data streams from health, social media, work, spending patterns, we still don't seem to have the data tools we need to help prevent this.
- or do we?

samaritans radar - who has heard of this?



Samaritans are here for people who need a listening ear. We've created Samaritans Radar to give you a chance to help friends who may need support.

Samaritans Radar will flag potentially worrying tweets that you may have missed, giving you the option to reach out to those who may need your support.

But don't worry, you won't be on your own in this — Samaritans Radar will help you to reach out to someone you're worried about by offering guidance and suggestions.

Samaritans Radar is completely free so activate it now by clicking on the button below to connect it to your Twitter account.

what could possibly go wrong?

one of the biggest, messiest social media backlashes for an app with the best intentions



And so, after two years of designing, testing, at the weekend, Samaritans have pulled the plug on the app,

and yet, we would be foolish to think that all our online data isn't already harvested, analysed, categorised and sold, its just mostly this is hidden from our knowledge

Another example of where social media data, and clinical health data is being used in an automated textual analysis and comparison with others is the DURKHEIM project

these tools however, were built with clear privacy controls, an informed consent process and clear info stream back to the end user detailing how their data is being used

DURKHEIMPROJECT Home Our Project Demo Research Results Our Team News Contact Us

OUR PROJECT

Current State of Art:
Diagnosis of psychological health and the prediction of negative events, such as suicide, or their ideation are limited by:

- A lack of broad societal attention to problem & clearly defined signals for interventions.
- A lack of real time, near real time reaction capability to large volumes of data.

The Need:
Broader coverage of suicide risk detection and better understanding of the expression of suicide ideation through data mining of text and images.

Proposed Solution:
Continuous monitoring of social network user behavioral intent enabling intervention, facilitated by social/online/mobile data sources.

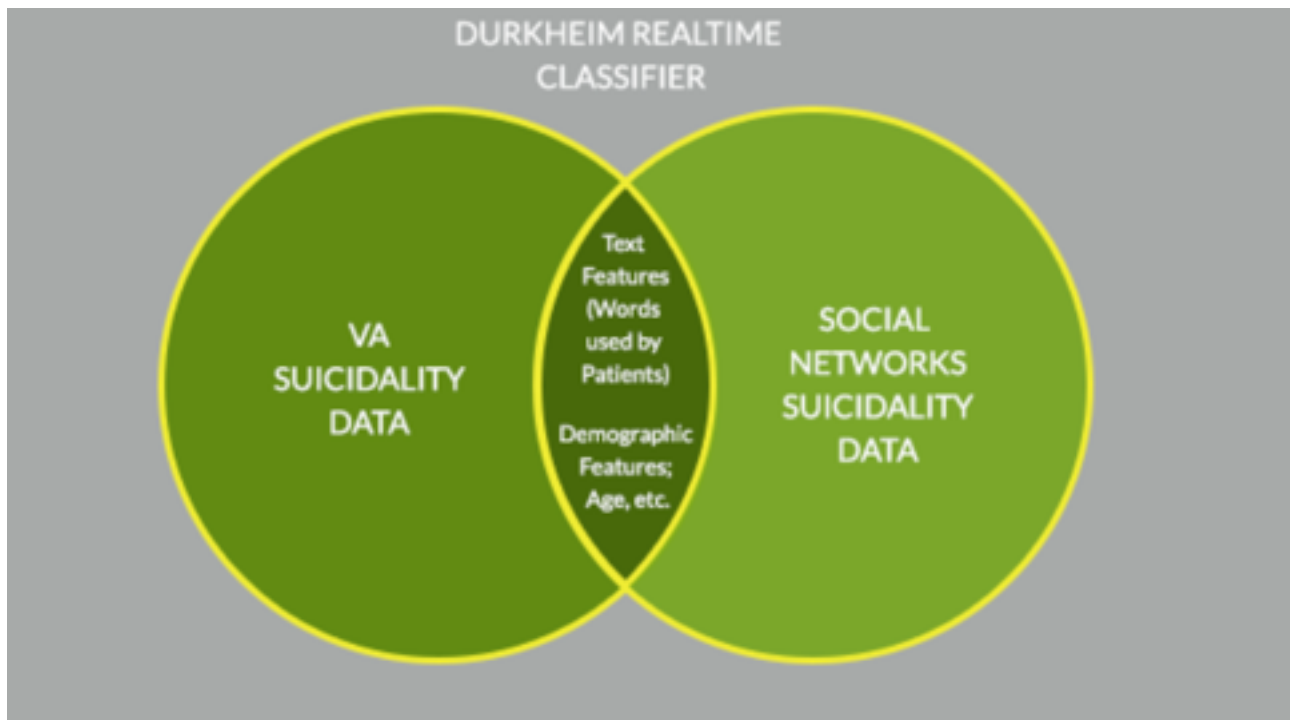
DURKHEIM REALTIME CLASSIFIER

VIA SUICIDALITY DATA

SOCIAL NETWORKS SUICIDALITY DATA

Text Features (Words used by Patients)
Demographic Features: Age, etc.

what does this look like?



a linguistics driven prediction model to estimate the risk of suicide

tweets being cross referenced to Facebook being cross referenced to clinical notes, being cross referenced to anonymised data from other similar people, that gets presented with an over all risk score. A risk score above certain amount gets flagged and monitored at minute intervals, added to this is a confidence score (or a probability rating) that the data is right. This is then tracked over time.

The clinician can then access this data, hundreds of data points, visualised in a meaningful way, that can inform a human, 1:1 interaction

tweets + facebook + clinical notes + patient reported data could be used to continually monitor seemingly healthy individuals at a primary care level and to continually evaluate the suicide risk among psychiatric patients

And this is where it is quite different from SamaritansRadar - rather than taking a single tweet, without consent, and informing the persons network, the Durkheim Project, takes a high volume of data, with the persons consent, and feeds it into a 1:1 human interaction

So we have flown through how stories make data, and really I am assuming that that was nothing new,

To finish up, I want to briefly touch on how data contributes to and shapes stories, and what is the role of the media in this?

We have a cycle of health/life events making the media, which in then makes data which directly influences health / life events - well, that's what some of the theorists seem to think

popular belief (popular press?) would have us think that we tend to mimic what we see on TV or read in the newspapers...

This view has developed from 'Social Learning Theory', this theory has featured in the media / violence debate, and somehow has gained dominance in the media and suicide prevention field (this is the topic of my masters, but for another talk!). In summary, Social Learning Theory has become a dominant model for trying to understand what and why people do what they do.

At its most basic, it goes like this: the media report, or films portray, and people imitate.

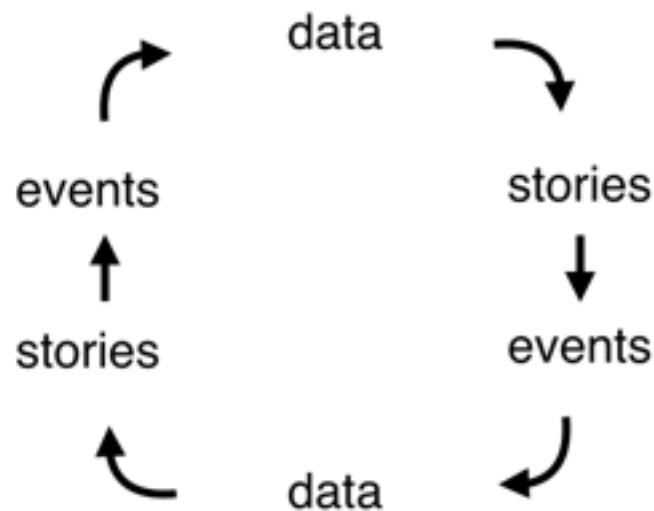
And then on this policy decisions are made and sometimes its bad because we have this and sometime this is good and we have we have media guidelines for the reporting of suicide.



So to be fair, there sometimes *is* a correlation between what people read or watch and their subsequent behaviour - but **its complicated**,

The issue is that many studies and published research appears to ignore the large and very conflicted field of mass media research.

So we have this kind of cycle



In the last decade research in this area appears to have become more closely aligned with demonstrating harm than in attempting to unravel the relationship between media and audiences and subsequent behaviour.

We know that the media are not constant as a potential or actual influence over time and between places (McQuail, 2008:464). We know that media influences can be transitory.



The other big issue we have is that research often sets out to address different forms of media communication (including films, television, fiction and non-fiction print), these media are still divided up, studied and viewed independently. While this, in part, must be for pragmatic reasons, no explanation or evidence is offered to support such a segmented understanding of the role of media in peoples lives. The work of media scholars

(Meyrowitz (1985) Briggs and Copley (2002) and McQuail (2008)) demonstrate, people assimilate information from multiple sources, fiction and non-fiction.

My parting plea is this:

We know that complex multidirectional relationship exists between people, the social world they live in, social media and mass media.

The creation of media artefacts (from news stories to tweets) is inescapably linked to the social, political and cultural context in which they are written and consumed and therefore these data streams and the things we do with them needs to be located within these contexts.