

How to tell your boss story points are a terrible metric

Chris Wilkinson and Liz Ince



DevOps & DevEx: Remove friction, ship code and add value
QCon London 2019





***One metric to rule them all,
One metric to bring them all,
One metric to find them,
and in the darkness bind them***



Chris Wilkinson

Lead Software Engineer and Architect

chris.wilkinson@capgemini.com



Liz Ince

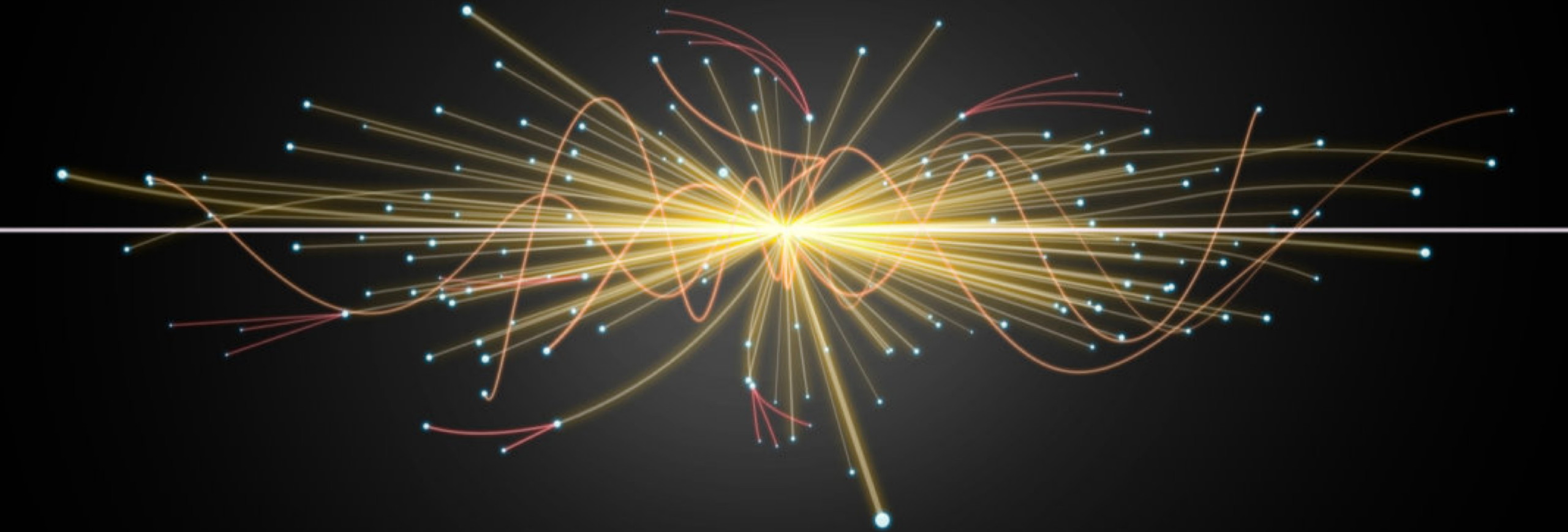
Enterprise Architect

liz.ince@capgemini.com



Metrics and Measurements in general

Heisenberg - *Any attempt to measure precisely the velocity of a subatomic particle, such as an electron, will knock it about in an unpredictable way, so that the act of measuring causes a change*





Target

noun, often attributive



- a. a mark to shoot at
- b. a target marked by shots fired at it
- c. something or someone fired at or marked for attack
- d. a goal to be achieved**



"When a measure becomes a target, it ceases to be a good measure."

- Goodhart's Law



Metric Fixation



1

Put in place
standardised metrics

2

Make the results
public – league tables
and rankings

3

Reward or punish
on the results



Gaming the system



So what DOES drive people?



Frederick Herzberg's two-factor theory – Motivation/Hygiene



... and the motivation factor?



So what does motivate people?



Recap

Goodharts Law

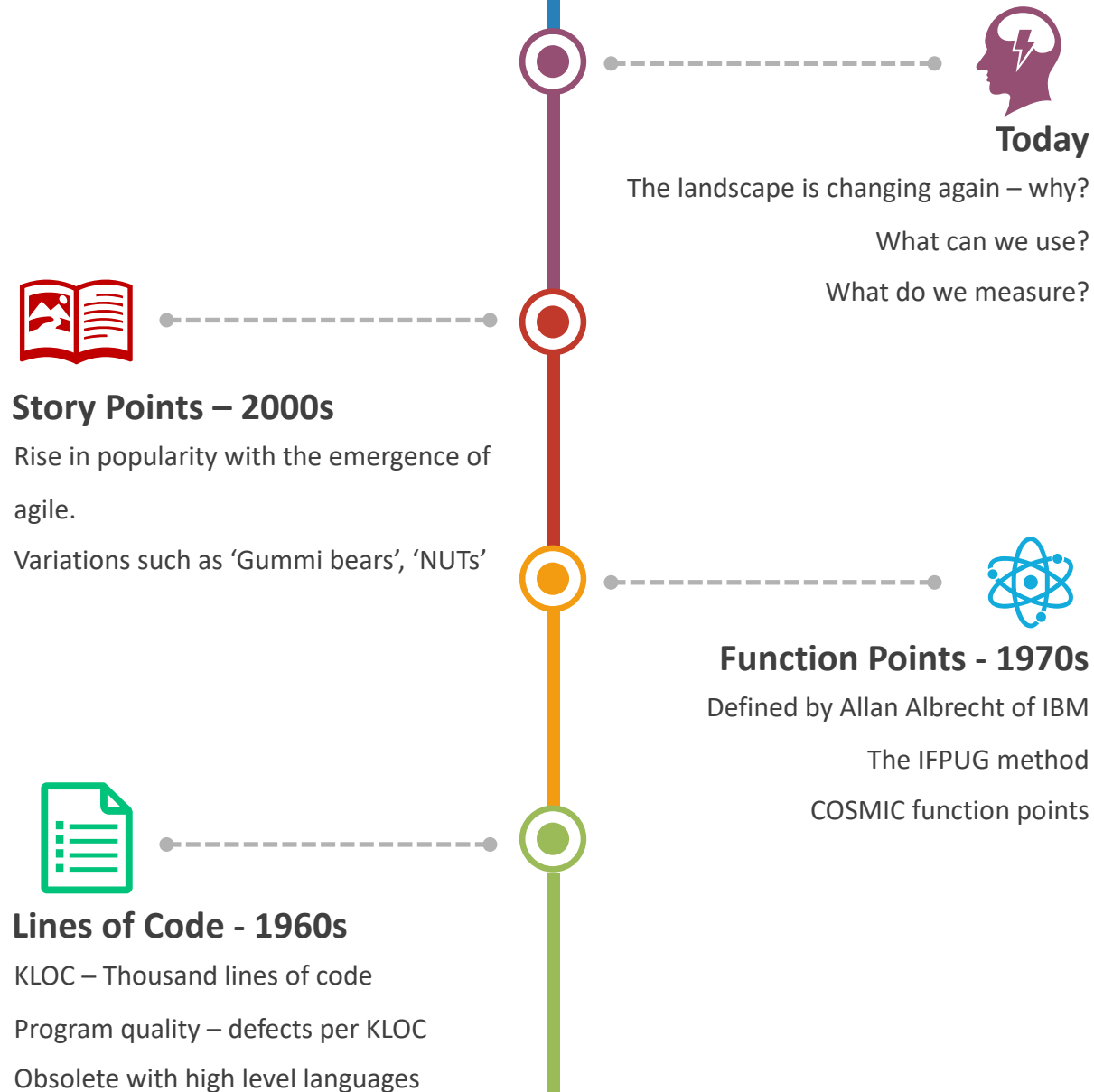
Reward & Punishment

Gaming the system

Motivation

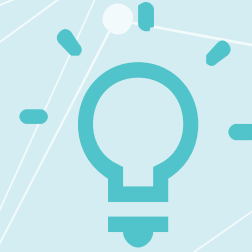


History





Working software is the
primary measure of progress.
—Agile Manifesto



The issues that are most
important, long term, cannot be
measured in advance.
—W. Edwards Deming



Question



Rachel Thomas

@math_rachel

co-founder [fast.ai](#) + professor @[usfca_msds](#) |
past: Duke math PhD, quant, early Uber eng |
now: making neural nets uncool

San Francisco, CA

[fast.ai](#)

Joined May 2013



1 30



Rachel Thomas

@math_rachel

Follow

Metrics are just a proxy for what you really care about, and unthinkingly optimizing a metric can lead to unexpected negative results, such as YouTube heavily promoting conspiracy theories. [fast.ai/2019/01/29/fiv ...](#)

... is an example of [the tyrann](#)
... just a proxy for what you rea
... unthinkingly optimizing a met
... ted, negative results. One ana
... n the UK began publishing the
... rgeons, [heart surgeons began](#)
... (but necessary) surgeries to
... s as high as possible.

... leads to negative outcomes.
Worldwide, people watch [1 billion hours of YouTube per day](#) (yes, that says PER DAY). A large part of YouTube's success has been due to its recommendation system, in which a video selected by an algorithm automatically begin playing once the previous video is over. Unfortunately, [these recommendations](#) are disproportionately for conspiracy theories promoting [white supremacy](#), climate change denial, and [denial of the mass shootings](#) that plague the USA. What is going on? YouTube's algorithm is trying to [maximize how much time](#) people spend watching YouTube, and **conspiracy theorists watch significantly more YouTube than people who trust a variety of media sources**. Unfortunately, a recommendation system trying only to maximize time spent on its own

1:17 AM - 9 Feb 2019

39 Retweets 94 Likes



What happens when it goes wrong?



Project Planning

**Risk of failure
and abuse**

Delivered Software



Agile things



What People See



How many story points have been completed?
Why is an individual completing fewer points this sprint?
Why is the team delivering fewer than last sprint?

What People See Don't See



“BS”



Subjective



Velocity



Relative

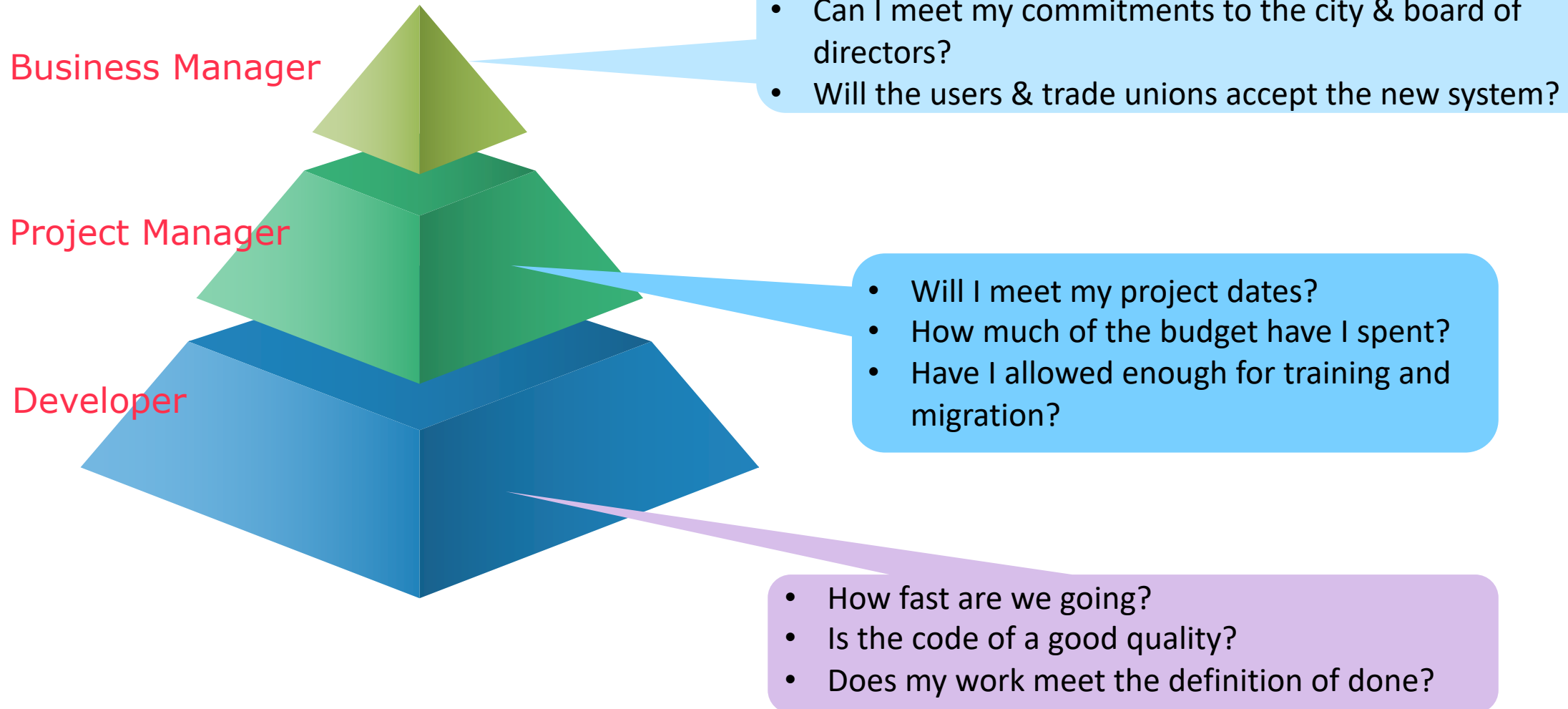


Bringing it all together



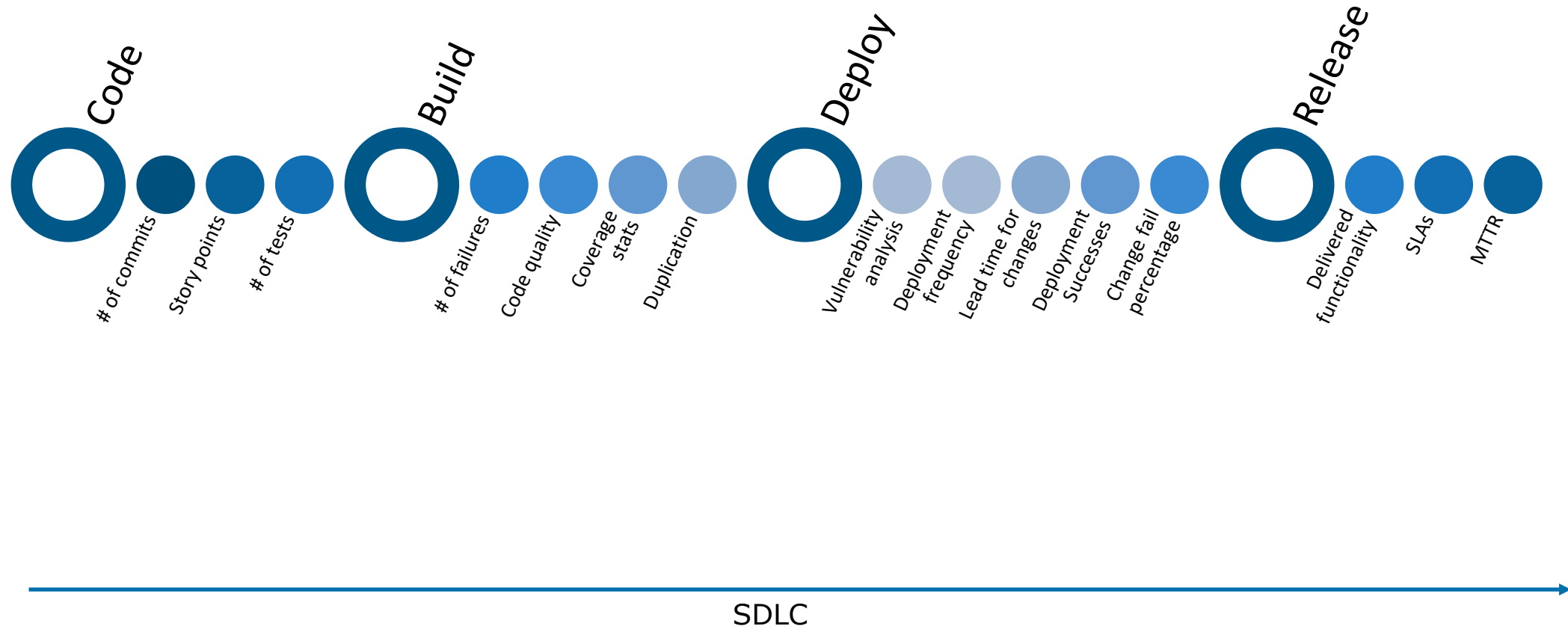


What is keeping them up at night?





Metrics within the SDLC

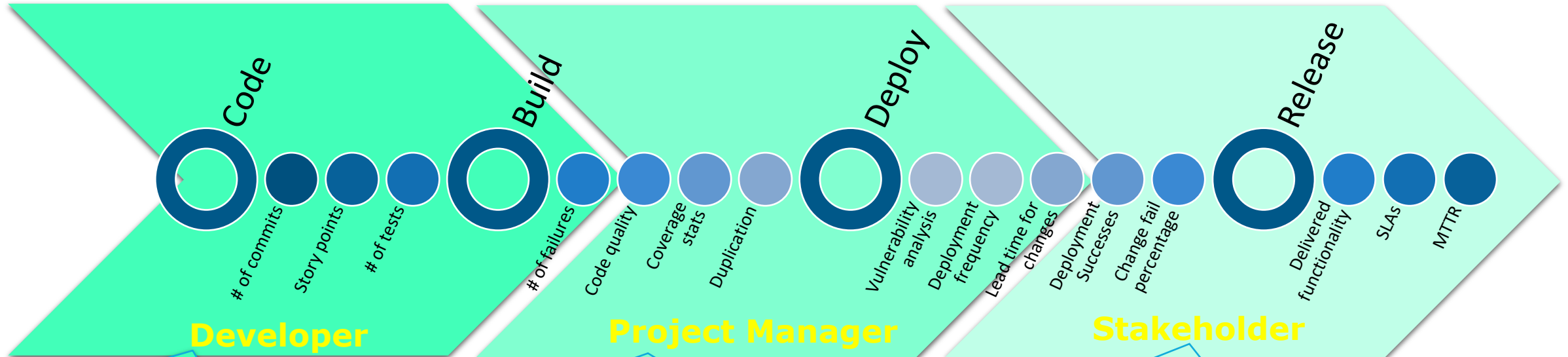


Correlation....



How many days have we used?
Are we ahead? Behind?
What issues are there?

How much of the budget have we spent?
Will we meet our deadlines?

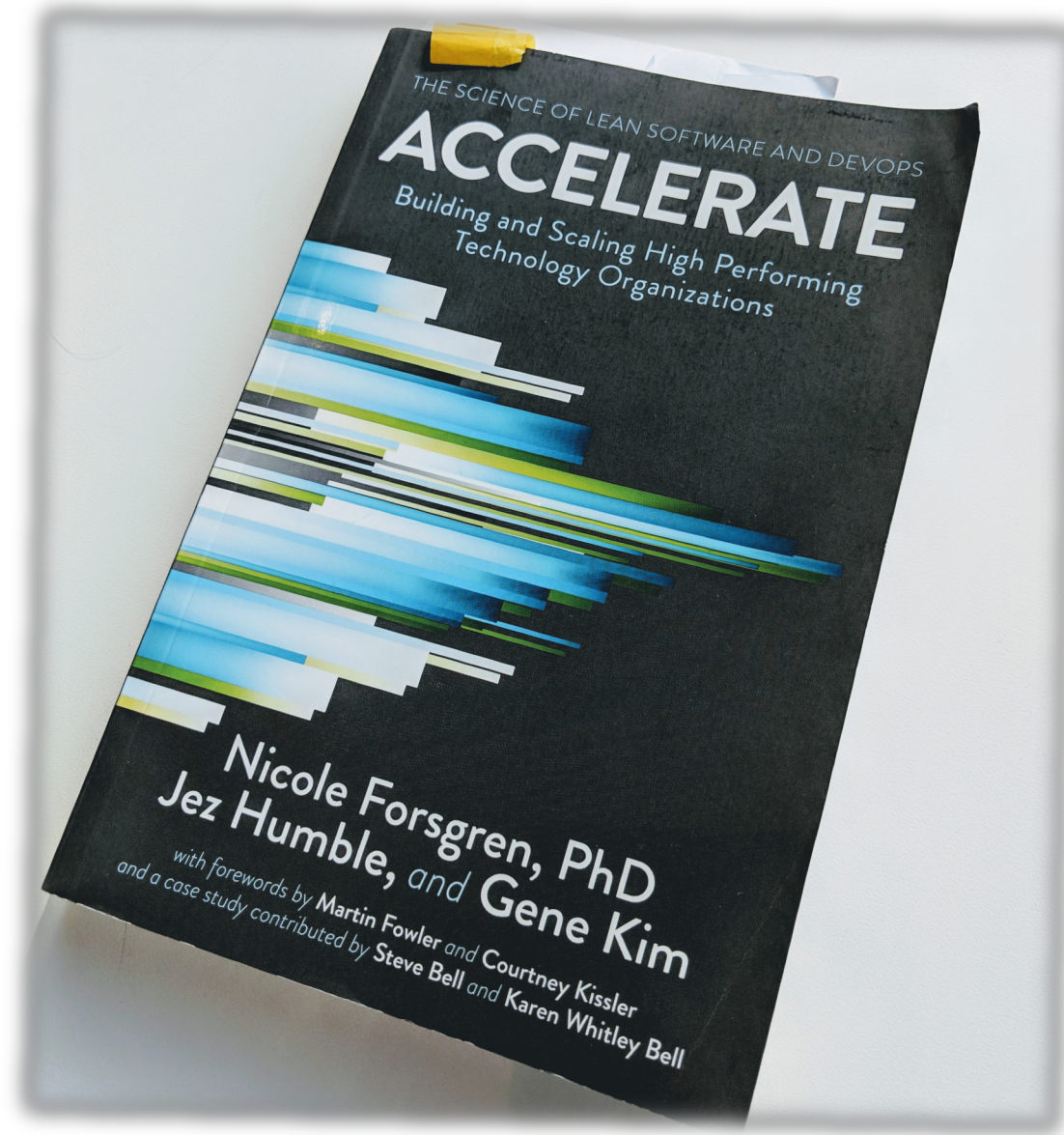


- Will I meet my sprint goals?
- Is the code of a good quality?
- Am I committing my code to master?

- Will I meet my project dates?
- How much of the budget have I spent?
- Have I allowed enough for training and migration?

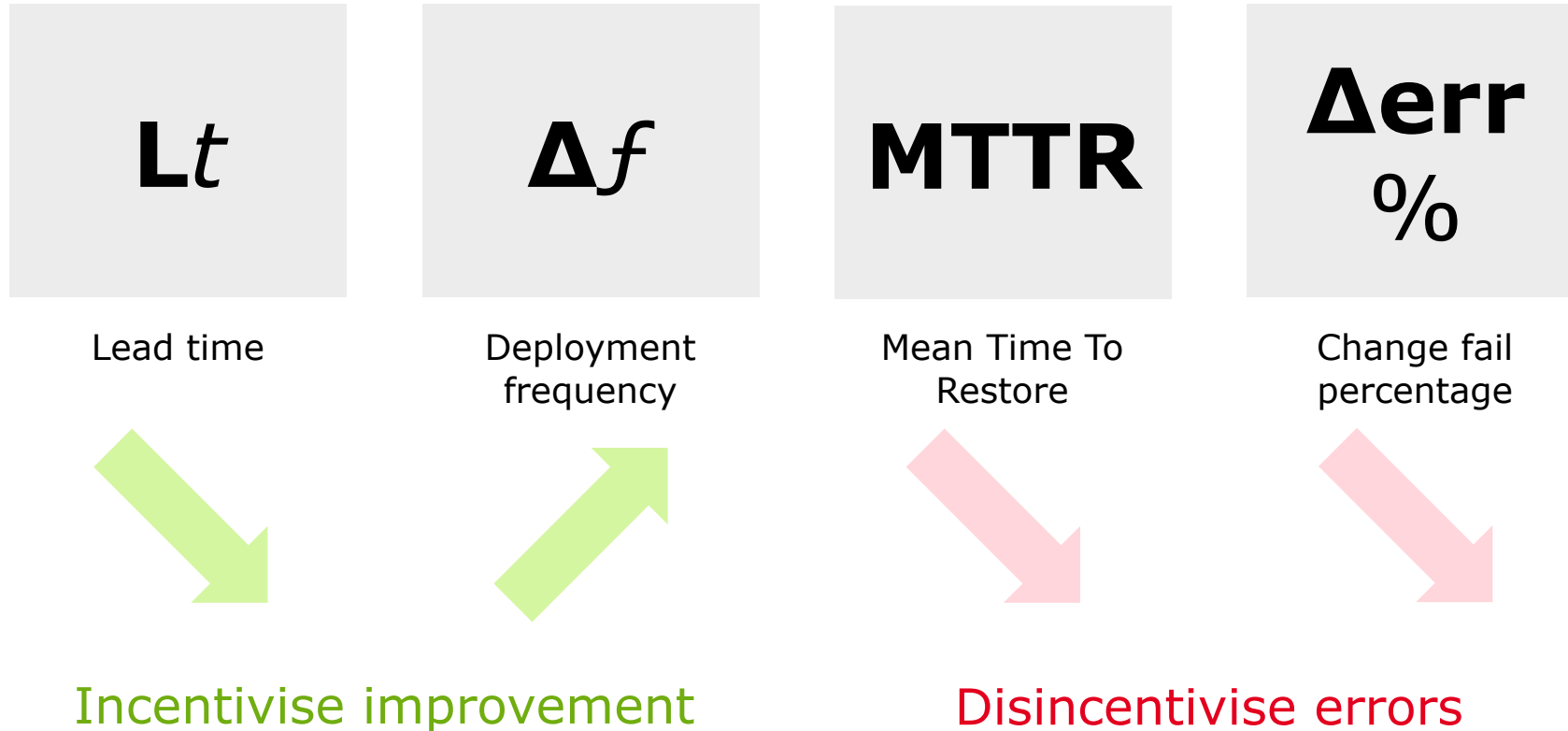
- Can I meet my commitments to the city & board of directors?
- How long will it take to restore service in a DR scenario?

SDLC





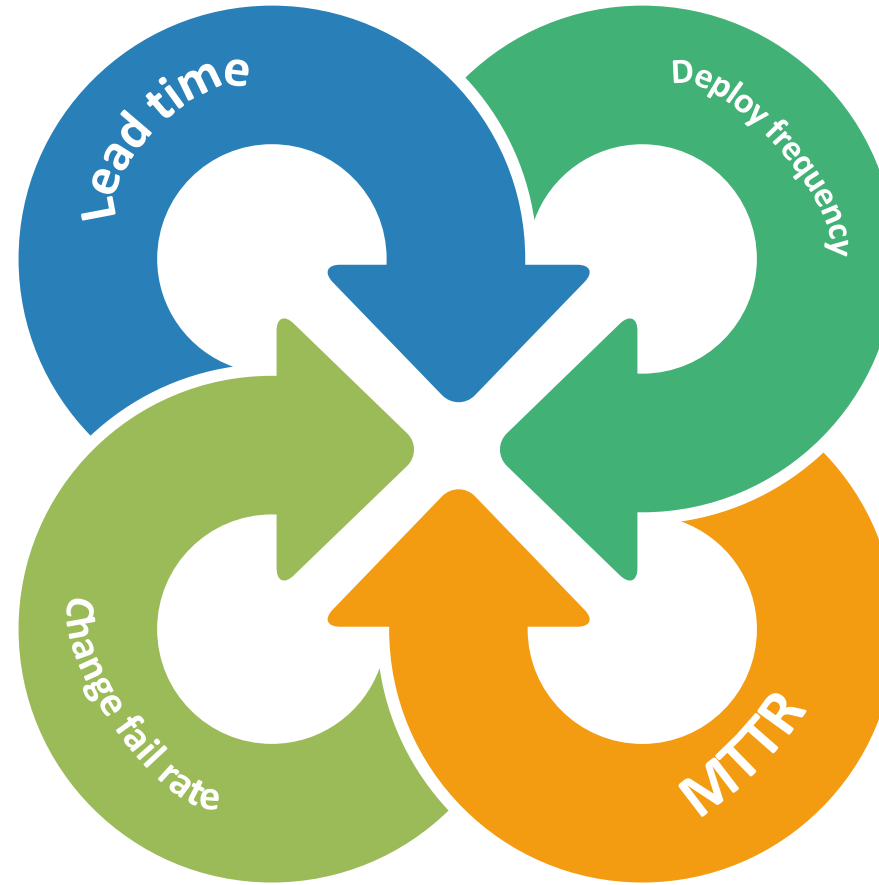
Measures that correlate with high performance



The Virtuous Cycle



Lead time for changes
The time it takes to go from code committed to code in production.



Deployment frequency
The frequency which you deploy your changes to production or to an app store.

Change fail rate
A measure of how often deployment failures occur in production that require immediate remedy.



Mean time to restore
The average time it takes to restore service in a DR like scenario.

Measuring Culture





Westrum Typology of Culture

Pathological (power-oriented)	Bureaucratic (rule-oriented)	Generative (performance-oriented)
Low cooperation	Modest cooperation	High Cooperation
Messengers shot	Messengers neglected	Messengers trained
Responsibilities shirked	Narrow responsibilities	Risks are shared
Bridging discouraged	Bridging tolerated	Bridging encouraged
Failure leads to scapegoating	Failure leads to justice	Failure leads to inquiry
Novelty Crushed	Novelty leads to problems	Novelty implemented



Robert Karasek –

Professor, Department of Work Environment
and

Director of JCQ Center - University of
Massachusetts Lowell (Boston)

Professor Emeritus, Copenhagen University,
Institut for Psykologi, Denmark








Westrum Typology of Culture

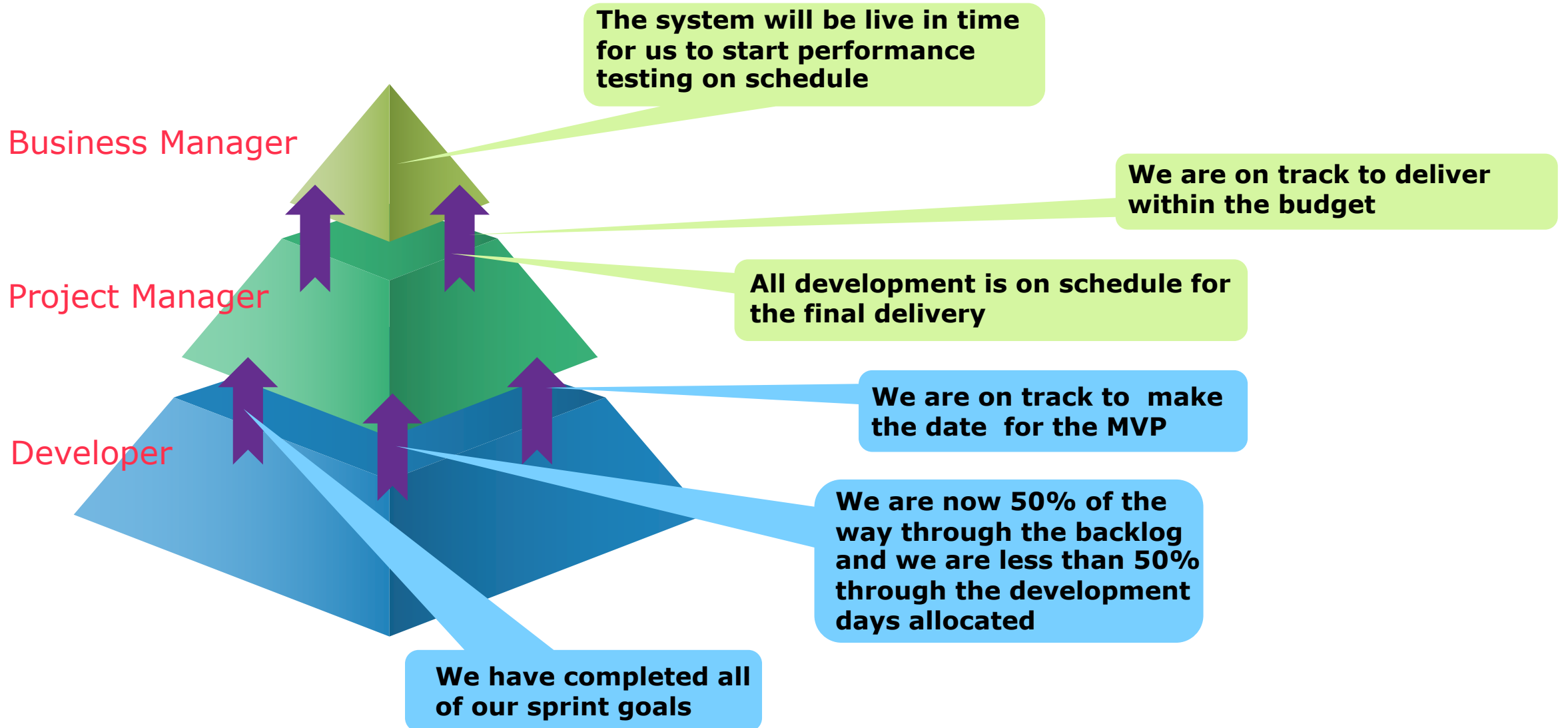
Pathological (power-oriented)	Bureaucratic (rule-oriented)	Generative (performance-oriented)
Low cooperation	Modest cooperation	High Cooperation
Messengers shot	Messengers neglected	Messengers trained
Responsibilities shirked	Narrow responsibilities	Risks are shared
Bridging discouraged	Bridging tolerated	Bridging encouraged
Failure leads to scapegoating	Failure leads to justice	Failure leads to inquiry
Novelty Crushed	Novelty leads to problems	Novelty implemented



Stakeholder metrics

-  Will it help us become top performers?
-  Will it improve customer satisfaction?
-  Will we get our invoices paid quicker?
-  What is the return on investment?
-  Will the new system increase sales?

Managing upwards – give them the relevant information



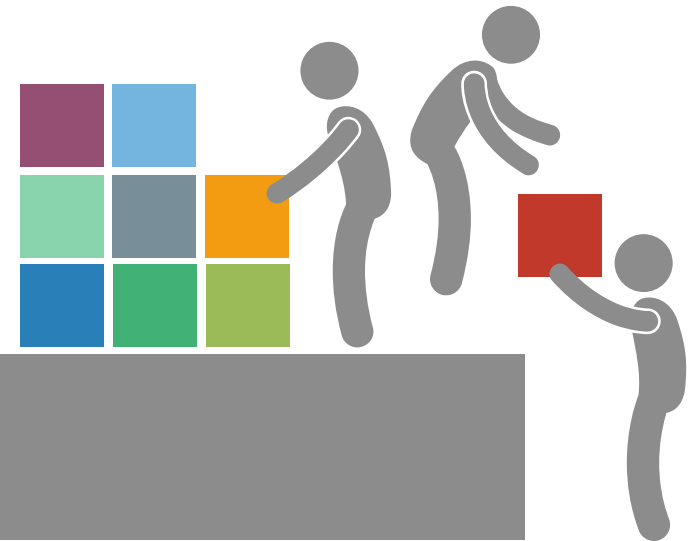


Finally...



What to discuss with your boss

- Apply reasonableness tests
- Recognise great teamwork
- People are not as motivated by money as business managers might think they are
- Unmeasurables - loyalty, team spirit, honesty, helpfulness, dedication...





To summarise

- Story points *DO* have their place – but they indicate **capacity** not **productivity** and they're unique to a group of people
- Optimising for a single metric is likely to have unintended consequences!
- **Metrics != Targets**
- Aim for a combination of **balanced** metrics
- Only report metrics that are **meaningful** between layers
- **Working software** is the **primary measure** of progress



Q&A



Reference material



Reference material

- Motivation to Work by Frederick Herzberg (ISBN - 978-1560006343)
- Money is not the best Motivator: <https://www.forbes.com/2010/04/06/money-motivation-pay-leadership-managing-employees.html#79e532533e13>
- The War Managers by Douglas Kinnard (ISBN - 978-1591144373)
- Kill anything that moves by Nick Turse (ISBN - 978-1250045065)
- The Tyranny of Metrics by Jerry Z. Muller (ISBN - 978-0691174952) and <https://www.c-span.org/video/?441850-2/the-tyranny-metrics>
- The concept of “Why” by Simon Sinek; TED talks (<https://www.youtube.com/watch?v=7zFeuSagktM>)
- Building Less Flawed Metrics by David Manheim: https://mpra.ub.uni-muenchen.de/90649/1/MPRA_paper_90649.pdf
- Wells Fargo Banking Scandal - <https://www.theguardian.com/business/us-money-blog/2016/oct/07/wells-fargo-banking-scandal-financial-crisis>



Reference material

- <https://devops-research.com/>
- Accelerate: The Science of Lean Software & DevOps - Nicole Forsgren, Jez Humble, Gene Kim
- Ron Westrum
- Mulder, P. (2017). *Job Demand Control Model by Robert Karasek*. ToolsHero: <https://www.toolshero.com/stress-management/job-demand-control-model/>
- Demand/Control Model: a Social, Emotional, and Physiological Approach to Stress Risk and Active Behaviour - Robert Karasek <http://www.iloencyclopaedia.org/component/k2/item/12-psychosocial-factors-stress-and-health>
- George M . Mark – The relationship between workplace stress and job characteristics, individual differences and mental health <https://orca.cf.ac.uk/54730/1/U585119.pdf>
- <https://continuousdelivery.com/implementing/culture/>
- <https://www.andykelk.net/devops/using-the-westrum-typology-to-measure-culture>
- <https://www.thoughtworks.com/radar/techniques/four-key-metrics>