A Case Study

London, June 2010

C. C. Shelley
OXFORD SOFTWARE ENGINEERING Ltd
9 Spinners Court, 53 West End,
Witney,
Oxfordshire
OX28 1NH
www.osel.co.uk
shelley@osel.netkonect.co.uk
Tel. +44 (0) 1993 700878

Contents

- **Background**
- **Story**
- **Results**
- **Next Steps – more value**

Background

- ***bwin*** **is a s/w company providing online services**
  - **multi-user (250,000 concurrent users)**
  - **multi- transactional international (20 languages) real-time financial transactions (over 1 million transactions per day)**
  - **require secure implementation**
  - **require rapid response (to changing commercial opportunities and legal environment)**

- **Two sites Vienna and Stockholm**
  - **Different cultures (Vienna acquired Stockholm organization)**
  - **Dev. approach is super agile (Scrummish, with interrupts to sprints allowed)**

The Improvement Story: 1

The need:

- **Organization wanted a novel agile capability:**
  - **to co-ordinate delivery of system components produced by delivery teams…**
  - **…to deliver major system changes rapidly…**
  - **…teams to be dependable 'delivery units' working together to a common 'rhythm'.**

- **Delivery teams were (and are) dependent on each other:**
  - **there were delays**
  - **it was unclear what was holding up delivery 'upstream'**
  - **team activity was opaque to other teams and business – inter team and business/team 'friction'**
  - **cascade further delays**

The Improvement Story: 2

- **Performance characteristics and objectives set (in priority order)**

1. **higher delivery predictability**
2. **increased responsiveness**
3. high frequency and regularity of delivery
4. while sustaining systems quality

The Improvement Story: 3

- **Specifically:**

  - **Predictability**
    - **When: 90% of all teams to deliver within 7 days of planned delivery**
    - **What: 90% of teams to deliver 90% of storypoints* committed to**

  - **Responsiveness**
    - **Increased ability to respond rapidly by committing (mid sprint) to deliver**
      - (storypoints delivered that were requested mid sprint)

  - **Frequency and Regularity**
    - **Enabled by predictability**
    - **Establishing a dependable organization wide delivery 'rhythm'**

  - **Don't compromise quality**
    - **As judged empirically from very rapid response from very many users**
    - **(what is quality? – they recognize it when they see it)**

The Improvement Story: 4

How?

- **Require a single process shared across all teams**
  - **…a 'standard' and 'defined' process (no tailoring)…**
  - **…that is understood by all…**
  - **…making a sprint's status, including emerging issues, at any time, visible and understandable to all**

- **Using 'rule 11'**
  - ***"Document processes, practices and activities 'as is' and use this to establish a credible process baseline which can then be developed. (Do not - ever - document processes as you would wish them to be and attempt to introduce them, or roll them out.)"***

- **Third time lucky (this wasn't easy)**

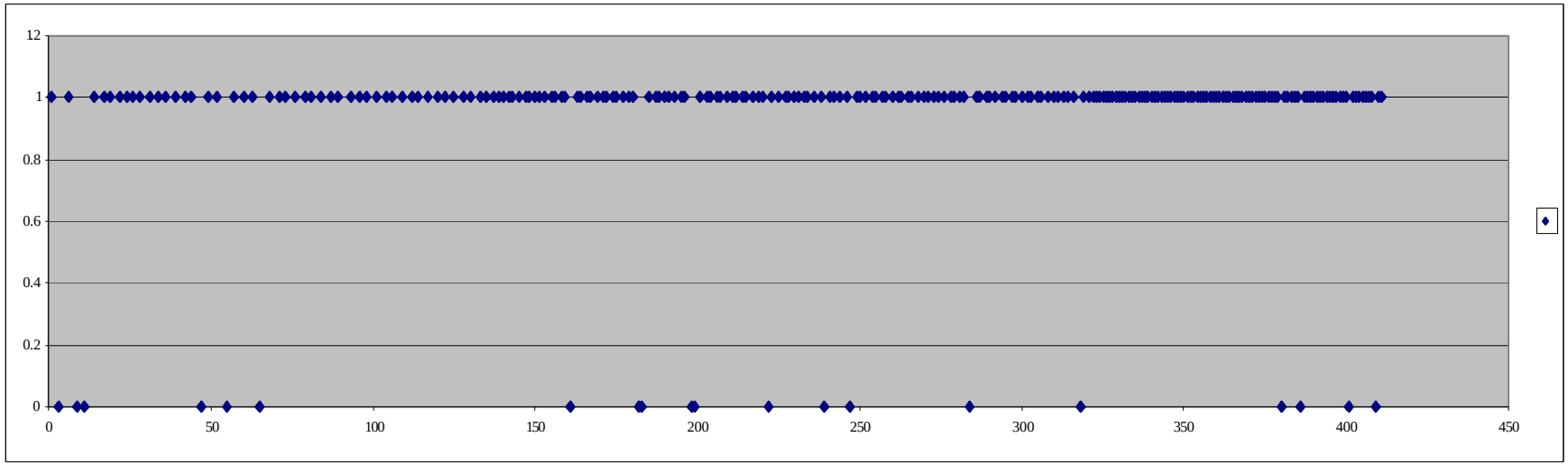
- **In effect introducing 'process discipline' to an agile environment**
  - **cf Systematic Software Engineering - introducing agile to a CMMI ML5 organization reported in *'Scrum and CMMI Level 5: The Magic Potion for Code Warriors'***

Results: 1

- **Data collected from mid 2007 to end of 2009**
- **6 teams using defined process in 2007, 10 in 2008, 16 (all teams) in 2009**
- **Analysis performed using day-to-day operational data (as it should be)**
- **Some additional data wanted by analyst, but not available**
  - **(too bad!)**
- **No filtering or stratification of data performed, outliers retained, and analyses kept as simple as possible (with one, exploratory exception)**

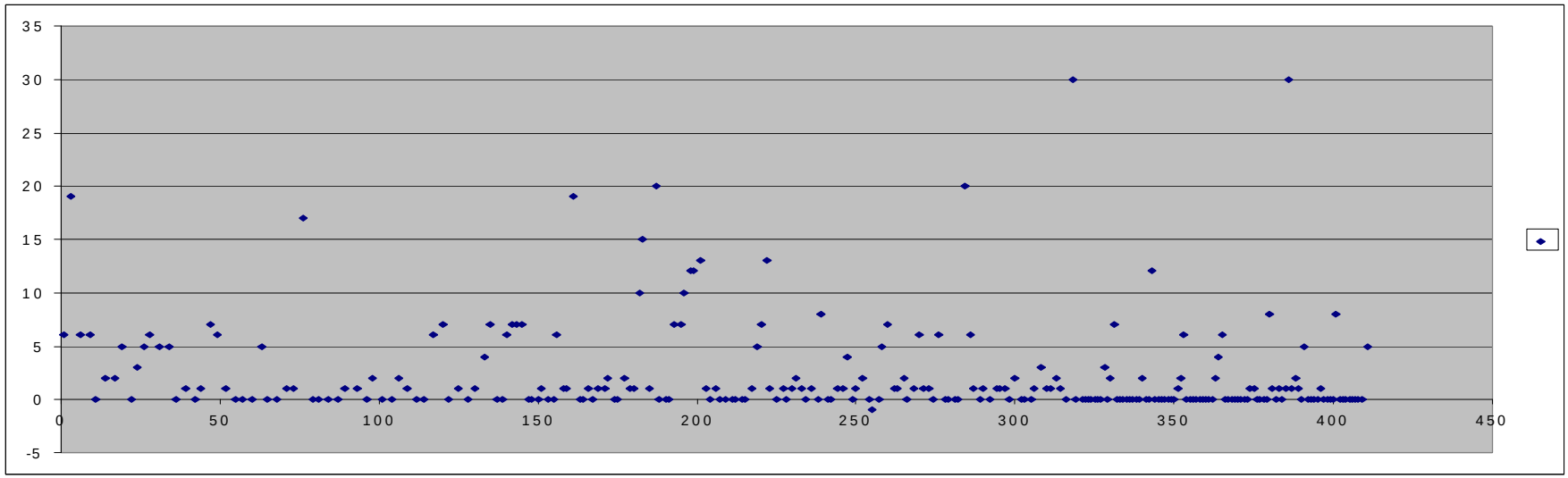
Results: 2

- **Schedule predictability (1) KPI :**
  - **‘1’ indicates delivery within 7 days of planned, 0, greater than 7 days**
  - **Note increasing density of ‘1’ data points with consistent density of ‘0’ points**



Results: 3

- **Schedule predictability (2) Days deviation:**
  - **Plotted with actual days deviation**
  - **Note number of deliveries with zero deviation towards and of reporting period**
  - **(performance better than indicated by KPI)**



Results: 4

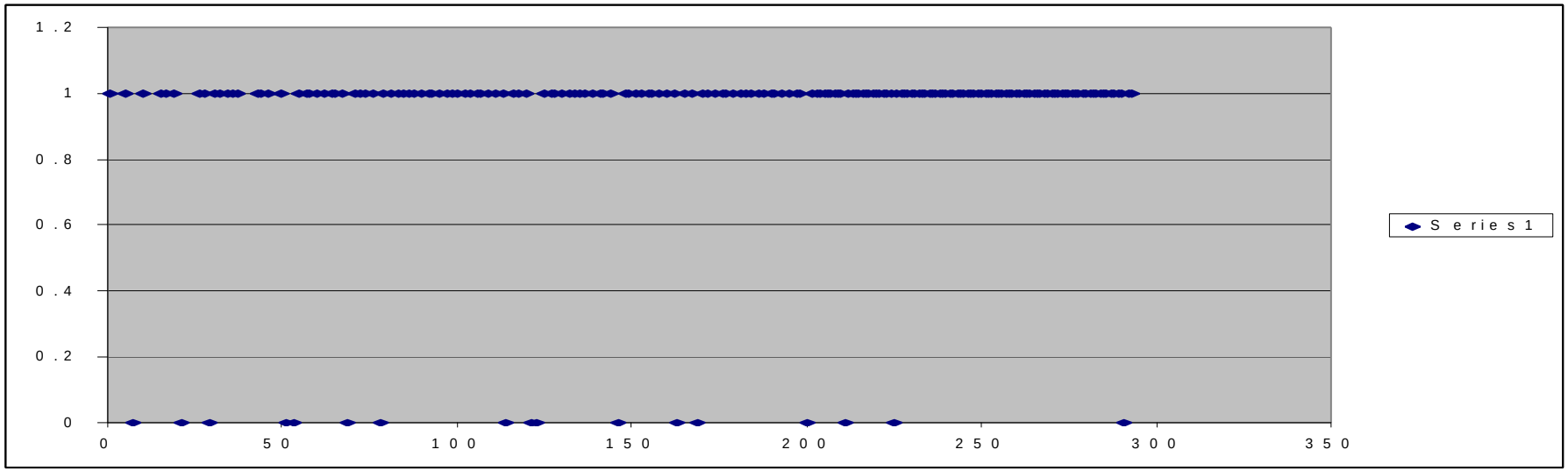
If the number of days delay for releases in the first four months are considered then the average delay per release is 4.25 (85 days delay/20 releases).

For the last four months the average is 1.45 days per release (61/42), and if the exceptional delay of 30 days, for SP 09-06 FRB 3.25, is removed this reduced to 0.74 days per release.

The ratio between the first four and last four months is a 2.93 times improvement in schedule predictability, or, excluding a SP 09-06 FRB 3.25, ***a 5.74 increase in schedule predictability***

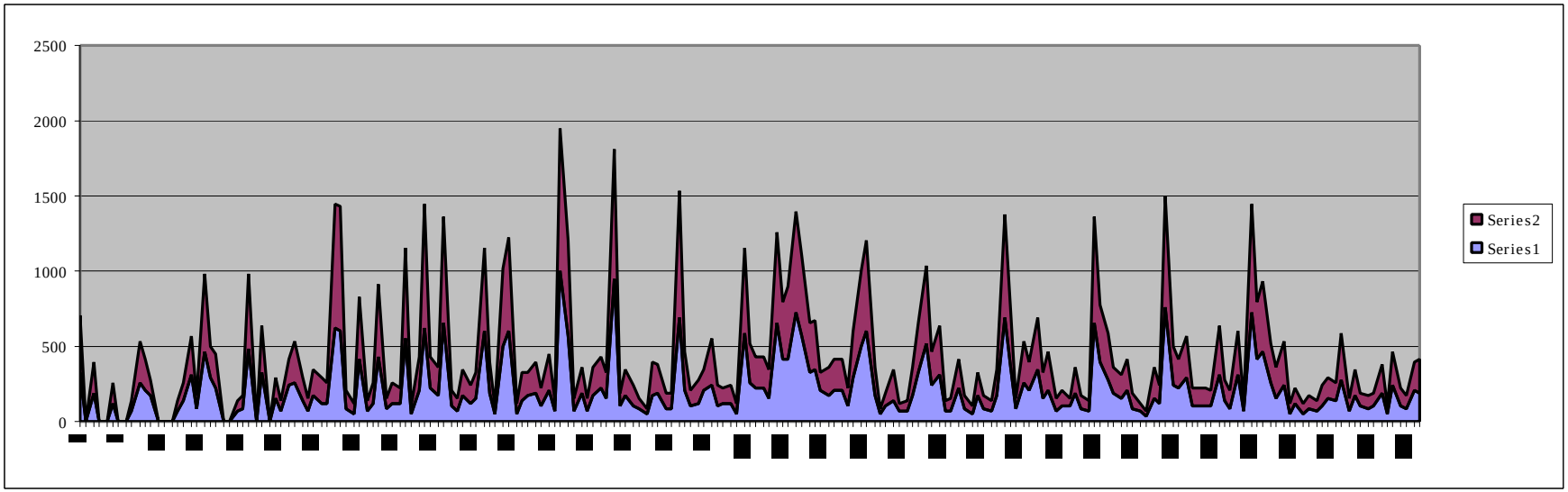
Results: 5

- **Scope predictability (1): KPI**
  - **‘1’ indicates 90% or more of committed storypoints delivered, ‘0’, less**
  - **Note the increasing density of ‘1’ over time, and decreasing number of ‘0’s**
  - **(NB – period from beginning 2008 to end 2009)o**



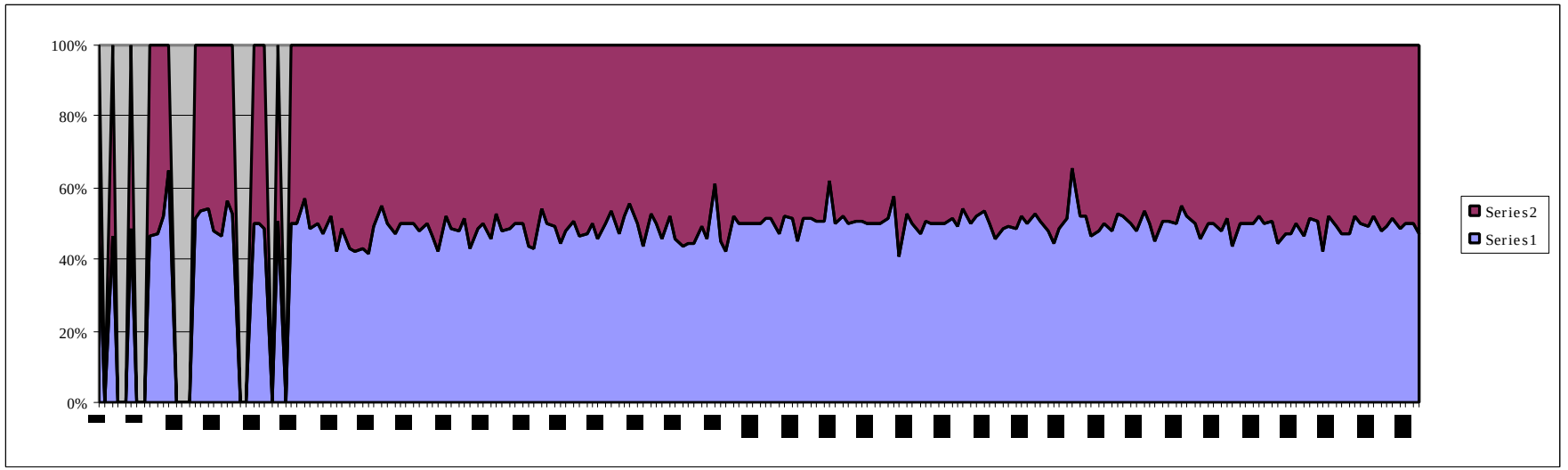
Results: 6

- **Scope predictability (2) Storypoints:**
  - **Shows the numbers of committed and delivered storypoints**
  - **Storypoints are calibrated locally (in teams)**
  - **Difficult to discern any pattern or trend**



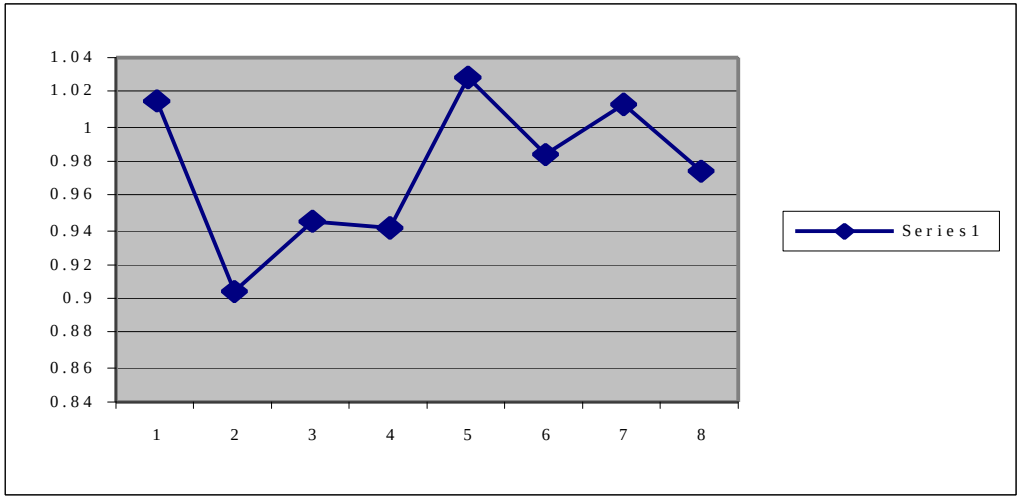
Results: 7

- **Scope predictability (3) Storypoints:**
  - **Plot the ratio of committed/delivered storypoints (normalize!!)**
  - **Still not much of a pattern, although it looks like delivery is more or less (literally) what was committed**



Results: 8

- **Scope predictability (4) Storypoints:**
  - **Plot the average of data points 'binned' in three monthly periods**
  - **Pattern emerges, but with odd first data point**



Results: 9

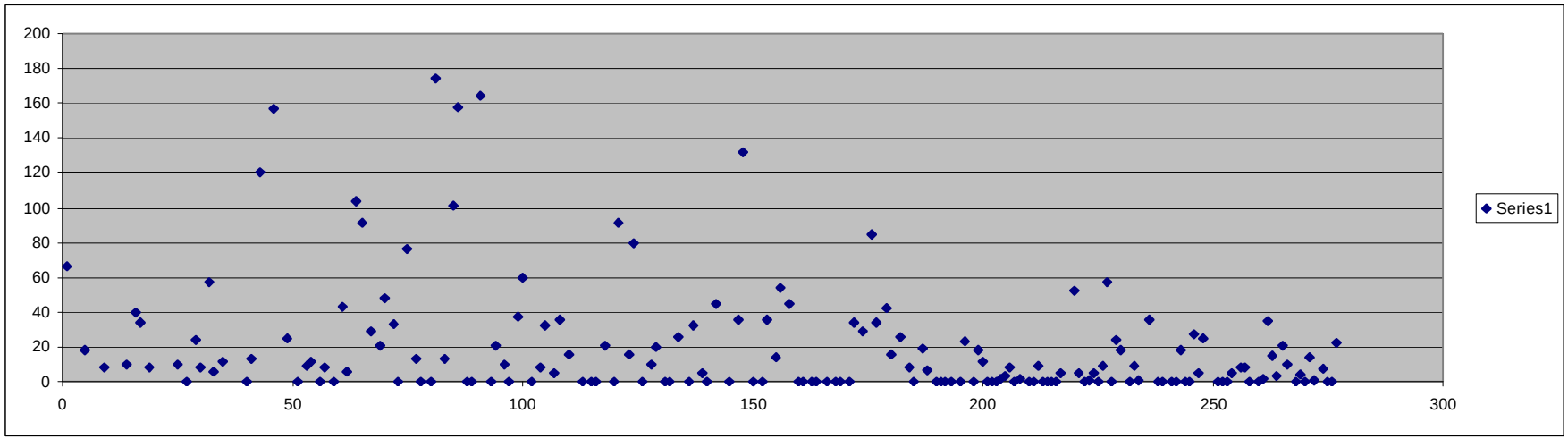
In the first four months for which we have data (January – April 2008) 14 of 20 releases achieved the 90% scope target (with 6 failing and no data for 7 others recorded.)

For the last four months for which data is available (May to August 2009) 39 of 40 releases succeeded.

This gives a ratio of about 12:1 - a twelve times improvement in scope predictability.

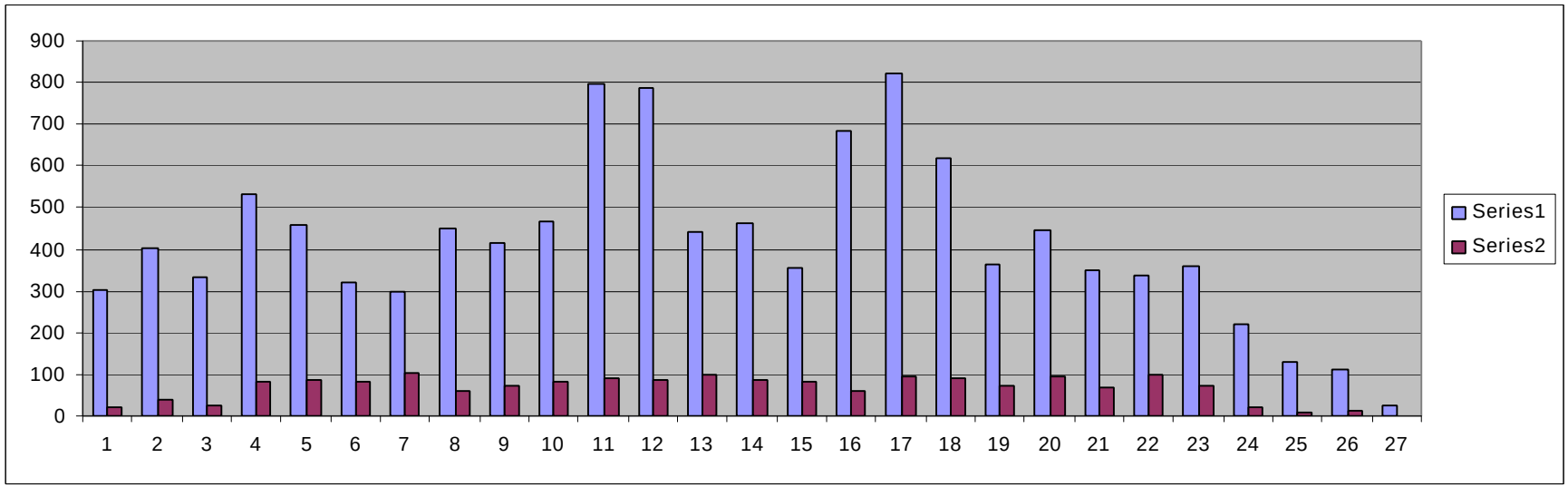
Results: 10

- **Responsiveness**
  - **Evaluated by the number of storypoints committed to in mid sprint and delivered**
  - **Number *declines* of this period**
  - **Interpreted as evidence of increasing use of a well understood, predictable and trusted process by teams and business, and only interrupted by genuine business need**



Results: 11

- **Quality (1);**
  - **Defects from live operation (red bars) as *proxy* measure of quality**
  - **(only data available for all teams over entire period)**
  - **Indicates consistent level of quality over 2007 – 2009 period**



Results: 12

- **Defect reporting levels suggested high quality**
- **How to establish level of quality?**
- **Defect densities have been used:**

Released S/W	Defect Density
NASA	0.1 defects / KLOC
Leading Edge	0.2 defects / KLOC
Critical systems	1.4 defects / KLOC
Military	5 – 55 defects / KLOC

Results: 13

- **What is bwin defect density?**

- **A defect model:**
  1. **bwin's software is heavily used and most defects will be found quickly.**
  2. **If bwin did not make changes to these systems, but simply fixed defects as they were found through day to day use the number of defects being found would decrease over a short period of time – say six months.**
  3. **At present bwin *does* make changes and the number of defects has not decreased over time.**
  4. **Therefore the defects being found (about 60 – 100 per month) are due to changes being made.**
  5. **Therefore, if the number of changed lines of code can be estimated (extracted from the SCM system) the defect density of these changed lines of code can be estimated.**

– **(Note: If this model is reasonable it also suggests that software quality could be improved quickly and easily by introducing a carefully selected set of quality controls to changed software.)**

Results: 14

- **One third of the defects reported from live operations are high priority (about 500 of about 1800). With live defects being reported at about 60 to 100 per month this means about one high priority defect being reported per day.**

- ***Bwin* software assets were about 400 KLOC (thousands of net, executable lines of code) in December 2007. This had increase to about 600 KLOC in December 2008 and to about 1 MLOC in September 2009.**

- **This increase in LOC, together with consistent defect reporting levels suggests that, in practice, the quality of bwin software is improving.**

- **In addition, if it is assumed that defects reported are mostly a result of new or changed code then the level of defects reported implies exceptional levels of code quality:**

- **200 KLOC of new code (plus changed lines of code) in 2008 gave rise to about 1000 defect reports that year, of which about 400 were high priority. This gives a defect density of about 5 defects/KLOC.**

- **Note: The data indicate that quality has been sustained (and improved) over this period and has not been 'traded off' against improved delivery predictability.**

Next Steps: 1

A number of candidate 'process improvements' identified:

1. **develop SPI infrastructure (SPG group)**
2. **analyse and reduce numbers of defects**
3. **develop technical reviews**
4. **'balance' estimation errors**
5. **refine understanding of software size**
6. **quantify system attributes**
7. **investigate CMMI generics for useful ideas**
8. **revive retrospectives**
9. **introduce value engineering into product and sprint backlog management**

- **1- 8 may improve efficiency of delivery (they are bounded), 9 will improve effectiveness (and is unbounded), and is therefore interesting**

Next Steps: 2a

Number 9

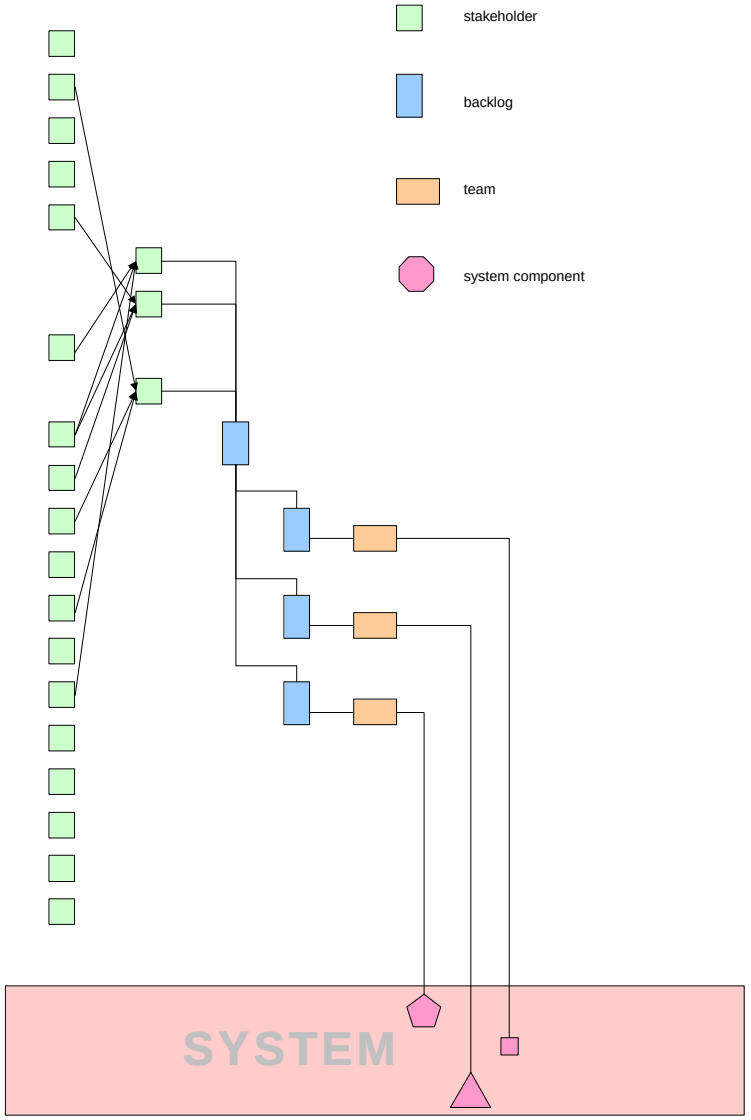
- **Situation:**
  - **Ongoing, un-interruptible stream of 'requirements'/stories*…**
  - **…to product backlog, then assigned to teams' sprint backlog**
  - **Want to enable exploration of business context to find the value opportunities**
    - **('feeling lucky' button costs $100M/year ***
    - **Where are our 'feeling lucky' buttons? (one found))**
  - **What requirements are missing (missed opportunities)?**
  - **Cannot/will not apply heavy weight techniques to each requirement**
    - **No bottlenecks allowed**

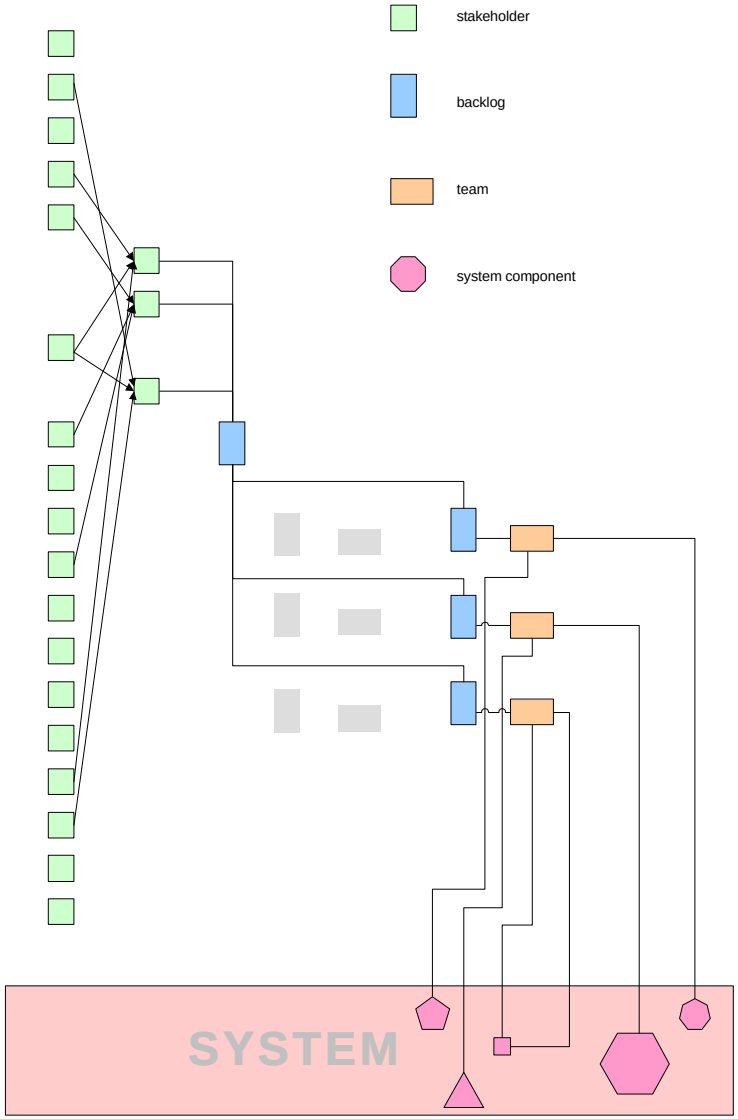
(* and see www.uie.com - $300M button)

Next Steps: 2b

Number 9

- **Stakeholder/Perspective types:**
  - **other (dependant) teams**
  - **team**
  - **product manager (s)**
  - **customers (system purchasers)**
  - **marketing – (image/look and feel)**
  - **multiple national law makers**
  - **multiple and hybrid user types – sporting, gaming, info, services**
  - **financial services providers**
  - **Society (societal impact, e.g. addiction monitoring)**





Next Steps: 3

- **Objective:**
  - **"To double the value delivered by delivery teams within the next two years."**

- **How?**

  - **Two phases:**

    1. **Apply rule 11 – document 'as is' backlog management process**

    2. **Refine this to enable :**
       - evaluation of requirements/stories value*
       - (begin by establishing a value delivery baseline)
       - then filter out low value requirements
       - and discover and promote high value opportunities
       - managing 'value delivery'
       - (and minimize production of code? – do *not* try to improve 'productivity')

Next Steps: 4

- **Want to:**
  - **Distinguish stakeholder needs (/customer requirements) from system requirements (this will be unattractive when existing systems are in place – people like to focus on the tangible) – 'two stage', multiple entry process (standard CMMI wisdom)**
  - **Introduce or refine 'evaluation' tools**
  - **Introduce 'discovery tools' (to reveal missed reqs)**
  - **Introduce 'explorer' tools (to search 'reqs space')**
  - **Introduce 'opportunity hunter' tools (opposite of discovery tool)**
  - **Free up 'prioritization' flag or introduce a 'value index'**
  - **Map 'value' to business units (Euro)**

Next Steps: 5a

- **Value is:**
  - **is an extrinsic property evaluated/judged by the stakeholders**
  - **is a mix of (un)desirability and (dis)utility**
  - **comprises a ~~spectrum~~ collection of concepts and attributes**
  - **so comparisons of value have several dimensions/aspects**
  - **that may not be (are probably not) orthogonal * ****

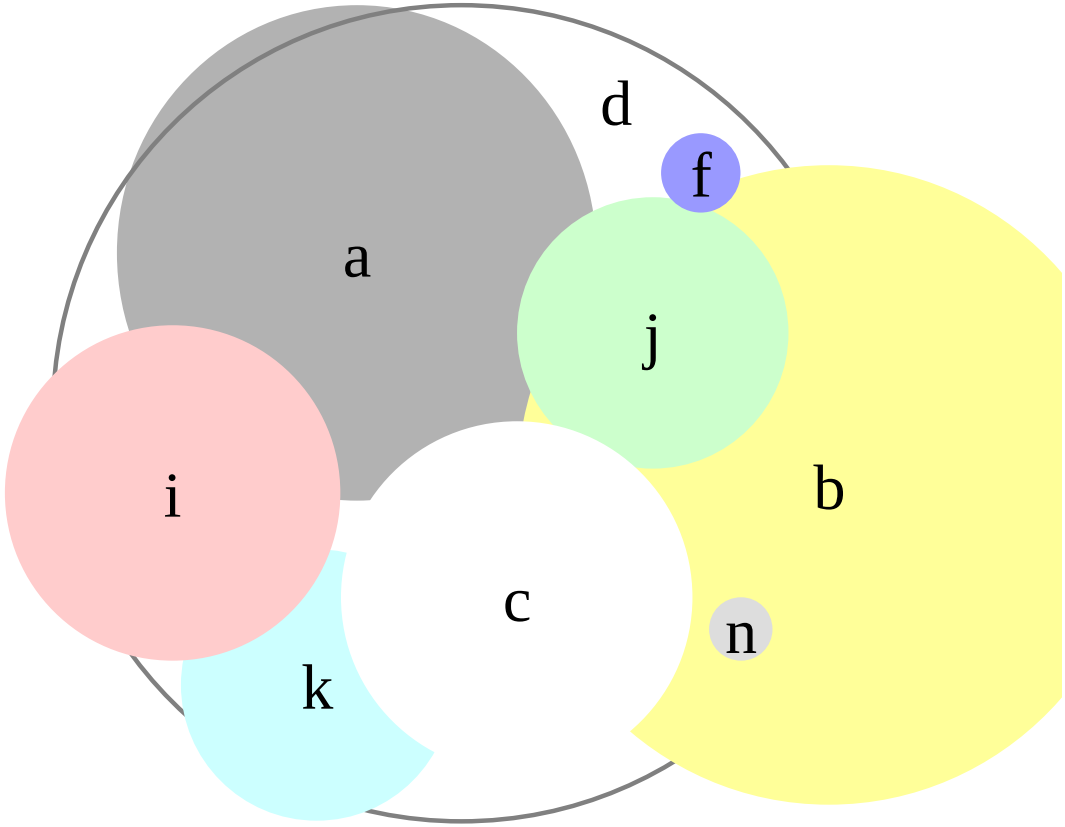
*** Multi-dimensional Trustworthiness Measure for Software*
HingWei Tao & Yixiang Chen

Next Steps: 5b

- **Value has:**
  - **many stakeholders with different perspectives**
    - **with different weights**
  - **each has an interest and weighting of several attributes**
  - **Which may be**
    - **incomplete,**
    - **overlapping,**
    - **conflicting,**
    - **implicit,**
    - **latent,**
    - **unknown.**
  - **This is complex**

Next Steps: 6

- **Value comprises:**
  - **attribute <a>, it contributes 0.333**
  - **attribute <b>, it contributes 0.333**
  - **attribute <c>, it contributes - 0.25**
  - **attribute <d>, it contributes 0.0**
  - **Attribute <e>, it contributes 0.4**
  - **....**
  - **attribute <n>, it contributes <x>**
  - **....**



Next Steps: 7

- **I'm looking for ideas, and looking at experts, so…**

…do you have any views, or know of candidate approaches that we can investigate?

(visions, strategies, methods, techniques, tools and hints all welcome)

Next Steps: 8

- **Need a mix of tools to manage complexity, provide oversight and enable insight…**

- **Candidates ideas components (in no particular order):**

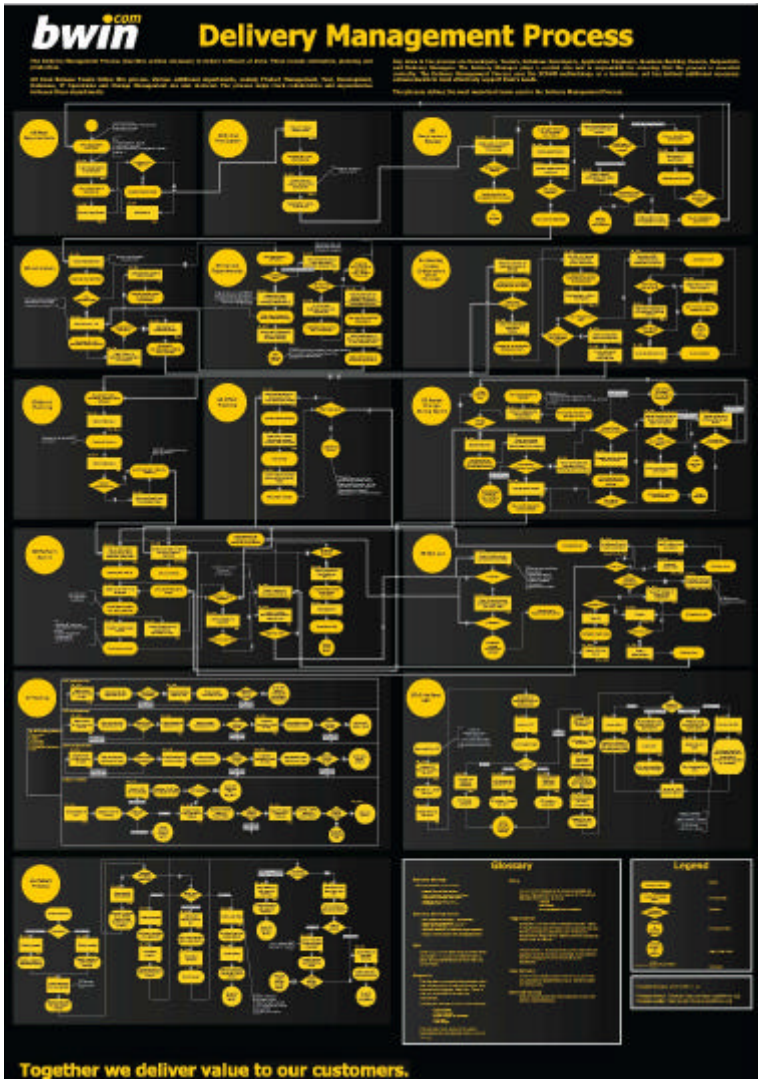
  - **Benjamin Franklin's 'moral or prudential algebra'**
  - **80:20 / Pareto**
  - **game theory**
  - **Alexander's 'discovery' methods to find easy to miss opportunities**
  - **Kano**
  - **Ad hoc grids and matrices (to show absence of information – which is information)**
  - **Not so ad hoc matrices - impact estimation/QFD**
  - **a logarithmic value index (like a Richter scale, as supplement or replacement of ordinal prioritization flags)**

- **But still no suggestion of how to map value**

OXFORD SOFTWARE ENGINEERING LIMITED

9 Spinners Court, 53 West End,
Witney,
Oxfordshire
OX28 1NH

www.osel.co.uk
shelley@osel.netkonect.co.uk
Tel. +44 (0) 1993 700878


bwin.com
Delivery Management Process
Glossary
Legend
Together we deliver value to our customers.

I cannot, for want of sufficient premises, advise you on *what* to determine, but if you please I will tell you *how* ... My way is to divide half a sheet of paper by a line into two columns; writing over one *Pro* and over the other *Con*. Then, doing three or four days' consideration, I put down under the different heads short hints of the different motives, that at different times occur to me *for* or *against* the measure. When I have thus got them all together in one view I endeavour to estimate the respective weights... [to] find at length where the balance lies... And, though the weight of reason cannot be taken with the precision of algebraic quantities, yet, when each is thus considered, separately and comparatively, and the whole matter lies before me, I think I can judge better, and am less liable to make a rash step: and in fact I have found great advantage for this kind of equation, in what may be called *moral* or *prudential* algebra

In a letter from Benjamin Franklin to Joseph Priestly