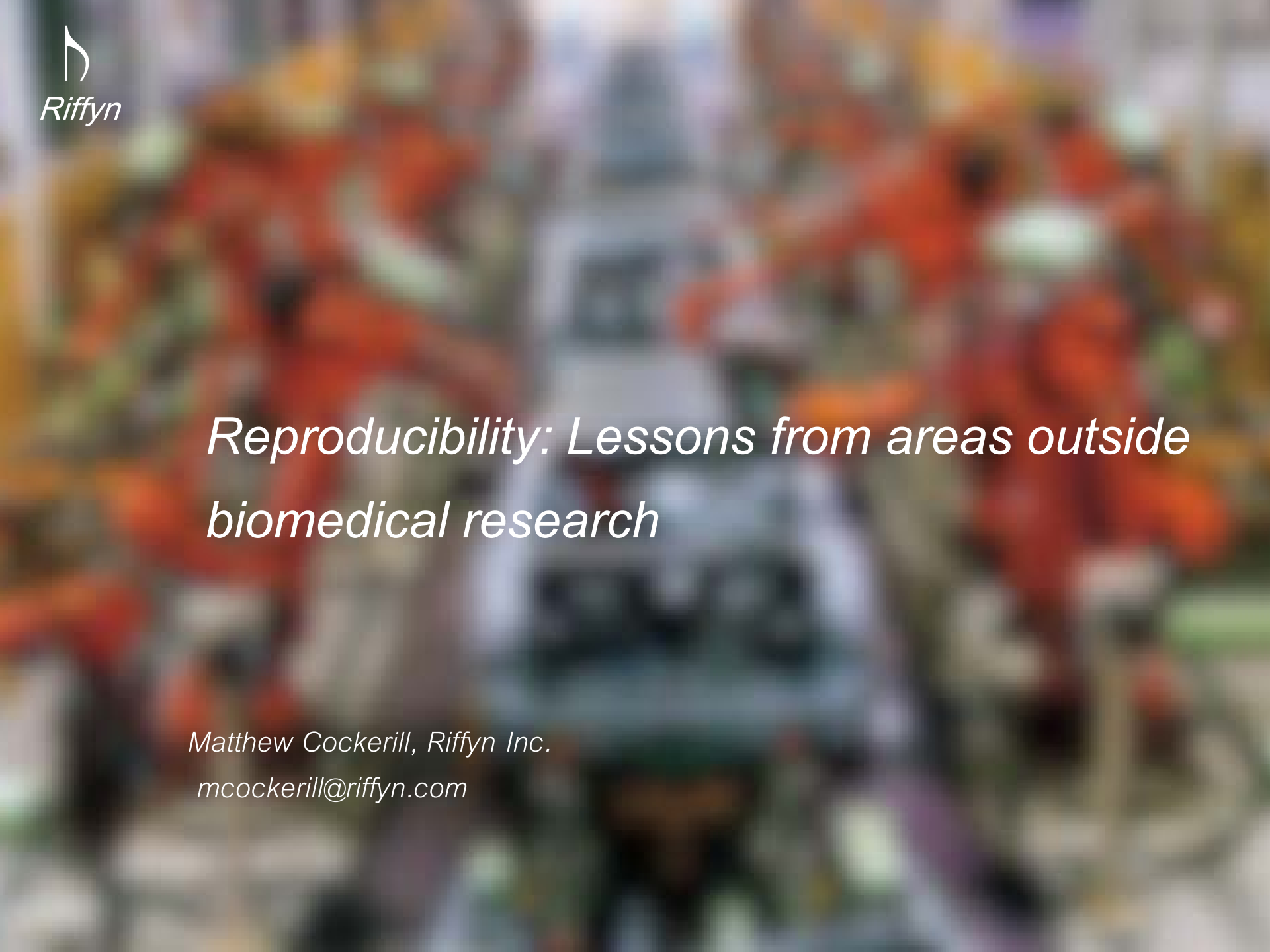


Reproducibility: Lessons from areas outside biomedical research

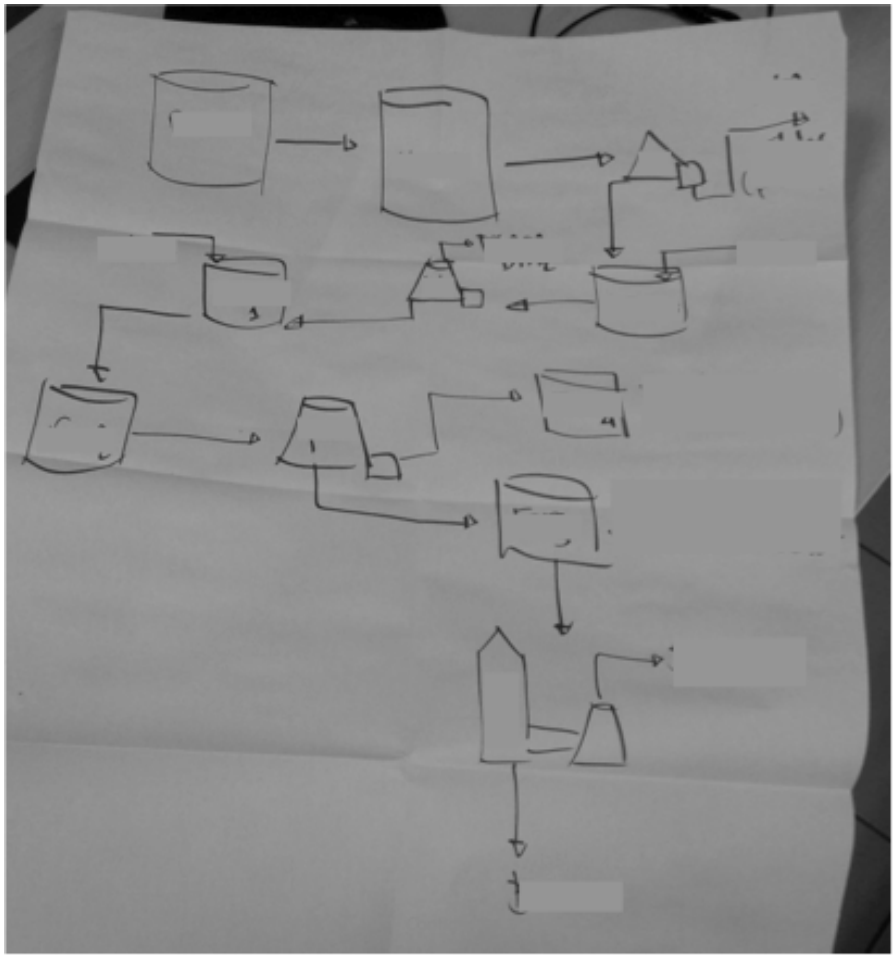
Matthew Cockerill, Riffyn Inc.
mcockerill@riffyn.com

The underlying causes of poor reproducibility

- Biomedical research has traditionally taken a “blacksmith shop” approach to methodology
- Craftsmanship remains key
- Essential methodological knowledge is implicit and undocumented
- To get something to work, typical advice is “go and spend time in the relevant lab”

Not just an academic problem…

Below: Process design sent to manufacturing

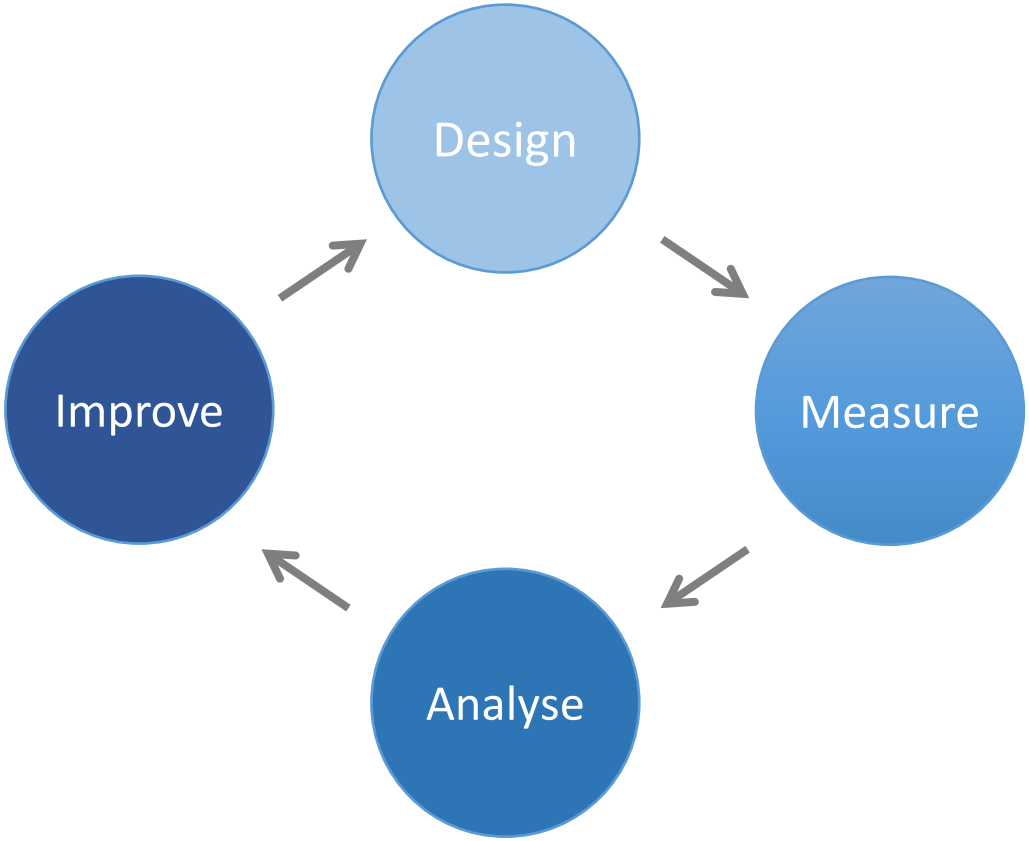


When we first transferred our manufacturing process to full-scale operations, our yields dropped 90%.
- CEO of a diagnostics company

We lost four months tracking-down a problem at the demo scale. Root cause was a parameter change that a new employee thought did not matter.
- VP Process Development of an algae company

Certainly it is a problem that we don't write down our processes - people change what they know, not what necessarily matters.
- Head of Process Development of a biopharma company

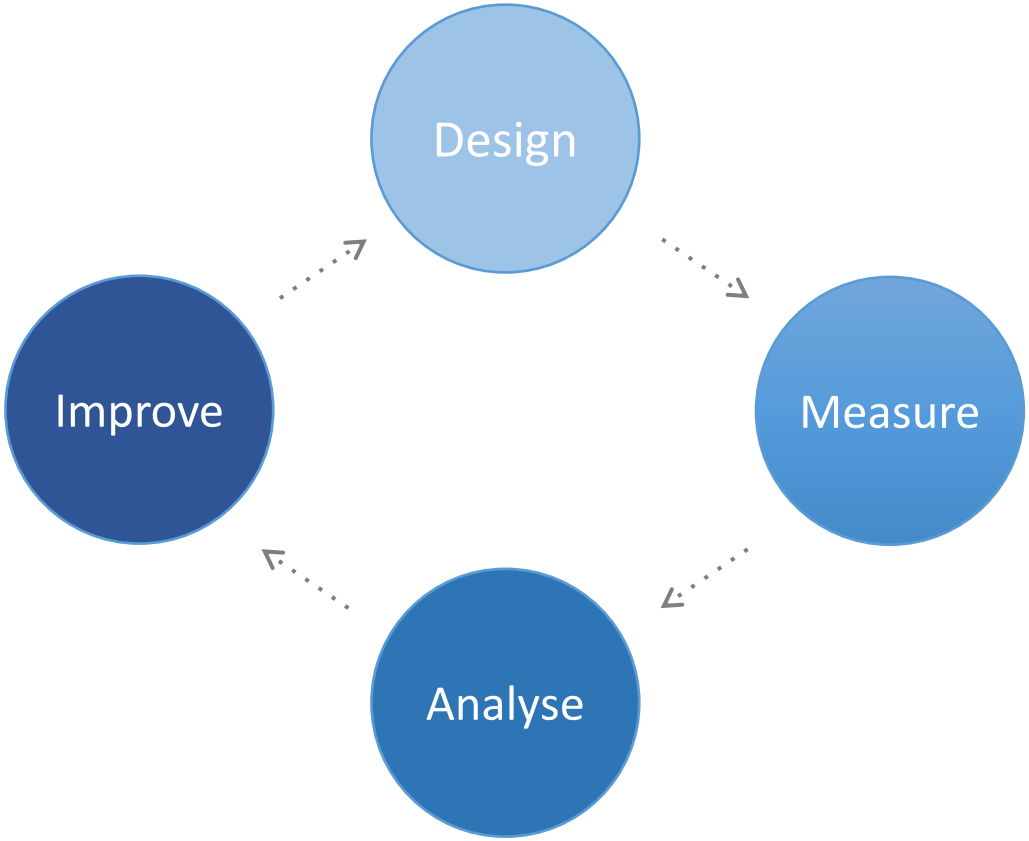
The problem of noise



Process improvement cycle

The problem of noise

Lack of reproducibility creates a fog of noise which slows progress



Process improvement cycle

Consequences of lack of reproducibility

In academia

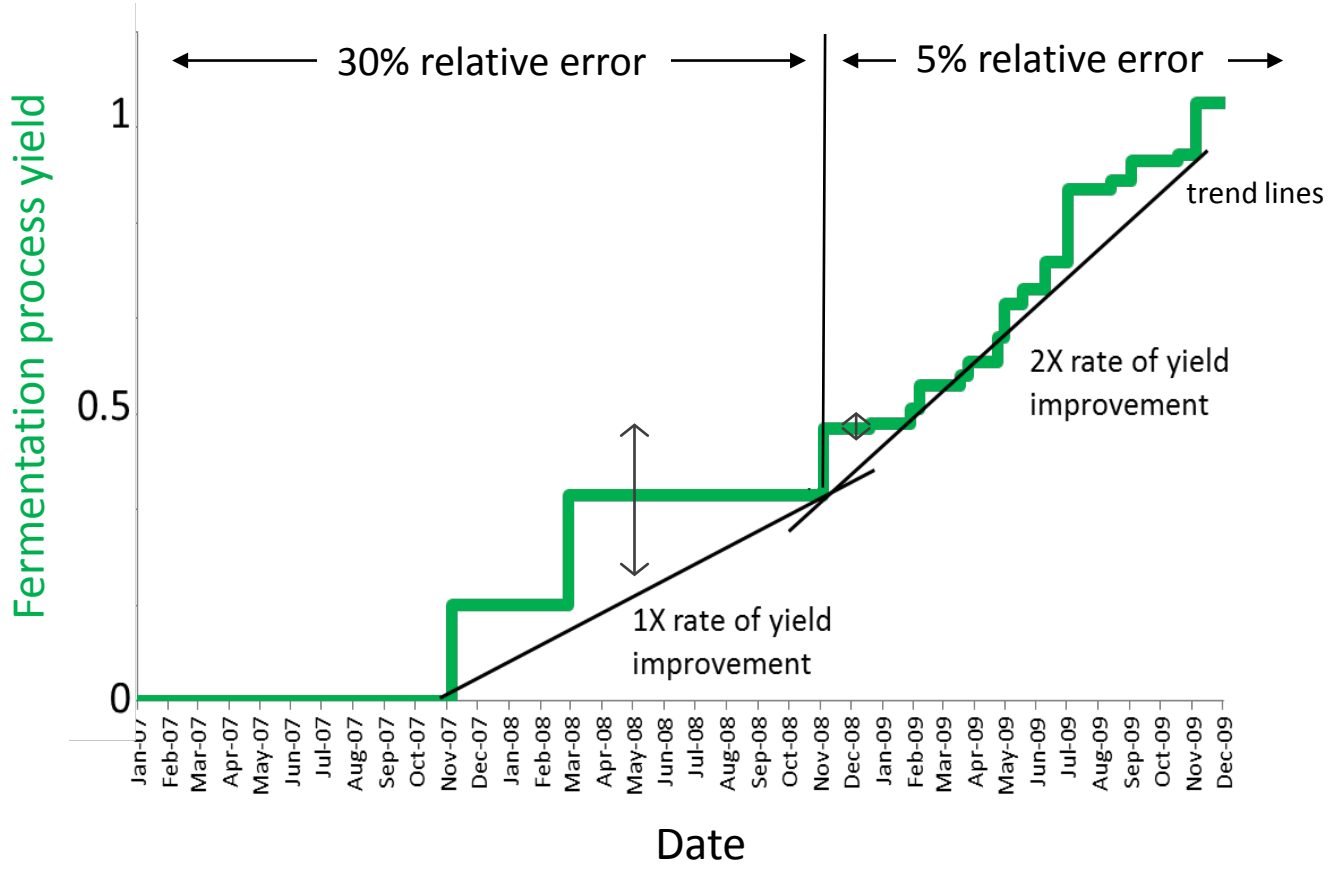
- Published results do not provide a solid foundation for future research, leading to wasted public research funding

In industry

- Tech transfer is slow and failure-prone
  - Academia -> Industry
  - Between business units and in wider collaborations
- Companies get low return on their R&D investment

Real-world biotech example

Quality improvement cuts error by 6X. R&D productivity doubles overnight.



Reproduced from Gardner TS, Trends in Biotechnology, March 2013, Vol. 31, No. 3
Data from product development at leading biotech firm

Is there a technical solution?

Reduce/eliminate noise, increase reproducibility by:

- Unambiguous process specification
- Tight specification of inputs
- Capture of all relevant data streams and experimental parameters
- Correlation/root cause analysis using process specification to make sense of data over time

Is there a technical solution?

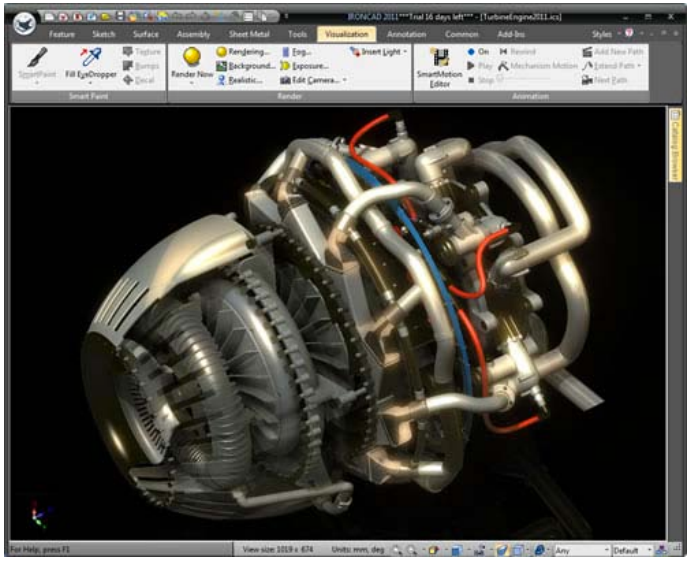
Reduce/eliminate noise, increase reproducibility by:

- Unambiguous process specification
- Tight specification of inputs
- Capture of all relevant data streams and experimental parameters
- Correlation/root cause analysis using process specification to make sense of data over time
- **Is this feasible in practice?**

Other fields have managed it!

Parallels from manufacturing

- **CAD/CAM**
  Computer Aided Design **(**CAD) provides unambiguous digital specification of a design. Combined with reproducible computer-aided manufacturing (CAM) process, facilitates reliable iterative improvement cycle for components

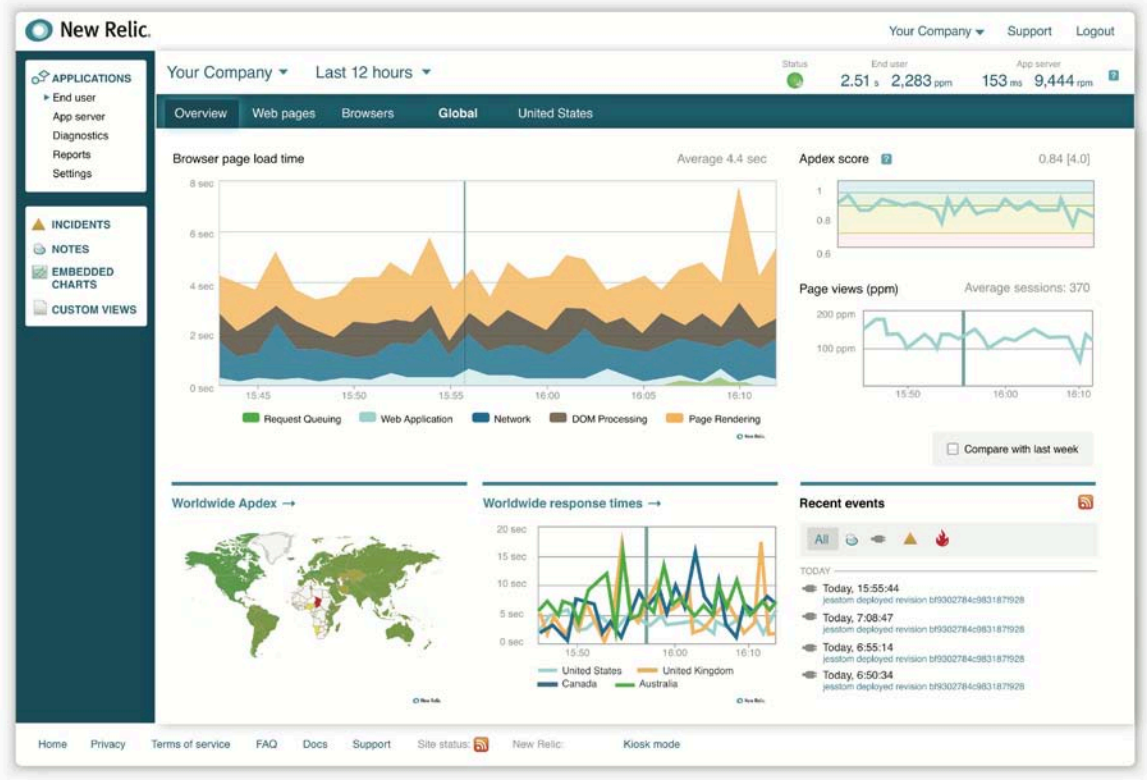


- **Auto assembly**
  Toyota Production System. Manufacturing process fully specified, and instrumented so that analytics can determine potential root causes (e.g. a new batch of rivets) which correlate with downstream problems (mysterious rattles)

Parallels from tech

Application Performance Management

Tools such as *New Relic* and *AppDynamics* instrument, capture and analyse performance data from software applications. Problems can be rapidly tracked down to their source.



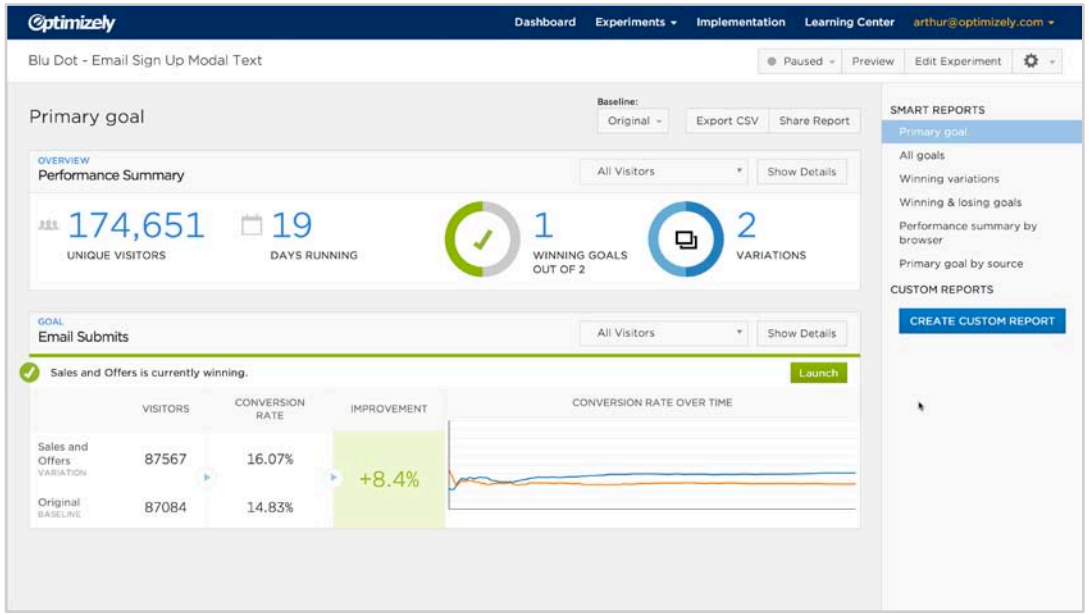
Parallels from tech…

DevOps

Sysadmin work is increasingly replaced by configuration management systems based on ‘recipes’ which reproducibly deliver a fully configured server which behaves predictably

Parallels from online marketing

- **Marketing analytics**
  Explicit upfront design of tests, with extensive instrumentation to capture results, leads to more efficient website optimisation



How does it work in manufacturing?

Design

- Everything starts with CAD files
- Manage complexity via modularity

Automated measurement

- Specs are measured & met
- Data acquisition is automated

Continuous improvement

- Continuous analytics & process improvement
- Absolute reproducibility

Everyone participates

- Make quality transparent and personal so everyone in organization is involved

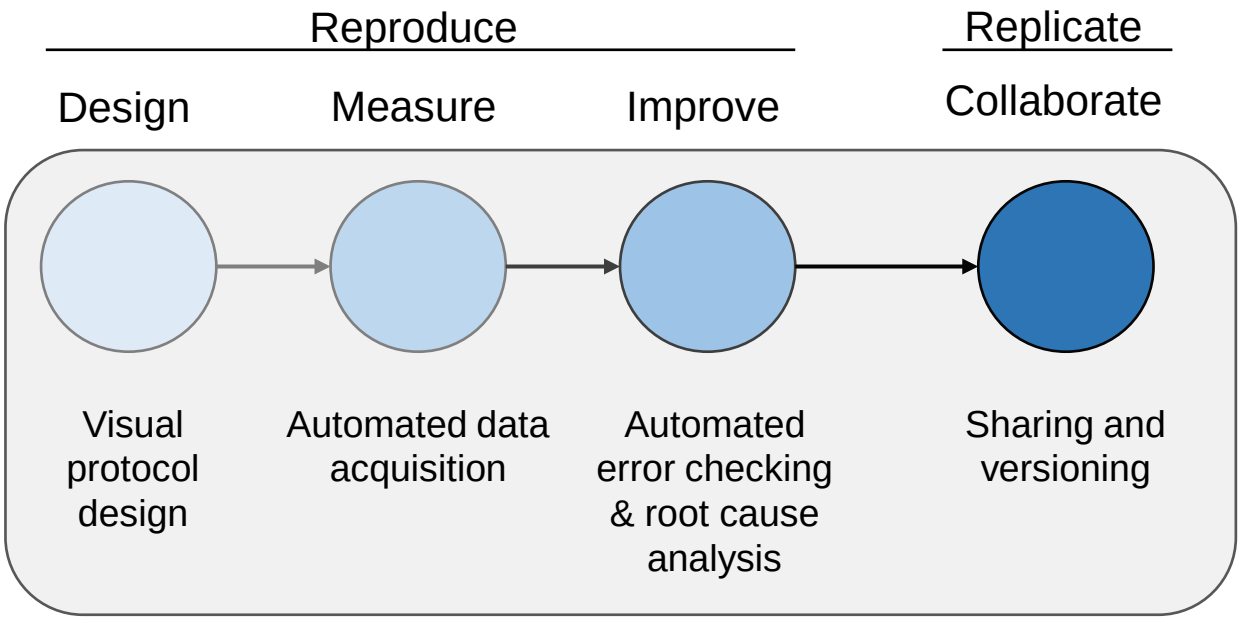
Can we do something similar in research?

- Computational researchers have demonstrated potential of such approaches (though mainstream adoption remains low)
- New data standards and improved tools are needed to extend this to bench researchers
- Emerging automated, outsourced lab systems may help to drive this standardisation, just as cloud computing drove standards for ‘run anywhere’ computing

Riffyn: Bringing CAD to the lab bench

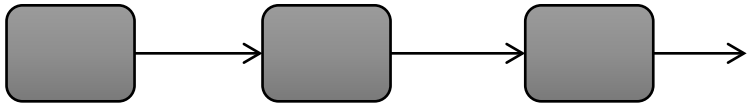
- Cloud-based software platform
- CAD for experiments
- Experimental process design at the heart of R&D
- Data stream capture from multiple sources
- Researcher gets unified & computable historical view of all their experimental data in the context of their experimental processes

Riffyn: a modular system



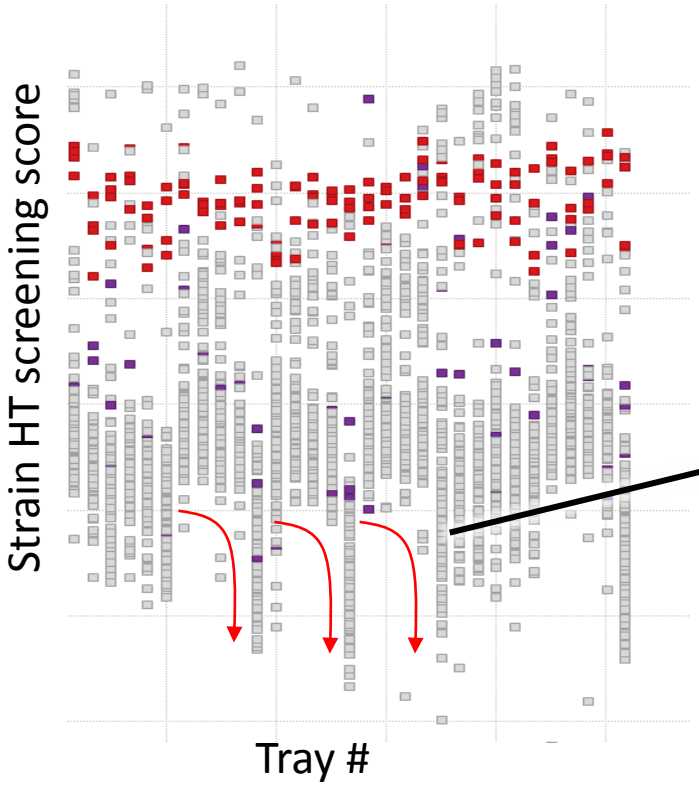
Data integration helps to track down root causes

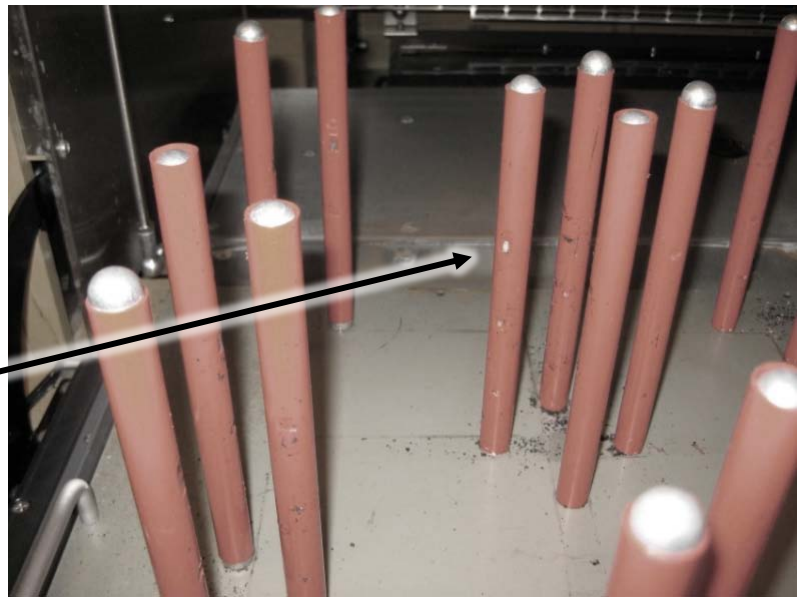
1 Define process inputs, outputs and critical variables



2 Measure inputs and outputs

3 Improve





Screenshots omitted

Standards for sharable, computable methods

- Need interoperable standards supported by:
  - Lab software suppliers
  - Instrument manufacturers
  - Outsourced research laboratories
  - Scientific publishers
  - Data repositories
- Existing standards including ISATAB, AniML, DICOM etc are good starting point but:
  - not (yet) sufficiently expressive or general
  - tools need to be easier to use
- For mass adoption, tools needs to be simple and visual

Drivers for adoption of new tools/standards

In industry

- Deliver measurable improvements to R&D productivity

In academia

- Make researchers life easier and more productive
- Demonstrate to funders that upfront experimental design and data integration will increase transparency and productivity and make better use of funding
- Persuade journals to support and ultimately require sharing of standard “computable” experimental process descriptions