



Solving the Big Data Problem: Smart Ways to Work Together

Dr Moira Smith, Tektonex Ltd
Dr Matthew Casey, University of Surrey

- Introduction and Backgrounds
- Big Data
 - Scope of the Challenge
 - Opportunities Presented
 - Societal and Commercial
 - Technology Advancements
 - Working Together
- Closing Remarks

Dr Moira Smith

- Director, Tektonex Ltd
 - BEng (Hons), PhD, CEng, MIET, MRAeS
 - Project and line management, 20 years experience, spanning:
 - Technology business creation, leadership and growth
 - R&D in industry (complex systems, software & processing)
 - Management (project and line)



Dr Matthew Casey

- Senior Lecturer, Computing Dep, University of Surrey
 - BSc PhD CEng MBCS CITP FHEA
 - 20 years experience, spanning:
 - Professional software engineering
 - Academic research (multi-sensory integration & processing)



Enabling Innovation

- ***Engineering & Technology Basis***
 - Advanced processing: smart algorithms for a wide range of markets
 - Systems engineering: full lifecycle
 - Software engineering: requirements capture to full V&V
 - Product development: rapid realisation of ideas and technology
- ***Business Knowledge***
 - Leadership, strategic vision and planning
 - Consultancy and management (programme, project, operations)
- **Co-Director & Founder, Dr Duncan Hickman**
 - BSc (Hons), MSc, PhD, MBA
 - 25+ years experience, inc. R&D in industry and academia
 - Engineering programme, project and line management





Lots of opportunities

- Potential for societal and commercial benefit

But where do we start?

- Issues with privacy, security, transmission, storage and processing...
- Not forgetting big data is only going to get bigger
- What can be learned from other domains where volume, variety, velocity and veracity are just as important?

Solutions will come from working together

- Smart, collaborative effort
- Processing in the right place at the right time

- The ever-increasing amount and availability of data, 24/7, presents enormous potential
- Benefits and impact have vast scope and reach

Societal

- Even greater global connectivity and instant, anywhere access to data through mobile
- Opportunities that affect major issues
 - Healthcare, Education, Environment, Security...
 - Lifestyle and Personal Information
 - Scientific discovery and endeavour

Commercial

- How to make sense of data for business benefit (e.g. smarter analysis, new ways of working)

And much more we don't even know about...

Data sharing and privacy

- Who owns what data?
- Greater opportunity in sharing (if willing)
- Open data standards and interoperability

Transmission, storage and processing

- Bandwidth limitations
- Size, location, backup and access to storage
- Processing the right information in the right way

Delivering innovative, cost-effective solutions

- Bringing together diverse data sources
- With solutions from multiple vendors (99.9% of private businesses are SMEs [1])

SOLUTION

Reliable

Veracity

Velocity

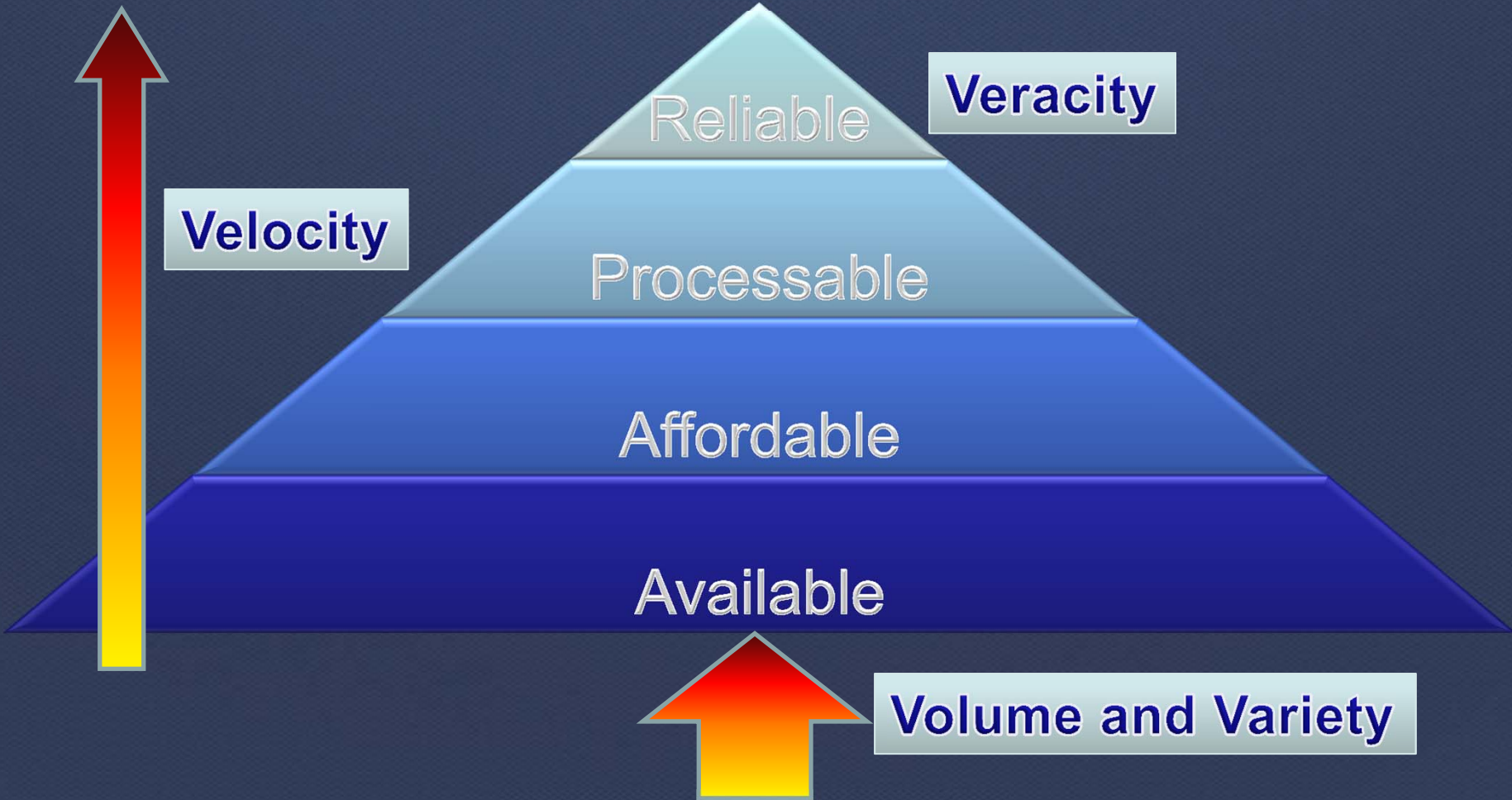
Processable

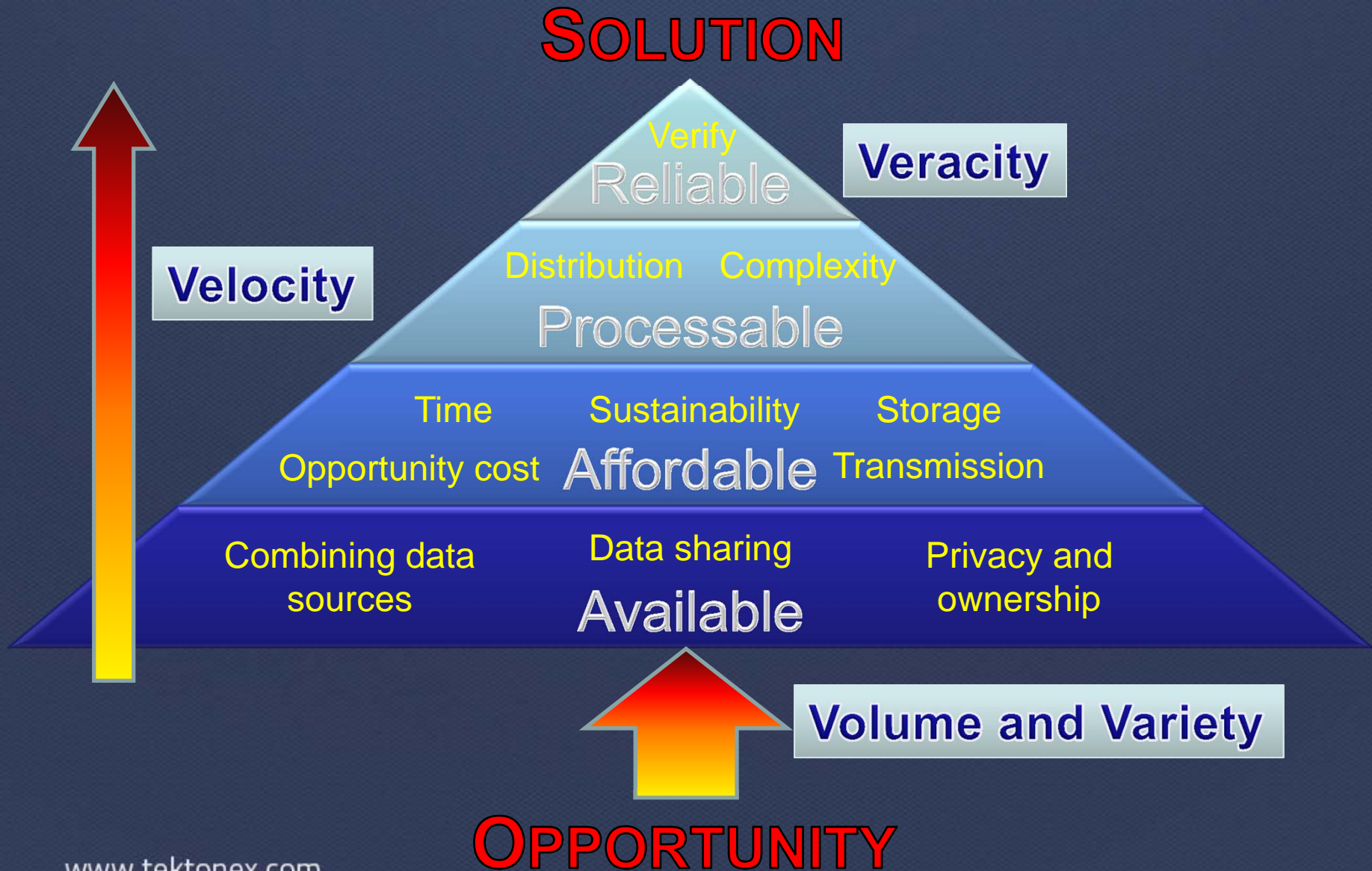
Affordable

Available

Volume and Variety

OPPORTUNITY





Existing data sources

- Health, communications, finance, environment, ...
- Social media, web analytics, consumer transactions, ...
- Personal information from diverse applications

Growing in volume and variety

- 7 billion people (growing by 2.4 people per second) [2]
- 6 billion web servers (up 5m in 1 month) [3]
- 6 billion mobile subscriptions (2011, up 0.5b per year) [4]
- 490m smartphone sales in (2011, 686m in 2012) [4]
- Internet of Things:
 - From smart cities to smart environments [5]



CCTV

- Estimated 1.9 million CCTV cameras in the UK [6]
- With compression at least 0.5TB per second
- Who is watching this data? (Why?)



Human Vision

- Each eye has 120M rods, 6M cones feeding 1M ganglion cells producing output of 1MB per second [7]
- Retina: contrast, on, off – salient data compression
- Mid-brain: rapid threat detection [8]
- Visual cortical areas: edges ... to perception (cf. [9])
- Process early and in parallel, transmit what is needed to dedicated systems which fuse and process





Biology vs. Computation

- Biological solutions tend to distribute processing
- Transmitting high volumes of data using multiple channels for hierarchical processing
- Computation provides fast, high-volume, reliable processing (for a fixed purpose)
- Relies on a central processing model using large-scale components (servers, cloud, ...)

Lessons to Learn?

- Why do we want to do everything centrally or in 'big bang' solutions?
- This causes problems with transmission, storage, processing, security...

Smart Processing and Distribution

- Place the processing in the right place
- Transmit lower bandwidth salient information
- Provide reactive and proactive control

Example

- Why transmit TB of unwatched CCTV data?
- Transmit salient information (smart detection)
- Provide automatic or on demand imaging

But this needs...

- Smart solutions and interoperability
- Organisations to work together on sharing data and processing



SOLUTION

Verifiability built
in at every stage

Reliable

Share processing
(services) to build
complex, flexible and
cost effective
solutions

Processable

Affordable

Available

Share data sources, which enforce
privacy and security, and which
are available via open standards



OPPORTUNITY





Recognising opportunity

- A single organisation's opportunity/problem or
- Collaborating organisations (NHS, agencies, ...)

Innovative solutions

- Universities: innovative timely, salient processing
- Institutes: standardising data sharing
- Specialist suppliers: rapid feasibility testing
- Solution providers: large-scale delivery
- Open Data Institute: a leading example

Long-term solutions in a changing world

- Embedding and adapting sustainable solutions
- Exploiting new opportunities



Standards, Policies and Support

- To solve problems irrespective of the data source, application or market
- Opportunities to bring initiatives together
- Open Data Institute as a hub?
- But what about *personal* data? Who owns *my* data? (cf. The Jericho Forum [10])
- Opportunities beyond corporate data sources

Technologies

- Data compression and distributed processing
- Intelligent algorithms for localised processing
- Event prediction for processing, transmission or storage (cf. National Grid)



1. Identify opportunity
2. Assess data sources and interoperability
3. Evaluate hierarchical data and processing
4. Develop distributed processing agents
5. Design support infrastructure
6. Deploy solutions
7. Evaluate, adapt and repeat



1. Identify opportunity
Stakeholders
2. Assess data sources and interoperability
Personal, corporate and partners
3. Evaluate hierarchical data and processing
Corporate, partners and vendors
4. Develop distributed processing agents
Vendors, specialists and universities
5. Design support infrastructure
Vendors and specialists
6. Deploy solutions
Vendors and specialists
7. Evaluate, adapt and repeat
Stakeholders

Online company handling video with BIG DATA challenges

The Objective

- To provide a robust and easily scalable service that allows video uploading, storage & retrieval, downloading and streaming
 - Anywhere worldwide, 24/7
 - On any fixed or mobile platform
 - With the best possible performance



Big Data Challenges

- Video storage
- Video streaming (upload and download)
- Fast indexing & searching
- Bandwidth
- Latency
- Traffic peaks
- Scalability

Risk Management

- Auto load balancing and failover, with data redundancy on RAID5
- Video backup on exact copy of video server on redundant part of system
- Auto detect and switch if loss of cloud connectivity

Optimal Performance

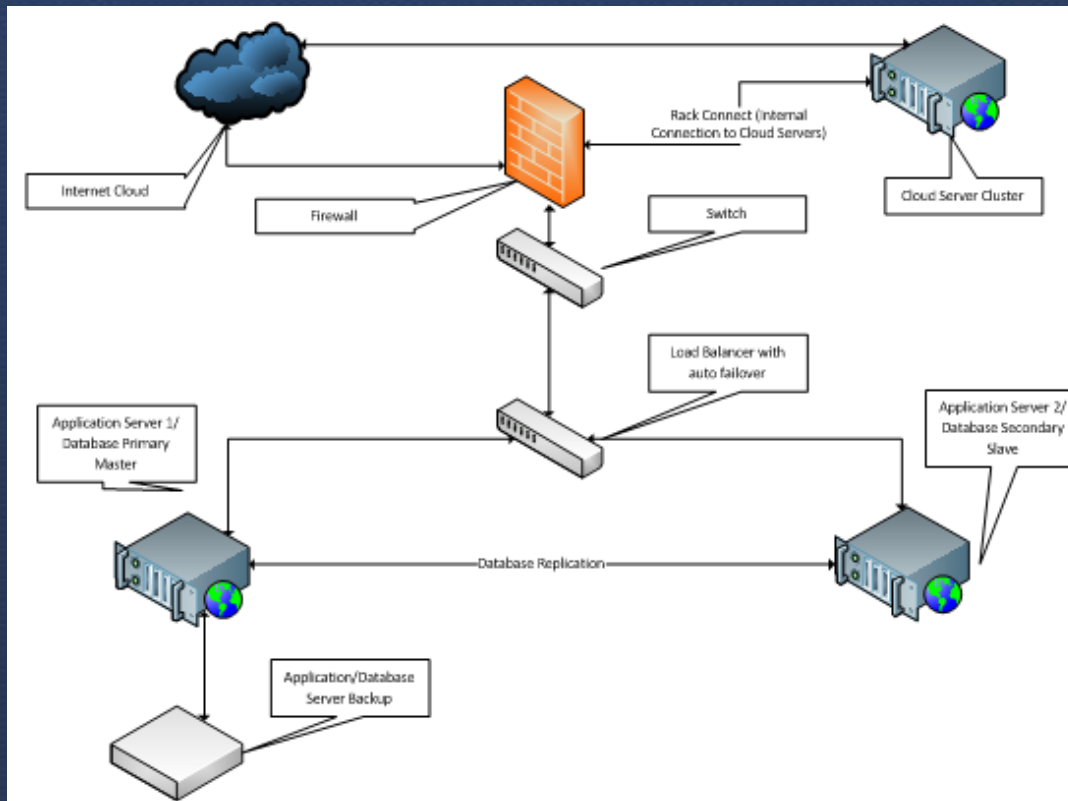
- Hybrid server solution
- Rapid scalability
- Fastest possible streaming worldwide

Indexing & Searching

- Hourly to flat files
- Live data held on DB
- 100's of millions of records searched in fractions of a second

Data Spikes

- Scripts communicate to cloud via APIs that auto-start new cloud servers as required
- Auto load detection



Video

- Cloud-based media server for recording incoming video
- Supports multi-platform (ios and mobile)
- Conversion server compresses and converts to all main video formats
- Standby image file can quickly be on cloud server if conversion demand peaks
- Scalable

Architecture

- Dedicated servers across web servers
- Clustered DB servers
- SAN data storage on RAID5



1. Identify opportunity
Stakeholders: Recruiters (SMEs, corporates, agencies)
2. Assess data sources and interoperability
Personal & commercial: All web & mobile, video & storage
3. Evaluate hierarchical data and processing
Corporate, partners and vendors: Hybrid architecture
4. Develop distributed processing agents
Vendors, specialists and universities: Key trusted partners
5. Design support infrastructure
Vendors & specialists: Leading global vendors
6. Deploy solutions
Vendors & specialists: Scalable delivery model
7. Evaluate, adapt and repeat
Stakeholders: Rapid response to provide enhancements

Recognising opportunity

- Data volume and availability will keep growing
- So too will irrelevant / superfluous data
- Harnessed intelligently, huge benefits possible

Grasping opportunity

- Needs innovative thinking and a multi-disciplinary approach
 - Technology (e.g. cross-cutting solutions)
 - Business level (inc. collaborative working)
- Impact so profound for Society that Government-level interactions and wider debate is required

1. Department for Business, Innovation and Skills (2012). Business Population Estimates for the UK and Regions 2012. <http://www.bis.gov.uk/assets/biscore/statistics/docs/b/12-92-bpe-2012-stats-release.pdf>. [Accessed 27-10-2012.]
2. Central Intelligence Agency (2012). CIA - The World Factbook. <https://www.cia.gov/library/publications/the-world-factbook/geos/xx.html>. [Accessed 10-11-2012.]
3. Netcraft Ltd (2012). Web Server Survey | Netcraft. <http://news.netcraft.com/archives/category/web-server-survey/>. [Accessed 10-11-2012.]
4. dotMobi (2012). Global mobile statistics 2012 Part A: Mobile subscribers; handset market share; mobile operators | mobiThinking. <http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats/a#smartphone-shipments>. [Accessed 10-11-2012.]
5. Libelium Comunicaciones Distribuidas S.L. (2012). Top 50 Internet of Things Applications – Ranking. http://www.libelium.com/top_50_iot_sensor_applications_ranking. [Accessed 10-11-2012].
6. Gerrard, G. & Thompson, R. (2011). Two Million Cameras in the UK. *CCTV Image*, 42, 10-12.
7. Koch, K., McLean, J., Segev, R., Freed, M.A., Berry II, M.J., Balasubramanian, V. & Sterling, P. (2006). How Much the Eye Tells the Brain. *Current Biology*, 16, 1428–1434.
8. Shi, C. & Davis, M. (2001). Visual pathways involved in fear conditioning measured with fear-potentiated startle: Behavioral and anatomic studies. *The Journal of Neuroscience*, 21(24), 9844–9855.
9. Hubel, D.H. & Wiesel, T.N. (1962). Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex. *Journal of Physiology*, 160, 106-154.
10. The Jericho Forum (2012). The Jericho Forum. www.opengroup.org/jericho/. [Access 14-11-2012.]



Solving the Big Data Problem: Smart Ways to Work Together

Dr Moira Smith, Tektonex Ltd
Dr Matthew Casey, University of Surrey