Cyber Security

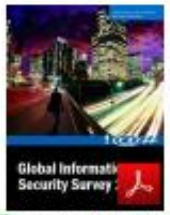
A Look Across Two Decades

A Quantitative Analysis of the Language of Security
2002-2019

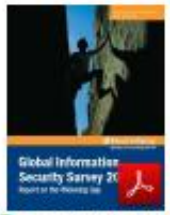
The Security Transformation Research Foundation

corix partners


2002_EY_GISS


2003_EY_GISS


2004_EY_GISS


2005_EY_GISS


2006_EY_GISS


2007_EY_GISS


2008_EY_GISS


2009_EY_GISS


2010_EY_GISS


2011_EY_GISS


2012_EY_GISS


2013_EY_GISS


2014_EY_GISS


2015_EY_GISS


2016-2017_EY_GISS


2017-2018_EY_GISS


2018-2019_EY_GISS

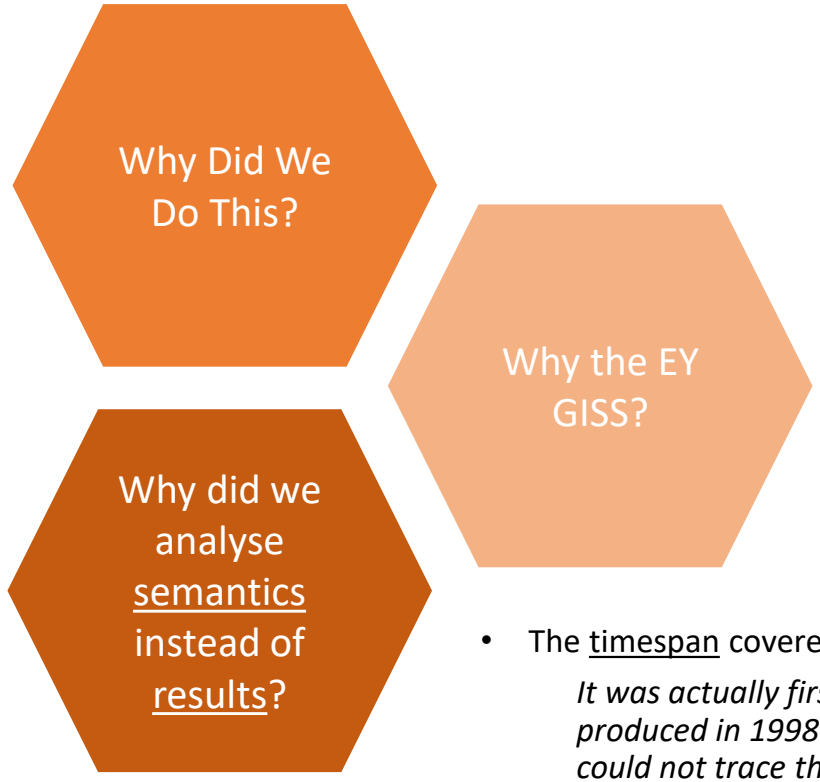
Cyber Security: A Look Across Two Decades

Results from the quantitative analysis of the semantics content of 17 annual **Global Information Security Surveys** from EY spanning the period 2002-2019

Cyber Security: A Look Across Two Decades

We performed a quantitative analysis of the frequency of keyword markers across a set of 77,750 meaningful words extracted from the text of the 17 GISS reports

To build a quantitative understanding on how the focus and priorities of the Security industry have evolved throughout the last 2 decades



No access to underlying data sets meant we could not compare or normalise results in a meaningful way

Semantics reveal the way the results were interpreted and the language used in such analysis is a good indicator of the industry focus points year after year

- The timespan covered

  It was actually first produced in 1998 but we could not trace the first 4 issues

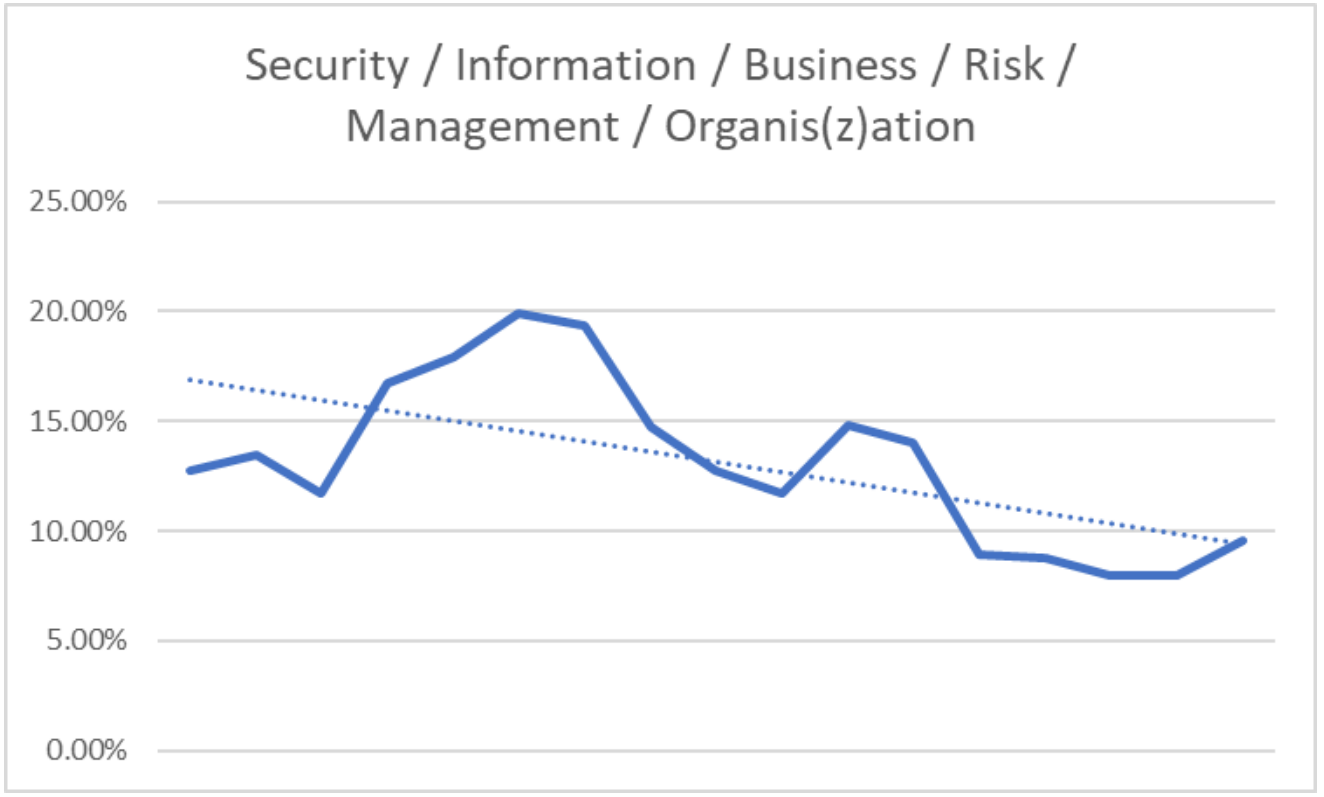
- The consistency in layout, size and approach
- The level of depth and general quality of the analysis

5 key findings, as indicators on how we (Security Practitioners) communicate with senior stakeholders and how our language has evolved over the past 2 decades

Cyber Security: A Look Across Two Decades

Finding 1

- The most common words are generic
- But while the language in the reports is dominated by generic terms, their overall proportion tends to be diminishing



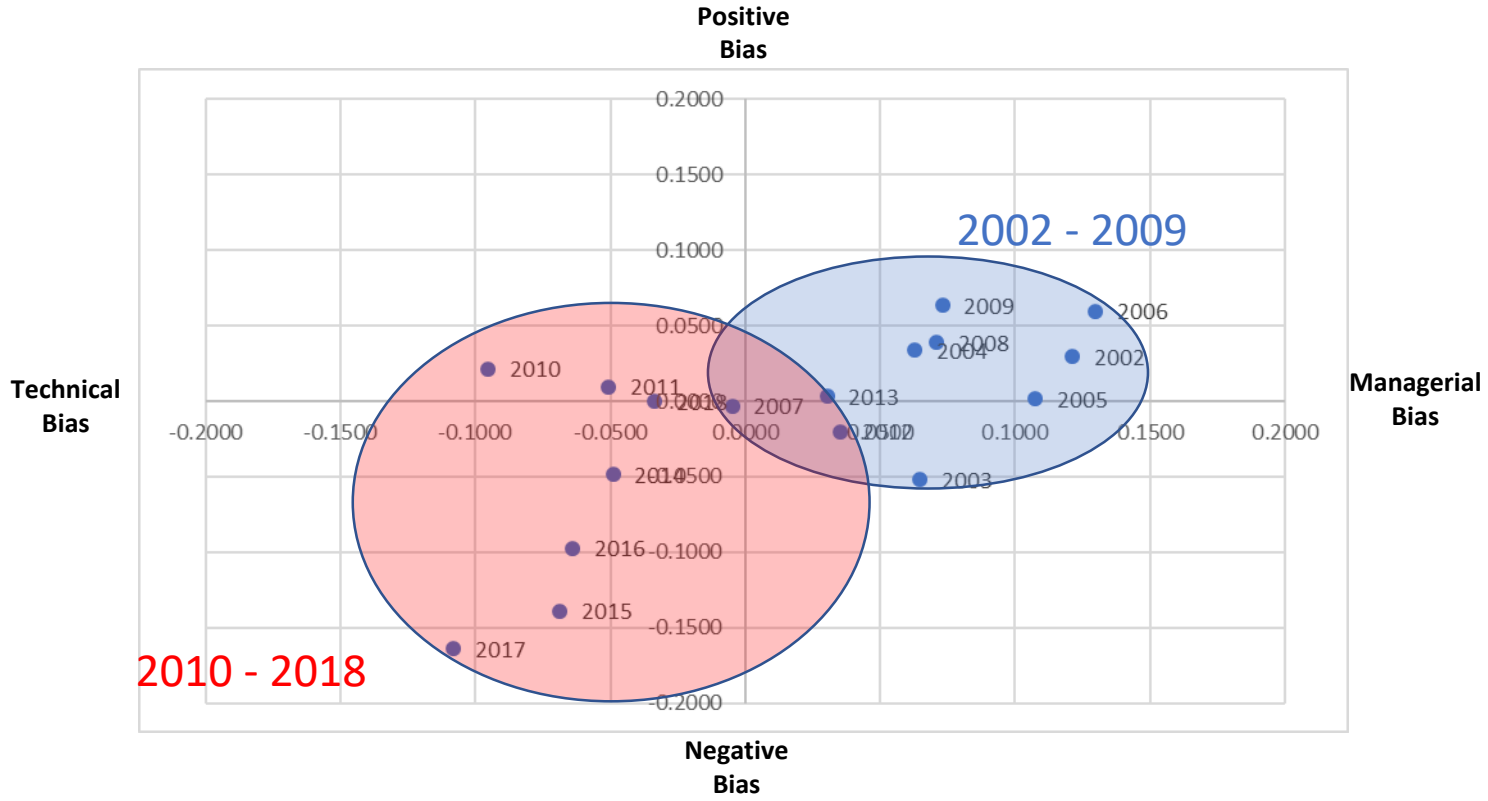
(*Frequencies of keyword markers in reports)*

We tend to talk about Security in more and more specific terms

Cyber Security: A Look Across Two Decades

Finding 2

When adding a sentiment analysis layer over the data, the 2 decades appear to be split by a clear semantic shift towards a more technical and more negative language



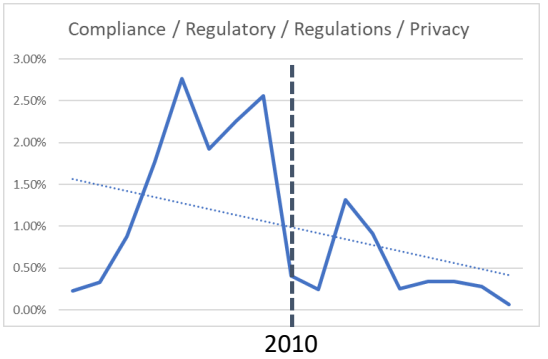
If our language around Security has become more specific, it has also become more technical and more negative

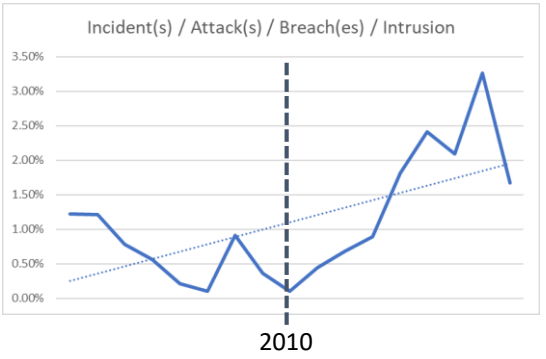
Cyber Security: A Look Across Two Decades

Finding 3

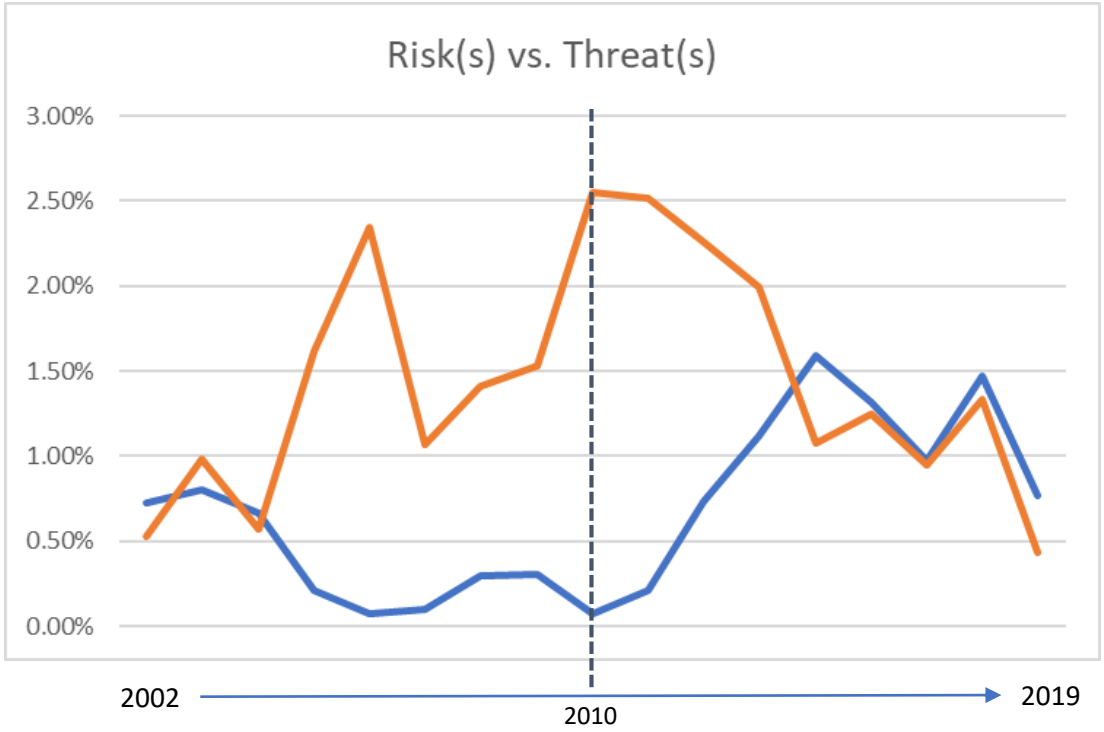
This split reflects a significant shift in focus across the 2 decades

The Compliance and Risk considerations which dominate the period 2002-2009 are clearly replaced by Incidents and Threats considerations during the following decade





(*Frequencies of keyword markers in reports)*



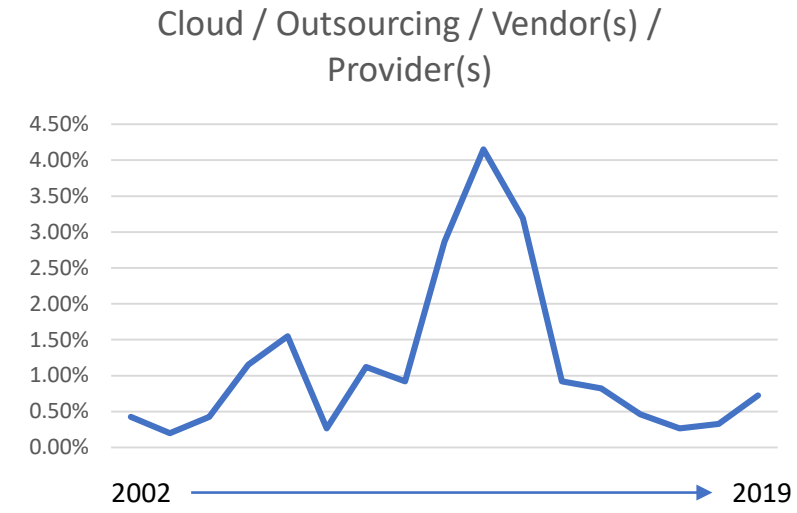
Cyber Security: A Look Across Two Decades

Finding 4

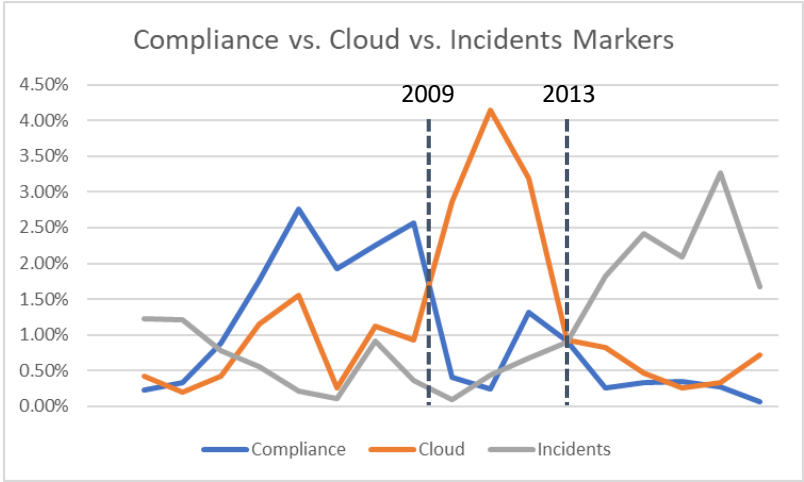
Outsourcing and Cloud considerations dominate sharply during a short middle period (2010-11-12) then vanish into acceptance

A sense of Realisation seems to dominate the junction between the 2 decades:

This is no longer JUST about Compliance and Risk: Tech is changing, Threats are real and Incidents do impact Business



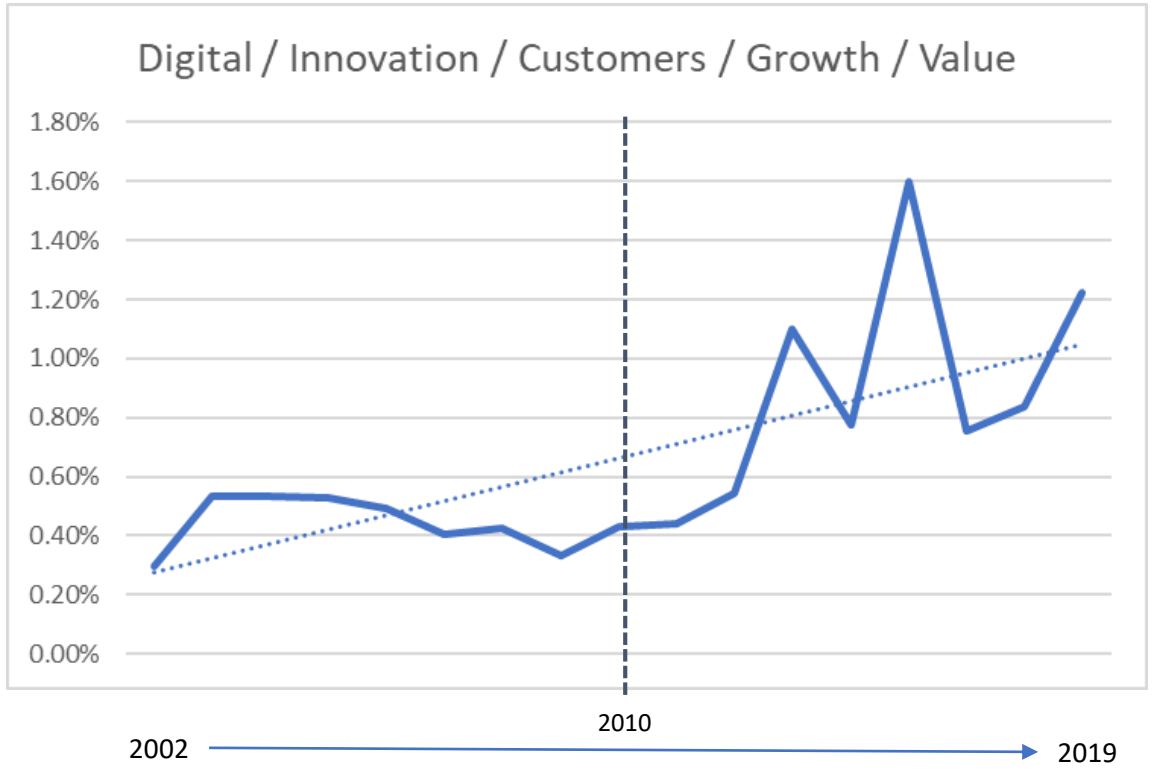
(*Frequencies of keyword markers in reports)*



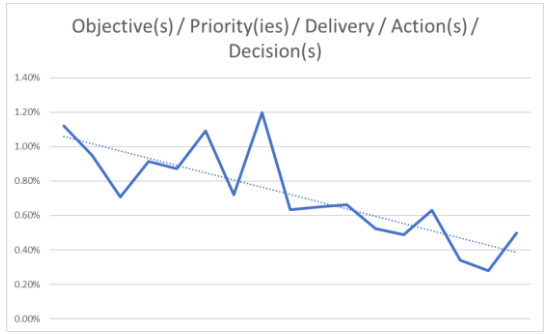
Cyber Security: A Look Across Two Decades

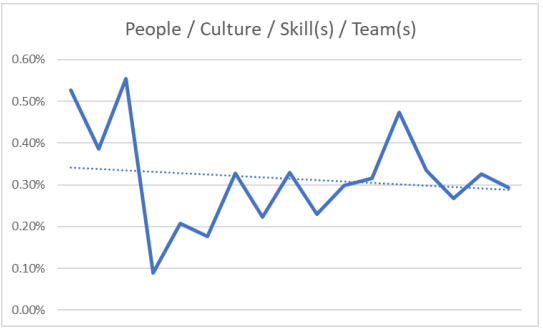
Finding 5

Our Business language tends to sharpen throughout the last decade but our focus on Execution and People tends to dwindle



(*Frequencies of keyword markers in reports)*





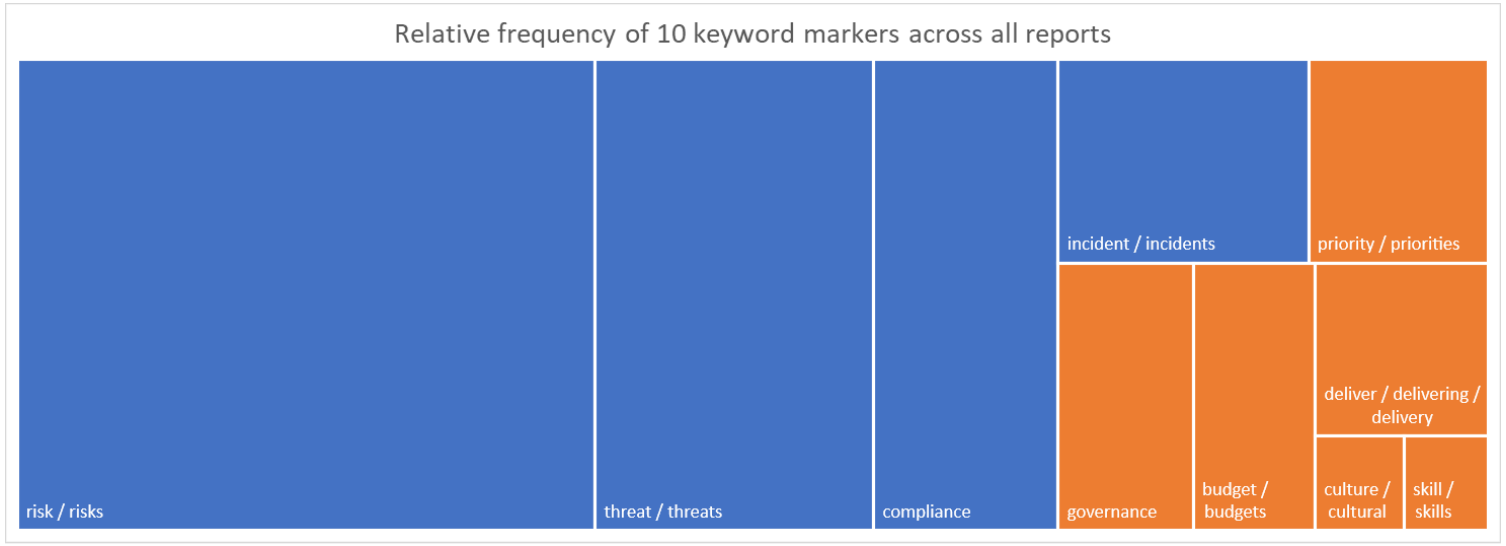
Cyber Security: A Look Across Two Decades

Keyword markers such as

risk / threat / compliance / incident

are 3.5 times more frequent across all reports than

governance / budget /delivery / priority / culture / skill



We tend to talk a lot about what could go wrong …
… but not as much about we could do to fix things

The semantics analysis shows the clear emergence of 2 periods

- Security as a balancing act between Compliance Requirements and Risk Appetite (and costs)

The CISO as Risk Manager

- Security as a necessary barrier against real Threats in a context of massive technological change (and the aftermath of a historical financial crisis)

The CISO as Fire Fighter

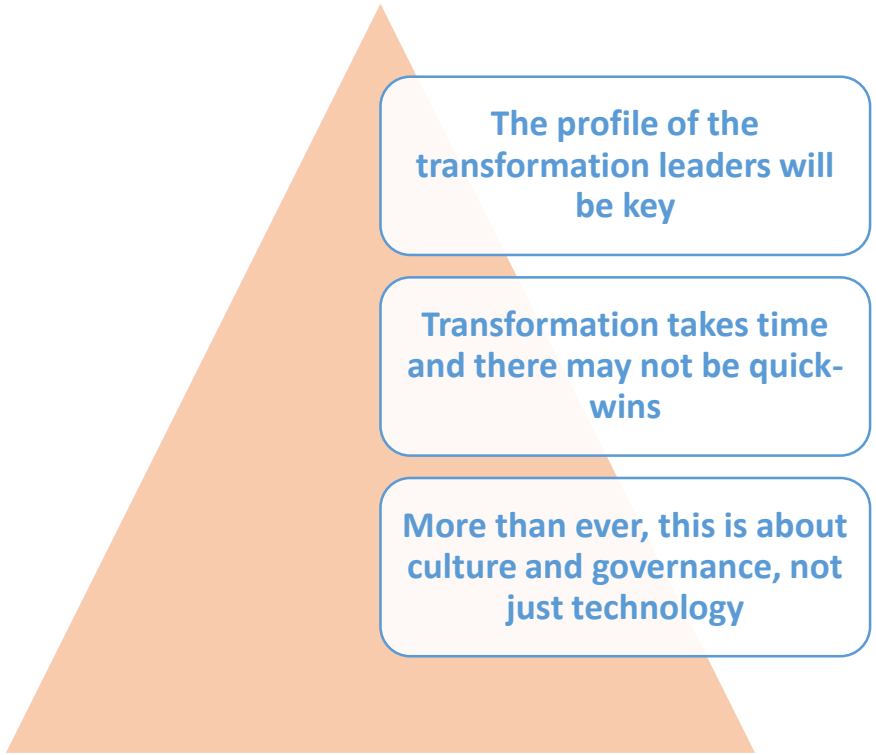
What the next decade must address

The Compliance Decade
Security as a balancing act between Compliance Requirements and Risk Appetite (and costs)

The Realisation Decade
Security as a necessary barrier against real Threats in a context of massive technological change (and the aftermath of a historical financial crisis)

The Execution Decade
Security as an imperative in the "when-not-if" era, in a context of significant maturity deficit in many firms (and potentially massive regulatory fines)

3 management considerations in conclusion ahead of the next decade



A good “fire-fighter” may not be a good “transformer”

Senior management must be able to look beyond the short-term and stay focused on transformational objectives

Throwing money at tech vendors will not build anything lasting without the right organisation and operating model

Methodology Summary

We gathered all EY Global Information Security Surveys from 2002 to 2019 in PDF format.

We read each PDF using the pdf_text function from the pdftools package in order to obtain the full text for each page in machine-readable format. Because of the nature of PDFs, some of the text could not properly be read (fancy headlines, non-standard font in some titles, etc.) but we were successful in getting more than 95% of the content of each report.

We then performed some amount of data cleaning – removing standard English stopwords (e.g. and, but, all, did, …), all one- and two-letter words, as well as some reports-specific uninformative words such as: ernst, young, annual, survey, percent, or respondents.

We used the quanteda package – the standard tool for managing and analyzing textual data in R – in order to turn the raw text into analyzable format called a document-feature matrix (dfm).

A dfm is simply a (typically very sparse) matrix where each row i is a different document (here, each row is a year), each column j is a word, and every entry [i, j] is the count of word j in document i. No stemming was performed at this stage.

We then computed the top 100 terms for each year and exported the final ranking (along with absolute counts and frequencies) to CSV for easy analysis in Excel. Stemming and grouping of terms was performed manually in Excel using domain-expertise.

After manually selecting the most interesting terms to the analysis, we went back to the dfm to complete the count for those terms in years in which they did not make it to the top 100.

Many thanks to Vincent Viers for his help with the research and the methodology

The Security Transformation Research Foundation

The Security Transformation Research Foundation is a dedicated think-tank and research body aimed at approaching Security problems differently and producing innovative and challenging research ideas in the Security, Business Protection, Risk and Controls space

www.securitytransformation.com

@Transform_Sec

The Security Transformation Research Foundation

FINAL – 1.0 – September 2019