You Won't Believe How Easy it is To Build an AI!

Rebuilding Inspec for the 21st Century

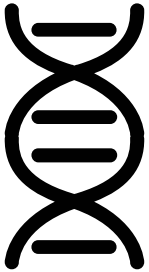
David Smith, Head of Technology Strategy & Product Solutions
The IET

About the IET

- The IET is one of the world's largest engineering institutions with over 175,000 members in 150 countries. It is also the most multidisciplinary – to reflect the increasingly diverse nature of engineering in the 21st century.
- The IET is working to engineer a better world by inspiring, informing and influencing our members, engineers and technicians, and all those who are touched by, or touch, the work of engineers.

INSPEC: A Bluffers Guide

- A highly curated A&I database covering Engineering, Computing and Physics (etc etc)
- For over 40 Years
- ~18 million abstracts
- So Much Metadata WOW!
- Several hundred years worth of Human Expertise keeps a very close eye on the metadata quality

Physics	Electrical and electronic engineering	Computing and control engineering	Information technology	Production, manufacturing & mech. engineering
				
9.84 million	**7.1 million**	**5.3 million**	**113K**	**2.1 million**

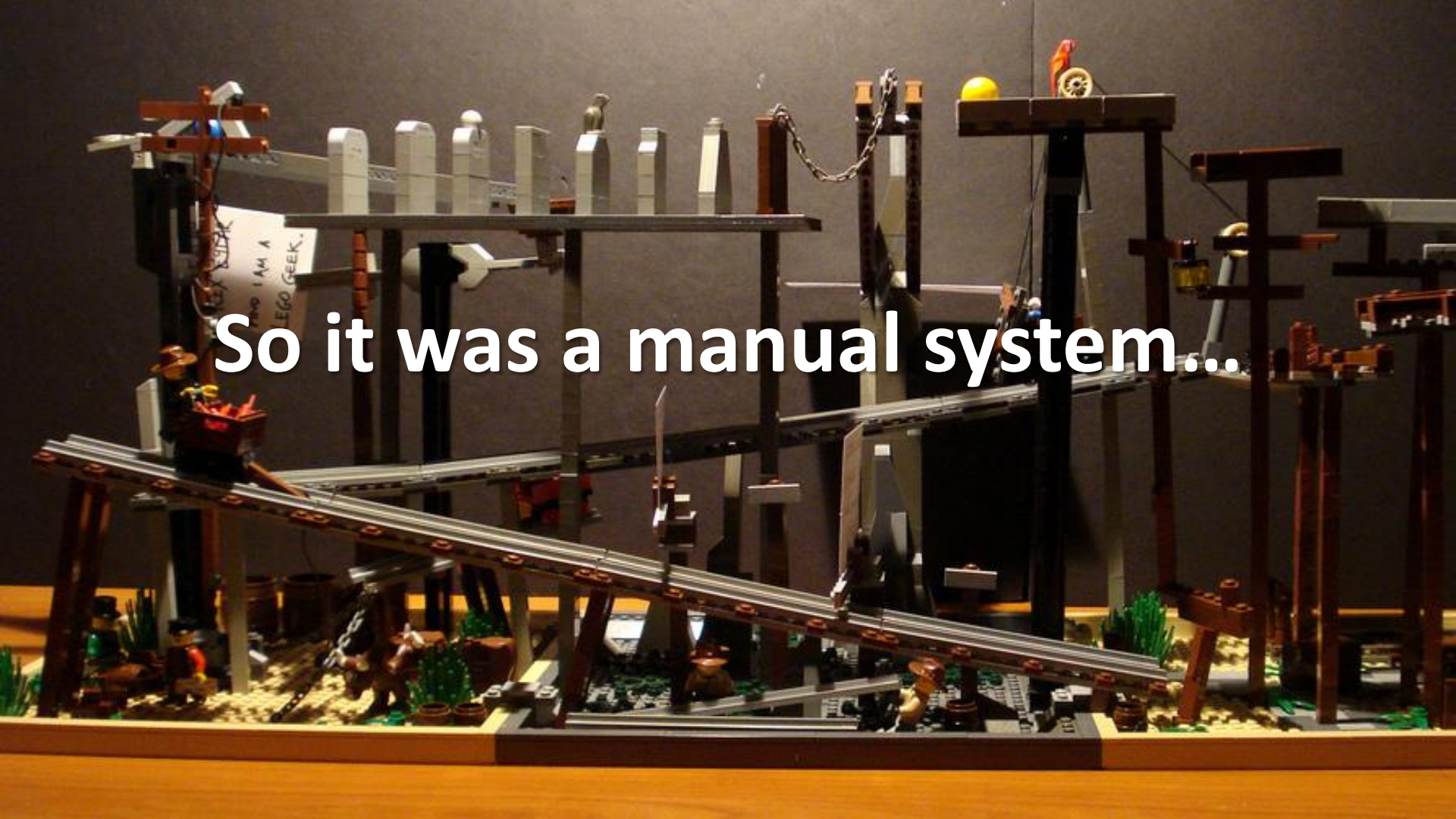
17.2M

records (Oct 2017)

- **870,000** records added in 2016
- Optional Archive adds another **873,699**
- Over **4,200** journals, and **3,000** other publications from 700 publishers

1969 — To date

- Or **1898** with the Archive


So it was a manual system…

And here’s how it worked…

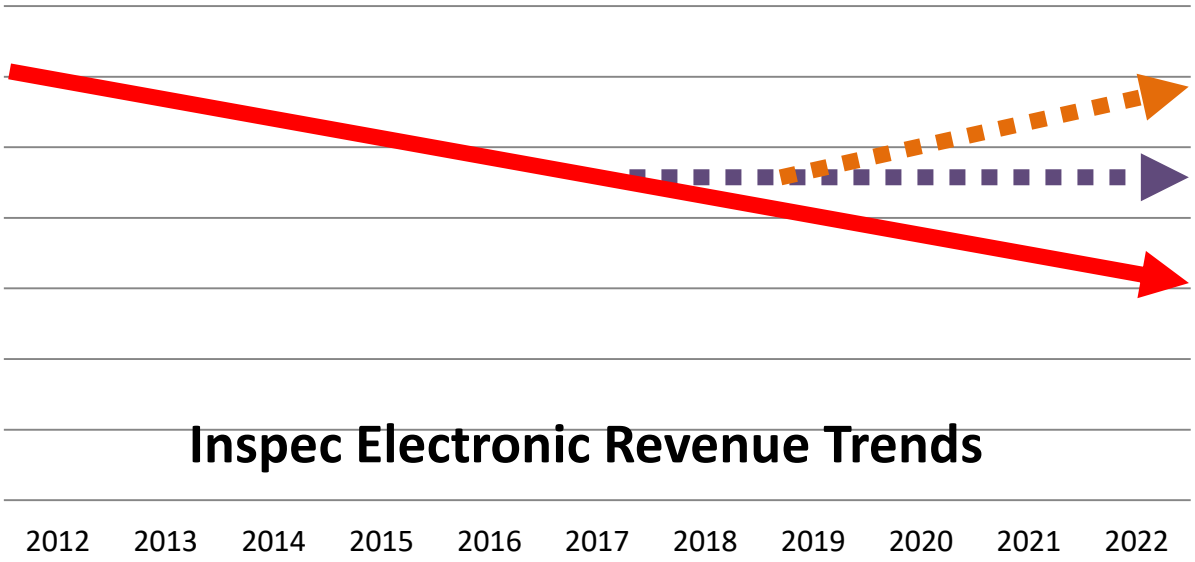
We needed to change this…

- The Tech was E.O.L
- The Manual methods were restrictive & expensive (but V High Quality)
- We had reached an upper limit on coverage and volume
- **There were clear opportunities to rethink what we were doing and why…**
- **Rebooting INSPEC production could open up new business avenues – if we got it right.**

Goals…

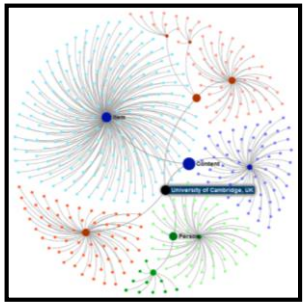
- **Deliver cost savings (ROI argument used)**
- **Move the human effort further up the value chain**
- Be able to extend coverage capabilities
- Be able to extend volume capabilities
- Reconfigure the data in INSPEC to allow new ways of asking questions of it.
- **Build a new IET IP asset**
- Focus on automation with human QA (‘Ground Truthing’)

Historic decline in A&I revenues



AI & Knowledge Graph Powered Inspec Production System

Humans	Machine	Machine & Human QA	Machine	Machine / Human QA
Acquisition	Normalisation	Metadata Application	**Product Generation**	Output
• Ingest Content • 3rd Party • IET	• Render to standard Inspec schema for onward processing	• The AI lives here…	• There are many flavours of Inspec	• Various outputs as needed



The Inspec Knowledge Graph covers all of the Inspec subject areas

The Inspec project has given the IET …

- A 21st C production process
  - Scalable
  - Extensible
  - Flexible

- A new POWERFUL and valuable asset: The Knowledge Graph – the IET's USP
  - The Inspec artificial intelligence (AI)
  - The Inspec domain models (& ontologies)
  - The reconfigured organisation of Inspec data

How does it work?

- It uses a mixture of
  - Heuristics
  - Natural Language Processors
  - Statistical analysis tools
  - And a selection of AI algorithms.
- We built a detailed domain model & Ontology for it to use
- It's been trained via directed learning of a golden corpus (circa 600K documents across INSPEC)

How does it work?

- And a selection of AI algorithms…
  - We looked at adaboost (go see wikipedia…)
  - Also word2vec (likewise)
  - And Tensorflow – the deep learning algo from Google. Interesting results… It did some rather odd things TBH so we abandoned that approach.

Inspec USP 1: What does the Inspec AI Do?

Title: Branching fraction measurements of $B \to \eta_c K$ decays

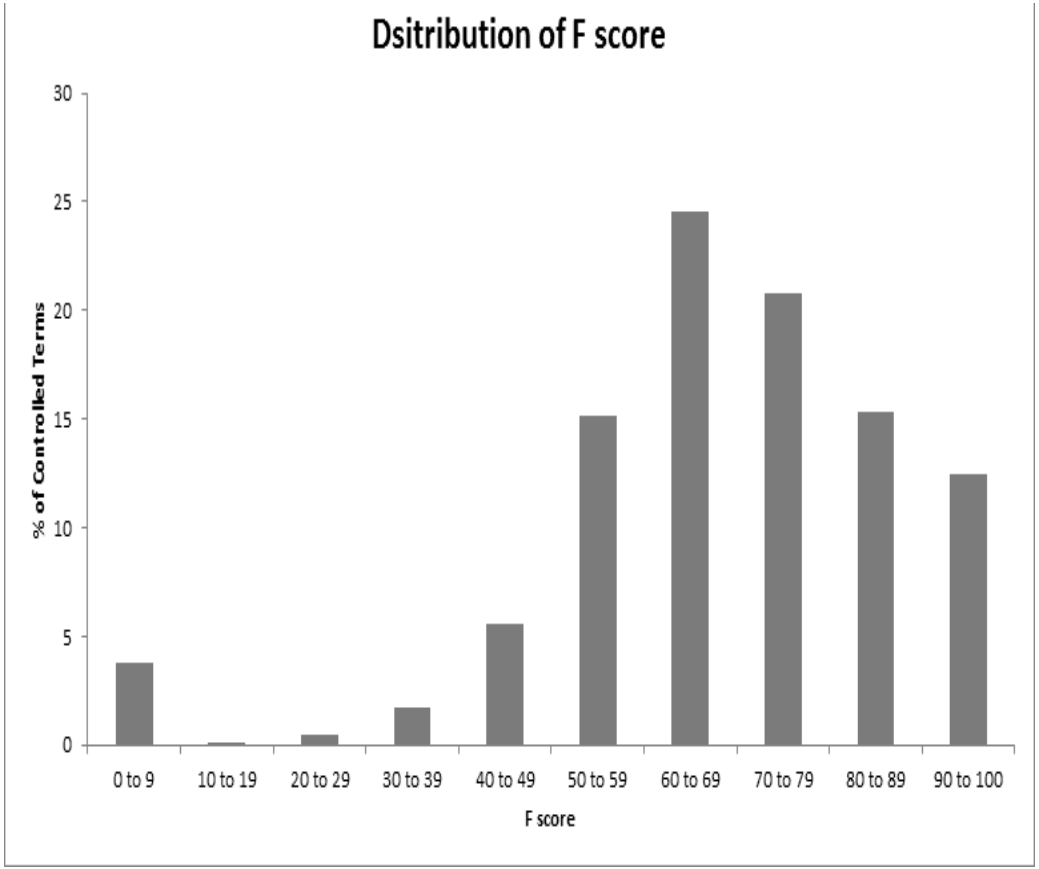
Abstract: We study the decays $B^+ \to \eta_c K^+$ and $B^0 \to \eta_c K^0$, where the η_c is reconstructed in the $K_S^0 K^\pm \pi^\mp$ and $K^+K^-\pi^0$ decay modes. Results are based on a sample of 86 million $B\bar{B}$ pairs collected with the BABAR detector at the SLAC e^+e^- B Factory. We measure the product of branching fractions $B(B^+ \to \eta_c K^+) \times B(\eta_c \to K\bar{K}\pi) = (7.40 \pm 0.50 \pm 0.70) \times 10^{-5}$ and $B(B^0 \to \eta_c K^0) \times B(\eta_c \to K\bar{K}\pi) = (6.48 \pm 0.85 \pm 0.71) \times 10^{-5}$, where the first error is statistical and the second is systematic. In addition, we search for $B \to \eta_c K$ events with $\eta_c \to 2(K^+K^-)$ and $\eta_c \to \phi\phi$ and determine the η_c decay branching fraction ratios $B(\eta_c \to 2(K^+K^-))/B(\eta_c \to K\bar{K}\pi) = (2.3 \pm 0.7 \pm 0.6) \times 10^{-2}$ and $B(\eta_c \to \phi\phi)/B(\eta_c \to K\bar{K}\pi) = (5.5 \pm 1.4 \pm 0.5) \times 10^{-2}$. (20 refs.)

Free Indexing: B^+ decay into etac+kaon+; B0 decay into etac+kaon0; B-antiB pairs; etac decay into kaon+antikaon+pion; B decay into etac+kaon; etac decay into kaon++kaon-; etac decay into phi+phi; branching fraction

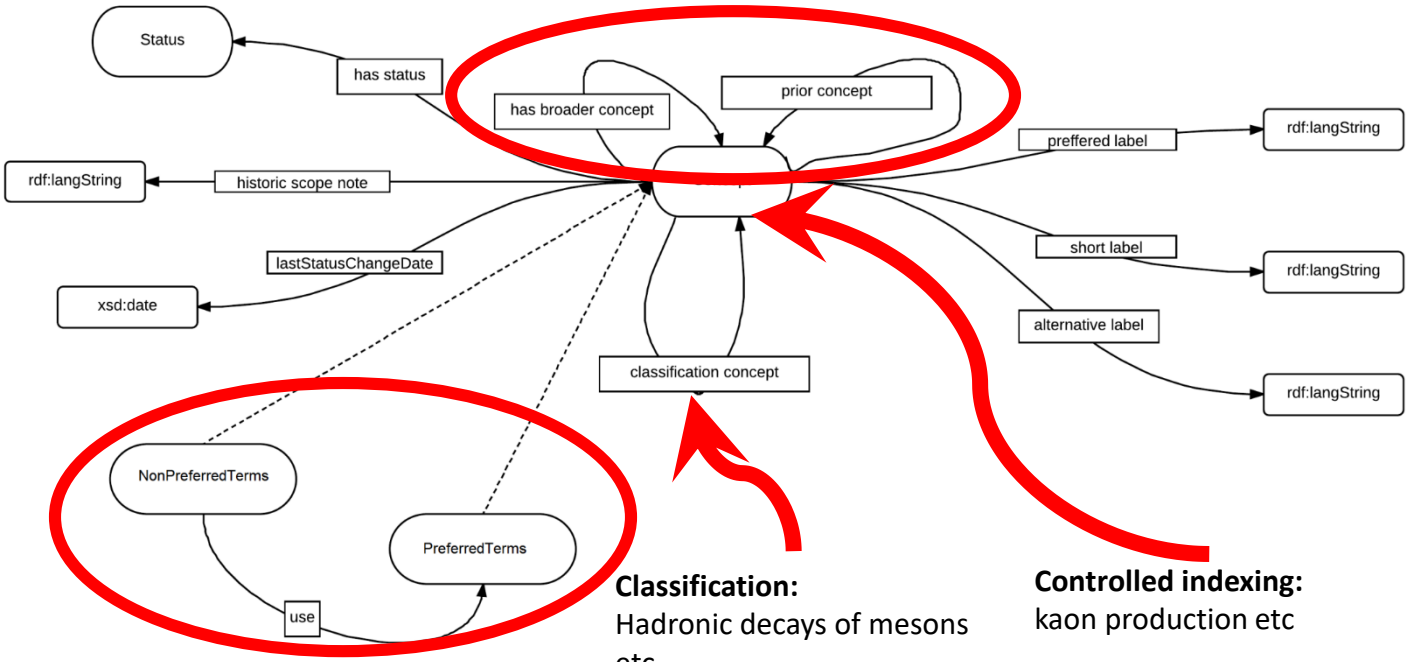
Controlled indexing: B mesons; eta mesons; kaon production; meson hadronic decay; phi mesons; pion production.

Classification: A1325 Hadronic decays of mesons; A1440M a and B mesons; A1440K rho, omega, and eta mesons; A1440N psi/J, upsilon, phi mesons

Is it any good?

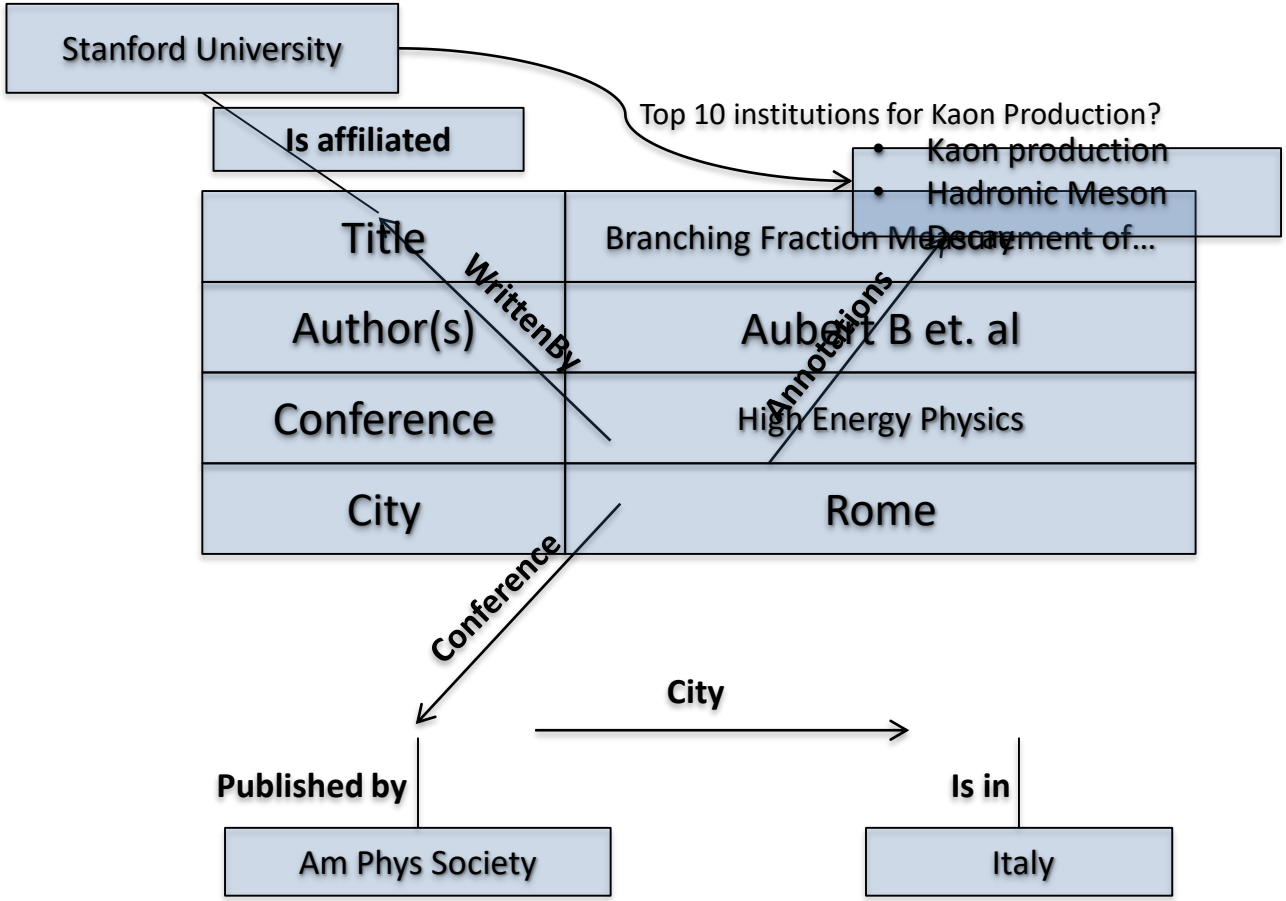


Inspec USP 2: Inspec Domain Models



Classification:
Hadronic decays of mesons etc

Controlled indexing:
kaon production etc


Stanford University
Is affiliated
Top 10 institutions for Kaon Production?
Kaon production
Hadronic Meson
Decayment of…
Title
Branching Fraction Measurement of…
Author(s)
Aubert B et. al
Conference
High Energy Physics
City
Rome
WrittenBy
Annotations
Conference
City
Published by
Is in
Am Phys Society
Italy

Inspec Analytics
Beta

Search for an organisation

Organisations

Subject Areas

Keywords

Authors

Organisations

University of Cambridge

Subject Classifications

Thesaurus Keywords

Collaborating organisations

Articles

Authors

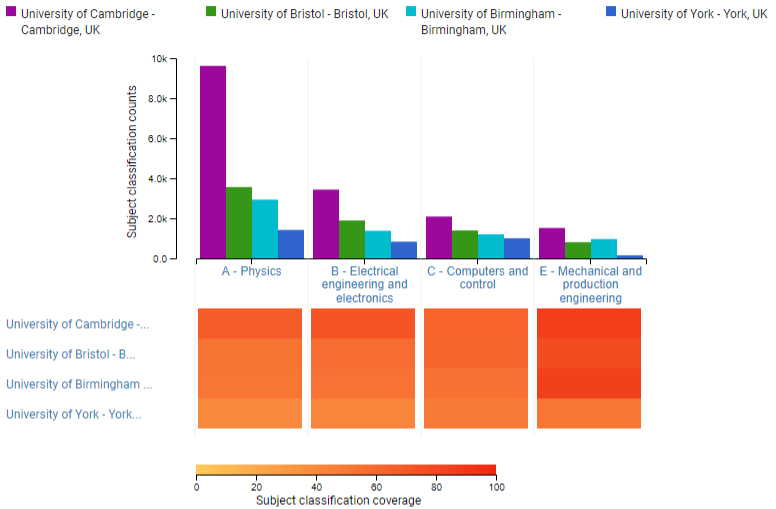
Detailed classification comparison

Controlled Terms

Detailed classification code comparison

From 2013 To 2017

Subject Classification



Organisation	Articles	Inspec Classification Counts							
		A - Physics		B - Electrical engineering and electroni...		C - Computers and control		E - Mechanical and production engineerin...	
		Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage
University of Cambridge Cambridge, UK	12914	9613	67.9%	3433	72%	2088	63.2%	1518	86.1%
University of Bristol Bristol, UK	5778	3560	52.9%	1892	57.5%	1393	62.1%	808	77.2%
University of Birmingham Birmingham, UK	4882	2932	52.2%	1377	53.2%	1198	53.7%	963	84.8%
University of York York, UK	2678	1419	39.9%	833	42.3%	1003	49.3%	152	51.9%

Organisations

University of Cambridge

Subject Classifications

Thesaurus Keywords

Collaborating organisations

Articles

Authors

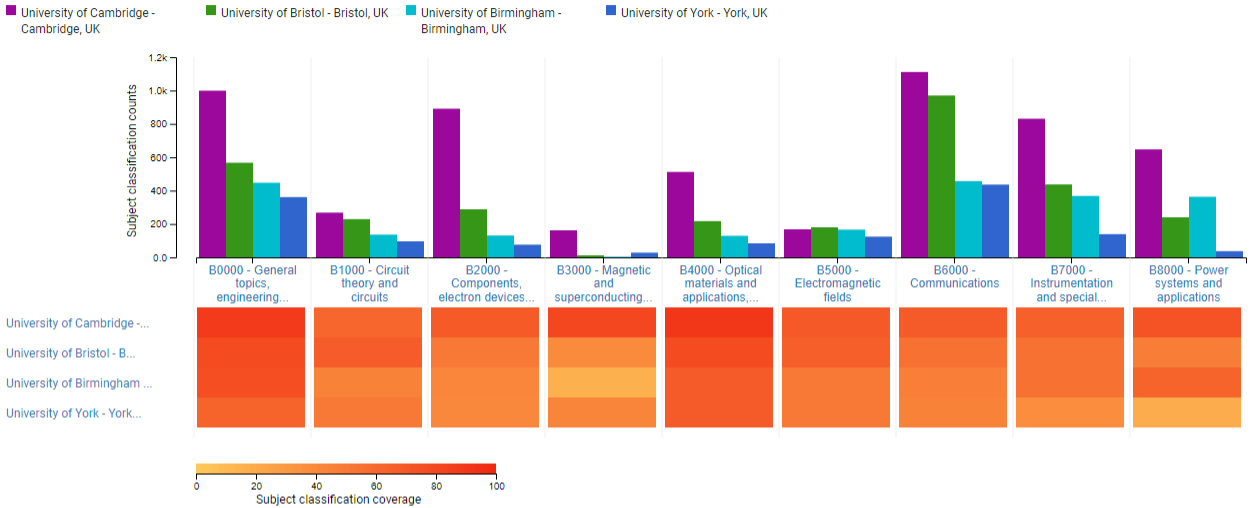
Detailed classification comparison

Controlled Terms

Detailed classification code comparison

From 2013 To 2017

Subject Classification ▸ B - Electrical engineering and electronics



Organisation	Articles	Inspec Classification Counts																	
		B0000 - General topics, engineering mathematics ...		B1000 - Circuit theory and circuits		B2000 - Components, electron devices and materia...		B3000 - Magnetic and superconducting materials a...		B4000 - Optical materials and applications, elec...		B5000 - Electromagnetic fields		B6000 - Communications		B7000 - Instrumentation and special applications		B8000 - Power systems and applications	
		Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage
University of Cambridge Cambridge, UK	12914	999	88.1%	268	61.5%	891	68.9%	162	80.8%	512	90%	168	69.2%	1111	68.4%	831	65%	647	72.7%
University of Bristol Bristol, UK	5778	567	78%	229	67.3%	288	49.5%	12	38.5%	217	78%	180	65.4%	970	54.7%	437	54.7%	240	46.5%
University of Birmingham Birmingham, UK	4882	447	76.3%	136	44.2%	131	41.7%	4	15.4%	129	68%	166	50%	457	46.3%	368	54.7%	363	62.6%
University of York York, UK	2678	361	62.7%	96	50%	76	40.8%	28	42.3%	84	68%	124	50%	436	44.2%	139	37.2%	37	18.2%

Organisations

University of Cambridge

Subject Classifications

Thesaurus Keywords

Collaborating organisations

Articles

Authors

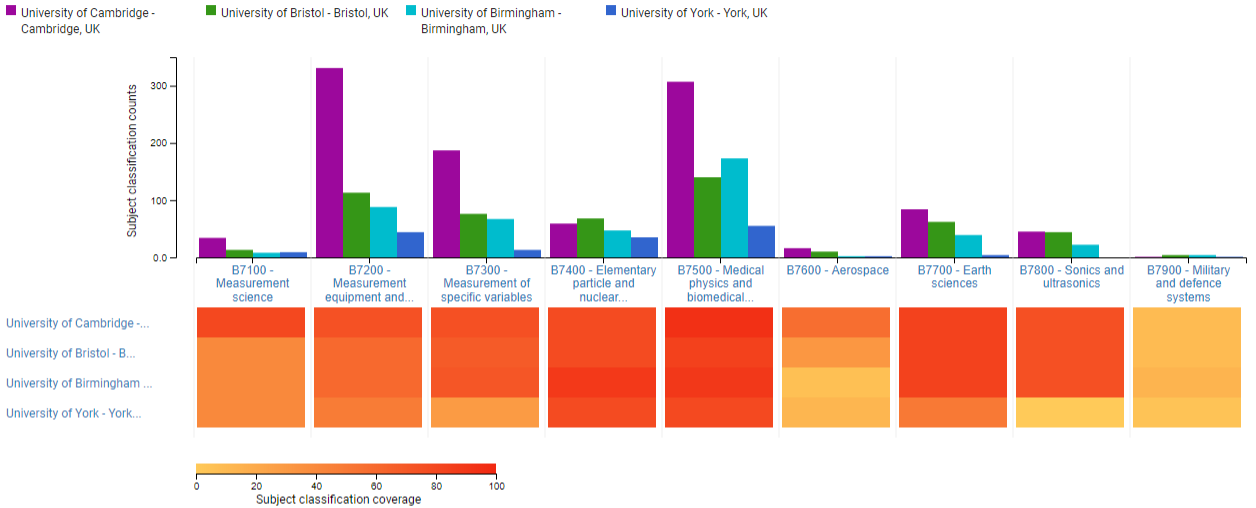
Detailed classification comparison

Controlled Terms

Detailed classification code comparison

From 2013 To 2017

Subject Classification ▸ B - Electrical engineering and electronics ▸ Instrumentation and special applications



Organisation	Articles	Inspec Classification Counts																	
		B7100 - Measurement science		B7200 - Measurement equipment and instrumentatio...		B7300 - Measurement of specific variables		B7400 - Elementary particle and nuclear instrume...		B7500 - Medical physics and biomedical engineeri...		B7600 - Aerospace		B7700 - Earth sciences		B7800 - Sonics and ultrasonics		B7900 - Military and defence systems	
		Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage
University of Cambridge Cambridge, UK	12914	34	80%	331	74.1%	187	75%	59	77.8%	307	94.7%	16	56.3%	84	83.3%	45	75%	1	9.1%
University of Bristol Bristol, UK	5778	13	40%	113	59.3%	76	67.9%	68	77.8%	140	84.2%	10	31.3%	62	83.3%	44	75%	4	9.1%
University of Birmingham Birmingham, UK	4882	8	40%	88	59.3%	67	71.4%	47	88.9%	173	89.5%	2	6.3%	39	83.3%	22	75%	4	13.6%
University of York York, UK	2678	9	40%	44	48.1%	13	28.6%	35	77.8%	55	78.9%	2	12.5%	4	50%	0	0%	1	4.5%

Organisations

University of Cambridge

Subject Classifications

Thesaurus Keywords

Collaborating organisations

Articles

Authors

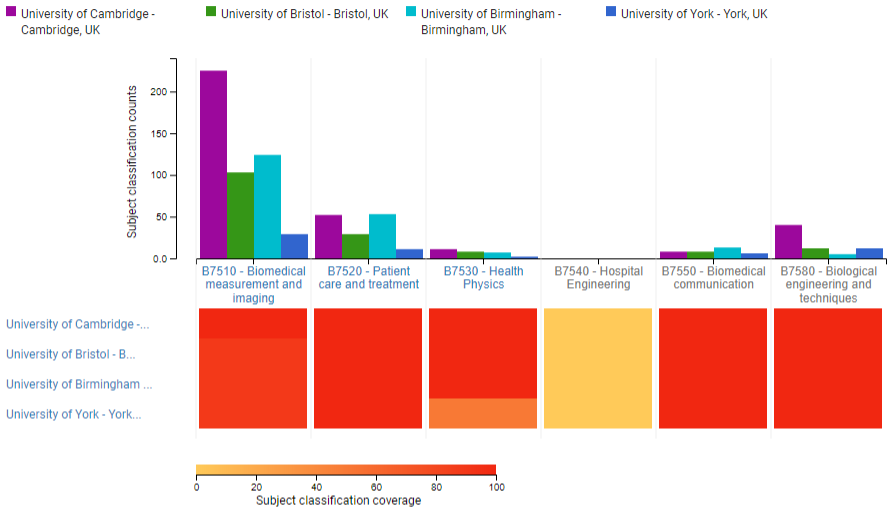
Detailed classification comparison

Controlled Terms

Detailed classification code comparison

From 2013 To 2017

Subject Classification ▸ B - Electrical engineering and electronics ▸ Instrumentation and special applications ▸ Medical physics and biomedical engineering



Organisation	Articles	Inspec Classification Counts											
		B7510 - Biomedical measurement and imaging		B7520 - Patient care and treatment		B7530 - Health Physics		B7540 - Hospital Engineering		B7550 - Biomedical communication		B7580 - Biological engineering and techniques	
		Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage
University of Cambridge Cambridge, UK	12914	225	100%	52	100%	11	100%	0	0%	8	100%	40	100%
University of Bristol Bristol, UK	5778	103	88.9%	29	100%	8	100%	0	0%	8	100%	12	100%
University of Birmingham Birmingham, UK	4882	124	88.9%	53	100%	7	100%	0	0%	13	100%	5	100%
University of York York, UK	2678	29	88.9%	11	100%	2	50%	0	0%	6	100%	12	100%

Organisations

University of Cambridge

Subject Classifications

Thesaurus Keywords

Collaborating organisations

Articles

Authors

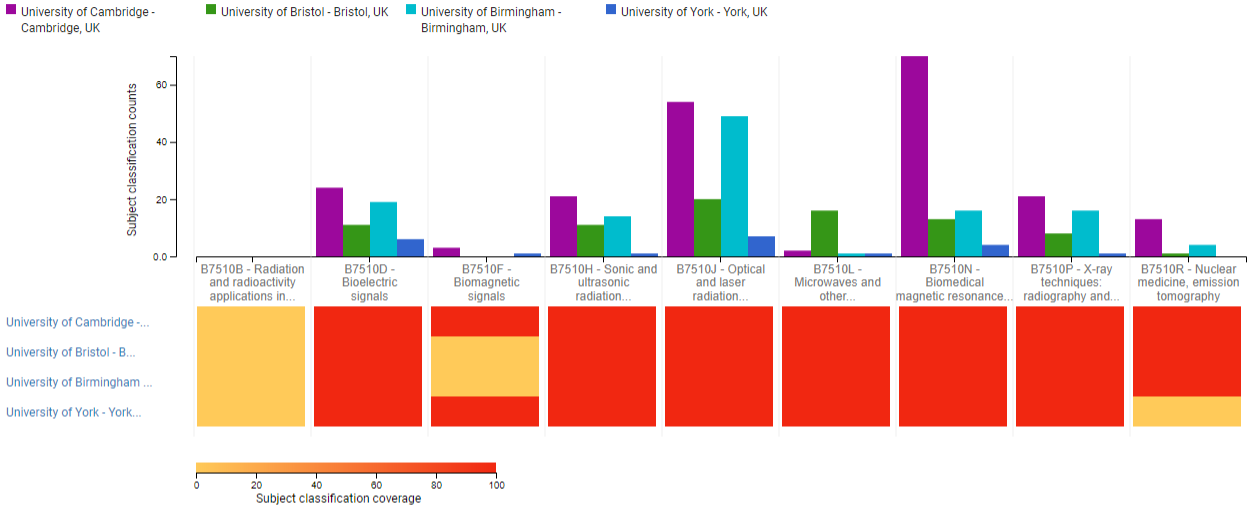
Detailed classification comparison

Controlled Terms

Detailed classification code comparison

From 2013 To 2017

Subject Classification ▸ B - Electrical engineering and electronics ▸ Instrumentation and special applications ▸ Medical physics and biomedical engineering ▸ Biomedical measurement and imaging



Organisation	Articles	Inspec Classification Counts																	
		B7510B - Radiation and radioactivity applications...		B7510D - Bioelectric signals		B7510F - Biomagnetic signals		B7510H - Sonic and ultrasonic radiation (biomedic...		B7510J - Optical and laser radiation (biomedical ...		B7510L - Microwaves and other electromagnetic wav...		B7510N - Biomedical magnetic resonance imaging an...		B7510P - X-ray techniques: radiography and comput...		B7510R - Nuclear medicine, emission tomography	
		Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage	Count	Coverage
University of Cambridge Cambridge, UK	12914	0	0%	24	100%	3	100%	21	100%	54	100%	2	100%	70	100%	21	100%	13	100%
University of Bristol Bristol, UK	5778	0	0%	11	100%	0	0%	11	100%	20	100%	16	100%	13	100%	8	100%	1	100%
University of Birmingham Birmingham, UK	4882	0	0%	19	100%	0	0%	14	100%	49	100%	1	100%	16	100%	16	100%	4	100%
University of York York, UK	2678	0	0%	6	100%	1	100%	1	100%	7	100%	1	100%	4	100%	1	100%	0	0%

Organisations
University of Cambridge
Subject Classifications
Thesaurus Keywords
Collaborating organisations
Articles
Authors
Detailed classification comparison
Controlled Terms

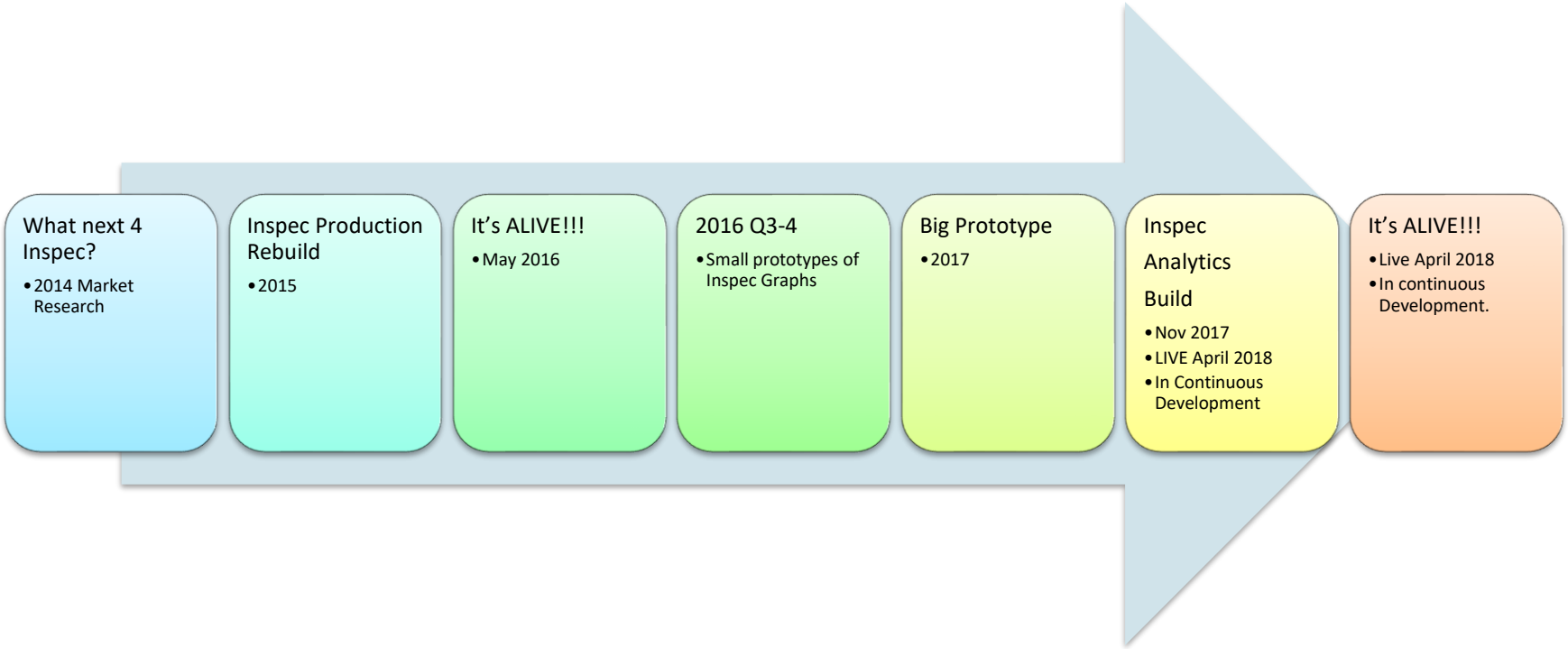
Trend Chart Galaxy Chart

Update graph 0 ✖ Reset

From 2013 To 2017

#	Subject classification	Articles
1	B7510-Biomedical measurement and imaging	225
2	B7520-Patient care and treatment	52
3	B7580-Biological engineering and techniques	40
4	B7530-Health Physics	11
5	B7550-Biomedical communication	8

The Inspec Journey




Thanks!
Q’s?