Topic modelling & visualisation of national research portfolios

RCUK Seminar, 23 October 2014

Mike Chantler, Heriot-Watt

RESEARCH PERSPECTIVES

EPSRC ICT Working Together Priority

ICT next decade, Oct 2010

ICT RESEARCH – THE NEXT DECADE REPORT

Issue date: 06 April 2011
Type: News
Related themes: Engineering, ICT

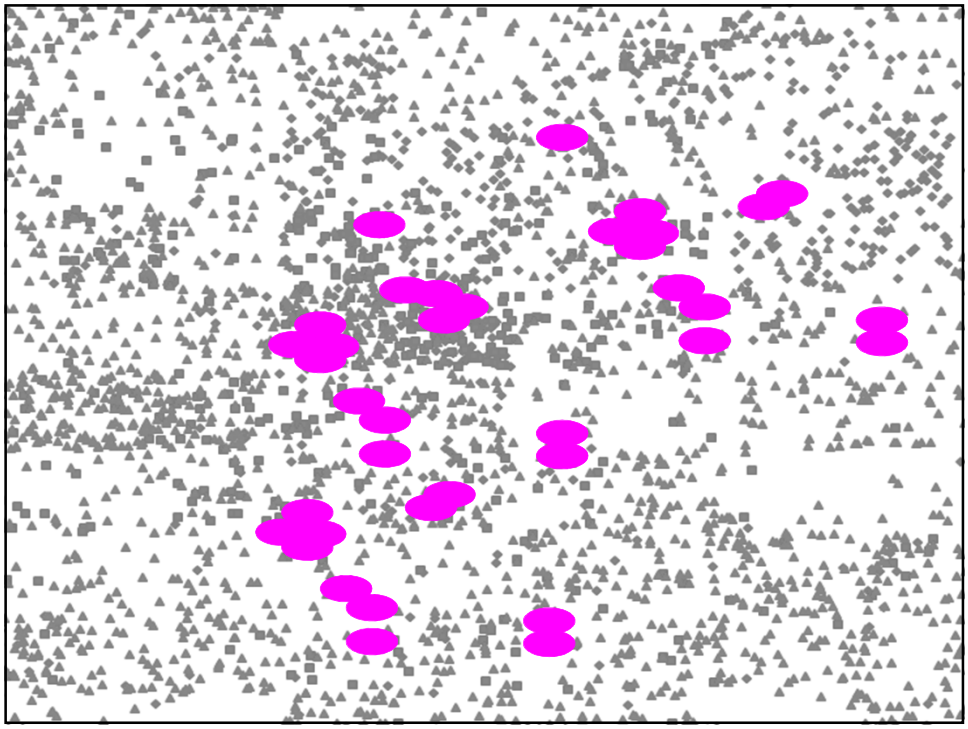
Tweet 6 Like 0 +1 0 Share 0

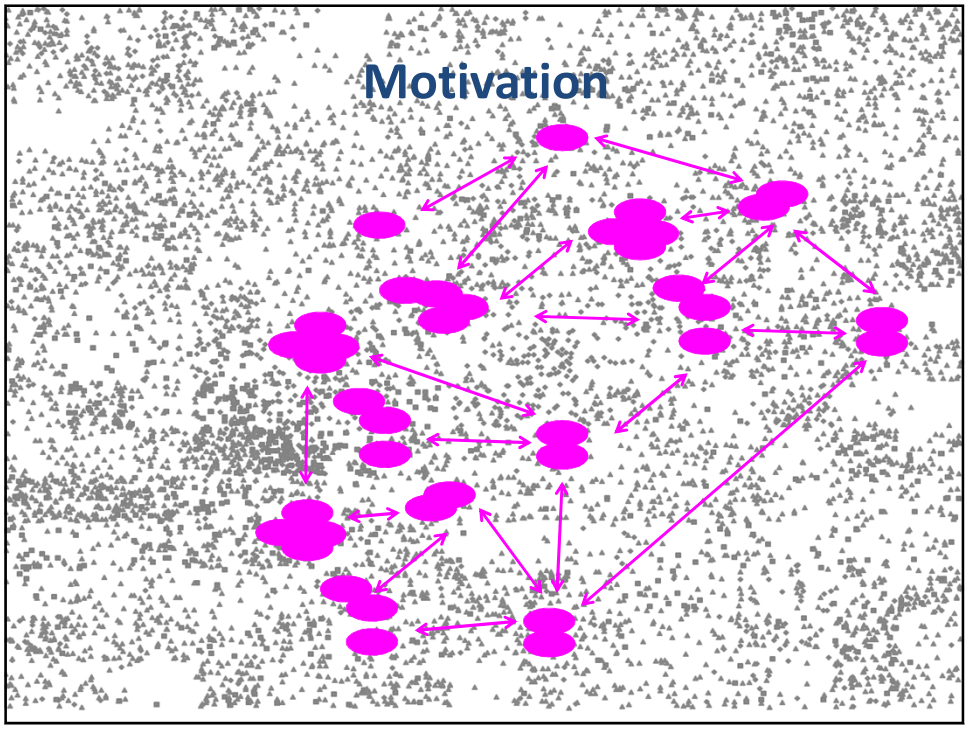
In October 2010 the ICT Programme held a workshop bringing together researchers from across the community to involve them in the development of strategy for the programme in the context of our new Delivery Plan. A report on this workshop has now been published on the website and is accompanied by a video made on the day featuring Liam Blackwell, Head of the ICT programme, and other delegates of the workshop.

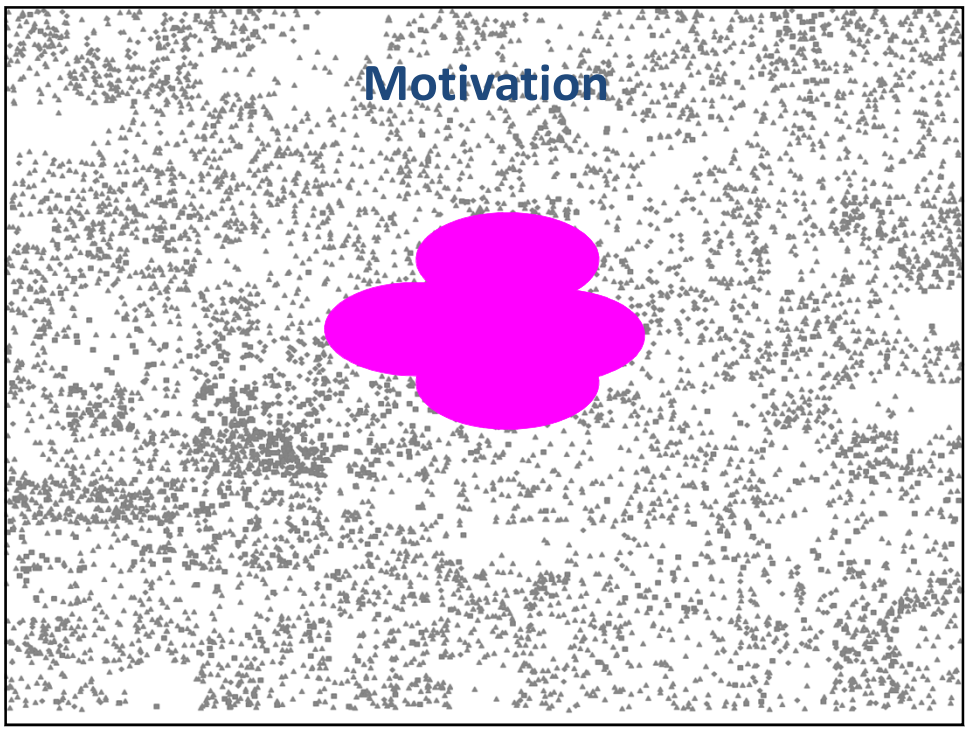
Since the workshop we have funded three community building research projects which arose as a direct consequence of the workshop. These are:

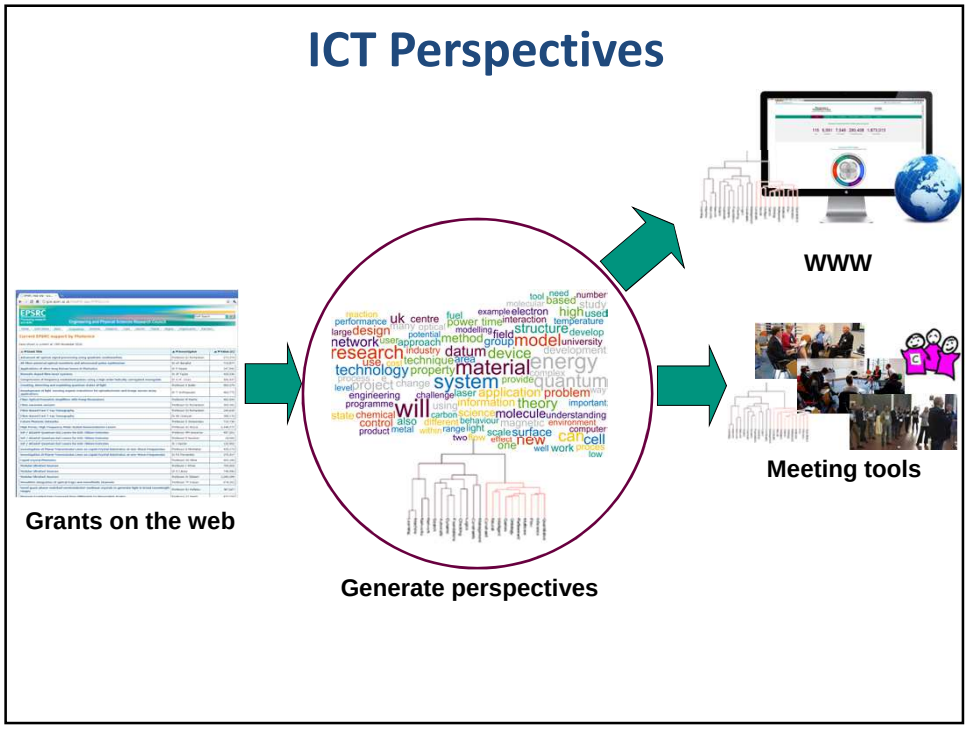
- ICT Perspectives (Mike Chantler, Heriot-Watt University)
- eFuturesXD - crossing the boundaries (Anthony O'Neil, Newcastle University)
- Framework for Ethics in ICT (Marina Jirotka, University of Oxford)

EPSRC ICT Working Together Priority




Motivation

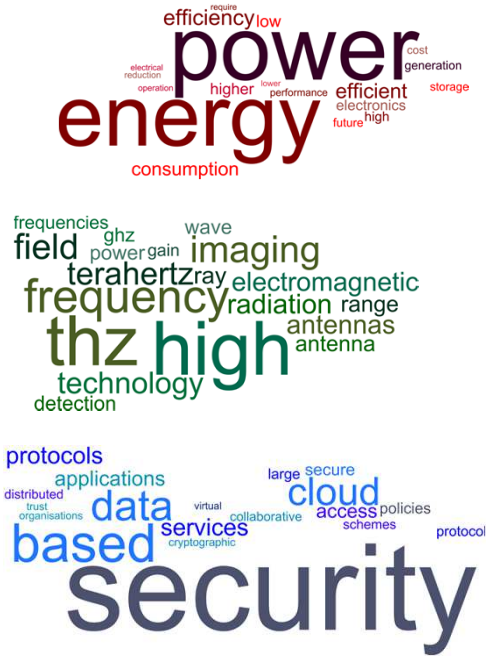

Motivation


ICT Perspectives
WWW
Meeting tools
Grants on the web
Generate perspectives

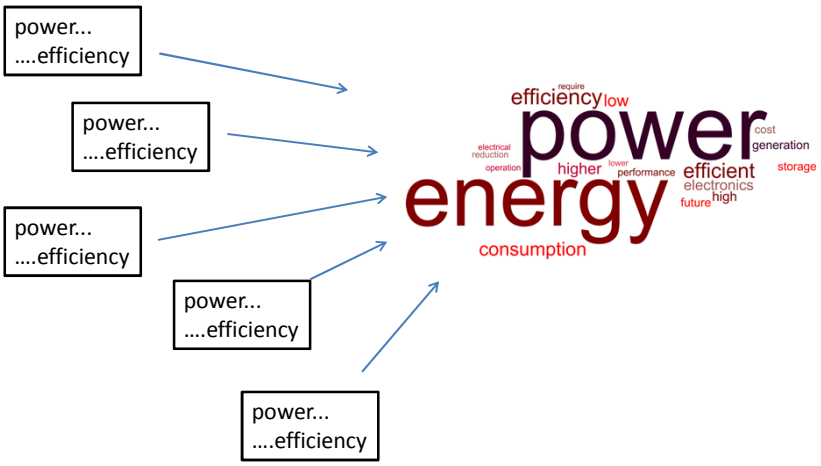
Topic modelling

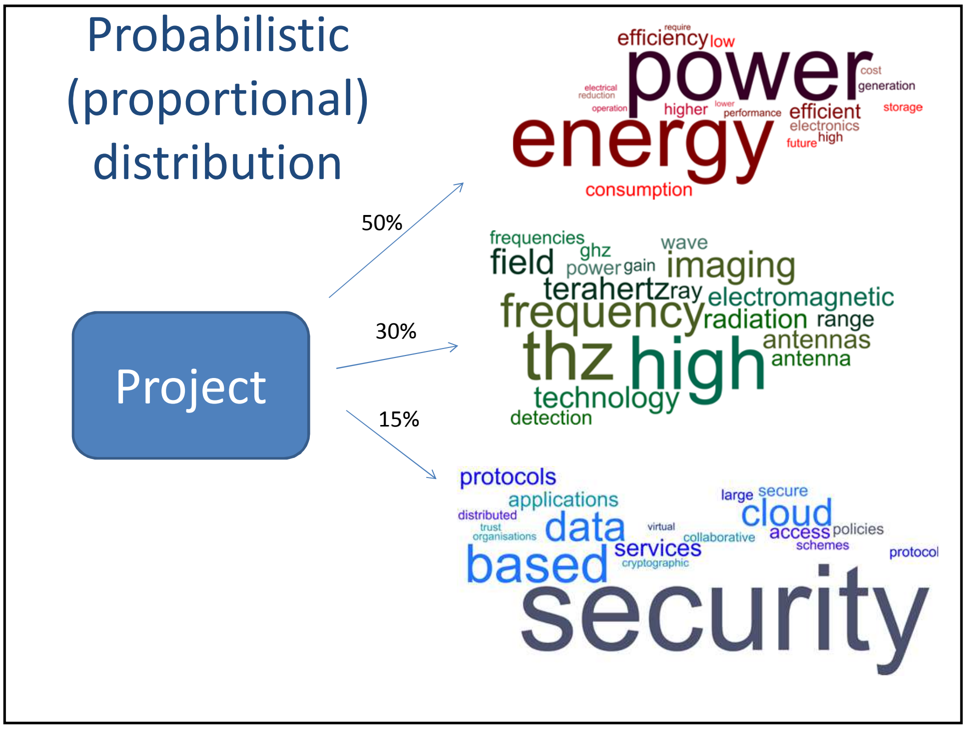
e.g. EPSRC
5591 Grants
7546 People
115 Research Areas
10 Themes

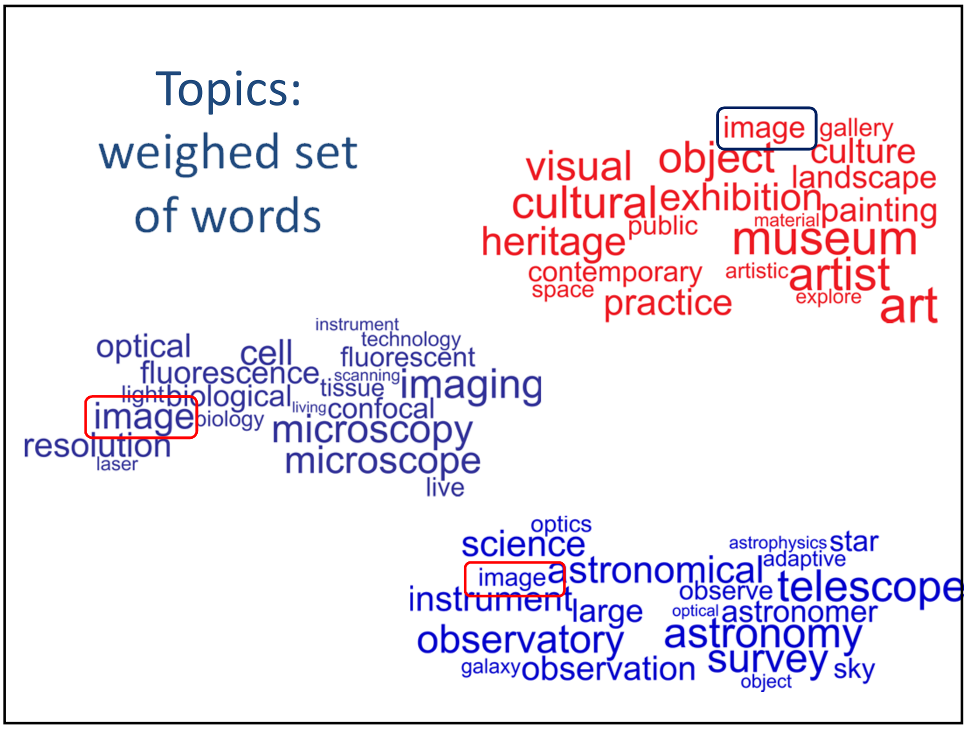
Topic modelling



How it works: Word co-occurrence




Probabilistic (proportional) distribution
Project
50%
30%
15%
efficiency low power energy higher efficient electronics generation storage future high consumption
frequencies ghz wave field power gain imaging terahertz ray electromagnetic frequency radiation range thz high antennas antenna technology detection
protocols applications large secure distributed data cloud access policies collaborative schemes services based protocol security


Topics: weighed set of words
image gallery visual object culture landscape cultural exhibition painting heritage public material museum contemporary artistic artist space explore practice art
instrument technology optical cell fluorescent fluorescence scanning tissue imaging light biological living confocal image biology microscopy resolution laser microscope live
optics science astrophysics star adaptive image astronomical instrument observe telescope large optical astronomer observatory astronomy galaxy observation survey sky object

Examples from AHRC
30 Topics


screen
adaptation
british
festival
film
image
cinema
star
interview
filmmaker
analysis
narrative
history
documentary
production
cinematic
spanish
audience
director
genre


audience
public
commercial
bbc
radio
market
technology
company
creative
cultural
television
broadcast
economy
production
industry
global
business
fashion
content
media


practice
historical
culture
question
term
focus
examine
relationship
form
experience
space
issue
understanding
understand
place
explore
contemporary
role
social
cultural

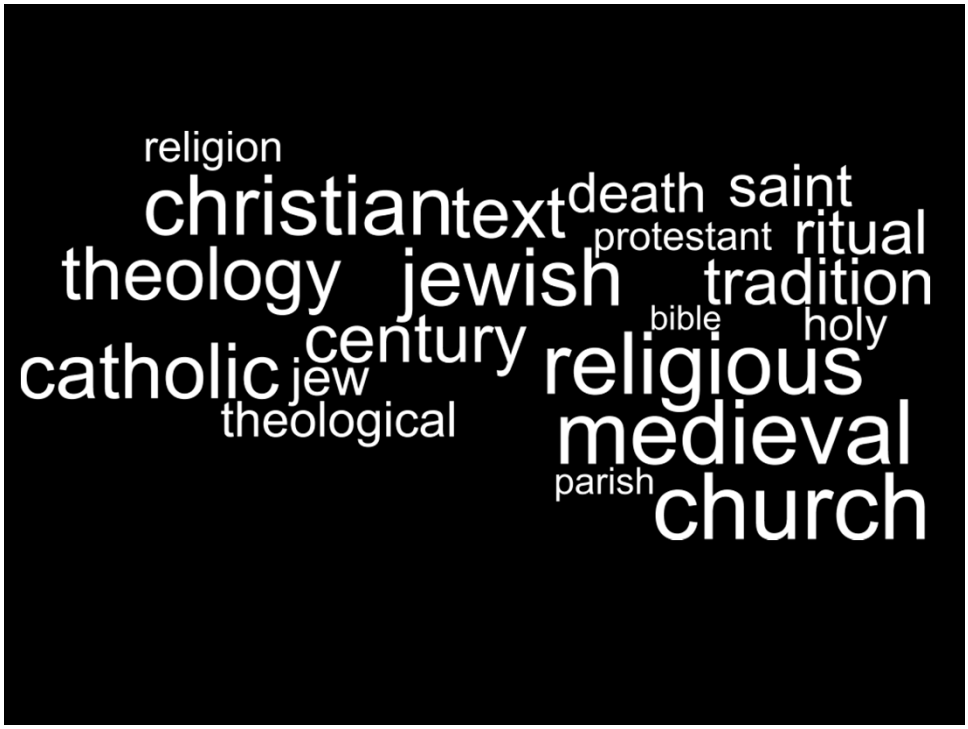

tool
software
designer
sensitive
personal
game
data
user
digital
material
web
search
detail
create
technology
textile
online
computer
design
environment

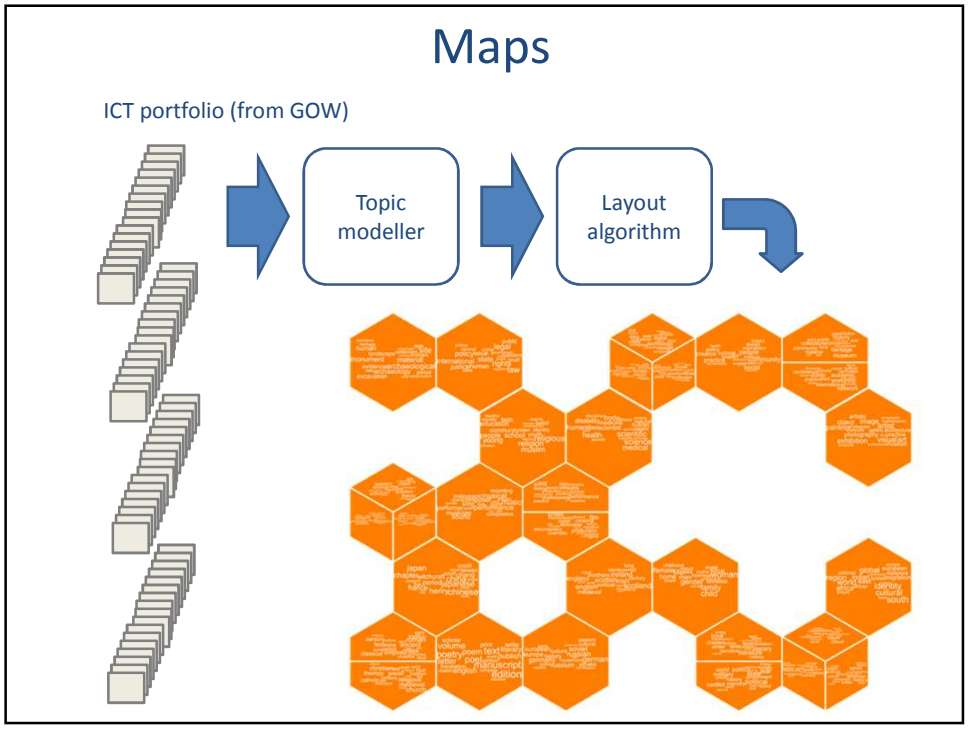

foreign
international
british
world
politics
force
war
military
state
government
power
history
political
conflict
memory
violence
policy
party
civil
national

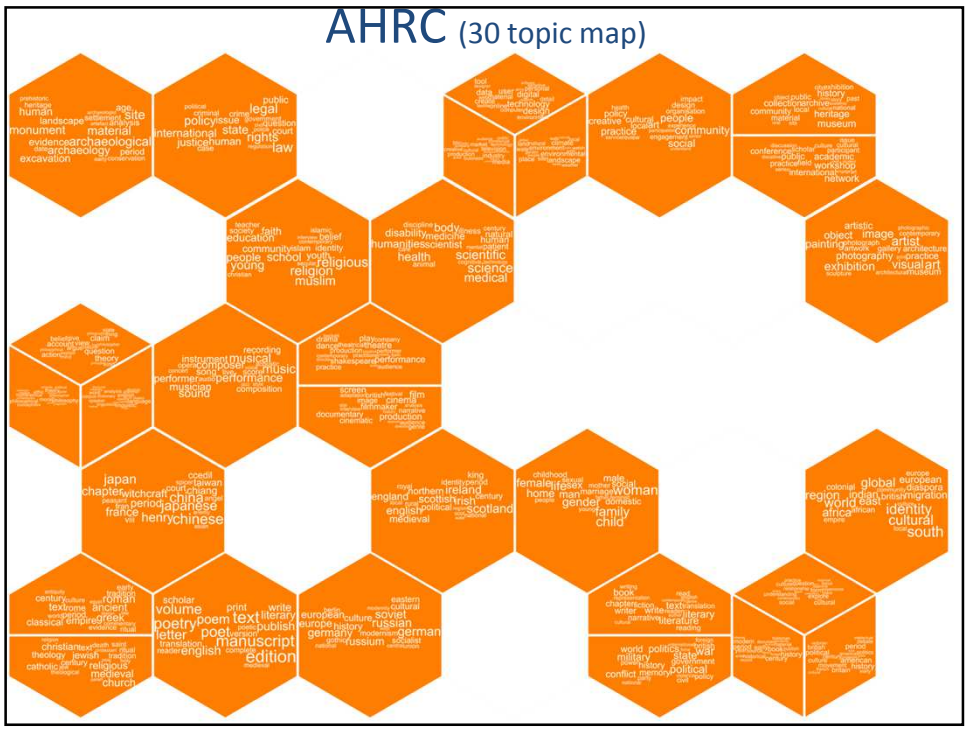

childhood
male
female
sexual
life
sex
mother
social
home
man
marriage
woman
father
feminist
people
gender
domestic
young
family
child


intellectual
debate
victorian
british
period
political
book
politics
america
century
black
nineteenth
culture
american
movement
history
french
britain
early
cultural

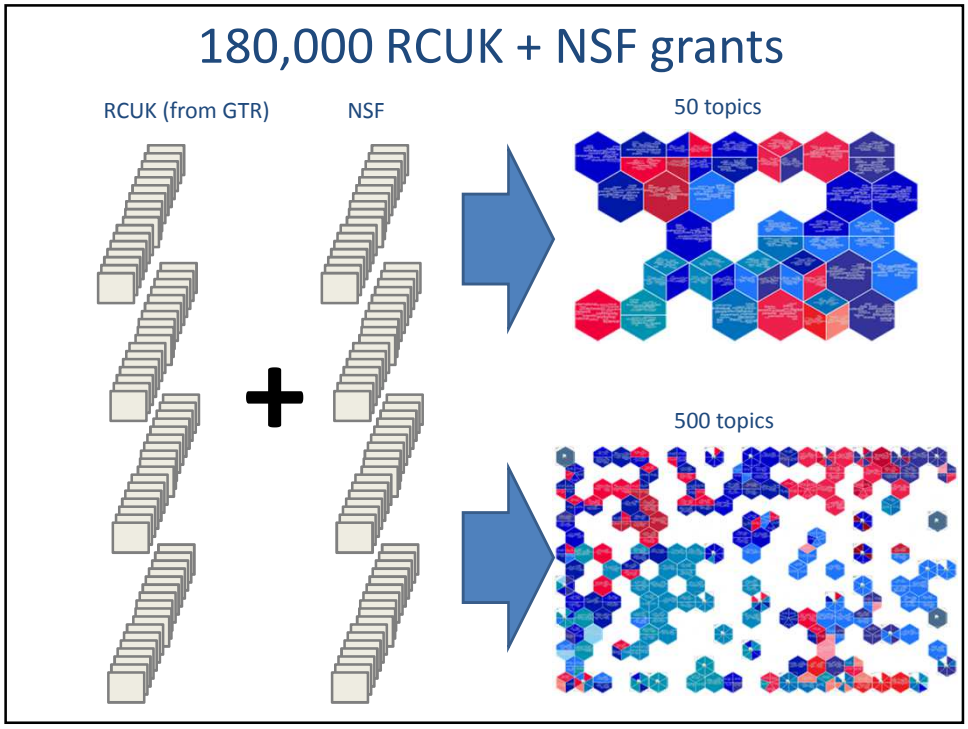

eastern
cultural
modernity
berlin
european
culture
soviet
europe
history
russian
west
germany
modernism
german
gothic
russium
socialist
national
central
union

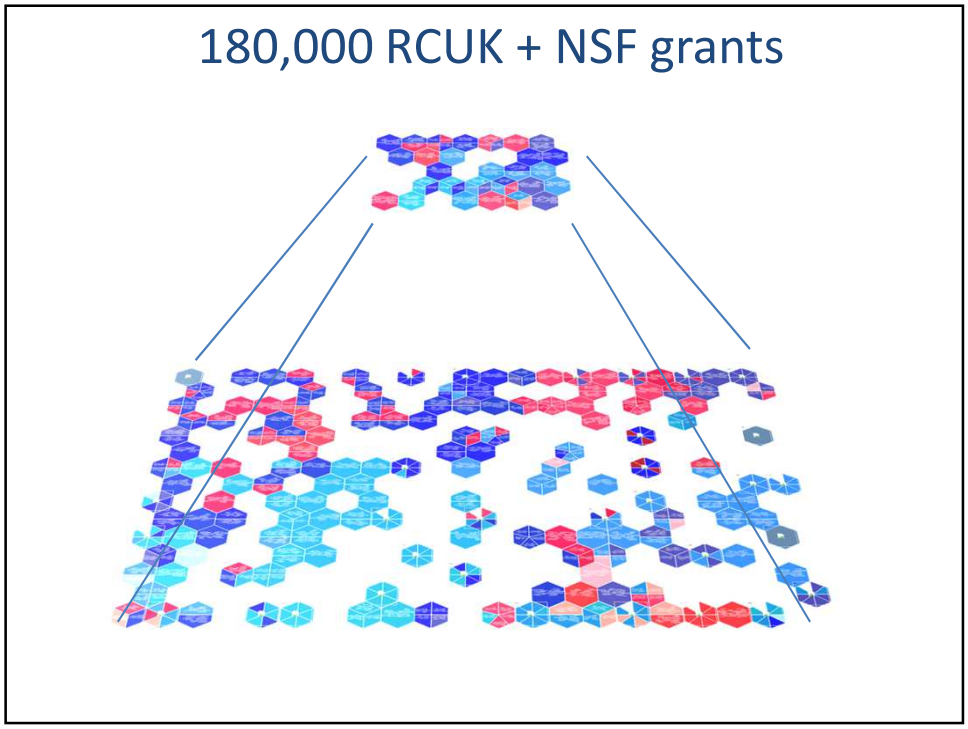

religion
christian
text
death
saint
protestant
ritual
theology
jewish
tradition
bible
holy
century
catholic
jew
religious
theological
medieval
parish
church


Maps
ICT portfolio (from GOW)
Topic modeller
Layout algorithm



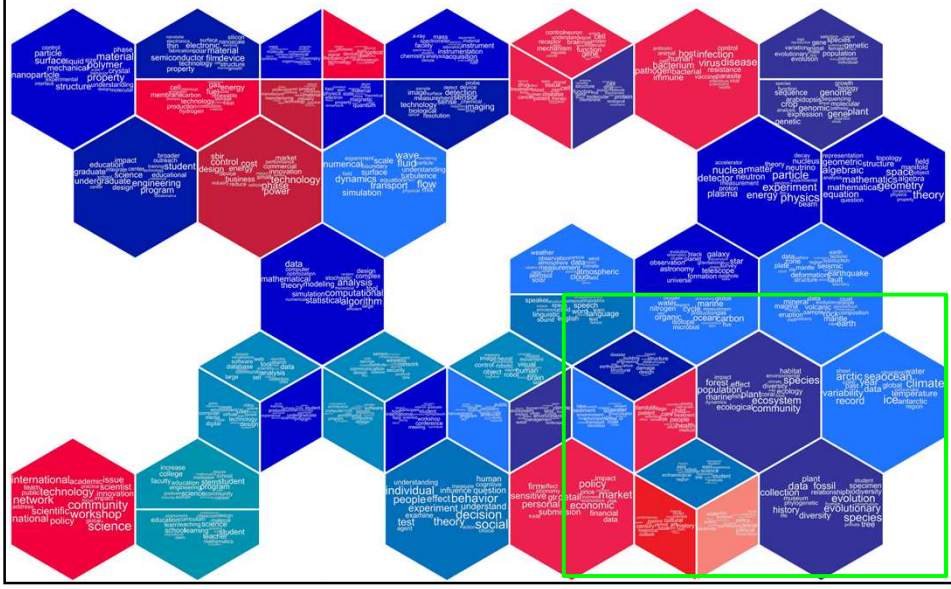
Different Granularities

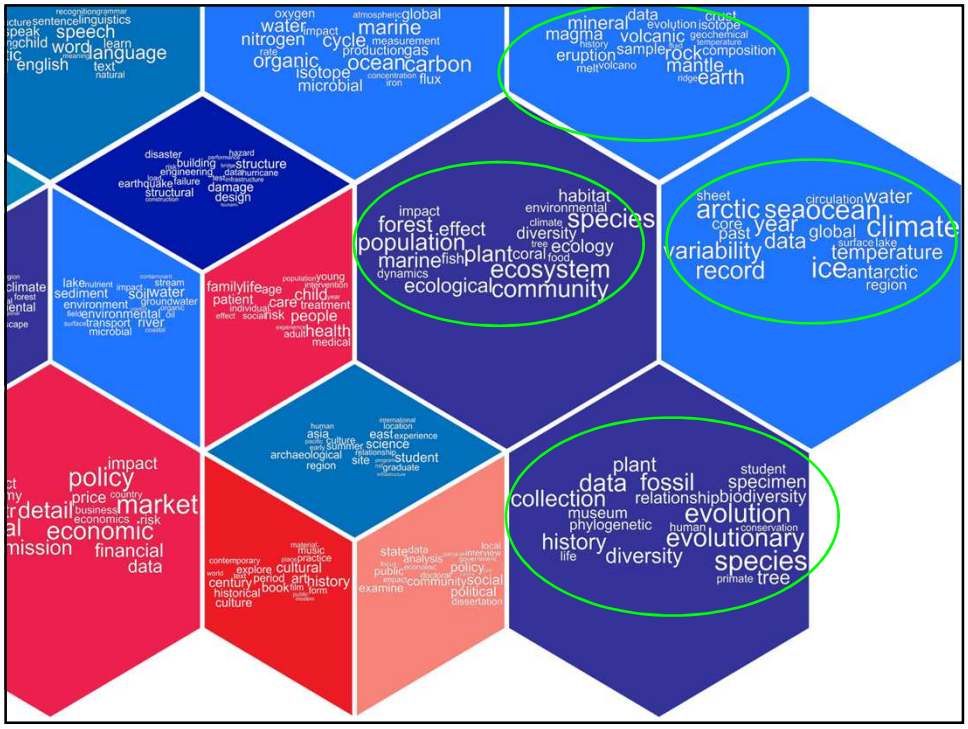

180,000 RCUK + NSF grants
RCUK (from GTR)
NSF
50 topics
500 topics

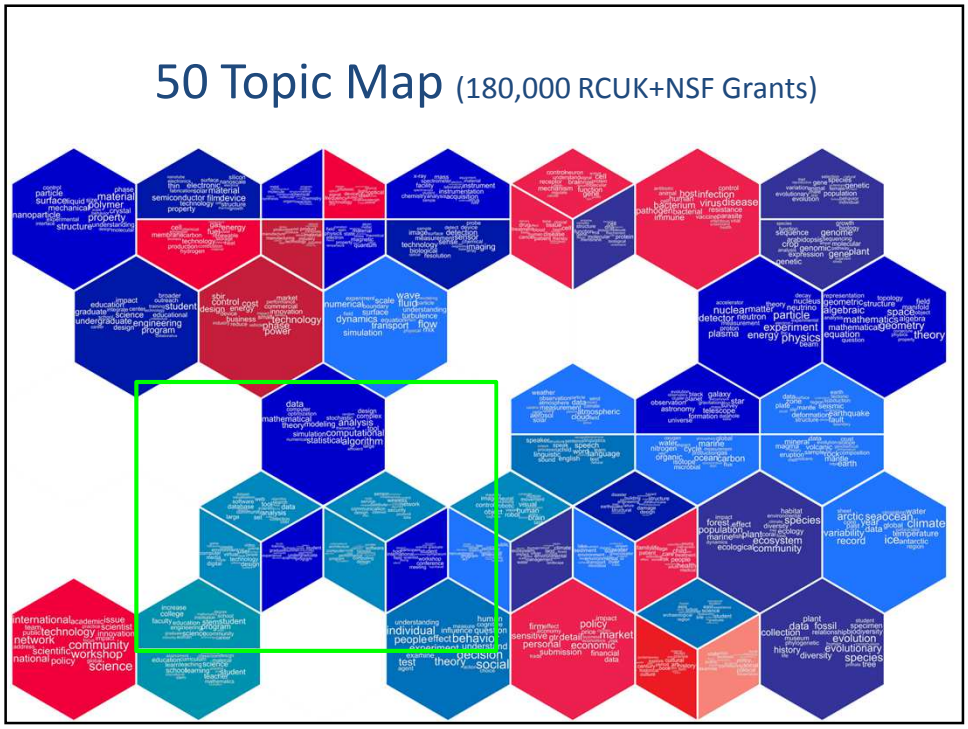

180,000 RCUK + NSF grants

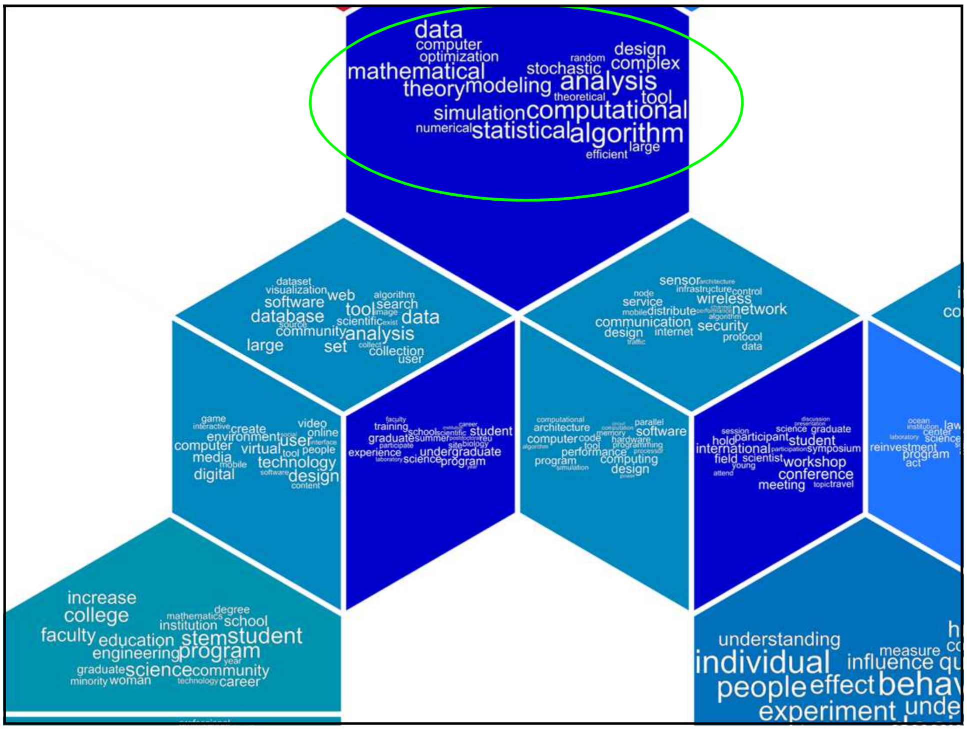
Top Level

50 Topic Map (180,000 RCUK+NSF Grants)

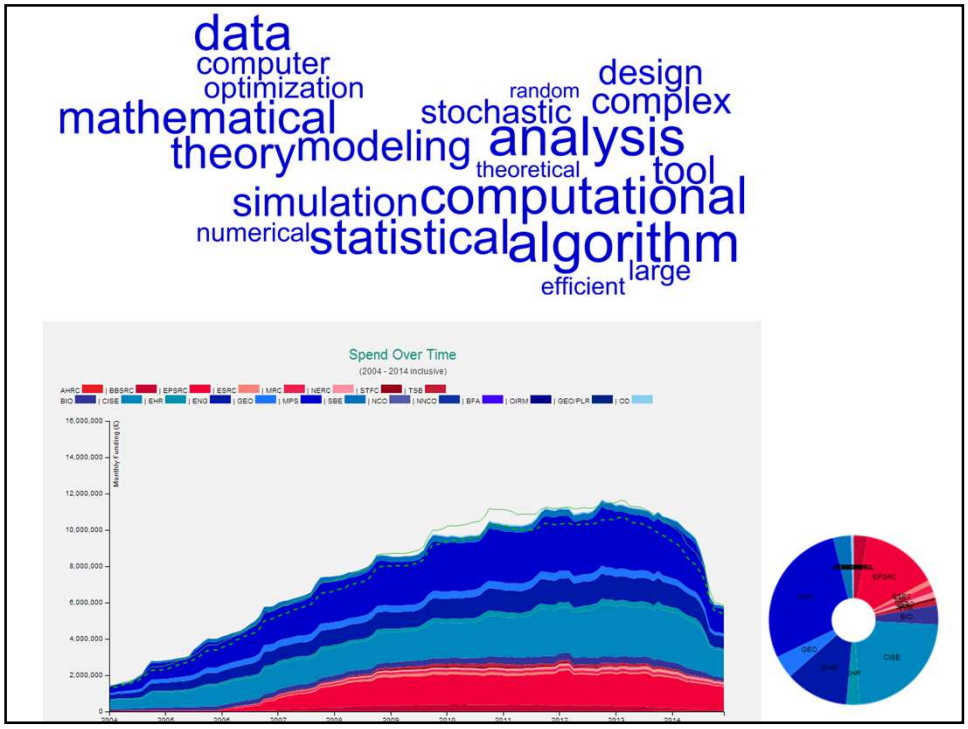



speech
language
word
learn
english
text
oxygen
water
nitrogen
cycle
marine
carbon
ocean
organic
isotope
microbial
mineral
magma
volcanic
rock
mantle
earth
eruption
data
disaster
building
structure
engineering
earthquake
damage
design
habitat
species
forest
effect
population
plant
ecology
ecosystem
community
marine
ice
arctic
sea
ocean
climate
temperature
variability
record
water
lake
soil
sediment
environment
river
family
child
care
people
health
impact
policy
market
economic
financial
data
asia
culture
science
archaeological
region
student
plant
fossil
specimen
collection
museum
phylogenetic
evolution
evolutionary
species
history
diversity
tree
music
practice
cultural
century
period
arthistory
historical
culture
data
analysis
policy
community
social
political
examine

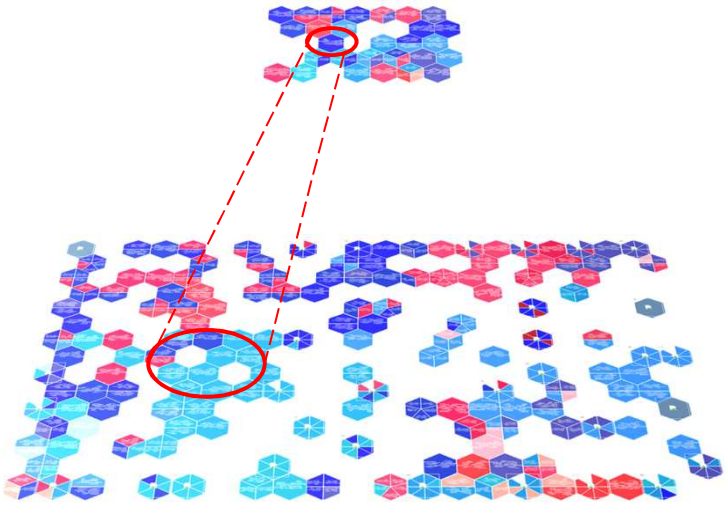

50 Topic Map (180,000 RCUK+NSF Grants)


data
computer
optimization
design
random
complex
mathematical
stochastic
theory
modeling
analysis
theoretical
tool
simulation
computational
numerical
statistical
algorithm
large
efficient
increase
college
faculty
education
stem
student
engineering
program
science
community
career
understanding
individual
people
effect
experiment

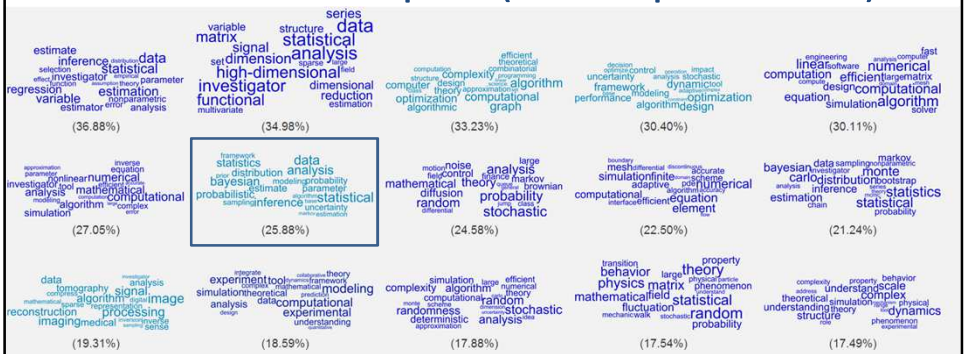
Drill Down

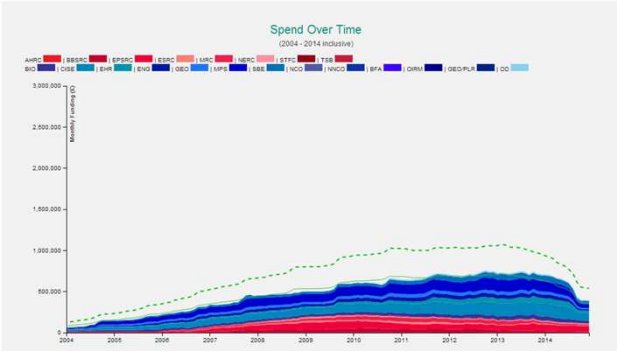


Drill Down to sub-topics

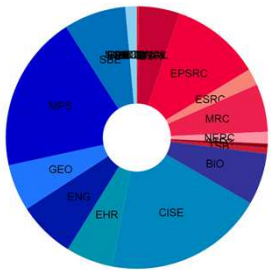


Related Sub-topics (500 topic model)

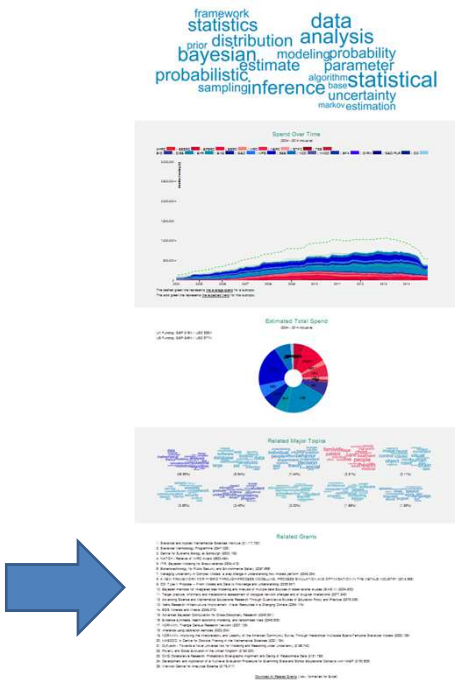




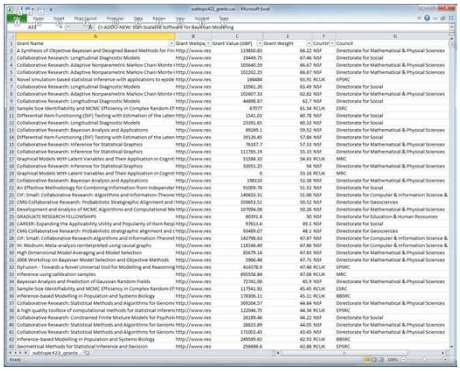
Monthly spend: last 10 years



Distribution across RC & NSF Directorates



Drill down to related grants



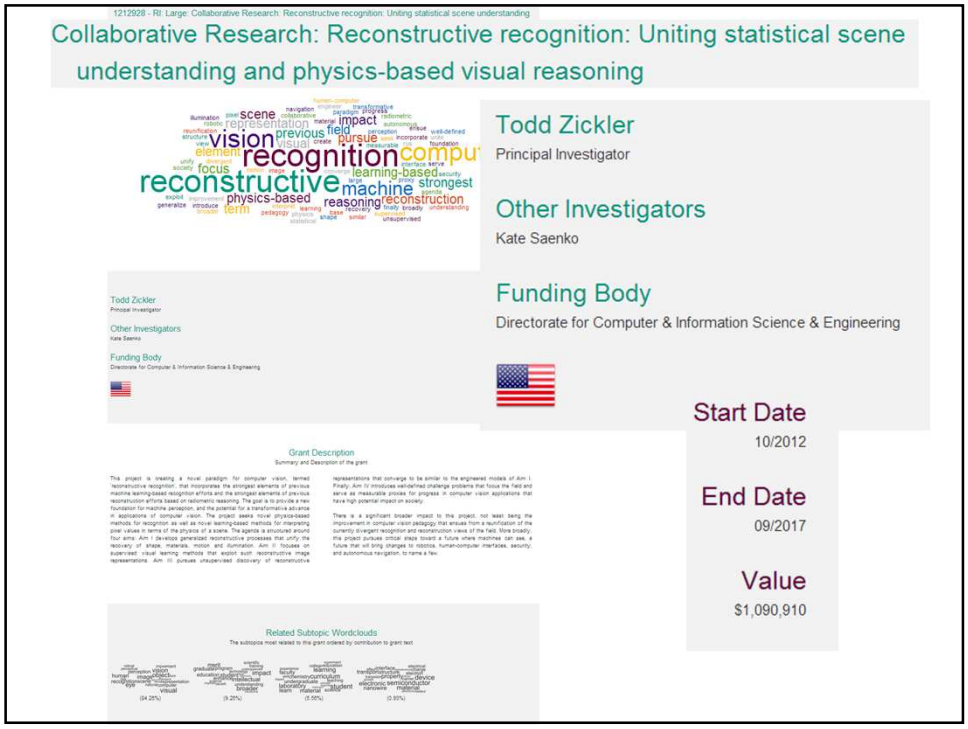
Downloadable CSV

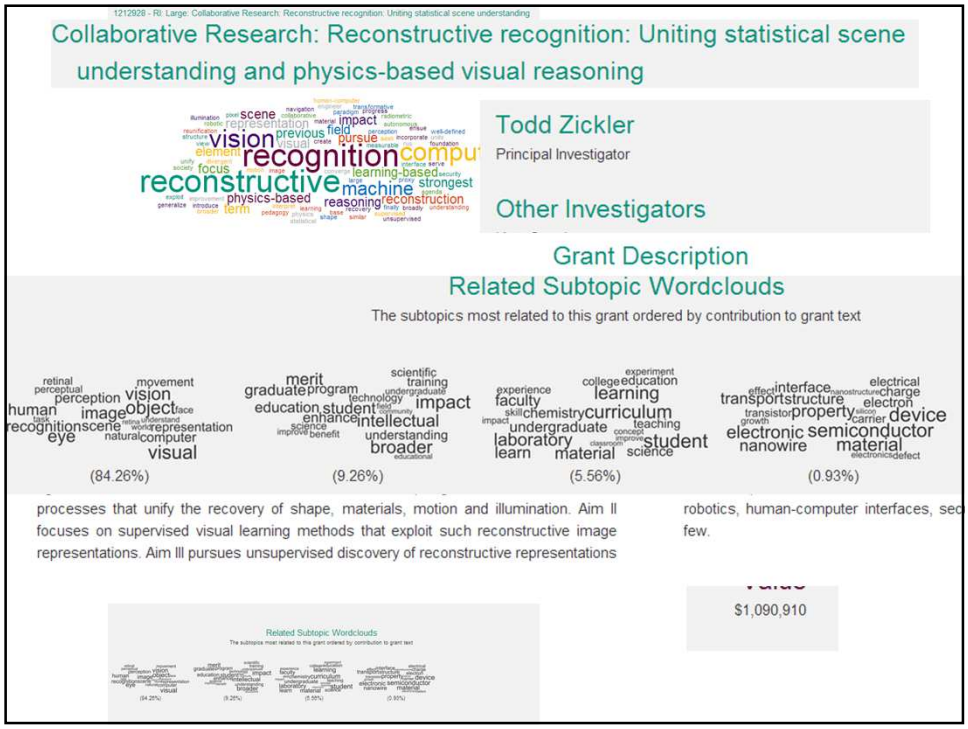
Drill down to related grants

Sub-topic

Related Grants

1. RI: Large: Collaborative Research: Reconstructive recognition: Uniting statistical scene understanding and physics-based visual reasoning (84.26%)
2. Investigating binocular coordination during viewing of stereoscopic images. (81.82%)
3. RI: Large: Collaborative Research: Reconstructive recognition: Uniting statistical scene understanding and physics-based visual reasoning (79.63%)
4. RI: Small: Collaborative Research: Detecting Abnormalities in Images (78.86%)
5. RI: Large: Collaborative Research: Reconstructive recognition: Uniting statistical scene understanding and physics-based visual reasoning (76.85%)
6. RI: Small: Collaborative Research: Detecting Abnormalities in Images (72.36%)
7. Collaborative Research: From Edge Pixels to Recognition of Parts of Object Contours (70.75%)
8. The role of self-cueing in visual organization (70.67%)
9. An investigation of fixational eye movement patterns during three-dimensional object recognition: (70.49%)
10. Collaborative Research: From Edge Pixels to Recognition of Parts of Object Contours (70.07%)
11. CAREER: Object Recognition with Hierarchical Models (69.39%)
12. Collaborative Research: Visual Cortex on Silicon (68.48%)
13. Collaborative Research: Visual Cortex on Silicon (68.48%)
14. RI: Small: Collaborative Research: A Scalable Architecture for Image Interpretation (68.24%)
15. Collaborative Research: Visual Cortex on Silicon (68.12%)
16. Collaborative Research: RI: Small: A Scalable Architecture for Image Interpretation (67.81%)
17. Collaborative Research: Visual Cortex on Silicon (67.39%)
18. Collaborative Research: Visual Cortex on Silicon (67.39%)
19. Collaborative Research: Visual Cortex on Silicon (67.39%)


1212928 - RI: Large: Collaborative Research: Reconstructive recognition: Uniting statistical scene understanding
Collaborative Research: Reconstructive recognition: Uniting statistical scene understanding and physics-based visual reasoning
Todd Zickler
Principal Investigator
Other Investigators
Kate Saenko
Funding Body
Directorate for Computer & Information Science & Engineering
Start Date
10/2012
End Date
09/2017
Value
$1,090,910
Grant Description
Related Subtopic Wordclouds

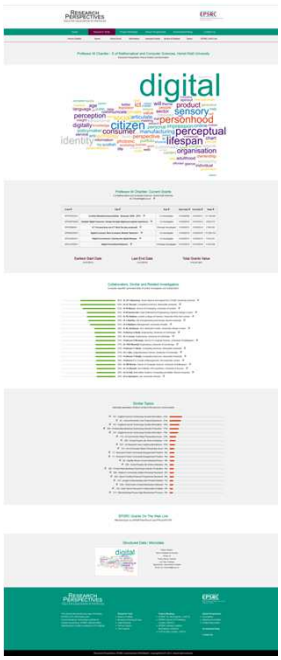

1212928 - RI: Large: Collaborative Research: Reconstructive recognition: Uniting statistical scene understanding
Collaborative Research: Reconstructive recognition: Uniting statistical scene understanding and physics-based visual reasoning
Todd Zickler
Principal Investigator
Other Investigators
Grant Description
Related Subtopic Wordclouds
The subtopics most related to this grant ordered by contribution to grant text
(84.26%)
(9.26%)
(5.56%)
(0.93%)
processes that unify the recovery of shape, materials, motion and illumination. Aim II focuses on supervised visual learning methods that exploit such reconstructive image representations. Aim III pursues unsupervised discovery of reconstructive representations
robotics, human-computer interfaces, sec
few.
$1,090,910

Investigators

1. Visualisation
2. Grants
3. Similar investigators / collaborators
4. Similar topics
5. GOW link
6. Meta Data

Requests by universities

http://researchperspectives.org



Search by Abstract

Learning to Recognise Dynamic Visual Content from Broadcast Footage

This research is in the area of computer vision - making computers which can understand what is happening in photographs and video. As humans we are fascinated by other humans, and capture endless images of their activities, for example home movies of our family on holiday, video of sports events or CCTV footage of people in a town center. A computer capable of understanding what people are doing in such images would be able to do many jobs for us, for example finding clips of our children waving, fast forwarding to a goal in a football game, or spotting when someone starts a fight in the street. For Deaf people, who use a language combining hand gestures with facial expression and body language, a computer which could visually understand their actions would allow them to communicate in their native language. While humans are very good at understanding what people are doing (and can learn to understand special actions such as sign language), this has proved extremely challenging for computers.Much work has tried to solve this problem, and works well in particular settings for example the computer can tell if a person is walking so long as they do it clearly and face to the side, or can understand a few sign language gestures as long as the signer cooperates and signs slowly. We will investigate better models for recognising activities by teaching the computer by showing it many example videos. To make sure our method works well for all kinds of setting we will use real world video from movies and TV. For each video we have to tell the computer what it represents, for example throwing a ball or a man hugging a woman . It would be expensive to collect and label lots of videos in this way, so instead we will extract approximate labels automatically from subtitle text and scripts which are available for TV. Our new methods will combine learning from lots of approximately labelled video (cheap because we get the labels automatically), use of contextual information such as which actions people do at the same time, or how one action leads to another (he hits the man, who falls to the floor), and computer vision methods for understanding the pose of a person (how they are standing), how they are moving, and the objects which they are using.By having lots of video to learn from, and methods for making use of approximate labels, we will be able to make stronger and more flexible models of human activities. This will lead to recognition methods which work better in the real world and contribute to applications such as interpreting sign language and automatically tagging video with its content.

Search by Abstract

Portfolio Similarity Search

ase copy and paste an abstract you would like to search for similarities in the EPSRC portfolio

Computer vision algorithms for individual tasks such as objectrecognition, detection and segmentation has now reac
m of scene understanding. The problem of scene understanding involves explaining thewhole image by recognizing
e the research will be the problem of automated understanding of cities from video usingcomputer vision, inspired by
elp/maps/streetview/, Yotta http://www.yotta.tv/index.php (who have agreed to supply OxfordBrookes with data) and
s: a van drivesaround the roads of the UK, in the van are GPS equipment and multiple calibrated cameras, synchron
ze objects of interest in the video, fromroad signs and other street furniture, to particular buildings, to allow them to be
ng for indoor scenes such as home or office, with video taken froma normal camera and Z-cam.

Search by Abstrat

Result – similar grants

Relevance	Title	Principal Investigator
92%	Learning to Recognise Dynamic Visual Content from Broadcast Footage	Bowden, R (Professor)
52%	inTouch: A video link system to improve social inclusion for people with dementia	Harris, N (Dr)
47%	Doctoral Training Centre in Digital Music and Media for the Creative Economy	Sandler, M (Professor)
44%	Illuminating the Path of Video Visualization	Chen, M (Professor)
44%	Advanced Formal Verification Techniques for Heterogeneous Multi-core Programming	Donaldson, AF (Dr)
43%	Scene Understanding using New Global Energy Models	Torr, PH (Professor)
42%	Liberating housebound obese individuals using augmented virtual reality	Darzi, AW (Professor)
42%	Enabler for Next-Generation Mobile Video Applications	Wang, Q (Dr)
42%	COMPPACT: Compression of Video using Perceptually Optimised Parametric Coding Techniques	Bull, D (Professor)
41%	EPSRC Network on Vision and Language (V&L Net)	Belz, A (Dr)

Result – similar persons

Relevance	Investigator		Department	Organization
91%	Bowden, R (Professor)		University of Surrey	Vision Speech and Signal Proc CVSSP
52%	Harris, N (Dr)		University of Bath	Department for Health
52%	Jones, RW (Professor)		University of Bath	Department for Health
52%	Panteli, N (Dr)		University of Bath	School of Management
51%	Plumbley, M (Professor)		Queen Mary, University of London	Sch of Electronic Eng & Computer Science
49%	Healey, PGT (Professor)		Queen Mary, University of London	Sch of Electronic Eng & Computer Science
48%	Baddeley, R (Dr)		University of Bristol	Psychology Dept.
46%	Chen, M (Professor)		University of Oxford	Oxford e-Research Centre
46%	Sandler, M (Professor)		Queen Mary, University of London	Sch of Electronic Eng & Computer Science
45%	Sadrzadeh, M (Dr)		Queen Mary, University of London	Sch of Electronic Eng & Computer Science
45%	Bull, D (Professor)		University of Bristol	Electrical and Electronic Engineering
45%	Waller, A (Professor)		University of Dundee	School of Computing
45%	Jones, NS (Dr)		Imperial College London	Dept of Mathematics
44%	Thornton, IM (Professor)		Swansea University	School of Health Science
44%	Nikolopoulos, D (Professor)		Queen's University of Belfast	Electronics Electrical Eng and Comp Sci
43%	Kroening, D (Professor)		University of Oxford	Computer Science

Result – similar Research Areas

Relevance	Research Area
56%	Human Communication in ICT
56%	Image and Vision Computing
56%	Graphics and Visualisation
54%	Natural Language Processing
51%	Architecture and Operating Systems
50%	Verification and Correctness
50%	Software Engineering

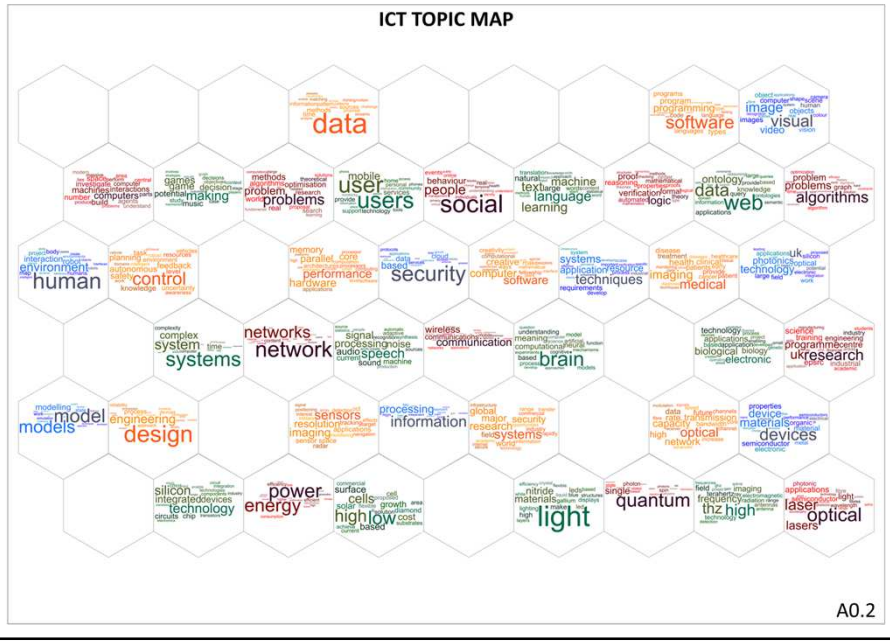
Example:
Network of Networks

ICT Network of Networks

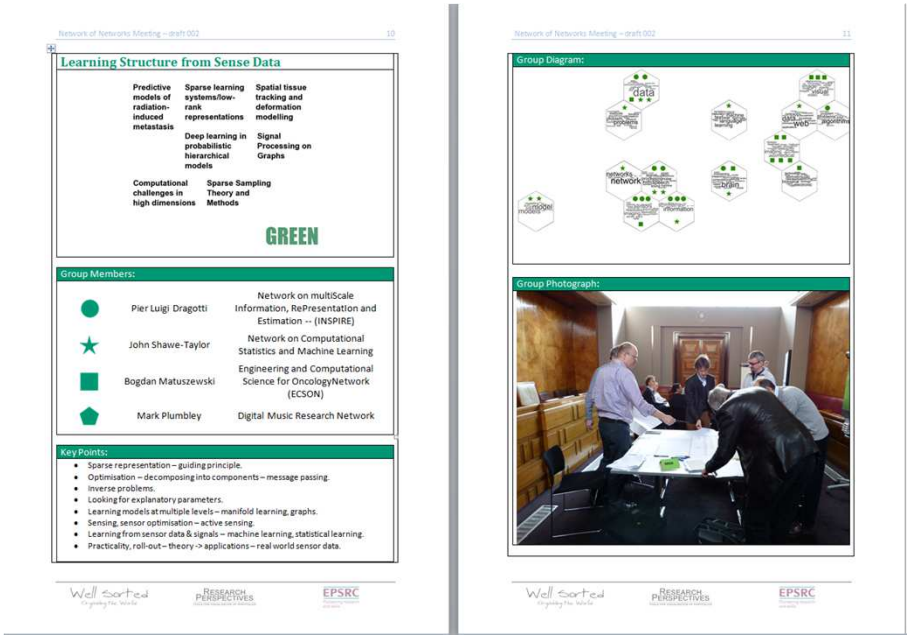
Portfolio coverage?

http://www.researchperspectives.org/meetings/NoN2014/

ICT Example



Learning Structure from Sense Data



Learning Structure from Sense Data

Predictive models of radiation-induced metastasis	Sparse learning systems/low-rank representations	Spatial tissue tracking and deformation modelling
	Deep learning in probabilistic hierarchical models	Signal Processing on Graphs
Computational challenges in high dimensions	Sparse Sampling Theory and Methods	

GREEN

Group Members:

Pier Luigi Dragotti	Network on multiScale Information, RePresentation and Estimation -- (INSPIRE)
John Shawe-Taylor	Network on Computational Statistics and Machine Learning
Bogdan Matuszewski	Engineering and Computational Science for OncologyNetwork (ECSON)
Mark Plumbley	Digital Music Research Network

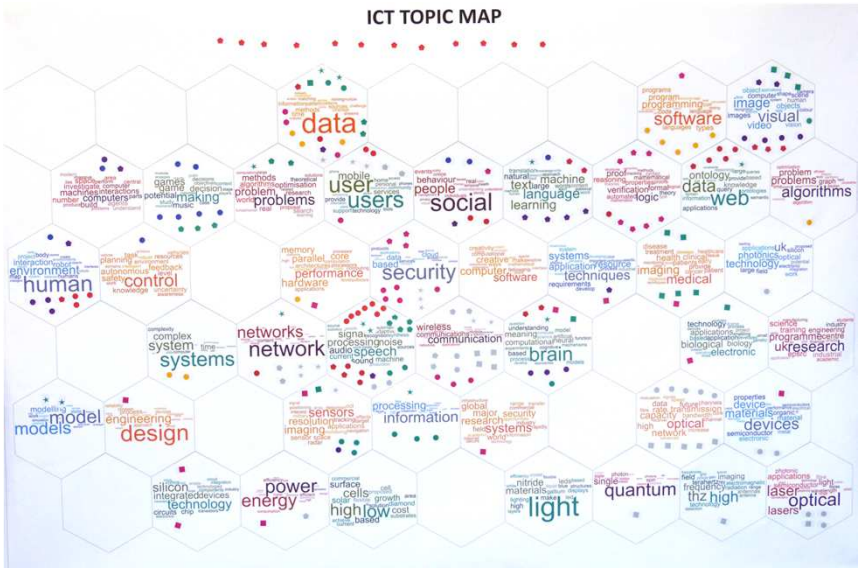
Key Points:

- Sparse representation – guiding principle.
- Optimisation – decomposing into components – message passing.
- Inverse problems.
- Looking for explanatory parameters.
- Learning models at multiple levels – manifold learning, graphs.
- Sensing, sensor optimisation – active sensing.
- Learning from sensor data & signals – machine learning, statistical learning.
- Practicality, roll-out – theory -> applications – real world sensor data.

Group Diagram:

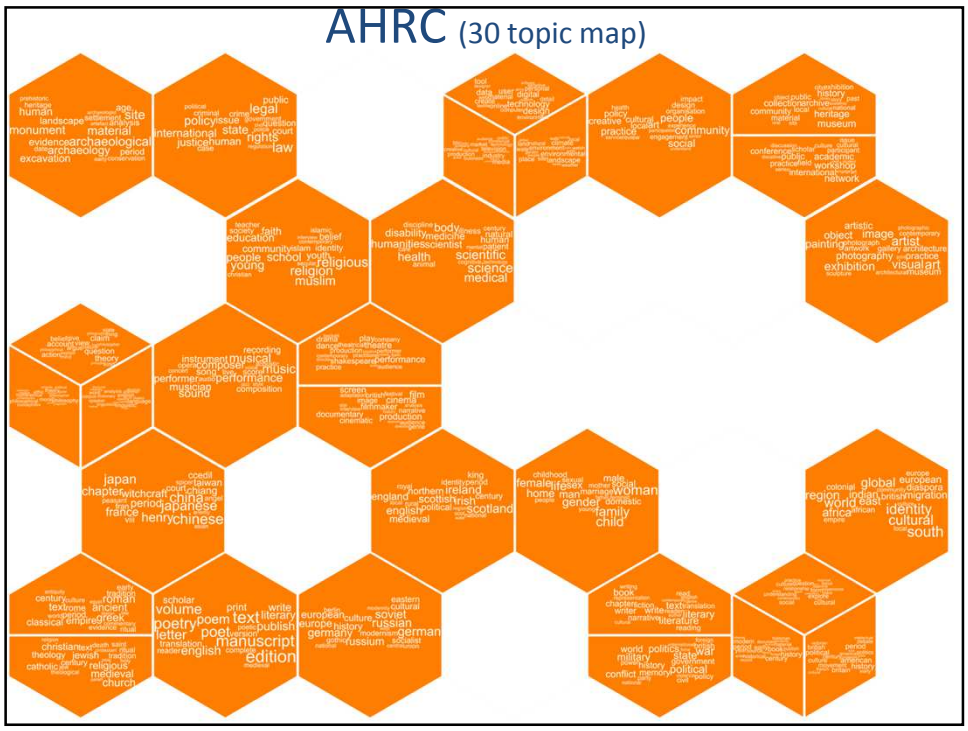
Group Photograph:

Networks' Coverage

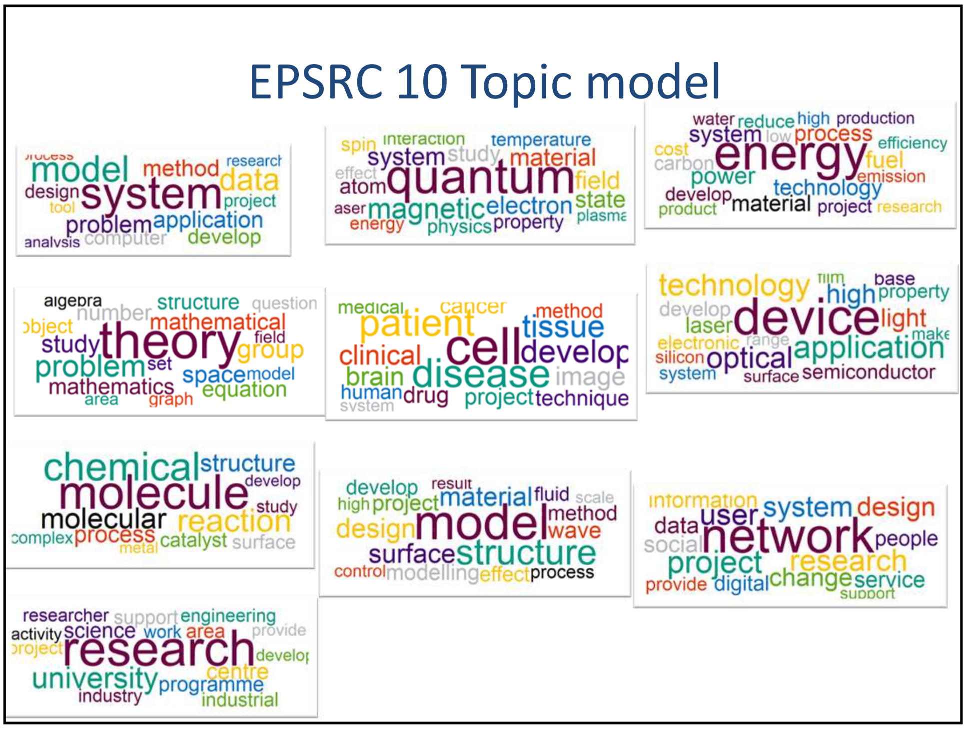


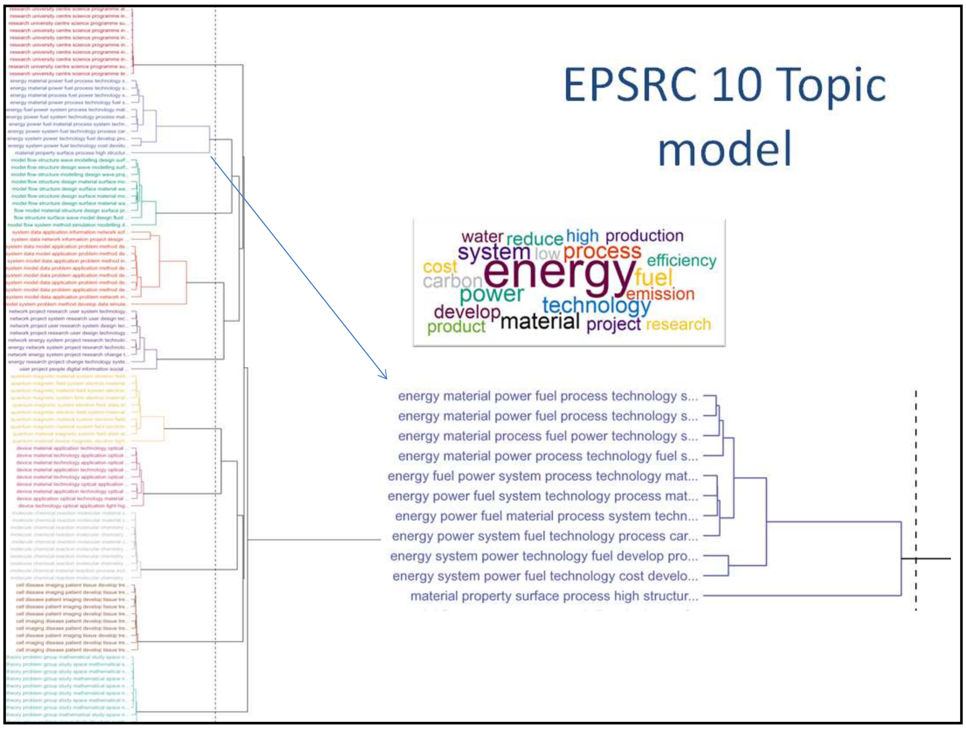
Example: AHRC

AHRC (30 topic map)



Example:
Abstract (10 topic) model of EPSRC


EPSRC 10 Topic model


EPSRC 10 Topic model
energy material power fuel process technology s...
energy material power fuel process technology s...
energy material process fuel power technology s...
energy material power process technology fuel s...
energy fuel power system process technology mat...
energy power fuel system technology process mat...
energy power fuel material process system techn...
energy power system fuel technology process car...
energy system power technology fuel develop pro...
energy system power fuel technology cost develo...
material property surface process high structur...

Example: Comparing Research Communities

CHI, CSCW, GROUP, ECSCW

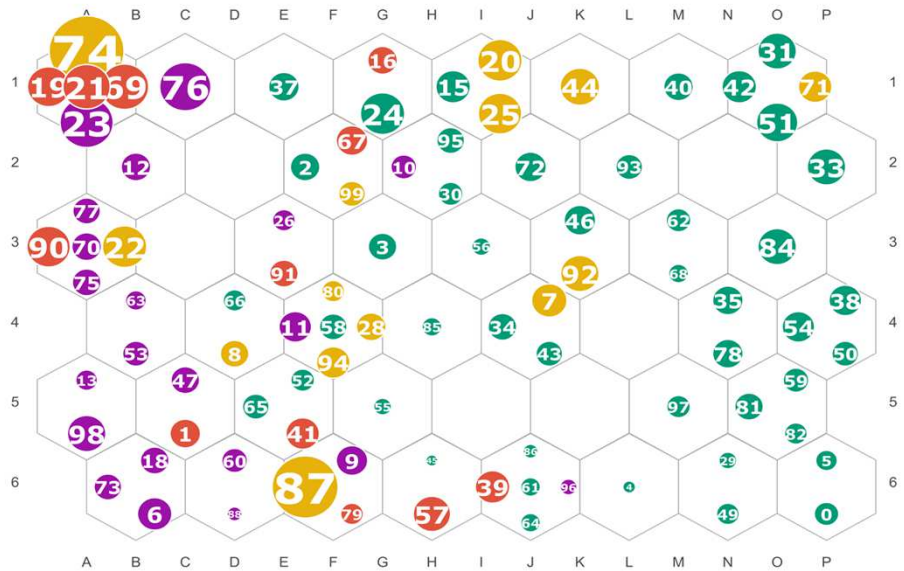
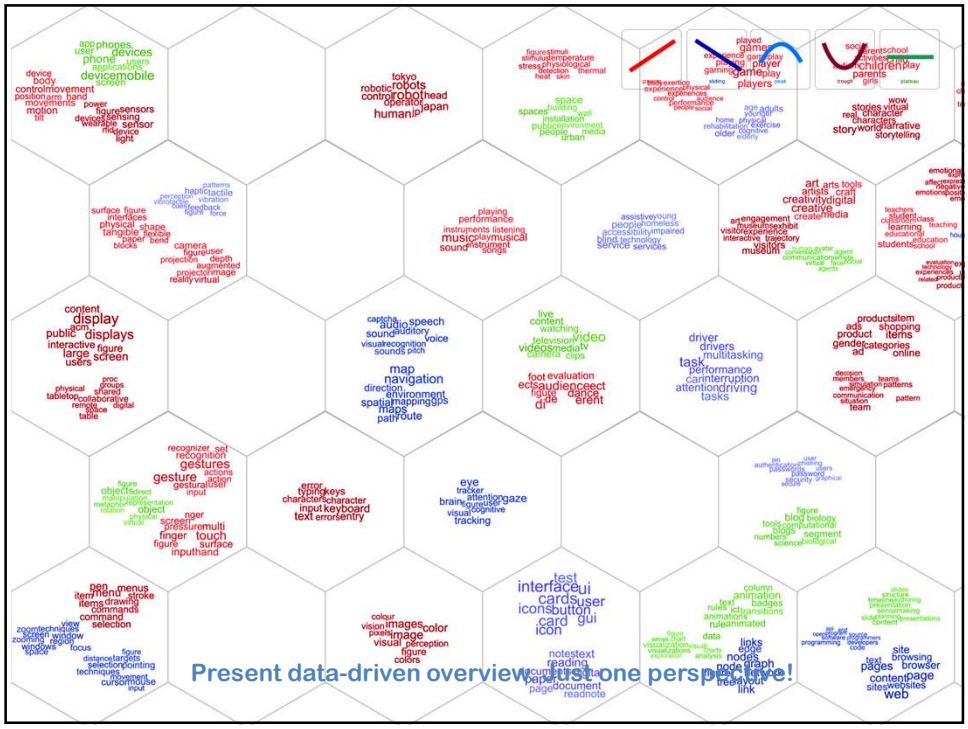
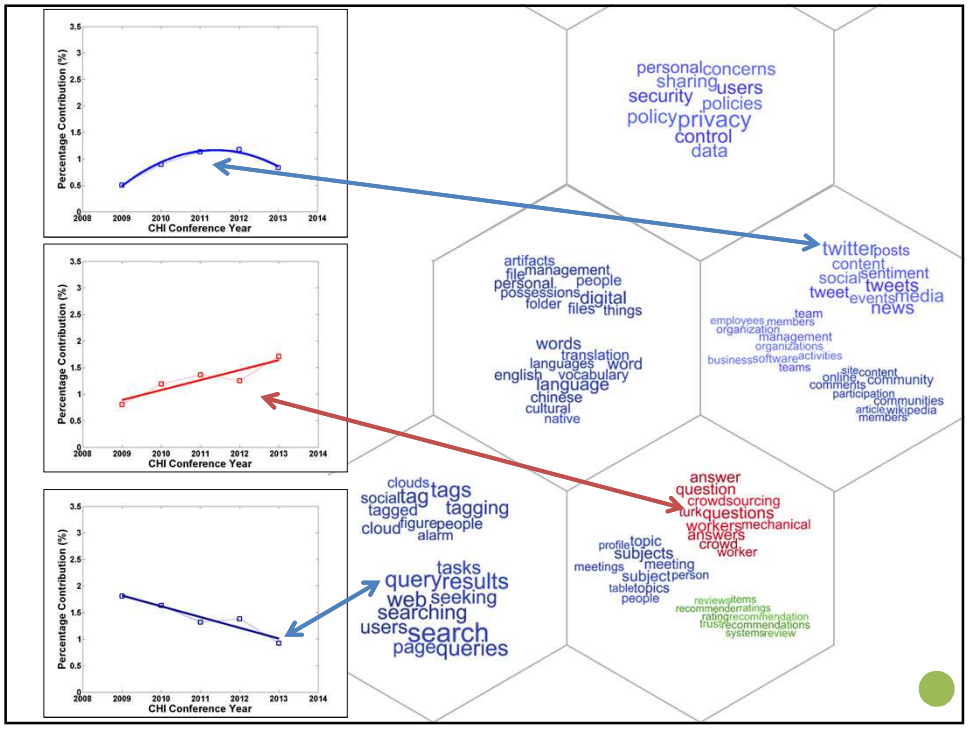


Figure 10: Topic map showing the venue where each topic was the most popular.

Example: Research trends in a particular research community: CHI


Present data-driven overviews of one perspective!


Percentage Contribution (%)
CHI Conference Year
2008
2009
2010
2011
2012
2013
2014
personal concerns
sharing users
security policies
policy privacy
control
data
twitter posts
content
social sentiment
tweet events media
tweets
news
artifacts
file management
personal people
possessions digital
folder files things
words
translation
languages word
english vocabulary
language
chinese
cultural
native
clouds
social tag tags
tagged tagging
cloud figure people
alarm
tasks
query results
web seeking
searching
users search
page queries
answer
question
crowdsourcing
turk questions
workers mechanical
answers
crowd
worker
profile topic
subjects
meeting
meetings
subject person
table topics
people

Gotcha's!

1. It has difficulty classifying grants without summaries
2. It's just word co-occurrence

EPSRC Internal project

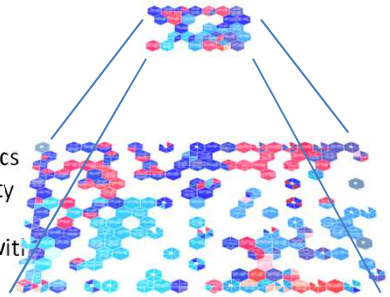
EPSRC Internal Project

Summary

www.researchpesrspectives.org

Summary

- Powerful data-driven technique
- Exploits natural structure
- Very intuitive
- Provides 'independent' classification
- Many levels of abstraction
- Cuts across domains, organisations and politics
- Could change the way UK research community collaborate
- Could change the way UK research engages with research strategy discussions. (and EU)
- Other nations starting similar work
- Caveat
  - Provides perspectives….
  - Only as good as the data
  - it's a decision aid tool….
- Human side critical
- Many different uses
- Please talk to us



Thanks due to

EPSRC ICT

Stefano Padilla, Tom Methven, Craig Mackie, Dave Corne

www.researchpesrspectives.org
www.well-sorted.org