UK Health Data Analytics Workshop: Well Sorted Materials

6th January 2016

Contents

Introduction 1

Dendrogram 2

Tree Map 3

Heat Map 4

Raw Group Data 5

For an online, interactive version of the visualisations in this document, go here:

www.well-sorted.org/output/UKHDANHealthcareOpportunities

Introduction

Dear participant,

Thank you for taking part in submitting and sorting your ideas.

This document contains several visualisations of your ideas, grouped by the average of your online sorts. They are:

Dendrogram - This tree shows each submitted idea and its similarity to the others. The lower two ideas 'join' the more people grouped those two ideas together. For example, if two ideas join at the bottom, every person grouped those two together.

Tree Map - This visualisation presents an 'average' grouping. It is calculated by 'cutting' the Dendrogram at the dashed line so that any items which join lower than that line are placed in the same group. In addition, rectangles which share a side of the same length are more similar to each other than their peers.

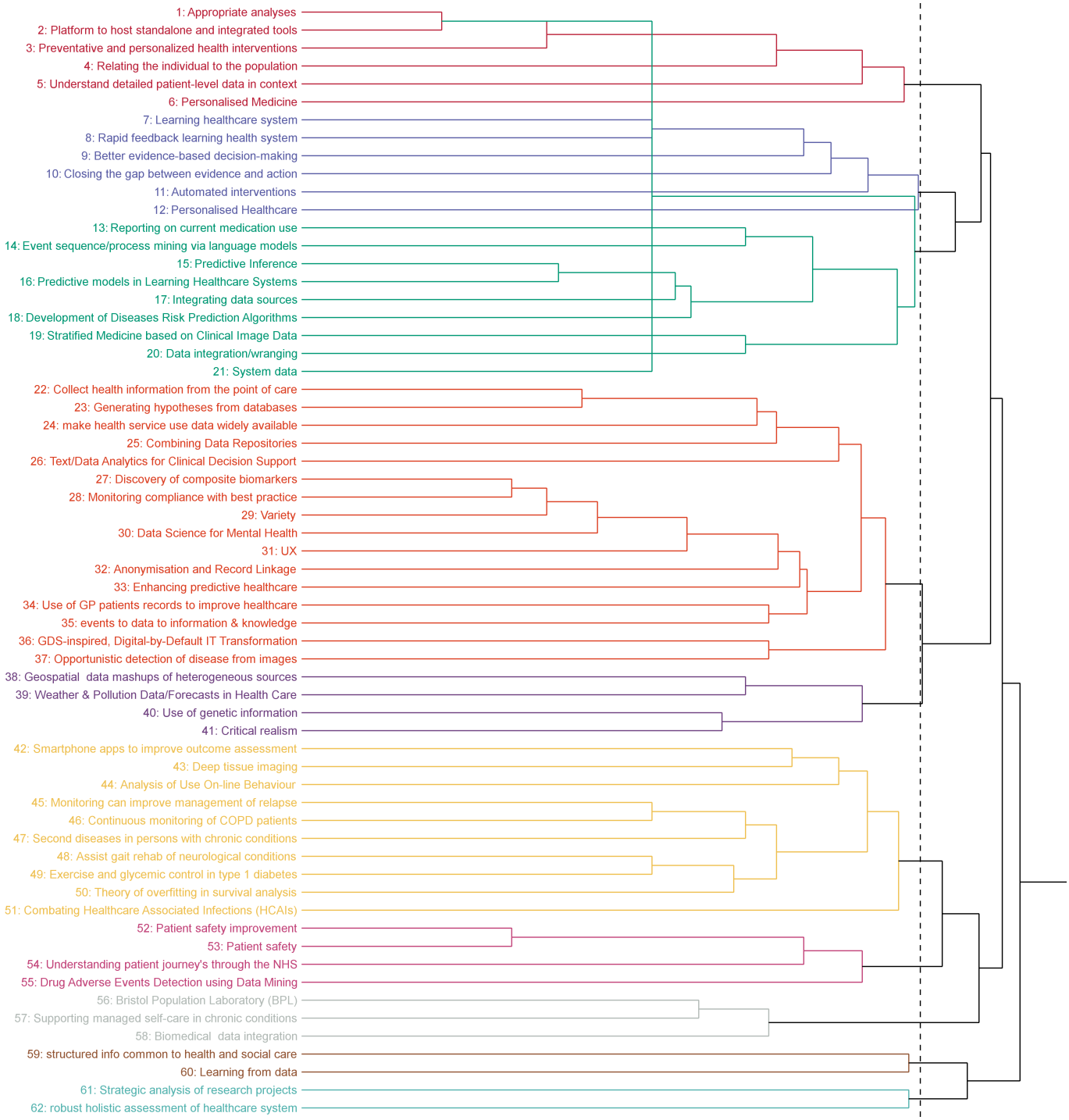
Heat Map - This visualisation shows a similarity matrix where each idea is given a colour at the intersection with another idea, showing how similar the two are. This is useful to see how well formed a group is. The more red there is in a group (shown by the black lines), the more similar the ideas inside it were judged to be.

Raw Group Data - This table shows every submitted idea and its longer description. They are shown in the same order as the Dendrogram (so similar ideas are close to each other) and split into the coloured groups used in the Tree Map. In addition, each idea has been given a unique number so they are easier to find.

References

[1] Methven, T. S., Padilla, S., Corne, D. W., & Chantler, M. J. (2014, February). Research Strategy Generation: Avoiding Academic 'Animal Farm'. In Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing (pp. 25-28). ACM. doi>10.1145/2556420.2556785

Dendrogram


1: Appropriate analyses
2: Platform to host standalone and integrated tools
3: Preventative and personalized health interventions
4: Relating the individual to the population
5: Understand detailed patient-level data in context
6: Personalised Medicine
7: Learning healthcare system
8: Rapid feedback learning health system
9: Better evidence-based decision-making
10: Closing the gap between evidence and action
11: Automated interventions
12: Personalised Healthcare
13: Reporting on current medication use
14: Event sequence/process mining via language models
15: Predictive Inference
16: Predictive models in Learning Healthcare Systems
17: Integrating data sources
18: Development of Diseases Risk Prediction Algorithms
19: Stratified Medicine based on Clinical Image Data
20: Data integration/wranging
21: System data
22: Collect health information from the point of care
23: Generating hypotheses from databases
24: make health service use data widely available
25: Combining Data Repositories
26: Text/Data Analytics for Clinical Decision Support
27: Discovery of composite biomarkers
28: Monitoring compliance with best practice
29: Variety
30: Data Science for Mental Health
31: UX
32: Anonymisation and Record Linkage
33: Enhancing predictive healthcare
34: Use of GP patients records to improve healthcare
35: events to data to information & knowledge
36: GDS-inspired, Digital-by-Default IT Transformation
37: Opportunistic detection of disease from images
38: Geospatial data mashups of heterogeneous sources
39: Weather & Pollution Data/Forecasts in Health Care
40: Use of genetic information
41: Critical realism
42: Smartphone apps to improve outcome assessment
43: Deep tissue imaging
44: Analysis of Use On-line Behaviour
45: Monitoring can improve management of relapse
46: Continuous monitoring of COPD patients
47: Second diseases in persons with chronic conditions
48: Assist gait rehab of neurological conditions
49: Exercise and glycemic control in type 1 diabetes
50: Theory of overfitting in survival analysis
51: Combating Healthcare Associated Infections (HCAIs)
52: Patient safety improvement
53: Patient safety
54: Understanding patient journey's through the NHS
55: Drug Adverse Events Detection using Data Mining
56: Bristol Population Laboratory (BPL)
57: Supporting managed self-care in chronic conditions
58: Biomedical data integration
59: structured info common to health and social care
60: Learning from data
61: Strategic analysis of research projects
62: robust holistic assessment of healthcare system

Tree Map

1: Appropriate analyses
2: Platform to host standalone and integrated tools
3: Preventative and personalized health interventions
4: Relating the individual to the population
5: Understand detailed patient-level data in context
6: Personalised Medicine

RED

7: Learning healthcare system
8: Rapid feedback learning health system
9: Better evidence-based decision-making
10: Closing the gap between evidence and action
11: Automated interventions
12: Personalised Healthcare
13: Reporting on current medication use
14: Event sequence/process mining via language models
15: Predictive Inference
16: Predictive models in Learning Healthcare Systems
17: Integrating data sources
18: Development of Diseases Risk Prediction Algorithms
19: Stratified Medicine based on Clinical Image Data
20: Data integration/wrangling
21: System data

BLUE

22: Collect health information from the point of care
23: Generating hypotheses from databases
24: make health service use data widely available
25: Combining Data Repositories
26: Text/Data Analytics for Clinical Decision Support
27: Discovery of composite biomarkers
28: Monitoring compliance with best practice
29: Variety
30: Data Science for Mental Health
31: UX
32: Anonymisation and Record Linkage
33: Enhancing predictive healthcare
34: Use of GP patients records to improve healthcare
35: events to data to information & knowledge
36: GDS-inspired, Digital-by-Default IT Transformation
37: Opportunistic detection of disease from images

ORANGE

38: Geospatial data mashups of heterogeneous sources
39: Weather & Pollution Data/Forecasts in Health Care
40: Use of genetic information
41: Critical realism

PURPLE

42: Smartphone apps to improve outcome assessment
43: Deep tissue imaging
44: Analysis of Use On-line Behaviour
45: Monitoring can improve management of relapse
46: Continuous monitoring of COPD patients
47: Second diseases in persons with chronic conditions
48: Assist gait rehab of neurological conditions
49: Exercise and glycemic control in type 1 diabetes
50: Theory of overfitting in survival analysis
51: Combating Healthcare Associated Infections (HCAIs)

YELLOW

52: Patient safety improvement
53: Patient safety
54: Understanding patient journey's through the NHS
55: Drug Adverse Events Detection using Data Mining

PINK

56: Bristol Population Laboratory (BPL)
57: Supporting managed self-care in chronic conditions
58: Biomedical data integration

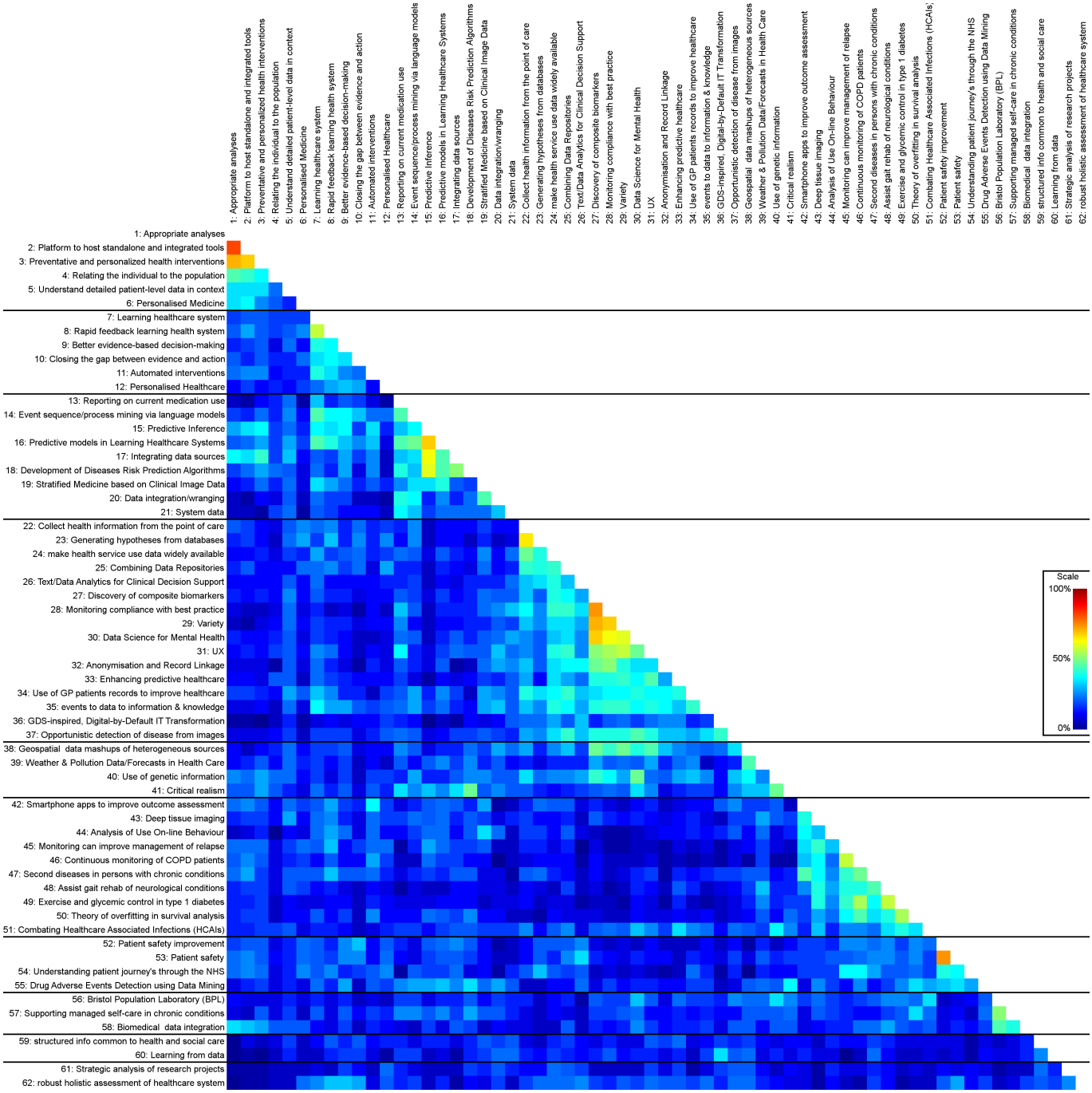
SILVER

59: structured info common to health and social care
60: Learning from data
61: Strategic analysis of research projects
62: robust holistic assessment of healthcare system

BROWN

CYAN

Heat Map


1: Appropriate analyses
2: Platform to host standalone and integrated tools
3: Preventative and personalized health interventions
4: Relating the individual to the population
5: Understand detailed patient-level data in context
6: Personalised Medicine
7: Learning healthcare system
8: Rapid feedback learning health system
9: Better evidence-based decision-making
10: Closing the gap between evidence and action
11: Automated interventions
12: Personalised Healthcare
13: Reporting on current medication use
14: Event sequence/process mining via language models
15: Predictive Inference
16: Predictive models in Learning Healthcare Systems
17: Integrating data sources
18: Development of Diseases Risk Prediction Algorithms
19: Stratified Medicine based on Clinical Image Data
20: Data integration/wranging
21: System data
22: Collect health information from the point of care
23: Generating hypotheses from databases
24: make health service use data widely available
25: Combining Data Repositories
26: Text/Data Analytics for Clinical Decision Support
27: Discovery of composite biomarkers
28: Monitoring compliance with best practice
29: Variety
30: Data Science for Mental Health
31: UX
32: Anonymisation and Record Linkage
33: Enhancing predictive healthcare
34: Use of GP patients records to improve healthcare
35: events to data to information & knowledge
36: GDS-inspired, Digital-by-Default IT Transformation
37: Opportunistic detection of disease from images
38: Geospatial data mashups of heterogeneous sources
39: Weather & Pollution Data/Forecasts in Health Care
40: Use of genetic information
41: Critical realism
42: Smartphone apps to improve outcome assessment
43: Deep tissue imaging
44: Analysis of Use On-line Behaviour
45: Monitoring can improve management of relapse
46: Continuous monitoring of COPD patients
47: Second diseases in persons with chronic conditions
48: Assist gait rehab of neurological conditions
49: Exercise and glycemic control in type 1 diabetes
50: Theory of overfitting in survival analysis
51: Combating Healthcare Associated Infections (HCAIs)
52: Patient safety improvement
53: Patient safety
54: Understanding patient journey's through the NHS
55: Drug Adverse Events Detection using Data Mining
56: Bristol Population Laboratory (BPL)
57: Supporting managed self-care in chronic conditions
58: Biomedical data integration
59: structured info common to health and social care
60: Learning from data
61: Strategic analysis of research projects
62: robust holistic assessment of healthcare system
Scale
100%
50%
0%

Raw Group Data

Colour	#	Title	Description
Red	1	Appropriate analyses	There will always be missing or incomplete data within healthcare records. Being able to apply the "best" analytical techniques to highlight the clinically relevant data would be of great benefit to the public and HCPs using/contributing data.
	2	Platform to host standalone and integrated tools	Provide an open reference platform having containers for versioned software with dependencies, easing the on ramp for less technical researchers to discover, reuse, integrate and repurpose open source tools for health research.
	3	Preventative and personalized health interventions	Large amounts of data are routinely captured for patients from interaction with health service that can be used by clinicians and medical professionals to drive preventative, personalised health interventions.
	4	Relating the individual to the population	When self-managing, people want to find "people like me". How to identify the key "fingerprints" in data for individual and population to facilitate clustering of people with similar conditions and experiences?
	5	Understand detailed patient-level data in context	Health questions often involve the analysis of large quantities of data that are of unknown quality. Visualization allows people to assess fine-grained, high-dimensional data in the context of underlying patterns.
	6	Personalised Medicine	The volume of genetic data available has vastly increased without a proportionate level of impact from the health research community. This is an opportunity to to build on huge amounts of data if we can establish the necessary research infrastructure.

Colour	#	Title	Description
Blue	7	Learning healthcare system	To distil relevant information from routinely collected data, feed that back to patients / clinicians and to test whether that makes a difference.
	8	Rapid feedback learning health system	Data is available to provide rapid feedback on many aspects of health care quality, safety and effectiveness, but we lack effective frameworks to ensure analysis, use and implementation of findings
	9	Better evidence-based decision-making	Access to well presented historical data / analysis will help health practicioners to make better informed decisions by exploring data from cohorts that are more widely sampled but also more specific to the diagnosis.
	10	Closing the gap between evidence and action	There is a cycle of data production to knowledge generation to action. How can we make sure the knowledge is produce in a timely fashion from the data and the knowledge delivered to the actor(s) who can take action?
	11	Automated interventions	Using sensing and real-time data analytics to detect when to trigger a treatment. e.g. to release a drug, not on a fixed timetable, but when the body is detected to be most receptive to it.
	12	Personalised Healthcare	Our digital presence, though inconsistent and varied, provides opportunities for traditional approaches to be improved and complemented towards a personalised, interactive and proactive approach.

Colour	#	Title	Description
Green	13	Reporting on current medication use	Creation of prescribing datasets provides a unique opportunity to report to clinicians and individuals on their current medication use. This would allow for more accurate reconciliation but could also provide a number of value added views of the data.
	14	Event sequence/process mining via language models	Standard data mining classifies and clusters unordered sets of data. health data has ordered sequences of events, more suited to language models from linguistics, e.g. n-gram taggers, Brill taggers and Chart parsers for tagging Part-of-Speech sequences.
	15	Predictive Inference	Most healthcare technologies are not predictive, and merely identify deterioration as it occurs; there is an opportunity to exploit fully-predictive systems for improving patient outcomes.
	16	Predictive models in Learning Healthcare Systems	Develop predictive models that respond to continuous feedback - i.e. coefficients that dynamically update based on secular trends, changing outcome distributions, changing coding practices.
	17	Integrating data sources	The use of multiple data sources and types to better model and understand health care problems and issues. I.e. using all available data regarding a patient to better diagnose and understand underlying conditions.
	18	Development of Diseases Risk Prediction Algorithms	Validated, applicable to clinical practice, predictive algorithms to calculate the risk of individuals for a specific disease to decide on lifestyle modifications and preventive medical treatments
	19	Stratified Medicine based on Clinical Image Data	Providing an integrated infrastructure across PACS and Life Sciences Image and related data sets offers novel tissue and organ-centric data analytics opportunities for clinical and translational sciences, including in stratified medicine.
	20	Data integration/wranging	Bringing together (disparate, large, multi-modal, sensitive) data sources is estimated to take 80% of the "analysis" process. It is critical to achieve effective heath data analytics to address this upstream activity to enable value-added analytics.
	21	System data	Errors in medical device use and EHR use cause harm, but we are not yet collecting use data, nor are logs accurate. We should use data to help design safer systems, and could mine it with patient outcomes. Currently we have no idea which things are safer.

Colour	#	Title	Description
Orange	22	Collect health information from the point of care	... Using pre-existing transactional opportunities. One of the biggest challenges to front line clinicians as the population ages and hospital care becomes more expensive is seeing multi disciplinary,multi organizational care information at presentation.
	23	Generating hypotheses from databases	We need better statistical methods (possibly combined with expert systems) for examining large databases for e.g. genetics and disease. This should mean reducing false positives and allowing formation of, rather than just testing, hypotheses.
	24	make health service use data widely available	administrative records of patients' use of services and prescriptions should be firstly, recorded more accurately, and secondly, made widely available to researchers in order to assess what specific groups within the population may have unmet needs
	25	Combining Data Repositories	There are many growing data repositories associated with experiment/modelling. There is value in combining these with data derived from patient records etc. Centralising these in a regional data centre opens great opportunity for further analysis. Discuss
	26	Text/Data Analytics for Clinical Decision Support	Current text analytics technologies can provide the necessary tools to extract information from unstructured text sources. Such an approach can exploit exemplar cases to help the provision of the right intervention to the right patient at the right time.
	27	Discovery of composite biomarkers	Risk assessment (precursors, biomarkers) for high-prevalence disease based on data analysis with measures from genotype and phenotype info. What groups of 'measures' associate best with outcome (disease)?
	28	Monitoring compliance with best practice	Variance in service delivery has positive and negative clinical and cost implications. EHR data presents an opportunity to identify correlations between service delivery patterns and outcomes to monitor compliance with established best practice.
	29	Variety	Exploiting, joining, and synthesizing a large variety of medical, healthcare and many other types of social data which are originally disconnected. This has provided a great chance of discovery!
	30	Data Science for Mental Health	How to use wearable devices, ubiquitous sensing, data analytics to track markers of mental health issues and provide real-time recommendations to people with mental health problems as which coping strategies to use
	31	UX	Moving well-understood techniques in UX/HCI into the healthcare domain. Especially around data presentation, interaction, and user modelling.
	32	Anonymisation and Record Linkage	The release of health records is dependent on how well we can anonymise the data for research

			purposes. For this, we need state-of-the-art statistical disclosure control techniques. Record linkage and statistical matching can fuse sources of data.
	33	Enhancing predictive healthcare	Integration of multi-omic data along with the extensive patient phenotype data available facilitates predictive modelling of healthcare outcomes. This approach is also central for personalised prediction.
	34	Use of GP patients records to improve healthcare	Longitudinal data is available from patients visiting their GPs. Data mining methods can be used to extract useful information and to potentially identify new drug adverse reactions/symptoms.
	35	events to data to information & knowledge	the technical infrastructure 4 the appropriate distribution of information in the healthcare sector is improving, a lot of the information is still in paper format & not shared; organisational processes to convert events into data, information need help
	36	GDS-inspired, Digital-by-Default IT Transformation	Commercial systems have largely failed to engage with the needs of the end user in health care. Poor IT design has huge costs in primary, 2ndary & social care. GDS-led agile devt. would transform health sector efficiency & transform clinician-engagement.
	37	Opportunistic detection of disease from images	All clinical images are recorded in hospital databases. Though taken to help diagnose a particular disease, they may show signs of other (unsuspected) disease. Automatic analysis algorithms could scan all images and warn if they detect relevant signs.

Colour	#	Title	Description
Purple	**38**	Geospatial data mashups of heterogeneous sources	Data science by mining associations and relations in multiple variables related to a domain can help in discovering important facts that healthcare needs to work better. Visualising these geographically give rise to better analytics and decisions.
	39	Weather & Pollution Data/Forecasts in Health Care	Weather influences human health (e.g. the risk of heat exhaustion, falls on ice, respiratory illness during high pollution episodes, weather influences on arthritis). New data and forecasts being collected can help diagnoses and hospital staff management.
	40	Use of genetic information	Genetic data can be used to aid in diagnosis and refined care pathways.
	41	Critical realism	Critical realism offers exciting prospects in shifting attention toward the real problems that we face and their underlying causes, and away from a focus on data and methods of analysis. As such, it offers a robust framework for the use of a variety of me

Colour	#	Title	Description
Yellow	42	Smartphone apps to improve outcome assessment	Smartphones offer opportunities to collect data directly from patients regarding the symptoms of their disease, and the benefits and harms of treatment, with the potential to improve how the information is used during routine clinic visit.
	43	Deep tissue imaging	We are now in a position to map out tissues using advanced image analysis and machine learning methods. This provides the opportunity to link genomic and proteomic data with the tissue hence putting it into the spatial context of the tissue.
	44	Analysis of Use On-line Behaviour	Log and analyse user on line behaviour to detect early signs of mental health problems, dementia and other pathologies. Data mining of keystrokes, gesture, and mouse input combined with text from email and Internet activity
	45	Monitoring can improve management of relapse	Real time symptom monitoring has the potential to pick up on early warning signs of relapsing mental health problem such as psychosis and therefore result in earlier and more effective and efficient treatment management.
	46	Continuous monitoring of COPD patients	COPD is common, increasing, and not well treated. Ideally treatment should reflect infections-exacerbations and recovery; patient activity levels; and environmental variation. How do we track 24/7 cough and activity to personalise therapy?
	47	Second diseases in persons with chronic conditions	Persons with chronic conditions (e.g. diabetes) are at increased risk of second (often fatal) diseases (e.g. cancer). There is a need to better predict these taking account of age, disease severity, duration and treatment, to develop optimal surveillance.
	48	Assist gait rehab of neurological conditions	People suffering from neurological conditions often experience symptoms affecting their gait. Using wearables to monitor gait and provide assistance when needed in home-based or outdoor settings may help to reduce visits to the clinic and overall cost.
	49	Exercise and glycemic control in type 1 diabetes	The effect of exercise on glycemic control for type 1 diabetics is complex, is specific to the exercise undertaken and is specific to the individual. Data science has the potential to quantify correlations and improve patient quality of life.
	50	Theory of overfitting in survival analysis	Overfitting (many variables, too few samples) is an increasing problem in medical outcome prediction. There is no method yet for predicting required sample sizes for the main statistical method used in the analysis of time-to-event data, Cox regression.
	51	Combating Healthcare Associated Infections (HCAIs)	Recent years have seen rapidly evolving, next-generation whole-genome-sequence (WGS) technologies becoming widely available. There is an exciting opportunity for developing new principled methods to jointly analyse WGS and epidemiological

			data.

Colour	#	Title	Description
Pink	52	Patient safety improvement	Risk in healthcare is dynamic and needs to be assessed continuously as knowledge about patient conditions, treatments and environment increases. Data science would provide proactive/structured means for evolving our understanding of patient safety risks.
	53	Patient safety	Primary health care data has the potential to be used to provide more information on adverse events and long-term health complications for specific treatments.
	54	Understanding patient journey's through the NHS	At present in the NHS in England trying to understand the interaction with the full range of NHS services is difficult as not all services have the capacity to share data and different coding systems are used by different NHS contributors.
	55	Drug Adverse Events Detection using Data Mining	Some drugs can cause serious adverse events on patients. Discovering new adverse events of drugs is important for pharmaceutical companies. Data mining algorithms have been applied to discovering new adverse events of drugs from electronic health records.

Colour	#	Title	Description
Silver	56	Bristol Population Laboratory (BPL)	A range of data is relevant to understanding causes of health problems and to evaluate health interventions. BPL will enable different scales of data (at population macro level, individual level and molecular level) to be linked and worked across together
	57	Supporting managed self-care in chronic conditions	Using data from a range of sources, including home and wearable sensors, to provide personalised advice and feedback, and detect deterioration early enough to avoid unplanned hospital admission and long-term decline.
	58	Biomedical data integration	A big progress has been made in utilizing the genetics or imaging clinical data (histology, MRI) to support diagnostic. On the other hand the integration of multiple source of the information still remains a challenge.

Colour	#	Title	Description
Brown	59	structured info common to health and social care	An agreed representation for the information commonly shared between health and social care, for use in research and direct patient care where appropriate. This would simplify the asking of research questions and the effective use of research results
	60	Learning from data	Using appropriate tools to support evidence-based medicine

Colour	#	Title	Description
Cyan	61	Strategic analysis of research projects	Billions of dollars are spent by research directorates across the globe on healthcare. Yet there is no single uniform view that provides decision makers with an overview of this research or how, for instance, RCUK and NSF spend and trends compare.
	62	robust holistic assessment of healthcare system	some major issues in healthcare are knowledge sharing, establishment of multi-agency, provision of multidisciplinary services and early diagnosis by evidence-based assessment tools. A robust holistic analysis of healthcare data helps to improve them.