REF 2014 Impact Case Studies and the BBSRC

July 2015

The Glasshouse Building, 68 Wharfdale Road

LONDON N1 9SR, UK

http://www.digital-science.com

consultancy@digital-science.com

REF 2014 Impact Case Studies and the BBSRC

This report describes analyses exploring the diversity of impact case studies that are associated with project funding from the Biotechnology and Biological Sciences Research Council (BBSRC). This is background for any further work developing a better understanding of where and how research funding leads to economic, technological and social impacts.

The Research Excellence Framework (REF) is the successor to the Research Assessment Exercise (RAE) cycle run by the UK HE Funding Councils since 1986. In 2014 the REF was restructured to include four-page case study descriptions of the social and economic impact of research, extending the previous focus on academic excellence. Case study documents consist of five sections: summary, research, research references, impact description, evidence for impact. Digital Science has normalized the text, tagged case studies with additional metadata and extracted and standardized the references. Some case studies are confidential but 6,679 of 6,975 original documents are available for analysis.

This is a rich and unique data source describing the outcomes of research work in UK universities. Some but not all REF case studies acknowledge one or more specific funding sources for the underpinning research. Where a specific principal public-sector body (such as a Research Council) was mentioned then the case study has been tagged accordingly. Because this information was not an obligatory component it was not consistently presented, so it is likely that a substantial number of case studies that were in fact RC supported do not explicitly state this.

The consultants

Digital Science is an operating division of the Holtzbrinck Publishing Group. Its consultancy works closely with Nature Publishing Group, a sister-company. Digital Science supports a growing range of analyses drawing on a diversity of data in the research process. Portfolio companies include Altmetric.com (on media attention to research publications), ÜberResearch (on funding), ReadCube (on publication content), figshare (on data, tables and figures within publications), BioRAFT (research management, compliance and productivity), LabGuru (on laboratory management) and Symplectic (on research management information). The consultancy capability is led by Jonathan Adams, formerly Director of Research Evaluation at Thomson Reuters and founder of UK-based consultancy Evidence Ltd.

Digital Science consultants focus on projects around research evaluation, whilst maintaining close links to academic research and institutions. Recent projects include a study for Nesta on the use of alternative metrics in relation to identification of innovative biomedical research; a review of data around conference proceedings for the Australian Research Council; work for HEFCE to analyze and host the REF2014 impact case studies; and on-going work on Science 2.0 for the European Commission. A series of other reports discuss research data and analytical possibilities in a practical, applied context, drawing on the data and expertise within Digital Science, see http://www.digital-science.com/blog/posts/digital-science-launches-digital-research-reports.

Methodology

The REF case study database includes **220 case studies** that explicitly reference BBSRC (or variants). This set (A) formed the first focus of analysis. It was evident that there was a disparity between the institutional distribution of BBSRC research funding and the distribution of case study acknowledgments. There is a testable assumption that there should be some correlation between the number/value of source funding and the number of mentions in impact case studies. If this is wrong, then that indicates that the source funding is not equitably represented in the database. This is likely to be a result of local expectations and practice, driven by researchers and research managers drafting case studies.

For example: Cambridge (385 grants worth £274m indexed in the *Dimensions* research funding database) has 13 case studies that mention BBSRC, whereas Manchester (453, £311m) has only one mention. Reading (104, £44m) has 5 'BBSRC' case studies while Leicester (82, £47m) has just one.[1] *Dimensions* data do not cover the full REF census period. However, the overall difference between the counts of case studies and the counts of awards for which data are immediately available is sometimes substantial, leaving further work to establish more links between funding and impact.

BBSRC used a variety of methods, including a review of grants data to identify matches between named Principal Investigators (PIs) and named case study lead researchers, a manual search of selected UoAs to identify other funding links, as well as matching outputs and impacts data (e.g. publication DOIs and details of spin-out companies known to be linked to BBSRC funding) to case study content. This identified a further set of **211 case studies** that do not explicitly reference BBSRC but can be associated with specific funding. This set (B) forms an adjunct focus for analysis.

The most concentrated clusters of set (A) case studies are found in two Units of Assessment: UOA5 Biological Sciences and UOA6 Agriculture, Veterinary and Food Science. There are **211 case studies** in these UOAs that have neither explicit nor implicit BBSRC links but could reasonably be associated with BBSRC by the nature of their research. This set (C) forms a comparator set for analysis, but may be influenced by BBSRC-support rather than directly engaged.

There are thus a total of 642 case studies in three subsets of which two are positively linked to BBSRC funding and one is linked by association.

Analysis

The initial work was a review and descriptive summary analysis akin to that applied to the entire REF case study database but adapted for the specific dataset. This describes the spread and main clusters of case studies within the BBSRC set by analyzing metadata associated with each case study (e.g. the authoring institution or the Unit of Assessment (UOA) to which the case study was submitted). The text content of the case studies was also processed to assign to a finer-grained field of research than

[1] It is unknown whether disparities in counts are because of selections made by the universities about case studies to submit, so that some BBSRC-funded PIs were not covered, or whether it is due to a failure to see source funding as relevant evidence.

the UOAs allow (by analyzing text in the section describing research background) or the location of impact (by searching for place names in the section describing research impact).

The principal case study elements reporting summary data for each of the subsets A, B and C are:

- Case studies by Units of Assessment (UOAs) and as a share of all case studies in that UOA
- Case studies by institution (HEI), noting frequent UOAs for denser clusters
- Case studies by ANZ Fields of Research (FoRs)

The data are also visualized as:

- Chord diagram summarizing overlaps between FoRs within case studies
- UK map of impact locations with commentary on HEI clusters
- World map of impact locations with commentary on HEI clusters

An analysis of topical clusters among the BBSRC case studies is also developed and visualized. This differs substantially from that reported for the full REF database because it is based on a smaller number of documents, hence a smaller text pool, and on a select disciplinary subset, hence a subject-limited text pool.

Funding overlap across case studies between BBSRC and other UK research funding bodies, insofar as researchers acknowledged any funding sources) is discussed in regard to the case study sub-sets.

Results

Case studies by UOA

BBSRC related case studies occur in 23 of the 36 UOAs in REF 2014. Comparison of the distribution across UOAs of case study subsets A, B and C confirms that the profile of B is similar to A and that it would be reasonable to regard a combination of these as a coherent object for further analysis. The principal difference is that A contains relatively more case studies from biology and agriculture whereas B contains substantially more case studies from chemistry. Since B is funder-identified rather than researcher-identified, this difference may reflect the perspective of the chemistry researchers as to their primary funding source. Subset C is inevitably different in UOA profile as it is based solely on UOA as source.

Table 1 Case study count by Unit of Assessment (UOA) for subsets A, B and C. Values are subset as %ge total case studies in each UOA. A fuller set of UOAs is shown in the Annex.

UOA	A	B	C
Agriculture, Veterinary and Food Science	28.6	19.8	51.6
Biological Sciences	25.7	17.5	56.8
Chemistry	10.4	25.6	0.0
Clinical Medicine	3.9	7.3	0.0
Physics	5.0	5.5	0.0
Allied Health Professions, Dentistry, Nursing and Pharmacy	5.5	5.0	0.0
Psychology, Psychiatry and Neuroscience	3.2	5.4	0.0
Aeronautical, Mechanical, Chemical and Manufacturing Engineering	4.2	3.3	0.0
Computer Science and Informatics	3.6	2.8	0.0
Electrical and Electronic Engineering, Metallurgy and Materials	3.2	3.2	0.0

UOA	A	B	C
Mathematical Sciences	2.4	3.3	0.0
General Engineering	3.3	1.7	0.0
Sociology	3.1	1.0	0.0
Earth Systems and Environmental Sciences	2.9	1.2	0.0
Geography, Environmental Studies and Archaeology	2.1	0.9	0.0

Case studies by institution

There were 68 HE institutions (44% of 154 total in REF2014) that referenced BBSRC funding in at least one case study (subset A) and a further eight were identified by BBSRC as funding recipients (subset B). Some institutions appeared much more frequently in subset B than subset A (e.g. Edinburgh, Manchester, Oxford), which suggests that the reporting of source funding may have been more consistently promoted in some institutions. Looking across subsets A and B, about 10 institutions have a relatively high concentration of BBSRC related case studies, collectively accounting for slightly more than half the total. These are listed in Table 2; the full list is in Annex A.

Apart from SRUC, which is a relatively small organization, no institution with a high frequency of case studies in subset C does not also feature in subsets A and B. This supports the proposition that the case studies in UOAs 5 and 6 are often likely to be influenced if not directly supported by BBSRC funding.

Table 2. Counts of case studies by submitting HE institution for subsets A, B and C. A complete list of institutions is shown in the Annex.

Submitting institution	A	B	C
University of Nottingham	16	6	7
Imperial College London	13	12	5
University of Cambridge	12	19	8
University College London	8	13	12
University of Sheffield	8	5	2
University of Warwick	8	7	1
University of Bristol	7	4	7
University of Leeds	7	3	3
Birkbeck College	6	3	12
University of Aberdeen	6	3	8
University of Edinburgh	6	23	15
University of Oxford	6	18	12
University of Glasgow	4	9	6
SRUC	2	6	11
University of Manchester	1	12	7

Case studies by Fields of Research

BBSRC related case studies are found in significant numbers across of 25 of the 157 4-digit group-level Fields of Research (FoRs) listed in the Australia - New Zealand classification system. Single outliers are found in other FoRs but the core set, where a field is identified in multiple case studies, is more informative.

Four FoRs (Biochemistry, Genetics, Clinical Sciences and Neurosciences) figure prominently for all three subsets. Clinical Sciences is slightly problematic in the text analysis tool because it is difficult to define the area in a specific way as clinical research tends to be wide-ranging. Three other FoRs are also common to the three subsets: Pharmacology, Medical Microbiology and Public Health.

Subset C has a relatively high frequency in Environmental Science and in Ecology; in fact, at the Division (2-digit level) 63 of 94 case studies in Environmental Sciences and 13 of 16 in Earth Sciences are in subset C. This suggests that these case studies would very likely be associated with NERC funding since these fields accord with its mission.

Table 3. Counts of case studies by Field of Research (based on text analysis of research description) for subsets A, B and C.

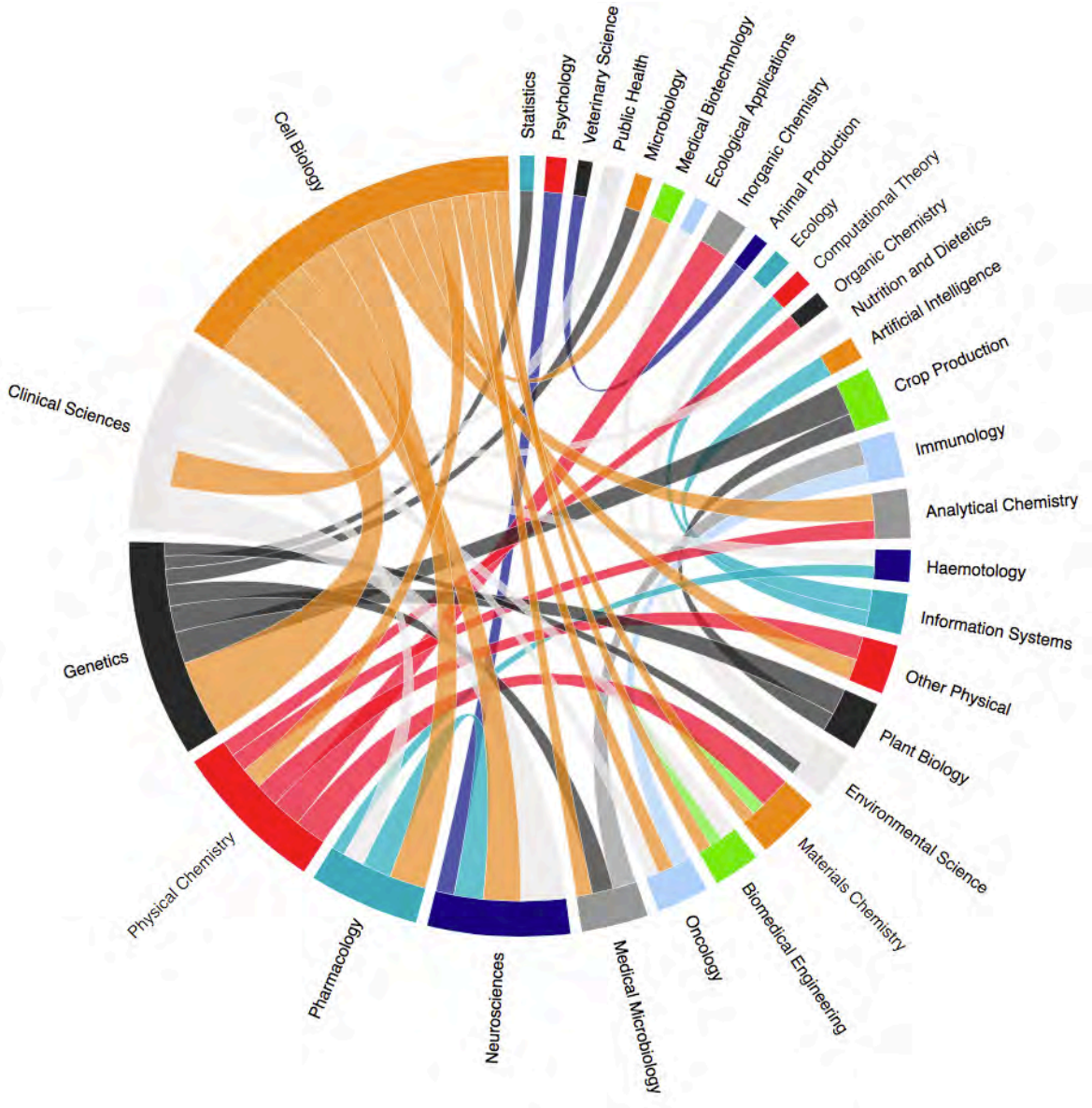
ANZ Field of Research Name	A	B	C
Biochemistry and Cell Biology	55	53	31
Genetics	39	37	46
Clinical Sciences	31	33	31
Neurosciences	22	37	28
Environmental Science and Management	11	7	49
Pharmacology and Pharmaceutical Sciences	17	25	12
Medical Microbiology	16	17	20
Public Health and Health Services	16	14	20
Ecology	5	5	37
Physical Chemistry (incl. Structural)	18	23	1
Immunology	7	17	14
Oncology and Carcinogenesis	7	12	17
Cardiorespiratory Medicine and Haematology	5	18	11
Other Physical Sciences	19	9	2
Analytical Chemistry	12	14	3
Macromolecular and Materials Chemistry	12	13	4
Crop and Pasture Production	15	4	9
Statistics	8	15	4
Artificial Intelligence and Image Processing	13	10	2
Organic Chemistry	9	11	2
Plant Biology	9	4	9
Veterinary Sciences	6	2	14
Medical Biotechnology	8	9	2
Biomedical Engineering	12	6	
Animal Production	7	4	7

B and C overlap with one another more than with A in Immunology, Oncology and Cardiorespiratory research, which are areas usually funded by MRC and the Wellcome Trust.

By contrast, C is rare in chemical and physical sciences. Artificial Intelligence and Statistics are also areas where subsets A and B are both present while C is less frequent.

The FoRs are linked by shared case studies and the network thereby created may add more information than is captured in the summary table of FoR frequency. The chord diagram summarises the FoR links (for A and B subsets, not C). This is an over-simplification as it does not use the spatial relationships that link the case studies (i.e. genetics might be deemed 'closer' to physiology than are either to astronomy) but it provides a rapid reference. The colouring is driven by the net frequency of each FoR. Cell Biology can immediately be seen to have its commonest link to Genetics, and then to Pharmacology and Neurosciences.

Figure 1 Chord diagram showing the overlap between Fields of Research (FoRs) for case studies in subsets A and B. Only FoRs that overlap at least five times with at least one other FoR are shown.



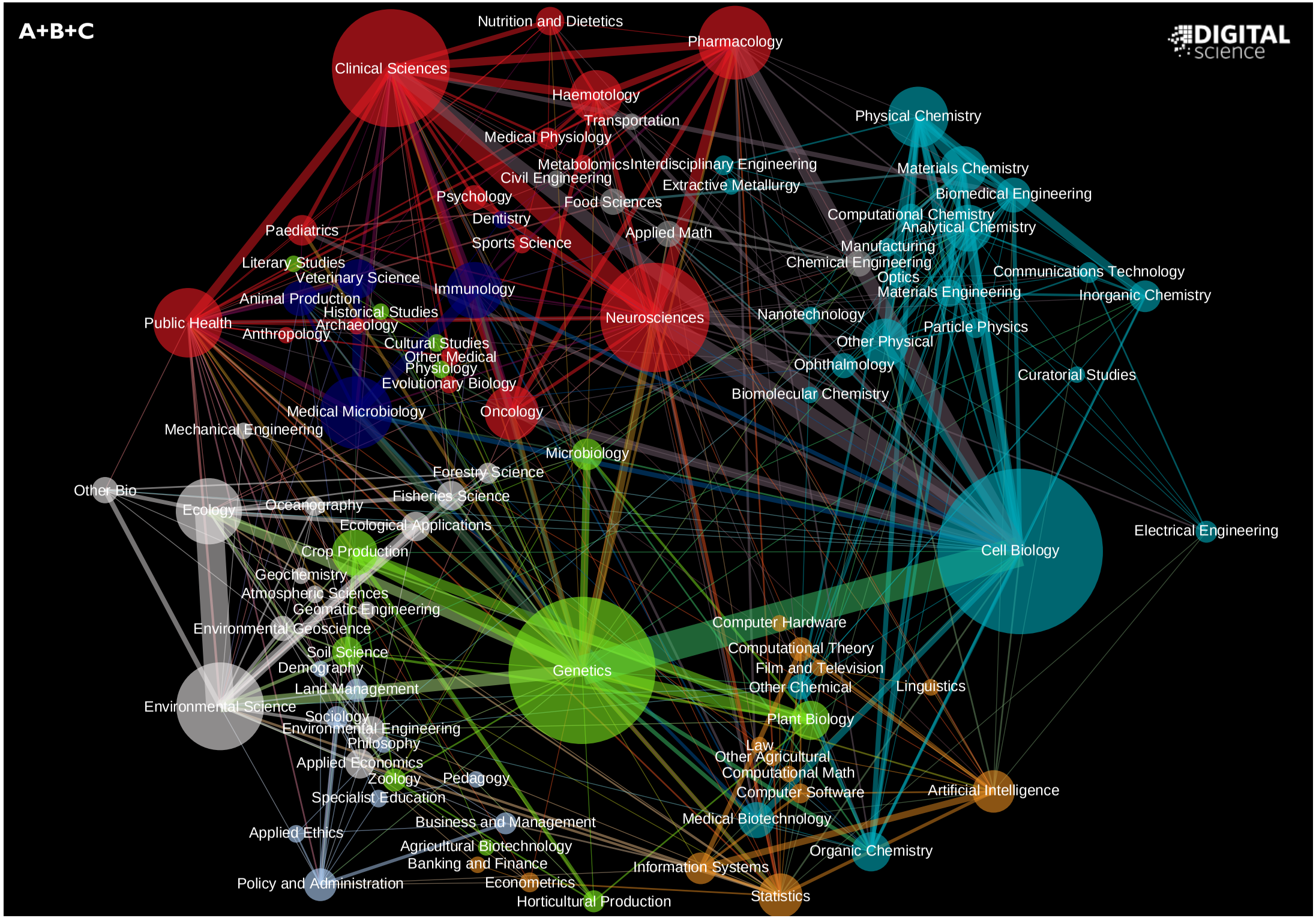
Some linkage sets, such as Genetics-Plant and Biology-Crop Production, provide an informative sense-check on the underlying analysis. Veterinary Science and Animal Production link only to one another. Note also that Ecology/Ecological Adaptations FoRs only link to Environmental Science and through that to Genetics.

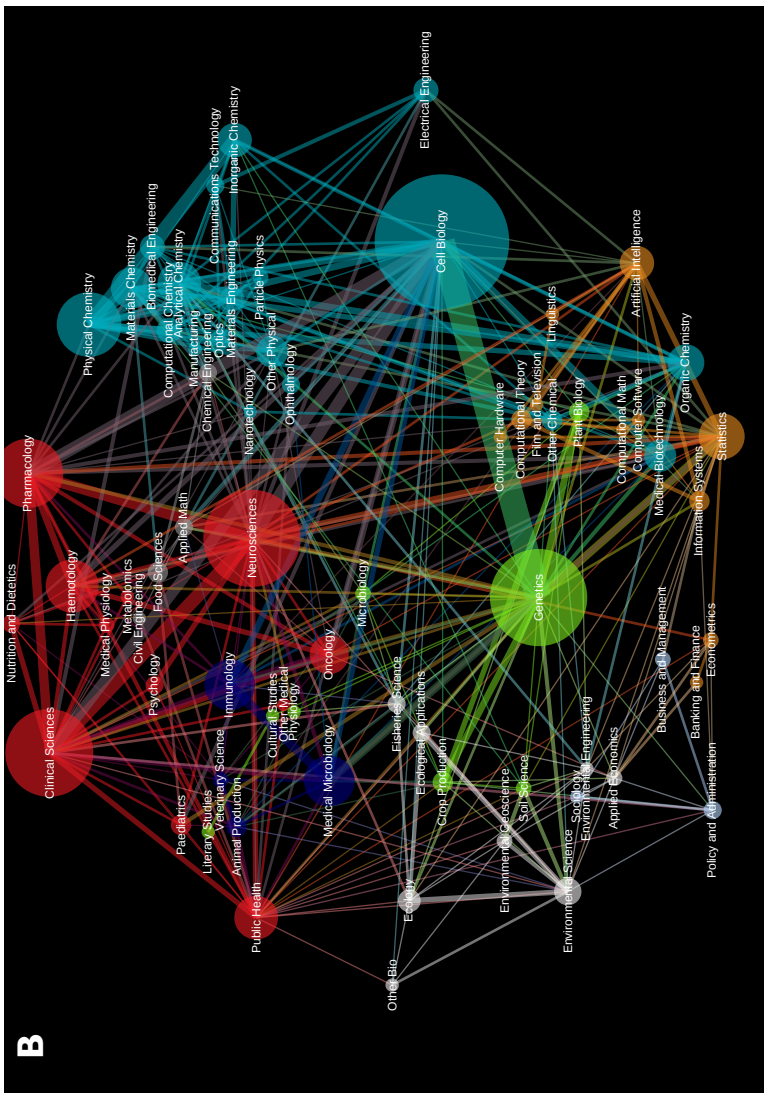
Further analysis makes use of network visualization where the 'distance' between the FoR nodes is informed by the frequency with which nodes have case studies in common. This is, of course, a complex picture to balance since any one node is pulled in multiple directions. The final picture is a convenient reduction to a page of what is in reality a highly multidimensional space. This should be regarded as a mathematically informed compromise rather than a definitive picture of how the FoRs group.

There are four network pictures: one created by the complete dataset of 642 case studies; and three created by the subsets of 220 A case studies (BBSRC referenced as a source), 211 B case studies (identified by BBSRC as receiving funds) and 211 C case studies (present in core BBSRC UOAs).

The complete dataset is used to set a common structure to which the other subsets are then mapped for visual comparison. Colours identify clusters of FoRs with a high frequency of case studies in common, which may overlap other clusters in the indicative space because of strong bilateral links. These clusters appear to be: Biomedical; Cell and materials; Animal sciences; Genetics; Ecology/environment; Education/dissemination; and IT.

The principal difference between the three subset networks is the prominent position of the Ecology/environment group in the C picture. The C picture also has a much lower IT content; this agrees with the previous observation regarding this subset.


A+B+C
DIGITAL science
Nutrition and Dietetics
Pharmacology
Clinical Sciences
Haemotology
Transportation
Medical Physiology
Physical Chemistry
Metabolomics
Interdisciplinary Engineering
Civil Engineering
Materials Chemistry
Extractive Metallurgy
Psychology
Food Sciences
Biomedical Engineering
Dentistry
Computational Chemistry
Analytical Chemistry
Paediatrics
Applied Math
Sports Science
Manufacturing
Literary Studies
Chemical Engineering
Veterinary Science
Optics
Communications Technology
Immunology
Materials Engineering
Inorganic Chemistry
Animal Production
Historical Studies
Neurosciences
Nanotechnology
Public Health
Archaeology
Particle Physics
Anthropology
Other Physical
Cultural Studies
Other Medical
Ophthalmology
Physiology
Curatorial Studies
Evolutionary Biology
Biomolecular Chemistry
Medical Microbiology
Oncology
Mechanical Engineering
Microbiology
Forestry Science
Other Bio
Fisheries Science
Ecology
Oceanography
Ecological Applications
Electrical Engineering
Crop Production
Cell Biology
Geochemistry
Atmospheric Sciences
Geomatic Engineering
Computer Hardware
Environmental Geoscience
Computational Theory
Soil Science
Film and Television
Demography
Genetics
Other Chemical
Linguistics
Land Management
Environmental Science
Plant Biology
Sociology
Environmental Engineering
Law
Philosophy
Other Agricultural
Applied Economics
Computational Math
Zoology
Pedagogy
Computer Software
Artificial Intelligence
Specialist Education
Medical Biotechnology
Business and Management
Applied Ethics
Agricultural Biotechnology
Organic Chemistry
Banking and Finance
Information Systems
Policy and Administration
Econometrics
Statistics
Horticultural Production



Figures 2 A-C

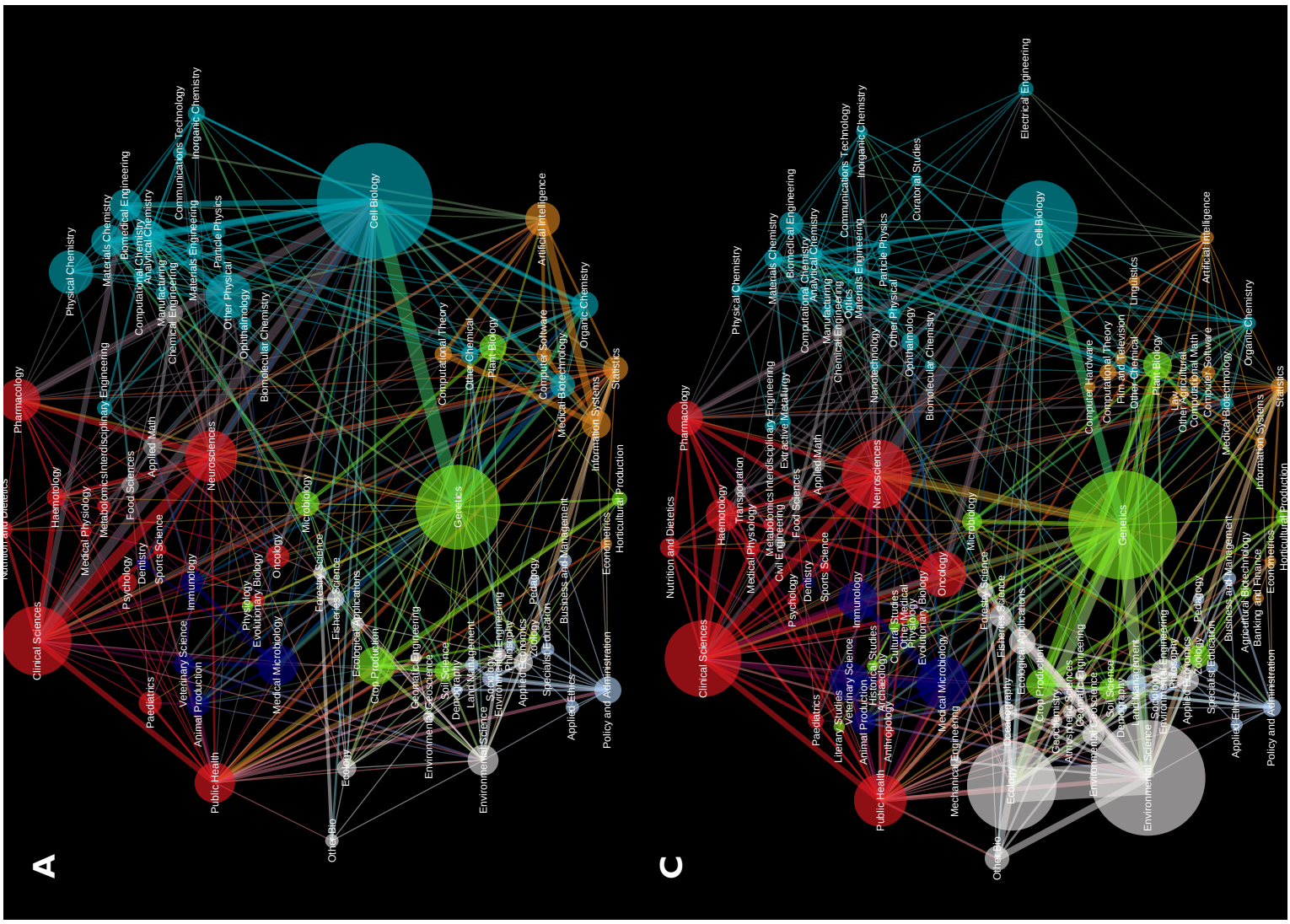
Network diagrams (based on structure in Figure 2) for Fields of Research shared between case studies in three subsets linked to BBSRC funding.

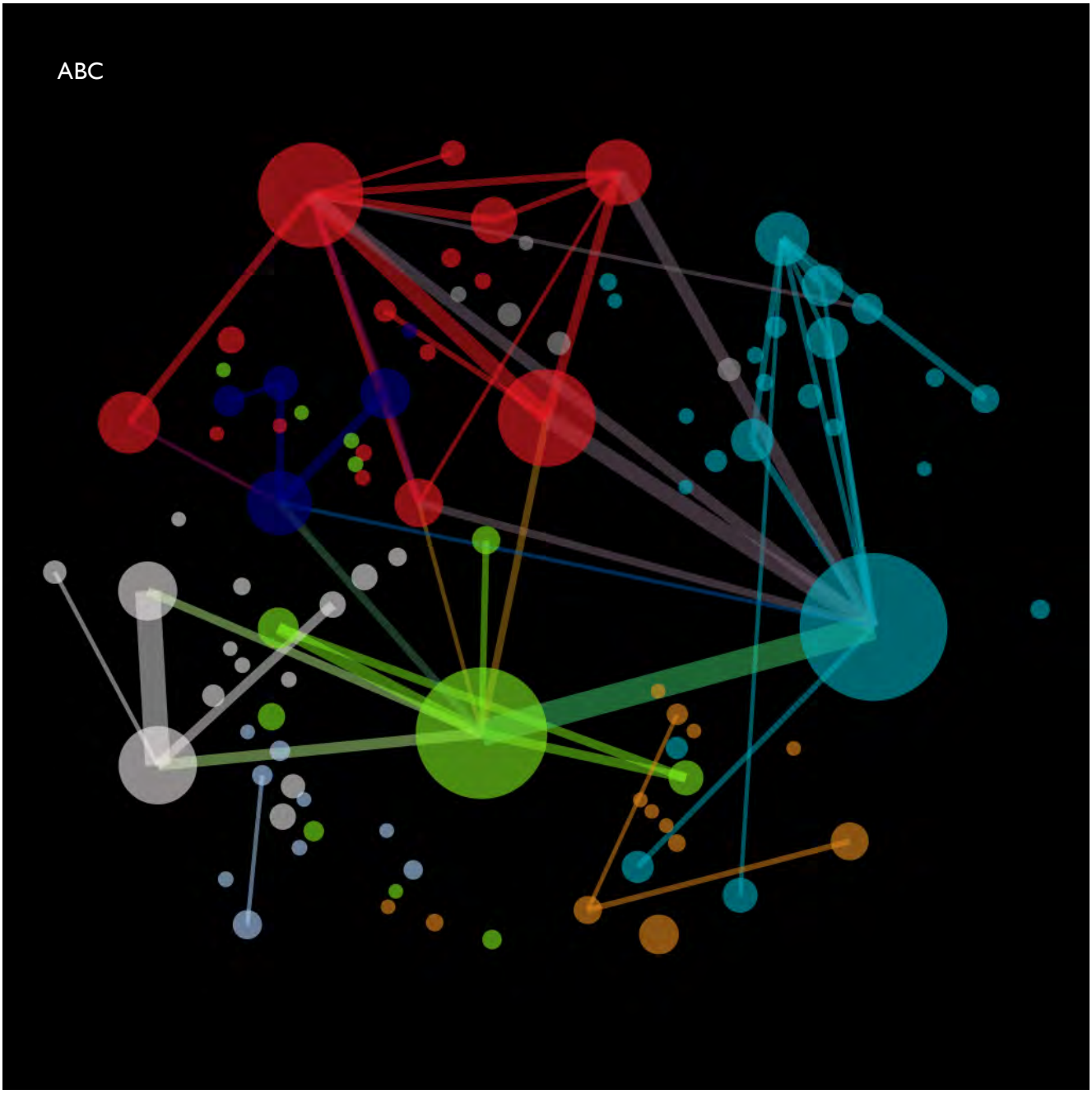
A BBSRC referenced as a source

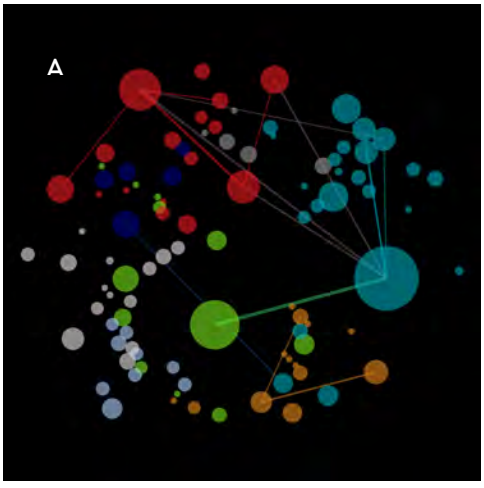
B identified by BBSRC as receiving funds

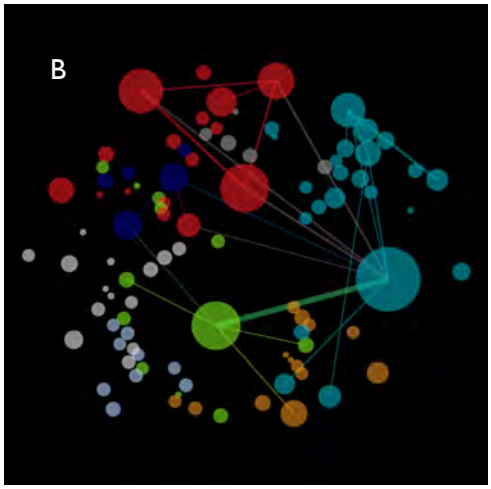
C present in core BBSRC UOAs

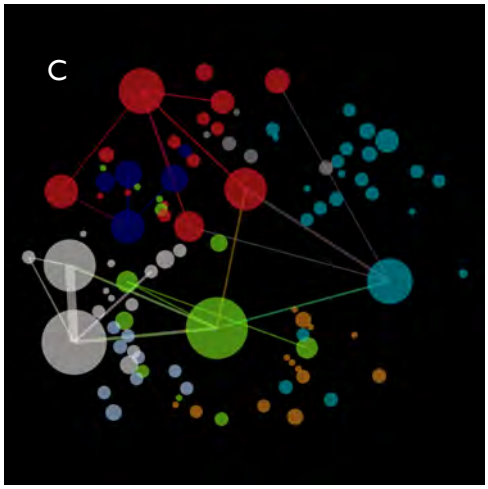
These are repeated on the next page without labels and with only the heaviest edges shown, for clarity.




ABC


A


B

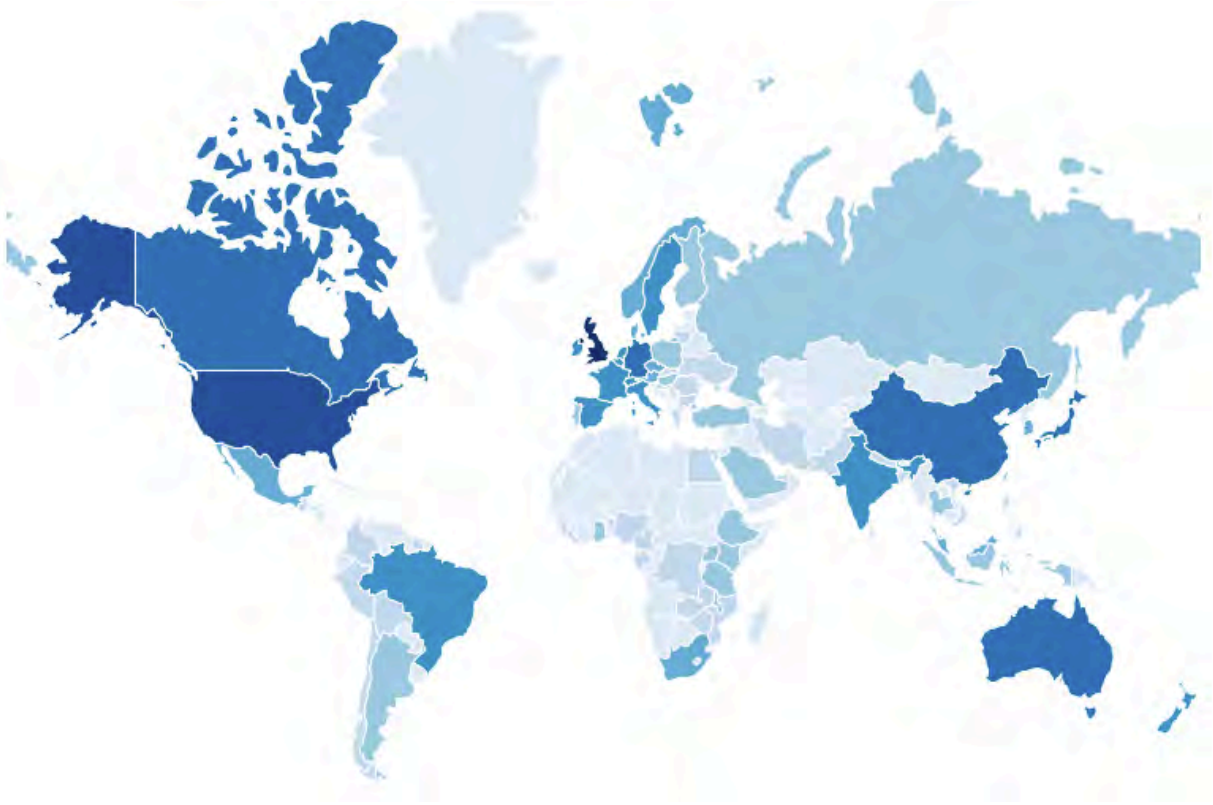

C

Global distribution of the impact of BBSRC-supported research

The case studies include comment on and references to the places where the research has had impact. Case studies are tagged with one or more locations on the basis of places (place names at town, city, region and national level as found in the GeoNames database http://www.geonames.org) referenced in the text of either Section 1 (Summary of impact) or Section 4 (Details of the impact) of the document.

For global locations, the location tags are aggregated up to country level for display purposes but more detailed searching of the documents is entirely feasible in support of further analysis.

Figure 3 World map of BBSRC impact. Colour is graduated on a log scale. Place names are tagged from case study subsets A+B (not C).



Geotagging is an indicative guide designed primarily to aid text searching. It is not a definitive identification of where impact has occurred as some text makes passing references to associated locations and other text refers to impact beneficiaries without identifying a specific location. It should also be noted that automated indexing cannot distinguish those instances where a geo-name is used for a street or a personal surname, e.g. Brazil (country, surname) and Dover (town, street).

Case studies identifying BBSRC-supported research (subset A = 220 instances) include 11 widely distributed examples in Africa and 59 in Asia, where there are clusters in China (23), India (14) and Japan (24). These are more frequently associated with Biological sciences than other UOAs. In Europe, the Netherlands (17) stands out with a cluster associated with Agriculture. There are six case studies associated with Brazil, across various UOAs. There is only one case study associated with Russia, but that country has relatively frequent links only to AHRC and the British Academy.

The combined A+B subsets have a similar relative frequency of case studies for each of these more common locations.

Table 4. Count of case studies in subsets A and B (431 case studies) that refer to locations associated with countries, for the 20 most frequently mentioned countries.

Country	Subsets A+B count	Country	Subsets A+B count
UK	334		
USA	178	Spain	31
Australia	62	Ireland	26
Germany	55	Italy	26
Japan	54	India	24
Canada	50	New Zealand	22
China	43	Belgium	20
France	41	Brazil	19
Netherlands	39	Sweden	18
Switzerland	36	Denmark	16

UK distribution of the impact of BBSRC-supported research

Case studies also make references to places where the research has had impact within the UK. For UK locations, the location tags are aggregated up to GeoNames' Admin 2 version of the Ordnance Survey 'Administrative Geography'. More detailed searching of the documents by city and town is entirely feasible in support of further analysis.

Table 5. Count of case studies in subsets A and B (431 case studies) that refer to UK locations grouped within UK Local Authority Admin 2 areas, for the 20 most frequently mentioned areas.

Area	Subsets A+B count	Area	Subsets A+B count
Greater London	51	Leeds	6
Bristol	21	Oxfordshire	6
Cambridgeshire	14	Cardiff	6
Southampton	14	Norfolk	5
Newcastle upon Tyne	12	Aberdeen City	5
Edinburgh	12	Devon	4
Glasgow City	12	Kent	4
Birmingham	11	Manchester	4
Sheffield	9	Portsmouth	4
Hertfordshire	7	Stirling	4

There is redundancy in UK place names for analysis because the submitting institutions are physically located in the UK. For example, in subset A seven case studies refer to Newcastle upon Tyne, and five of these were submitted by Newcastle University. There is no doubt that universities do have a significant impact on their cities and regions, but separating reference to impact from simple self-identification is inevitably a manual task.

Topic modeling

Topic models provide a way of analyzing the content of full text documents. The process looks for clusters of terms that appear together in a statistically significant way across a set of documents. The derived topics are collections of terms, where each term has a numerical weight in the topic. Each topic is then associated through these terms with documents, where again each topic has a numerical weight in the document. Here, we present analysis of topics identified through Latent Dirichlet Allocation (LDA) using the MALLET[2] topic model package.

For subsets A and B of the BBSRC-identified case studies, 75 topics were identified. Some of these topics are evidently pulling out words specific to areas of impact ("policy", "UK", "government", "evidence", "impact"…), and some denote specific research content ("dairy", "milk", "cattle", "fertility", "farm"…). Other topics are clusters of words used generically to describe research process ("research", "university", "development", "funding", "led"…). This categorization of topics is by no means perfect, but allows users to focus on those topics that are descriptive of the tangible research and impact described in case studies. The 36 topics that the following analysis draws on are listed in Annex B; each topic is labeled manually (by reading associated terms and documents), and the three most highly weighted terms that form the topic are given.

In the following analysis, each case study either *is* or *is not* associated with each topic. In reality this information is drawn from a sliding scale. The topic modeling process assigns a statistical weight that quantifies the extent to which a topic is present in each case study document. By applying a consistent cutoff on this weight to assign (or not) all topics to all case study documents the derived analysis retains mathematical thoroughness whilst the binary connection between any pair of [case study document] and [topic] aids interpretability.

The topic analysis produced, as noted, three types: socio-economic and policy areas (impact topics); specific research areas (research content); and generic research terms (research process). Figure 4 illustrates the 36 topics associated with impact and research content. The impact topics tend to be the largest while those with research content are smaller; many of these are labelled by number, which can be referred to topic labels in Table 6.

[2] McCallum, Andrew Kachites. "MALLET: A Machine Learning for Language Toolkit." http://mallet.cs.umass.edu. 2002.

Figure 4 Bubble diagram showing relative representation of topics amongst BBSRC case studies. Bubble size and colour depth are scaled by the number of associated case studies. For clarity, only the largest topic clusters are given name labels. See Table 6 for labels corresponding to numbered bubbles.

Table 6 Labels for smaller topics labelled by numbers on bubble diagram (Figure 4)

Bubble index	Topic label
16	Biological systems modelling
17	Polymer devices/surface materials
18	Agricultural; plant resistance/pesticides
19	Neuroscience/cognitive health
20	Genetic disease and diagnostics
21	Livestock breeding
22	DNA sequencing and analysis
23	Dairy farming
24	Imaging
25	Vector-borne disease
26	Conservation and biodiversity
27	Poultry health and welfare
28	Antimicrobial resistance
29	Vaccinology

Bubble index	Topic label
30	Therapeutic antibodies
31	Aging population
32	Aquaculture
33	Prions/TSE
34	Human health/Dermatology
35	Dentistry and health
36	Sheep health and welfare

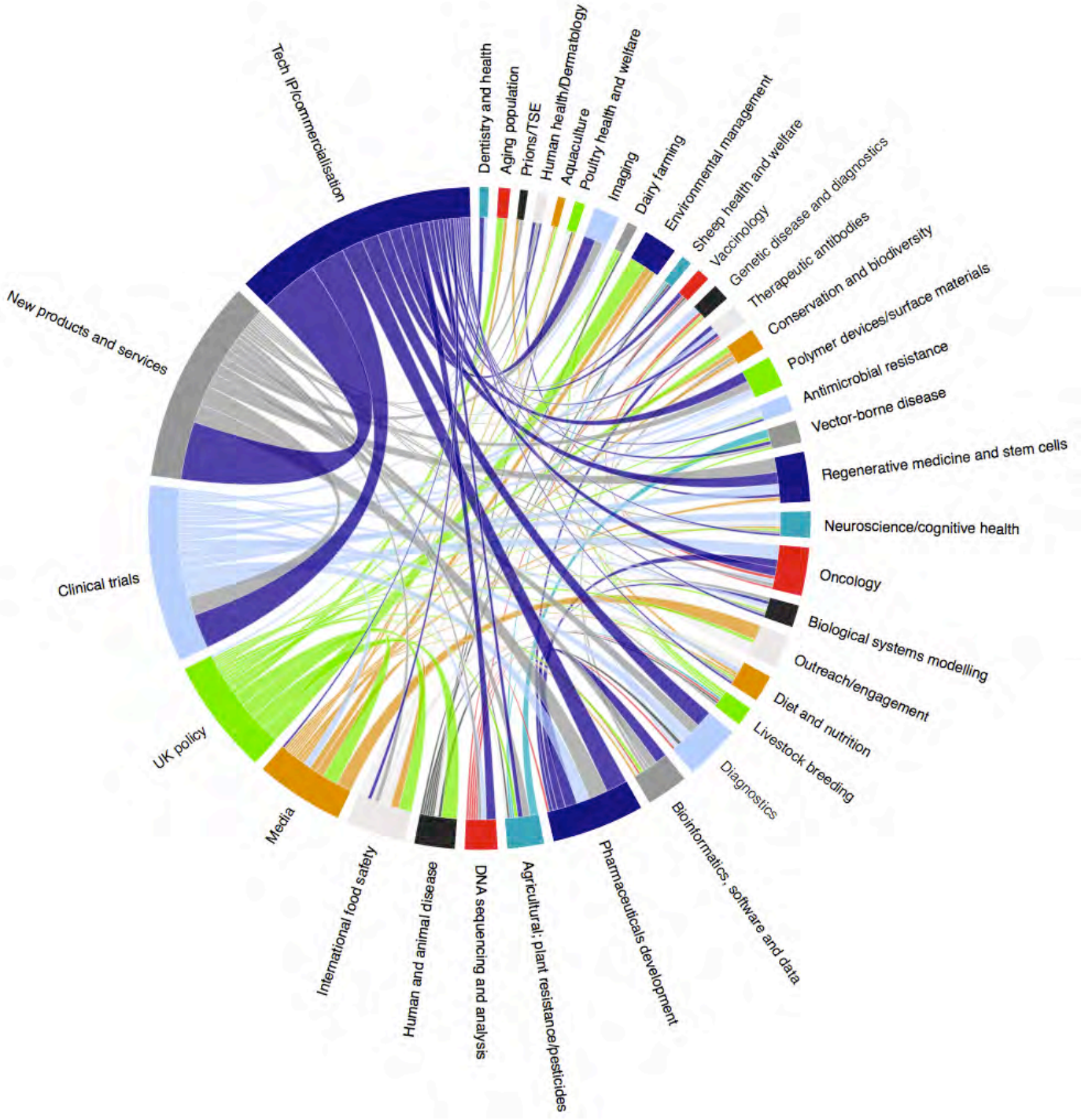
There is a strong association between BBSRC-funded case studies and three main areas:

- Technology IP/commercialization linked to New products and services
- Clinical trials linked to Pharmaceutical development
- UK policy

It is informative to look at the chord diagram (Figure 5) as an adjunct to interpretation of the significance and distinctiveness of the topic clusters in Figure 4. This makes it evident that the UK policy topics are distinct from the other large clusters.

The Media cluster is linked to but distinct from that for Outreach/engagement. The other impact clusters cover a spectrum of biotechnology and biomedical areas associated with the major clusters. For example, both Diagnostics and Polymer devices/surface materials can be seen to be strongly linked to Technology IP/commercialization.

Figure 5 Chord diagram illustrating overlap between topics; the width of each arc shows the extent to which BBSRC case studies are associated with the two topics it links. The chord diagram only reveals the relative size of these two-way overlaps and does not provide definitive information on the extent to which a single topic appears on the same set of documents as another topic.



The chord diagram (Figure 5) reflects the number of case study documents that are shared across topics. As noted, this is most useful when seen alongside the bubble diagram and the two help to interpret the significance of the topic clusters and the way they refer to one another. The topics with the greatest degree of overlap to other topics are highlighted individually in specific chord diagrams shown in Annex C.

The next set of diagrams (Figure 6) summarize the frequency at which FoR divisions (the 2-digit level) are associated with each of the larger topic clusters. Because they cover the full range of academic disciplines, a number of FoRs have little relevance to BBSRC core activity although they may occasionally be associated with a BBSRC case study in part. A threshold is therefore used to focus the analysis on the more widely linked FoRs and to filter out the background. Table 7 indicates which FoRs are included (and excluded).

Table 7. Frequency of the appearance of each Field of Research (2-digit level of aggregation) across the BBSRC case study dataset. FoRs in red have fewer than 50 links to case studies and are not shown in the Impact Wheels illustrated in Figure 6.

FoR division	Count of case study assignments
01 Mathematical Sciences	155
02 Physical Sciences	168
03 Chemical Sciences	686
04 Earth Sciences	13
05 Environmental Sciences	153
06 Biological Sciences	1,003
07 Agricultural and Veterinary Sciences	238
08 Information and Computing Sciences	226
09 Engineering	279
10 Technology	115
11 Medical and Health Sciences	1,519
13 Education	17
14 Economics	54
15 Commerce, Management, Tourism and Services	23
16 Studies in Human Society	81
17 Psychology and Cognitive Sciences	32
19 Studies in Creative Arts and Writing	6
20 Language, Communication and Culture	10
22 Philosophy and Religious Studies	14

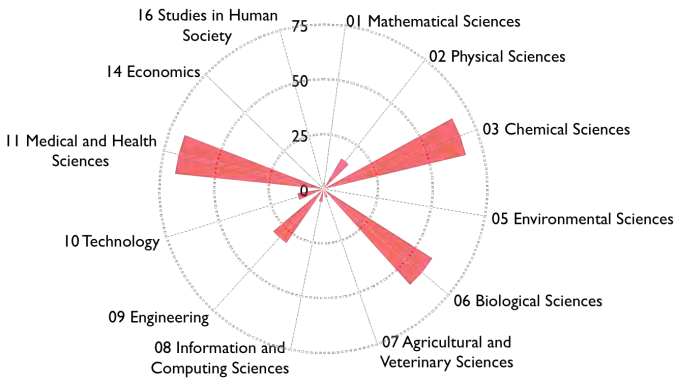
The diagrams in Figure 6 are impact wheels, with the FoRs arranged on the spokes. The thick coloured spokes show the extent of the association between FoRs and topics.

Two major spokes are evident: 06 Biological Sciences and 11 Medical and Health Sciences. These are present in most topics and often occur at a relatively high frequency. It is important to recall that health sciences are rather generic in profile and that biotechnology can be very readily linked to health areas, sometimes for coincidental reasons.

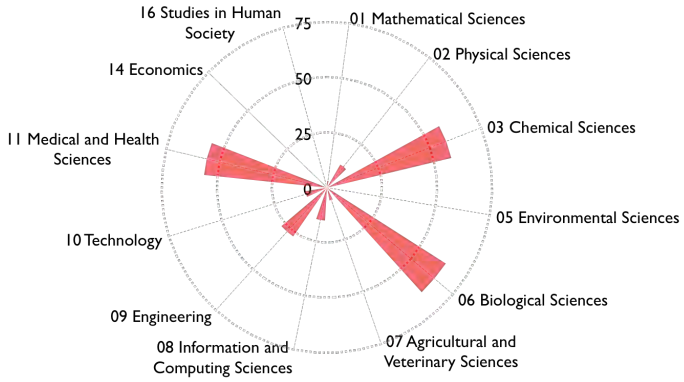
The largest clusters (Technology IP/commercialization and New products and services) look very similar and include 03 Chemical sciences and 09 Engineering as well as 06 and 11. In fact, there are slight differences in balance for the main spokes: New products draws more on 06 Biological Sciences and includes 08 Information and Computing.

Figure 6 A-F Impact Wheels showing the frequency with which the more common FoRs contribute to the research supporting the larger BBSRC impact-orientated topic clusters. Each wheel shows just one topic, and N is the number of case studies that have significant association with that topic. The count on the radial axis indicates the number of times each FoR division is assigned to case studies associated with the topic.

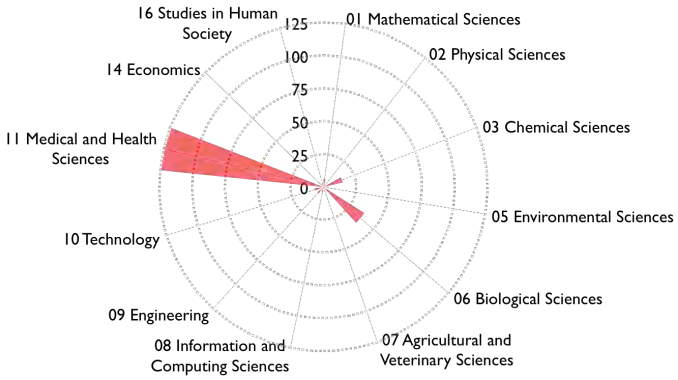
Tech IP/commercialisation, N=105



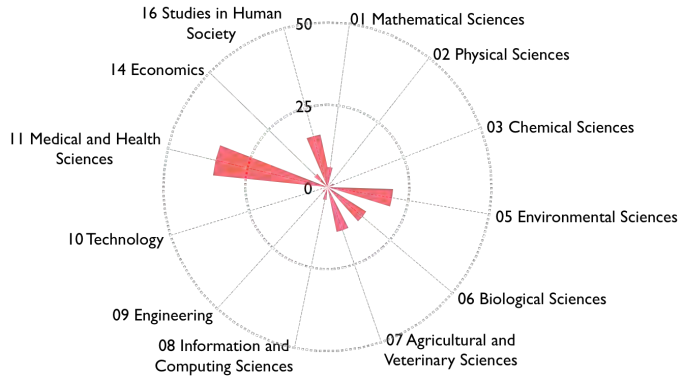
New products and services , N=105



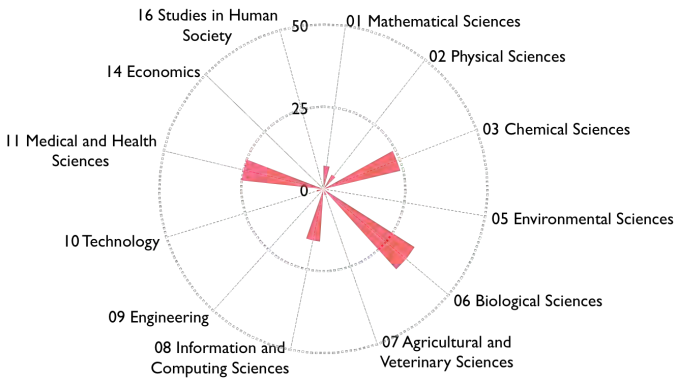
Clinical trials , N=89



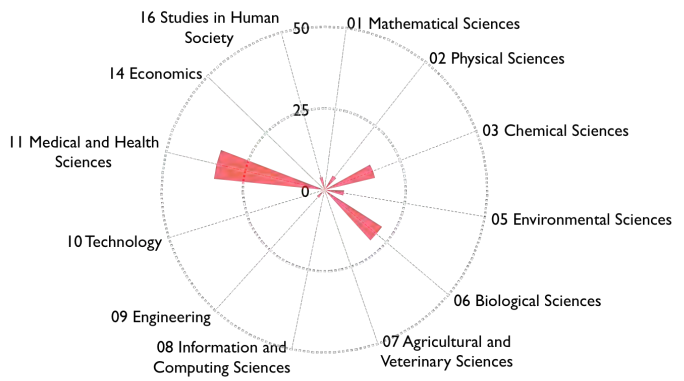
UK policy, N=55



Pharmaceuticals development, N=46

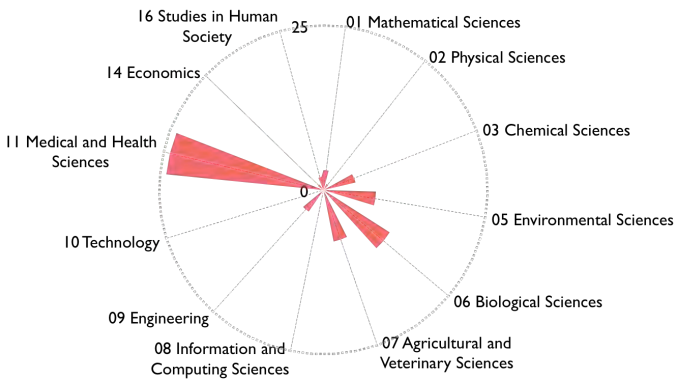


Media, N=43

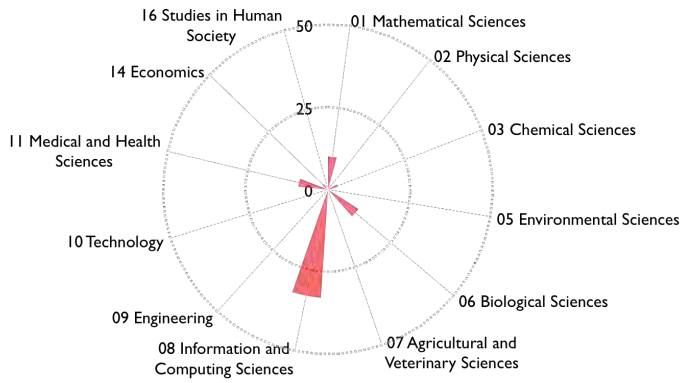


In the Clinical trials topic, the main research spoke is towards FoR 11 Medical and health, and this is also the case for Diagnostics, whereas Pharmaceuticals development includes more basic research from Biology and Chemistry. Bioinformatics draws strongly on 08 Information and computing (as well as 01 Mathematics though to a smaller extent) while the Human and animal disease topic draws on 07 Agriculture and veterinary as well as 06 and 11. These combinations seem to make sense in terms of the original cluster identification, which was based on keyword frequency. They confirm the distinctiveness of the impact topics to which BBSRC research is contributing.

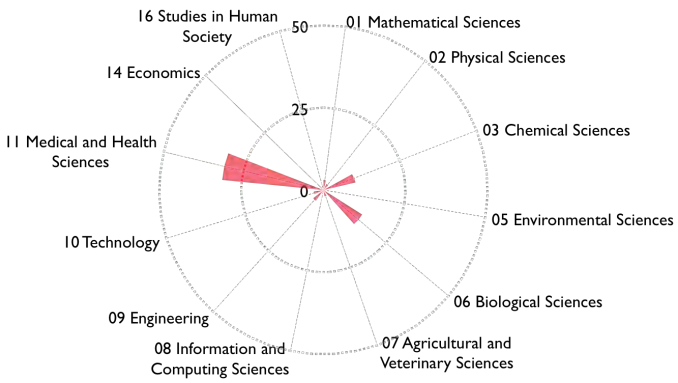
International food safety N=30



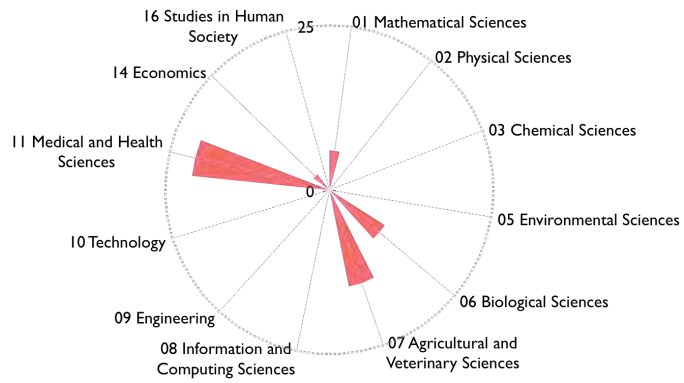
Bioinformatics, software and data, N=25



Diagnostics, N=29



Human and animal disease , N=25



Overlap with other research funders

Both BBSRC and other funding bodies are referenced by case study authors but, as noted, some institutions appear to be more consistent in this than others. Analysis of coincidence is therefore indicative rather than complete.

In subset A, BBSRC overlaps most frequently with EPSRC, MRC and the Wellcome Trust. It overlaps least with STFC, R Acad Eng, AHRC, CCLRC and not at all with the British Academy,

Subset B has a similar pattern of funders to A, excepting the complete absence of BBSRC itself. The two RCs that frequently overlap (EPSRC and MRC) are also relatively less frequent so this tends to confirm the supposition that some authors or institutions simply failed to acknowledge funding.

The Royal Society is about equally frequent in all three subsets, which may reflect a tendency to pick out prestigious funding from that body as supporting evidence of research performance.

NERC is most frequent in subset C. MRC and Wellcome are moderately frequent in that subset, though less than in the other subsets. However, alongside other data, on the pattern of FoRs, this supports the idea that subset C is capturing research that is largely outside BBSRC direct-funding although it may well benefit from BBSRC-funded research occurring in the same institutions.

Table 8. Frequency with which other funding bodies are also identified by authors of case studies associated with BBSRC funding.

Funding body	**A**	**B**	**C**
Total case studies	220	211	211
Arts and Humanities Research Council	3		
Biotechnology and Biological Sciences Research Council	220		
British Academy		1	1
Council for the Central Laboratory of the Research Councils	1	1	
Economic and Social Research Council	16	5	2
Engineering and Physical Sciences Research Council	63	35	5
Medical Research Council	59	43	36
Natural Environment Research Council	22	6	25
Research Councils UK	20	8	4
Royal Academy of Engineering	3	4	
Royal Society	34	33	28
Science and Technology Facilities Council	6	1	
Wellcome Trust	50	46	27

Concluding remarks

The full text analysis of 642 REF impact case studies linked to BBSRC funding shows that the work is relatively concentrated by UOA, as would be expected, but case studies spread across 24 UOAs in total and across 68 universities including many large, research-intensive institutions. They draw on a diversity of Fields of Research (FoRs), where a network analysis reveals clusters of linked FoRs around principal areas in BBSRC's objectives.

These outcomes are somewhat different to those seen in the report produced for HEFCE by Digital Science and King's College, London, as the number of source documents is both highly selective and much smaller. However, the global impact map for BBSRC funded research remains as global in its spread as does the 'parent' map for the total case study portfolio.

Topic modelling indicated three major areas of impact: (1) Technology IP/commercialization linked to New products and services; (2) Clinical trials linked to Pharmaceutical development; and (3) UK policy. Further analysis of the content and research influence suggested that these were relatively distinct clusters of impact and shows how the impact draws on the underlying research.

BBSRC funding mentions overlap with mentions for MRC, EPSRC and the Wellcome Trust. The profile of case studies inferred but not proven to be BBSRC funded indicates that its direct funding overlaps relatively little with NERC.

Full text analysis draws on a wider range of information than e.g. publication coauthorship and citation data. Many of the analyses created here would be infeasible without this full text. This report is therefore complementary to and extends the information available through historical publication-based analyses.

Annex A

BBSRC subsets of case studies as % for total count in Unit of Assessment

Unit of Assessment	Subset A	Subset B	Subset C
Agriculture, Veterinary and Food Science	28.6	19.8	51.6
Biological Sciences	25.7	17.5	56.8
Chemistry	10.4	25.6	0.0
Clinical Medicine	3.9	7.3	0.0
Physics	5.0	5.5	0.0
Allied Health Professions, Dentistry, Nursing and Pharmacy	5.5	5.0	0.0
Psychology, Psychiatry and Neuroscience	3.2	5.4	0.0
Aeronautical, Mechanical, Chemical and Manufacturing Engineering	4.2	3.3	0.0
Computer Science and Informatics	3.6	2.8	0.0
Electrical and Electronic Engineering, Metallurgy and Materials	3.2	3.2	0.0
Mathematical Sciences	2.4	3.3	0.0
General Engineering	3.3	1.7	0.0
Sociology	3.1	1.0	0.0
Earth Systems and Environmental Sciences	2.9	1.2	0.0
Geography, Environmental Studies and Archaeology	2.1	0.9	0.0
Civil and Construction Engineering	2.0	0.0	0.0
Architecture, Built Environment and Planning	0.7	0.7	0.0
Anthropology and Development Studies	0.0	1.3	0.0
Politics and International Studies	1.2	0.0	0.0
Business and Management Studies	0.7	0.2	0.0
Public Health, Health Services and Primary Care	0.0	0.6	0.0
Social Work and Social Policy	0.5	0.0	0.0
Art and Design: History Practice and Theory	0.0	0.4	0.0

Counts of case studies by submitting institution

Submitting institution	Subset A	Subset B	Subset C
University of Nottingham	16	6	7
Imperial College London	13	12	5
University of Cambridge	12	19	8
University College London	8	13	12
University of Sheffield	8	5	2
University of Warwick	8	7	1
University of Bristol	7	4	7
University of Leeds	7	3	3
Birkbeck College	6	3	12
University of Aberdeen	6	3	8
University of Edinburgh	6	23	15

Submitting institution	Subset A	Subset B	Subset C
University of Oxford	6	18	12
Aberystwyth University	5	4	4
Bangor University	5	3	4
Newcastle University	5	3	4
Royal Veterinary College	5	2	4
University of Kent	5	2	
University of Reading	5	2	5
King's College London	4	4	5
University of Dundee	4	1	5
University of Durham	4	1	2
University of East Anglia	4	4	3
University of Glasgow	4	9	6
University of Liverpool	4	2	5
University of Southampton	4	5	
Cardiff University	3	4	2
Keele University	3		2
Lancaster University	3	1	
Queen Mary, University of London	3	1	3
Queen's University Belfast	3	3	4
Sheffield Hallam University	3		1
University of Exeter	3	5	4
University of Greenwich	3		2
University of St Andrews	3	3	5
University of Stirling	3	1	2
University of Surrey	3	2	
Aston University	2		
Cranfield University	2	1	2
Oxford Brookes University	2	1	2
SRUC	2	6	11
University of Bath	2	2	1
University of Birmingham	2	3	4
University of Hertfordshire	2		2
University of Strathclyde	2	2	
University of Sussex	2	1	5
Brunel University	1	2	
Edinburgh Napier University	1	1	
Glasgow Caledonian University	1		
Heriot-Watt University	1	1	1
Institute of Zoology, London	1		
London School of Hygiene & Tropical Medicine	1		
London South Bank University	1		
Loughborough University	1		
Manchester Metropolitan University	1		
Open University	1		

Submitting institution	Subset A	Subset B	Subset C
Plymouth University	1	1	2
St George's, University of London	1		
University of Bedfordshire	1	1	
University of Brighton	1		
University of Essex	1	1	2
University of Huddersfield	1		2
University of Hull	1		
University of Leicester	1		4
University of Manchester	1	12	7
University of Portsmouth	1	2	
University of Salford	1	1	2
University of York	1	3	3
University of the West of England, Bristol	1		
Anglia Ruskin University		1	2
Canterbury Christ Church University			2
Harper Adams University			3
Institute of Cancer Research		3	3
Liverpool School of Tropical Medicine		1	
Middlesex University		1	
Northumbria University Newcastle		1	
Royal Agricultural University			2
Royal Holloway, University of London			3
Swansea University		1	
University of Abertay Dundee			1
University of Chester		1	
University of Derby			2
University of Lincoln			2
University of Sunderland		1	
University of Worcester			2
Writtle College			2

Annex B

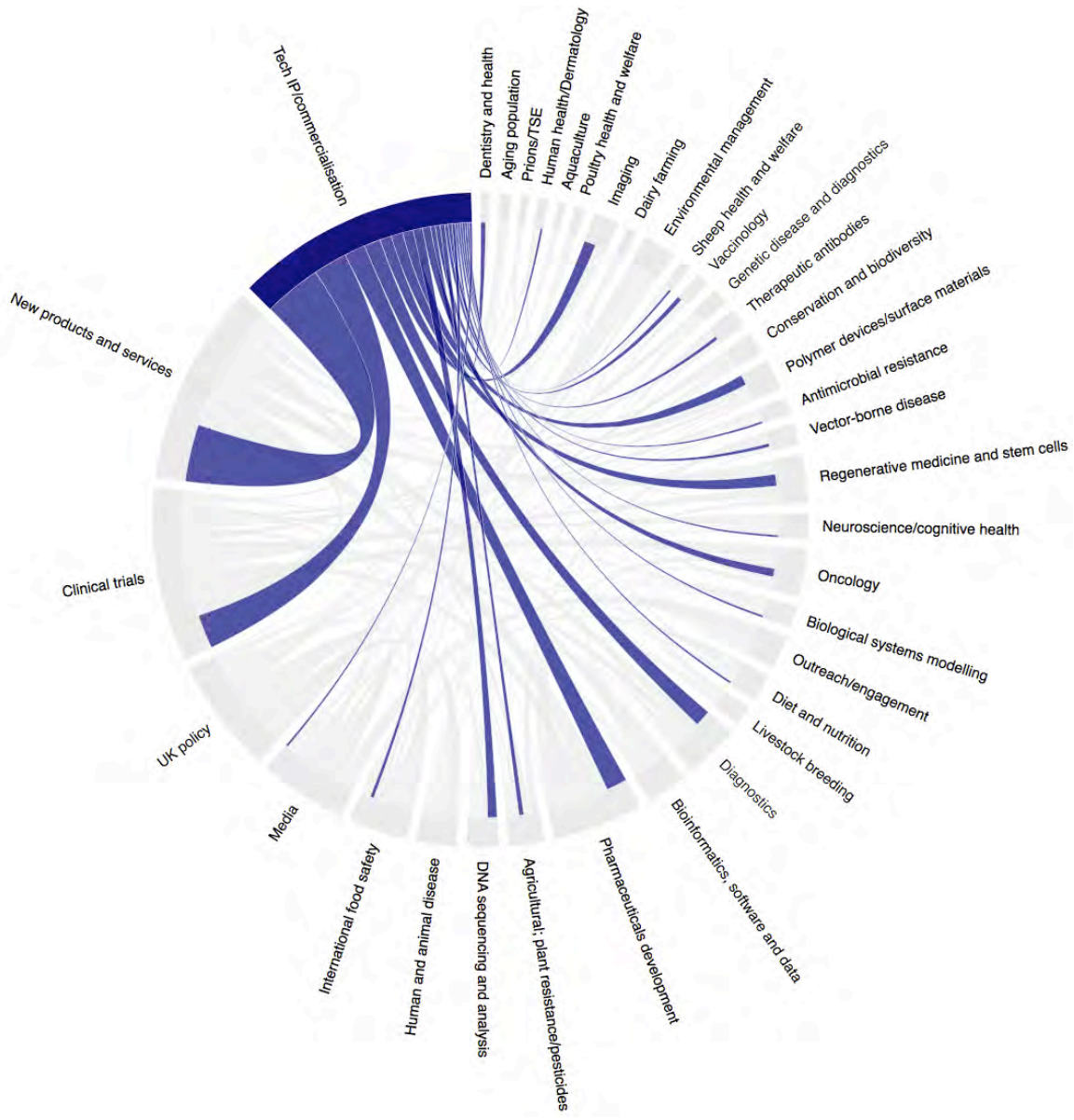
Topic modeling labels and top three terms

Topic label	terms		
Tech IP/commercialisation	technology	company	patent
New products and services	company	products	product
Clinical trials	clinical	patients	treatment
UK policy	policy	uk	government
Pharmaceuticals development	protein	drug	discovery
Media	public	science	media
International food safety	food	european	eu
Bioinformatics, software and data	data	software	analysis
Outreach/engagement	science	school	public
Diet and nutrition	food	health	research
Environmental management	management	environmental	land
Dairy farming	dairy	milk	cattle
Poultry health and welfare	welfare	animal	rvc
Aging population	ageing	people	age
Diagnostics	test	detection	assay
Human and animal disease	disease	control	infection
Regenerative medicine and stem cells	cell	cells	stem
Oncology	cancer	ref	phase
Biological systems modelling	modelling	systems	model
Polymer devices/surface materials	surface	materials	release
Agricultural; plant resistance/pesticides	seed	crop	plant
Neuroscience/cognitive health	brain	cognitive	disease
Genetic disease and diagnostics	disease	genetic	gene
Livestock breeding	genetic	breeding	uk
DNA sequencing and analysis	gene	expression	dna
Imaging	imaging	microscopy	high
Vector-borne disease	control	malaria	oil
Conservation and biodiversity	species	conservation	dna
Antimicrobial resistance	bacterial	mrsa	antibiotic
Vaccinology	vaccine	virus	viral
Therapeutic antibodies	antibody	dna	antibodies
Aquaculture	fish	salmon	light
Prions/TSE	uoe	health	blood
Human health/Dermatology	skin	human	uk
Dentistry and health	leeds	bone	dental
Sheep health and welfare	sheep	farmers	sarissa

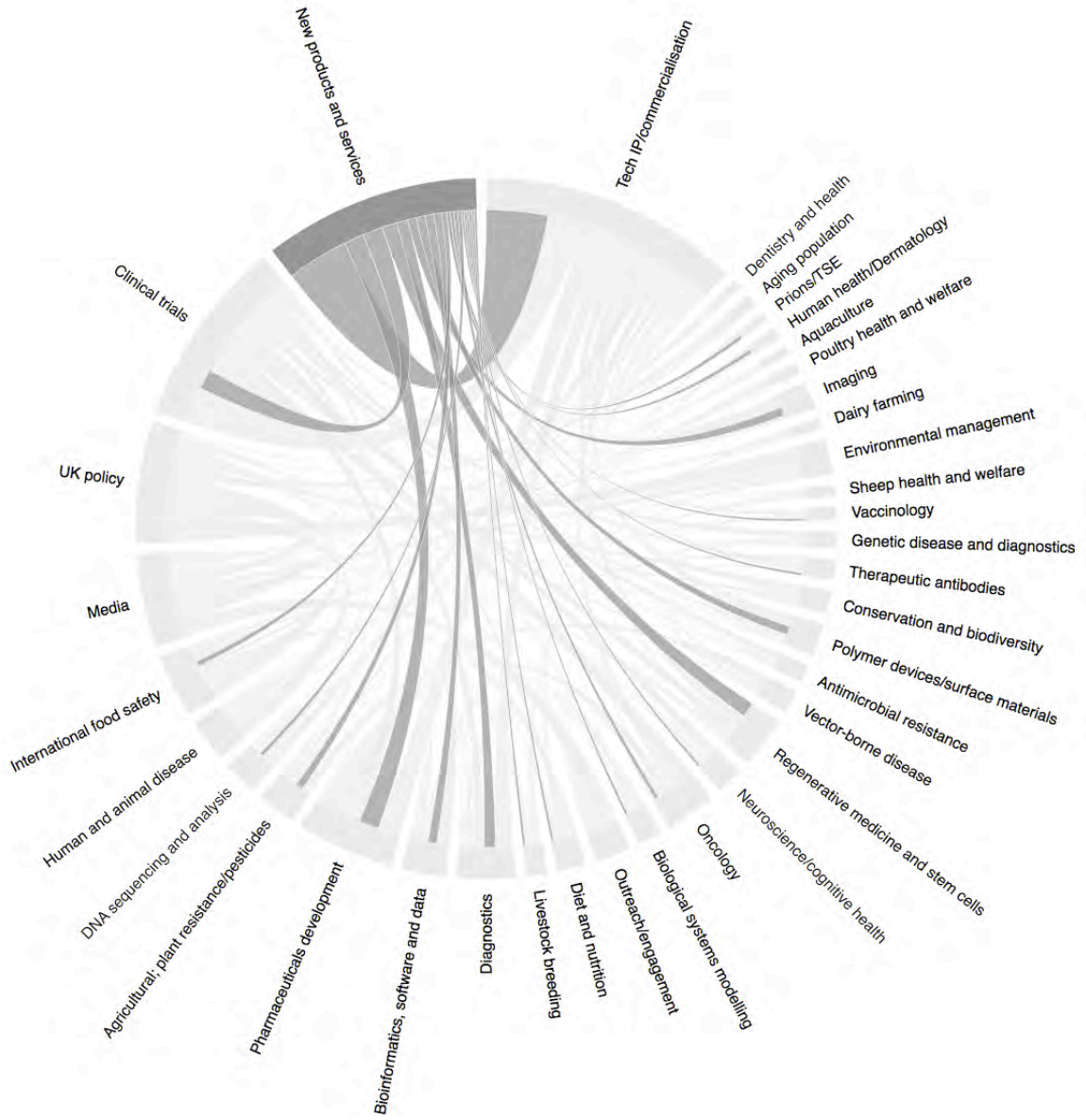
Annex C

Topic overlap chord diagrams

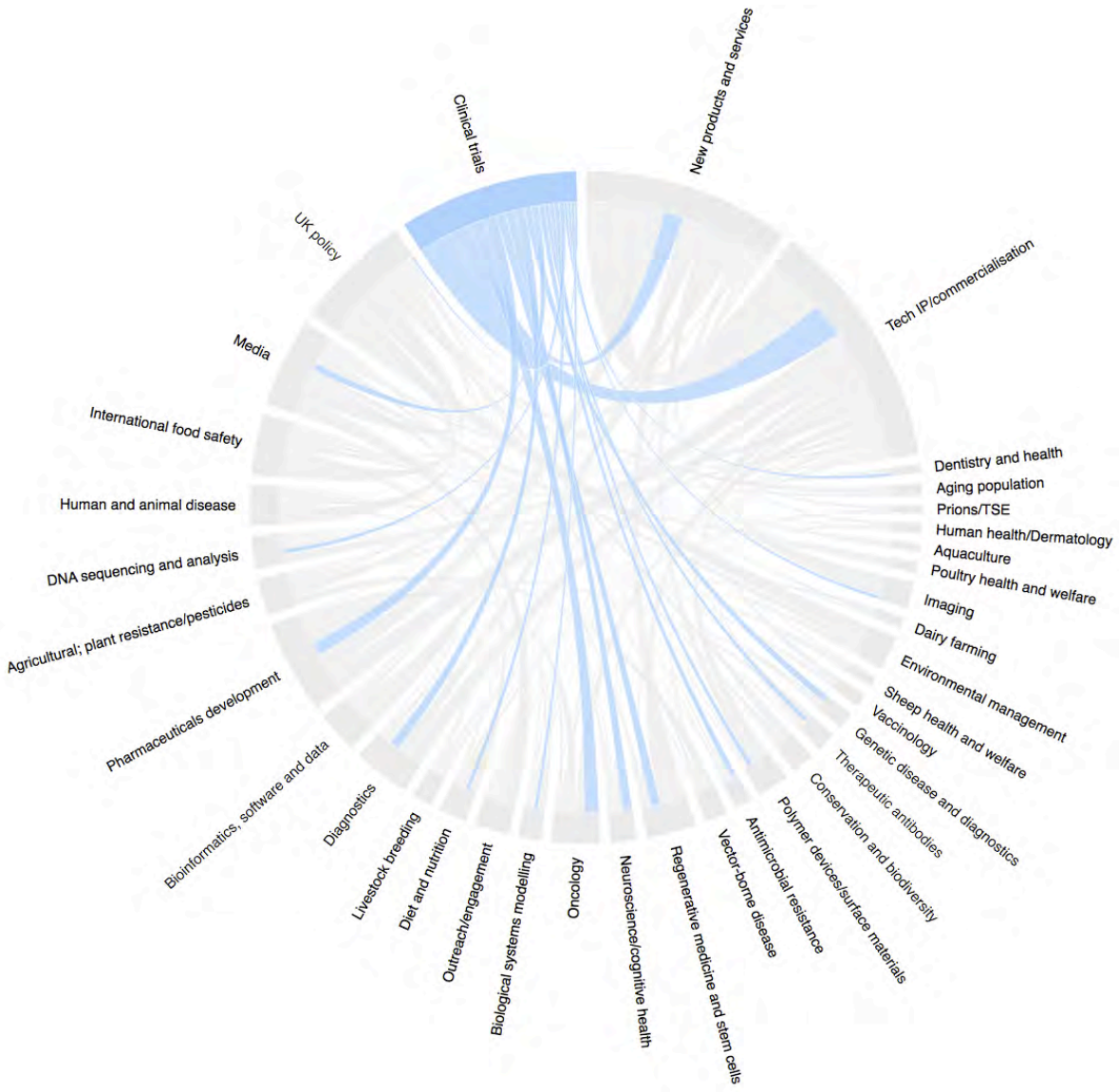
This is the topic overlap chord diagram as in figure 5, with links greyed out unless they involve the Tech IP/commercialization topic.



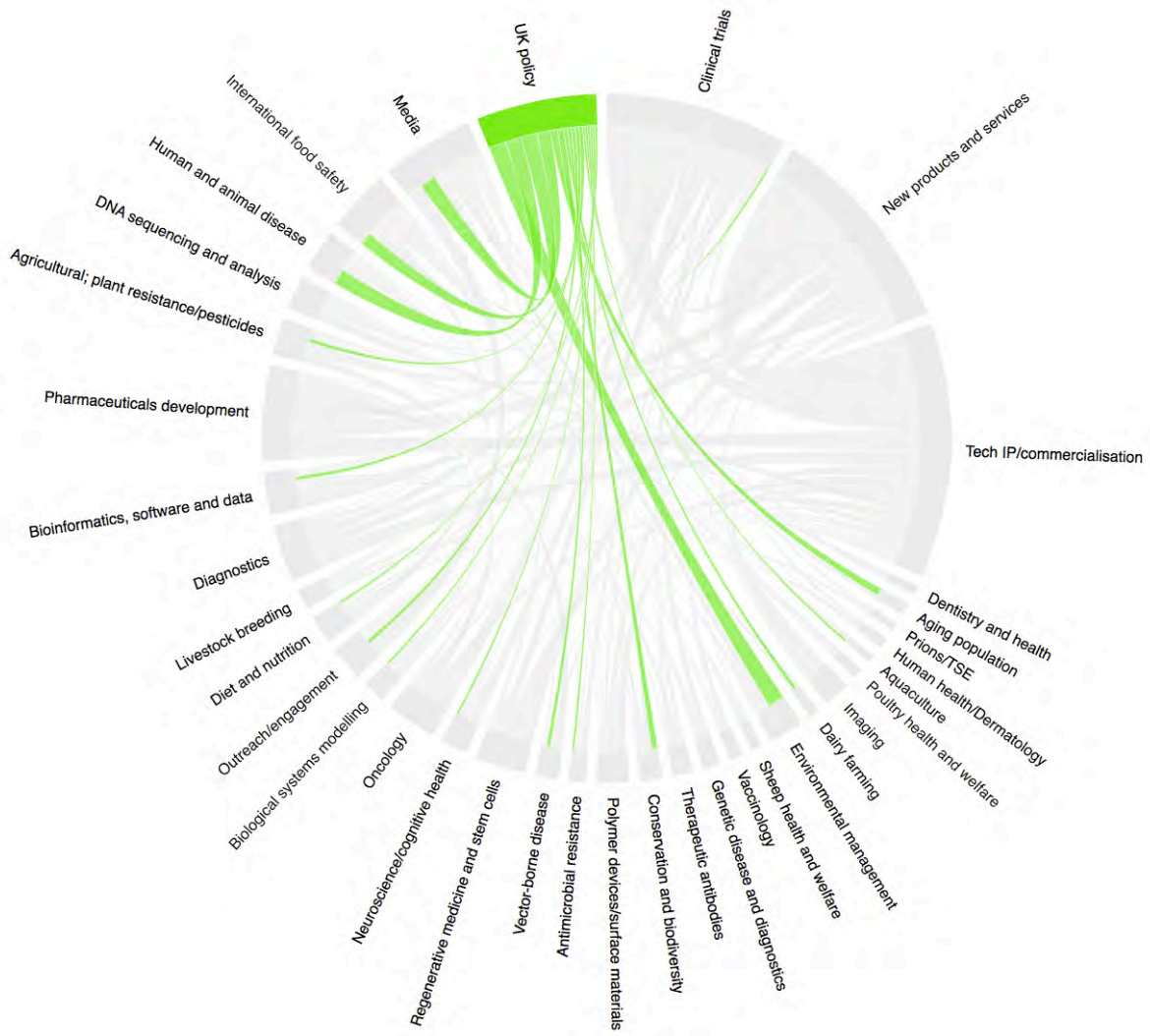
This is the topic overlap chord diagram as in figure 5, with links greyed out unless they involve the Tech products and services topic.



This is the topic overlap chord diagram as in figure 5, with links greyed out unless they involve the Clinical trials topic.



This is the topic overlap chord diagram as in figure 5, with links greyed out unless they involve the UK policy topic.



This is the topic overlap chord diagram as in figure 5, with links greyed out unless they involve the Media topic.