Science Industry Partnership

DATA, COMPUTATIONAL AND INFORMATICS SKILLS FOR THE SCIENCE-BASED ECONOMY

A SIP Discussion Paper

October 2017

SIP

Introduction

Some technologies change the world. The introduction of iron, steam, and electricity and then information technology caused step-changes in the way in which economies function. This earned these sectors the title of General Purpose Technologies (GPTs). Also often presented as the "four industrial revolutions".

As the dialogue has developed, the term Key Enabling Technology (KETs) subsequently emerged to reflect some of the "enabling" qualities of GPTs that support widespread industrial development – and the fourth industrial revolution.

And there is no doubt skills needs of the future will be driven by the adoption of these new technologies, combined with underpinning core science skills.

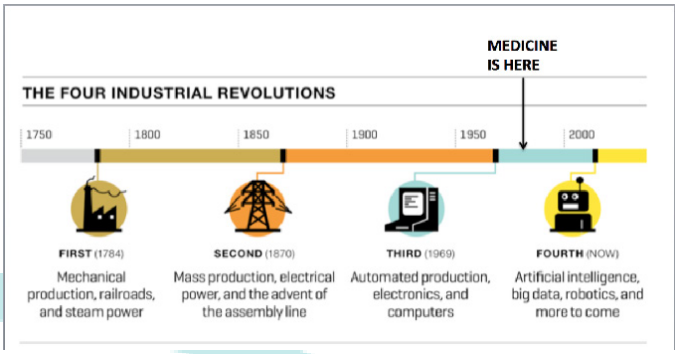
Through its Skills Strategy 2025, the SIP has identified five Key Enabling Technologies (KETs) which will underpin the future success of the science-based sector.

- big data and informatics
- synthetic biology and biotechnology
- advanced manufacturing
- formulation technology
- materials science.

Without doubt, data science in its widest sense and effective data mining is now a critically important Key Enabling Technology across the life sciences sector. PwC's Global Data and Analytics Survey 2016, (Big Decisions™) saw more than 2,100 executives surveyed on the topic (including from the health industries). Two-thirds (61%) acknowledge their companies should rely on data analysis more and intuition less. Highly data-driven companies are significantly changing how they make decisions, improve operations, or use analytics to deliver products and services.

Organisations that understand how to manage and analyse "big data" relevant to their products, to their markets and to their customers and patients, will gain the insights required to innovate and provide personalised approaches.

Through extensive consultation and analysis, the SIP has identified a 'red list' of shortage occupations, requiring immediate action to increase the availability of people with these skills in the labour market. These include informaticians, data and computational scientists (in particular bioinformaticians) and formulation scientists.



Alongside the requirement for greater computational skills, comes a general need for improved quantitative skills, and the mathematical and statistical techniques needed for accurate analysis and interpretation of data. We also know that industry access to electronic health record data (EHR) will support the sector in making medical advances, allowing skilled researchers to track the impact of therapies on patients over the long term and improving patient.

The SIP has also recognised that if educational institutions, employers, and individuals aren't keeping up with these changes, the workforce will be left behind. We are making it a priority to better understand these gaps and deliver solutions that close them.

This SIP Discussion Paper is aimed at bringing a focus to the debate. It outlines what is needed to boost skills in the increasingly important disciplines that populate the wider field, as well as work currently underway and provides the SIP with the evidence which will underpin the solutions required.

Dr Malcolm Skingle, Director GSK and SIP Chair

What the research tells us

SIP Skills Strategy 2025

The SIP's consultations and analysis have identified those areas and occupations with immediate issues, requiring urgent action.

This is our 'Red List'.

Bioinformaticians, Cheminformaticians & Health Informaticians
Increasing demand from across the economy for 'informaticians'. The specialist combination of informatics with scientific or healthcare knowledge is in short supply.

Computational Scientists
Mismatch of skills in the current workforce, with large portions educated before computational sciences formed a significant part of curricula, and current graduates lacking the type of skills sought by employers.

Health Economists
An area with shortages in supply of skilled people, requiring similar mathematical, statistical and computational underpinning skills sets as informatics areas, combined with specialist health .

Http://www.scienceindustrypartnership.com/media/529053/5202fd_sip_skills_strategy_2015_final_low.pdf

ABPI: Bridging the Skills Gap in the biopharmaceutical industry

The Association of the British Pharmaceutical Industry's (ABPI's) Bridging the skills gap in the biopharmaceutical industry provides further evidence to develop the right skills initiatives.

The report, which supersedes its 2008 survey notes that there have been a number of mergers, downsizing or closure of UK R&D activities. So medicines often arise following collaboration with academics/SMEs and that in future the biopharmaceutical company may be less likely to be the discoverer of the drug molecule.

An increasing proportion of new medicines are biologics rather than chemically-processed small molecule medicines, and new medicines are now more likely to have been created following mining of data. The medicine may also only be intended for a small subgroup of patients, i.e. a "stratified medicine" and these changes will also have greatly influenced the skills required by the industry.

Disciplines in demand (ABPI)

- **Bioinformatics/computational systems biology**
  Computational systems biology integrates experimental and computational research to understand complex biological systems. Bioinformatics uses statistical techniques to simulate biological systems to predict the activity of medicines. This discipline was rated as medium or high priority by 90% of the ABPI's respondents.

- **Data mining**
  The process of analysing data to find patterns in large sets of data. This is an emerging discipline that was not rated in 2008, but is a high priority in 2015. More than 60% of respondents agreed that urgent action is needed, while only around 10% thought it was a low priority area.

- **Chemoinformatics**
  None of the survey participants marked chemoinformatics as less than a medium priority. Like most other subjects in this section, there is a problem with both quality and number of candidates.

- **Computational chemistry**
  Companies employing computational chemists had problems finding talent mostly at higher qualification levels including PhDs, post-docs and experienced staff.

- **Metabolomics**
  Metabolomics is the "systematic study of the unique chemical fingerprints that specific cellular processes leave behind". The international Metabolomics Society and ELIXIR-UK have conducted a global assessment of the training needs in metabolomics and concluded that there is a need to urgently develop training courses to fill the knowledge gaps.

http://www.abpi.org.uk/our-work/library/industry/Documents/Skills_Gap_Industry.pdf

Example Applications

- **Precision Medicine:**
  Linking genomic information to health records to determine likely drug response in individuals.
- **Drug Discovery:**
  Both genome sequencing and the analysis of data on the structure and properties of pharmacologically active molecules is leading to the discovery of new medicines.
- **Cell and Gene Therapies:**
  Genomic information is fundamental to the research and development of these new advanced therapies.

Elixir UK: Skills gaps in the existing workforce

ELIXIR-UK is the UK node of ELIXIR, the European infrastructure for life science data. The UK node promotes the skills, training and data resource expertise of the UK scientific community in the European and Global context of ELIXIR. It is also tasked with enhancing the skills and training opportunities of UK life scientists.

As part of the broad development of training strategy for the ELIXIR project, the UK node carried out a survey of industry scientists. The complete results of this industry survey are available from the ELIXIR-UK website, but some key indicators are extracted from below.

Around 70% of bioinformaticians would like training in statistics and data-analysis methods, with a specific focus on sequencing and genomics.

Similarly, around 60% of wet-lab biologists would like to acquire skills in data visualisation, data manipulation and general statistics.

The majority of wet-lab scientists (74%) have no programming experience, and 60% perform their data analysis in Excel.

Although most respondents do have a bioinformatician with whom they can collaborate,
34% do not have a bioinformatician/statistician to whom they can turn for support.

In an attempt to specifically address the skills requirements for statistics training, ELIXIR UK, along with stakeholders such as the MRC, BBSRC, ABPI and the SIP began a project called the Statistics Training Signposting (StaTS) Project, resulting so far in a web-based resource collection (www.statschoices.org.uk) and BBSRC-funded "skills schools" run by the University of Cambridge/ELIXIR UK.

Defining the disciplines

Dr Lee Larcombe, Higher Education lead for the SIP HE Working Group has been exploring the needs of the sector, with the SIP HE Group's Members. He says "If we want, for example, to develop Apprenticeship Standards for Bioinformatics we need to be clear on the topic, skills, people and roles we are talking about."

Despite its current popularity, Lee Larcombe asserts that the term data science is not always appropriate in the context bioinformatics. Data science is distinct field - a "concept to unify statistics, data analysis and their related methods". Whilst a bioinformatician may employ a sub-set of data science techniques as part of a larger biological science skill-set, a true data scientist and a bioinformatician are very different people. This challenge of definition is further complicated by the trend to refer to the activity of business analytics as data science – again, using a sub-set of data science techniques within a business or commercial context, but distinct from true data science (and certainly from bioinformatics).

Indeed a Data Scientist (level 6) Degree Apprenticeship Standard has now been given the green light for development by the Institute of Apprenticeships (IfA) and it is the Office for National Statistics (ONS) which is leading the development of this particular new apprenticeship Standard. This builds on the pre-existing level 5 Data Analyst Apprenticeship.

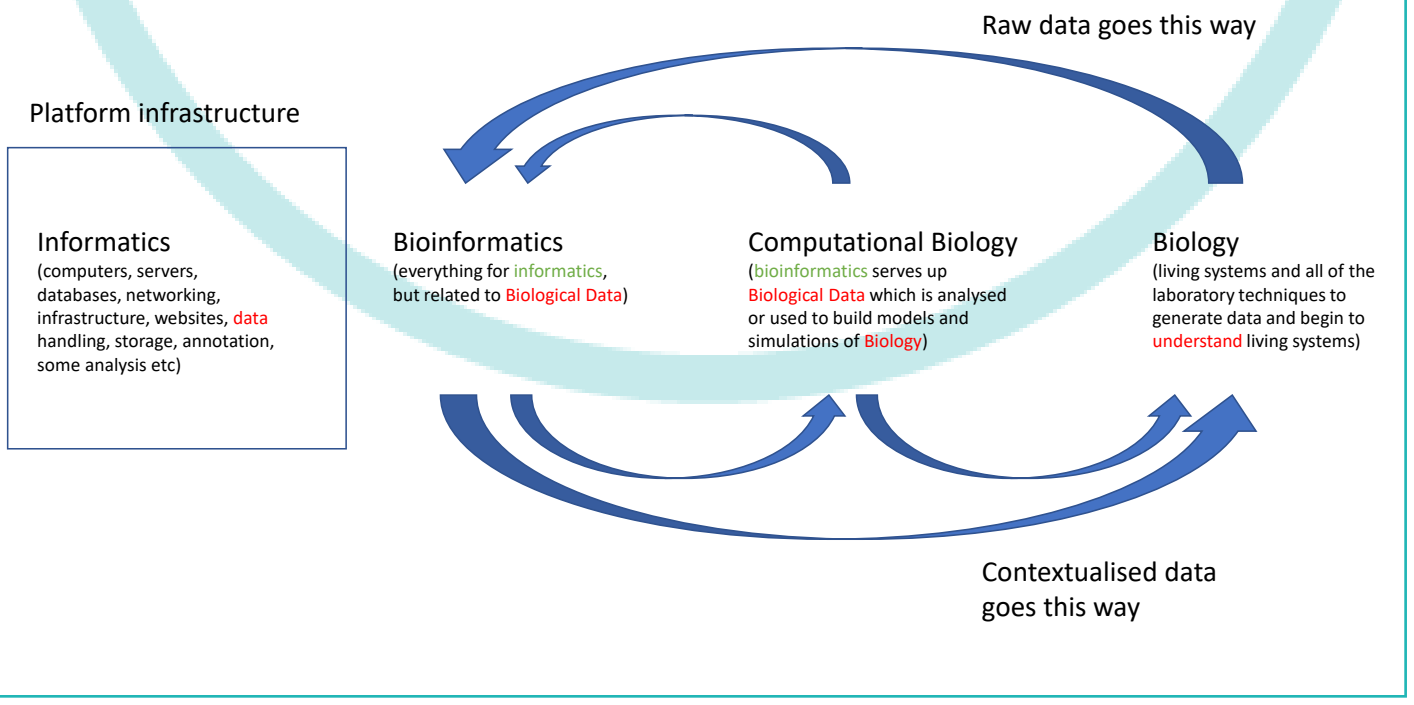
The value of bioinformatics (and related subjects such as chemoinformatics) to the life sciences comes from feeding contextualised data back into the system, "the ability to add context comes from having far broader life science subject-specific knowledge and competencies than just the data informatics/analysis skills piece." (see figure 1)

Dr Larcombe also explains that a bioinformatician cannot be expected to have sufficient expertise in all topics related to Bioinformatics as it is a diverse field – not a distinct subject - indeed there is no such thing as a universal bioinformatician (see figure 2).

He adds "an individual needs to develop skills and competencies in some of these specific areas as well as computing, coding and statistical skills, and that these will often be industry/sector/job-specific."

To this end the proposed Bioinformatics Scientist at Level 7 (Degree) under development by the SIP is clear that Bioinformaticians are life scientists who use computer and informational techniques which are applied to a range of problems in the life sciences, for example, in pharmaceutical companies in the process of drug discovery.

What is Bioinformatics?



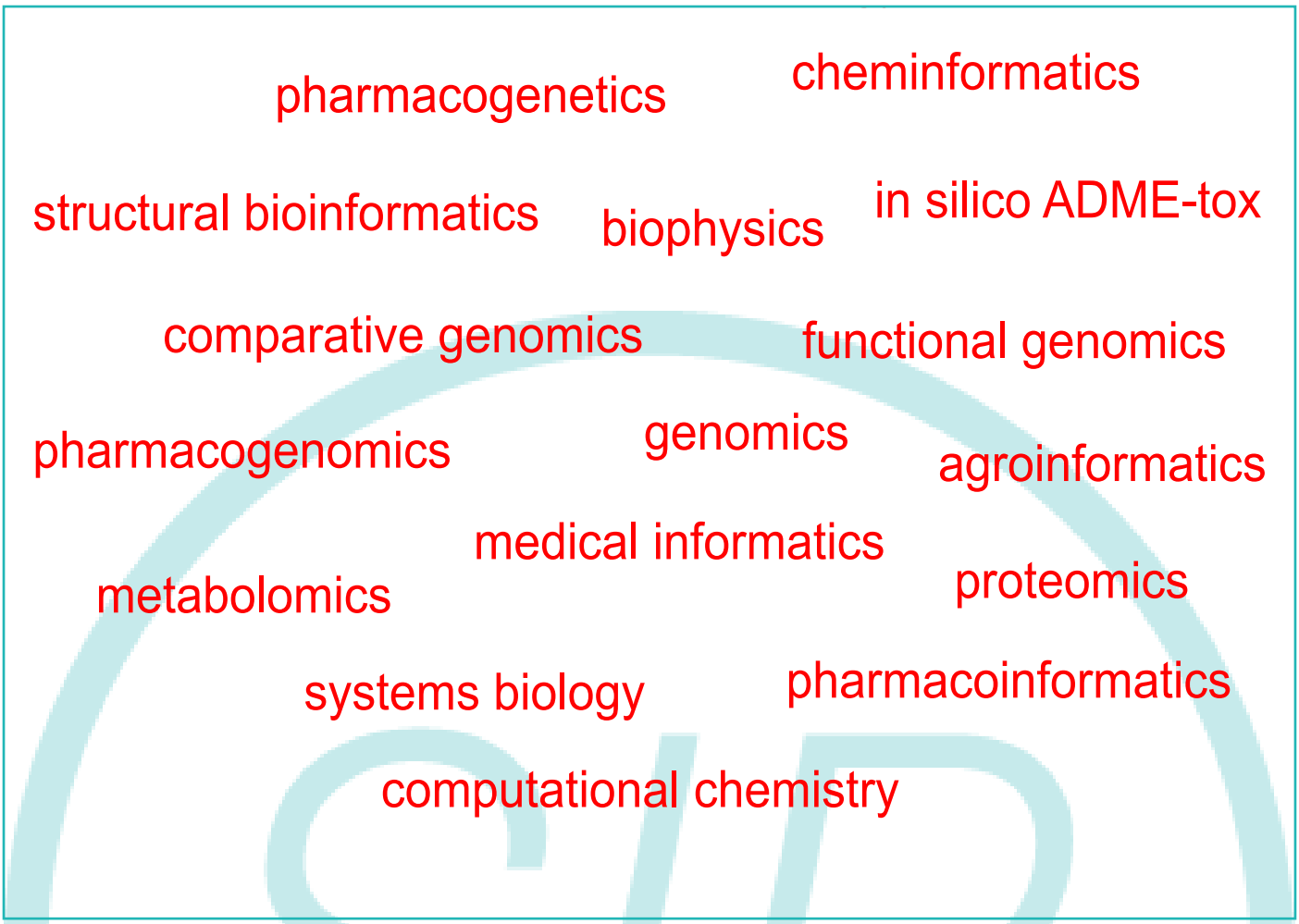


Figure 2: bioinformatics is a diverse field

Definitions

- Computer Scientist – expert in IT systems and development, uses IT tools (e.g. Oracle, Java)
- Data Scientist – expert in manipulating and analysing data, uses data science tools (e.g. Anaconda, Zoomdata, R, Python)
- Informatician – has double knowledge of a scientific area (e.g. biology) as well as IT and data methodologies – uses specialist tools (e.g. BLAST, Omicsoft, ChemDraw)
- Formulation Scientist - develops/tests formulations. Part chemist, part material scientist.

Closing the Gap

Development of a new Bioinformatics Scientist Standard – Level 7 (Degree)

Degree apprenticeships combine work, on-the-job learning and funded part-time university education.

SIP members and others are contributing to the development of a Degree Apprenticeship in Bioinformatics, which will need to meet Trailblazer standard requirements and be approved by Government. The programme will comprise a combination of mentored workplace learning, practical experience and separate study, will lead to a university degree.

Who's involved?

Unilever
GSK
MedImmune
AstraZeneca
Pfizer
FujiFilm Diosynth Biotechnology
Syngenta
UCB
Eagle Genomics
MRC
Wellcome Trust Sanger Institute

There will also be stakeholder input from ABPI, RSB, BioChemSoc and EMBL-EBI.

Occupation Title	**EXTRACT FROM THE DRAFT** **Bioinformatics Scientists Standard Level 7**
Role Overview	Bioinformaticians are scientists who use computer and informational techniques which are applied to a range of problems in the life sciences, for example, in pharmaceutical companies in the process of drug discovery. Roles require scientists who can work computationally with data derived from different life science activities, and role names and descriptions often reflect this by using slightly different names for what is essentially the same skill-set. For example, bioinformatics, computational biology, chemoinformatics, computational toxicology, Health informatics, Medical informatics, Agri-informatics.
Role Activities	• applying knowledge of computing (including coding), biology, statistics and mathematics appropriate to the discipline • having a detailed understanding of the scientific discovery process and of the role of bioinformatics in it • applying statistical research methods in the contexts of molecular, chemical or cellular biology, genomics, other 'omics disciplines, medical and population genetics research (as examples) • applying their knowledge of general biology, in depth knowledge to at least one area of biology, and understand related biological data generation technologies.
Qualifications	Bioinformaticians will generally have a background in a life sciences subject, or occasionally informatics/computer science, to BSc or equivalent degree.
Key Skills	An individual would not be expected to have all of these skills, but must have specialised knowledge in at least one life sciences area to be complemented by a range of skills across the other areas as necessary for their job role. • **Life Sciences** • **Bioinformatics** • **Computational** • **Statistics & Mathematics** • **General** - time management; project management; scientific concepts; analytical reasoning; scientific creativity etc.
	Role activities, competencies and skills adapted from: Welch L, Lewitter F, Schwartz R, Brooksbank C, Radivojac P, Gaeta B, et al. (2014) Bioinformatics Curriculum Guidelines: Toward a Definition of Core Competencies. PLoS Comput Biol 10(3): e1003496. https://doi.org/10.1371/journal.pcbi.1003496

Developing new National Occupational Standards for the science sector

Cogent Skills, the SIP's partner, has successfully secured funding from Skills Development Scotland (SDS) to develop a number of new National Occupational Standards (NOS). SDS manages the database for UK-wide occupational standards and manages the contracts with delivery partners on behalf the devolved administrations. The new NOS will be in the following areas:

1. Bioinformaticians
2. Cheminformaticians
3. Health Economists
4. Qualified Persons
5. Scientific Project Management
6. Scientific Quality

NOS are statements of the standards of performance individuals must achieve when carrying out functions in the workplace, together with specifications of the underpinning knowledge and understanding. They underpin qualification and apprenticeships in the devolved nations.

£16 million drive to boost maths skills

The Government has announced a £16 million investment to increase the quality of teaching in post- 16 maths. While maths continues to be the most popular subject at A level, almost three quarters of students with an A*-C in GCSE maths at age 16 choose not to continue studying the subject.

The investment follows a Government-commissioned review by Professor Sir Adrian Smith about how to improve 16-18 maths education in England. Adrian Smith's recommendations have been heralded as a positive step towards enabling the UK to rise to the challenge of becoming a data-literate nation.

https://www.gov.uk/government/news/16-million-drive-to-boost-maths-skills-for-post-brexit-britain

What are the current gaps and what should the UK be doing in relation to data science?

Leading academic researchers working in health data science, clinicians, industry leaders, and representatives from research funders and regulatory bodies met in February 2016. The discussions focussed on current and future capabilities in data science research.

Meeting participants identified several gaps in data science in the UK, including:

Lack of strategies for bringing data together: the group recommended that the solution is to develop the technology, governance systems and incentives to bring data together whenever possible.

Shortage of data science skills: there are currently not enough people being trained to use, process and analyse data, and there is also a lack of further training for people working in the field – as this paper notes, this needs to be increased.

Lack of clarity from regulators: Participants recommended that regulators should be much clearer on what data they will accept.

Lack of public and patient engagement in sharing their data: meeting participants considered it important to establish ways to share data and expertise, such as with an "e-Lab".

Empowering patients: Initiatives should be set up to empower patients to share their data and engage them in research. This should include reporting back to patients so they can understand the value of sharing their data.

What should the UK be doing in data science?	What should the UK not be doing?
• Improve collaborative working • Encourage patient and public engagement in sharing data • Establish funding mechanisms and support • Increase training and skills in data science • Apply best research practice • Involve regulators in clarifying data requirements • Develop quality standards • Derive value from the existing data infrastructure • Scale up initiatives that work • Further develop the national infrastructure for data science • Think globally	• Focus solely on development of drugs Work in separate silos • Wait for perfection in data science • Put national interests above opportunities to collaborate internationally • Forget the end user • Continue with a negative environment for data collection • Move away from centralising data into single databases

Extracted from Data Science for Health and Care Excellence Report

UK data developments

This Discussion Paper does not attempt to compile or endorse UK provision available in the area of bioinformatics or data science in the context of life sciences, however there are a number of major developments which are of particular note in supporting industry developments in this arena.

NHS – data skills for Personalised Medicine
NHS England is increasingly concerned with "Personalised Medicine" which it has described as move away from a 'one size fits all' approach to the treatment of patients to one which uses approaches in areas such as diagnostic tests, functional genomic technologies, data analytics and real time monitoring of conditions To achieve this it will need a large number of different healthcare scientists with different skill sets and entirely different roles including bioinformaticians and molecular pathologists. Its training is changing to deliver this new workforce as well as upskilling existing healthcare professionals. It recognises that partnership with industry is also critical.

The Genomics Education Programme (GEP) is a funded NHS initiative that ensures professionals have the genomics knowledge they need for the future. The aim is to ensure that the health service remains a world leader in genomic/ precision medicine – particularly for NHS England Genomic Medicine Centres (GMCs) and the contribution to the 100,000 Genomes Project.

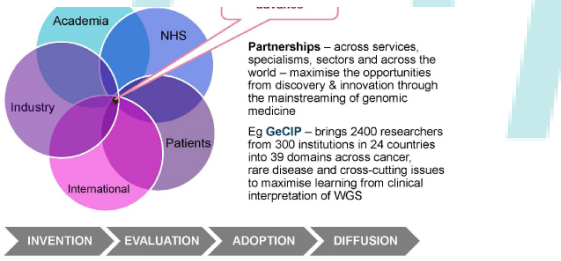


Figure 3 Healthcare Science Education: where are we now and where are we going by Ruth Thomsen, Scientific Director NHSE

Access to health data will ultimately mean that the life science industry will be able to pioneer new and suitable treatments. And analysis of NHS data can also support the personalisation of medicines and the ability to monitor the outcomes. Analysis of anonymous medical records provides for an enhanced understanding about risks and benefits of different treatments.

https://www.genomicsengland.co.uk/the-100000-genomes-project/

The Farr Institute
The Farr Institute is a UK-wide research collaboration involving 21 academic institutions and health partners in England, Scotland and Wales. Publically funded by a consortium of ten organisations led by the Medical Research Council, the Institute is committed to delivering high-quality, cutting-edge research using 'big data' to advance the health and care of patients and the public. It is positioning the UK as a world leader in health informatics research through scientific discovery and the enhancement of patient and public health.

http://www.farrinstitute.org/

Health Data Research UK (HDR UK)
The Medical Research Council (MRC) is leading on the establishment of a multi-funder UK institute for health and biomedical informatics research to transform the UK medical informatics research landscape. It will be a national, interdisciplinary research institute that will capitalise on the UK's renowned data resources and research strengths. The Institute will develop the capacity (people and skills) and methods to accelerate the pace and scale of health and biomedical data science to deliver a step change in UK capabilities.

https://www.mrc.ac.uk/about/institutes-units-centres/uk-institute-for-health-and-biomedical-informatics-research/

EMBL-EBI
The European Bioinformatics Institute (EMBL-EBI), helps scientists realise the potential of 'big data' in biology. It champions open data in life sciences, providing freely available data and bioinformatics services to the scientific community. It also provides advanced bioinformatics training to scientists at all levels, in particular providing hands-on bioinformatics training to help experimental biologists make the most of the wide range of data resources. EMBL-EBI maintains the world's most comprehensive range of freely available molecular data resources.

http://www.ebi.ac.uk/

SIP Strategy recommendations

The ability to meet the skills challenge will determine the extent to which the Key Enabling Technologies will spread through the science sector and into the wider economy.

Skills research and policy needs to come to terms with the implications of these changes.

The analysis presented here shows that there are important quantity and quality issues that need addressing across the entire workforce. This observation leads in turn to important implications for prioritisation and targeting.

In practical terms it emphasises the importance of higher level technical cross sector roles such as bioinformatics and health economics, as well as the wider skills sets that all scientists need to communicate and spread innovation.

Key recommendations

More opportunities to develop practical skills for the workplace in higher education: Employers frequently raise concerns about the practical skills of new graduates. Not all employer can recruit people with 2+ years' experience. More employers need to take a role in developing practical and cross cutting skills for the workplace, including apprenticeships, internship and placement schemes, engaging students in live workplace projects, and offering graduate training/ mentoring schemes.

Build mathematical & statistical skills throughout the education system: Mathematical, data analysis and statistical skills underpin many areas of science. It is essential that a good grounding in mathematics and statistics is embedded from school age. In addition, training in maths and stats must be available and accessible to address identified weaknesses in the existing scientific workforce.

Build practical computing skills across stem education: Data handling, computing, programming and software using skills are touching roles at all levels. The need for these skills is growing and should be core elements of STEM related education. This would provide learners with the skills to work with company specific software and equipment as soon as possible.

Facilitate transferability between public and private scientific workforces: Facilitating the transfer of staff between industry, NHS and academia has many benefits. These include developing multidisciplinary individuals and teams, increasing cross-sector collaboration, enhancing transfer of ideas and innovation, and creating a more flexible workforce with better opportunities.

Support SMEs in engaging with education and skills delivery: Whilst the skills issues for SMEs are the same as those facing the industry as a whole, SMEs have significantly higher barriers in engaging with education and skills development. With an increasingly fragmented industry, it is important that small companies are supported in playing their part in developing the skilled workforce of the future. This includes offering manageable work based learning opportunities - including shorter placements, live project opportunities, apprentice sharing schemes, and developing financial incentive packages.

Develop cutting edge apprenticeship standards

The SIP's Occupational Map continues to demonstrate the skills gaps the sector faces, and is a particularly useful tool in identifying the need for new apprenticeship standards. These new standards take the focus away from gaining qualifications and put the emphasis on developing an agreed and end-assessed set of skills, knowledge and behaviours necessary to perform key job roles in the sector. By developing standards in newly developing disciplines, individuals are offered a genuine alternative pathway to university study and employers get the high level skills they require.

References

1. Cogent Skills Biofutures: the life sciences workforce in context 2014
2. Copyright: © 2017 Larcombe L et al. This is an open access article distributed under the terms of the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
3. Training needs in metabolomics, authors and affiliations: Ralf J. M. Weber, Catherine L. Winder, Lee D. Larcombe, Warwick B. Dunn, Mark R. Viant
4. McKinsey (2011) Big data: The next frontier for innovation, competition, and productivity
5. NIH (2000) Working Definition of Bioinformatics and Computational Biology

T. 01925 515220
E. sipmembers@cogentskills.com
W. www.scienceindustrypartnership.com

Supported by
Cogent skills
for science industries