NERC Environmental Omics: A Community Vision

Mission: To deliver innovative omics solutions for a changing world.

Vision: To enable world-leading environmental research with real-world impact through the application of omics approaches

Author Prof. Peter Kille

1 Introduction

1.1 Executive Summary

Technological advances have provided researchers with an unprecedented ability to comprehensively characterise living systems across scales from the molecular level to entire ecosystems. In particular, omics approaches that consider entire layers of biological organisation (e.g., genomics, metabolomics, phenomics, etc) have matured, enabling fundamental research as well as providing solutions that improve environmental management. The assessment of ecosystems through time and at all levels of organisation, from individuals to populations and communities, allows us to monitor the impact of a changing environment and plan appropriate mitigation, adaption and restoration. This community vision is designed to maximise the potential of these developments in supporting the breadth of NERC research and ensuring translation of these innovations into solutions that deliver economic and environmental benefits. Our delivery plan will consider technological advances, human resources, business drivers, partnerships and infrastructure required to realise our mission 'to deliver innovative omics solutions for a changing world'.

1.2 Community Action Plan

NERC requested the community develop its vision on env omics requested that the research community develop its vision for a future strategy for environmental omics research following on from the 2010 NEOMICS report [1]. To achieve this, NERC established an expert steering group (Annex 1) to advise on the key areas to be included within the vision, to determine the process of community engagement and to contribute to articulating how the vision can support NERC's delivery plan. In additional to steering group meetings, a workshop was held (17th April 2018) allowing the community an opportunity to directly provide evidence and advice on the future direction of environmental omics and to consider the capabilities needed to support a future strategy in omics research. The recommendations emerging from this consultation process are summarised below:

- **Enabling world-leading environmental research** by supporting research programmes that utilise omics to deliver the NERC strategy.
- **Develop the scientific expertise and stakeholder skill base** required to allow the environmental community (researchers, policy makers, regulators and industry) to fully exploit omics tools.
- **Converting innovations to solutions:** Work in partnership with Innovate UK, UK government departments, devolved administrations, stakeholders and industry to ensure omics technologies are converted into innovations that assist environmental management and monitoring.
- **Promote cross-UKRI interdisciplinary strategic programmes** that utilise omics approaches to address environmental global challenges and to promote the discovery of novel biological compounds for societal gain (e.g. Blue Biotech).
- **Identify** immediate and future **infrastructure requirements** (**data generation, integration and analysis**) to support omics research and input long-term priorities into the UKRI Infrastructure Roadmap.

1.3 Overview

The 'NERC Environmental Omics Strategy' (NEOMICS, 2010)[1] was developed to ensure that emerging technologies supported NERC strategy (2007-2012) [2] *Next Generation Science for Planet Earth*. Considering ongoing technological advances, together with the maturation of various omics platforms and approaches, it was felt timely to update in the context of the current NERC Delivery Plan (2016-2020) *The Business of the Environment [3]* as well as opportunities arising from the formation of UKRI.

NERC has invested in omics over the last 20 years through a series of thematic programmes (Environmental Genomics - EG, Post Genomics and Proteomics - PGP, Mathematics & Informatics for Environmental 'Omic Data Synthesis - EOS) and by supporting significant infrastructure (NERC Biomolecular Analysis Facility - NBAF). The current strategy update aims to provide a coherent vision of how the long-term investment in omics can be harnessed to address environmental global challenges and benefits for society.

Central to this current mission is the delivery of innovative omics solutions that allow us to better understand and monitor the changing world. Through exploiting state-of-the-art approaches such as: **aDNA** (ancient DNA) to study the effects of historical natural events (e.g. climate change) and human interventions (e.g. historical deforestation) on ecosystems; **eDNA** (environmental DNA) and metabarcoding to provide powerful tools to better understand non-native species invasion, species conservation, and ecosystem processes. These new approaches have delivered outputs that have transformed archaeological and ecological research agendas. Metagenomics and eDNA have also yielded legislative regulatory tools for monitoring threatened species (DEFRA - Great Crested Newt Detection WC1067 [4]) and water quality assessment (EA - Diatom Metabarcoding SC140024/R [5]). **Omics tools provide us with both understanding and capacity to predict how our planet works and tools to assist us in responsibly managing its precious resources**.

2 Delivery Plan

2.1 Omics is enabling world-leading environmental research

Omics approaches are being widely used in NERC blue-skies science programmes in addition to being central to several strategic programmes. The underpinning nature of the omics approaches restricts the ability to directly report on its full foot-print within NERC science, however, since all work supported at NBAF exploits omics we can assess those funded applications which exploit NBAF facilities, these have, on average, attracted £50m over 17 awards per annum between 2012-2018 (Figure 1). Highlight Topics (HT), such as '*eDNA: A Tool for 21st Century Ecology'* have been designed to encourage the development of omics tools within specific areas. Many other HT and science programmes have drawn on omics tools to deliver their objectives with the recent '*Emerging Risks of Chemicals in the Environment'* drawing on omics to both assess the ecological impact of chemicals by providing high-throughput tools for food-web and community analysis as well as exploiting omics to provide mechanistic impact of chemicals to range of sentinel species. Omics approaches are essential for the delivery of high-quality

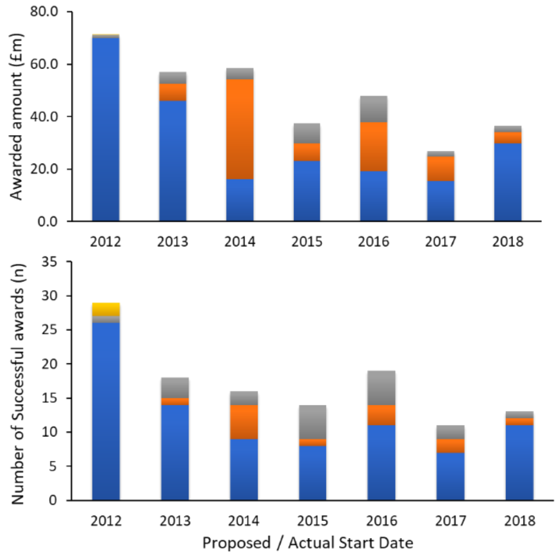

Figure 1: Number and value for successful NERC grants that were supported by NERC Biomolecular Analysis Facility (NBAF).

NERC science. Of the 240 grants applications rated at international competitive (research score of 7 or higher) incorporating omics submitted between 2012-2018 124 (52%) have attached funding. Omics continues to deliver high quality publications outputs showing an above average normalised citation impact when compared to environmental science research generally. Key examples of how omics approaches are being exploited to support NERC delivery plan are provided in Box 1.

While some areas of environmental science have substantial expertise and capacity in omics, others need investment to stimulate activity and ensure NERC science continues to be internationally leading. As our understanding of living systems increases, new directions of research emerge. It is important that novel areas are identified and supported both through discovery science and strategic funding routes.

> **Box 1: World Leading Research to Support the NERC Delivery Plan.**
>
> **Benefiting from Natural Resources**
> - **Develop** omics tools to assess natural capital.
> - **Characterise** ecological networks that support heathy waters and heathy soils.
> - **Mine** pathways underlying novel natural product chemistry.
>
> **Resilience to Environmental Hazards**
> - **Mitigate** pollution impact on ecosystems and health through application of 'Precision Environmental Science'.
> - **Deploy** eDNA tools as sentinels for invasive species, vulnerable populations and infectious disease.
> - **Predict** interactions between natural and human-generated hazards.
>
> **Managing Environmental Change**
> - **Understand** the biology underpinning global biogeochemical cycles.
> - **Identify** functional evolutionary drivers: how the microbiome, genome and epigenome interact to deliver phenotype.
> - **Conserve** biodiversity by supporting sustainable ecosystems.

New insights to fundamental questions: From Fragments to Facts

Ancient DNA: learning from the past

Technological and methodological innovations continue to yield omics data from unprecedented sources, thereby transforming established areas of research. The ability to recover fragments of DNA from archaeological and paleontological material (**aDNA**) [6] provide insights on evolution of individual species as well as allowing us to reconstruct complete ecosystem (**sedimentary ancient DNA - sedaDNA**) [7]. These approaches provide a powerful temporal perspective in which to study evolutionary processes (including plant and animal domestication), together with previous climate change events. Archaeogenomics can contextualise and characterise individuals while also allowing us to reconstruct the ecosystems in which they lived. The potential of these techniques is only now being realised, and there remains a vast untapped potential to learn from the past to understand the changing world in which we live. To ensure international leadership, this area requires both targeted infrastructural support as well as development of the skill base to exploit these unique data.

Environmental DNA (eDNA): monitoring ecosystems

The ability to recover fragmented DNA from environmental samples (**eDNA**) is also having profound impact in understanding ecosystems (examples can be found in special issue of biological conservation 2015 [8]). Applications ranging from non-invasive sampling of organisms through to the reconstruction of food-webs using gut or faecal material is furnishing ecologists and biologists with an unparalleled suite of tools in which to study key communities (special issue Molecular Ecology 2019 [9]). The ability to track individuals, characterise communities and identify the interactions within communities, without requiring direct observation, can transform these disciplines. From assessing population numbers of a species of potentially high conservation risk inhabiting inaccessible terrain to characterising plants visited by specific pollinators, exploitation of omics approaches to deliver answers to previously intractable questions.

Uncovering microbial community biodiversity

Metagenomic approaches are now well-established and provide a valuable tool for studying the structure and function of microbial communities. However, there are significant limitations in studying a homogenised system [10]. A key limitation is the inability to assign functional pathways to a specific microbe or microbial group. With the continual exchange of useful genetic material between bacteria, the mobilome, a systems-level understanding of functional potential of these communities can only be achieved by determining the genetic composition at an individual level [11]. The empowering technology of **single cell genomics and transcriptomics** is enabling this insight by determining hundreds to thousands of individual microbial genomes [12].

Unlocking Biological 'Dark matter': Exploring life

Omics has supported an explosion in our ability to explore biology revealing previously undiscovered organisms (even whole taxa) [13] and demonstrating the functional importance of biomolecules previously thought to unimportant. Global projects focused on microbial content of sea (Global Ocean Sampling Expedition) [14] and land (Earth Microbiome project [15]) continue to reveal new phylogenetic groups as well support novel functional pathways.

As genomics is applied to the full diversity of life it generates a secondary challenge, that of understanding "dark matter" transcriptomics, all those transcripts, which are currently annotated as "unknown" in so many environmental specimens. Developing comparative analysis platforms for sequence data to identify which are species-specific and which are process-specific and the development of tools such as CRISPR-Cas9 for environmental species to unlock these secrets and develop next generation synthetic biology solutions for industry. Unlocking these pathways has significant biotechnological potential. Processes such as mollusc biomineralization which produce incredibly hard and durable materials and where synthetically mimicry of such a natural-processes requires hugely complex technology and lots of energy have the potential to deliver natural products of the future.

Extreme environments as sources of novel compounds for the benefit of society

Microbes are found in the most extreme environments on earth and have found ways to adapt and thrive, from the deep ocean biosphere to volcanic calderas [16]. The majority of these organisms cannot currently be cultured in laboratories, leaving genomics as the primary tool to unlock their extraordinary abilities [13]. Furthermore, natural systems are in a continual arms-race for resources, with this struggle being especially profound in microbial communities [17]. Evolutionary processes have resulted in an array of strategies designed to liberate resources and provide their hosts with a competitive advantage. The succession of fungal and prokaryotic microbes responsible for driving nutrient cycling has have evolved multiple **natural product chemistries** [18]. Mapping out the chemical ecology underlying these novel pathways and the associated vast array of novel small molecules / secondary metabolites, has significant potential for exploitation within industrial biotechnology. A prime example is that these organisms host a plethora of novel antibiotics as well as providing a reservoir for antimicrobial resistance (AMR) [19].

Disease monitoring in wild populations

The omics tool-kit delivers methods for surveillance of **wildlife pathogens** (e.g., an ability to unlock life cycles) together with comparative genomic insights into combating these infectious diseases. Pathogens have catastrophic effects on specific populations, together with entire classes of organisms, resulting in impacts to ecosystems globally. For example, white-nose syndrome in North American bats has disrupted trophic cascades, leading to an increase in insect pests [20]; chytrid fungal infection has caused global declines of amphibians [21]; and *Hymenoscyphus fraxineus* infection resulting in ash dieback has left a major aesthetic mark on the countryside, while devastating associated ecosystems [22]. The emergence of novel diseases within threatened species with small genetic diversity, such as Tasmanian devil facial tumour disease (DFTD), pose significant threats that can lead to extinction [23]. These natural hazards are exacerbated through the interaction with added stressors (e.g., climate change or pollution), resulting in increased pathogenicity. Application of omics technology helps us to better understand what is happening as disease outbreaks occur and also informs strategy for mitigation of their consequences.

Quantifying the World's biodiversity

The upward trajectory in sequencing capability and capacity shows no sign of slowing. The announcement of a biological 'moon-shot', the Earth Biogenome Project (EBP), aimed at unifying the international community behind sequencing every Eukaryotic genome on the planet, is both timely and achievable (https://www.earthbiogenome.org/) [24]. The **comparative genomics** resource that this initiative will generate will transform evolutionary biology, while having untold benefits in all areas of science. From application, such as providing templates for synthetic biology, to fundamental understanding of evolutionary process, the EBP will change our understanding of every living system. The UK component of the project, provisionally entitled 'The Darwin Tree of Life Project' (https://www.sanger.ac.uk/news/view/genetic-code-66000-uk-species-be-sequenced), will have a primary focus on sequencing species from the British Isles and UK protectorates. Linked to this will be a focus around key observatories (e.g. St Kilda, Priests Pot and Wytham Woods), yielding immediate value-added benefits by complementing conventional long-term datasets. Significant uplift can be achieved by strategically adding to resources with selected re-sequencing detailed studies to investigate the relationship between genome architecture and a changing environment.

Genotype-environment interactions in wild species

Understanding how the environment helps shape phenotype has been significantly enhanced through **multi-omics analysis** [25], which allows us to characterise genome architecture, epigenetic modifications (miRNA, DNA methylation and histone modifications), transcriptomics and associated microbial communities. Unravelling the causative relationships that link environmental changes through alterations at the level of genome, epigenome and transcriptome with the phenotype, represents a significant challenge for integrative biosciences. Exploiting Omics for examining adaptive and acclimative processes will allow us to model evolutionary trajectories predicting resilience of individuals, populations and ecosystem to global change events.

Understanding community relationships

Many organisms share intermittent interactions: some being one-way dependencies (e.g. parasitism); some interactions are mutually beneficial, to such a degree that the partners have become interdependent (e.g. lichens); while others have less closely coupled relationships (e.g. microbiomes or rhizosphere interactions) where benefits are derived but not essential (commensalism). Omics approaches are enabling us to dissect **symbiosis** in all of its forms [26], characterising the relationships, benefits, co-dependencies and co-evolutionary processes that may be involved. Critical interaction support some of the most iconic ecosystems such as marine reefs which really on the interaction between coral and their algal symbiosis, an interaction that is under specific pressure given global warming [27]. These complex interactions extend our understanding into how complex communities interact to yield sustainable ecosystems. This is especially pertinent to soils, where the rhizosphere represents interactions between plant, microbe (bacteria & fungal) and macro/micro invertebrates and that maintain sustainable terrestrial environments by supporting natural and agricultural ecosystems [28]. The excitement that surrounds the more loosely coupled interactions, such as the microbiome and the rhizosphere, is that we can dissect and even manipulate the composition of these communities to combat pollution or engineer a more resilient host/ecosystem.

Pollution and environmental health

A recent analysis revealed that 16% of global human deaths are linked to pollution, and our ecosystems are at no less risk [29]. Biomes delivering essential ecosystem services are being assailed by cocktails of synthetic chemicals in the form of agrochemicals, industrial chemicals, consumer products, pharmaceuticals and veterinary medicines, all of which reach our green and blue spaces in significant amounts. Added to this are various particles and gases in the air, the by-products from the combustion of waste products. Pollutants range from these atmospheric contaminates to the visible and invisible plastics in our soils, rivers and seas. A complex array of legislation attempts to protect both human health and ecosystems from the effects of pollution, but it has two significant shortfalls. Firstly, evidence is generated on single compounds under largely laboratory conditions and secondly only a small number of representative organisms are tested. These limitations are partially addressed by applying safety margins to compensate for differential species sensitivities, mixture effects, changes between laboratory and field conditions and longer exposure durations in the real world. Application of omics technologies, combined with a comparative mechanistic understanding of biological systems, has the potential to detect the full spectrum of pollutants in our environments identify the associated changes and hence generate predictive frameworks that can more accurately identify risks to both human health and ecosystems for species with different physiologies. These novel approaches together represent a vision of ***Precision Environmental Health*** *[30]*.

Towards real time monitoring

The combination between microfluidics and nanopore sequencing provides the realistic future possibility for field omics analysis and real-time automatic genomics monitoring of the environment. These tools may benefit research where field-based analysis will dynamically informing monitoring regimes, but they are also important to support human and environmental protection. For example current microbiological test for beach-waters take 48 hours and exclude viral load, issue that can be addressed given development in sampling and sequencing technologies. However, data streams containing such information provided challenges demand develops in artificial Intelligence or neural networks to interpret the real time change to provide human accessible monitoring information.

2.2 Develop the scientific expertise and stakeholder skill base

There is a need to make omics training accessible to academic researchers and well as governmental and industrial partners. This requires an integrated approach to ensure omics education is available, at an appropriate level, to researcher of all types, ages and skill levels (Box 2). Current capacity has been developed through programs such as the NERC '*Advanced Training Short Courses*', however, the dynamic nature of omics technology development poses challenges to ensure that training reflect current state-of-the art.

Training the next generation of researchers (DTPs and CDT training)

Omics analysis is becoming pervasive to environmental research on living systems. Ensuring that the future generation of NERC researchers are appropriately informed about the breadth of state-of-the-art omics approaches that is available and can access appropriate training in experimental design, technical execution and informatic analysis of these tools is essential. Many NERC DTPs already incorporate introductory omics and informatics courses. However, there is significant variability in access to support for detailed execution. We recommend, in consultation with the DTP network, the development and commissioning of a central suite of training to cover the majority of the applications of environmental omics that can be offered to the full cohort of NERC PhD students. There is a significant demand for postgraduates with well-developed skills in environmental informatics and omics data analysis. This demand would be best addressed through supporting a Centre of Doctoral Training (CDT) in "Ecological and Environmental Informatics", which would create a critical mass of highly skilled postgraduates that would enhance both research and data underpinning decisions by government agencies and industry.

> Box 2: Investing in People
>
> Postgraduate training
> - Embedding omics training in DTPs
> - Develop omics CTD program
> - Continuing education to support industry and legislative uptake
>
> Targeted training
> - Upskilling to support professional development.
> - Innovation instruction to ensure rapid uptake of new technical breakthroughs.
> - Speciality instruction delivered by experts for the community.
>
> Exchange programmes
> - Secondments into industry and government
> - Interdisciplinary exchange programmes
> - Technology transfer fellowships, industry-academia and academia to industry

Targeting training, 'Upskilling' and speciality instruction

The continued development of new omics platforms, approaches and analytical tools demands that researchers and principal investigators are continually upskilled as part of their professional development. There are two distinct requirements – conceptual and practical: information on practical delivery as to how omics can be applied within a research area and instruction in the techniques for execution and analysis. These speciality courses are best delivered by experts who actively use the approaches in question. These courses can also be offered to DTP students where appropriate (see DTPs and CDT above).

Cross-disciplinary end-user exchange programmes

Many successful omics projects benefit from the contributions from interdisciplinary team best executed through cross-disciplinary exchange. Exchange between researchers and industry/end users are essential so that researchers gain a better understanding of end user needs ensuring research will be more aligned to impact whilst end users need to understand new developments and potential provided by omics. These exchanges are straightforward to arrange either as part of a PhD programme or through targeted secondments. Postdoctoral researchers who do not have the exchange funded as part of their programme can apply for targeted funding. The logistics for these exchanges for principal investigators is much more challenging given their other academic commitments. Supporting these exchanges has significant benefit for ongoing and future research and for building legacy collaborations and the creation of UKRI should facilitate such schemes across the different research councils.

2.3 Converting innovations to solutions

The importance of protecting the natural environment is recognised at every level of government. It is explicit in the UN Sustainable Development Goals and embodied in international, national and local legislation. In the UK all devolved governments have stated their commitment to environmental protection through several key planning documents including Defra's 25 Year Environment Plan, e.g. DEFRA's *A Green Future* [31], The Environment (Wales) Act 2016 [32] and the Scotland's *State of the Environment* report [33]. Omics innovations have the potential to deliver solutions to support the delivery of every aspect of environmental management (Box 3). These solutions can support the business sector that has developed to support the environmental aspirations defined by legislation. The Environmental Management and Assessment (IEMA) and UK environmental goods and services sector (EGSS) representing ~2% of the UK's national GDP in 2015 [34]. This sector is predicted to grow as environmental testing becomes privatised and new environmental protection legislation is implemented. Additional markets aimed at early toxicity/ecotoxicology screening for pharmaceutical companies, chemical companies and companies related to food products are estimated to rise to >$20 billion globally in 2021, with equivalent food safety platforms projected to inject $17 billion into the global economy. Thus, environmental protection represents big business.

> **Box 3: Omics innovations to solutions**
>
> Measure
> - **Define** abnormal against normal
> - **Design** tools that are rapid, accurate, robust and cost effective.
> - **Develop** real time autonomous platforms.
>
> Monitor
> - **Quantify** biodiversity and assess ecosystem function.
> - **Survey** wild-life pathogens, zoonotics and AMR.
> - **Regulate** invasive species and vulnerable populations.
>
> Manage
> - **Easily communicated** and legally defensible.
> - **Inform** remedial action.
> - **Track** efficacy of intervention.
>
> Mitigate
> - **Protect** nature and enhance landscapes.
> - **Secure** productive and biologically diverse ecosystems.
> - **Reduce** pollution and waste.

DEFRA led Centre of Excellence (CoE) for DNA based applications is being developed to ensure that government takes full advantage of the opportunities to protect and improve the environment offered by molecular and omics level tools. A key focus of this would be supporting the UK's implementation of international environmental legislation, for example Water Framework Directive (WFD), Marine Strategy Framework Directive (MSFD), Habitats Directive (HD), Environmental Impact Assessment (EIA) and Invasive Non-Native Species (INNS) requirements.

As part of this initiative, consultation with researchers and end users in May 2018 identified clearly the priorities, benefits and barriers to exploiting DNA based methods for environmental and ecological assessment. The application areas currently exploiting these approaches fell into four broad headings, ecosystems and biodiversity, animal and plant health, food safety and environmental pollution (Box 4). Key barriers to translation of research tools into this sector were identified as lack of co-ordinated funding to assist technology transfer together with the need for bioinformatics skills and knowledge transfer.

The formation of the UKRI with the integration of Innovate UK and business driven programs such as Industrial Strategy Fund provides an opportunity to drive further development of omics tools against government priorities. In addition, partnering with IEMA and EGSS industrial sectors will realise the economic benefits stemming from these technological advances. The omics Strategy mission will be to ensure that these exciting technical innovations are not limited to the research arena but have impact to society through regulatory and commercial application.

2.4 Promote partnership, interdisciplinarity and globalization

Interdisciplinary opportunities

Omics as a discipline represents a common language linking biological communities from human health through the bioeconomy to the environmental sciences. A key strategic objective is to ensure Environmental omics contributes to the significant synergies within UKRI (Box 5).

Beyond a shared core technology, the 'complementarity' between various antecedent

Box 4: Research and development of DNA based methods by Environment Protection organisations

Ecosystems and biodiversity
- Species detection and community analysis in freshwater, estuarine, marine and terrestrial systems
- Invasive non-native and protected species surveillance
- Pollinator analysis
- Population and evolutionary genetics
- Fungi
- Crop wild relatives

Animal and Plant Health
- Plant health and pathogens
- Genome sequencing of disease and pathogens

Food Safety
- Food traceability and authenticity

Enviornmental Pollution
- Pollution source tracking
- Bioaerosol monitoring

Cross Cutting
- Technological advances
- Bioinformatics
- Standardisation and validation
- Quality assurance and proficiency testing

research councils' remits in the omics area is significant. For example, deriving the genomic basis of phenotype is important in understanding human genetic disease (MRC), the development of new crops (BBSRC) and in understanding evolutionary and adaptive processes (NERC). Shared science questions and grand challenges form the foundations on which to realise the benefits of shared knowledge between different communities.

Environmental science is fundamentally an interdisciplinary pursuit, for example, exploiting physical science to support remote sensing and earth observation, social science to map out the human–environment interface and engineers to deliver the solutions to major environmental challenges. UKRI will provide an extended opportunity for interdisciplinary teams to work together in pursuit of common goal, integration of omics approaches will be key to the success of these initiatives. Omics provides a powerful tool to aid in the understanding of conservation challenges but often the drivers are social, ranging from competition for natural resources to economic growth. Greater integration between researchers from different fields will provide better continuity from identification and understanding of environment issues to tangible solutions.

Of the 17 UN Sustainable Development Goals 6 map directly onto the Environmental agenda; sustainable soils (15. Life on Land), healthy waters (14. Life below water/6. Clean water and sanitation), climate change (13. Climate Action), agroecology (12. Responsible consumption and production) and pollution and chemical hazards (11. Sustainable cites & communities / 12. Responsible consumption and production). In contributing to these aspirational goals, research can directly realise societal impact. Omics forms a powerful component of projects attempting to provide solutions to these global challenges. From tools for monitoring and assessing biodiversity in soils and waters, to understanding biome resilience to climate change, through to providing the evidence base for better understanding of pollution impact on ecosystems, omics has a major contribution to make. Advocating the integration of omics techniques when addressing these societal challenges is important. Many GCRF hubs have already integrated these approaches into their programmes, but additional advocacy is needed to ensure the full societal benefits of the omics tool box is realised.

Box 5: UKIR areas of current and future synergy

Shared science questions
- Sequencing life on earth: Earth BioGenome Project.
- Organism Interactions: Symbiosis.
- Genotype to phenotype.

Interdisciplinary teams to achieve common goals
- Bioremediation: From pollution management to ecosystem restoration.
- Product value chains – from use to disposal
- Understanding our past to inform our future.

Achieving shared societal impact
- One Health Global Network: Linking human, animal and environmental health
- Ecosystem service provision: Species and habitat recovery and restoration.
- Providing sustainable water management for people and ecosystems.

Opportunities for business: Natural product discovery

Translation opportunities from environmental sequencing: generation of cleaner, greener alternative solutions for industrial processes and generation of novel products. For example: global market for industrial enzymes is expected to grow from nearly $5.0 billion in 2016 to $6.3-7.2 billion in 2020 with annual growth rates of between 4.6-5.8% (other examples: enzymes for food and beverage industry, animal feed and biofuels). With our enhanced capacity to sequence whole genomes cheaply, we are now in a position to exploit environmental species for novel enzymes and novel compounds in a way not previously possible, especially if linked to high through-put screening. This would be a real cross-council research field.

2.5 Infrastructure: Data generation, integration and analysis

Empowering the environment research community to exploit the opportunities provide by innovation in omics is the provision of **experts and enablers**, access to an expert skill-base together with provision of training and tools to support independent capability. Defining a one-size-fits-all strategy that addresses the infrastructural requirements for Environmental omics is challenging, given several variables: (a) the **rate** of technological advance; (b) **access** to appropriate local infrastructure; (c) the **maturity** of different analytical platforms; (d) **diversity** of applications; and, (e) the **expertise** level of researchers and research communities.

New platforms appear, and technologies retire at an unprecedented rate. This raises two issues; the continued capital investment needed to provide researchers with access to state-of-the-art technologies and the dissemination of best practice for the use of these new technologies. The institutional provision of omics infrastructure is extremely variable, with some researchers having access through dedicated local facilities whilst others rely on provision by regional or national centres. The area of genomics has matured rapidly, whilst other disciplines, such as **metabolomics and proteomics**, have not experienced the same developmental trajectory, leaving **expertise and equipment at a premium**. The diversity of applications mirrors the strength and breadth of environmental research, with projects ranging from the transcriptomic analysis of organisms from extreme environments to the characterisation of DNA from 8,000-year-old sediments. **eDNA, dietary DNA and aDNA** all require specialist facilities, robust sample protocols and rigours data analysis pipelines. Expertise in both the physical preparation of samples and analysis of resulting data is extremely inconsistent with some of the best environmental researchers being inexperienced in the application of omics in their fields whilst have world leading expertise. It is essential that the environmental community develop infrastructure network to share best practice and facilitate access to specialist facilities and expertise.

Addressing this heterogeneity in availability represents the key strategic infrastructure requirement of 'Environmental Omics'. Three approaches are key to delivering enhanced research outputs, these include:

a) **provision of national centres of excellence** with state-of-the-art platforms for data generation, analytical pipelines (established and new) and an expert knowledge base;
b) **delivery of training** (informatics and wet-lab) to expand the knowledge/skill base within the environmental community;
c) **development of tools** to support community empowerment and democratisation of omics;
d) **access to informatic capability and capacity** to support analytical requirements;
e) **promotion of collaboration** to share expertise and support interdisciplinarity.

However, any infrastructural investment must exploit the established national capability providing support for environmental specific resource and developing fields, whilst exploiting regional or national centres and commercial providers where appropriate.

Data generation

Access to DNA/RNA sequencing capacity is not a current limitation to research aspirations. Selecting the appropriate technology (platforms and methodologies) that can assist delivery of a specific research objectives is often complex, requiring access to expertise. Specific specialist facilities, experienced user base and physical infrastructure are required for the preparation of specific sample types, e.g. single cell analysis or preparation of archaeological samples. It is important that these specialist infrastructures be linked to expert users, but available to the wider community. Furthermore, neglected areas in environmental science, such as environmental metabolomics and proteomics, need significant support to assist further development and exploitation. However, these techniques are heavily used in the BBSRC and medical communities. UKRI is ideally placed to facilitate greater co-ordination and exchange of these facilities between the different research communities, with a dynamic distributed infrastructure where a hub coordinates access to a range of expertise across the spectrum of different science disciplines.

Key to realising the full potential of Omics is data

Each technological advance allows us to acquire more data at lower cost. The increase in genetic sequence data capture has exceeded Moore's law for the past 14 years (www.genome.gov/sequencingcostsdata/). The opportunities this provides for environmental research are far reaching, however, the challenge posed by handling the data is also significant. Careful consideration must be given to the infrastructure used for storage, policies/mechanisms for data sharing, analytical tools and integration of omics data with the full spectrum of environmental metadata. In isolation, omics data has significant value but when combined with the full spectrum of related data, whether this be land-use acquired by remote sensing through to detailed phenotypic measurements, the value of the omics data is significantly enhanced. In 1999, the astrophysics community identified the direct cost to research caused by having no coordinated data structure for their observations at 333 FTE / annum [35] and thus justified the development of the platform that now allows all astrological data to be access through a single portal (ALADIN [36] / SIMBAD portals [37]). The recent explosion of environmental omics data places our community in a parallel situation, although publication should ensure data disclosure of raw data or metadata is not included or not digitally accessible (this therefore does not represent 5* open data). The recent announcement of the 'Constructing a Digital Environment' provides an ideal platform on which to build an omics data layer that can be integrated with

Therefore, investment in the long-term infrastructure to support the integration omics environmental data into the 'Digital Environment' together with the related ecological and geophysical data is a priority.

Big Data Infrastructure

The requirement for access to high performance computing (HPC) to support environmental omics data analysis is essential. The heterogeneity of data type, the need for specialist analytical pipelines and the dynamic nature of bioinformatic software development possess significant challenges for classical centrally managed HPC infrastructures. No one hardware configuration will support all types of omics analysis, with genome assemblies demanding terabytes of directly accessible RAM whilst other informatic processes can efficiently use parallel processing clusters. The emergence of specialist processor architectures dedicated toward fulfilling specific information tasks may assist a restricted suite of applications but will not address the heterogeneity of applications. The exponential increase in the size of reference data sets (bacterial metagenome resource held at NCBI now exceeds 2 petabytes) presents challenges for researchers who wish to interrogate these repositories. The integration of omics data with complementary environmental data requires standardisation and ontologies whilst the need for transparent reproducibility of information analysis needs novel innovation and approaches. However, many of these computational issues are not limited to 'Environmental Omics' data and therefore solutions should be addressed through the development of an integrated UKRI e-infrastructure to be integrated into the UKRI infrastructure roadmap.

A range of e-infrastructure solutions for bioinformatics outside the classical HPC system have been developed. Systems such as the Genomic Virtual laboratory (MRC CLIMB implemented infrastructure) and CyVerse exploit a core open cluster architecture providing users with the ability to customise and scale their computing requirements. These systems can be applied at a local institutional level or as distributed networks and can be easily transferred to large-data centres. The NERC Omics Community successfully trialled an equivalent system (EOS cloud), demonstrating the viability of this type of computer architecture to deliver for the breadth of its community. New approaches to delivering High-Performance Computing (HPC) capability that have been customised to support the requirement of the individual researchers provide exciting developments. One large advantage that these implementations provide is the ability to record and share the complete configuration used to analyse a specific dataset, promoting reproducibility of these complex pipelines.

It is essential that we drive forward towards 5* open data: this will require data to be accessible but also linkable. The development of Open Data Cubes in environmental science provides a vehicle for anchor omics data within the physical environment, providing the temporal and spatial reference to allow integration of the data with remote sensing and wider earth observation information i.e. land use or weather pattern. However, there is an increasing need requirement for conserve vocabularies and controlled ontologies (i.e. Genome Standards Consortium) if we are to ensure that environmental omics data is truly linkable and discoverable.

Box 6: Requirements of a Scalable e-infrastructure for Bioinformatics

Performance

- Scalable
- Flexible
- Customisable

Portability

- Locally accessible
- Compatible with local infrastructure
- Mobile to data repositories

Provenance

- Reproducible
- Towards 5* open data
- Linkable

3 Bibliography

1. Kille, P., Field, D., Bailey, B., Blaxter, M., Morrison, N., Snape, J., Turner, S., and Viant, M. (2010). NERC Environmental 'Omics Strategy. (https://nerc.ukri.org/research/funded/programmes/omics/neomics-report/: NERC).
2. Gaffney, O., and Jeffery, H. (2007). Next Generation Science for Planet Earth 2007-2012. (Natural Environment Research Council: Natural Environment Research Council).
3. NERC (2013). The Business of the Environment (Natural Environment Research Council).
4. Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Griffiths, R.A., Foster, J., Wilkinson, J., Arnett, A., Williams, P., and Dunn, F. (2014). Analytical and methodological development for improved surveillance of the Great Crested Newt. In DEFRA WC1067.
5. Kelly, M., Boonham, N., Juggins, S., Kille, P., Mann, D., Pass, D., Sapp, M., Sato, S., and Glover, R. (2018). A DNA based diatom metabarcoding approach for Water Framework Directive classification of rivers. (Environment Agency).
6. Kistler, L., Ware, R., Smith, O., Collins, M., and Allaby, R.G. (2017). A new model for ancient DNA decay based on paleogenomic meta-analysis. Nucleic Acids Res *45*, 6310-6320.
7. Smith, O., Momber, G., Bates, R., Garwood, P., Fitch, S., Pallen, M., Gaffney, V., and Allaby, R.G. (2015). Archaeology. Sedimentary DNA from a submerged site reveals wheat in the British Isles 8000 years ago. Science *347*, 998-1001.
8. Goldberg, C., Strickler, K., and Pilliod, D. (2015). Special Issue: Environmental DNA: A powerful new tool for biological conservation. Biological Conservation *183*, 1-102.
9. Creer, S., Jonsson, M., Roslin, T., Stone, G., Symondson, B., and Traugott, M. (2019). Special interactions, ecological networks and community dynamics.
10. Prosser, J.I. (2015). Dispersing misconceptions and identifying opportunities for the use of 'omics' in soil microbial ecology. Nat Rev Microbiol *13*, 439-446.
11. Siguier, P., Gourbeyre, E., and Chandler, M. (2017). Known knowns, known unknowns and unknown unknowns in prokaryotic transposition. Curr Opin Microbiol *38*, 171-180.
12. Gawad, C., Koh, W., and Quake, S.R. (2016). Single-cell genome sequencing: current state of the science. Nat Rev Genet *17*, 175-188.
13. Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F., Darling, A., Malfatti, S., Swan, B.K., Gies, E.A., et al. (2013). Insights into the phylogeny and coding potential of microbial dark matter. Nature *499*, 431-437.
14. Yilmaz, P., Kottmann, R., Pruesse, E., Quast, C., and Glockner, F.O. (2011). Analysis of 23S rRNA genes in metagenomes - a case study from the Global Ocean Sampling Expedition. Syst Appl Microbiol *34*, 462-469.
15. Gilbert, J.A., Jansson, J.K., and Knight, R. (2014). The Earth Microbiome project: successes and aspirations. BMC Biol *12*, 69.
16. Hedlund, B.P., Dodsworth, J.A., Murugapiran, S.K., Rinke, C., and Woyke, T. (2014). Impact of single-cell genomics and metagenomics on the emerging view of extremophile "microbial dark matter". Extremophiles *18*, 865-875.
17. Ghoul, M., and Mitri, S. (2016). The Ecology and Evolution of Microbial Competition. Trends Microbiol *24*, 833-845.
18. Smanski, M.J., Zhou, H., Claesen, J., Shen, B., Fischbach, M.A., and Voigt, C.A. (2016). Synthetic biology to access and expand nature's chemical diversity. Nat Rev Microbiol *14*, 135-149.
19. Moloney, M. (2018). Natural Products as a Source for Novel Antibiotics. 37 *8*, P689-710.
20. Alves, D.M., Terribile, L.C., and Brito, D. (2014). The potential impact of white-nose syndrome on the conservation status of north american bats. PLoS One *9*, e107395.
21. O'Hanlon, S.J., Rieux, A., Farrer, R.A., Rosa, G.M., Waldman, B., Bataille, A., Kosch, T.A., Murray, K.A., Brankovics, B., Fumagalli, M., et al. (2018). Recent Asian origin of chytrid fungi causing global amphibian declines. Science *360*, 621-627.
22. McMullan, M., Rafiqi, M., Kaithakottil, G., Clavijo, B.J., Bilham, L., Orton, E., Percival-Alwyn, L., Ward, B.J., Edwards, A., Saunders, D.G.O., et al. (2018). The ash dieback invasion of Europe was founded by two genetically divergent individuals. Nature Ecology & Evolution *2*, 1000-1008.
23. Woods, G.M., Fox, S., Flies, A.S., Tovar, C.D., Jones, M., Hamede, R., Pemberton, D., Lyons, A.B., and Bettiol, S.S. (2018). Two decades of the impact of Tasmanian Devil Facial Tumour Disease (DFTD). Integr Comp Biol.
24. Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., Durbin, R., Edwards, S.V., Forest, F., Gilbert, M.T.P., et al. (2018). Earth BioGenome Project: Sequencing life for the future of life. Proc Natl Acad Sci U S A *115*, 4325-4333.
25. Acharjee, A., Kloosterman, B., Visser, R.G.F., and Maliepaard, C. (2016). Integration of multi-omics data for prediction of phenotypic traits using random forest. BMC Bioinformatics *17*, 180.
26. Chaston, J., and Douglas, A.E. (2012). Making the Most of "Omics" for Symbiosis Research. The Biological Bulletin *223*, 21-29.
27. Baker, D.M., Freeman, C.J., Wong, J.C.Y., Fogel, M.L., and Knowlton, N. (2018). Climate change promotes parasitism in a coral symbiosis. ISME J *12*, 921-930.
28. van der Heijden, M.G., de Bruin, S., Luckerhoff, L., van Logtestijn, R.S., and Schlaeppi, K. (2016). A widespread plant-fungal-bacterial symbiosis promotes plant biodiversity, plant nutrition and seedling recruitment. ISME J *10*, 389-399.
29. Landrigan, P.J., Fuller, R., Acosta, N.J.R., Adeyi, O., Arnold, R., Basu, N.N., Balde, A.B., Bertollini, R., Bose-O'Reilly, S., Boufford, J.I., et al. (2018). The Lancet Commission on pollution and health. Lancet *391*, 462-512.
30. Bradburne, C., and Lewis, J.A. (2017). Personalizing Environmental Health: At the Intersection of Precision Medicine and Occupational Health in the Military. J Occup Environ Med *59*, e209-e214.
31. Department for Environment Food & Rural Affairs (2018). A Green Future: Our 25 Year Plan to Improve the Environment. (HM Government).
32. Welsh Government (2016). Environment (Wales) Act. W. Government, ed.
33. Critchlow-Watton, N., Dobbie, K.E., Bell, R., Campbell, S.D.G., Hinze, D., Motion, A., Robertson, K., Russell, M., Simpson, J., Thomson, D., et al. (2014). Scotland's State of the Environment Report. (Scotland's Environment Web.).
34. Statistics, O.f.N. (2015). UK environmental goods and services sector (EGSS): 2010 to 2015. (OGL).
35. Kurtz, M.J., Eichhorn, G., Accomazzi, A., Grant, C.S., Murray, S.S., and Watson, J.M. (2000). The NASA Astrophysics Data System: Overview. Astron. Astrophys. Suppl. Ser. *143*, 41-59.
36. Bonnarel, F., Fernique, P., Bienaymé, O., Egret, D., Genova, F., Louys, M., Ochsenbein, F., Wenger, M., and Bartlett, J.G. (2000). The ALADIN interactive sky atlas. Astron. Astrophys. Suppl. Ser. *143*, 33-40.
37. Wenger, M., Ochsenbein, F., Egret, D., Dubois, P., Bonnarel, F., Borde, S., Genova, F., Jasniewicz, G., Laloë, S., Lesteven, S., et al. (2000). The SIMBAD astronomical database. Astron. Astrophys. Suppl. Ser. *143*, 9-22.

4 Annex 1: Steering committee and editors

Omics Steering Committee:

Chair: Prof. Peter Kille (Cardiff University)

Prof Thomas Meagher (St Andrews)

Prof Mark Bailey (CEH)

Prof Terry Burke (Sheffield University)

Prof Richard Nichols (QMUL)

Dr Louise Johnson (Reading University)

Prof David Lunt (University of Hull)

Prof Neil Hall (Earlham Institute)

Dr Jake Bundy (Imperial College)

Dr Andy.Nisbet (Natural England)

Dr Federica di-Palma (Earlham Institute)

Expert Editorial comments provided by EOS Advisory Group:

Dr Félix Forest (Kew)

Dr Ioanna Katsiadaki (CEFAS)

Dr Robin Allaby (Warwick)

Dr Danielle Ashton (Environment Agency)

Dr Melody Susan Clark (BAS)

Professor James Prosser (University of Aberdeen)

NERC Staff

Emma Payne - Programme Manager, National Capability Commissioning

Simon Kerley - Head of Terrestrial Ecosystems

Hannah Pearson - Senior Programme Manager, National Capability Commissioning