THE VALUE OF DIGITISING NATURAL HISTORY COLLECTIONS

Report prepared for the Natural History Museum, London

Danail Popov, Priyanka Roychoudhury,
Helen Hardy, Laurence Livermore, Ken Norris

OPEN ACCESS

https://doi.org/10.3897/rio.7.e78844

The Value of Digitising Natural History Collections

Danail Popov[‡], Priyanka Roychoudhury[‡], Helen Hardy[§], Laurence Livermore[§], Ken Norris[§]

[‡] Frontier Economics, London, United Kingdom
[§] The Natural History Museum, London, United Kingdom

Corresponding author: Helen Hardy (h.hardy@nhm.ac.uk)

Received: 03 Dec 2021 | Published: 07 Dec 2021

Citation: Popov D, Roychoudhury P, Hardy H, Livermore L, Norris K (2021) The Value of Digitising Natural History Collections. Research Ideas and Outcomes 7: e78844. https://doi.org/10.3897/rio.7.e78844

CONTENTS

1. EXECUTIVE SUMMARY 3

2. INTRODUCTION 11

3. APPROACH 13

4. FINDINGS 16

5. CONCLUSION 41

FUNDING PROGRAM 43

AUTHOR CONTRIBUTIONS 43

REFERENCES 44

SUPPLEMENTARY MATERIAL 49

ENDNOTES 49

THE VALUE OF DIGITISING NATURAL HISTORY COLLECTIONS

Abstract

The Natural History Museum, London has been creating digital data about collections for many years, with a formal Digital Collections Programme since 2014. Efforts to monitor the outcomes and impact of this work have focused on metrics of digital access, such as download events, and on citations of digital specimens as a measure of use. Digitisation projects and resulting research have also been used as impact case studies, highlighting areas such as human health and conservation. In 2021, the Museum decided to explore the economic impacts of collections data in more depth, and commissioned Frontier Economics to undertake modelling, resulting in this report. While the methods in this report are relevant to collections globally, this modelling focuses on benefits to the UK, and is intended to support the Museum's own digitisation work, as well as a current scoping study funded by the Arts & Humanities Research Council about the case for digitising all UK natural science collections as a research infrastructure. This study focuses on digitisation in the round, not distinguishing between different collection types or levels of data creation at this stage. Three methods have been used: first, analysing five key thematic areas or sectors where data from natural science collections are likely to lead to benefits; secondly, analysing typical returns on investment in scientific research; and thirdly, examining the efficiency savings that can be reinvested in research if data are available freely and openly. Together, these methods confirm benefits in excess of £2 billion over 30 years, representing a seven to ten times return on investment.

Keywords

natural history collections, natural science collections, value of research, value of collections, economic benefits, digitisation, biodiversity conservation, invasive species, medicines discovery, agricultural research and development, crop wild relatives, mineral exploration

1. EXECUTIVE SUMMARY

The value of research enabled by digitisation of natural history collections can be very large, creating benefits in excess of £2 billion (Fig. 1)

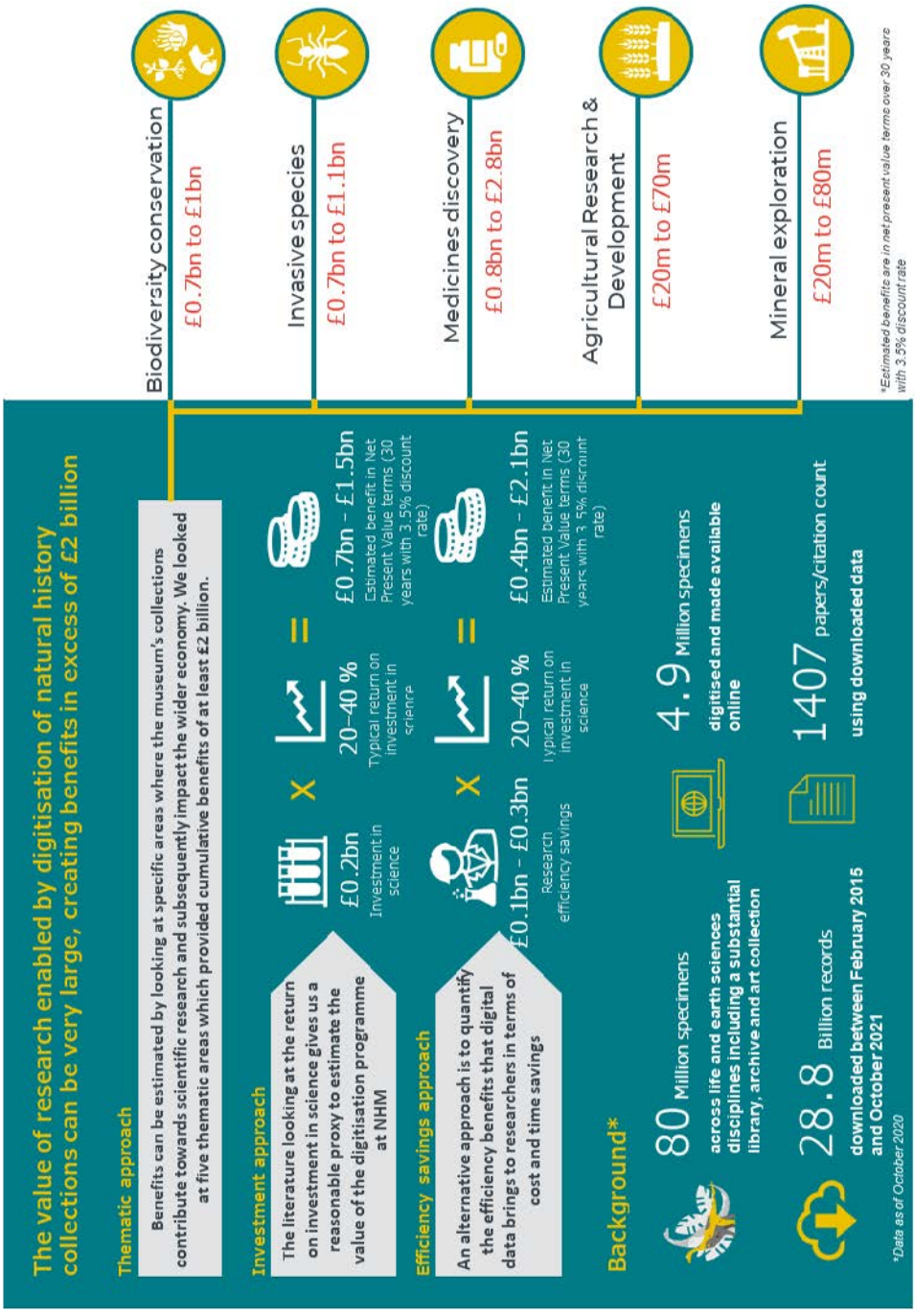

Figure 1. Overview of thematic, investment and efficiency savings approach including five key thematic areas: biodiversity conservation, invasive species, medicines discovery, agricultural research and development, and mineral exploitation.

Natural history collections are an invaluable resource enabling fundamental scientific research with wide applications

Natural history collections are hugely important as a scientific resource and infrastructure, as well as a public and cultural experience. They help scientists to study fundamentally important issues for humankind such as what is happening to biodiversity, how to develop new medicines, improvement of agricultural crops and the impacts of climate change on animals and plants. Collections contain historical data which serve as spatial and temporal baselines against which modern observations and collections can be compared, and serve as a repository where animals and plants cannot be readily found in nature, allowing us to understand how and why nature is changing and plan a more sustainable future.

The Natural History Museum, London (the Museum) has a vast collection of over 80 million objects, one of the largest and most historically and geographically diverse collections in the world. The collection covers multiple disciplines from botany and zoology through to palaeontology and mineralogy, as well as a substantial library, archive and art collection. The Museum collection is used by scientists all over the world to address fundamental questions about the past, present and future of life on Earth and of the solar system. Museum scientists (of whom there are more than 300) have, for example, been studying the genetic data of historical wheat samples to determine if ancient varieties of wheat are more resistant to pests, diseases and climatic conditions such as drought (Walkowiak et al. 2020). Similarly, by studying the Museum's vast collection of bat specimens (of which around 10,000 specimens of three major families are currently being digitised), scientists can discover new information about known and unknown viral threats including where viruses originate and how they transfer to humans (Jacklin 2021).

The Museum collection can currently be accessed by scientists and the public but less than one per cent is on public display, and use of collections behind the scenes is primarily constrained to physical visits. For instance, in order to study a specimen, scientists have to come to the museum and find the item(s) (with the support of curators), then spend time to generate their specific data set before they can begin their analysis. Physical visits are clearly necessary and important, but they are also very time consuming and expensive. Analysis of SYNTHESYS+ project data, shows that the average project for researchers visiting the Museum involves around 10 days of researcher time and costs the researcher more than £3,500. This includes travel costs, which apart from taking a significant portion of the researcher's overall visit time, also tend to be carbon intensive.

Digitising collections would improve access to crucial data enabling more research in a variety of domains from human health to the environment

Digitisation of a collection is the process of creating digital data relevant to collections objects. There are different levels to which a collection can be digitised depending on the medium, purpose, availability of equipment for digitisation and budget among other things. Digitisation may include transcribing data to help collections be discovered, e.g. what something is, when it was collected, where and by who; creating 2D and 3D images; and further detailed analyses that generate digital data such as microscopy and computerised tomography scans, chemical, and molecular or genomic analyses. The Museum doesn‘t distinguish between ‚catalogue‘ data and fuller digital representations of collections objects, they are treated as a continuum of scientifically useful information. This study does not differentiate the impact of different levels of digitisation.

Digitisation of collections provides benefits which vary according to the intended purpose of the digitisation process for an organisation. Benefits include:

- **Accessibility**: Physical collections can be made available to a global audience (which includes researchers and the general public) so that access isn't restricted to those who are able to visit, saving travel costs and time.
- **Searchability**: The data associated with a digital copy of the collection through the process of digitisation enable users to find relevant content more easily and accurately, which helps to increase research efficiency, collaboration between researchers, and integration with other collections.
- **Preservation**: Using a digital copy can mean that users don't need to interact with the physical objects as much as they would without digitisation, therefore reducing the costs incurred due to damage from handling and transportation of specimens. Additionally, where digital discovery increases demand for physical access or where only physical access can meet a research need, this access can be more efficiently and effectively targeted when more is shared digitally in advance.
- **Interaction**: As a digital copy isn't restricted by the need for physical space for an audience to interact with it, digitisation helps to open up larger parts of the collections than previously possible to a wider audience.

Wider accessibility and searchability of the data held in specimen collections enables researchers to extract valuable information for conducting research activities and innovation in the economy across multiple areas such as biodiversity conservation, climate change and human health.

Figure 2. The Museum collections of which ~6% are digitised and already have high usage in scientific publications. Data for period February 2015 to October 2021.

The Museum has already made progress in this area – as it stands, around 4.9 million specimens are available freely and openly on the Museum's Data Portal, and there is evidence that these are being used extensively by the scientific community as illustrated by Fig. 2. Indeed, over five years more than 28 billion records have been downloaded in over 427,000 download events via the Data Portal and/or major aggregator the Global Biodiversity Information Facility (GBIF), and more than 1,400 scientific publications worldwide have cited Museum specimen data via GBIF (GBIF 2021).

The value to the UK of research enabled by digitisation can be very large, running into billions

The digitisation process creates valuable data for research which then leads to benefits for the economy. Previous literature has looked at the benefits of having greater access to data generally and the impact of digital data for science, using various approaches such as rapid evidence assessments, user-based surveys and cost benefit analyses. Analysis of data equity in the UK summarised by the OECD (2015b) suggests that greater access to data was worth £25 billion to businesses in the UK a decade ago and this was projected to increase by a factor of eight over the following five years.

The economic value of research data and the related contribution to economic development has also been studied. For example, Houghton and Sheehan (2009) looked at the effects of increasing accessibility to public sector research outputs in Australia, and estimate that increased accessibility generates a return of AUD 9 billion over 20 years (equivalent to £4.9 billion[*1]). Houghton et al. (2010) estimate that the open access archiving mandate for US Federal Research Agencies over a transitional period of 30 years may be worth around USD 1.6 billion–USD 1.75 billion (equivalent to £0.9 billion–£1 billion[*1]). These figures would be significantly higher than the estimated

cost of implementing open access to the archives. Jisc (Beagrie and Houghton 2014) conducted a study on the economic impact of three UK data centres*16 and estimated that each of them could bring a twofold to tenfold return on investment over 30 years. A recent study by Beagrie and Houghton (2016) looking at the impact of the European Bioinformatics Institute (which makes life science data freely available to the global research community) estimated future research impacts worth £6.9 billion.

There are billions of pounds of value from digitising the Museum collection

In order to value the benefits of digitisation, we have examined a handful of specific areas in which digitisation of the Museum's collections can be expected to boost research, ultimately leading to benefits for the UK society and the world. Although these are only a subset of research areas through which the Museum's collections impact the UK and the global economies, they are some of the key areas where impact is most likely to be seen*2. These thematic areas provide specific examples of how the benefits from digitisation can be expected to materialise in practice. The cumulative estimated value across these areas provides an aggregate estimate of the value of digitisation. We have looked at five such areas:

- **Biodiversity conservation**: digitisation can enhance taxonomic knowledge, which can improve the detection of threatened species. This in turn enables conservation efforts to be put in place which reduce the rate at which species decline, maintaining the ability of ecosystems to provide vital services for humanity.
- **Medicines discovery**: digitisation can improve accessibility of samples and consequently the range of samples tested for the drug discovery and commercialisation. There are very large economic benefits from successful drug discovery. Even if digitisation leads to a very small increase in the rate of discovery, the benefits are large.
- **Invasive species**: digitisation allows for better and faster detection of invasive species which have costs for the UK economy estimated to be £2 billion a year. Reducing the frequency of losses from threats can thus lead to substantial economic benefits.
- **Agricultural R&D**: Digitisation can help in the discovery and/or improve the understanding of the genetic traits of Crop Wild Relatives (CWR), wild plant species that are genetically related to cultivated crops. This can enhance breeding of crops which are environmentally friendly, have higher yields and are disease resistant.

- **Mineral exploration**: digitisation can improve the accuracy of existing geological data and help develop enhanced geological data from the museum's rich collections of minerals, ores and rocks. This accelerates the geological discovery process and reduces the risks and thereby costs of exploration significantly.

The estimated total benefit in these areas exceeds £2 billion, as summarised in Fig. 3.

These estimates can be corroborated by applying a more top-down approach, looking at the aggregate returns to investment in science. A comprehensive literature review by Frontier Economics (2014) found that the typical private rates of return are in the region of 20%–25% with social returns being two to three times higher. Applying these numbers to an investment in digitisation of £200 million[*3], we obtain benefits ranging from £0.7 billion to £1.5 billion in present value over 30 years[*4]. This level of investment would generate discovery level data for the majority of UK natural science collections (an estimated 130–150 million specimens in total) or could be used to provide

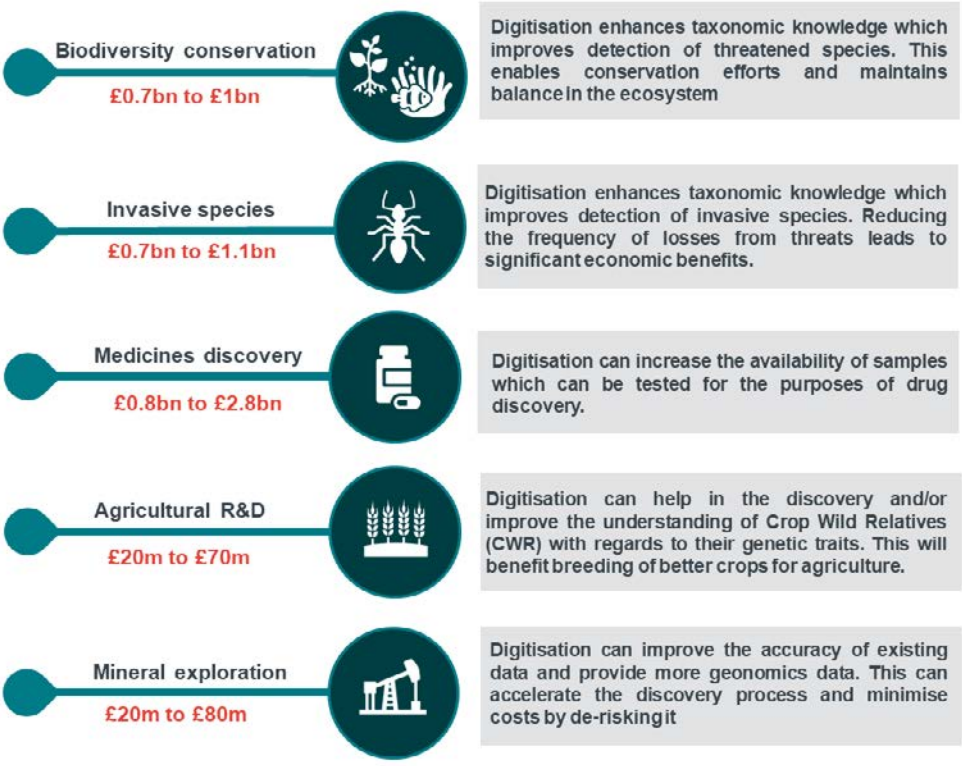

Figure 3. Valuing pathways to impact across five key areas: biodiversity conservation (£0.7bn–£1bn), invasive species (£0.7bn–£1.1bn), medicines discovery (£0.8bn–£2.8bn), agricultural research and development (£20m–£70m), and mineral exploitation (£20m–£80m). All estimates are in NPV terms over 30 years using a 3.5% discount factor.

enhanced data for the 80 million items at the Natural History Museum – for example fully geo-referenced records, with images and additional analyses where relevant and associated data infrastructure costs.

An alternative approach is to quantify the efficiency benefits that digital data brings to researchers in terms of cost and time savings. Exact cost savings will differ between researchers based on the extent to which the availability of digital data replaces the need for physical visits or simply reduces their duration. On top of efficiencies for existing researchers, digital data can spur new research in multiple disciplines which would not have otherwise happened.

We use available data on digital downloads of Museum specimen records in combination with SYNTHESYS+ cost information for physical visits to produce indicative estimates of value. Our approach is to estimate what it would cost to recreate the digital data used by researchers (based on download statistics) by conducting physical visits. We estimate benefits ranging from £0.4 billion to £2.1 billion in present value over 30 years (Fig. 4).

It is worth noting that all our estimates of value are based entirely on data from secondary sources. They should be interpretated as approximate values. Furthermore, one of the most significant benefits of digitisation is that it can foster blue skies research in any number of areas which are impossible to predict and value *ex-ante*. These can be expected to come about through more and better linking of data across collections in the UK, EU and rest of the world including associated datasets in the fields of climate, population, health and others.

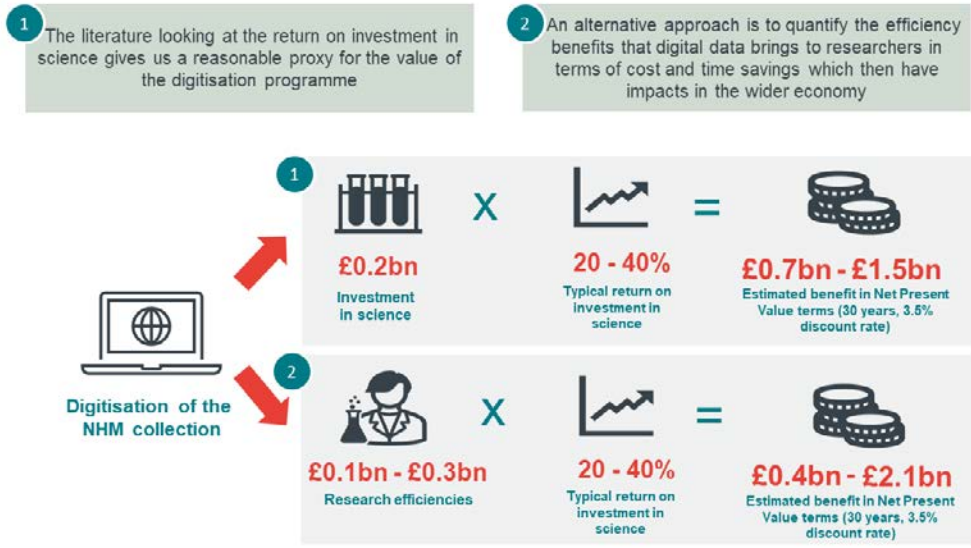

Figure 4. Summary of investment and efficiency approaches to valuing digitisation of Museum collection.

We have looked at only a small number of areas in which tangible estimates of economic value could be expected. There are many more areas which could and would benefit from digitisation but have not been considered in this study, including for example climate, and many areas of humanities research including art and design, and decolonisation.

Future research could add significant value to this work by studying the existing and potential user base for digital natural science collections data more closely, and attempting to track the ultimate impacts of the data use which goes beyond academic publications and onward citations. Survey work and case studies with users could be used to examine the impacts of specific digitisation initiatives. This would help to understand how digitised data is useful in practice and the exact pathways through which it can produce benefits for the user groups. In addition, this work deliberately focuses on the research benefits of digital data, but there are of course extensive wider possible benefits including education and public engagement.

2. INTRODUCTION

Natural history collections are an extremely valuable resource which enable fundamental scientific research with wide applications from human health to climate change and the environment. They help us to study important issues for planet Earth. The historical data contained in natural history collections serve as spatial and temporal baselines against which we can compare observations and new collections today, which in turn allows us to understand what is happening around us and the implications for our prosperity and wellbeing. The Natural History Museum is one of the largest UK museums. It attracts around 5 million visitors a year. As well as being a major attraction for the tourism and education sectors, the museum is a leading science research centre with 300 scientists producing around 700 scientific papers a year with their international collaborators.

The Museum has one of the largest natural history collections in the world with over 80 million objects covering multiple disciplines from entomology, botany, zoology through to palaeontology and mineralogy. The Museum collection is used by scientists to address fundamental questions about life on Earth and further afield, including the past and future of our solar system. Museum scientists have, for example, been studying the genetic data of its historical wheat samples to determine if ancient varieties of wheat are more resistant to pests, diseases and climatic conditions such as drought. By studying the Museum‘s vast collection of bat specimens (of which three key families – around 10,000 specimens – are currently being digitised), scientists can discover new information about known and unknown viral threats, including where viruses originate and how they transfer to humans (Jacklin 2021).

Physical space constrains how much of its collection the Museum can display at any one point in time – only around 27,000 specimens (significantly less than one per cent of the collection) are currently on display. Further, given that the vast majority of its collections have not yet been digitised at specimen level, locating specimens can be a time-consuming physical process even within the Museum, and even more challenging for those who are not able to visit the museum in person. As a result, a large proportion of specimens may be underused, when otherwise they could form part of current or future research or allow the general public to enjoy the full wealth of the collections. Further, the COVID-19 pandemic has heightened the need for remote/digital access to museum collections (Atkinson 2020) and indeed there is evidence that the process of digitisation is being accelerated across Europe (Zuanni 2020).

Digitisation is the process of creating digital data relevant to collections objects, from basic data enabling collections discovery (e.g. what, when, where), through 2D and 3D imaging to deeper analyses such as chemical and molecular analyses. By digitising its collection fully, the Museum aims to allow the vast natural wealth contained in its collections to reach a wider audience. By becoming more accessible to both the national and international research community and the general public, this has the potential to unlock substantial benefits. Providing access to a much larger group of scientists and other experts could significantly increase the benefits to multiple sectors of the economy. Given that there have been nearly 29 billion digital record downloads and over 1,400 publications directly citing the Museum's digital collections[*5] with only about 6% of its collections digitised, the potential benefits of expanding access are very large. Current papers cover areas such as climate change, citizen science and agriculture which lie beyond the more traditional areas of research associated with the research at the Museum (e.g. taxonomy), allowing a broad range of disciplines to benefit from the collection.

As well as improving access, digitisation improves the searchability of data as users can find relevant content more easily and quickly. An additional benefit of digitisation is the preservation of the collection: using a digital copy will mean researchers would not necessarily need to interact with the physical collection as much as they would without digitisation, or when physical access is needed it is better targeted, therefore reducing the costs incurred due to damage caused by handling and transportation of specimens. Finally, digitisation enables multiple users to access the same specimen simultaneously thus opening the collection to an even wider audience.

To achieve the Museum's digital collections ambitions and unlock these potential benefits, a step change in the current speed of digitisation is required. One key part to achieving this was the announcement in March 2020 awarding the Museum £180 million from the UK government to build a new Science and Digitisation Centre at Harwell in Oxfordshire to open in 2026. This grant will enable the Museum to build the

physical infrastructure to support its digitisation programme and future partnerships, forming a key component of its transition to a 21st century digital museum and further cementing its place at the heart of the international ecosystem of scientific collections. While the grant by the UK Government is an essential enabler for the digitisation programme to progress, the funding will not finance operations once the physical infrastructure is in place. Significant additional investment going forward will be needed in order to support the Museum's digital operations, which will ultimately become a key part of the UK's research and data infrastructure.

In this study, we consider what the monetary benefits of digitising the Museum collection are for the UK. We do this in several steps. First, we develop a theory of change which sets out the theoretical channels through which digitisation can lead to economic impacts. We do this at a high level but also in specific areas of research which are hypothesised to benefit most from digitisation. Second, we evidence the theory of change with existing evidence from the academic and grey literature and draw all the findings together in a return on investment (ROI) model which provides estimates of economic return under different scenarios.

The remainder of this report is organised as follows:

In Section 3 we outline the approach to the study;

Section 4 includes our headline findings; and

Section 5 provides some concluding remarks and suggests some further areas for research.

3. APPROACH

Identifying the benefits of digitisation

We carried out an initial review of information shared by colleagues at the Museum followed by an evidence review looking at grey and academic literature on the benefits of open natural history collections data. These helped us to develop a theory of change/logic model which sets out the different pathways to impact/benefit applicable to investments in digitisation, how these are expected to materialise (and over what time period), how significant they are likely to be and who the ultimate beneficiaries are (e.g. visitors, scientists, taxpayer, society at large).

The theory of change shows (Fig. 5) the categories of inputs that go into digitisation activities along with the outputs, outcomes and ultimately the impacts that flow to the wider economy through the contribution of these activities to research. We define the components of the logic model as follows:

Figure 5. Components of the theory of change.

- **Inputs** are the resources that are used to deliver the digitisation of the museum's collection for research activities. We expect inputs to include financial, physical and human resources.
- **Activities** are the actions taken, using the resources described in the inputs. We have kept the activities open to recognise that there can be different levels of digitisation activities feasible based on inputs.
- **Outputs** are the direct products of the activities i.e. the data generated as a result of the digitisation process. We have listed a few examples of what the outputs can look like but our list is not exhaustive.
- **Outcomes** are the short run effects of the activities. They can reach beyond immediate users of the outputs and can include changes in efficiencies, behaviours and new activities in adjacent sectors.
- **Impacts** describe the long run lasting effects of the outputs. These describe changes to the different areas that the research activities are intended to contribute to.

As well as understanding pathways to impact, the evidence review we carried out also allowed us to identify methodologies for valuing the impacts of digitisation and data which was useful for producing monetizable estimates of impact.

We reviewed the latest academic literature which looks at the benefits of open data and what increases in scientific research it leads to. We drew on a range of academic databases including Google Scholar, Econlib, JSTOR. We also examined grey literature including publications from governmental and non-governmental bodies as well as Frontier's own previous work looking at the impact of scientific research (Frontier Economics 2014) or work by other agencies looking at the value of institutions which have undergone similar transformation initiatives such as the European Bioinformatics Institute (Beagrie and Houghton 2016).

As well as the evidence review, we consulted with experts at the Museum who provided input into the theory of change development in different areas within Life and Earth sciences. Feedback was gathered during a virtual workshop as well as follow up interviews with relevant experts. In total, we consulted with more than 20 experts at the Museum covering areas such as biodiversity conservation, climate change, mineral exploration and others.

Valuing the benefits of digitisation

Once the key pathways to impact were identified, we assembled all available evidence and data in a model which was used to produce estimates of the value of digitisation.

Estimates of value were produced in two ways. A top-down approach was used to estimate benefits at a high-level drawing on the literature looking at the typical returns to investment in scientific research. These were applied to the expected level of investment required to digitise the entire Museum collection[*6] and/or the research efficiencies that are expected to arise as a result of digitisation[*7]. These are summarised in Fig. 4.

The high-level impacts were underpinned by examination of specific areas in which digitisation of the museum's collections is expected to boost research, ultimately leading to benefits for the UK society and the world. These thematic areas helped to provide a more detailed and specific view into how the benefits from digitisation can be expected to materialise in practice. The cumulative estimated value across these areas provided further support to the value from digitisation estimated at a high level above. We looked at five specific areas: biodiversity conservation, medicines discovery, invasive species, agricultural R&D and mineral exploration. The end-to-end approach is summarised in (Fig. 6).

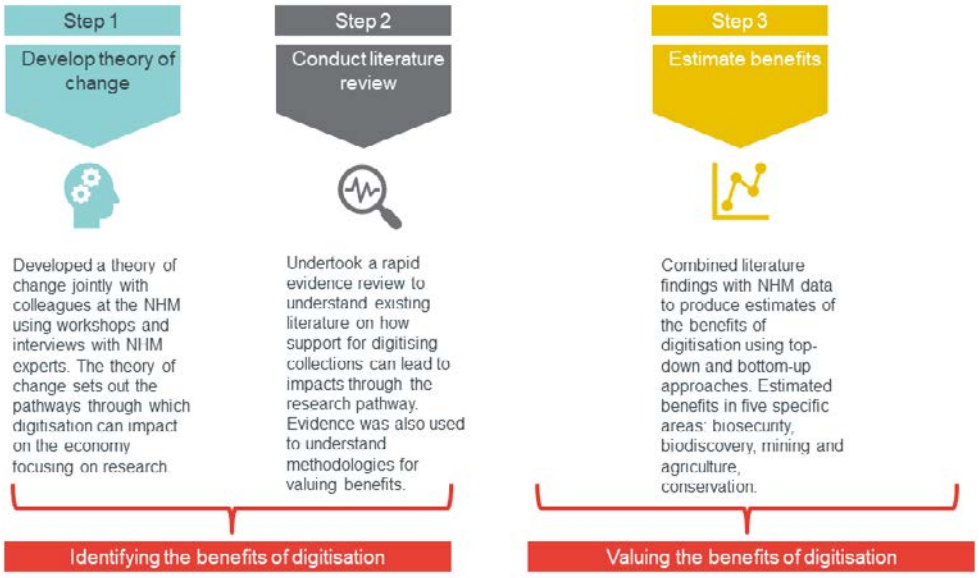

Figure 6. Summary of the approach, using a theory of change to guide a literature review and benefits modelling.

4. FINDINGS

Theory of change

Digitisation of the Natural History Museum's extensive collections can have an impact on the economy through multiple pathways that reflect the vast array of activities the museum is involved in – collecting, preserving, researching, educating, exhibiting. Digitisation can support and potentially enhance the museum's work across all these areas, enriching the experience of the users who interact with the collections. The research community were identified through our workshop as a user groups who could benefit significantly from the digitisation of the collection.

EXAMPLE OF A RESEARCH PROJECT FACILITATED BY THE NATURAL HISTORY MUSEUM

Sequencing the wheat genome (Pavid 2020)

Wheat contributes about a fifth of the total calories of human consumption every year and is one of the most valuable food crops in the world alongside rice and corn. Since humans started farming, humans have bred wheat plants to improve the quality of crops over time. This is known as selective breeding, leading to higher yields and greater resistance to diseases.

But selective breeding also reduces genetic diversity, making wheat crops more similar over time and therefore less resilient to disease and less adaptable at population level. This is a threat to global food supplies, and can also lead to increased use of pesticides and other measures that directly impact the ecosystem, and artificial fertilisers that negatively impact the health of the soil by changing the balance of nutrients in it. These issues have become particularly pertinent in the wider context of climate change and food security for a growing population.

In 2020, researchers published sequenced genomes for 16 wheat species, including samples from the Natural History Museum's collection. This will allow breeders to better select traits to improve yield and harness genetic immunity to pathogens, reducing the need for pesticides which can have devastating effects on wider ecosystems and environments.

Research has increasingly started to integrate digitised data, which have transformed the production of knowledge by making the process of planning, conducting, disseminating, and assessing research activities more efficient. At the same time, *open data* have encouraged the sharing and interlinking of heterogenous research data sets, providing

efficient ways to model data for new and cross-cutting insights. Researchers are now better able to use data resources from diverse sources which improves the accuracy of scientific findings and helps to identify future directions for research.

The Natural History Museum's digitisation programme aims to contribute to the open data movement by giving the global research community free and open access to the wide array of data contained in its collection of 80 million items.

Logic model

At a high level, the theory of change posits that digitisation enables more (i.e. accelerates existing research and enables brand new research) and potentially higher quality research, which ultimately leads to a range of societal benefits in a number of areas from agriculture to health. The theory of change also describes the chain of events that are expected to bring about the ultimate impacts (see Fig. 7).

The wider and greater accessibility of data is expected to save time and money for the research community, who otherwise would have to rely on physical visits and/or spend

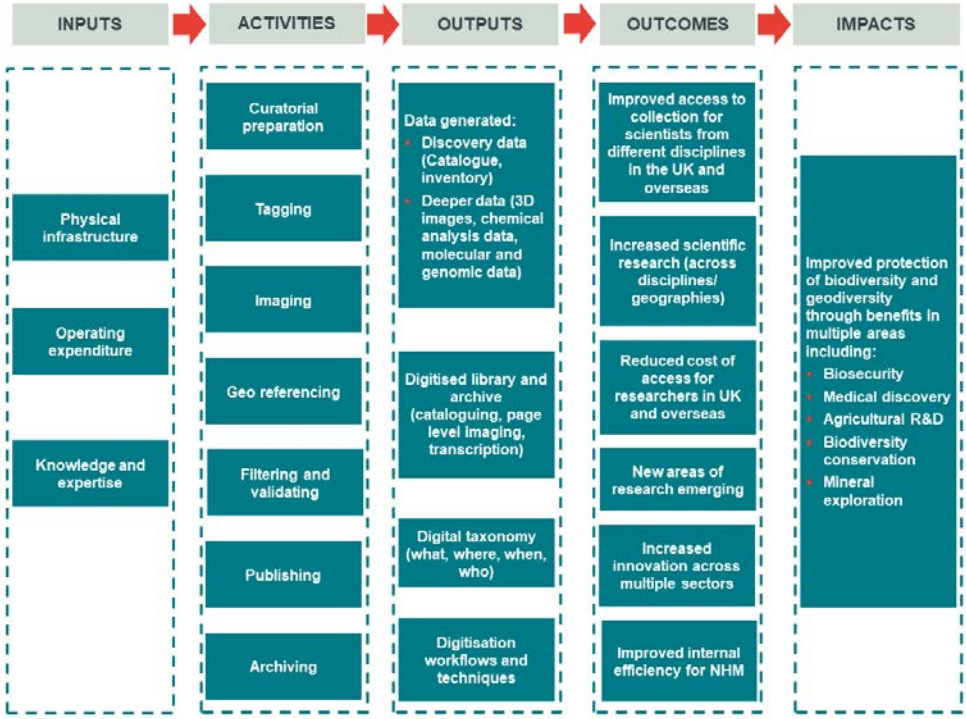

Figure 7. Theory of change showing the four components (inputs, activities, outputs and outcomes) with examples that lead to impact.

more time searching, compiling and creating data relevant for their respective research activity.

The availability of digitised data leads to more and better research which subsequently can have impact in different areas. We have looked at five such thematic areas:

- **Biodiversity conservation**: digitisation can enhance taxonomic knowledge which can improve the detection of threatened species. This in turn enables conservation efforts to be put in place which reduce the rate at which species decline, maintaining the ability of ecosystems to provide vital services for humanity.
- **Medicines discovery**: digitisation can improve accessibility of samples and consequently the range of samples tested for drug discovery and commercialisation. The economic value of commercialised drugs for human health is huge. So even if digitisation leads to a very small increase in the rate of drug discovery, the benefits are large.
- **Invasive species**: digitisation allows for better and faster detection of invasive species which have significant costs for the UK economy estimated at £2 billion a year. Reducing the frequency of losses from threats can thus lead to substantial economic benefits.
- **Agricultural R&D**: digitisation can help in the discovery and/or improve the understanding of Crop Wild Relatives (CWR)[*8] with regards to their genetic traits. This can enhance breeding of crops which are environmentally friendly, have higher yields and are disease resistance.
- **Mineral exploration**: digitisation can improve the accuracy of existing geological data and help develop enhanced geological data from the museum's rich collections of minerals, ores and rocks. This could accelerate the geological discovery process and reduce the costs of exploration significantly by de-risking the process.

The value of digitising the museum's collections

Digitisation of a collection is the process of creating digital representations of physical specimens. There are different levels to which a collection can be digitised depending on the medium, purpose, availability of equipment and budget among other things. Creating a digital copy is one part of the process; digital records need to be annotated and categorised to make them useful to the audience.

Digitisation of collections provide benefits in different ways based on the intended purpose of the digitisation process for an organisation, some of which include:

- **Accessibility**: Collections held in a physical collection can be made available to a global audience (which includes researchers and the general public among others) so that access isn't restricted to those who are able to physically access the collection, saving travel costs and time.
- **Searchability**: The data associated with a digital copy of the collection through the process of digitisation enable users to find relevant content more easily and accurately which helps to increase research efficiency, collaboration between researchers, and integration with other collections.
- **Preservation**: Using a digital copy can mean that users don't need to interact with the physical objects as much as they would without digitisation, therefore reducing the costs incurred due to damage from handling and transportation of specimens. Additionally, where digital discovery increases demand for physical access or where only physical access can meet a research need, this access can be more efficiently and effectively targeted when more is shared digitally in advance.
- **Interaction**: As a digital copy isn't restricted by the need for physical space for an audience to interact with it, digitisation helps to open up larger parts of the collections than previously possible to a wider audience.

As already noted above, one of the direct pathways through which digitisation can create value in the economy is through research activities. Wider accessibility to and searchability of the data held within specimen collections enable researchers to extract valuable information and/or enhance existing datasets for conducting research activities. This has the potential to lead to innovation in the economy across many important areas such as biodiversity conservation, climate change, human health etc.

CASE STUDY: BENEFITS OF DIGITISING COLLECTIONS

How digitising bat specimens in museum collections is helping research into coronaviruses

Around 75% of all emerging infectious diseases are those where viruses are transmitted from animals to humans (Jacklin 2021). Bat populations have been identified as one of the high-risk viral reservoirs i.e. they are populations which host viruses naturally. Increased human activity such as deforestation and intensive farming have brought humans closer to wildlife, creating the conditions that mean diseases can transfer to humans.

The recent global pandemic has generated renewed interest in bats as research shows that genome sequences of the COVID virus found in humans are 96%

identical to that of a bat coronavirus (Natural History Museum 2021d). Scientists are studying this and related bat families globally to unlock the mystery of the origins of the COVID-19 pandemic, answers to which could prove to be invaluable to mitigating the risks from future viral spillover events. Museum collections of natural specimens provide crucial data for such research activities.

The study of specimens held in natural history collections is a huge resource that can help to build a knowledge base for research to understand the origins of zoonotic diseases[*10] such as COVID-19. For example, the Natural History Museum alone has around 40,000 bat specimens, and are digitising around 10,000 from three key families. Digitising these collections enables researchers to easily access and collate information associated with the specimen such as when and where it was found, building a picture of distribution over time.Greater understanding of species distribution and ecosystem change from specimen data can help researchers to link events such as the outbreak of the COVID-19 pandemic to specific environmental conditions. This can help to predict when a future viral spillover might occur.

Valuing the impact of digitisation on research activities

Research is increasingly becoming data intensive, making the accessibility of usable data a key enabler for scientific projects and innovation. Previous literature has looked into the benefits of digital data for science using various approaches such as rapid evidence assessments, user-based surveys and cost benefit analyses. While some have looked at the benefits of open data for scientific purposes (Fell 2019), others have attempted to estimate the value created from specific scientific databases for its specific user base (e.g. Beagrie and Houghton 2016). These studies provide a helpful foundation to understand the ways one can look into valuing the benefits of digitising the collections at the Natural History Museum.

The economic value of research data and the related contribution to economic development has also been studied. For example, Houghton and Sheehan (2009) looked at the effects of increasing accessibility to public sector research outputs in Australia, and estimate that increased accessibility generates a return of AUD 9 billion over 20 years (equivalent to £4.9 billion). Houghton et al. (2010) estimate that the open access archiving mandate for US Federal Research Agencies over a transitional period of 30 years may be worth around USD 1.6 billion–USD 1.75 billion (equivalent to £0.9 billion–£1 billion), significantly higher benefits than the estimated cost of implementing open access to the archives. Jisc commissioned a study on the economic impact

of three UK data centres*[16] and estimated that each of them could bring a twofold to tenfold return on investment over 30 years (Beagrie and Houghton 2014). A recent study by Beagrie and Houghton (2016) looking at the impact of the European Bioinformatics Institute (which makes life science data freely available to the global research community) estimated future research impacts worth £6.9 billion.

The literature looking at the return on investment in science gives us a reasonable proxy for the value of the digitisation programme. A comprehensive literature review by Frontier Economics found that the typical private rates of return are in the region of 2025% with social returns being two to three times higher. In the area of medical research in particular, a study estimated the economic returns, both in terms of health gains and GDP gains, to be 25% on public investment in UK medical research (Health Economics Research Group et al. 2008). Along with estimating the returns on scientific research, there have also been previous attempts at identifying the benefits of widening the knowledge base related to the natural world. An important study in this respect is the cost-benefit analysis of the Taxonomy Australia Mission (Deloitte Access Economics 2020) which estimated that every one Australian Dollar spent on improving the discovery and documenting of species in Australia could result in benefits to Australia alone ranging from four to 35 Australian Dollars.

The value for the UK of digitising the entire Natural History Museum collection is likely to be very large

There is significant existing literature which makes the case that scientific research has been crucial in the progress of economies globally. Open access to data has helped the research community to accelerate this progress in the past couple of decades. But to articulate an economic case for investing in digitisation — one that involves collections as geographically extensive and historically significant as the Museum — we need to show that the economic return from digitisation is worth the investment.

There is no perfect method for estimating a return on investment in this context and the inherent uncertainties around the scientific breakthroughs that may be enabled by digitisation make it difficult to produce precise estimates. We have looked at this issue both from a top down and a thematic perspective. The top down entails estimating at the aggregate level the expected returns an investment in science is likely to generate. The thematic approach is about valuing specific benefits in a particular research area which might be expected to arise with the help of digitisation.

One approach to value the impact of digitisation is to look at the returns expected to be generated as a result of investment in the programme to digitise the museum's entire collection. This can provide an indication of how important digitisation is for

Table 1. Impact on R&D due to digitisation.

	Low	Medium	High
Investment	£200m	£200m	£200m
Discount rate*9	3.5%	3.5%	3.5%
Return	20%	30%	40%
Present value of additional research over 30 years (NPV)	**£735m**	**£1.1bn**	**£1.5bn**

increasing research activities which have been shown to be an important factor in economic growth (Aghion and Howitt 1992). As shown in Table 1, assuming that an initial investment of £200 million is made into the digitisation programme, expected benefits range between £730 million and £1.5 billion in present value terms over 30 years.

It is important to note that the level of investment is set at £200 million for the purposes of this estimation. This is deliberately ambitious to reflect the variety of both collection object types and levels of digitisation that might be desirable for research, which it has not been possible to differentiate in this initial research. While the Natural History Museum's collections could be digitised to discovery level (basic data on e.g. what/where/when) for a smaller investment, £200 million could for example enable either digitisation of the majority of UK natural science collections (incuding and beyond the Natural History Museum) to discovery level, or of the Natural History Museum collections to an enhanced level, potentially including enhanced imaging (microscope imaging, CT and 3D scanning for instance) and enhanced analyses in some cases e.g., genetic species barcoding or chemical data for mineral collections.

Another approach for valuing the benefits of digitisation is to look at the value created for the direct users of the collection for research. We can quantify the value of digitisation to research activities by estimating the efficiency benefits that digital data brings to researchers in terms of cost and time savings. In particular, this would involve:

- estimating the cost savings in terms of researchers not having to travel to a physical location and search for specific specimens which can then be used for new research activities; and
- estimating the amount of new research that is made possible by the availability of digital data.

Exact cost savings will differ between researchers based on the extent to which the digital database is sufficient for their purposes. However, we expect researchers to save at least a proportion of their costs (travel as well as time) if they are able to access the data digitally.

Table 2. Value of digitisation from increased research.

	Low	Medium	High
Fraction of costs saved by researchers due to using digitised data relative to a physical visit	5%	10%	12.5%
Value of extra time available to researchers if entire collection is digitised	£120m	£230m	£290m
Average returns to science R&D	20%	30%	40%
Annual impact of additional research	£20m	£70m	£120m
Discount rate	3.5%	3.5%	3.5%
Present value of additional research over 30 years (NPV)	£0.4bn	£1.3bn	£2.1bn

Noting the data limitations that prevent making precise estimations, we use available data to identify an approximate indicator for new research using the number of downloads of specimen records data from the available digitised database (which accounts for only 6% of the entire collection) assuming that each download event corresponds to a unique research project. We then make an assumption that 75% of these projects would not have been undertaken if the data was unavailable digitally. We look at different scenario cases which extrapolate the above estimation of new research from digital data to the entire collection and hypothesise different levels of cost savings to researchers and returns to R&D. Based on this, we find that the benefits from the digitisation of collections could range between £0.4 billion and £2.1 billion in present value terms over 30 years (see Table 2 for details).

In the following sections we present thematic estimates of value associated with five distinct areas where digitisation can have an impact.

Invasive species

Invasive non-native species (INNS) or invasive species are those that have moved outside of their natural range due to human activity and have negative effects on native wildlife and ecosystems. They are the second biggest threat to global biodiversity after habitat loss (RSPB 2021). In the UK, invasive species have been identified as one of the top five threats to the natural environment, which is growing with the expansion of international trade, transport and travel (House of Commons Environmental Audit Committee 2019). These threats have been recognised in the UK's biosecurity regime and the significance of their economic costs is reflected in major international agreements (e.g., Bern Convention - Council of Europe (1979), the Convention on Biological Diversity - United Nations (1993)) ratified by the UK, which deal with the prevention of invasive species being introduced and established.

EXAMPLES OF INVASIVE SPECIES: THE ASIAN HORNET AND FLOATING PENNYWORT

Asian Hornet

The Asian Hornet is an INNS which got imported accidently into France in 2004. From there, it made its way to different parts of Europe before arriving in the UK sometime around 2016 (Osterloff 2018). They are an invasive predator of medium to large insects and pose a serious threat to native pollinators such as honeybees and hover flies. Additionally, they can also have adverse effects on human health as they carry a venomous sting causing severe reaction and in some cases fatalities.

Floating Pennywort

The Floating pennywort, native to the Americas, was introduced to the UK in the 1980s (CABI 2021). They form dense mats of leaves which float on the surface of water, leading to depletion of oxygen levels and of light for photosynthesis. This poses a threat to aquatic life and competes with native water plants. Additionally, they cause economic losses by impacting commercial fisheries, damaging waterworks etc.

Biosecurity refers to a set of precautions aimed at preventing the introduction and spread of harmful organisms. There can be several pathways through which these species can be introduced into an area. For example, international movement is a primary source of invasive species as 'stowaways' or contaminants occur through cargo and passenger travel. Recently, climate change has also been recognised as a catalyst for introducing invasive species (Ashworth 2021). Biosecurity involves detecting such species during routine surveillance on common routes and in incidents with as much certainty as possible, in order to form an appropriate biosecurity response.

Digitisation of specimen collections can provide easier access to researchers to form a more comprehensive and updated database to identify species. In particular, faster and easier access to specimen databases can help to inform biosecurity checks and responses, and identify threats from invasive species correctly.

HOW CAN DIGITISATION OF SPECIMEN COLLECTIONS HELP?

Overview

During biosecurity surveillance, diagnosis of species involves identifying and classifying them with as much certainty as possible, since a misdiagnosis can have significant costs. A better understanding of species and of inter-species interactions can help with developing the most minimally interventionist and/or sustainable

responses. This requires a good understanding of species by biosecurity diagnosticians which hinges upon their ability to access information on invasive species (and on species that might be mistaken for them). Digitisation of collections help to provide faster access to specimens and other research datasets, facilitating faster detection and diagnostic certainty.

Illustrative example of the use of digitised specimens

The Natural History Museum has digitised more than 6,700 specimens of bumblebees with the oldest specimens dating back to the nineteenth century (Pullar 2018). Records of species for an extensive timeline such as this, which includes information such as location and time of sightings, provide unparalleled information for various research topics. These data can be used for biosecurity responses – identifying invasive species such as the Asian hornet sightings correctly or avoiding misdiagnosing another species as a threat.

Valuing the impact of digitisation on biosecurity

One of the ways to value the benefits of digitising the museum's specimen collections would be to understand the costs that are avoided from invasive species by a successful biosecurity response. Studies have aimed to estimate the annual costs to the UK from invasive species and these range at the lower end in the millions (Cuthbert et al. 2021) and at the higher end in the billions (Wildlife and Countryside Link 2020b). Biosecurity threats can have significant negative impacts on the use values of ecosystem services such as agriculture, forestry, tourism and recreation as well as economic costs such as to transportation services and human health. These costs can be avoided by addressing late detection biosecurity threats, and/or misdiagnoses of non-genuine threats. Ideally, the estimation of these avoided costs would involve estimating:

- the reduction in time by avoiding delay/uncertainty in detecting threats; and
- the reduction in damages due to the faster detection of threats or prevention of misdiagnoses, with greater certainty.

However, quantifying each of the above is complicated and requires data which is not easily available. An alternative approach is to estimate the impact by looking at the costs saved due to a reduction in the frequency of the occurrence of losses through biosecurity threats caused by invasive species. For the purposes of our estimation, we take the annual economic losses to the UK in a business-as-usual (BAU) environment i.e. where there is no digitisation of species collections, to be £2.2 billion annually (Wildlife and Countryside Link 2020b). We assume that digitisation of the Museum

collection can affect only a relatively small proportion of these costs (for the purposes of this estimation we assume that only 5% of annual losses can be affected by digitisation). We look at three different scenarios of reduced frequency of losses incurred – low impact (once in 2 years), medium impact (once in 3 years) and high impact (once in 4 years). We then estimate the net benefits by comparing these scenarios with the BAU case. **These benefits range between £0.7 billion and £1.1 billion in present value over 30 years** (see Table 3 for details).

It is important to note here that these benefits are a conservative estimate as they take into account the impact to the UK only. The overall benefits are expected to be greater

Table 3. Detailed value calculation of the potential benefits of digitisation for reducing economic losses from the threat of invasive species.

Variable	Description	Value	Calculation
a	Economic losses to the UK due to invasive species annually (based on value inWildlife and Countryside Link (2020a))	£2,000,000,000	n/a
b	Proportion of losses that digitisation of collections can help reduce	5%	n/a
c	Effective economic losses digitisation can help reduce	£100,000,000	a * b
d	Discount rate	3.50%	n/a
e	Value of losses to invasive species which digitisation can help reduce under BAU (NPV over 30 years)	£1,839,204,541	NPV(c) over 30 years @3.5%
Scenarios			
f	Value of losses to invasive species which digitisation can help reduce under Low Scenario (frequency of losses reduced to every 2 years)	£1,151,741,090	NPV (f) over 30 years @3.5%
g	Value of losses to invasive species which digitisation can help reduce under Medium Scenario (frequency of losses reduced to every 3 years)	£831,660,532	NPV (g) over 30 years @3.5%
h	Value of losses to invasive species which digitisation can help reduce under High Scenario (frequency of losses reduced to every 4 years)	£687,395,554	NPV (h) over 30 years @3.5%
Net benefits (PV over 30 years)			
i	Value of digitisation (Low Scenario)	£687,463,451	e - f
j	Value of digitisation (Medium Scenario)	£1,007,544,009	e - g
k	Value of digitisation (High Scenario)	£1,151,808,987	e - h

as the museum's collections include specimens from across the world and will benefit geographies outside the UK as well.

Benefits of digitisation to medicine discovery

Biodiscovery is the exploration of biological material for commercially valuable genetic and biochemical properties in various fields such as agriculture, medical and pharmaceutical products, and cosmetics. It encompasses a stage in the medical R&D value chain that involves locating potentially valuable bioactive compounds in nature, collecting samples of native biological materials (such as plants, marine sponges, and microorganisms) and testing for chemical compounds.

Biodiscovery is an important source of leads for new medicines and is widely recognised as the most successful class of drug leads in the process of drug discovery and development (Harvey 2008). Around 35% of all drugs today are derived from natural compounds (Calixto 2019). In particular, in the area of cancer, over the timeframe 1940–2014, 49% of all new approved drugs by the FDA were either natural products or directly derived therefrom (Newman and Cragg 2016). The discovery and development of these drugs involves the collection of species of interest, and the study of them to identify compounds which have the potential to be a new drug. Samples of these species are maintained for further research as repeated collections are usually resource intensive.

ILLUSTRATIVE EXAMPLE: ANTI-CANCER DRUGS

A species of sponge — *Tecitethya cypta* (de Laubenfels, 1949) — produces two chemicals which were used to develop the first anti-leukaemia drug and were also part of the breakthrough drug first administered in the 1980s to people living with HIV (Osterloff 2017). Blue green algae, or Cyanobacteria are a type of photosynthetic bacteria which produces a chemical compound shown to have potential as a novel anti-cancer drug with future research (Leclerc 2021). Finding the new compounds involves testing samples and usually these are collected and screened after which voucher samples of these species are sent to be stored for further examination and study. Botanical voucher samples, like that of the blue green algae, document the source material of a drug discovery and help to recreate the chemicals in the lab. They help to ensure that new agents can be found and properly identified over and over again.

Generally, biodiscovery for medical purposes is a multi-staged process which involves biotechnology researchers in the public or private sector working collaboratively with

pharmaceutical companies. Value from the biodiscovery value chain is generated when the compounds from species are successfully commercialised and can be estimated by looking at two aspects:

- **Research value**: In most cases, the first step in the biodiscovery process is the collection of samples and identifying the ones which have the potential to be explored for bioactive compounds which can be used pharmacologically further down the value chain.
- **Development value**: This stage of the biodiscovery value chain represents the commercialised value of the species sample. This includes both the market value and health benefits realised through the commercialisation of the drugs. There are two important components of this process – 1) selecting samples of pharmaceutical interest and 2) successfully commercialising the drugs made using the natural components of the samples to generate health benefits.

Digitisation of specimen collections can accelerate the discovery and targeting process of specimens with useful bioactive compounds that might have health benefits by improving the speed and accuracy of identification. This then improves the accessibility of the samples and consequently the range of samples tested for the purposes of drug discovery and commercialisation. A direct impact of digitally accessible specimens will be with regards to the research value generated from the samples, with increased number of species being explored for bioactive compounds. This can then be expected to have a knock-on effect on the number of successfully commercialised samples which will generate increased health benefits to society.

Valuing the impact of digitisation on biodiscovery for health

To value the impact of digitisation of specimens on biodiscovery for health we follow a framework developed in by Deloitte Access Economics (2020) where we need to:

- Estimate the impact of digitisation on the number of samples available for testing; and
- Estimate the health benefits due to the increased number of commercialised samples.

In order to estimate the increased value brought about by digitisation of specimen collections, we need to identify the rate at which digitisation of specimen collections increases the number of samples available for testing. We do not have data on the number of samples tested annually in the UK but Deloitte Access Economics (2020) estimated this to be ca. 20,000 for Australia. Given that business expenditure on R&D in the pharmaceutical sector is much larger in the UK (according to the OECD (2015a) the relevant

figures are 0.3% of GDP for the UK and 0.03% for Australia), we would expect the number of samples tested in the UK to be significantly higher, potentially 10–20 times higher if this is proportionate to levels of R&D expenditure. For the purposes of this benefits valuation, we have assumed that as a result of digitisation an additional 1,000 can be tested annually[*11] with a higher than standard probability of commercialisation.

Estimating the health benefits due to increased research is complex. One way would be to look at the benefits realised through the commercialisation of natural product-based drugs. Firstly, this would involve the selection of samples which are pharmacologically bioactive through clinical testing and development. Secondly, these samples must be successfully commercialised to generate health benefits to patients.[*12]

In the business as usual case, we estimate an overall probability of any species derived sample being successfully commercialised in the market to be as low as 0.0001% which is a conservative lower bound (Deloitte Access Economics 2020). We then assume a hypothetical scenario where this probability increases to 0.0002% (low scenario) and 0.0003% (high scenario) due to an increase in the number of bioactive samples of interest selected given the overall increased number of samples tested.

To derive the value of health benefits arising from successfully commercialising drugs derived from natural products in the future, we apply this conservative lower

Table 4. Detailed Value calculation for health benefits to the UK. Assumptions are based on (b) and (c) which are adapted from Deloitte Access Economics (2020) based on estimates from Beattie et al. (2005).

Variable	Health benefits to the UK	Low	High	Formula
a	New samples tested annually due to digitisation	1,000	1,000	n/a
b	Health benefits per successful commercialisation	£52,743,776,515	£104,099,558,910	n/a
c	Probability of successful commercialisation of a species derived sample (above baseline)	0.0001%	0.0002%	n/a
d	Annual health benefits from increased samples and higher probability of successful commercialisation due to digitisation	£52,743,777	£197,789,162	a * b * c
e	Discount rate	3.50%	3.50%	n/a
f	**PV of health benefits attributable to digitisation**	**£749,615,823**	**£2,811,059,338**	NPV over 30 years @3.5%

Table 5. High level estimates of value based on identification of threatened species and conservation efforts. Benefits of ecosystem services based on Wilson (2008).

	Low bound	High bound
Annual benefit to UK citizens of ecosystem services	£1,600,000,000	£3,000,000,000
Discount rate	3.50%	3.50%
Proportion of value affected by digitisation	1.00%	2.00%
Value affected by digitisation	£16,000,000	£60,000,000
Total discounted value attributable to digitisation	£**294,272,727**	£**1,103,522,725**

bound probability to measures of economic benefits of improved health due to these drugs. **Using existing literature, we estimate the net benefits attributable to digitisation to be in the range of £0.8 billion to £2.8 billion in present value over 30 years**[*13] (see Table 4 for details).

Benefits of digitisation to agricultural research

Agricultural Research and Development (R&D) is a crucial component in making improvements to agricultural production systems which contribute towards shaping the future of global food production. Scientists and agricultural industries are continually undertaking research in plant and animal sciences to discover procedures that will increase livestock and crop yields, improve farmland productivity, reduce loss due to disease and insects, develop more efficient equipment, and increase overall food quality. In general, agricultural R&D takes places at the intersection of several primary disciplines such as microbiology, ecology, botany and zoology, as well as secondary enabling sciences such as chemistry and climate science. As part of this process, plant and animal specimens form a foundational element in the research pathways, and researchers increasingly need to draw on the expansive knowledge embodied in natural specimen collections. An increased understanding of plant, animal, fungal and microbial species will aid ongoing research activities through several pathways, some of which include:

- Discovery of species and genes that contribute to the transition to bio-manufactured foods;
- Greater understanding of micro-organisms enhancing yield productivity;
- Greater understanding of pests and pathogens;
- Discovery of new and more effective biological control agents; and
- Discovery and/or better understanding of native crop plants and animals with genetic traits that enhance breeding e.g. disease resistance, yield productivity.

Digitisation of natural specimen collections is expected to accelerate the rate at which researchers are able to discover and improve their understanding of different natural species for the purpose of agricultural R&D. Databases composed of digitised specimen data enable faster and easier access to crucial information (which, if not digitised, researchers may not be aware exist or at the very least may find difficult to access) that can speed up the research process. This can subsequently translate into substantial economic and social value as demonstrated by economic impact assessments of benefits of agricultural R&D.

Valuing the impact of digitisation on agricultural R&D

To value the impact of digitisation on conservation we need to:

- Identify the rate at which digitisation increases the discovery and/or understanding of natural species for the purposes of agricultural R&D; and
- Identify how this increased research creates economic value.

Given the complexity in the different pathways through which knowledge from natural specimens can contribute towards agricultural R&D activities, for the purposes of this assessment we focus on the benefits of digitisation on the discovery and/or understanding of Crop Wild Relatives.

CROP WILD RELATIVES (CWR)

Overview

Crop wild relatives are wild plant species that are genetically related to cultivated crops. They continue to evolve in the wild with a greater genetic diversity than their cultivated cousins, developing traits such as drought tolerance or pest resistance without active human interference to the process. These can be crossed with existing domesticated crop species to produce new varieties and promote breeding.

Benefits of Crop Wild Relatives

Ensuring that agriculture is able to adapt to the demands of the planet's growing population and extreme weather events brought about by climate change is an urgent challenge of our times. Today, approximately 80 per cent of the human calorie intake comes from just 12 plant species (Wilding and Cockel 2019). There is an urgent need to ensure that crops, upon which global food security rests, are resilient in the face of these changes as well as show yield improvements. The diversity in CWR means that they are a rich source of adaptive characteristics such

as heat or drought tolerance, disease resilience or the ability to thrive in saline soils. These important traits can be bred back into domesticated crops to make our crops resistance to the impacts of climate change and see an uplift in their yield.

How can digitisation of specimen collections help?

There are estimated to be between 50,000 to 60,000 species of CWR in the world (Wilding and Cockel 2019). A recent study shows that 35% of selected CWR taxa face the threat of extinction and there is an urgent need to increase knowledge and research around different aspects of CWR (Goettsch et al. 2021). The collection, conservation and preparation of CWR to tap into their genetic diversity requires better understanding of these species such as their morphological and /or genetic traits. Collections of plant specimens are a rich source of such information – for example – identifying the closeness of the relationship between a crop and its CWR as defined in terms of the Gene Pool concept (Fielder et al. 2015). Digitisation of plant specimen collections is expected to accelerate research into CWRs by enabling phylogenetic analyses at a scale which would be difficult to undertake if the data had to be collected from scratch. It can also facilitate faster and easier access of CWR specimens for further research by providing data for genomic analyses.

Previous research commissioned by the Millennium Seed Bank sought to understand the value of CWR. Based on an analysis to understand gaps in collections of priority CWR, 29 crops were identified. The study went on to estimate the value of the economic benefits from future improvements for these 29 crops from the use of CWR material in breeding programs based on gross production value*[14] to be USD120 billion in perpetuity (PwC 2013). Using this as the basis to evaluate the benefits from CWR, we can estimate that for any one crop, improvements due to cross with CWR could generate a lower bound of annual economic benefits of £30 million. It is important to note that these benefits are global but can be expected to have significant value to the UK due to the interconnected global food supply chain that we see today.Digitisation of the plant specimen collections held by Natural History Museum could see improvements in pre-breeding research worldwide which will then have knock on effects across the rest of the global production value chain through an acceleration of the realisation of economic benefits. We construct three scenarios – low, medium, high – wherein we hypothesise that digitisation accelerates the commercial lead time by one, three and five years respectively.Hypothetically, we assume that under a business-as-usual (BAU) scenario, a crop currently under improvement through agricultural R&D into its wild relative has a 10-year commercialisation lead time. As seen in Table 6, the present value

Table 6. Economic value from research into Crop Wild Relatives.

Scenarios	Discount rate = 3.5
Business As Usual (10 years commercialisation lead time)	£425m
Low (9 years commercialisation lead time)	£440m
Medium (7 years commercialisation lead time)	£467m
High (5 years commercialisation lead time)	£493m
Value of digitisation (difference between scenarios and BAU)	
Low (1 year less lead time than BAU)	£15m
Medium (3 years less lead time than BAU)	£42m
High (5 years less lead time than BAU)	£68m

of economic benefits in this case would be £425 million using a 3.5% discount rate. Considering the impact of digitisation, the earliest that benefits could start occurring under the high change scenario case is five years earlier. Under this scenario, the present value of benefits would be £493 million for 30 years. Under the medium case, in which commercialisation lead time is reduced by three years, the present value of benefits would be lower at £467 million, and under the low case, in which commercialisation lead time is reduced by only one year, the present value of benefits would be at £440 million. In this case, a lower bound of the impact of digitisation can be seen through the net benefits which are given by the difference in scenarios compared to the business as usual case. These are different scenarios are summarised in Table 6. For example, in the medium scenario, the present value of net benefits from digitisation are £42 million. However, it is important to keep in mind that, as demonstrated by this exercise, these estimates are sensitive to the assumptions and value inputs used to estimate the benefits. They are a function of the extent to which digitisation can accelerate the actual realisation of benefits from crops improved by CWRs.

Benefits of digitisation to biodiversity conservation

Biodiversity is essential for the functioning of ecosystems, underpinning the provision of ecosystem services that ultimately affect human well-being. Ecosystem services can be defined as the benefits people obtain from ecosystems and can be divided into several categories:

- Provisioning services such as food, water, timber and fibre;
- Regulating services such as the regulation of climate, floods, disease, waste and water quality;

- Cultural services such as recreation; aesthetic enjoyment and spiritual fulfilment; and
- Supporting services such as soil formation, photosynthesis and nutrient cycling.

Biodiversity is in decline and there is evidence that the rate of decline may be accelerating. According to the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES 2019) "around 1 million animal and plant species are now threatened with extinction, many within decades, more than ever before in human history." (United Nations 2019).

In order to enable conservation efforts to slow or reverse nature's decline, one needs to understand which species are endangered. This requires one to assess changes in populations of species over time in different geographies. This can only be done if the data are there. Museum collections can enable the assessment of populations because they have data associated with specimens (particularly where and when they were collected) and this can be used to assess changes in geographical ranges over time.

The digitisation of collections therefore accelerates the rate at which species may be identified as threatened and action can be taken to protect them before it is too late. Digitally available data enables faster and easier access to crucial information (which if not digitised researchers may not be aware exists or, at a minimum, may be difficult to find and access). This can speed up the identification of threatened species. The impacts of digitisation would therefore be highest for countries rich in biodiversity but poor in biodiversity data. And there are many species for which good data is lacking. The International Union for Conservation of Nature's Red List of Threatened Species contains over 20,000 species which are data deficient (Suppl. material 1) – these are species for which insufficient data is available to make a direct or indirect assessment of their risk of extinction.

Valuing the impact of digitisation on conservation

To value the impact of digitisation on conservation we need to:

- Estimate the value UK citizens place on preventing species declining anywhere in the world; and
- Estimate the rate at which digitisation accelerates the identification of threatened species.

Regarding the first point, existing evidence suggests that the possible gains from digitising collections data are large. Even if digitisation leads to very modest improvements in the identification of threatened species (enabling conservation efforts), societal benefits can run in the hundreds of millions if not billions. Previous studies estimating the value

of biodiversity to UK citizens have shown that the value of individual relevant categories run into the billions. (Wilson 2008). For example, marine ecosystem services are shown to be worth at least £1.6 billion per year and tropical forest ecosystem services are shown to be worth at least £1.3 billion per year.

If digitisation affects only a small proportion of this value (say 1% of marine ecosystem services value) by accelerating the identification of threatened species and enabling productive conservation efforts, the discounted benefits to UK citizens would be close to £300 million (see low bound in Table 5).

It is difficult to predict precisely how far digitisation can accelerate the identification of threatened species but existing work like the 'Plants under pressure' project (Natural History Museum 2021c) give us clues as to how digitisation might affect the process in practice (see case study below for details).

Overview

The Plants Under Pressure project measures how many plant species are threatened with extinction, where these plants grow and why they are threatened.

Together with the Royal Botanic Gardens, Kew, the Natural History Museum has produced a first IUCN Sampled Red List Index (SRLI) for Plants, using a sample of thousands of species from around the world drawn at random from major plant lineages. The work is helping to mobilise information about the plant world and highlight gaps in knowledge.

The assessments involve inference of the appropriate IUCN Red List Category at an earlier point in time using information only known at a later point in time, a technique known as back-casting.

The work suggests that many more plant species are threatened than previously thought. According to the Museum's work one in five plant species worldwide are threatened. This is significantly higher than the 3% of plants which are currently on the Red List.

Why digitisation matters?

Since the majority of plant species lack any documentation on population sizes or dynamics, for most species the most comprehensive, easily accessible and reliable information on which to base a conservation assessment is the known distribution of that species. The best source of this distribution information is the collection of plant specimens held in the world's herbaria. Herbarium specimens provide verifiable records indicating the existence of a species at a given time and place. Historical specimens may play a role in assessments as they can indicate that a decline

has occurred, if contemporary surveys reveal that the species is no longer extant at a locality where a historical collection was made. The herbarium collections at the Royal Botanic Gardens, Kew contain some seven million plant and fungus specimens, and the Natural History Museum, London contains six million plant specimens from all over the world. These were systematically searched for all available specimen information for a given species, and all collections databased and geo-referenced, together with other specimen data available online. These specimen data were combined with information about the species from scientific literature, analysis of the species range using Geographical Information Systems (GIS) together with widely available global datasets.In such a significant undertaking, having access to digital data is extremely important — close to 1 million of the plant specimens in the collections at the Museum have been digitised for discovery so far (Natural History Museum 2021b). Indeed without access to digitised species data, the research would have required significantly longer to complete.

There are around 350,000 plant species in the world (Antonelli et al. 2020) and the Plants Under Pressure project has suggested that around one in five (ca. 80,000) are threatened, almost 6.7 times more than the number of plants on the Red List.

In order to estimate the value of digitisation for ecosystems services we assume that digitisation increases the identification of threatened species by 5 percentage points and that this enables conservation efforts which prevent half of these threatened species from declining. Further, we assume that the value of ecosystems services is proportional to the number of species in existence – specifically we assume that for every 1% of species prevented from declining 0.5% of ecosystem services value decline is avoided. Based on these assumptions we estimate that digitisation could create value to UK citizens in the order of £670 million –£1 billion (see Table 7).

Table 7. Detailed value calculation for ecosystem services to UK citizens. Data behind values of variables a, d, e, and f are from Wilson (2008) , Antonelli et al. (2020), Natural History Museum 2021c , and the IUCN Redlist respectively.

Variable	Description	Low bound	High bound	Formula
a	Annual benefit to UK citizens of ecosystem services (tropical forests, marine services)	£2,900,000,000	£4,500,000,000	
b	Discount rate	3.50%	3.50%	
c	Threatened speces detected with support from digitised information	78,200	78,200	d * e

Variable	Description	Low bound	High bound	Formula
d	Total number of plant species	350,000	350,000	
e	% of threatened species uncovered	20%	20%	
f	Baseline % of threatened plant species	3%	3%	
g	Additional threatened species detected with digital data (over and above business as usual)	5%	5%	
h	Additional species where decline is prevented/slower (over and above business as usual)	2.5%	2.5%	g / 2
i	Ecosystem services value which could in principle be prevented from declining	£72,500,000	£112,500,000	a * h
j	Ecosystem services value preventable with support from digitisation	£36,250,000	£56,250,000	j / 2
k	Total discounted value attributable to digitisation	£666,711,646	£1,034,552,554	NPV j over 30 years @ 3.5%

Benefits of digitisation to mineral exploration

The Natural History Museum houses vast collections of meteorites, rocks, minerals, and fossils. They help researchers in furthering understanding across different areas – defining the evolution of the planetary surface, characterising the climate system, and discovering mineral and energy resources among others. In particular, these collections can play an important role in identifying deposits of valuable natural resources for the UK and the world. The use of museum collections, often collected for different purposes, has enabled the discovery of mineral deposits which have significant economic value.

USE OF MUSEUM'S COLLECTION TO RESEARCH LITHIUM MINERALS

Importance of new materials to tackle climate change

Use of renewable energy is at the forefront of solutions to help the shift towards low carbon economies. Batteries have emerged as a key part of the solutions to support low carbon transport and stable electricity supplies in a net zero world, and lithium ion batteries are currently the most viable short term battery technology available.

Museum records of rocks and minerals are helping researchers to tap the lithium potential in the UK

Currently, there is no commercially viable lithium production in UK and the rest of Europe. It is increasingly becoming important to investigate possible sources to develop a sustainable domestic lithium supply chain. The Li4UK project is looking into the feasibility of producing lithium compounds from lithium found in rocks and geothermal waters of the UK which can be used for battery production. The Museum's Earth Sciences collections were used to identify Lithium deposits in Cornwall and in Scotland (Natural History Museum 2021e). The domestic production of Lithium will support UK manufacturing, particularly in the context of rules of origin regulation in the aftermath of Brexit[*15]. The UK's potential annual demand for Lithium is forecast at over 75,000 tonnes by 2035 (The Faraday Institution 2020) and these new deposits have the potential to deliver almost all of it (Bliss 2021).

Collections are a valuable data base against which researchers can cross check samples in order to identify deposits of known valuable minerals as illustrated above (e.g. by searching samples from old closed mines which may no longer be accessible). The economic benefits that this brings are associated with:

- **Efficiency of discovery**: by having a reliable and accurate database to cross check against, the discovery of valuable deposits should occur at a lower cost and with lower environmental impact; and
- **Efficiency of processing**: once a deposit has been discovered, collections can enable efficiency in how this is exploited – e.g., copper typically accounts for only 1% of rock mass with the remaining 99% being waste. Collections can allow us to understand what other elements of value may be present such as Critical Raw Materials (within the 99% of rock mass) and bring economies of scale and scope to processing. Efficiencies of recovery will be a significant factor in reducing the environmental impact of mining.

As well as maximising the value of known materials, collections can also help in expanding the understanding of available materials which we currently know little about, thus supporting their use including in fast growing industries such as green energy generation. For example, a recent study involved researching the mineral incidence and behaviour of Scandium — a rarely used metal — which has enormous potential to be used in the aerospace industry and in sustainable energy resources (as a catalyst in fuel cells and in environment friendly lightbulbs) (Natural History Museum 2021a).

The Museum collection of mineral samples including rocks, gems, minerals and meteorites can be valuable resources for economic geology and scientific research. Digitisation can accelerate the process by:

- **Improving the accuracy of the existing database**: geological work requires extremely precise geo-location data in order to achieve its maximum impact. Relatively small deviations (as small as several meters) can increase costs of discovery by hundreds of thousands of pounds. Digitising the collection can ensure that information catalogued by the Museum are accurate, and that data are consistently recorded to be used in geological exploration projects commercially.
- **Providing additional information on the specimens**: more and better economic geology data can be captured for specimens such as their chemical and structural information: i.e. crystallographic group, elemental composition, and association with other minerals etc. which can again accelerate the discovery process and minimise costs by de-risking discovery, allowing the exploration projects in the industry to know when to stop).

Finally, having digitised information which is easily accessible can enable fundamental scientific research to take place and help level the playing field by providing open data to firms in the geological exploration industry, thereby enabling innovation in the sector.

Valuing the impact of digitisation

To value the impact of digitisation on mineral exploration we need to:

- Estimate how digitisation affects the discovery process and/or fundamental scientific research; and
- Estimate the value of any efficiencies it helps to achieve during discovery and/or try to value the fundamental research that might take place.

One of the industries which directly depends on successful exploration and discovery of mineral deposits is the mining industry. Mineral exploration requires high investment and sustained time with inherent high risk and is the foundation of all value creation in mining. The process involves exploration drilling, a technique used in the sector to 1) explore for new mineral prospects 2) evaluate land for economic mining, and 3) augment additional ore reserves and resources in the mine. Drilling is used to obtain detailed information about rock types, mineral content and rock fabric. These provide the basic information before a commercial case for mining or extraction can be made at any given location.

Digitisation of collections can impact the discovery process of minerals, ores and rocks as museums hold large numbers of valuable geological samples. These can serve as es-

sential research material to understand the viability of mineral exploration in locations where sites for deposits have closed down or become inaccessible, or where the costs of collecting new samples from scratch far exceeds the benefits, because of a high degree of uncertainty of successful discovery.

To estimate the impact of digitisation on the discovery process we would ideally need evidence regarding the improved accuracy of geological information brought about due to digitisation. Subsequently, estimating the efficiency benefits from this improved sample data would require us to evaluate the costs savings to all affected exploration companies who now will need to either drill less far into the ground and/or de-risk the drilling process due to higher probability of a successful discovery.

Due to data limitations, we derive the benefits by assuming that digitisation leads to a small fraction of cost savings through more and better geological data available to all drilling projects currently in operation globally. The data can be expected to benefit new exploration projects as well which will lead to a future stream of annual benefits in the mining industry. To understand the UK share in these benefits we estimate UK's share of employment in the global mining industry and apportion the global benefits accordingly.

Assuming that digitisation led to a 0.5% cost savings for each drilling project, we can see benefits to the UK ranging between £17 million to £84 million in present value over 30 years (see Table 8).

Table 8. Detailed value calculation for cost savings in mineral exploration. Lower and upper bounds based on drilling 100 kilometres and 500 kilometres in total respectively. Assumptions based on: (b) drilling projects to require drill holes ranging from 100km to 500km to capture variability in types of projects, (d) S&P Global Market Intelligence (2020), (g) S&P Global Market Intelligence (2020), (k) Share of UK as a proportion of total employment in mining industry globally based on UK EITI (2021) and UNEP (2020) .

Variable	Description	Lower bound	Upper bound	Formula
a	Cost of drilling holes per metre	£700	£700	
b	Drilling cost per exploration project	£70,000,000	£350,000,000	
c	Impact of digitisation through cost savings	0.5%	0.5%	
d	Cost savings due to digitisation per project	£350,000	£1,750,000	b * c
e	Number of global active drilling projects currently	1,261	1,261	

Variable	Description	Lower bound	Upper bound	Formula
f	Average number of new drilling projects annually	100	100	
g	Discount rate	3.5%	3.5%	
h	Current potential savings for exploration activities globally	£441,350,000	£2,206,750,000	d * e
i	Potential savings in the future annually (global)	£35,000,000	£175,000,000	d * f
j	Total discounted value attributable to digitisation	£1,036,330,285	£5,181,651,425	NPV (i) over 30 years @3.5%
k	UK share of potential savings	1.6%	1.6%	
l	Total discounted savings for drilling projects for UK	£16,805,356	£84,026,780	i * k

It is worth noting that apart from the potential efficiency savings for the exploration industry, a possible significant benefit from digitisation is the enabling of local discovery of valuable minerals which could minimise supply chain issues going forward. Global shortages of Lithium for example have been extensively reported on by media in recent days and concerns raised about how these can affect the adoption of electric vehicles in the UK and abroad. Further, although we don't value this explicitly, making the Museum minerals collection more easily accessible can facilitate discovery overseas leading to benefits across the globe.

5. CONCLUSION

This project has looked at the value that digitising the entire Natural History Museum collection can unlock for the UK economy, while also recognising that much value will also be generated abroad. We have drawn on the best available data and evidence to produce estimates of monetary value which capture a range of potential areas of research, from human health to biodiversity conservation, enabled or facilitated by digitisation of the Museum collection.

Our estimates indicate that the value of investing in digitisation can be very large, consistent with existing literature. Specifically, we estimate that investing around £200 million in digitisation can unlock benefits in the region of £1.4 billion–£2.2 billion (Fig. 4). That is, for every one pound invested in digitising the Museum collection, we can expect to receive a return of seven to ten pounds in societal benefits.

These high-level impacts are underpinned by specific areas in which digitisation of the museum's collections is expected to boost research, ultimately leading to benefits for the UK society and the world. Examining these thematic areas help to provide a more detailed and specific view into how the benefits from digitisation can be expected to materialise in practice. The cumulative estimated value across these areas provides further support to the value from digitisation estimated at a high level above. We have looked at:

- **Biodiversity conservation**: digitisation can enhance taxonomic knowledge which can improve the detection of threatened species. This in turn enables conservation efforts to be put in place which reduce the rate at which species decline maintaining the ability of ecosystems to provide vital services for humanity.
- **Medicines discovery**: digitisation can improve accessibility of samples and consequently the range of samples tested for the drug discovery and commercialisation. There are very large economic benefits from successful drug discovery. Even if digitisation leads to a very small increase in the rate of discovery, the benefits are large.
- **Invasive species**: digitisation allows for better and faster detection of invasive species which have costs for the UK economy estimated to be £2 billion a year. Reducing the frequency of losses from threats can thus lead to substantial economic benefits.
- **Agricultural R&D**: Digitisation can help in the discovery and/or improve the understanding of the genetic traits of Crop Wild Relatives (CWR), wild plant species that are genetically related to cultivated crops. This can enhance breeding of crops which are environmentally friendly, have higher yields and are disease resistant.
- **Mineral exploration**: digitisation can improve the accuracy of existing geological data and help develop enhanced geological data from the museum's rich collections of minerals, ores and rocks. This accelerates the geological discovery process and reduces the risks and thereby costs of exploration significantly.

Fig. 3 summarises these benefits which can exceed £2 billion according to our estimates.

These are just a small number of areas in which tangible estimates of economic value were explored as part of this study. There are many more areas which could and would benefit from digitisation but have not been considered in this study. The estimates have been based entirely on data from secondary sources. The ranges reflect the uncertainty in those such estimates. One of the most significant benefits of digitisation is that it can foster blue skies research in any number of areas which are impossible to predict

and value beforehand. These can be expected to come about through more and better linking of data across collections in the UK, EU and rest of the world including associated datasets in the fields of climate, population, health and elsewhere.Future research could add significant value to this work by studying more closely the user base of the Museum's digital data and tracking the ultimate impacts of the data beyond academic publications and onward citations. Survey work and case studies with users are a good option for examining these wider impacts of digitisation. This would help to understand how digitised data is useful in practice and the exact pathways through which it can produce benefits and for whom.

FUNDING PROGRAM

This study has been carried out by Frontier Economics Ltd under contract to the Natural History Museum, London. The study is funded by the Museum.

AUTHOR CONTRIBUTIONS

Authors:

Frontier Economics*: **Danail Popov** - conceptualization, formal analysis, investigation, methodology, project administration, validation, and writing- original draft. **Priyanka Roychoudhury** - conceptualization, formal analysis, investigation, methodology, project administration, validation, and writing- original draft.

Natural History Museum: **Helen Hardy** - conceptualization, resources, supervision, and writing-review & editing. **Laurence Livermore** - resources, and writing-review & editing. **Ken Norris** - conceptualization, resources, supervision, and writing-review & editing.

Contributors:

The following Natural History Museum staff were consulted during the conceptualisation period, provided resources and/or data to Frontier for analysis, and reviewed/commented on drafts of the manuscript: Dr Robin Armstrong, Dr Neil Brummitt, Dr Matt Clark, Kathryn Gibbons, Prof Richard Herrington, Dr Tim Littlewood, Dr Giles Miller, Kathryn Packer, Dr Ben Price, Jennifer Pullar, Prof Andy Purvis, Mike Rumsey, Dr Vincent Smith, Jo Wilbraham.

The following Frontier Economics staff supervised the work and reviewed drafts of the manuscript: Matthew Bell, Sarah Snelson.

Contribution types are drawn from CRediT - Contributor Roles Taxonomy.

*Frontier Economics Ltd is a member of the Frontier Economics network, which consists of two separate companies based in Europe (Frontier Economics Ltd) and Australia (Frontier Economics Pty Ltd). Both companies are independently owned, and legal commitments entered into by one company do not impose any obligations on the other company in the network. Views expressed in this document are the views of Frontier Economics Ltd.

REFERENCES

Aghion P, Howitt P (1992) A Model of Growth Through Creative Destruction. National Bureau of Economic Research https://doi.org/10.3386/w3223

Antonelli A, Smith RJ, Simmonds MS, Kersey PJ, Pritchard HW, et al (2020) State of the World's Plants and Fungi 2020. Royal Botanic Gardens, Kew. https://doi.org/10.34885/172

Ashworth J (2021) Dragonfly species are colonising Britain 'at a greater rate than ever before'. https://www.nhm.ac.uk/discover/news/2021/september/dragonfly-species-are-colonising-the-uk-at-a-greater-rate.html. Accessed on: 2021-11-24.

Atkinson R (2020) How can digital platforms help museums connect to audiences during Covid-19 emergency? https://www.museumsassociation.org/museums-journal/news/2020/03/19032020-how-digital-platforms-be-used-to-support-musems-during-covid-19-crisis/#. Accessed on: 2021-11-24.

Beagrie N, Houghton J (2014) The Value and Impact of Data Sharing and Curation: A synthesis of three recent studies of UK research data centres. Jisc. URL: https://repository.jisc.ac.uk/5568/1/iDF308_-_Digital_Infrastructure_Directions_Report%2C_Jan14_v1-04.pdf

Beagrie N, Houghton J (2016) The Value and Impact of the European Bioinformatics Institute. EMBL-EBI. URL: https://beagrie.com/static/resource/EBI-impact-report.pdf

Beattie A, Barthlott W, Elisabetsky E, Farrel R, Kheng C, Prance I, Rosenthal J, Simpson D, Leakey R, Wolfson M, Laird S (2005) New Products and Industries from Biodiversity. In: Hassan R, Scholes R, Ash N (Eds) Ecosystems and Human Well-being: Current State and Trends. 1. Island Press, Washington, DC. URL: https://www.millenniumassessment.org/documents/document.279.aspx.pdf [ISBN 1-55963-227-5].

Bliss D (2021) In Cornwall, ruinous tin and copper mines are yielding battery-grade lithium. Here's what that means. https://www.nationalgeographic.co.uk/science-and-technology/2021/05/in-cornwall-ruinous-tin-mines-are-yielding-battery-grade-lithium-heres-what-that-could-mean. Accessed on: 2021-11-26.

CABI (2021) Floating pennywort. https://www.invasive-species.org/species/floating-pennywort/. Accessed on: 2021-11-25.

Calixto JB (2019) The role of natural products in modern drug discovery. Anais da Academia Brasileira de Ciências 91 https://doi.org/10.1590/0001-3765201920190105

Council of Europe (1979) Convention on the Conservation of European Wildlife and Natural Heritage. Bern, Switzerland. URL: http://conventions.coe.int/Treaty/EN/Treaties/Html/104.htm

Cuthbert R, Bartlett A, Turbelin A, Haubrock P, Diagne C, Pattison Z, Courchamp F, Catford J (2021) Economic costs of biological invasions in the United Kingdom. NeoBiota 67: 299-328. https://doi.org/10.3897/neobiota.67.59743

Deloitte Access Economics (2020) Cost benefit analysis of a mission to discover and document Australia's species. URL: https://www2.deloitte.com/content/dam/Deloitte/au/Documents/Economics/deloitte-au-economics-aas-rapid-cost-benefit-analysis-tas-mission-060521.pdf

Fell M (2019) The Economic Impacts of Open Science: A Rapid Evidence Assessment. Publications 7 (3). https://doi.org/10.3390/publications7030046

Fielder H, Brotherton P, Hosking J, Hopkins J, Ford-Lloyd B, Maxted N (2015) Enhancing the Conservation of Crop Wild Relatives in England. PLOS ONE 10 (6). https://doi.org/10.1371/journal.pone.0130804

Frontier Economics (2014) Rates of return to investment in science and innovation. Frontier Economics Ltd.. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/333006/bis-14-990-rates-of-return-to-investment-in-science-and-innovation-revised-final-report.pdf

GBIF (2021) GBIF Cited Literature Resources, Publisher=Natural History Museum. https://www.gbif.org/resource/search?contentType=literature&publishingOrganizationKey=19456090-b49a-11d8-abeb-b8a03c50a862. Accessed on: 2021-11-24.

Goettsch B, Urquiza-Haas T, Koleff P, Acevedo Gasman F, Aguilar-Meléndez A, Alavez V, Alejandre-Iturbide G, Aragón Cuevas F, Azurdia Pérez C, Carr J, Castellanos-Morales G, Cerén G, Contreras-Toledo A, Correa-Cano ME, De la Cruz Larios L, Debouck D, Delgado-Salinas A, Gómez-Ruiz E, González-Ledesma M, González-Pérez E, Hernández-Apolinar M, Herrera-Cabrera B, Jefferson M, Kell S, Lira-Saade R, Lorea-Hernández F, Martínez M, Mastretta-Yanes A, Maxted N, Menjívar J, Ángeles Mérida Guzmán M, Morales Herrera A, Oliveros-Galindo O, Orjuela-R. MA, Pollock C, Quintana-Camargo M, Rodríguez A, Ruiz Corral JA, Sánchez González JdJ, Sánchez-de la Vega G, Superina M, Tobón Niedfeldt W, Tognelli M, Vargas-Ponce O, Vega M, Wegier A, Zamora Tavares P, Jenkins RB (2021) Extinction risk of Mesoamerican crop wild relatives. PLANTS, PEOPLE, PLANET 3 (6): 775-795. https://doi.org/10.1002/ppp3.10225

Harvey A (2008) Natural products in drug discovery. Drug Discovery Today 13: 894-901. https://doi.org/10.1016/j.drudis.2008.07.004

Health Economics Research Group, Office of Health Economics, RAND Europe (2008) Medical Research: What's it worth? UK Evaluation Forum. URL: https://mrc.ukri.org/publications/browse/medical-research-whats-it-worth/

H.M. Treasury (2020) The Green Book: Central Government Guidance on Appraisal and Evaluation. H.M. Treasury, London, 152 pp. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/938046/The_Green_Book_2020.pdf [ISBN 978-1-912225-57-6]

Houghton J, Sheehan P (2009) Estimating the Potential Impacts of Open Access to Research Findings. Economic Analysis and Policy 39 (1): 127-142. https://doi.org/10.1016/s0313-5926(09)50048-3

Houghton JW, Rasmussen B, Sheehan P (2010) Economic and social returns on investment in open archiving publicly funded research outputs. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.187.226&rep=rep1&type=pdf

House of Commons Environmental Audit Committee (2019) Invasive species. First Report of Session 2019. House of Commons. URL: https://publications.parliament.uk/pa/cm201919/cmselect/cmenvaud/88/88.pdf

IPBES (2019) Summary for policymakers of the global assessment report on biodiversity and ecosystem services. Zenodo https://doi.org/10.5281/zenodo.3553458

Jacklin L (2021) How natural history museums can help fight future pandemics. https://www.nhm.ac.uk/discover/how-natural-history-museums-can-help-fight-future-pandemics.html. Accessed on: 2021-11-24.

Leclerc E (2021) Scientists Find Blue-Green Algae Chemical with Cancer Fighting Potential. https://www.smithsonianmag.com/blogs/national-museum-of-natural-history/2021/03/04/scientists-find-blue-green-algae-chemical-cancer-fighting-potential/. Accessed on: 2021-11-29.

Natural History Museum (2021a) Scandium in geological systems. https://www.nhm.ac.uk/our-science/our-work/sustainability/scandium-in-geological-systems.html. Accessed on: 2021-11-26.

Natural History Museum (2021b) Data Portal query on 1 resources created at 2021-11-26 14:22:31.036325 PID. Natural History Museum. URL: https://doi.org/10.5519/qd.xzz3wtut

Natural History Museum (2021c) Plants Under Pressure. https://www.nhm.ac.uk/our-science/our-work/biodiversity/plants-under-pressure.html. Accessed on: 2021-11-26.

Natural History Museum (2021d) Natural History Museum contributing to international project to help prevent future pandemics. https://www.nhm.ac.uk/press-office/press-releases/natural-history-museum-contributing-to-international-project-to-.html. Accessed on: 2021-11-25.

Natural History Museum (2021e) Li4UK: Securing a domestic lithium supply chain for the UK. https://www.nhm.ac.uk/our-science/our-work/sustainability/li4uk.html. Accessed on: 2021-11-26.

Newman D, Cragg G (2016) Natural Products as Sources of New Drugs from 1981 to 2014. Journal of Natural Products 79 (3): 629-661. https://doi.org/10.1021/acs.jnatprod.5b01055

OECD (2015a) Health at a Glance 2015: OECD Indicators. OECD Publishing, Paris. https://doi.org/10.1787/health_glance-2015-en

OECD (2015b) *Data-Driven Innovation: Big Data for Growth and Well-Being*. OECD Publishing, Paris. https://doi.org/10.1787/9789264229358-en

Osterloff E (2017) Can sponges cure cancer? https://www.nhm.ac.uk/discover/can-sponges-cure-cancer.html. Accessed on: 2021-11-25.

Osterloff E (2018) Why Asian hornets are bad news for British bees. https://www.nhm.ac.uk/discover/why-asian-hornets-are-bad-news-for-british-bees.html. Accessed on: 2021-11-25.

Pavid K (2020) New insights will help the world protect future wheat supplies. https://www.nhm.ac.uk/discover/news/2020/november/new-insights-will-help-the-world-protect-future-wheat-supplies.html. Accessed on: 2021-11-24.

Pullar J (2018) Digitising British Bumblebees |Digital Collections Programme. https://naturalhistorymuseum.blog/2018/08/13/digitising-british-bumblebees-digital-collections-programme/. Accessed on: 2021-11-25.

PwC (2013) Crop wild relatives: A valuable resource for crop development. PwC Valuations. URL: https://pwc.blogs.com/files/pwc-seed-bank-analysis-for-msb-0713.pdf

RSPB (2021) Invasive non-native species. https://www.rspb.org.uk/our-work/policy-insight/species/invasive-non-native-species/. Accessed on: 2021-11-24.

S&P Global Market Intelligence (2020) PDAC Special Edition: World Exploration Trends 2019. S&P Global Market Intelligence. URL: https://www.spglobal.com/marketintelligence/en/news-insights/research/report-world-exploration-trends-2019

The Faraday Institution (2020) Lithium, Cobalt and Nickel: The Gold Rush of the 21st Century. Faraday Insights (6). URL: https://faraday.ac.uk/wp-content/uploads/2020/12/Faraday_Insights_6_Updated_Dec2020.pdf

UK EITI (2021) Sector Data. https://www.ukeiti.org/sector-data. Accessed on: 2021-11-29.

UNEP (2020) Why Does Extractives Matter? https://www.unep.org/explore-topics/extractives/why-does-extractives-matter. Accessed on: 2021-11-30.

United Nations (1993) The Convention on Biological Diversity. URL: https://treaties.un.org/doc/Treaties/1992/06/19920605%2008-44%20PM/Ch_XXVII_08p.pdf

United Nations (2019) UN Report: Nature's Dangerous Decline 'Unprecedented'; Species Extinction Rates 'Accelerating'. https://www.un.org/sustainabledevelopment/blog/2019/05/nature-decline-unprecedented-report/. Accessed on: 2021-11-26.

Walkowiak S, Gao L, Monat C, Haberer G, Kassa M, Brinton J, Ramirez-Gonzalez R, Kolodziej M, Delorean E, Thambugala D, Klymiuk V, Byrns B, Gundlach H, Bandi V, Siri JN, Nilsen K, Aquino C, Himmelbach A, Copetti D, Ban T, Venturini L, Bevan M, Clavijo B, Koo D, Ens J, Wiebe K, N'Diaye A, Fritz A, Gutwin C, Fiebig A, Fosker C, Fu BX, Accinelli GG, Gardner K, Fradgley N, Gutierrez-Gonzalez J, Halstead-Nussloch G, Hatakeyama M, Koh CS, Deek J, Costamagna A, Fobert P, Heavens D, Kanamori H, Kawaura K, Kobayashi F, Krasileva K, Kuo T, McKenzie N, Murata K, Nabeka Y, Paape T, Padmarasu S, Percival-Alwyn L, Kagale S, Scholz U, Sese J, Juliana P, Singh R, Shimizu-Inatsugi R, Swarbreck D, Cockram J, Budak H, Tameshige T, Tanaka T, Tsuji H, Wright J, Wu J, Steuernagel B, Small I, Cloutier S, Keeble-Gagnère G, Muehlbauer G, Tibbets J, Nasuda S, Melonek J, Hucl P, Sharpe A, Clark M, Legg E, Bharti A, Langridge P, Hall A, Uauy C, Mascher M, Krattinger S, Handa H, Shimizu K, Distelfeld A, Chalmers K, Keller B, Mayer KX, Poland J, Stein N, McCartney C, Spannagl M, Wicker T, Pozniak C (2020) Multiple wheat genomes reveal global variation in modern breeding. Nature 588 (7837): 277-283. https://doi.org/10.1038/s41586-020-2961-x

Wilding E, Cockel CP (2019) Conserving the cousins of our crops. https://www.kew.org/read-and-watch/conserving-crop-wild-relatives. Accessed on: 2021-11-26.

Wildlife and Countryside Link (2020a) Prevention is better than cure. Wildlife and Countryside Link. URL: https://www.wcl.org.uk/assets/uploads/img/files/Prevention_is_Better_than_Cure_Report_2020_1.pdf

Wildlife and Countryside Link (2020b) Spend £6 million a year to save £2.7 billion on nature invader costs to the economy by 2040. https://www.wcl.org.uk/nature-invader-costs-to-economy-2020.asp. Accessed on: 2021-11-25.

Wilson S (2008) Benefits of Global Biodiversity Assets to UK Citizens: A Literature Review. Defra. URL: http://randd.defra.gov.uk/Document.aspx?Document=WC0707_7444_FRP.pdf

Zuanni C (2020) Mapping museum digital initiatives during COVID-19. https://pro.europeana.eu/post/mapping-museum-digital-initiatives-during-covid-19. Accessed on: 2021-11-24.

SUPPLEMENTARY MATERIAL

Suppl. material 1: Query summary of IUCN Red List Data - data deficient species

Authors: IUCN Red List

Data type: IUCN assessments

Brief description:

Query summary of IUCN data using the following parameters on 26/11/2021:
Geographical scope='Global'
Red List Category='DD - Data Deficient'
Include='Species'

Download file (3.19 MB)

ENDNOTES

[*1] Exchange rates from xe.com correct as of 08/11/2021.

[*2] Based on engagement with colleagues across different areas of expertise at the Museum and wider literature review.

[*3] This is a higher-end estimate for the costs of digitising the entire Museum collection.

[*4] The low range is based on a return on investment of 20% while the high estimate is b.ased on a 40% return on investment.

[*5] Based on GBIF Resources, GBIF 2021.

[*6] We assumed that digitisation would involve a mix of basic digitisation (for discovery) as well as more enhanced levels of digitisation which are required in order to unlock some of the benefits examined in the study.

[*7] We used current usage of Museum collections data available digitally to estimate the value of research efficiencies made possible by digitisation.

[*8] CWRs are defined as wild plant species which are genetically related to crops but have not been domesticated.

[*9] For the purposes of estimating benefits, we use a discount rate of 3.5% throughout this report based on the Green Book's Social Time Preference rate (STPR) (H.M. Treasury 2020).

[*10] Diseases caused by germs which spread between animals and humans.

[*11] This would be equivalent to an increase of 0.25% assuming the number of samples tested in the UK are higher than those in Australia in proportion to R&D expenditure in the pharmaceutical sector.

[*12] It is important to note here that health benefits and market value cannot be added as they represent value flowing from the same transaction between patients and pharmaceutical firms.

[*13] This takes into account the length of the drug discovery lifecycle from successful discovery to commercialisation i.e. a 10 year delay in realising benefits.

[*14] This is the value received by farms from the sale of agricultural produce at the farm gate. It does not take into account the value derived from trading of agricultural produce or form the processing of agricultural produce into consumer foods.

[*15] In order to be eligible for duty free trade vis a vis the EU, UK exporters need to show that their products originate in either the UK or the EU. Because modern value chains are complex and involve inputs from many sources, countries that are party to a Free Trade Agreements (FTA) set Rules of origin which stipulate the threshold amounts of content (usually as a share of value added) that must originate from within the countries that are party to the FTA. In this case the UK and the EU.

[*16] The Economic and Social Data Service (now the UK Data Service), the Archaeology Data Centre (now the Archaeology Data Service) and the British Atmospheric Data Centre (part of the Centre for Environmental Data Analysis).