City University London

Informational Interfaces: A case study of the impact of discovery systems on biodiversity research and search at the Natural History Museum, London, UK

Sarah A. Stewart

January 2015

Submitted in partial fulfilment of the requirements for the degree of M.Sc. in Library Science

Supervisor: Dr. Lyn Robinson

Abstract

The Natural History Museum is one of the world's leading research institutions for biological diversity research with collections comprising some 8 million specimens representing a vast wealth of data. It also holds one of the foremost collections of biodiversity-related literature in its Library and Archives, with over one million items catalogued, ranging from artworks and historical manuscripts to born-digital material such as journal articles. As part of its strategic plan, the Natural History Museum is transitioning towards a digital environment, and the implementation and launch of the Library and Archives discovery system, Primo, and the Informatics Data Portal reveal this transition towards an informational museum environment.

Information in biodiversity is heterogeneous and fragmented, but also is vital for research into understanding the natural world. There has been an international call for improved management of information in biodiversity, with the consequent rise of cybertaxonomy and biodiversity informatics, including the use of single-search data portals to manage information in biodiversity. Concurrently, discovery systems have proliferated in academic libraries, where they have increased the use of the collections and facilitated efficient search and access to information through single-search interfaces.

This study investigates the impact of discovery systems on search and research at the Natural History Museum with the objectives: to determine exact attitudes to searching for biodiversity information resources using the Library and Archives discovery system and the Data Portal; to explore how biodiversity information is discovered and used through the discovery system; to examine the impact of discovery systems on search and research; and to document the transition toward a digital, networked and aggregated discovery system approach to managing biodiversity information. To date, this is the first case study of the implementation of a discovery system in a natural history museum library with a focus on biodiversity research.

A case study approach was utilised, with evidence collected from document analysis, surveys and interviews, including interviews with key informants. The results from this study have revealed that while it may be too early since the implementation to gauge impact on search and research, particularly for the Data Portal, which was launched only in December 2014, there is some ambivalence surrounding the impact on search and research among the system users, which may point to the need for further outreach by the Library and Archives. Library staff were optimistic that the new system would facilitate improved, efficient searches by researchers, and therefore have a positive impact on research at the Natural History Museum. Although it was not possible to investigate the impact of the Data Portal owing to its late launch, the initial feedback suggests the possibility that the Data Portal will gobeyond discovery, pushing information resources created in the Museum to a wider, external audience.

Table of Contents

Abstract	2
Acknowledgements	5
Chapter 1 - Introduction	6
1.1 Aims and Objectives	6
1.2 Rationale	6
1.3 Scope and Definitions	8
Chapter 2 – Biodiversity and Discovery Layers: A Review of the Literature	10
2.1 Biodiversity	10
2.2 Research and Information in Biodiversity	10
2.3 Informational Crisis in Biodiversity	11
2.4 Towards a web-based Cyberscience	12
2.5 Biodiversity on the Web	12
2.6 Biodiversity Informatics and Cybertaxonomy	14
2.7 Data Portals and the Centralisation of Information	15
2.8 Informational Role of Museums and Libraries	17
2.9 Towards the Virtual Library Environment	18
2.10 Search Interfaces and the Virtual Library	18
2.11 Discovery Systems and Interfaces	19
2.12 Discovery, Search and Research Impact	21
Chapter 3 - Methodology	22
3.1 Rationale and Research Framework	22
3.2 Limitations to the Case Study Approach	23
3.3 Initial Courses of Investigation and Design	23
3.4 Ethical Considerations	24
Chapter 4 – The Natural History Museum Library and Data Portal in Context	26
4.1 Introduction	26
4.2 History and Development of the Natural History Museum Library	26
4.3 Natural History Museum Library Collections	27
4.4 The Virtual Library Project	28
4.5 Initial Library and Archives User Survey	31
4.6 The Natural History Museum Data Portal	33
Chapter 5: Discovery System User Survey	35
5.1 Introduction	35
5.2 On-line Survey: Usage and Impact of the Library Discovery System	35
5.2.1 Preparation and Design of the Survey	35
5.2.2 Testing the Survey	38
5.2.3 Distribution and Launch	38
5.2.4 Detailed Results of the On-line Surveys	38
5.3 Data Portal Survey	42

5.4 Summary of Survey Results	43
Chapter 6 – Interviews with Library Users and Key Informants	45
6.1 Library and Archives User Interviews	45
6.1.1 Question Production and Protocol	45
6.2 Results from User Interviews	45
6.3 Summary of Evidence provided by Interviews	47
Chapter 7 - Conclusion	56
7.1 Summary of Findings	56
7.2 Attitudes to Searching for Biodiversity Information Resources using Primo	56
7.3 Information Discovery in Biodiversity Research	56
7.4 Evaluation of Impact	57
7.5 Documenting the Transition to a Virtual Informational Environment	57
7.6 Synthesis of Evidence	57
7.7 Limitations of the Study and Directions for Future Research	58
Bibliography	59
Appendix 1: Dissertation Proposal	65
Appendix 2: Natural History Museum Digital Strategy Internal Communication	71
Appendix 3: Project Information Sheet and Consent Forms for Interviews	76
Appendix 4: Sample Interview Questions	79
Appendix 5: Dissertation Reflection	80

Acknowledgements

I gratefully acknowledge financial support received from the Natural History Museum and the opportunity it provided to study Library Science at City University London.

I acknowledge the support, mentorship and guidance of my supervisor, Dr. Lyn Robinson, throughout the course of this research. She guided me through the research process and assisted me with any problems that I encountered.

I would also like to thank Dr. Vincent Smith and Mr. Ben Scott of the Natural History Museum Informatics Group for their time, enthusiasm, advice and input into my research.

I would like to thank my colleagues at the Natural History Museum Library and Archives for their interest and contributions to my work, particularly Jane Smith, Dr. Mel Smith, Sarah Vincent, Sarah Sworder and Michael Loran who agreed to be interviewed for this work. I also thank my line manager, Angela Thresher, and Julie Harvey of the Centre for Arts and Humanities Research for their insights into the research process.

I thank Andrew Preater of Imperial College Library for his input and advice on discovery systems and usability in a library setting, for providing me with an initial forum to explore ideas on discovery and interfaces at the Pi and Mash Libraries and Technology Unconference at Senate House Library (August 9th 2014), and also for his highly insightful comments on my research methodology and participation in an interview for the dissertation work.

I thank my family and friends, particularly my parents, Jim and Grace Stewart, for their support and encouragement.

Finally, I thank all of my colleagues at the Natural History Museum who participated in the surveys and interviews, without whom, this research would not have been possible.

Chapter 1: Introduction

1.1 Aims and Objectives

The primary aims of this study are to investigate the impact of discovery system interfaces on search and research in biodiversity at the Natural History Museum. This study also aims to produce a case study documenting the implementation and uptake of discovery system interfaces in both the Library and Archives (Ex Libris Primo) and the newly-developed Data Portal within the context of the Natural History Museum's greater strategic mandate of developing a virtual museum.

To accomplish the research aims, the following objectives have been proposed:

1.) Determine exact attitudes to searching for biodiversity information resources using the discovery layers of the library and the informatics data portal.

2.) Explore how biodiversity information is discovered and used between the library and the informatics department.

3.) Examine the impact of the discovery systems on search and research.

4.) Document the transition towards a digital, networked and aggregated discovery layer approach to managing biodiversity information.

1.2 Rationale: Towards a Virtual Museum and Cybertaxonomy

The Natural History Museum (London, UK) is one of the primary centres for research into biological diversity. It holds growing collections of over 79 million specimens, collected over the last 300 years and representing the world's diversity of animals, plants and geological specimens. It also holds one of the largest collections of taxonomic literature, spanning several centuries, including not only monographs, but also serials, manuscripts, correspondence, field notebooks and artworks.

The Natural History Museum Library and Archives holds a growing collection of over one million catalogued items. It has recently joined an international consortium including Kew Gardens, Leiden University, the Missouri Botanical Gardens and the Smithsonian in creating an on-line repository for legacy (out of copyright) taxonomic literature, the Biodiversity Heritage Library or BHL (http://www.biodiversitylibrary.org/). The BHL provides free public access to digitised materials from the collections of member institutions to enable further research into biological diversity. This is only one of many initiatives across the world to make collections of specimens and literature available for wider research purposes.

Researchers are working towards a more unified platform for the storage and exchange of data and information related to biodiversity research through biodiversity informatics and the building of a cybertaxonomy. There is a real "data deluge" in the biological sciences, not only in the field of genetics, with its high-throughput next generation sequencers which can sequence whole genomes of organisms, but also in biodiversity information, with a plethora of taxonomic records, and related data types

(Godfray, 2002b, 2002c; Hine, 2008; Sarkar, 2007, 2009) Data and information types in biodiversity research are heterogeneous, ranging from taxonomic nomenclature and description of characteristics defining a taxon, to phylogenetic trees gauging similarity and evolutionary relatedness of taxa. Geospatial information is also valuable, as it will outline the geographic distribution of a taxon. Genetic sequence data too is increasingly important, and with the advent of next-generation high-throughput sequencing, and the ability to sequence whole genomes, the amount of genetic data has risen exponentially (Hine 2008; Sarkar, 2007, 2009). In light of this "data deluge" a new field, biodiversity informatics, has emerged to capture, store and curate the burgeoning data and information related to biodiversity. Biodiversity informatics is also tied closely to the development of cybertaxonomy, "taxonomy on the web", a crucial part of biodiversity research (Hine, 2008; Sarkar, 2007, 2009).

In contrast to the vast amount and richness of information in biodiversity research is the "informational crisis" in the discipline of taxonomy during the last decade. This is related to the fragmentary, disparate, and heterogeneous nature of information related to biological diversity (Godfray, 2002b, 2002c; Hine, 2008; Raven, 2004). Taxonomists, the principal researchers in biodiversity, expend the majority of their time and effort trying to search for, retrieve and piece together information during the course of their work (Godfray, 2002b, 2002c; Hine, 2008; Raven, 2004, Scoble, 2004) This lack of cohesiveness and organisation in biodiversity information has often resulted in redundancy of work, in the duplication of studies on the same taxonomic groups, and the proliferation of poorly- or incompletely-defined taxa, hampering progress in the field (Select Committee on Science and Technology, 2002a, 2002b). The lack of uniformity and accessibility of information crucial to biodiversity research has resulted in an international call to make taxonomy a web-based science, which can easily be accessible to researchers worldwide. Furthermore, countries with high biodiversity are often in the developing world, and cannot afford to access the expertise and sources of information such as museum specimens and journal articles held in the collections of developed countries, further hampering progress in biodiversity studies (Godfray, 2002b, 2002c; Moritz, 2002; Select Committee on Science and Technology, 2002a, 2002b).

The Natural History Museum has a great wealth of informational resources held both within its collections and in its Library and Archives. It must support its researchers in facilitating access to these informational resources, which are vital to biodiversity research. In October 2013, the Natural History Museum announced its "Digital Strategy" calling for the development of a virtual museum (2013, Appendix 2). This strategy was met by the Natural History Museum's Informatics Group, which was developing a new Data Portal, and by the Natural History Museum Library and Archives, which was implementing a new, single-search discovery system, Ex Libris Primo, in order to facilitate search and discovery of informational resources within the library collections.

Both the Library discovery system and the Data Portal are notable for their use of single-search boxes enabling natural language, full-text searching, and this can be envisioned as emulating the interfaces of web-scale search engines such as Google, Yahoo or Bing. The informational interfaces for the Library and Archives discovery system and the Data Portal have been developed to be simple and intuitive for use in searches, with the user experience in mind. The use of a single-search box, ubiquitous in web-based search engines such as Google, may indicate, or imply that there has been a change in how

search and information retrieval is being conducted. This study will investigate whether these "Google-like" discovery interfaces impact on the search for, and information retrieval of, biodiversity information at the Natural History Museum.

1.3 Scope and Definitions

For the purposes of the dissertation work, the research will take place only at the Natural History Museum, London, U.K. The study will focus on scientific researchers in the biological or earth sciences who conduct research into biological diversity, including geological and mineral diversity.

In the dissertation work, a discovery interface, which also may be termed a "discovery layer" or "discovery system", is defined as a system or interface used to search for information held in library or collection catalogues or databases that utilises a single search box entry point to search the entirety of a database or catalogue. According to Kennedy (2014), any search query entered into a discovery interface will "scan the institution's central index, all metadata about the various resources they provide access to and return a list that includes items matching or related to the search query" (Kennedy, 2014). Discovery interfaces provide a "new, simpler way of searching…with a Google-style interface and new and improved user features" (British Library, 2010). Marshall Breeding (2010, 2012) defines the discovery system or interface as systems which "aim to provide access to all aspects of the library collections not just those managed by the traditional library catalogue which is limited to the content managed by the integrated library system. Ideally, a discovery system should channel resources managed by multiple systems into fewer numbers of user interfaces, providing a single-search entry point to all library resources" (Breeding, 2010, 2012). In addition to a simplified single-search interface, other major features of discovery systems include an expanded and consolidated scope of search, faceted navigation of search results, sophisticated search technologies, including full-text search, relevancy-ranking of results, enriched displays and other features such as the ability for users to add content (eg. tagging, sharing resources through social media or pushing resources into reference management software such as EndNote) (Breeding, 2010).

Biological diversity or biodiversity is defined as the description of the variety of life on earth, "encompassing the diversity of all life from humans to micro-organisms, the types of habitats in which they live and the genetic diversity of individuals within species" (Natural History Museum Biodiversity website, 2014). Biodiversity research is defined as a domain of study in the biological and earth sciences (palaeontology) that encompasses organismal diversity across both time and space. It is largely described in terms of systematics (classification of organisms into taxa, taxonomic groups such as genera and species), biogeography (the spatial or geographic distribution of extinct and extant taxa) and synecology (the interaction of organisms in ecological communities) (Berendsohn et al., 2011).

Biodiversity Informatics is a newly emergent field, originating in 2002, that draws upon information technology and data sciences that focuses on data, from preserved collections such as museums and herbaria, living collections such as zoos, botanical gardens and culture collections data derived from both professional research and citizen science initiatives such as bio-blitz surveys in addition to taxonomic literature (Sarkar, 2009; Berendsohn et al, 2011). It involves data curation,

information management, and data integration through the development of data standards platforms for data and information exchange.

Taxonomy is a branch of the biological sciences and a core component of the domain of biodiversity research focussing on naming and classification of life into discrete units based on shared characteristics. Modern taxonomy utilises the nomenclatural system developed by Carl Linneaus and is based on evolutionary relationships between organisms both extant and extinct. Although taxonomy may relate strictly to the naming of biological entities, taxa, it can encompass investigations of evolutionary relationships (systematics), biogeography, the study of species distributions and species ecology. This broader definition will be utilised in light of the richness and variety encompassed in this broader definition of biodiversity research.

Chapter 2: Biodiversity and Discovery Layers: A Review of the Literature

2.1 Biodiversity

The 1992 United Nations Convention on Biological Diversity defines biodiversity as the "variability among living organisms from all sources including…ecological complexes of which they are part: this includes diversity within species, between species and of ecosystems". Biological diversity is a global asset that is of tremendous value to present and future generations (United Nations, 1992), and is considered necessary for planetary well-being, having not only aesthetic and intellectual, or scientific value, but also significant economic, medical and agricultural value and the potential for other functions not yet known (Wilson, 1998, 2001), including acting as a fundamental part of the Earth's life support system through supporting fresh water, soil, and clean air (Natural History Museum Biodiversity webpage, 2014). It is therefore necessary for the survival of the planet (Natural History Museum Biodiversity webpage, 2014). The knowledge of biological entities gained through taxonomy is also vital for the biological sciences, particularly with the development of genetics and genomics, as the knowledge of evolutionary history and taxonomic status can be vital to research endeavours. Common identification of parts is important for universal knowledge and communication, therefore, biological taxonomy forms the nomenclatural foundation for work and research in the biological sciences.

Since the signing of The 1992 Convention on Biological Diversity at the Rio Earth Summit, biological diversity has become a political benchmark for both conservation and research (United Nations, 1992; Hine, 2008; Mace, 2004). One of the major articles of the Rio Biodiversity Convention includes statements that the extent of biological diversity must be surveyed and documented, and any "taxonomic impediment" such as a lack of taxonomic knowledge or expertise to obtaining full knowledge of biodiversity be removed (Hine, 2008). However, one year prior to the Rio Earth Summit and the signing of the Convention, a 1991 Report by the UK House of Lords Select Committee on Science and Technology revealed that taxonomy and systematics, as the branches of biological science responsible for surveying, describing and documenting biological diversity, was underfunded and lacked resources in relation to its strategic importance and mission to describe the variation of life on Earth (Select Committee on Science and Technology, 1991). Taxonomy, although one of the oldest and most fundamental branches of the biological sciences, was suffering from a lack of scientific prestige, being considered "old-fashioned, out of date" or mere "stamp collecting" (Select Committee on Science and Technology, 1991, 2002a, 2002b; Godfray 2002b, 2002c; Hine, 2008). This poor image of the taxonomic discipline may result from the fact that taxonomic information was not easy to find or understand, particularly for non-taxonomic end-users such as ecologists or agriculture and forestry scientists who would need or want to identify taxa. Furthermore, access to information in taxonomy was difficult, as information in taxonomy was often difficult to obtain, being fragmented, distributed and held in specialist institutions where few could readily access it (Godfray, 2002b, 2002c).

2.2 Research and Information in Biodiversity

Historically, the study of biological diversity typically involved an individual researcher or small team of researchers, often based in an institution such as a natural history museum or herbaria,

collecting specimens, comparing these specimens to known "type" or reference specimens, and consulting the taxonomic literature in order to determine relatedness to known taxa (Allnott, 2012; Godfray, 2002b; Hine, 2008). The formalised rules for establishing taxa were not established until the 19th Century, when to avoid an earlier informational crisis resulting from the unregulated proliferation of taxonomic names, taxonomists invented a complex set of rules that determine how a species is named and related to a type specimen, how genera and higher taxonomic categories such as orders should be treated and how conflicts over the application of names should be resolved (Godfray, 2002b). These rules also explicitly governed the publication of names in scientific monographs and journals, codifying biological nomenclature. The "type" or holotype specimen is significant in that it is a physical specimen nominated to be the representative to which a taxonomic name is formally attached. This original species name stays with the type specimen, even if the taxonomic status changes. Specimens, particularly type specimens, were considered to be the primary sources in taxonomy, but could only be useful if the informational context of the specimen were preserved, such as specimen labels, museum collection identifiers and collection or specimen data (Allnott, 2012; Hine, 2008).

2.3 Informational Crisis in Biodiversity

The sources of information for the study of biodiversity, whether specimens contained in museum or herbarium collections, or taxonomic literature, were often disparate, and researchers would often have to visit collections, libraries and archives in many institutions. This could involve travelling to other countries at great financial cost and time taken in order to locate and access the information sought. Furthermore, information in biological taxonomy is fragmentary and heterogeneous, consisting of many different information types. These range from preserved voucher specimens, to descriptions of morphology, historical artwork, photographs or other images, geospatial information such as population ranges, correspondence, collectors' field notes, museum specimen labels, and increasingly, DNA sequence data. Taxonomic information has exceptional longevity (Thresher, 2012) and remains useful to researchers looking for the first instance of a taxonomic description (Thresher, 2012). This can present a challenge, as taxonomic literature spans many centuries. The domain of biodiversity literature has been estimated to include over 5.4 million books dating back to 1469 and over 40,000 journal titles. Legacy taxonomic literature in the form of historic manuscripts, natural histories or field notebooks is often of vital importance for holding the description and definition of the representative, the "type" or holotype, specimen upon which a taxon is based. Taxonomic literature is of crucial importance, particularly for groups of organisms for which the type specimen has been lost or accidentally destroyed (Hine, 2008). More recently, with the rise of molecular taxonomy, biogeography and phylogenetic systematics, the taxonomic literature may also now include genetic sequence data, geospatial or distribution data and phylogenetic or phylogenomic trees (Hine, 2008; Sarkar, 2007, 2009).

Information management in biodiversity was piecemeal, and searching for relevant information amongst fragmented, disparate sources took researchers considerable effort and time, particularly before these resources began to become available on-line (Hine, 2008). The lack of any unified source for biodiversity information was seen as hampering progress in the field, occasioning an informational crisis.

2.4 Towards a Web-Based Cyberscience

Outside forces soon placed pressure on the field of taxonomy, pushing it towards becoming an information science. In 2002, the UK House of Lords Select Committee on Science and Technology issued a report strongly recommending that taxonomy and systematics become a web-based science (2002a, 2002b). This report explicitly stated that "species descriptions and interactive identification keys should be published on the web, and increased funding should be made available to digitise collections of specimens and make taxonomic information available to people all over the world" (Select Committee on Science and Technology, 2002a, 2002b; Hine, 2008). Surprisingly, as Christine Hine points out in her investigation of the use and uptake of information and communication technology in systematics (2008), the call to make systematics a web-based science did not come from within the discipline at all, rather, the pressure to "modernize" one of the oldest branches of the biological sciences came from international government, as knowledge of biological diversity was becoming an increasingly important benchmark for environmental conservation.

Concurrently, Oxford-based population biologist Charles Godfray published a letter in *Nature* calling for an open access, web-based taxonomic forum moderated through the internet through which a consensus or unitary taxonomy could be achieved (Godfray, 2002b, 2002c). While there is still considerable debate as to whether consensus taxonomy can be achieved (Scoble, 2004; Raven 2004), Godfray proposed that taxonomy and systematics go on-line. He stated:

"One of the astonishing things about being a scientist at this particular time in history is the vast amount of information that is available, essentially free, via one's desktop computer. I can download the sequences of millions of genes, the positions of countless stars. Yet, with a few wonderful exceptions, the quantity of taxonomic information available on the web is pitiful, and what is present (typically simple lists) is of little use to non-taxonomists. But surely taxonomy is made for the web: it is an information-rich subject, often requiring copious illustrations. At present, the output of much taxonomy is expensive printed monographs, or papers in low-circulation journals available only in specialized libraries. These are not attractive 'deliverables' for major-research funders." (Godfray, 2002b).

2.5 Biodiversity on the Web

Godfray originally proposed a moderated "web revision" for taxonomy, which would include a web page bearing the traditional description of each taxon and the location of type material, in addition to supplementary material not currently required for a formal taxonomic description, such as identification keys, images such as photographs or illustrations and even genetic sequence data. The web revision would also include a treatment of any known synonyms derived from the taxonomic literature, therefore maintaining ties with historical legacy literature. The web page for each taxon would be mounted on the web for further revision and comment, and would become the reference unitary taxonomy for the group. This web revision would be freely accessible on-line, and would become a first point of reference for taxonomic enquires and information seeking. Godfray stated: "the only way to organise a unitary taxonomy and make it widely available is on the web…this would replace the distributed, printed taxonomy" (Godfray, 2002b). Furthermore, Godfray added that taxonomy and

systematics should not only include nomenclature and evolutionary relationships, but also should "become an information science that curates our accumulated knowledge of that species in the way a gene annotation in a genome database organises our knowledge of a particular protein". Godfray has observed that the collections and type material that underpin distributed taxonomy requires administration, and that museums and herbaria undertake this task whilst embracing the new technologies of the internet. "Hosting web revisions is something I see as becoming…modern information storehouses" (2002b). Godfray predicted that taxonomy should evolve to become more encompassing and web-based, rather than distributed, an informational and digital "active discipline at the heart of modern biology" rather than side-lined as "distributed, a sad waste of over 250 years of knowledge learned in the course of study of the world's biota". The 2002 Report of the UK House of Lords Select Committee and the statements of Charles Godfray in *Nature* were a call to arms to push taxonomy into the informational realm of the World Wide Web, heralding the emergence of "cybertaxonomy" and biodiversity informatics (Godfray, 2002b; Select Committee on Science and Technology, 2002a, 2002b).

With the advent of the World Wide Web, some taxonomic information was available on the internet, but this information was still difficult to access, being primarily taxonomic lists of species in a particular geographic region geared towards an audience of taxonomists working on particular groups, or lists of type specimens held in natural history collections. Although distributed computing networks had been used by museums since the 1970s or 1980s to store collections information in databases, this information was also not readily accessible. The information held by museums about collections represented a great wealth of information which could be used to investigate a broad range of topics in evolution and ecology (Graham, 2004). Natural history collections such as those housed in natural history museums and herbaria represent over 2.5 billion specimens, each with an associated recorded collection event representing a temporal and spatial point (Graham, 2004), so information from museum collections can be used not only uncovering gaps in current collections, thereby aiding curation and directing collections policies, but also representing a great wealth of untapped research, particularly in the domain of biodiversity studies. Specimens contained in natural history collections are "vouchers" that enable verification of identification and the potential updating of taxon identity as the nomenclature is revised. Taxa are confirmed when compared with documented representative specimens, or type specimens. Type specimens are those presenting the characteristics for which the taxon, whether species or genus was named, and any published documentation will confirm these specimens as representative. While, the development of electronic catalogues for natural history collection holdings began during the 1970s, access to this information for the purposes of biodiversity research remained fundamentally unchanged until the advent of distributed networking and the World Wide Web (Graham, 2004).

The number of projects gathering data in the domain of biodiversity has increased (Berendsohn et al., 2011), facilitated by the World Wide Web and its capabilities for fostering collaboration and data sharing through increased communications and networking, which has greatly benefitted scientific research more generally. However, with the rise of biodiversity research shared and communicated on-line, collections must become digitally accessible and conform to a need for uniform data standards and

interoperability of databases for further exchange of data. This changing and expanding digital information landscape has had a strong impact on biodiversity study, with the consequent drive towards a consensus "cybertaxonomy", literally taxonomy on the web (Hine, 2008; Sarkar, 2009), with the development of platforms for the capture and storage of biodiversity information such as uBio (http://www.ubio.org/), the GBIF consortium (http://www.gbif.org/) and data standards such as those developed by the Taxonomic Databases Working Group (TDWG, http://www.tdwg.org/) and initiatives such as Scratchpads for taxonomic data curation (http://scratchpads.eu/), the EDIT platform for cybertaxonomy and biowikifarm (http://biowikifarm.net/meta/). Citizen Science initiatives such as Bio-Blitzes and local taxonomic surveys are also important sources of data and information for biodiversity research (Hine, 2008; Sarkar, 2009; Hardisty et al., 2012; Smith, 2009a, 2012, 2013).

2.6 Biodiversity Informatics and Cybertaxonomy

As biodiversity information became digital, the discipline of cybertaxonomy (Hine, 2008; Sarkar, 2009; Smith, 2009, 2013; Hardisty et al., 2012) emerged in an attempt to standardize, organise and manage web-based biodiversity information resources, platforms, projects and tools. Cybertaxonomy, defined literally as "taxonomy on the web", or "the use of computers and the World Wide Web in a program of taxonomic research" is part of a greater movement of biodiversity informatics, a "new suite of methodologies and tools that extends contemporary computer science and informatics principles within biodiversity data" (Sarkar, 2009). It differs from bioinformatics, in that rather than being concerned only with data analysis in molecular biology, biodiversity informatics "requires frameworks and approaches that can accommodate the full range of biological information from molecules to morphological features – populations, habitats – collectively developing the ultimate computational web of knowledge about life on Earth" (Sarkar, 2009). Biodiversity informatics therefore works towards an infrastructure upon which to organise, manage and disseminate biodiversity information.

The term "biodiversity informatics" was first used to describe the application and use of informatics techniques to biodiversity information, typically taxonomic, ecological and biogeographic information stored in digital form, to discover new ways of using this information to view and analyse biological diversity information (Sarkar, 2009; Smith, 2009, 2013, Hardisty et al., 2012). It involves a suite of informatics tools to "extend computer science and informatics principles within biodiversity data". While the term was initially coined in 1992, the field has emerged and burgeoned. In 2000, the journal Science published a special issue devoted to the "Bioinformatics for Biodiversity", and later, in 2004, a journal devoted to the field, Biodiversity Informatics, was published. There is now a Masters course at Imperial College London. Biodiversity informatics requires a framework and approaches that can accommodate the full range of biological information, from molecules to populations, habitats and ecosystems, using computational approaches to manage complex data. This field has evolved substantially as technologies have developed and progressed. Cybertaxonomy is a subset within biodiversity informatics, concerned with making taxonomic content accessible on-line and also with creating a unitary or consensus taxonomy. Likewise, biodiversity informatics is concerned with a global list of taxonomic names and is working towards this through the Encyclopedia of Life and the Barcode of Life international initiatives.

2.7 Data Portals and the Centralisation of Biodiversity Information on the Web

The ultimate aim of biodiversity informatics and cybertaxonomy is the organisation, management, and dissemination of biodiversity information to further knowledge of the diversity of life on Earth. Technological advances in information technology, concurrent developments and advancements in museum information management and in academic libraries have enabled the facilitation of these goals, while also mirroring the evolution of taxonomy and systematics into information sciences. Not only is there a challenge in making all of the collections information accessible on-line, there is also the challenge of standardisation of taxonomic information over a distributed network (Smith, 2009, 2013; Hardisty et al., 2012).

With the increasing availability and accessibility of biodiversity information on-line, there is also a concurrent drive to "standardise" or "unify" the plethora of taxonomic classifications to ultimately arrive at unified or consensus taxonomy. While some dispute exists as to whether a consensus taxonomy for all life may or may not be readily achieved (Scoble, 2004; Raven, 2004, Hardisty et al., 2012), numerous web-based projects such as the Universal Biological Names (www.ubio.org), the Encyclopedia of Life (http://eol.org/), the Tree of Life Web Project (www.tolweb.org) and the Global Biodiversity Information Facility (www.gbif.org) have worked towards creating on-line, single-search portals for taxonomic information. These web-based biodiversity information portals are a radical departure from previous ways of conducting research in biodiversity, where a great deal of researchers' time is consumed tracking down historical or legacy literature and specimens in disparate museums, libraries and archives (Hine, 2008; Sarkar, 2009; Hardisty et al., 2012). Furthermore, many of these sites offer a single-search interface for efficiently finding all indexed information related to the topic of search, thereby acting as aggregators for biodiversity information. This could save time for researchers looking for specific information. These data portals for biodiversity research have seemingly evolved from information and search strategies used on web-scale search engines, simplifying search for users. The major data portals currently in use for biodiversity research are described and summarised below.

The Universal Biological Names Index (UBio), began as an initiative within the science library community to create a comprehensive and collaborative catalogue of the names of all extant and extinct organisms. UBio based its foundation on the Taxonomic Names Server (TNS) which catalogued both names and classifications so that researchers could search for information on any organism using any of the names that might be related to that organism, such as any synonymies or superseded names. UBio recognised the importance of the taxonomic name and the linking of information to that name, which might potentially create problems in information search and retrieval, particularly for a taxon with many names, or a name referring to many taxa. Thus, the Taxonomic Names Server acted as a "names thesaurus". In addition to a list of names (held in the Name Bank), there was also a "Classification Bank" that stored multiple classifications and extended the functionality of the Name Bank. The Name Bank thus promoted "the emergence of a layered biological informatics infrastructure that allows different expert systems to share common information", supporting biodiversity research (UBio).

The Tree of Life Web Project (http://tolweb.org/tree/) was another large collaborative web-based project in which biodiversity information was linked hierarchically, reflecting evolutionary

relationships and essentially recreating a consensus taxonomy of life, based on genetic relationships. Each organism had its own webpage, which contains information about its characteristics evolutionary history, and phylogeny, portrayed through text, image and any other information. Contributors can upload information and content using customized web authoring tools which would upload a tree structure, images and other materials to the webpages. This information was then stored in a series of databases, and web pages can be created dynamically. Users can then browse the information contained in the Tree of Life hierarchically, following the branching nodes of the phylogeny. All information stored in the Tree of Life databases, such as taxon names, were treated as objects attached to these phylogenetic nodes or branching points in the phylogenetic hierarchy. Although the idea of organisation is sound, phylogenetic hypotheses and taxonomic nomenclature can change; they are subject to revisions. This inherent flexibility and "fluidity" of classification makes management of data in biodiversity difficult, so the "phylogenetic" information architecture can change to accommodate alternative nomenclatures and phylogenetic arrangements, as web pages are generated dynamically from information stored in the Tree of Life databases. Thus, the Tree of Life web project "facilitates retrieval of phylogenetically-structured data for display of biological diversity in an appropriately synthetic evolutionary framework" (http://tolweb.org/tree/).

Like the Universal Biological Names Index and Tree of Life Web Project, the Encyclopedia of Life (http://eol.org/) is a collaborative web-based portal providing single-search access to over 1.9 million webpages of images, text and databases, launched in 2008 with the vision of aggregating and providing open public access to biodiversity information. "Our knowledge of the many life-forms on Earth - of animals, plants, fungi, protists and bacteria - is scattered around the world in books, journals, databases, websites, specimen collections, and in the minds of people everywhere. Imagine what it would mean if this information could be gathered together and made available to everyone – anywhere – at a moment's notice". The Encyclopedia of Life therefore attempts to aggregate webpages and databases from multiple sources, including existing databases and contributions from both professional biologists and citizen scientists alike. In January 2014, the Encyclopedia of Life launched TraitBank, an open access repository for taxonomic datasets, such as morphological measurements, interactions and other facts that is searchable. The Encyclopedia of Life is literally a web-based encyclopedia, with a page on each species, with the initiative dependent on linking content from already existing taxonomic databases and datasets such as FishBase, a taxonomic encyclopedia of fish diversity. The Encyclopedia of Life also includes indexed information from the Biodiversity Heritage Library (BHL) an international consortium of institutions including the Natural History Museum, of digitized legacy literature important to taxonomy which is no longer in copyright, or which the BHL has received special permission to scan and distribute electronically.

The Global Biodiversity Information Facility (http://www.gbif.org/) plays a significant role in biodiversity informatics, being one of the largest aggregator of biodiversity data, aiming to provide open access biodiversity data to researchers all over the world. GBIF was founded in 1999 and officially established in 2001 following a meeting of the Megascience Forum of the Organization for Economic Cooperation and Development (OECD). The OECD panel's recommendations in their report stated that "an international mechanism is needed to make biodiversity data and information accessible worldwide.

Such a mechanism would produce many economic and social benefits, enabling sustainable development through provision of sound scientific information". Currently, the GBIF data portal and web services provide a single point of access to more than 500 million records shared freely by hundreds of institutions worldwide, with evidence relating to more than 1.5 million species "collected over three centuries of natural history exploration…including the current observations of citizen scientists, researchers and automated monitoring programs". GBIF functions through its participating institutions, which publish data according to defined standards. The use of standards enables data to be shared and used, facilitating research and informing policy decisions, thereby furthering GBIF's vision of "A World in which biodiversity information is freely and universally available for science, society and a sustainable future".

Although there are numerous web-based projects and portals to aggregate biodiversity information on-line, one of the major challenges is to maintain data standards. The Biodiversity Informatics Standards, formerly the Taxonomic Databases Working Group (TDWG) now sets and maintains the standards for biodiversity data, including metadata standards such as Darwin Core XML (http://www.tdwg.org/).

2.8 The Informational Role of Museums and Libraries

The uptake of the new technological medium of the internet by a branch of the biological sciences also echoes trends in the management of information in museums, libraries and archives. These "heritage" institutions are considered to be "memory repositories" which have been profoundly influenced by digital media. Where formerly, they could be considered repositories for material culture (whether this material culture be a man-made object such as a work of art or an artefact, or a specimen from a natural history collection), museums have increasingly become both creators and repositories for digital information. Andrew Roberts, in a paper outlining the changing role of information professionals in museums stresses increasing responses and trends towards the informational, rather than the material, in museum collections (2013). These trends evolved from government and public perception of the importance of an information society, with an increased emphasis on access, combined with the widespread availability and proliferation of information technology and systems and the development of external networks, particularly in the early 2000s. Information technology is now pervasive and ubiquitous across museums and libraries. A full account of the history of information technology in museums can be read in Williams (2013), but will not be outlined here. There is an increasing move in museums from a focus on inventories to information, including the development and delivery of access (Roberts, 2013).

This move towards the informational in museums also heralds a convergence in the interests of museums and libraries in providing increased access to information and information management, not only through making it available on the World Wide Web, but also through the use of content and syntax standards and an emphasis on supporting the use of networks and information for both staff and users (Roberts, 2013). These statements are echoed by Orne and Pettit (2013), who state: "if a museum is to make productive and profitable use of information, it needs…to understand itself as a community of users of information, to recognise the stakeholders in information and to provide them with the

means of negotiating over the use of information (Orne and Petitt, 2013). Museums have embraced the digital through the development of numerous virtual museums, on-line databases or portals which further both access and management of information. As Heath stated, the "idea of a place-based research collection is behind us" (Heath and Sul, 2012) and what is needed are discipline-based electronic repositories capable of supporting several different layers of collaboration, including peer-reviewed reporting of research findings (Heath and Sul, 2012).

2.9 Towards the Virtual Library Environment

Libraries, like museums, have also embraced the digital, and clearly the rapid change in the informational landscape with the rise of technology and the "explosion" (Connaway and Dickey, 2010) and proliferation of electronic resources such as electronic publishing, mass digitisation projects and the internet and World Wide Web have wrought profound change and brought with it challenges for information management in academic and specialist research libraries. These changes have been seen primarily as "disruptive", heralding a shift from physical space and collections provided by the library to the library as a virtual digital environment. Increasingly, libraries are opting for electronic and digital journals and e-books in place of print materials, both to save space taken up by storage of physical collections and also to provide increased access to researchers and students who may often not use the physical space that the library provides (Kennedy, 2014; Dempsey, 2006;). Libraries are now spending up to 85% of their materials budget on e-resources (Kennedy, 2010) but whether these materials are readily "findable" by library patrons may be of some debate (Brophy and Bawden, 2005; Connaway and Dickey, 2010; Fast and Campbell, 2005).

2.10 Search Interfaces and the Virtual Library

The on-line tools and electronic resources for finding scholarly information in a library's holdings, such as A-Z e-journal lists, federated search tools, specialised indices, bibliographic databases and the library's on-line public access catalogue (OPAC) may cause confusion in library users, as there are many points of access and multiple interfaces to search for information. Having to search in multiple electronic locations for a reference in response to their information need is not efficient, even for experienced researchers. The internet and World Wide Web have facilitated ready access to a vast volume of information, and web-based search engines such as Google have made searching for information on the web faster and more efficient, returning potentially huge amounts of results for a single search query. Google's single search box and full-text searching of content on public web pages has made information research and retrieval very simple and efficient on the world wide web. Furthermore, Google Scholar provides ready access to scholarly publications and has indexed full-text publications from many of the major academic publishers, including Open Access scholarly materials and resources. It is estimated to contain over 160 million documents as of May 2014 (Brophy and Bawden, 2005; Fast and Campbell, 2005). Google has therefore become the search engine of choice for many users searching the web, to such a great extent that "to Google" has become a verb meaning to search for information on the internet (Brophy and Bawden, 2005). User studies from academic and higher-education libraries have shown that students at many levels prefer the ease and speed of search provided by Google and other search engines to the more "sophisticated, but time-consuming" search

provided by the library, with separate searches of the on-line catalogue and databases of journals. The idea that library search is more sophisticated is contentious; however, as the search system ultimately relies on the quality of the metadata used to index the library materials, this may vary greatly. Furthermore, not all searches of library content would provide access to full-text materials, and this was found to be a great source of frustration for researchers in studies by the Research Information Network (2006, 2007).

In a comparison between Google searches and library system searches, Brophy and Bawden (2005) determined that Google is "superior for coverage and accessibility", while library systems are superior for their quality of results (Brophy and Bawden, 2005). Google and Google Scholar, which provide access to Open Access academic material, and other similar search engines such as Bing, have made searching for digital information less time-consuming and have increased efficiency, but have also brought with them the increased expectation and perception that information can be searched for and retrieved without effort, and also that all of the information is readily available "on the internet" (Brophy and Bawden, 2005). Users are often dismissing the formal search services provided by the library and its e-resources databases in favour of this "quick fix" in meeting their information needs, following Zipf's principle of least effort, or "satisficing" in order to save time and energy (Brophy and Bawden, 2005; Dervin et al., 2007), even though library databases may hold information of a higher informational quality. While both library databases and Google returned relevant results, users favoured accessibility and ease of use when choosing a starting point for their information-seeking activities. Roger Schonfeld (2014) writes that the "vision of the library as a starting point for research" is often in conflict with the actual practices of library users such as students and faculty, in results from a 2013 survey of academic libraries. Information scientist Lorcan Dempsey (2012) also stated that "discovery happens elsewhere" and that libraries would have to adopt a more "inside-out approach" in revealing their institutional assets, whether this involves facilitating improved access to the collections, though improved metadata which could be indexed by external search engines such as Google or Google Scholar, or highlighting the expertise of library staff. Most importantly, Schonfeld (2014)highlights the importance of the introduction of "various systems designed to bring together as high a share as possible of the library's collections into a single-search interface" (Schonfeld, 2014).

2.11 Discovery Systems and Interfaces

Discovery systems are interfaces and systems which enable a library user to find and access materials in its collections, providing access to multiple types of materials independently of the library management platform or system involved. The discovery system interface is often presented as a single search box through which the user can search for and retrieve content, often with features such as relevancy ranking, or ordering of search results. It may also include facets that can be utilised to narrow and refine the results returned, such as publication date, contributors or formats of materials (library terms website). Discovery systems access centrally-indexed scholarly material and content to which libraries are subscribing and to which they are adding their own resources (Dempsey, 2012). They provide a unified view across resources such as local archival management systems, institutional repositories, and the catalogue component of the integrated library management system. Discovery systems also index data lying outside of the library's immediate catalogue, including publishers content

such as web-based content stored remotely, metadata for works in copyright or content held in other libraries. Because they utilise a central index and cover multiple sources of information, discovery systems offer a greater scope for search than a search of the library's catalogue. They also act as portals to information stored in databases. Discovery systems, by facilitating access to resources such as journal databases, provide a more efficient search than using a library catalogue and associated journal databases, which take time and numerous clicks to navigate, in addition to presenting varying user search interfaces (ref). They also provide full-text access through single-search interfaces. Discovery systems scan the institution's central index of all accessible resources, both internal and external, including metadata, and return a list of items that match the search terms. Because most discovery systems are proprietary and mediated by third parties, the algorithms used in search and relevance ranking are not available.

Starting in 2002 with the release of AquaBrowser by ProQuest, libraries, particularly in academia and higher education, have been implementing and using discovery systems to make content, particularly digital content such as e-resources and databases, easier to find and access, and are increasingly offering additional options for users such as social media tagging and exporting of references and citations to user-mediated reference-management software. Discovery systems are heralded as the "next generation library systems", and are comprehensive platforms for the "discovery" of information, rather than simply a new version of the library catalogue with web features. They have been defined as a "new, simpler way of searching…with a 'Google-style' interface and new and improved user features" (British Library, 2010). The major features of a discovery system include an intuitive, single-search entry point to the library's indexed resources, an expanded and consolidated scope of search with more sophisticated search technologies such as natural language and full-text searches, faceted navigation and relevancy-ranking of search results (Breeding, 2010). The implementation of these systems by libraries can be seen as a response to perceptions of user expectations of a more integrated, instantaneous library search experience (Dempsey, 2012; Levine-Clark et al, 2014). A study commissioned by UKSG in 2013 has revealed that "resource discovery systems are becoming a major element of the academic library landscape", with 77% of survey respondents having already implemented a discovery system, and 11% in the process of doing so at the time of the survey. Discovery systems have "fundamentally changed the way that users search for information" (Levine-Clark et al., 2014).

There are currently four major proprietary discovery systems in use at the time of writing: Summon by Serials Solutions, Primo by Ex Libris, EBSCO Discovery Service (EDS) and World Cat Local (OCLC). In addition to these proprietary systems, several Open Source discovery systems exist, including VuFind (URL) and BlackLight. These systems are often combined with access to a vendor's "big index" such as the Summon or EDS index, and although they can be envisioned as "layers" overlying a library system, they are often thought of as catalogue replacements in their own right. Although these systems each differ in their own way, they share the similar characteristic of enabling a search across multiple sources at once, mediated through a single search box. Single search through discovery systems is now almost ubiquitous across most academic and research libraries. A web-scale discovery system can efficiently and effectively connect users to electronic content. The basic goal of a discovery layer is to

enhance the catalog so it functions as a “Google-like interface that serves as a single-search box entry point for the entirety of a library’s collection” (Kennedy, 2014). A discovery system should therefore provide seamless access to the many electronic resources held by the library.

2.12 Discovery, Search and Research Impact in Academic Research Institutions

Although uptake of discovery systems has been rapid, and is escalating in academic and research libraries, and although discovery systems promise effective and efficient searches of the library’s collections, particularly e-resources, providing a single-search point of access, they may also bring certain issues or problems with them. Usability studies of discovery systems in academic and higher education libraries have shown that although discovery systems enable access to resources, users of these systems may have difficulty sorting through the overwhelming amount of search results that are returned, and therefore end up “satisficing” rather than evaluating every result returned (ref). Relevancy ranking of search results returned by the discovery layer is increasingly important and is of increasing concern. It is important for the best content to appear high up in the list of search results returned, particularly as studies have shown that users will become overwhelmed by the amount of results returned and are likely to settle on whatever resources appear first on the list. There may also be vendor bias inherent in discovery systems, as search algorithms and relevancy ranking factors are proprietary and confidential. Proprietary companies of discovery layers have no incentive to work with competitors, and since vendors provide both their own content and that of other publishers, it may be possible that they rank their content above that of a competitor (ref). Relevancy rankings should ideally be accurate, logical and have a clear search algorithm. Another issue affecting discoverability of materials, is that the metadata used to index content to a discovery system may be of insufficient quality, which will inhibit its “discoverability” by the discovery system. Finally, although users may be used to using a search engine for general web searches, they may not have the same mental concept of discovery when using a library-based system. A study by Preater (2012) found that users brought with them their prior experiences of the library OPAC when searching for materials suitable for university-level study, but found obstacles to discovery beyond the use of the single-search box (Preater, 2012).

Although several case studies have been conducted on implementation, usage and impact of discovery systems in the academic and higher education library sectors (refs), only one case study has been written about the implementation of a discovery system in a special library (in this case, a museum for Jewish Heritage), but there have to date been no studies of the use and impact of discovery systems on a special library, in this case, a library within a natural history museum in the context of biodiversity research. This case study proposes to fill this gap in knowledge.

Chapter 3: Methodology

3.1 Rationale and Research Framework: Building a Case Study

The objectives of the dissertation research, to investigate user attitudes and behaviours in seeking biodiversity information resources using both the library discovery system and the informatics data portal, to assess impact of the discovery system on search and research, and to document the transition towards a virtual library environment at the Natural History Museum will be accomplished as part of the overarching aim of building a case study of the implementation and use of the discovery system and data portal at the Natural History Museum , utilising a case analysis strategy, combined with impact evaluation utilising a mixed methods approach consisting of surveys and interviews.

The case study strategy facilitates in-depth investigation of user information-seeking behaviours within a strictly-bounded or defined context (Case, 2012; Pickard, 2007). This research approach places emphasis on the "context of what is being studied" (Case, 2012), including many of its key processes and interactions. The investigative, exploratory approach advocated by Yin (1994, 2011) will be utilised, with conclusions being drawn from the data as it is collected and analysed. Data will also be evaluated in light of what has been found in previous studies drawn from the background literature; however, given that this is the first (to date) case study of the implementation and impact evaluation of discovery systems in a museum library, if not museum-wide setting, it is also hoped that new information may come to light which will provide insight into how discovery systems impact upon search and research in a museum-setting.

Various types of evidence and various methods may be utilised to build a case study, ranging from document analysis to interviewing key participants in the process. Although the case study approach, as outlined by Pickard (2002, 2007) and Yin (1994, 2011, 2014) is somewhat "intuitive", involving emergent investigative methods to encompass a wide range of evidence types, it is suitable for investigating user information-seeking behaviour and system implementation and usage within the delimited context of biodiversity research at the Natural History Museum. Yin (2014) recommends six sources of evidence: documentation, archival records, interviews, direct observations, participant observation and physical artefacts (Yin, 2014, pp.103-130). In the course of this study, due to constraints on direct observation, as much of the research behaviour takes place outside of the library in the various laboratories and collections areas of the museum and is therefore not directly observable, and likewise, although the system itself may be considered an artefact, the algorithms and coding that make up this artefact are proprietary and therefore unavailable for use. Similarly, because the implementation of both the discovery system in the library and the Data Portal are both very new, no archival documentation exists at the time of writing. Therefore, in the course of this work, three sources of evidence will be utilised: documentation (presentations and publically available project planning documents), interviews and participant observation through surveys and interviews.

3.2 Limitations to the Case Study Approach

There are some important considerations to take into account when performing a case analysis study. Case, in his review of research on information seeking, needs and behaviour (2012), raises some important points regarding the case study as a research strategy, namely that it may not be possible to generalise upon case studies because findings may be difficult to establish and may not be reliable, as they are limited in the number of entities that are being investigated (Case, 2012). Secondly, the length and timing of the investigation may also introduce bias, as it will essentially be a "snapshot" at one point in time, rather than a series of observations and evidence-gathering sustained over a period of time (Case, 2012). While this investigation has been limited by time and the timing of events (in particular, the late launch of the Informatics Data Portal in December 2014), every effort has been made to obtain as wide a range of evidence possible ranging from user surveys and interviews to key informant interviews and document analysis (Chapter 4 for a documentary description of the Natural History Museum Library and Data Portal in Context). A diverse range of evidence and a holistic, process-oriented emphasis involving multiple times of observation may make the case study research strategy more rigorous, as stated by Case (2012) and also Yin (2011) and Stake (2008). The case study approach will be embedded in a grounded theory mode of investigation and analysis. As emphasised by Pickard (2007), echoing the statements of Charmaz (1995), "a grounded theory approach is solely focussed on discovery…the only focus is the process and the particular user group" (Pickard, 2007, pg. 157). Therefore, the data obtained from the research will be utilised to inform the development of the theoretical conclusions drawn from the information obtained during the research process as it unfolds.

3.3 Initial Courses of Investigation and Research Design

The case study approach is similar to the ethnographic study approach, as it will involve investigative work *in situ* within a specific context. Ethnographic studies have been conducted of communication and the uptake of information technology in taxonomy on the web by Hine (2009) and discovery in a higher education environment by Lanclos (2012); however, a complete, rigorous ethnographic study would require a longer time frame in order to establish the contextual framework for the dissertation research, which is currently outside the scope and timeframe of the current work. However, an ethnographic study might be a possibility for future research, particularly in order to investigate the temporal impact of the discovery system on search and research at the Natural History Museum. This possibility will be discussed further on.

Initially, a usability test of the Library discovery system was also considered; however, given the time constraints of the dissertation, the lack of suitable resources such as testing software, and the potential invasiveness and incursion on researcher time, this test was abandoned following the recommendation of Dr. Vincent Smith (*pers. comm.*) during an initial consultation as to the study design. Likewise, although a log of web activities would also have been a useful gauge of systems usage on-line, and would have provided an unbiased, nonintrusive record of actual use (Bollen and Luce, 2002), it would not have provided an assessment of impact (Thresher, 2012, pg. 9). Furthermore, it was also abandoned due to internal systems constraints and confidentiality at the Natural History Museum.

In this study, the context will be established through an analysis of strategic planning documents outlining the implementation of the new discovery system for the Library and Archives at the Natural History Museum, and include data from a "historical" user survey that was conducted prior to the implementation of the new discovery system. Insights from "key informants" will be obtained following the case study approach outlined by Yin (2011) and Case (2012). For comparison, an external point of reference from an academic library in higher education will be examined, as this institution also utilises Ex Libris Primo and Alma, the same discovery system implemented by the Natural History Museum. The case study research strategy will be complemented by impact evaluation in the form of user surveys and interviews, which will provide a source of both qualitative and quantitative information supplementing the information collected from the interviews with key informants following the case study approach. Two user surveys will be released, one for the users of the Library and Archives discovery system, and another for the beta test group of the NHM Informatics Data Portal, which will examine attitudes towards discovery and user information-seeking behaviours.

3.4 Ethical Considerations

Any study involving human behaviour and systems evaluations within an institutional setting poses ethical considerations. In undertaking the collection of data using both the on-line surveys and interviews, it was important to be certain that respondents and interview candidates were aware of, and fully understood the purpose of the dissertation research. A copy of the dissertation research proposal was distributed to all senior management at the Natural History Museum Library and to the head of Informatics, and written approval received for the work. Furthermore, and most importantly, it was ensured that the work being conducted for the study was transparent and explicit at all stages of the research process, and that the results would be made available to all stakeholders once the study concluded.

Participants were explicitly informed that their confidentiality and anonymity would be respected throughout the course of the dissertation work. They were also informed of the purpose of the dissertation research, both in the introductory statement in the on-line survey, and in the invitations to participate in the surveys. Although names were requested, these identifiers would not be presented as part of the final report. Participants were also informed that they could withdraw from the dissertation research at any time, although responses received prior to their withdrawal would be used in the final work. Contact details of the principal investigator were provided with all communication should the participants have any questions, comments or concerns regarding the dissertation research process.

The interviews conducted for the purposes of the dissertation work were of two types. The first type involved interviews with willing participants who had explicitly consented to an interview following the completion of the Library discovery system user survey. These interviews would be kept anonymous, with only a general job title being utilised. The participants were informed of this prior to the interviews and provided with a consent form and an information form prior to the interview taking place. The second type of interview involved interviews with key stakeholders from the Library and Archives who would not remain anonymous, as identifiers such as their name and job title would appear alongside any

information they provided for the purposes of the dissertation work. These interview participants also received a copy of the Information Sheet outlining the purposes of the dissertation research and a consent form, as their responses would not remain anonymous. Copies of these consent forms may be found in Appendix 3.

Chapter 4: The Natural History Museum Library and Data Portal in Context

4.1 Introduction

The collections held by the Natural History Museum Library are unsurpassed in their focus on biodiversity and earth sciences, and are a national resource for British systematic biology in particular. The collections form a unique resource for national and international scientific and historical research, and support and supplement the equally rich specimen collections of the Natural History Museum.

4.2 History and Development of the Natural History Museum Library

While a full treatment of the history of the Natural History Museum is beyond the scope of this work, a more detailed, comprehensive history may be found in Thackray and Press (2001).

The Natural History Museum Library and Archives were essentially "built from scratch" following the move of the natural history departments of the British Museum from Bloomsbury in London to South Kensington in 1880, forming the British Museum (Natural History). The first librarian was Bernhard Barham Woodward (1853-1930), who selected volumes for inclusion within the library. The library at that time consisted of a General Library, which held collections that served the needs of all of the scientific staff, and departmental libraries, such as Botany, which held specialist research materials. Woodward was initially aided in his selection of materials by Charles Davies Sherborn (1861-1942), who was compiling an index of all of the scientific names of all animals in use since 1758, and who listed many European monographs and journals containing descriptions of new genera and species.

The library collections grew rapidly and "organically". By 1900, the General library contained 19,395 volumes and 5569 sheet maps. The Departmental libraries had also grown, with Botany containing 14,980 volumes, Geology 9395 volumes, Mineralogy 6339 volumes and Zoology 17167 volumes. These volumes were indexed in a central card catalogue, which was kept up to date, and published in five volumes between 1903 and 1940. In 1975, the Museum unified its library staff into a Department of Library Services under the management of Maldwyn Jones Rowlands (1918-1995). At that point, the library had a staff of 42 and operated 6 reading rooms, which received over 8500 visitors per year.

Currently, the Library and Archives operates a single combined reading room, where visitors can consult materials from the General and Departmental Libraries. The Library also has an off-site store located in Wandsworth, South London. The library collections now comprise almost one million catalogued items, including journals, artworks, maps, monographs and historic manuscripts, including correspondence and field notebooks which are invaluable to biodiversity research. The collections had previously been indexed on a card catalogue system, with a system of bookboards to record loans to staff. Staff would complete a bookboard and file it in place of the monograph or serial part on the shelf. The card catalogue was eventually "digitised" and made into the first electronic catalogue in the mid-

1990s. This information was then migrated to a Sirsi Dynix library management system, starting first as Unicorn and then changing to Symphony in 2012.

4.3 Natural History Museum Library Collections

At the time of writing, the collections held by the Natural History Museum Library comprise some **(x)** number of items, with over one million items relating to the history and work of the Museum held by the Archives alone. A breakdown of the collections can be seen in Figure 1. There is a range of physical items in the collections, dating from 1469 to the present day. The majority of the collections (59%) are "modern collections", materials such as journals and monographs published after 1920. While born-digital materials comprise only 1% of the collections, this percentage is growing rapidly, as journal subscriptions are transferred to digital subscriptions only. Although the majority of the collections are still in print format, digitising the collections has also been a key goal of the Library's strategy, and they have been involved in numerous mass and singular projects over the last five years, including being a founding member of the Biodiversity Heritage Library. Over 3 million items have been digitised from the Natural History Museum collections, including archival and special collections materials such as Darwin's correspondence and library collection for the Darwin's Library project, and recently, the correspondence of evolutionary biologist Alfred Russell Wallace. The majority of the collections have been catalogued at least to collection level description, and are searchable on-line.

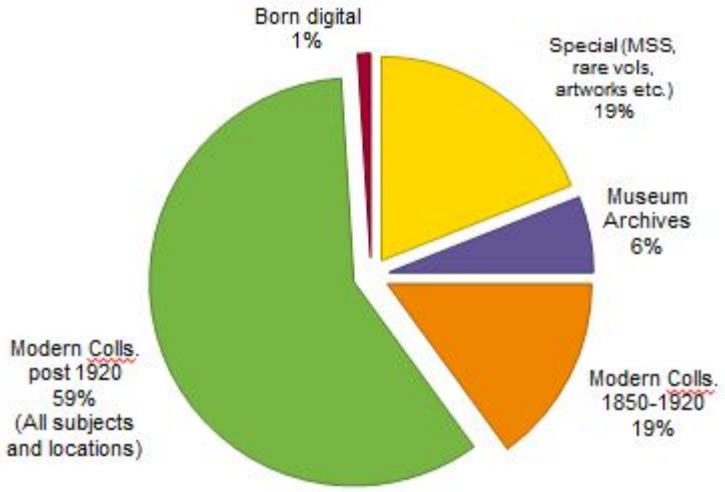


Fig. 1. Breakdown of the collections held by the Natural History Museum Library. Source: Jane Smith, Library & Archives, The Natural History Museum (2014).

4.4 The Virtual Library Project

In 2012, barcoding of the print monographs began, starting with the General and Zoology libraries. At the same time, the number of electronic journals began to rise, as numerous subscriptions were transferred to digital copies only.

In 2013, the Natural History Museum unveiled its Digital Strategy, which would involve the "exploitation of opportunities presented by digital change will be central to the Museum's work, enabling the delivery of world class experience for a global audience of visitors, learners and science peers and transforming science research and access to collections" (Digital Strategy Communication, Appendix 2) and would deliver a technological infrastructure that would support public engagement, scientific research, collaboration and effective collections management through new methods of storing and sharing information.

As part of the museum-wide Digital Strategy, the Library and Archives adopted an increasingly digital content for the collections and made a shift towards on-line delivery of information, through the Virtual Library Project. In July 2014, the electronic catalogue, Sirsi Dynix Symphony, was replaced by Ex Libris Alma and the discovery system, Primo, were implemented as part of the Virtual Library Project.

The aim of the Virtual Library project is to improve the discovery of resources "regardless of their current format or source" (Jane Smith, presentation). A survey of library users released in 2013 prior to the implementation of the new library system, Alma, with Primo Central for discovery, pointed to the overarching desire for single search to improve the efficiency and speed of search and the ability to process large volumes of data in order to support the research activities of the Natural History Museum. The Virtual Library Project, alongside the other digitisation projects such as the Biodiversity Heritage Library is one of the key projects required to deliver on the Museum's overarching strategy to become a virtual as well as a physical museum (Jane Smith, internal presentation, October 2014). The implementation of the discovery system "aims to transform the way our research users can access Library and Archives services and content". Prior to the implementation of the new discovery system, the library had been primarily focussed on managing and providing access to the physical collections, while the demand was increasingly for on-line resources, such as access to on-line journals or digitised material such as artworks. The born-digital and digitized materials were held in various databases requiring multiple user interfaces, which tended to silo the information in unconnected systems and workflows. For example, some holdings may be available as full-text on-line articles, but this was not obvious from a search of the library's on-line catalogue. Furthermore, it was observed that some users were abandoning the library in favour of local networks and more general search tools, particularly Google Scholar, which was seen as the "go-to research tool".

The Virtual Library Project (the implementation of a discovery system) would involve the following:

- Make the Library collections discoverable within a single interface
- Increase the visibility of digitised resources
- Reveal connections between collections
- Provide a system that can manage the digital space as effectively as the print.

The new discovery system would also link to the Biodiversity Heritage Library (BHL), which was not possible pre-implementation, although the museum library had contributed over 3 million pages. Pre-implementation, the BHL was only accessible via a separate website interface.

The BHL averages around 70,000 hits and downloads per month, which represents access to over 600 titles, and is a prime example of how exposing and promoting content on-line can widen access for a wider audience than the physical version alone could achieve. The discovery system will enable linking to other resources through open application program interfaces (APIs) and web services which will enable the re-use of metadata for indexing purposes. Jane Smith states in the overview presentation of the Virtual Library Project, "We are finding the more we expose our collections digitally, the more demand there is to view and research the physical collections".

Following a tendering process, in which four discovery systems (Ex Libris Primo, Innovative Interfaces, Sirsi Dynix and OCLC) were evaluated, Ex Libris Primo was chosen for the following reasons. A full breakdown of the scoring system can be observed in Fig. 2.:

- It provided the best solution for integrating BHL into the discovery layer.
- Primo would provide the ability to display archival records from the Archives catalogue, CALM. Although the hierarchy of the archives catalogue would not be retained, search facets could be configured to reflect the hierarchical structure of the catalogue holdings, thereby enabling search.
- Use of an Ex Libris system would enable the library to continue to purchase materials from other vendors, rather than being restricted to a particular vendor (such as EBSCO).
- Primo user interfaces are configurable to represent different user types. This will allow us to set up different environments for internal vs. external users, including graded access to eResources, different support and service information and distinct requesting functionality for each user group.

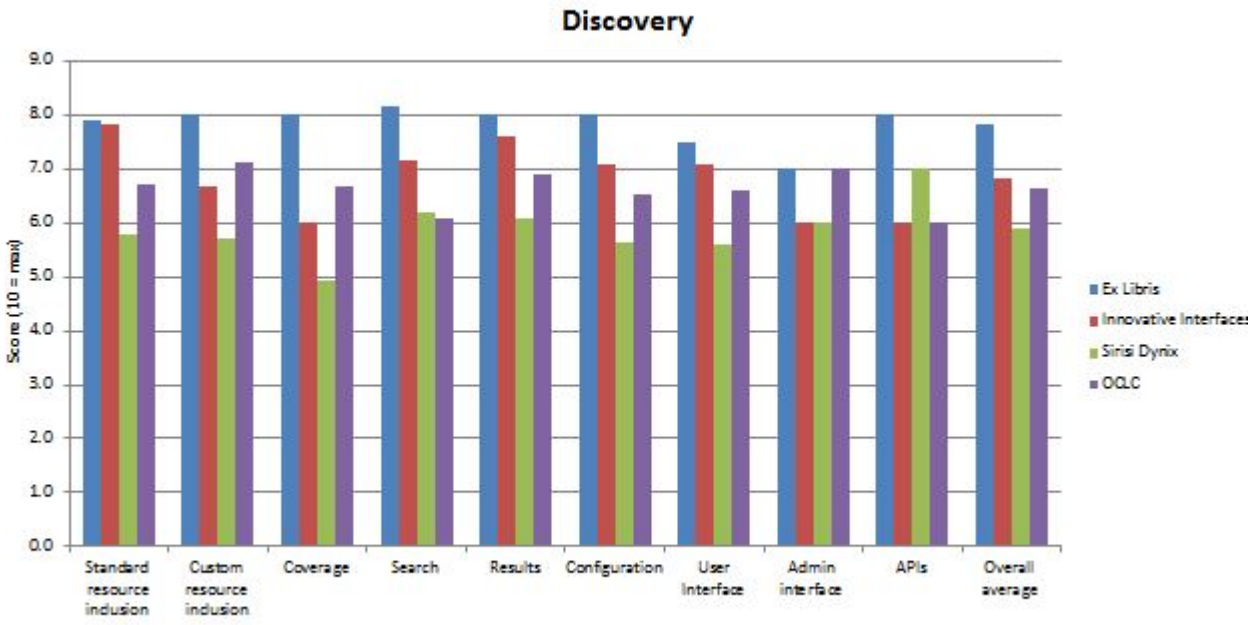


Fig. 2.: Comparison of the four candidate discovery systems with scoring system based on the following desired functionalities: Standard resource inclusion, custom resource inclusion, coverage, search, results, configuration, user interface, administrative interface, APIs and Overall Average Score. Ex Libris Primo consistently outperformed the other three systems in each area. Source: Jane Smith, Library & Archives, The Natural History Museum.

Following the tendering process, Ex Libris Primo was implemented following the following plan, with the successful "go-live" date of June 30th, 2014. The full plan is depicted in Fig. 3 (overleaf).

Fig. 3.: Plan showing the timeline of the implementation process of Ex Libris Alma and Primo at the Natural History Museum Library. Source: Jane Smith, Library & Archives, The Natural History Museum.

4.5 Initial Library and Archives User Survey: Historical Findings

Prior to the launch of the Virtual Library Project, a survey of user attitudes was carried out by the Digital Services Librarian, Sarah Vincent. A visualisation of the results of this survey can be seen in Fig. 4 (overleaf).

NATURAL HISTORY MUSEUM

Library and Archives User Survey Results

In 2013, we sent a feedback survey out to Library and Archives users. These graphics show a few of the things we discovered about our users, the way they use our services and how we can continue developing our collection to support research at the Museum in the future.

WHO ANSWERED THE SURVEY?

Science 82% Public Engagement Group 12% Corporate Services 6%

HOW DO YOU USE THE LIBRARY AND ARCHIVES?

65% RESPONDENTS USING ONLINE SERVICES AT LEAST ONCE A WEEK

- 57% use the library catalogue*
- 57% use Article Search*
- 53% read the new books list*
- 43% visit the Library and Archives intranet pages*

*at least once a week

35% RESPONDENTS USING IN-PERSON SERVICES AT LEAST ONCE A WEEK

- 55% use the reading room*
- 37% visit the enquiry desk*

*at least once a month

THESE ARE A FEW OF YOUR FAVOURITE THINGS*:

ENQUIRY DESK	98%
NEW BOOKS LIST	71%
READING ROOM	70%
LIBRARY CATALOGUE	65%
ARTICLE SEARCH	61%

*based on levels of reported satisfaction

POPULAR FIRST STOPS FOR LITERATURE SEARCHING:

*Includes subject-specific databases, the Archives Catalogue and Wikipedia

WHAT DO YOU THINK OF US?

WORDS USED TO DESCRIBE LIBRARY AND ARCHIVES...

WHAT ARE YOU...

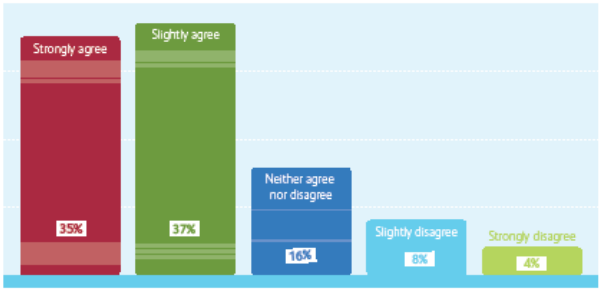
MISSING?

- 63% of staff have never emailed Archives
- 63% of people don't submit inter-library loan requests
- 82% of people don't regularly request off-site items
- But the majority of people who do use these services are very happy with them

IF FOUND, PLEASE CALL 020 7942 5460

ARE WE FULFILLING OUR MISSION STATEMENT?

"THE LIBRARY AND ARCHIVES DEPARTMENT IS SUPPORTING THE NATURAL HISTORY MUSEUM RESEARCH BY DELIVERING RELEVANT LITERATURE AND INFORMATION EFFECTIVELY"



WHAT DO YOU WANT NEXT?

TOP PREFERENCES FOR THE NEW LIBRARY AND ARCHIVES SYSTEM

More forgiving/fuzzy search 31% Single search across all resources 29% Other* 15% Clear and easily accessible content 11%
Search beyond NHM resources 9% Personalisation of alerts and services 5%

*Including easier requesting, more eResources and single sign on capabilities

SUGGESTIONS FOR FUTURE DEVELOPMENTS:

#1 eBooks #2 Literature searching #3 Liaison librarians

Comments or questions?
Email us at **library@nhm.ac.uk** or visit your nearest library office.

Fig. 4.: (Previous page) Visualisation of Data obtained from a Library User Survey initiated prior to the launch of the Virtual Library Project. Source: Sarah Vincent, Digital Services, Library & Archives, The Natural History Museum.

The survey respondents tended to favour on-line Library services, with 65% using the on-line services per week, as opposed to 35% who would visit the Library Reading Rooms. 57% of respondents would use the the Library OPAC to find information resources, however, in searching for literature, both the Library OPAC and Google Scholar were used to find articles, with 22% of respondents using Google Scholar, and 22% using the Library search. Other commonly searched resources included Web of Science (15%), independent journals (13%) and the Library's Article Search, which was used by 9% to find specific articles in the e-journal A-Z.

Findings from this survey further suggested that the majority of scientific researchers at the Natural History Museum wanted a unified single search to save time and provide increased efficiency for their research efforts (29% of respondents), a more forgiving "fuzzy" search (31%) and the capability to search beyond the Natural History Museum's collections (9%). These desired qualities were to be met with the implementation of Primo as a discovery system.

4.6 The Natural History Museum Informatics Data Portal

As part of the museum-wide Digital Strategy, the Informatics Group at the Natural History Museum would create and launch a data portal, as part of its commitment to open access and open science. The Informatics Group, like the Library, are responsible for information management at the Natural History Museum, except that their focus is primarily on the data generated by researchers working at the Natural History Museum. They have developed innovative tools for analysis of datasets produced by the Natural History Museum, and have pushed many of these tools into wider use, such as the Scratchpads for managing information on biodiversity and taxonomy (ref). Datasets produced by the museums 600+ curators and research staff would be made publically available. The Data Portal can be accessed at the following link: http://data.nhm.ac.uk/.

The Data Portal was built using the open source data publishing software platform CKAN, and presents the user with a single-search box through which datasets can be searched and accessed through a natural language, full-text search. Facets enabled browsing and filtering of the returned search results. Dataset can be searched by taxonomic name, geographical range, citation by literature and author. Datasets can be made available through direct download, or through the use of an application program interface (API).

Ben Scott, the lead architect of the Data Portal, provided an overview of one of the early versions of the Data Portal (Ben Scott, August 2014, *pers. comm*.). At the time of writing, the Data Portal consisted of a single search field which can look up any of the data contributed by researchers at the Natural History Museum. The datasets include collections data on specimens, pulling in 2.8 million records from the Museum's collections database, KE Emu. Much of this data was mapped to DarwinCore, a metadata format related to taxonomic nomenclature that makes the data discoverable by search engines. A geospatial interface also enabled mapping of datapoints onto a geographical map

so that specimen collecting locations and species ranges could be mapped, both spatially and temporally, pulling data from the historical collection dates associated with each record. The datasets will change research, as they not only enable discovery through the searching facilities of the Data Portal, but will act as a permanent data repository, which is particularly important for published datasets. A persistent Digital Object Identifier (DOI) is assigned to cited data.

In addition to its powerful search and discovery capabilities, the Data Portal offers the capability of uploading datasets, tagging and tracking changes and use, thereby enabling collaboration. Metrics are calculated for each dataset using a Datacite service, which tracks where and how the datasets are used, thereby gauging its impact in research.

At the time of writing, the Data Portal is currently still in its beta test stage, and as of December 15, 2014, has now been publically launched. To date, the Data Portal contains 3.5 million records, including 2,499,355 specimen records, and 8 datasets.

Chapter 5: Discovery System User Survey

5.1 Introduction

Two surveys were prepared for the dissertation research; the first was a survey to investigate search and research impact at the Library and Archives using the discovery system, Primo. The second was a survey to gauge the impact on search and research using the newly launched Data Portal. Owing to the late public launch of the Data Portal, and the fact that the users would have had insufficient experience of the live system to respond to a survey, the survey prepared for the Data Portal was abandoned, but it still exists for use in future research, therefore, only the Library and Archives survey is discussed in this chapter.

The Library and Archives user survey was released twice; the first time in combination with a second Library and Archives user survey authored by the Digital Services Librarian, Sarah Vincent, and the second time independently, to boost response rate.

Copies of the surveys and the complete results are available to view at the following link: https://www.surveymonkey.net/MySurveys.aspx?ut_source=header (Log-in using the username: nhmdigitallibrary and password: l6inqb1*dk)

5.2 On-Line Survey: Usage and Impact of the Library and Archives Discovery System

5.2.1 Preparation and Design of the Survey

The purpose of the survey was to investigate the issues surrounding the use and impact of the implementation of Primo, the discovery system, as it related to all three research objectives. An on-line survey was selected, as providing the most time- and cost-effective means with which to obtain responses from the whole Science department, currently estimated at 300 staff. The survey design was informed by Couper (2000) and Fink and Kosecoff (1998).

A web-based survey service, SurveyMonkey (www.surveymonkey.com) was used to prepare the on-line survey, as it was freely available on-line and enabled the design and creation of surveys with varying question formats which also provide the opportunity to receive responses from over 500 respondents per survey, which was more than sufficient for the purposes of the study.

Questions prepared for the survey were mainly of the multiple-choice variety or required the respondent to rank, score or order a set of responses. These types of questions have the advantage of being simple and quick to complete, and also facilitate quantitative analysis, as the responses can be scored or counted (Pickard, 2007, pp. 181-200). Although some open-ended, or written answer questions were included, these were minimal, as they might potentially discourage respondents as they require more effort and time to complete.

The initial draft of the User Survey consisted of 23 questions falling into three sections. The first section included questions to establish baseline data from the respondents; these asked for personal data such as their name, length of time working at the Natural History Museum and department, in

addition to the type of role that they held at the Natural History Museum. The second section included questions assessing usage of informational resources and services including the discovery system. The third section posed questions about the perceived impact of the discovery system on search and research in biodiversity. The survey utilised a mixture of closed questions to facilitate quantitative analysis, and open questions to enable descriptive responses. The order and language of the questions was carefully considered in order to make the survey unbiased, logical and clear to the study respondents. More straightforward questions were placed at the beginning of the questionnaire and descriptive, open questions requiring more thought or effort were located near the end of the questionnaire. A variety of question types was included to reduce respondent fatigue.

As a result of a meeting with the Head of Library, Head of Modern Collections and the Digital Services Librarian, the dissertation survey was combined with a survey authored by the Digital Services Librarian, the order of the questions was consequently altered to account for areas in which both surveys overlapped. The final draft of the survey included questions grouped roughly into the following sections:

1.) Baseline information, including an assessment of historical usage of the library and its resources and services: questions 1-7
2.) Current research practices: questions 8 and 9
3.) Attitudes to searching using the Library discovery system: questions 10-12
4.) Impacts on research practices: questions 13-16
5.) Impacts on search practices: questions 17-22

The first set of questions (questions 1-4) establishes a baseline of information about the respondents, such as the department in which they work and the major focus of their work, whether they are primarily researchers, or hold another position within the museum such as a curator or administrator. As the targeted respondents for this study are scientific researchers, these questions will ensure that only the targeted pool of respondents is selected for the study.

Questions 5 through 7 examine behaviours and attitudes towards the library and the services that it offers, including determining whether the respondents utilise primarily electronic resources or tend to prefer to visit the library in person.

Questions 8 and 9 examine research practices, such as whether researchers would initiate research utilising the library's search system as a starting point. This has often been a point of discussion in the higher education literature surrounding discovery, as Lorcan Dempsey states, "discovery may not always happen in the library"(Dempsey, 2012). While question 8 examines where research starts, by asking respondents to indicate and rank their preferred information sources, question 9 continues by asking which information sources respondents prefer to keep up to date with developments in their area of research.

Question 10 asks respondents to compare the new search system with the previous library system (Sirsi Dynix Symphony with associated e-journal databases linked through a separate webpage). This question was designed to assess attitudes towards utilising the new discovery system for searching

for resources within the library. Question 11 also determines library users' attitudes to the new search interface, but provides a more detailed breakdown of specific features of the discovery system, Primo, including the ability for the system to provide specific content alerts in response to saved searches, the ability to search for items external to the library's collections, and the ability to easily share resources by e-mail or by pushing them into reference management systems, or social media. Question 12 provides a very general assessment of levels of satisfaction regarding the library's ability to provide research support in response to users' information needs.

Questions 13 and 14 complement questions 1-7 in establishing the types of resources to which researchers most commonly resort for the purposes of their research. Thus, it establishes a pattern of research practices which will then inform search using the library's discovery system. Previously, print and electronic holdings could be searched only by using separate on-line search systems, the library's OPAC and the Article Search and e-journal A-Z, which provided a web page with a list of alphabetically-ordered links to each e-journal and database title.

Question 15 is an open-ended question soliciting the general opinion of the respondent about the overall impact of the new discovery system on their research. This question would be utilised to inform the interview questions should the respondent consent to the interview following the survey completion.

Questions 16-22 examine the impact of the discovery system on search and information-seeking behaviours. Question 16 investigates the awareness and knowledge of information sources that can be accessed through the discovery system. Question 17 encourages respondents to determine whether the discovery system's single search interface has generally changed or impacted on their search behaviours. This is followed in question 18 by a more detailed breakdown of search features including the usability of the interface for providing efficient, effective search. It is here that the discovery aspect of the interface is explored, as relevance and the linking of new information is examined. Questions 19 and 20 again look at the impact of the discovery system on search and information-seeking behaviour, this time with a focus on the results returned from a typical search query. Literature in the higher education sector **(refs)** implied that often too many search results were returned, which was potentially overwhelming for users, who might "satisfice" by selecting only the top ten results returned. Question 21 echoes question 11 in its focus on discovery and the serendipitous discovery of new information found through the discovery system's capacity to link information related to the search terms. Question 22 assesses the impact of the discovery system on the use of the print and digital collections, as the discovery system should potentially pinpoint information resources held in the collections that are relevant to the users' searches, and this may increase the use of the collections in these areas.

In a separate final question, respondents were asked if they would be willing to participate in a brief, anonymous one-on-one interview following the survey. If respondents agreed, then their e-mail address was collected and they were contacted to arrange the interview.

Deduplication would be accomplished through the use of the participants' name, which they supplied in the first question to the survey.

5.2.2 Testing the Survey

Before the survey was released to any participants, it was tested in two ways. The first stage of testing involved the automated survey test on SurveyMonkey, which populates the survey with random data in order to ascertain that the survey logic is sound. The second stage of the pilot testing involved e-mailing the survey to colleagues working in the Library and Archives at the Natural History Museum who would not be involved as participants in the final release of the survey. They were asked to send feedback on the time taken to complete the survey, any technical difficulties encountered when loading, viewing or inputting answers into the survey, and whether the language utilised for the instructions and questions was clear, flowed logically, was explicit and avoided bias through the presentation of leading statements. Eleven colleagues from the Library and Archives who would not be included in the study trialled the survey and reported feedback including changes to language and ordering of questions. The average time taken to complete the survey was 15 minutes.

5.2.3 Distribution and Launch of the On-Line Survey

Upon completion of the implementation of the new library system, Alma with its integrated discovery system Primo on July 1st 2014, the Library and Archives was interested in obtaining feedback on the use and user satisfaction with these new systems. Following a meeting with the Digital Services Librarian, the Head of Collections and the Head of Library, it was agreed that the dissertation user survey should be combined with a Library and Archives survey, in order to avoid participant fatigue in response to several surveys being sent separately. The combined survey was released on October 22, 2014 to all Natural History Museum staff members. Some modifications were made to the dissertation survey to avoid redundancy and duplication in the questions, as some of the questions overlapped in scope. A skip clause was inserted between the Library and Archives survey and the dissertation research survey, and an introductory statement was prepared to explicitly state the purposes of the dissertation survey. The survey remained open for two weeks, after which point, it was closed and responses collected, with a reminder e-mail sent out half-way through the survey period. In order to increase the response rate and improve statistical significance, the dissertation survey was released a second time independently in the hopes of boosting the response rate. The independent Library and Archives survey was released on November 10, 2014 via a link sent in an e-mail to members of the Science Group. The survey was then open and available for two weeks before being closed again and responses collected and analysed. De-duplication in both surveys was accomplished through the use of the participants' names, which they supplied in the first question, and also in the automatic settings of the SurveyMonkey survey tool, which would only accept one response per computer.

5.3 Detailed Results of the On-Line Surveys

The results obtained from both the combined and independent Library Users surveys were consistent with each other; however, some questions differed slightly in their answers. A full summary of the results of both the Combined and Independent surveys can be viewed on the SurveyMonkey site (log-in required).

Questions 1-3: Baseline Information

The total number of respondents who responded to both the combined and independent surveys was 42, with 20 respondents viewing the combined survey and 22 respondents viewing the independent survey, respectively. As the current estimated number of scientific staff is 300, including visiting and associate researchers, this represents a response rate of approximately 7%. Of these, the majority were researchers who had worked at the museum between 1-5 years (50% of respondents for the Independent survey), with the second largest group of respondents having worked at the Museum for over 16 years (23% in the Independent Survey), and were science staff, with the majority working in the Life Sciences Department (64% in the Independent Survey). The majority were researchers (67% in the Independent Survey) followed by curators (14%).

Questions 4-7: Behaviours and Attitudes to the Library and its Services

Findings from the Combined and Independent surveys revealed that 57% of respondents use the Library and Archives search weekly, with 30% visiting the library reading rooms weekly. The Library and Archives web pages were visited weekly by 39%, and 20% regularly used remote access to e-resources on a weekly basis. Surprisingly, 2% never used the Library and Archives search, and 33% never used remote access to e-resources.

Results from the Independent survey indicated that 65% of respondents used the Library and Archives on-line services weekly, with 35% visiting the Reading Rooms to consult the physical collections, which is consistent with the Library user survey released in 2013. Surprisingly, the independent survey also revealed that 10% of respondents did not use the Library at all for their research, and it is likely, but assumed, that these researchers primarily relied on external search engines such as Google Scholar or accessed journal databases directly without going through the Library and Archives search.

Questions 8-9: Research Practices

Results from both surveys pointed to external search engines or databases as the starting point for research, rather than the Library and Archives discovery system. Interestingly, Google Scholar ranked as the most important resource for search when starting a new research project for both surveys, and was also seen as a source through which researchers could keep up-to-date with current work in their fields.

In Question 8, both surveys ranked Google Scholar first (65%), followed by Web of Science (29%) and then the Library and Archives search (47%) as the first source to search for information when initiating a new research project.

In Question 9, both surveys ranked Google Scholar first (41%), followed by individual journals (29%) and Science Direct (21%).

Questions 10-12: Attitudes to Search

Question 10 asked how the discovery system compared to the previous system in terms of search. The majority of respondents (56% and 52%) regarded the discovery system favourably; however, there was some ambivalence expressed, as 19% and 16% felt that the performance was about the same as the previous Library search on the OPAC and separate e-journal databases.

Question 11 examined the various characteristics of the discovery system and asked respondents to rate how successful these aspects of the system were and whether they had an impact on research. Most respondents agreed that the search was successful in being more forgiving (37.5%), but 25% of respondents felt that they did not know whether this aspect of the discovery system was successful. Likewise, while 37.5% found the single-search across all library collections favourable, 13% did not know whether this aspect of the discovery system was successful. Similarly, the ability to search for resources beyond the Library and Archives collections was regarded favourably by 25% of respondents and 31% of respondents felt that the capability of the discovery system to provide easy and immediate access to digital content was very successful, with 31% of respondents agreeing.

A few features of the discovery system were mostly unknown to respondents. The ability to save searches and search results and set up alerts was a feature that was not known to most respondents (57%), likewise, the ability to share resources by e-mail or social media was not known by 69% of respondents and 56% of respondents did not know about the direct request feature of the discovery system, which enables immediate access to the article searched for. This question also revealed some ambivalence surrounding certain aspects of the discovery system, such as the accessible content such as the BHL was unknown by 25% of respondents.

Question 12 asked respondents whether the discovery system supported research at the Natural History Museum. While 38% of respondents agreed with this statement, 25% disagreed and another25% did not know.

Questions 13-14: Research Sources and Information Types Used

Question 13 compared the use of print vs. digital formats for biodiversity research. While digital content was selected by the majority of respondents as the major format of information used, print monographs were selected by 40% of respondents for use in print (as opposed to 47% who used both print and digital formats. Legacy books were utilised by 53% of respondents in digital form, which may indicate use of the Biodiversity Heritage Library digitised content, while 60% of respondents used both print and digital legacy journals (published before 1950). Interestingly, 47% of the respondents did not use archival materials in their research.

Question 14 examined what types of datasets informed biodiversity research. The majority of respondents (47%) utilised taxonomic data in both print and digital formats, while 36% used species distribution ranges in both print and digital formats. Surprisingly, while 42% of respondents reported using digital genetic datasets in the form of sequence data, 50% of respondents did not use this type of

data in their research. This may be unusual, given that taxonomic practices utilise molecular phylogenetic analysis (Hine, 2008).

Question 15: General Opinion on Research Impact

Question 15 solicited brief, general comments on the impact of the discovery system on research. Some of the comments are presented below. These comments range from positive, to ambivalent.

"Improved remote access when working from home."

"Easy to access needed resources.."

"No obvious changes."

"No negative impact."

"I still end up with a similar percentage of requests where I need the help of our colleagues from the library to arrange access."

"No obvious changes."

"Made external research more accessible."

"Made it easier to access on-line material in the same search as for print material – saves time."

"Minimal. All I can do is search fewer journals to download manuscripts. I have no idea how to use Science Direct, and I cannot find out how either."

"It has helped reduce the number of searches by combining the searched areas."

"Easy access to up-to-date info."

"It has made my life easier as I search on only one website."

"It has had much as it is a bit too comprehensive."

Questions 16-22: Impact of the Discovery System on Search and Information-Seeking Behaviours

Question 16 reported on the knowledge of resources accessed through the discovery system. This question uncovered a lack of knowledge of the resources linked through the discovery system. For example, 50% of respondents did not know about access to the BHL through the discovery system. Likewise, 50% did not know about the linked abstracts and unsubscribed full-text resources, and 92% did not know about any datasets linked through the new system.

Question 17 solicited a general response about the impact on search. Here, respondents felt that the discovery system did not have a perceivable impact, with only 21% agreeing that their

information –seeking behaviours had been changed through the use of the discovery system, while 58% felt that there was no change, and 21% were undecided.

Question 18 examined the use and characteristics of the discovery system with regard to its use in search. The majority of the respondents felt that the discovery system enabled quick, easy searches (43%), returned relevant results (42%), increased the efficiency of search (43%), was simple to navigate (35%) and had a clear, intuitive interface (57%). However, the "discovery" aspect of the discovery system produced some ambivalence, with 21% disagreeing and 21% unknown as to whether the system facilitated discovery of new information resources. Similarly, 36% did not know whether the system linked information resources in new ways. Access was also problematic, with 29% of respondents stating that they did not agree that the discovery system enabled access to resources.

Question 19 examined whether respondents found the number of search results overwhelming, however 50% of respondents felt that the number of search results returned was adequate, with only 25% feeling that too many results were returned, and 25% too few.

Question 20 examined the post-search filtering behaviours of respondents. While most (57%) reported that they would search through the whole list of results returned from a typical search, 35% would make do with only the first page, and 8% with the first ten results returned, therefore providing some evidence of satisficing.

Question 21 gauged the "discovery" aspect of the discovery system, and whether respondents felt it would return new or novel information resources. Only 23% felt that the discovery system would enable the discovery of new resources, while 77% felt that no new or novel resources had been found during the course of a typical search.

Question 22 asked respondents whether they had increased their use of the print or digital collections as a result of using the discovery system. The majority of respondents disagreed, with 46% and 42% disagreeing that the discovery system had increased their use of print and digital resources respectively. However, some respondents felt that their use of collections had increased as a result of using the discovery system, with 8% of respondents reporting an increase in their use of the print collections, and 14% of the digital resources.

5.4 Data Portal Survey

An on-line survey was created using SurveyMonkey in order to assess the impact and usage of the Data Portal, but owing to the late launch of the Data Portal within the time frame of the dissertation research, and the subsequent Christmas holidays, it was not deemed practical to expect a great response to the Data Portal survey, particularly as its impact and usage could not be fully ascertained so close to the launch date. The logic and questions utilised for the survey are similar to those of the Library User Survey; however, questions about impact had been removed on the advice of Dr. Vincent Smith. The Informatics Data Portal survey consisted of 12 questions, with the first half of the survey obtaining baseline information about the survey respondents, and the second half of the survey

assessing impact and usability. On the advice of Dr. Vincent Smith, the survey was largely abandoned, but the survey instrument exists and can be utilised at a later date for future research purposes.

5.5 Summary of Questionnaire Results

The surveys aimed to meet three of the research objectives posed by the dissertation work: to examine the impact of the discovery system on search and research for biodiversity information, to determine attitudes to searching for biodiversity information using the discovery system; and to explore the discovery and use of biodiversity information through the discovery system.

Although caution must be made in extrapolating general conclusions from the results of the surveys owing to the small sample size and potential for self-selecting bias in the respondents, the results from the surveys indicate that it may still be too early to gauge a specific impact of the discovery system on search and research, as many of the users are still in an "uptake" stage and learning how the system operates in conjunction with their own individual patterns of information-seeking behaviour. Mental models of the system and its operation may come into play here, and this may be a course for future research. A percentage of the respondents felt that there was some positive impact in that they were able to find and access resources quickly and efficiently from a single interface, but a similar percentage felt that there was no real impact. Furthermore, a similar percentage was undecided and did not know whether the discovery system had any impact on their searching behaviour and research.

While the greater part of the respondents were satisfied with the Library and Archives discovery system, there was some ambivalence and uncertainty around the features and capabilities that the discovery system offered that could aid in furthering research and search for information resources. Furthermore, many respondents reported that they utilised external search services such as Google Scholar, individual journals or Web of Science as the first point of search when starting a new research project or keeping up to date in their field of research. Interestingly, while most researchers found the discovery system efficient to use, they also reported a lack of awareness of resources that were linked through the new system, such as direct access to the Biodiversity Heritage Library or to specific datasets created by Natural History Museum staff as part of published work.

The surveys pointed out that it may still be too early on in the uptake of the discovery system to observe any changes in search behaviours. Respondents also reported that they would look through the entirety of search results rather than just "satisficing" and looking through the first ten or the first page of results returned from a search. The respondents provided positive feedback through the survey about various aspects of the discovery system, such as the single-search interface, the ability to save searches and set up alerts, and the ability to share resources, although it seems that more outreach may be required by Library and Archives staff to make researchers aware of the resources that are available through the discovery system and its features and capabilities. The ambivalence expressed by the users in the survey may point at the short time elapsed between the implementation and launch of the discovery system and users still learning how to navigate and use the system effectively for their research, so it may be too early yet to effectively gauge an impact on research at this point.

In addition to providing feedback about the information-seeking behaviours of the researchers, the survey provided some interesting feedback about what types of information are used in biodiversity research at the Natural History Museum. Legacy literature, both books and journals, are primarily used in their print form, whilst more modern articles (published after 1920) are primarily used in their on-line or digital form. This may also be indicative of two putative populations that were identified as being in the majority of the respondents to the survey: researchers who had been at the Museum between 1 and 5 years, and researchers who had been at the Museum for more than 16 years. While it may be assumed that researchers who had spent a longer period of time at the Natural History Museum may be more used to searching print resources, the majority of respondents revealed that they tended to use both print and electronic resources, particularly taxonomic resources for their research.

Clearly, the responses from the survey have revealed that the respondents are still in an uptake stage as they learn how to use the discovery system to find information resources to meet the demands of their research. The survey also points to some ambivalence surrounding the capabilities and facilities of the discovery system that might be met with increased outreach and information literacy training, which may assist in addressing any lack of understanding and orient the researchers to the discovery system.

Chapter 6: Interviews with Library Users and Key Informants

6.1 Library and Archives User Interviews

6.1.1 Question Production and Protocol

The main purpose of the interviews were to increase detail and clarify the responses provided by the participants in their survey answers, but the interviews also had the additional aim of providing a more in-depth look at how they used the discovery system to search for information, and also to encourage participants with the chance to elaborate on their opinions about the impact of the discovery system both for their own research and more generally.

Because of the time constraints of both the researcher and the interview participants, an asynchronous e-mail design was employed, in order to enable respondents to both answer at their convenience and in order to provide adequate time to reflect on their responses. Computer-mediated interviews has several advantages in that they minimize any interruption to participant work commitments and does not require them to be available on-site; however there are also several disadvantages in that there would be little evidence of emotional responses from the participants, which might have provided further insight into their attitudes towards using the discovery system to find biodiversity information resources. The questions were deliberately open-ended to encourage the participants to describe their individual experiences and thoughts, providing a descriptive insight into the use and impact of the discovery system at the Natural History Museum Library.

Following the completion of the user surveys, any respondents who had agreed to participate in further interviews were contacted and invited to the interview using the e-mail addresses they had provided. They were provided with the interview questions, consent form and information sheet informing them of the dissertation research and its process. Copies of the research information sheet and consent form may be found in Appendix 3 and copies of the interview questions may be found in Appendix 4.

6.2 Results from User Interviews

The recruitment to the interview stage following the surveys was low, with a total of six survey respondents consenting to further interviews between both the Combined Library User survey and the Independent Library User survey. Of the interview respondents agreeing to further interviews, one was an administrative staff member, and so was excluded from the study, as they were not a scientific staff member, one was a visiting postdoctoral researcher who left the museum before the interview could be conducted, and four were senior researchers. Following the invitation to the interview, only two of the five consenting and eligible interview candidates responded and provided responses to the interview questions.

While caution must be exercised in extrapolating a general conclusion based on such a small sample size, both interviews pointed to the fact that the users required more time to learn how to use the discovery system before they could gauge its impact on their searching behaviour and research

overall. Both interview candidates expressed widely differing views of the discovery system and its impact on their research and information-seeking behaviours.

Both participants had contrasting views of the discovery system. One interview participant, a taxonomist working in botany, felt that single search interfaces such as those provided by Primo would not impact how research is conducted at the Natural History Museum. Furthermore, they admitted that conducting a search of the library's collections was not the first destination when searching for informational resources. As they stated: "The Library is not my first port of call – JSTOR (database of journal articles), BHL (Biodiversity Heritage Library) are usually where I go first – unless I know, or think I know, we have the item". Additionally, they stated that they felt that the library search was too general in scope, and felt that other systems "may be better tailored to find specific items of information". Finally, they felt that the Library and Archives search system "needs to be integrated with other systems so that it is multi-faceted".

This response contrasted greatly with the interview responses of an entomologist, who felt that the discovery system facilitated improved, increased access to the collections, and that it had a beneficial impact on biodiversity research at the Natural History Museum, particularly through the efficiency and time-saving qualities of the discovery system interface, in which formerly separate and multiple databases had been "unified"

"The major benefit will probably be in the speed of searching and not needing to look through several databases… I think the major benefit is in saving time. All library searching can be done from the desk (or from anywhere there is computer access) without the need to walk to several different locations….Those benefits apply to all areas of scientific research, including that on biodiversity. "

Furthermore, the discovery system would enable not only researchers in the Museum to access important information resources on biodiversity, but would potentially be an aid to those outside of the Museum as well. As they state:

"If virtual access is made available to researchers in countries rich in biodiversity yet poor in other forms of access to published material (e.g. hard-copy libraries) then the digital library has the potential to greatly improve the contribution those researchers can make to addressing biodiversity issues."

Rather surprisingly, although one of the key functions of the discovery system is to facilitate "discovery" of the library collections, the participant felt that the discovery system did not facilitate this serendipitous aspect. As they state:

"…the use of such databases might mean missing a few publications that are not on the database, especially older literature. Also, the serendipity of browsing library shelves is lost. Perhaps that will not have a negative impact on research, but a generation might miss on the joy of discoveries made by chance!"

Both participants therefore provided contrasting, but still insightful views on the perceived impacts of the Library and Archives discovery system on search and research at the Natural History Museum.

6.3 Key Informant Interviews

In order to establish and build a case study, various types of evidence must be collected, including the perspectives of key informants and participants in the process. The purposive sampling of these key informants will enable in-depth insights into processes and interactions that will facilitate documentation of the transition to the "virtual library" and the implementation of the library discovery system, Primo, and the Informatics Data Portal. The key informants will be identified and selected "by virtue of their position within the context..such informants can provide the inquiry team with an 'inside' view of the norms, attitudes, constructions, processes, and culture that characterize the local setting" (Lincoln and Guba, 2007, pg. 258). The key informants will therefore facilitate insights into the research aim and objectives as defined by Pickard (2007) and Lincoln and Guba (1985).

The key informants were selected for the key informant interviews based both on their professional role within either the Library and Archives or the Informatics Department, and on the level of participation in the implementation of either the Library and Archives discovery system or the Informatics Data Portal. In total, seven prospective interview candidates were selected from the Library and Archives, and two candidates from the Informatics Group. These candidates were formally invited to the interview process by e-mail. Five candidates from the Library and Archives and one candidate from the Informatics Group agreed to participate in the interviews, and the consenting candidates were then sent the information sheet and consent form, informing them that their personal details would be included in the final report on the dissertation research in accordance with the ethics considerations stated previously.

Whenever possible, the interviews were conducted face-to-face; however in a few cases, use of an asynchronous e-mail interview was utilised when it was not possible or convenient to meet. In the case of the face-to-face interviews, which lasted between 20-30 minutes, the interview was recorded and responses reviewed. Quotations were selectively transcribed as supporting evidence.

An external interview participant was also included to provide a point of comparison with the systems implemented at the Natural History Museum. Imperial College Library is an academic library in an institute of higher education which has a subject focus in the sciences. Imperial College Library implemented Alma in 2011 and Primo in 2013. Andrew Preater, Team Leader of Systems and Services provides an insight into the implementation, impact and current use of the discovery system, Primo, at Imperial College Library.

6.3.1 Evidence obtained from Key Informant Interviews

In order to supplement the information obtained from users through the surveys and interviews, a perspective on the implementation, uptake and usage of the discovery system in the library and the Informatics Data Portal was sought through the interviewing of "key informants". These interviews were conducted with six museum staff members, with five staff members from the Library and Archives and one staff member from the Informatics Group. One external participant was interviewed for a point of comparison. The interviews were semi-structured, involving the use of open-

ended questions and soliciting observations and opinions from each of the staff members on the implementation, use and uptake of the new system.

Interview with Ms. Jane Smith, Head of Library on the choice and implementation of the discovery system:

The Head of Library, Jane Smith, provided the following insights about the choice and implementation of the discovery system. According to Ms. Smith, a discovery layer-type system would enable the library to meet its strategic objectives, particularly those posed by the museum-wide digital strategy. Two "core strands" were identified that would be fulfilled by the implementation of the discovery system. The first was the improvement and development of services "to support both internal and external researchers". And the second was collections management and development. Ms. Smith states: "We needed systems that would enable us to manage both print and digital content effectively, and support the effective discovery of that content. We needed a new system that would underpin the Library and Archive's contribution to Science/Museum strategic goal to develop the 'digital museum' as more material will not only be available, but more importantly, discoverable, via the library systems, contributing to the goal of enhancing collection use and increasing research and public engagement with the collections." (Jane Smith, *pers. comm*.) The new library system would also fill a much-needed gap for collection management, as the earlier system was out of date and "inefficient, with a lot of bolted on functionality". The single search capability provided by the discovery system was also seen as providing an efficient search facility for researchers, but would also enable the library to manage its collections as usage data of both print and digital materials could be readily assessed, providing valuable information for collections management. Furthermore, the capability of the discovery layer to integrate both internal and external content such as the Biodiversity Heritage Library and eventually museum collections databases such as KeEmu (the Natural History Museum's collections database) once the facilities to integrate these resources becomes available. Finally, the discovery layer would provide information to help library staff deliver current awareness and tailored support. The implementation of the new system has already increased the usage of the physical collections. "Already the number of searches has doubled within the period since go-live. The capability to search across resource types (print/electronic; published/original material; in-house/external) reduces the risk of missing potentially relevant resources. The more visible/exposed the physical collections become via the new Search, the more likely it will be requested. " So according to the Head of Library, the implementation of the discovery system will provide benefit to both the library and to its users, the many researchers both internal and external, who will use the library and its collections.

Interview with Dr. Mel Smith, Head of Modern Collections

The Head of Modern Collections, Dr. Mel Smith, provided the following insights about the impact of the discovery system on the management of collections in the library. Management of collections was of particular concern when the system was implemented, as in conjunction with the museum-wide digital strategy, many of the print serial subscriptions were being changed over to on-line only, and the library was undertaking a digital "revamp" of the catalogue, which would better integrate print and digital resources. A discovery system, such as Primo, would be necessary in order to make

these resources visible and findable. “The new system is much easier to use and allows researchers to search both print and electronic at the same time. It should be much easier for users to find what they need on the system and then access it. The system is easier to search, the narrow down searches and allows for variations in spellings etc., so information should be more retrievable. The additional data that can be searched through Primo for material that is not held in the NHM Library also means that library users can discover across a broader base.” Furthermore, Dr. Smith mentioned that the new discovery system would mimic the “Google Scholar” search and research environment in an attempt to bring users back to the library and to envision the library and its collections as a starting point for research at the Natural History Museum. She also makes an interesting observation about researcher behaviour in the library, as embodied by long-time members of staff. As she states:

“Ideally, as librarians we would like to think that a research project starts with a search of the Library catalogue, but clearly Google Scholar and the like play a very important role these days. In choosing a new library system we were very keen to have a system that looked and felt much more like google environment, was more forgiving of misspellings and gave seamless access to both our print and electronic resources. The Museum library is slightly unusual in that many of its established users rarely use the catalogue; they simply know where the journals that they want are located. The Museum has many very established academics who have worked at the institution for a long time. The catalogue is only used by many when the Librarian moves something and they have to relocate it. Electronic journals were indexed separately in an A-Z list and therefore the catalogue represented the old printed stuff and the A-Z the modern electronic stuff. Obviously, this was not really the case as the NHM Library still takes many of its current journals in print. With the new Alma catalogue and Primo search system print and electronic are managed together and presented to the user in one interface. Primo also allows users to search material not held by the NHM Library itself. Hopefully as the new systems are bedded in and researchers realise its potential it will become more of a starting point for research than it has been in recent years. I hope that researchers will move back to seeing the library system as worth searching at an early stage in their research. The additional resources that Primo provides will hopefully ensure that their searches are more complete and therefore our researchers are more aware of other work going on. We are also developing our services through Alma so that eventually researchers will be able to order material directly through the catalogue and it will be delivered to their research space or they will be able to access digital material directly and seamlessly without needing to deal with messy logins and paywalls.”

Dr. Smith also described the impact that the new system (both Alma and Primo) has had on collection management in the library. Already since its implementation, it has had a direct impact on material circulation from the off-site store in Wandsworth, and it has affected workflows in acquisitioning and processing of new library materials.

“We have for a long time been keen to track material both on loan and as it moves around the library. We are currently barcoding our serials collections (in addition to our books) the ultimate aim is to be able to issue all items and track them. This will be a move forward from the current bookboard system where users simply leave a book board on the shelf where they have taken the item from. Library staff will be able to see who has borrowed the item and what items each user has

borrowed which will be particularly useful in the reading room where security is a concern. Hopefully many of the acquisitions processes will be more automated leaving more staff time for providing services to the users."

Finally, Dr. Smith concluded with the observation that as users adapt to the new system and develop their familiarity with it, she hopes that they will use the discovery system to find and use more of the collections in their research work. She concludes:

"So far the users seem very positive about the changes and usage of the system has gone up. Hopefully as we develop our user education and promote the system more the users will be able to exploit the system more fully. Library staff are still at the beginning in learning what the new system can do and potential ways to exploit it. Hopefully future enhancements will also provide greater benefits for users."

Interview with Mr. Michael Loran, Systems Librarian

According to Mr. Loran, the greatest overall impact of the discovery system, Primo, on biodiversity research at the Natural History Museum has been in the "ability to search the Biodiversity Heritage Library (BHL) and library print collections from a single interface, and the improved access to the range of online subscriptions the library offers in this area". This is a dramatic change from the previous system, in which the library catalogue offered print holdings and separate e-resources were presented, including an e-journal A-Z. Probably the greatest implication for the library from the new system (both Primo and Alma) will be related to metadata, which is crucial for the function of the discovery layer. As Mr. Loran pointed out, "I think there could be an impact on metadata creation – do we still need to create in-house records for articles that are already indexed on Primo Central. We may also need to look at creating more metadata for non-print material – if Primo begins to contain more in-house digitised material what is the role of the library in creating metadata for this material. The possible introduction of on-demand digitisation via Primo will also create workflow challenges that need to be addressed." As for any further impact on searching behaviours by the users, Mr. Loran thought that it might be too early to see any direct impact. The discovery system has only been implemented since July, just less than six months at the time of this investigation. As he states: "Bit early to tell I think – but we are perhaps finding that the search behaviour for our users is very different from the typical under/postgraduate approach where a particular citation is sought – our users are perhaps more likely to conduct searches in a specific area to see what is available in the published literature. Our users are perhaps more interested in the ability to browse in a particular subject area." Finally, Mr. Loran points out that it may not be the system that influences search, but rather the post-search behaviours. "I don't have any evidence to support this but I imagine that user search strategies are less based around which system is most appropriate to use initially and more about interpreting and filtering results post searching."

Interview with Ms. Sarah Vincent, Digital Services Librarian

Sarah Vincent is the Digital Services Librarian at the Natural History Museum. Her role involves determining the usage of the collections and improving the system in response to user needs and behaviours, so as such, she is well-placed to provide insights into the use and impact of the system. The

major impact of the discovery system has been the increased usage of collections, particularly formerly "invisible" digital resources and print holdings in off-site storage. She explains: "It's early days yet, but we're seeing indications of wider use of our collections and increased use of online resources and off-site print collections – obviously difficult to put all that down to the new system given how many other changes were happening at the same time, but I think it must be a significant part of the reason. Another key benefit we were looking for was that Primo should be less time-consuming for our users – they're researchers and shouldn't be spending more time than necessary…trying to find literature. Based on the feedback we received from our latest user survey, most respondents prefer the new system over the old one so we're succeeding in that respect…Still some bugs to iron out around off-site requesting but a lot of that's down to the quality of our metadata." According to her, the users can be grouped into two major groups, the "traditional" users, comprising researchers who have been active in the museum for many years (decades) who are used to using the library's numerous resources, and researchers who are more "at home" in a digital environment, so they will already have different inherent information-seeking behaviours. She explains: "For our more traditional users, I think it's quite a change. Anecdotally, they're the ones who are more likely to find the interface overwhelming/crowded and need more support when it comes to logging in and using the system to its full potential. I think a big part of that is habit, though – unlike HE libraries, some of our users have been here for decades and although (in my opinion) the old catalogue was a total nightmare to use, they were used to it…I've had a few complaints about the new system not being intuitive, which I think sometimes does just mean 'I'm not used to this yet'. For our users who already have good levels of online research skills and are more at home in a digital information environment, I don't think it's a particularly big change although I suspect a lot of them have just stuck to using Google Scholar. We're working to support those users as well of course – there's no reason Scholar and the L&A (Library and Archives) system can't co-exist and complement each other, I use both of them myself for different tasks. We've recently set up a mechanism whereby users working from home/in the field can use Scholar through the Museum network, so they maintain access to the Scholar Library Link indicators and can easily identify and access our subscribed content, despite being outside the NHM network. It's a simple thing, but if you're running dozens of searches a day and hitting paywalls all the time it can cut out a lot of frustration and save you time." As for impact on search behaviour, and any emerging patterns, Ms. Vincent states: "Frustratingly, our Primo reports haven't been working since we went live (it's supposed to be fixed in a future Primo release), so at the moment, I can't see the more granular detail behind the searches. We have Google Analytics in place at least, which shows us that searches and sessions are at a healthy level and have steadily increased since July. Apparently, the L&A homepage was the second most-visited page on the NHM website during the six months following the launch of Primo and the restructure of our webpages, which is great news."

Interview with Ms. Sarah Sworder, Reader Services Librarian

The Reader Services section at the Natural History Museum Library work directly with all users of the library, whether they are visiting members of the public or internal staff members. The Services department routinely handles requests from researchers and also assists them with searching for resources. As such, the Reader Services librarians would be able to provide a unique insight into the

impact of the discovery system, Primo, on search and research. The Reader Services librarian, Ms. Sarah Sworder, was able to provide the following insights about the impact of the discovery system on search:

"The greatest impact that the system has had is to bring all our collections searchable in one place so readers are able to search print, online and archival collections in a single search." Furthermore, she observed, "The information is there but sometimes our users struggle to find it. This is more of a user education issue rather than the fact that the information is not findable. Our users (staff in particular) are used to finding information in a different way. I believe in time as users get used to the discovery system it will become easier for them to locate what they need." She feels that the reaction to the discovery system has been mixed, and adds, "Some users find it much easier to use the one search system and some find it harder to navigate. I think in time researchers will get used to it and see the benefits and their searching habits will change as a result." She agrees that the discovery system has impacted research in that it has made electronic resources more visible and accessible. Finding electronic journals is undoubtedly more straightforward thus making it easier for researchers to find articles relevant to their work. As time goes on and users get more used to the system I feel the impact on NHM research will be greater as it is still early days for the library using Primo."

The Informatics Data Portal: Interview with Dr. Vincent Smith, Informatics Research Leader, Life Sciences Department, Natural History Museum

The Data Portal was devised in response to the fragmentation of taxonomic data. Taxonomy is an information science and the idea that there will be a single source of information such as a single book about all species is wrong! There is an enormous volume of published literature, specimen collections comprising over 1.5-3 billion specimens worldwide, with 80 million specimens held at the Natural History Museum alone, and the datasets accompanying the published literature – so this is a huge amount of information. Research happens at many different scales as well. What is required from on-line information is the stability of resources and persistent links, so the Museum is tackling this problem through providing a route for external access to these information resources. Because the information resources are so heterogeneous, it is not expected that research will always start in an institution such as the NHM Library or even the data portal. Research often starts in other places, such as Google or Yahoo. For taxonomy, GBIF, EOL and NCBI GenBank are often used as starting points, but there are a myriad of niche topic resources for coverage as well. The Data portal will be a starting point for the Natural History Museum, and will be about more than discovery alone – it will be a place of deposit for data that is generated, and will provide a persistent repository for these informational resources that will also have citability. So the Data Portal will not only aid researchers in discovery and providing information, but will also enable discovery to occur externally as well, supporting researchers in their work. The Data Portal is a mechanism for us to contribute to Global biodiversity resources. The Data Portal is an ecosystem of services that allows us to expose data. It will contribute data to GBIF, which will further discovery on a global scale. There is a lot of scope with the Data Portal to innovate. It is a research tool rather than just a source for information. The long-term potential of the Data Portal to act as an innovative space, serving more than niche needs. It is still in development, so will need to be

tweaked and adjusted as needs evolve, but providing a persistent, stable framework for this information will make it easier and make search more efficient. It is currently in its beta testing stages, and is undergoing a soft-launch, being released museum wide on December 15th, but there will be a public launch to GBIF on January 20th. There is currently a high level of usage, which may cause instability. Most importantly, the Portal will be a better use of data that we have currently in our museum collections. The Data Portal will not replace physical specimens, but the digital data has a unique role and enormous reach.

The External Perspective: Discovery Systems in a Higher Education Library: Interview with Mr. Andrew Preater, Team Leader, Systems and Services, Imperial College Library

Imperial College London Library implemented Alma in 2011 and Primo in 2013. As a point of comparison with the Natural History Museum Library and its specialist collection, Mr. Preater, Team Leader, Systems and Innovation Support Services, was interviewed. The focus of the interview was on how the researchers at Imperial College London utilise the discovery system for search and research at a large academic library.

Imperial College Library has recently increased its spending on electronic resources. In 2013, the annual resource budget on electronic resources was 53%. By 2014, this had increased to 91%. The proportion of the budget spent on print declined steadily during this period. The primary focus at Imperial are the STEM subjects (sciences, technology, engineering and maths). As the sources for STEM subjects are increasingly electronic, the library collection was extensively weeded, with print journals and over 20% of the print book collection discarded. However most on-line articles start in the 1970s, and earlier published literature is still relevant, particularly for maths. So content still needs to be accessible to researchers, particularly to the academics, who also produce research outputs, and the library must support them in this. Although Primo provides a Google-like single search, the library will not out-Google these commercial search engines, so is still not the starting point for research, however, this will vary for users based on their field of study and their level of experience. For instance, graduate students will often have a good level of recall and precision, but more senior academics tend to focus on core journals and on authors publishing in their areas of interest, so there are different routes into information seeking and retrieval, and the library is not always a starting point. The discovery system search supports various types of citation searches, such as authors, titles and dates. The search must be forgiving, unlike previous OPACs. The main objective is to serve up the good stuff (for the researchers) when they come to us. Although the discovery system is good for this, one of the areas that might need improvement is the experience of serendipitous browsing, as you don't get this in discovery layers. Ex Libris is working on replicating browsing physical items such as a shelf browse (which shows images of book covers and items that are shelved near the title searched for) and Amazon-like recommendations based on the search terms. Finally, the library is taking an inside-out approach and pushing our content outside. Metadata becomes important in this aspect, as it will be indexed, particularly if it is Open Access. The discovery system has improved the findability of resources, but can also be used to push content externally. In the end, it is all about how to get things into the hands of the users.

6.4 Summary of Evidence provided by Interviews

While the library user interviews cannot be construed as being representative, owing to their small sample size, they expressed two widely differing views with regard to the impact of the Library and Archives discovery system on research in biodiversity and information seeking behaviours.

One user clearly felt that the new system was beneficial and provided a positive impact in terms of efficiency in finding print and digital items in the library's collections, but the other was ambivalent and felt that there was no real impact or change, and that they usually started research outside of the library. The views expressed by the library staff indicated that they hoped that the new system, with its "Google-like" search interface would entice researchers back to the library, but also that they would endeavor to work with those researchers who preferred to use these external search engines through supporting systems integration, as per the comments provided by the Digital Services Librarian, Sarah Vincent.

The user expressing a positive view of the new system in that it improved access to resources and enhanced efficiency of search is in accordance with the views expressed by Dr. Mel Smith, the Head of Collections, and Jane Smith, the Head of Library, who both stated that the discovery system should facilitate search through making the library's collections more "searchable" through a single interface. Similar views are expressed by Michael Loran, the Systems Librarian and Sarah Vincent, the Digital Services Librarian. Furthermore, Sarah Vincent has pointed out that the usage of the Library collections has increased, particularly for "invisible" parts of the collections such as the electronic resources and print holdings held off-site. While the library staff all agreed that the single search interface provided by the discovery system facilitated search and would therefore impact upon research through the efficient and effective delivery of resources, including the "discovery" of potential new resources by users, this view was not shared equally by the library user interview participants, who felt that the search was too general in scope and even prevented serendipitous browsing. However, this discrepancy in the viewpoints expressed by the library user interview candidates may be a result of the short period of time that has elapsed since the implementation of the discovery system. It is possible that they are still learning how to best use the system for their research.

Dr. Mel Smith and Sarah Sworder, the Reader Services Librarian, both pointed out that some of the users are more used to browsing the physical collections, so it may take time and further acclimatisation, as well as further information literacy training before these users are comfortable using the discovery system. This was certainly evidenced in the user who, while they felt that the discovery system facilitated more efficient search, felt that the serendipitous browsing of the library's collections was absent.

The interviews with both the users and with Sarah Vincent indicated that "discovery does not always happen in the library". One of the library users would begin their research with a specific database, such as JSTOR or the BHL rather than starting their search on the library system. Sarah Vincent also pointed out that many library users start their research using Google Scholar, so the library

discovery system will aim to integrate Google Scholar with Primo. The BHL, along with the formerly separate electronic journal databases, has already been integrated.

Finally, although it was too early on in the implementation of the Data Portal to examine its usage or impact, Dr. Vincent Smith provided a succinct summary of the proposed impact of the discovery system provided through the Natural History Museum's Data Portal. Its major impact will be in providing a citable, persistent repository for datasets produced by researchers at the Natural History Museum, and furthermore, will go beyond discovery in enabling interaction, sharing and communication of biodiversity datasets beyond the Museum itself, potentially furthering research in biodiversity elsewhere, and pushing the data produced by the Natural History Museum into a public, global scale.

Interestingly, the perceived impacts of the discovery system on search and research at a specialist research centre are similar to those observed in a large academic library such as Imperial College, where the discovery system has facilitated increased use of the collections, particularly for digital formats. Additionally, the requirement for improved browsing to facilitate serendipitous search was also seen to be a needed innovation with the current discovery system, as users often came to the system with preconceptions of search based around that of the previous OPAC.

The interviews therefore have provided an insight into the impact of the discovery systems, not only on search and information-seeking behaviours, but also on research through the potential to use more of the collections, and also, in the case of the Data Portal, to provide a repository for datasets that can be "discovered" through search for future use and research in biodiversity.

Chapter 7: Conclusion

7.1 Summary of Findings

There are two aims to the study. The first, to investigate the impact of discovery interfaces on research and search for information in biodiversity research at the Natural History Museum, and the second to build a case study of the Library and Archives discovery system, Primo, and the Data Portal within the Museum's greater strategic aim of developing a virtual museum.

This study had the specific objectives to:

1.) Determine exact attitudes to searching for biodiversity information resources using the Library and Archives discovery system and the Data Portal.
2.) Explore how biodiversity information is discovered and used.
3.) Examine the impact of the discovery systems on search and research.
4.) Document the transition towards a digital, networked and aggregated discovery systems approach to managing biodiversity information.

The major findings for each objective are summarized below:

7.2 Attitudes to Searching for Biodiversity Information Resources using Primo

The study has established that between 57-65% of scientific researchers use the Library and Archives search to look for information. However, while attitudes towards the Library discovery system are mostly positive, there is some ambivalence or uncertainty about using the system for search and also a lack of certainty surrounding the resources and services available through the discovery system such as access to the BHL. Furthermore, 22-28% of respondents do not use the discovery system as the first point of search or keeping up to date, preferring to use external search services such as Google Scholar, individual journals or Web of Science rather than the Library and Archives discovery system. Both the survey and the user interviews indicate that the discovery system may not enable serendipitous browsing and that this is an aspect of the discovery system that may need to be improved upon. Finally, the user survey and interviews have identified a gap in knowledge about the capabilities and research potentials of the new discovery systems, and thus further outreach by library staff is recommended.

While it was not possible to examine attitudes to searching for biodiversity information using the Data Portal, owing to its late launch, an examination of user responses to this service would be interesting future research.

7.3 Information Discovery in Biodiversity Research

The surveys, user interviews and Library staff interviews provided some insights into how information is discovered using the Library and Archives search. Most respondents relied on external systems such as Google Scholar for keeping up to date with research and also for conducting their own searches. They also stated that their searches were often serendipitous, with

information found whilst browsing in the Library Reading Rooms. This observation was also made by the Library staff, particularly in the interview with Sarah Sworder, the Reader Services Librarian, and with Dr. Mel Smith. These observations from the interviews and the survey also suggest that there is an observed divide between library and archives users who prefer to browse the physical collections and those who rely primarily on on-line resources, possibly reflecting the levels of experience of different generations of researchers at the Natural History Museum, although data from the survey suggested that the use of both digital and print resources was in the majority.

7.4 Evaluation of Impact

The study has revealed that it may be too early to gauge the impact of the discovery system on search and research at the Natural History Museum, as the users responding to the survey and in the interviews have expressed some ambivalence or uncertainty regarding the ability of the discovery system to impact search. This may indicate that further outreach is required by Library and Archives staff in order to increase awareness of the resources and services that are available through Primo. The interviews with Library staff; however, indicate that there is a great confidence regarding the capabilities for improved search and scope of resources available through Primo. Further research will be required at a later date to ascertain impact of the discovery system on search and research at the Natural History Museum.

7.5 Documenting the Transition to a Virtual Informational Environment

While it was not possible to obtain evidence from the Data Portal, owing to its late launch, the evidence obtained from the survey and interviews of both Library and Archives users points to a gradual transition towards a virtual museum environment for biodiversity information resources. However, the study also uncovered a reliance on the print collection, particularly the legacy literature in the form of books and pre-1920s journals. Taxonomic information that is critical for biodiversity research has a great longevity, so access to these resources must be assured. While digitised versions of some titles are available through the Biodiversity Heritage Library and indexed for discovery on Primo, the serendipitous browsing method of information retrieval favored by some of the research staff was not adequately accommodated for within the discovery system.

7.6 Synthesis of Evidence

This research work has revealed that while the implementation of the discovery system in the Natural History Museum Library and Archives enabled efficient searching of the collections through the single-search interface, users of the system are still undecided and ambivalent about the perceived impact of the discovery system, Primo, on their search and research, other than that the single-search interface provided a more efficient search system for facilitating access to digital resources held in the collections compared to the previous system, in which digital resources could only be found by searching multiple interfaces. This is supported by the observed increase in the usage of collections, particularly of the born-digital and off-site items. However, the Library and Archives staff were optimistic about the implementation of Primo and felt that the impacts of the discovery system on search and research would not only enable an increase in the use of the print and digital collections, but

also result in increased traffic to the library and use of the system, in spite of users opting for Google Scholar or other search sites instead of starting their research using the Library discovery system.

The evidence collected from the document analysis provided an insight into the goals and objectives of the implementation of the Library and Archive discovery system, which were met with Primo. This evidence was corroborated by the Library staff in the interviews, who felt that the discovery system made searches more efficient and effective across the library collections, and also entailed positive impacts from a collections management point of view. However, the users expressed an ambivalence or uncertainty about the impact of the discovery system on research, both in the survey responses and in the interviews. This may reflect the short time lapse between the implementation and launch of the discovery system and the uptake of the new system by library and archives users.

7.7 Limitations of the Study and Directions for Future Research

As case studies are essentially "snapshots" in time of a particular process, the timing of the research project to closely follow the implementation of both the Library and Archives discovery system, Primo, and the Natural History Museum Data Portal may have been too close to the time of implementation and launch of these systems, as was the case for the Data Portal. This may have resulted in the ambivalent response to the discovery system from Library and Archives users, and the advised abandonment of the survey and interviews related to the Data Portal.

Because of the low response rate, the surveys would introduce a high degree of self-selecting or sampling bias, so although the results provided are insightful, they would not provide a reliable basis upon which to evaluate the impact of a discovery system.

Finally, this research has opened up the possibility of an extended piece of research, utilising both usability testing and an embedded ethnographic approach examining the uptake and impact of the discovery interfaces at the Natural History Museum as it makes the transition to an "informational museum" in the service of biodiversity research and the fulfilment of its mission statement: to "encourage the discovery of the wonders of nature".

Bibliography

Allnott, G. (2012) 'What Natural Science Collections can tell us', *Museums Journal*, 112(12), pg. 19.

Berendsohn W., Güntsch A., Hoffmann N., Kohlbecker A., Luther K, Müller A. (2011) Biodiversity information platforms: From standards to interoperability. *ZooKeys* 150: 71-87. doi: 10.3897/zookeys.150.2166 (Accessed: 24 September 2014).

Bishop, A.P., Neumann, L, Star, S.L., Merkel, C., Ignacio, E. & Sandusky, R. (2000) 'Digital libraries: Situating use in changing information infrastructure', *Journal of the American Society for Information Science*, 51(4), pp. 394-413. [Online] DOI: 10.1002/(SICI)1097-4571(2000)51:4<394::AID-AS18>3.0.CO;2-Q (Accessed 13 July 2014).

Bollen, J. & Luce, R. (2002) 'Evaluation of digital library impact and user communities by analysis of usage patterns', *D-Lib Magazine*, 8(6) pp. 1-14. [Online] Available at: http://www.dlib.org/dlib/june02/bollen/06bollen.html (Accessed: 20 Aug 2014).

Breeding, M. (2010) 'State of the art in library discovery 2010', Computers in Libraries, 30 (1), pp.31-35. *Library Technology Guides* [Online]. Available at: http://www.librarytechnology.org/ltg-displaytext.pl?RC=14574 (Accessed: 13 July 2014).

Breeding, M. (2012) 'Looking forward to the next generation of discovery services', *Computers in Libraries* 32(2), pp. 28-32. [Online] Available at: http://www.librarytechnology.org/ltg-displaytext.pl?RC=16731 (Accessed: 20 August 2014).

Brophy, J. & Bawden, D. (2005) 'Is Google enough? Comparison of an internet search engine with academic library resources', *Aslib Proceedings*, 57(6), pp. 498-512. [Online] Available at: http://www.city.ac.uk/__data/assets/pdf_file/0010/79822/Is-google-enough.pdf (Accessed 13 September 2014).

Campbell, D. (2010) Library Discovery Services – A better way to find what you're looking for, *Wiley Exchanges*, Dec. 22, 2010. [Online] Available at: http://exchanges.wiley.com/blog/2010/12/22/library-discovery-services-a-better-way-to-find-what-youre-looking-for/ (Accessed Sept. 22, 2014).

Case, D.O. (2012) *Looking for Information: A Survey of Research on Information Needs, Seeking and Behaviour*, 3rd Ed. London: Emerald Group Publishing Ltd.

Chu, H. (2010) *Information Representation and Retrieval in the Digital Age*, 2nd Ed, Medford, New Jersey, U.S.A.: Information Today, Inc.

Connaway, L.S. & Dickey, T.J. (2010) *The Digital Information Seeker: Report of findings from selected OCLC, RIN and JISC User Behaviour Projects*, [Online] Available at: http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/media/documents/publications/reports/2010/digitalinformationseekerreport.pdf (Accessed: 20 September 2014).

Couper, M.P. (2000) 'Web surveys: a review of issues and approaches', *Public Opinion Quarterly*, 64(4), pp. 464-494. [Online] Available at: http://www.jstor.org/pss/3078739d (Accessed: 20 August 2014).

Dempsey, L. (2006) 'The Library Catalog in the New Discovery Environment: Some Thoughts'. *Ariadne*, Iss.48. [Online] Available at: http://www.ariadne.ac.uk/issue48/dempsey (Accessed: 20 September 2014)

Dempsey, Lorcan. (2012) 'Thirteen Ways of Looking at Libraries, Discovery, and the Catalog: Scale, Workflow, Attention.' *EDUCAUSE Review Online*, December 10, 2012. [Online] Available at: http://www.educause.edu/ero/article/thirteen-ways-looking-libraries-discovery-and-catalog-scale-workflow-attention. (Accessed: 20 September 2014).

Dervin, Brenda, and Carrie Lynn D. Reinhard. 2007. Sense-making the information confluence: The whys and hows of college and university user satisficing of information needs. Final Report. *Report on National Leadership Grant LG-02-03-0062-03, to Institute of Museum and Library Services*, Washington, D.C. Columbus, Ohio: School of Communication, The Ohio State University. http://imlsproject.comm.ohiostate.edu/imls_reports/PERFORMANCE_REPORT/PERFORM_ONLINE.pdf. (Accessed 23 October 2014).

Fast, K.V. and Campbell, D.G. (2005) '"I still like Google": university students perceptions of search the OPAC and the Web', *Proceedings of the American Society for Information Science and Technology*, 41 (1), pp. 138-146 Wiley Online Library [Online]. DOI 10.1002/meet.1450410116 (Accessed: 13 July 2014).

Fidel, R. (1993) ' Qualitative methods in information retrieval research', *Library and Information Science Research*, 15, pp. 219-247. [Online]. Available at: http://faculty.washington.edu/fidelr/RayaPubs/QualitativeMethodsinInformationRetrievalResearch.pdf (Accessed 20 Aug 2014).

Fink, A. and Kosecoff, J. (1998) *How to conduct surveys: A step-by-step guide*. 2nd edition. Thousand Oaks, CA: Sage.

Fleming-May, R. and Grogg, J. (2010) 'Assessing Use and Usage: The Imperative', Library Technology Reports, 46(6), pp. 5-10. [Online] Available at: http://web.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=2cc27c3a-91ca-4956-893e-5a6628368fe4%40sessionmgr10&vid=4&hid=15 (Accessed: 22 Aug 2014).

Gillham, B. (2000) *Case Study Research Methods*. New York: Continuum.

Godfray,H.C.J. (2002b) 'Challenges for taxonomy – The discipline will have to reinvent itself if it is to survive and flourish. *Nature* 417(6884): pp. 17-19.

Godfray, H.C.J. (2002c) 'Towards taxonomy's glorious revolution', *Nature* 420(6915): pg. 461.

Godfray, H.C.J, Clark B.R., Kitching, I.J., Mayo S.J. and Scoble, M.J. (2007) 'The web and the structure of taxonomy', *Systematic Biology*, 56(6), pp. 943-955. [Online] DOI: 10.1080/10635150701777521 (Accessed: 15 Oct 2014).

Graham, C.H., Ferrier, S., Huettman, F. Moritz, C. & Townsend Peterson, A. (2004) 'New developments in museum-based informatics and applications in biodiversity analysis', *Trends in Ecology and Evolution*, Vol. 19(9), pp. 497-503.

- Hardisty, A., et al (includes Smith, V.S. and 78 other authors). (2013). 'A decadal view of biodiversity informatics: challenges and priorities', BMC Ecology, 13(16), pp. 13-16. DOI: doi:10.1186/1472-6785-13-16.

Hine, C. 2000. *Virtual Ethnography*, London: Sage Publications.

Hine, C. 2008. *Systematics as Cyberscience: Computers, Change and Continuity in Science*, Cambridge, MA: MIT Press.

Hoeppner, 2012. The Ins and Outs of Evaluating Web-Scale Discovery Services. *Information Today*, [Online] Available at: http://www.infotoday.com/cilmag/apr12/Hoeppner-Web-Scale-Discovery-Services.shtml (Accessed: 14 May 2014).

Kennedy, S.P. (2014) 'Uncovering discovery layer services', Public Services Quarterly, 10(1), pp. 54-61. [Online] Available at: http://dx.doi.org/10.1080/15228959.2014.875788 (Accessed: 13 July 2014)

Kuhlthau, C. (1988). Longitudinal case studies of the information search process of users in libraries. *Library and Information Science Research*, 10(3), pp. 257-304.

Lanclos, D. & Le Cornu, A. (2012) 'Digital Visitors and Residents: Progress Report'. JISC. [Online] Available at:http://webarchive.nationalarchives.gov.uk/20140702233839/http://www.jisc.ac.uk/media/documents/projects/visitorsandresidentsinterim%20report.pdf (Accessed 26 October 2014).

Levine-Clark, M., McDonald, J. & Price, J. (2014) Examining the impact of discovery systems on usage, *Wiley Exchanges*, July 21, 2014 [Online] Available at: http://exchanges.wiley.com/blog/2014/07/21/examining-the-impact-of-discovery-systems-on-usage/ (Accessed: 21 September 2014).

Lincoln, Y.S. & Guba, E.G. (1985) *Naturalistic Inquiry*, London, Sage.

Lister, A. & Climate Change Research Group. (2011) Natural history collections as sources of long-term datasets, *Trends in Ecology and Evolution*, Vol. 26(4), pp. 153-154.

Mace, G.M. (2004) 'The role of taxonomy in species conservation', Philosophical Transactions of the Royal Society of London B, 359, pp. 729-730. [On-line] Available at: http://rstb.royalsocietypublishing.org/content/359/1444/711.abstract (Accessed 28 July 2014).

Marchiononi, P. (2009) 'Why are users so useful? User engagement and the experience of the JISC digitisation programme', Ariadne, 61, pp. 1-12 [Online] Available at: http://www.ariadne.ac.uk/issue61/marchiononi/ (Accessed: 13 July 2014).

Mawhinney, T. (2013) 'Impact of discovery tools on e-book usage', Research Societies Libraries, July 18, 2013. [Online] Available at: http://exchanges.wiley.com/blog/2013/07/18/impact-of-discovery-tools-on-e-book-usage/ (Accessed: September 21, 2014).

Meyer, E, Eccles, K., Thelwall, M. & Madsen, C. (2009) Usage and impact study of JISC-funded phase 1 digitisation projects & the toolkit for the impact of digitsed scholarly resources (TIDSR), Final Report, 20 July 2009, Oxford Internet Institute, University of Oxford. [Online] Available at: http://microsites.oii.ox.ac.uk/tidsr/system/files/TIDSR_FinalReport_20July2009.pdf (Accessed: 13 July 2014).

Moritz, T. (2002) 'Building the biodiversity commons', D-Lib Magazine, 8(6) [On-line] Available at: http://www.dlib.org/dlib/june02/moritz/06moritz.html (Accessed: 20 Aug 2014).

Natural History Museum Biodiversity. 2014 On-line. Available: [http://www.nhm.ac.uk/nature-online/biodiversity/] Accessed: January 6, 2015.

Natural History Museum Informatics Department. 2014. On-line. Available: [http://www.nhm.ac.uk/research-curation/research/informatics/index.html] Accessed: May 14, 2014.

Natural History Museum Library and Archives. 2014. On-line. Available: [http://www.nhm.ac.uk/research-curation/science-facilities/library/] Accessed: January 6, 2015.

Orna, E. & Pettitt, C. (2013) 'What is Information in the Museum Context?', in Parry, R. (ed). Museums in a Digital Age. London: Taylor & Francis.

Parry, R. (2013) 'The Practice of Digital Heritage and the Heritage of Digital Practice' in Parry, R. (ed). Museums in a Digital Age, London: Taylor & Francis.

Patterson, D.J., Cooper, J., Kirk, P.M., Pyle, R.L. & Remsen, D.P. 2010. Names are key to the big new biology, *Trends in Ecology and Evolution*, Vol. 25(12), pp. 686-691.

Pickard, A.J. 2007. *Research Methods in Information*, London: Facet Publishing.

Pilsk, S.C., Person, M.A., Deveer, J.M., Furfey, J.F. and Kalfatovic, M.R. (2010) 'The Biodiversity Heritage Library: advancing metadata practices in a collaborative digital library', Journal of Library Metadata, 10, pp. 136-155. [On-line] DOI: 10.1080/19386389.2010.506400 (Accessed Aug 20 2014).

Preater, A. 2010. Mental models and user experience of a next-generation library catalogue. [Available: http://www.preater.com/wp-content/uploads/2011/12/Andrew-Preater-MSc-Dissertation-final.pdf] Accessed: Oct 20, 2014.

Raven, P. (2004) 'Taxonomy: where are we now?', Philosophical Transactions of the Royal Society of London B, 359, pp. 729-730. [On-line]. Available at: http://rstb.royalsocietypublishing.org/content/359/1444/729.full.pdf+html (Accessed 30 July 2014).

Research Information Network. 2006. Researchers and discovery services: Behaviour, perceptions andneeds. London: Research Information Network. http://www.rin.ac.uk/our-work/using-and-accessinginformation-resources/researchers-and-discovery-services-behaviour-perc.

Roberts, A. (2013) 'The Changing Role of Information Professionals in Museums', in Parry, R. (ed). Museums in a Digital Age, London: Taylor & Francis.

Sarkar, I.N. (2007) 'Biodiversity informatics: organizing and linking information across the spectrum of life', Briefings in Bioinformatics, 8(5) pp.347-357. [On-line] Available at: http://bib.oxfordjournals.org/content/8/5/347.full (Accessed 25 Jul 2014).

Sarkar, I.N. (2009) 'Biodiversity informatics: emergence of a field', BMC Bioinformatics, 10, Suppl. 14. [Online] Available at: http://www.biomedcentral.com/1471-2105/10/S14/S1 (Accessed 25 Jul 2014).

Scoble, M.J. (2004) 'Unitary or unified taxonomy?' Philosophical Transactions of the Royal Society of London B, 359(1444), pp.699-710.

Schonfeld, R.C. (2014) 'Does Discovery still happen in the Library? Roles and Strategies for a Shifting Reality', Ithaka. [Online] Available at: http://www.sr.ithaka.org/sites/default/files/files/SR_Briefing_Discovery_20140924_0.pdf (Accessed Sept 24, 2014).

Select Committee on Science and Technology (1991) Systematic Biology Research Volume 1 Report. London: HMSO.

Select Committee on Science and Technology (2002a) What on Earth? The threat to the science underpinning conservation. London: House of Lords, Session 2001-02 3rd Report HL Paper 118.

Select Committee on Science and Technology (2002b) What on Earth? The threat to the science underpinning conservation. London: House of Lords, Session 2001-02 3rd Report HL Paper 118.

- Smith, V.S., Rycroft, S.D., Harman, K., Scott B. & Roberts, D. (2009) Scratchpads: a data-publishing framework to build, share and manage information on the diversity of life. *BMC Bioinformatics* **10(Suppl 14):S6 :** .doi:10.1186/1471-2105-10-S14-S6.

- Smith, VS (2009) Data publication: toward a database of everything. *BMC Research Notes* **2 :**113 .

- Smith, V.S., (2013) Cybertaxonomy. Edited by Debby Shorley and Michael Jubb,, In: *The Future of Scholarly Communication* . Facet Publishing 63-74 .

Soberon, J. (1999) Postscript: Linking biodiversity information sources, *Trends in Ecology and Evolution*, Vol. 14(7)

Thackray, J. & Press, B. (2001) The Natural History Museum: Nature's Treasurehouse, London: Natural History Museum.

Thresher, A. D. (2012) *The impact of the Biodiversity Heritage Library Digitisation Project on Scientific Research at the Natural History Museum, London*. London Metropolitan University.

United Nations (1992) 1992 Convention on Biodiversity, Rio. [Online] Available at: https://www.cbd.int/doc/legal/cbd-en.pdf (Accessed 15 August 2014).

Williams, D. (2013) 'A Brief History of Museum Computerization', in Parry, R. (ed.). Museums in a Digital Age, London: Taylor & Francis.

Wilson, E.O. (1998) *Consilience: The Unity of Knowledge*, New York: Knopf.

Wilson, E.O. (2001) *The Diversity of Life*, London: Penguin.

Yin, R.K. (1994) 'Discovering the future of the case study method in evaluation research, Evaluation Practice, 15, pp. 283-290.

Yin, R.K. (2011) Qualitative Research from Start to Finish, New York: Guildford.

Yin, R.K. (2012) Applications of Case Study Research, 3rd edition, Thousand Oaks, CA: Sage.

Yin, R.K. (2014) *Case Study Research: Design and Methods*, London: Sage Publications Ltd.

Appendix 1: Dissertation Proposal:

Investigating the role of the discovery layer in biodiversity research support by the Library and Informatics Department at the Natural History Museum, London: A Case Study

S.A. Stewart

Supervisor: Dr. Lyn Robinson

Introduction

The study of biological diversity, the variation of all life on earth, has resulted in a plethora or "deluge" of rapidly growing heterogeneous information, particularly as advances in high-throughput genetic and environmental technologies enable data to be obtained at a greater frequency. This paradigm shift in the study of biodiversity is echoed throughout the biological sciences as a whole, which are becoming increasingly data-intensive and data-driven "big science" (Patterson et al., 2010).

Biodiversity information comes in a variety of types, from geospatial information to genetic sequence data, museum specimens and scientific literature, in the form of journal articles, monographs, artworks and field and laboratory notebooks. This information has often been physically disparate, and researchers would often have to visit several institutions and collections, often in different countries, to gather the information required for their work. Furthermore, biodiversity information is unique in that it is historical, and therefore does not quickly become outdated, even if names of taxa are changed. Historical or legacy literature is therefore still an important source of information for researchers.

Biodiversity information, in spite of its importance, is often difficult to manage and access; however, with the advent of the internet and subsequent extensive digitisation not only of scientific literature, but also specimens held in museum collections, biodiversity information is becoming increasingly accessible to researchers.

The Natural History Museum (London, UK) is one of the leading research institutions for the study of biodiversity. It houses over 79 million specimens, and has a Library and Archives with one of the largest collections of natural history literature and artworks, which number over 1 million. The library is currently in the process of changing over many of its journal subscriptions to on-line access only, and will also launch a new library management system in July 2014, a combination of Ex Libris Alma and Primo, which offer the possibility of a discovery layer for finding information in a single search, rather than consulting several different systems in order to find various types of resources.

The Natural History Museum also has an informatics department, which is working to "aggregate biodiversity data from multiple sources" and provide a "single, searchable set of data". This data portal will also launch in 2014.

This study will be a case study of biodiversity information discovery and access through the interfaces created by both the library and the informatics department at the Natural History Museum. It will compare and contrast the research approaches to the biodiversity information held by both the

library and informatics collections and explore how these interfaces are used by researchers and information practitioners, the librarians and informaticians, at the Natural History Museum.

Most significantly, this dissertation project will document the transition towards a digital, networked aggregation of biodiversity information, which is of great significance not only to the Natural History Museum itself, but also to biodiversity researchers globally.

Aims

The principal aim of this project is to investigate, compare and contrast the role of the discovery layer in biodiversity research support at the Natural History Museum as mediated through both the Library and the Informatics Department, and to document the transition towards an integrated digital information system for biodiversity across the Natural History Museum as a whole.

Objectives

In order to meet the aim of the study, this project will involve the following objectives:

1.) Determine exact attitudes to searching for biodiversity information resources using the discovery layers of the library and the informatics data portal.

2.) Explore how biodiversity information is discovered and used between the library and the informatics department.

3.) Document the transition towards a digital, networked and aggregated discovery layer approach to managing biodiversity information.

Preliminary Literature Review

The study of biodiversity involves information drawn from numerous and varied sources, including museum specimens, ecological niche data, genetic sequence data, and a vast and varied literature spanning centuries, from illustrated accounts of voyages of discovery to the most recent article published in a journal. These sources of information about the variety of life are often disparate, and researchers will often have to visit multiple institutions in order to obtain the necessary information for their work, often necessitating travel to other countries. With the rise of cybertaxonomy, web-based taxonomy and systematics; the development of data portals for various taxonomic groups, and mass digitisation projects, not only of scientific literature, such as the Biodiversity Heritage Library (BHL), but also of museum specimens, often important type specimens held in museum collections, biodiversity information has become increasingly accessible for researchers. In spite of its increasing availability, the sources for this information are still disparate, and require searching through many on-line resources in order to find information to fulfil an information need. Therefore, the development of a "discovery layer", literally, a web-based, "preharvested central index…providing a single search across a library's local, open access and subscription collections" (Hoeppner, 2012) will enable natural language searching and will cover numerous information resources in a single search. The discovery layer is the user

interface and search system which will enable the user to interact with the information resources contained in the library system (Hoeppner, 2012).

As biodiversity becomes digital, the discipline of cybertaxonomy has emerged to organise and manage web-based biodiversity information resources in such a way that the information derived from natural history collections is made accessible. There have been numerous attempts to create an on-line single-search for taxonomic and biodiversity information, ranging from Universal Biological Names (http://www.ubio.org/), to the Encyclopedia of Life (EOL) (http://eol.org/) and Global Biodiversity Information Facility (GBIF) (http://www.gbif.org/) and the Tree of Life Web Project (http://tolweb.org/tree/). These web-based biodiversity information portals are a radical departure from previous ways of conducting research in biodiversity, where a great deal of researchers' time is taken up with tracking down historical or legacy literature and specimens in disparate museums, libraries and archives (Hine, 2012).

The Natural History Museum in London is one of the foremost institutions in the world for biodiversity research. It contains over 79 million specimens and contains one of the largest natural history library collections in the world with over 1 million items catalogued. To reflect the changes taking place throughout the biological sciences as a result of advances in technology and the internet, The Natural History Museum is implementing digital initiatives as part of its next 12-year strategic plan.

The Natural History Museum library and archives contain a range of information resources, from journals and monographs to artworks, manuscripts and notebooks. Until quite recently, the library has not had an electronic catalogue, and items in the library collections have only been barcoded for inclusion in the library's on-line catalogue as recently 2012. The library is currently undergoing another period of change as it implements a new library management system, Ex Libris Alma, which will replace the former SirsiDynix Symphony catalogue and separate e-resources databases. Alma will be integrated with Primo, which will provide a discovery layer enabling a unified search of the libraries local information resources in addition to web-based resources and "community" resources held by other institutions.

Case studies of natural history museum libraries are extremely rare, and, to date, no previous studies have been found that explicitly detail the role and use of the discovery layer within the context of biodiversity information in a natural history museum library.

The recently-formed Informatics Department is currently in the process of developing a data portal, though which data from over 79 million specimens will be made accessible on an open access platform. This data portal will also enable exploration and visualisation of data through a "clean and intuitive interface" (NHM Informatics website, 2014). The data portal will also enable users to access sub-portals to other museum collections, browse geospatial specimen information and upload their own data sets, whilst also providing persistent (NHM Informatics website, 2014). The Informatics Department has also been instrumental in developing Scratchpads, an open-source system for the creation of web-based databases integrating several sources of information including geospatial, literature and through the creation of taxonomies.

While both the library and the informatics department manage and organize biodiversity information, their strategies differ; however they are similar in that both will utilise a discovery layer in order to make their information accessible to users. This study aims to investigate how the discovery layer will be used in both the library and the informatics department to make biodiversity information accessible to its community of researchers.

Scope and Definition

The research and study comprising the dissertation will take place only within the Natural History Museum, although, should the project be successful, further exploration of the discovery and use of biodiversity information amongst the other institutional members of the natural history museum consortia (eg. Biodiversity Heritage Library partners or data portal partners) would be of considerable interest. Participants in the dissertation research will include research scientists and curators who utilise the informatics and library resources, in addition to the information practitioners, the librarians and informatics department team members who are developing the Natural History Museum Data Portal.

Methodology

Throughout the course of the dissertation work, the methodology will be informed by literature as much as possible. Therefore, a through literature review is necessary. The literature review will be on-going throughout the dissertation.

A survey will be developed to obtain data on how the discovery layers for both the library and informatics databases will be accessed and searched. A pilot survey will be tested beforehand in order to ascertain that the questions provide appropriate data and that the survey questions are clearly written and easily understood. Finally, interviews will be conducted with selected participants to clarify and add depth to the data collected from the survey.

Ethical and Legal Considerations

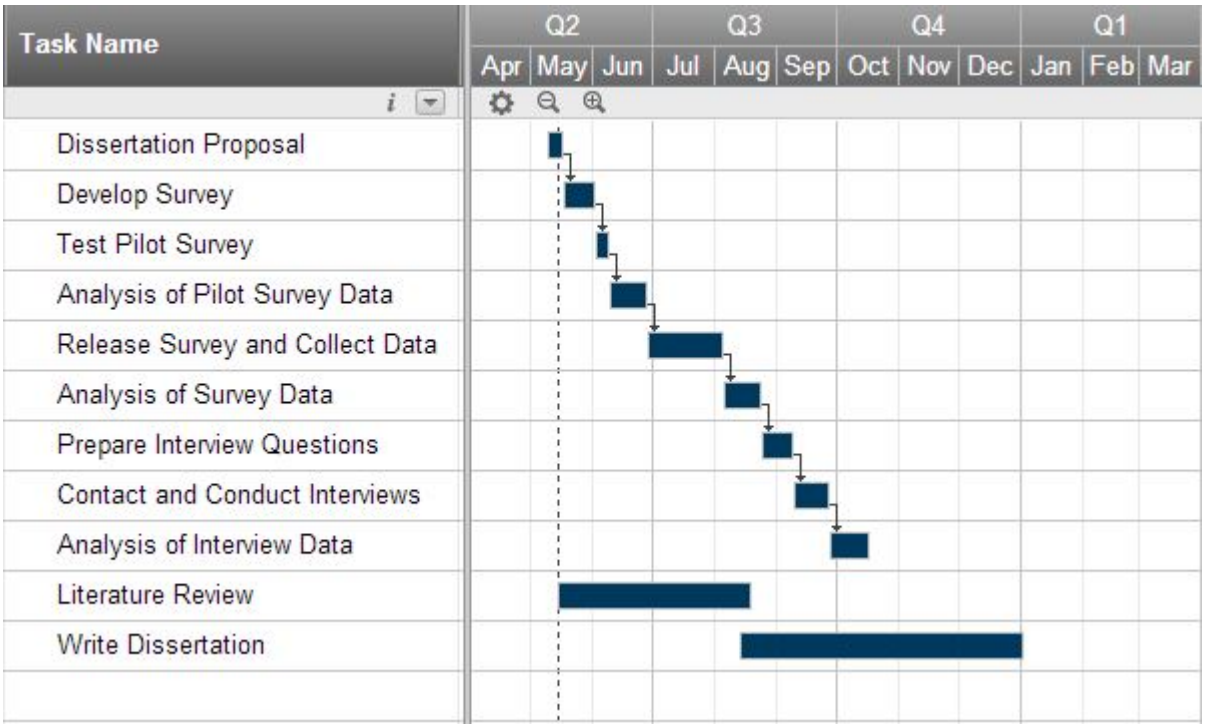
It is not expected that there will be any ethical or legal problems, as no sensitive data will be collected, and the dissertation project and its methodology will be fully disclosed to participants; however, it might be the case that some personal data (names, employee status, length of time working for the Natural History Museum) will be collected in the course of both the survey and the interviews. This data will be kept in a secure database, and will be anonymized. Participants will also have the opportunity to withdraw from the study at any time.

Resources and Materials

No special equipment will be required for either the survey or the interviews. The on-line survey will be mediated through Survey Monkey (http:www.surveymonkey.com), a free, on-line survey tool.

Interviews will be conducted face-to-face whenever possible, at the convenience of the interview participants.

Project Timeline



The Gantt Chart above visualises the dissertation project timeline.

References

Case, D.O. 2012. *Looking for Information: A Survey of Research on Information Needs, Seeking and Behaviour*, 3rd Ed. London: Emerald Group Publishing Ltd.

Chu, H. 2010. *Information Representation and Retrieval in the Digital Age*, 2nd Ed, Medford, New Jersey, U.S.A.: Information Today, Inc.

Graham, C.H., Ferrier, S., Huettman, F. Moritz, C. & Townsend Peterson, A. 2004. New developments in museum-based informatics and applications in biodiversity analysis, *Trends in Ecology and Evolution*, Vol. 19(9), pp. 497-503.

Hine, C. 2008. *Systematics as Cyberscience: Computers, Change and Continuity in Science*, Cambridge, MA: MIT Press.

Hoeppner, 2012. The Ins and Outs of Evaluating Web-Scale Discovery Services. *Information Today*, On-line. Available: [http://www.infotoday.com/cilmag/apr12/Hoeppner-Web-Scale-Discovery-Services.shtml], accessed May 14, 2014.

Lister, A. & Climate Change Research Group. 2011. Natural history collections as sources of long-term datasets, *Trends in Ecology and Evolution*, Vol. 26(4), pp. 153-154.

Natural History Museum Informatics Department. 2014. On-line. Available: [http://www.nhm.ac.uk/research-curation/research/informatics/index.html] Accessed: May 14, 2014.

Patterson, D.J., Cooper, J., Kirk, P.M., Pyle, R.L. & Remsen, D.P. 2010. Names are key to the big new biology, *Trends in Ecology and Evolution*, Vol. 25(12), pp. 686-691.

Pickard, A.J. 2007. *Research Methods in Information*, London: Facet Publishing.

Soberon, J. 1999. Postscript: Linking biodiversity information sources, *Trends in Ecology and Evolution*, Vol. 14(7)

Appendix 2: Natural History Museum Digital Strategy Internal Communication

The following is the text of an internal communications document from the staff intranet of the Natural History Muesum:

Digital Strategy programme update (Autumn 2013)

What do we mean by digital?

Digital is a very broad term covering:

- technology infrastructure and devices
- software applications and platforms
- content delivery channels

which enable new or enhanced ways of:

- communicating
- interacting
- collaborating
- storing and sharing information

What is the Digital Strategy programme?

Digital services and technology infrastructure are increasingly central to the Museum for supporting and enhancing:

- public engagement
- scientific research, collaboration and collections management
- everyday technology needs of Museum staff

The Digital Strategy programme was set up to:

- enable the Museum to exploit societal and technological changes
- provide a roadmap for delivering the Museum's digital vision
- initiate, support and resource significant digital projects and activities

It aims to:

- improve the planning and scheduling of digital projects
- support communications and decision-making on digital matters

Why is it important?

The digital world is fast changing and requires continuous forward-planning in order to prepare staff and technology infrastructure alike to meet future demands. Colleagues from across the Museum are working together to deliver exciting projects such as the digitisation of collections, redeveloping the Museum's websites and providing Public WiFi. There are also many more digital activities that the Digital Strategy is supporting that are required to improve and maintain the Museum's technology services.

The Digital Strategy programme has a Museum-wide view of digital projects and activities in order to:

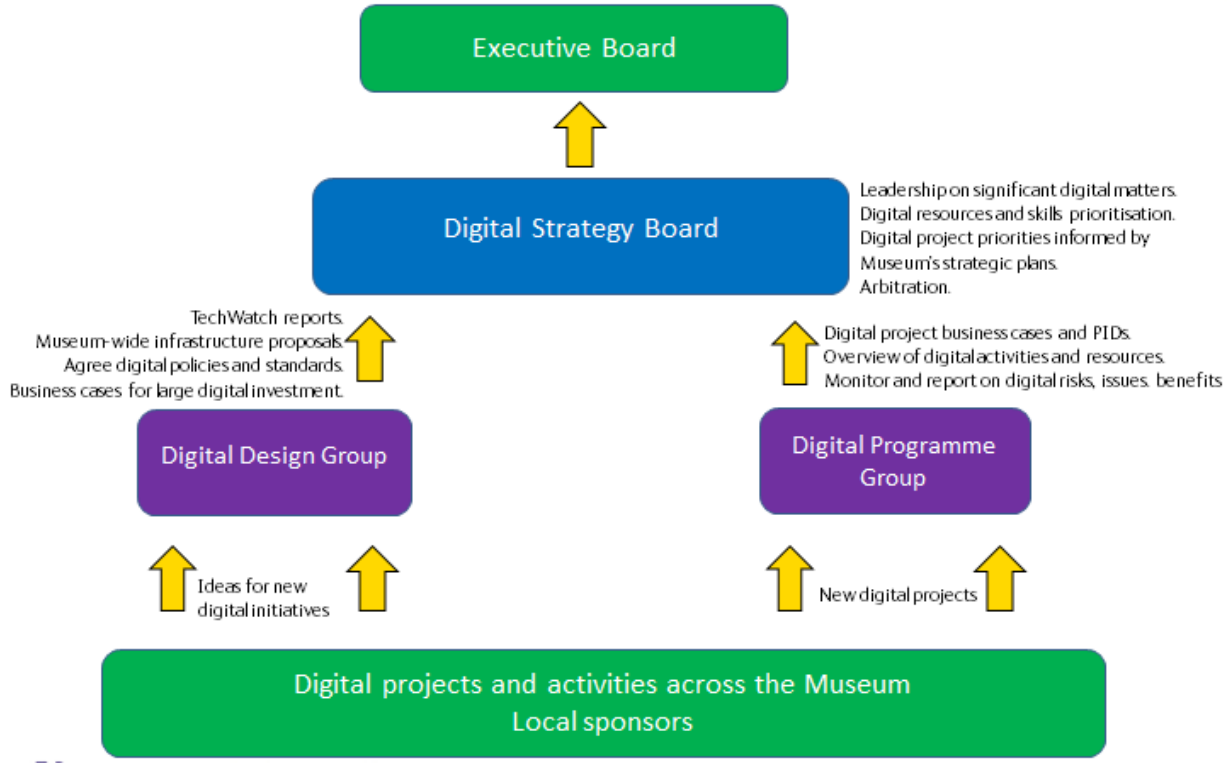
- better understand and use expertise available within the Museum
- identify digital resource commitments and pinch-points
- provide support for emerging projects and established digital projects
- maintain existing digital service whilst planning for digital change

The Digital Strategy programme builds on existing relationships between the digital resources across the Museum in order to provide:

- cross-representation of all the Museum's digital stakeholders
- a forum for discussion and issue resolution
- early awareness of new technologies

- early awareness of resource/ skills shortages
- momentum for change and innovation

How is the Digital Strategy programme run?



Digital Strategy Board provides Museum-wide leadership and arbitration on significant digital matters and is responsible for commissioning and sponsoring major investment in digital initiatives.

Members: Ailsa Barry, Lincoln Borgartz, Neil Greenwood (SRO), Justin Morris, Ian Owens, Vince Smith, Dave Thomas.

Digital Design Group constantly evaluates emerging technologies, providing suggestions to the Digital Strategy Board on investment in digital infrastructure to align Museum digital objectives and suggest how best to develop digital skills to meet expected future demands. The group promotes the sharing of knowledge and looks for opportunities for digital resources across the Museum to work together.

Members: Jez Burn, Alan Hart, Ken Johnson, Ben Rakison, Sheila Sang, Darrell Siebert, Chris Sleep, Vince Smith ,Ivan Teage, Dave Thomas, Julian Thomas, Annette Ure.

Digital Programme Group is concerned with the scheduling and funding of the numerous digital projects and activities across the Museum and identifying resource pinch-points.

Members: Lincoln Borgartz, Victoria Carter, Richard Hinton, Siorna McFarlane, Ivan Teage.

What does it mean?

All staff

- More effective technology and digital services
- Access to technology and digital services more appropriate to your needs

Digital staff

- More opportunities to work with leading edge technology
- Utilise transferable skills and develop new digital skills
- Work with different staff across the Museum on exciting digital projects

Project Owners/ Managers

- A clear route and process to evaluate and progress new digital projects
- Clearer view of potential skills and resources requirements and challenges

Management

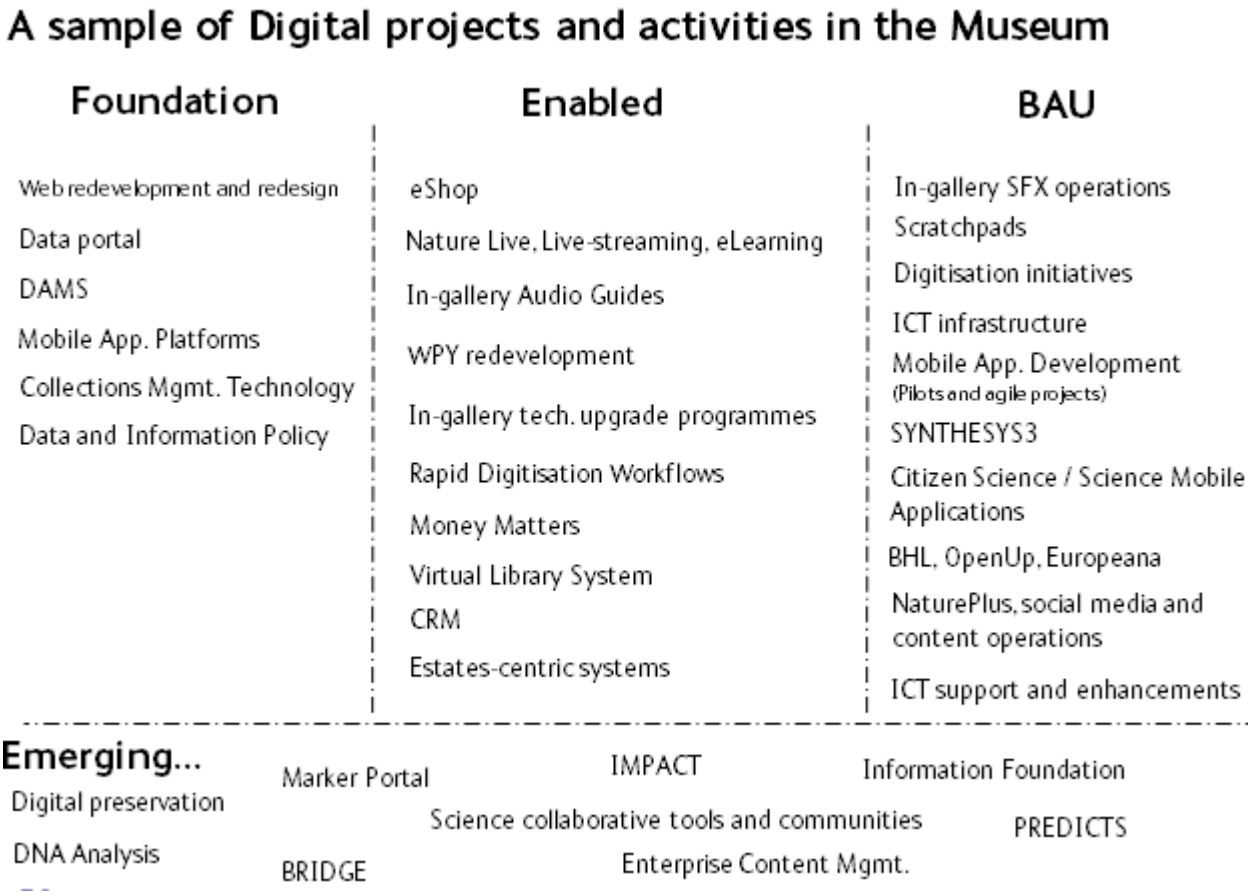
- Transparency of projects, plans and digital resources
- Cross-Museum prioritisation

Upcoming activities

The programme is currently working to deliver the following:

- Web redevelopment and redesign

- Data portal
- Digitisation initiatives
- Information /Data policy
- DAMS review
- Public WiFi rollout
- eShop



Appendix 3: Project Information Sheet and Consent Forms for Interviews

Project Information Sheet

The following is an example of the information sheet that was distributed to participants in the interviews for the dissertation work:

City University London, Department of Library and Information Science

Masters Dissertation Project Information Sheet

Informational Interfaces: Impact of Data Portals and Discovery Systems on Biodiversity Research at the Natural History Museum: A Case Study

Sarah A. Stewart

Outline

Research in biodiversity involves working with information resources that are often complex, heterogeneous and disparate. With the rise of the World Wide Web and the recent emergence of virtual library and museum environments, this information may become increasingly "findable". At the same time, advances in library management systems to incorporate discovery systems involving single-search interfaces may impact how search and information retrieval are conducted, potentially also impacting research. This investigation will examine the use and impact of the Natural History Museum Library and Archives discovery system, Ex Libris Primo, implemented July 2014, and the Informatics Data Portal which is set to launch publically in January 2015, on search and research by scientific staff involved in biodiversity research at the Natural History Museum. It will also build a case study, though interviews with key contact staff members, documenting the implementation and impact of the discovery system in the Natural History Museum Library and Archives.

Interviews

Two types of interviews will be conducted for the purposes of the dissertation project.

The first, with library users, will primarily investigate how they search for information using the Library and Archives discovery system, Primo, or the Informatics Data Portal. The only information that will be made public may be comments and quotations, and the users' identity, including name, kept anonymous. Consent will only be required to use quotations and comments during these interviews.

The second type of interview will be conducted with key contact staff involved in the implementation and design of the systems or workflows under investigation. In these interviews, consent is required, as identifiers such as name and job title may be published in conjunction with quotations and comments.

Ethics

A consent form will be provided along with this information sheet. Please sign the form if you agree to the use of any information for the purposes of the dissertation research. Please be aware that participation in the research is voluntary, and that you have the right to withdraw from the study at any time; however, any information that you submit before withdrawing may be used in the study. Please contact the investigator, Sarah Stewart, (sarah.stewart2@nhm.ac.uk) should you have any questions or concerns.

Thank you for taking part in my Masters dissertation research.

Interview Consent Form

The following is an example of the consent form that was distributed to interview participants:

Title of Study: *Informational Interfaces: Impact of Data Portals and Discovery Systems on Biodiversity Research at the Natural History Museum: A Case Study*

Principal Investigator: Sarah Stewart, Department of Informatics, City University London

Please initial box

1.	I agree to take part in the above City University London research project. I have had the project explained to me, and I have read the participant information sheet, which I may keep for my records. I understand this will involve [*researcher to add/delete as appropriate prior to use*]: • be interviewed by the researcher • complete questionnaires asking me about professional practices. • make myself available for a further interview should that be required	
2.	This information will be held and processed for the following purpose(s): I understand that I have given approval for my name and/or the name of my workplace to be used in the final report of the project, and future publications.	

	I understand that confidentiality cannot be guaranteed for information which I might disclose in the focus group(s)/group interviews(s). I consent to the use of sections of the interviews in publications.	
3.	I understand that my participation is voluntary, that I can choose not to participate in part or all of the project, and that I can withdraw at any stage of the project without being penalized or disadvantaged in any way.	
4.	I agree to City University London recording and processing this information about me. I understand that this information will be used only for the purpose(s) set out in this statement and my consent is conditional on the University complying with its duties and obligations under the Data Protection Act 1998.	
5.	I agree to take part in the above study.	

_____ _____ _____

Name of Participant Signature Date

When completed, 1 copy for participant; 1 copy for researcher file.

Appendix 4: Sample Interview Questions

The following questions were utilised to guide interviews with both system users and key informants. As the interviews were semi-structured and questions were open-ended with the goal of prompting interview participants to discuss their interactions with the discovery interfaces and systems, there was no rigid structure imposed on the interview process.

1.) Do you envision the library search system as the starting point for research?
2.) Do you think that the new library discovery system will improve access to the collections?
3.) What impact do you see the library discovery system having on research?
4.) What impact do you see the new discovery system having upon collections management in the library / or in the museum?
5.) What are your thoughts on the transition towards a more digital environment for research?
6.) What aspects of the library discovery system do you find most useful / not useful?
7.) How do you think the Library discovery system will support research in biodiversity at the Museum?
8.) How do you use the discovery system to search for information?
9.) Do you think that the discovery system has changed the way that you search for information?

For Library Staff (Key informant interviews):

1.) What impact do you see the discovery system, Primo, having upon research in the Library?
2.) Are there any implications for how the collections are managed?
3.) Has there been a change in usage of the collections?
4.) Do you see any changes to users behaviour when looking for information in the Library?
5.) How do you think users are searching the collections using the discovery system?

For Dr. Vincent Smith (Data Portal):

1.) What do you predict the impact of the Natural History Museum Data Portal will be on research at the Natural History Museum?
2.) What will the Data Portal wider impact/implications be on biodiversity research?
3.) Do you forsee an impact on search?
4.) What are your thoughts on the transition of the Museum from physical collections to informational?

Appendix 5: Dissertation Reflection

The dissertation research provided me with the opportunity to carry out some original research related to my interests in information management in biodiversity research and biodiversity informatics, and with relation to my place of employment as an Information Assistant at the Natural History Museum Library and Archives.

I hope that my employers may find benefit in my work; I have personally benefitted from the dissertation research through the research process itself, in learning how to plan and conduct library science research. It has also enabled me to combine theory and practice and has elucidated how the theoretical material learned through the Masters course at City University can be applied to practical, "on-the-job" situations, such as conducting surveys to gauge attitudes to library systems and services, such as the discovery system at the Natural History Museum Library. Furthermore, conducting research utilising a case study and gathering evidence through different means such as the surveys and interviews has now provided me with the conceptual framework and the confidence to conduct similar research should I need to do this in future. The research involved in this dissertation work enabled me to learn more about information search and retrieval, systems usability and discovery systems more generally. Most importantly, this work was beneficial in that I was able to observe the implementation and launch of a library management system and was able to observe some of the planning and project management involved in co-ordinating this work.

I believe that under the circumstances and within the time permitted, I have conducted the research work effectively and have met most of the aims and objectives of the dissertation research. With the exception of the late launch of the Data Portal and the consequent advised abandonment of the Data Portal surveys and interviews, I have still been able to collect enough data on the Library and Archives discovery system to meet all of my objectives and stated research aims. Although I have not been able to capture very much data from the Data Portal, a survey instrument has been prepared that may prove useful for future research in this area. Furthermore, the information gained from the interview with Dr. Vincent Smith and the system overview with Mr. Ben Scott, the systems architect, have been very informative and insightful with regard to how they expect the system to be used, and that the Data Portal will not be only a tool for discovery within the Natural History Museum, but will essentially provide a platform for datasets generated by the Natural History Museum and its collections, making this data available to a wider audience, and producing a benchmark in the transition of the Natural History Museum towards an informational virtual museum environment.

While the survey has provided useful points to my research and possibly to my employers, the low response rate introduces self-selecting biases, so reliance on the data obtained from the survey is uncertain. Predicting the response rate prior to releasing any survey is difficult; therefore, a better method of gauging information –seeking behaviour might have been to run usability tests with small groups of participants in the library setting, as would normally occur during a library induction. This may be a possibility for future research.

From the insights gained from the survey, and interviews with Library users and staff, I have been able to build a case study of the impact of the discovery system on search and research at the Natural History Museum Library, uncovering patterns of information-seeking behaviour among the researchers and providing a reference point whereby recommendations could be made as the uptake of the discovery system unfolds. This is also the first study of the implementation of discovery system in a Natural History Museum Library setting. I hope that this study will be useful to the Library and Museum staff and to any future researchers.

The dissertation process has been a positive experience overall, enabling me to develop the necessary research skills required to assess and evaluate a library system or service in light of its use and uptake. The dissertation process has also enabled me to expand my existing communication skills through arranging interviews and reporting on my research findings in this dissertation, and, most importantly, has furthered my knowledge and experience of teamwork involved to carry out large projects such as the development and launch of the Library and Archives discovery system or the Data Portal.