

Project Title: Darwin's Library
 Version: 01.0
 Contact: G Young
 Date: 15-Nov-09



Project Document Cover Sheet

Project Information			
Project Acronym	–		
Project Title	Darwin's Library		
Start Date	1 September 2009	End Date	28 February 2011
Lead Institution	University of Cambridge (Cambridge University Library)		
Project Director	Grant Young (UK Project Director)		
Project Manager & contact details	Grant Young, Digitisation and Digital Preservation Specialist Cambridge University Library West Road, Cambridge CB3 9DR		
Partner Institutions	Natural History Museum (NHM) American Museum of Natural History (AMNH) Biodiversity Heritage Library (BLH)		
Project Web URL	http://www.lib.cam.ac.uk/projects/		
Programme Name (and number)	JISC/NEH Transatlantic Digitisation – Phase 2		
Programme Manager	Alastair Dunning		

Document Name			
Document Title	Project Plan		
Reporting Period	–		
Author(s) & project role	Grant Young (UK Project Director), David Kohn (US Project Director), Jane Smith (NHM and BHL)		
Date	15-Nov-09	Filename	DarwinLibraryPlan01.pdf
URL	–		
Access	<input type="checkbox"/> Project and JISC internal		<input checked="" type="checkbox"/> General dissemination

Document History		
Version	Date	Comments
01.0	15-Nov-09	



JISC Project Plan

Overview of Project

1. Background

This 18-month project will construct an online version of Charles Darwin's working library, enabling researchers, students and the interested public to study the works that most directly informed his thinking. Many of the books in Darwin's possession are now rare, so the digital library will provide an important service to scholars by allowing them to access books he refers to in his own published works, notebooks and correspondence. However the digitisation of Darwin's library provides much more value than this – it offers important insights into the way Darwin read and engaged with the works of others.

The project will bring together existing and new digital content from both sides of the Atlantic in order to re-present Darwin's library as a major digital resource for the use of scholars. Cambridge University Library (CUL) will digitise the most heavily annotated volumes from Darwin's library, depositing these as a special sub-collection within the Biodiversity Heritage Library (BHL), a major international initiative to digitise and make freely accessible the world's biodiversity literature. Editions that Darwin owned which have already been made available within the BHL will be identified and marked out for inclusion within this sub-collection. The UK Natural History Museum (NHM) is a contributor to the BHL and will prioritise the digitisation of editions from its collection that are also found in Darwin's library. The American Museum of Natural History (AMNH) has access to the transcriptions of Darwin's annotations through the participation of the Darwin's Marginalia editors in the Darwin Manuscripts Project (DMP), which will contribute these to the BHL so they can be displayed alongside the digital facsimiles and surrogates.

Were Cambridge University Library or another party to attempt to digitise every single page of every work owned by Darwin and then preserve and make that data available via the Internet, the enterprise would take many years and cost many millions of pounds. The approach adopted by this project provides a pragmatic alternative, but one that avoids compromising the data. Through its combination of digitising originals, surrogates and co-opting existing content, this project will be able to build half of Darwin's Library within the time and budget constraints of this grant and it will undertake fundraising for a second phase to complete the remainder of the library. The project will deliver a significant and substantial resource for scholars. By leveraging the content, skills, and infrastructure from its partners, it will do so with great efficiency. In placing the content within the structure of the BHL it will ensure that the collection is sustainable and easy for scholars to discover, access and use. And by combining the digital facsimiles and surrogates with the transcriptions of Darwin's annotations prepared by the AMNH, it adds considerable value to both sets of data.

The significance of Darwin's Library as a part of Darwin's archival legacy was first recognised a century ago by his son Francis: "The chief interest of the Darwin books lies in the pencil notes scribbled on their pages, or written on scraps of paper and pinned to the last page."¹ Charles Darwin read with a pencil in his hand, adding thousands of comments and marks to the books on his shelves. He filled their inside covers with his notes, and inserted pieces of paper when he ran out of space in the margins. He would come back again and again to some texts, adding further comments and marks.

¹ H. W. Rutherford, *Catalogue of the library of Charles Darwin now in the Botany School, Cambridge*. Introduction by Francis Darwin. Cambridge: Cambridge University Press, 1908, p. viii.

The importance of Charles Darwin's working library is the very concrete insights that it gives us into Darwin's practice as a scientist and Victorian intellectual. In fact, the library's significance is materially linked to the books' condition. Many volumes are quite worn by their owner's heavy use. As Francis observed, Darwin "hardly ever had a book bound, and the collection retains to a great degree its original ragged appearance." Moreover, Darwin's pencil annotations, on soft paper, are often difficult to read both because of his handwriting and because of the smudging and abrasion caused by Darwin's use and reuse. Darwin's reading practice included making detailed lists and abstracts of pages with marginalia. Yet these well-worn volumes, with their often cramped and blurred marginalia, constitute precious documents that throw light on the intricate development of Darwin's evolutionary theory and its application to every aspect of natural history. Often, these notes in books written by other people provided the raw material out of which Darwin fashioned his own books. Often they highlighted the obstacles he knew his theory must face.

To fully gauge the significance of this project to reconstruct Darwin's working library online, it is necessary to put Darwin's books and their scribbles into context. Darwin's systematic habit of annotating his books formed one – albeit only one – of the largest elements of the surviving, massive record of Darwin's practice as a reader. That, in turn, was but one of the key elements of Darwin's scientific practice. Among other things, his reading supported a practice we might today call meta-analysis, where he would test crucial theoretical premises by compiling and tabulating evidence derived from his reading. A good example is the discussion of variability at the end of chapter 2 of the *Origin*, in which Darwin boiled down information, originally compiled in long tables, drawn from numerous monographs. His immediate intent was to relate patterns of geographic distribution to patterns of taxonomic variability. But his deeper objective was nothing less than to discriminate among competing explanations for how new species arise in nature – an issue unquestionably at the heart of his theory and essential to its overall plausibility and reception. This use of reading placed it on an equal footing with the other elements of his scientific practice: his own experimentation and observation, his private speculation, and his extensive communication with others through correspondence and conversation. In fact, all these elements have in common a style of disciplined note taking – a habit first pursued with full rigor on the *Beagle* – combined, more and more as the years passed, with the equally disciplined extraction of salient ideas and insights, which Darwin interpreted in the light of the interwoven 'secondary laws' and generalisations that his theory provided. This intense style of reading and thinking was the sort of thing Darwin meant when he once complained that his mind had become a machine for the production of facts. But that was only one of Darwin's reactions to this kind of reading and experimenting, and it rather belied the great excitement he often derived from such grinding, disciplined work. For those interested in the cognitive processes of imaginative, hard-working intellectuals and artists, Darwin's annotations, once they are digitised, online, and searchable, will offer a large and in many ways representative body of primary data for interpreting how this scientist thought and how he created knowledge.

Darwin's reading has been a subject of humanities interest. Dame Gillian Beer has explored his reading notebooks, where she discovered the depth and range of his general, particularly literary reading.² But the marginalia have yet to be examined from the perspective of the history of the book. Indeed, while Secord's *Victorian sensation*³, offers a book history of the *Vestiges of creation* (1845), a widely influential semi-popular book on evolution, published anonymously by Richard Chambers there is no book history of the most widely disseminated evolutionary book ever published, Darwin's *Origin of species* (1859). The digitally reconstituted Darwin's Library will be a *sine qua non* to such a major undertaking. By unlocking the possibility of understanding Darwin as a reader, scholars also stand to gain tremendous insight into Darwin as an influential writer, for he is well known to have strategically shaped his book to his expectations of his intended readers.⁴ Thus, much of both Darwin's practice

² Gillian Beer, *Darwin's plots: evolutionary narrative in Darwin, George Eliot, and nineteenth-century fiction*. London; Boston: Routledge & Kegan Paul, 1983.

³ James A. Secord, *Victorian sensation: the extraordinary publication, reception, and secret authorship of Vestiges of the natural history of creation*. Chicago: University of Chicago Press, 2000.

⁴ David J. Depew 'The rhetoric of the Origin of species,' in *The Cambridge Companion to the 'Origin of Species'*, Michael Ruse and Robert J. Richards, eds. Cambridge: Cambridge University Press, 2009, pp. 237-255.

and his reception remain unexplored and unexplorable because of the lack of basic access to Darwin's Library. This remains true even though DiGregorio and Gill published their work, *Charles Darwin's Marginalia*,⁵ in 1990 and it is a standard tool for all Darwin specialists. This monumental achievement in deciphering and transcribing the toughest of Darwin's handwriting is, however, a product of its pre-internet times. It is hard to use because you must have to hand a copy of the exact edition Darwin owned. Otherwise, you will be faced with a wall of codes that is as distracting from the context of the original work as Darwin's handwriting is difficult to decipher. The resource created by this project will provide a pragmatic, cost efficient, way to reveal whole new strata of Darwin manuscript material and will make it truly accessible to a broad range of scholars and to the educated public.

2. Aims and Objectives

As described in section 1, the broad aim of the project is to **provide online access to a substantial portion of Darwin's Library, displaying the full texts of the books alongside images and transcriptions of his annotations.**

In pursuit of this aim, the project has the following **specific objectives**:

- **Digitise originals and surrogates, and co-opt existing digital content, to represent a significant portion of Darwin's Library.** The volumes and page numbers are still being calculated, but are expected to include approximately 300,000 pages digitised by Cambridge University Library and the Natural History Museum or other BHL partners.
- **Provide the digital representation of Darwin's Library with sustainable preservation and delivery.** The content will be deposited within the Biodiversity Heritage Library (BHL), which takes advantage of preservation and delivery services of the Internet Archive.
- **Enable sophisticated resource discovery and access to the collection.** Darwin's Library will enjoy its own distinct identity as a sub-collection of the BHL, but also be discoverable through the standard BHL search. Further access points will include the AMNH's Darwin Manuscript Project catalogue and the library catalogue of Cambridge University Library.
- **Provide a pilot and model for a second phase of the project.** The JISC/NEH grant will enable more than half of Darwin's Library to be presented online. Fundraising for its completion will be greatly assisted by the existence of the content and infrastructure established during this current project.

3. Overall Approach

3.1. Strategy / Methodology

3.1.1. Identification of books to be scanned or selected (see Workpackage 2)

The funding available for this project will not enable the scanning of Darwin's entire library of 1476 works, or even the 706 with marginalia. The project bid envisaged a two-stage approach, with this current project as the first stage. The bid suggested scanning the most heavily annotated works from Darwin's own copies at Cambridge and providing surrogates from NHM or other BHL libraries. In some cases this would mean displaying the transcriptions of the marginalia alongside the 'clean' pages of surrogates.

⁵ Mario A. Di Gregorio and N. W. Gill, *Charles Darwin's Marginalia*. New York and London: Garland, 1990, vol. 1.

When the project was awarded we undertook a re-think of this approach to selection and display. This was required by lower than anticipated funding awarded on the UK side of the project, but also by careful consideration of what would be the most appropriate and beneficial approach for scholars. With our programme manager's approval, we have adopted a revised approach:

- Cambridge will prioritise the digitisation of entire works with the highest proportion of Darwin annotations
- Cambridge will additionally digitise individual annotated pages from works with minimal annotation that are available from the NHM or are already held within the BHL
- NHM will prioritise digitisation of surrogates for the works with minimal annotation, so that the individual page scans from Cambridge can be incorporated into full works.
- NHM will additionally seek to identify BHL books that Darwin made separate notes about, but did not mark, since he borrowed rather than owned them. We intend to present transcriptions of his notes alongside these books, enabling the project to extend beyond Darwin's Library and represent something of the broader unit of "Darwin's reading".

Adopting these approaches, we will be able to deliver more content than initially proposed, and show the transcriptions alongside the actual annotated pages rather than 'clean' surrogates, which will be of greater benefit to scholars. Extending the project to encompass Darwin's wider reading is also more intellectually satisfying, since these works are of arguably greater importance than books on Darwin's shelves that show no indication of engagement.

We have already developed a full inventory of Darwin's Library, with detailed metrics of page annotations. But before the full scanning lists can be drawn up for Cambridge and NHM, it is necessary to establish: (1) which digitised works are already available within the BHL; (2) which physical works are available within the NHM; and (3) of the NHM works, (a) which must be excluded due to their condition, (b) which must be scanned via the NHM studio (a more expensive process), and (c) which can be scanned via the Internet Archive/BHL scanning 'pod' (a less expensive process). Books must be carefully checked to confirm they are the same edition as the copies Darwin held.

Once the NHM has completed this work, we can finalise the list of Cambridge works to be scanned (a) in their entirety (less expensive per page) or (b) selectively (more expensive per page).

3.1.2. Digitisation of Darwin's Library originals by CUL (see Workpackage 3)

Once the scanning lists have been finalised, Cambridge University Library will begin scanning of Darwin's original copies. This will be handled within the Library's Imaging Services unit, which has a mix of camera and scanning equipment capable of capturing all but the most challenging of materials. The work will be overseen by the Keeper of Scientific Collections, who is responsible for this collection, and members of the conservation department will be on hand to support the process.

The selective page scanning, will be carefully timed to follow the checking work being undertaken in 3.1.4. below. As the editor has checked a page of transcriptions, he will put a slip within the volume to indicate which page needs to be captured.

3.1.3. Digitisation of surrogates by NHM (see Workpackage 4)

Once the scanning lists have been finalised, the Natural History Museum will also begin its scanning. Books will be routed via one of two workflows: (1) the Library's Digitisation Unit, which uses specialist camera and scanning equipment to capture valuable or difficult works; or (2) the BHL/Internet Archive pod, which adopts a faster, cheaper capture methodology.

3.1.4. Preparation of the annotations by AMNH (see Workpackage 6)

Darwin's marginalia has already been transcribed and exists electronically. In preparation for republication within Darwin's Library, they are being checked against the originals and additional indexing terms added to aid retrieval and display.

3.1.5. Development of metadata by AMNH (see Workpackage 6)

Metadata will be used to provide three main points of entry to the collection: (1) via BHL, (2) via the AMNH's DARBASE catalogue, and (3) via the Cambridge University Library catalogue.

Metadata has already been created for the books in Darwin's Library as part of the DARBASE project – a full catalogue of Darwin's manuscripts being prepared by the Darwin Manuscript Project. The project will map the current BibText-based metadata into a form suitable for the BHL's ingest and display, and also as simple MARC records for incorporation into Cambridge University Library's catalogue. Darwin's Library has never been included within the Library's main electronic catalogue, since it has been treated as a manuscript collection rather than a book collection. The project will now enable these works to be discovered and (for those included in stage 1) electronically accessed.

3.1.6. Development of the delivery interface by BHL (see Workpackage 5)

The BHL has an existing infrastructure for ingesting, preserving and delivering digitised books and periodicals. In order to deliver Darwin's Library, it will undertake further development work to effectively display the transcribed marginalia and make use of the indexing terms developed by the Darwin Manuscripts Project. Discussions on the interface design are well underway and we expect to have a development plan from BHL by the end of this year and a data entry interface available early in the new year.

Particular consideration is being given to (1) how the annotations will be aligned with the page images, (2) how the individual scans will be associated with the surrogate books, and (3) how Darwin's Library will sit within the wider BHL as a separate, yet integrated, collection.

The revised approach to the project (described in 3.1.1.) means there is less need to explicitly link the annotation with the page image (since Darwin's original annotations will be visible to the user rather than a 'clean' surrogate page). However, we do want to provide some assistance to the reader in untangling the sometimes very complex layering and interlinking of transcriptions on a single page.

It is also important that we establish the best way for Darwin's Library to sit within the BHL. It will need to take advantage of some of the content and functionality of the wider BHL, but at the same time have its own distinct identity, as a sub-collection, curated collection or special edition. This is necessary so that users can limit their exploration to Darwin's Library (if the wish), and we can give appropriate recognition to project funders and protection for copyright contributors.

3.2. Issues to be addressed

The previous section (3.1.) has highlighted most of the issues that remain to be addressed, chiefly: (1) the final selection lists required for the scanning to proceed; and (2) the way in which the collection will be delivered within the context of the BHL. Both issues are being actively addressed and should be clarified before the end of the calendar year.

3.3. Scope and boundaries

As outlined above, the project will be able to include a significant proportion of Darwin's annotated library, but not the entire collection. In the selection of material for scanning within this phase of the project, we are seeking to optimise the content we make available to users. The extension to include Darwin's reading notes will provide a further dimension to the project, enabling users to gain a wider appreciation of Darwin's reading.

3.4. Critical success factors

The following factors will be critical to the success of this project:

- Timely finalisation of book lists for scanning
- Timely delivery of digitised images to BHL by CUL and NHM
- Timely conclusion of the checking and further mark-up of annotations by AMNH
- Timely development of the interface for delivery of the content through BHL
- Reception of the collection: Darwin's library becomes well-known and used by its intended audiences (Darwin scholars, students and the interested public)

4. Project Outputs

4.1. Deliverables

The project intends to produce the following tangible deliverables:

- **Online Darwin's Library collection**, comprising an estimated 300,000 pages of text, preserved, delivered, and made discoverable through the Biodiversity Heritage Library
- **A methodology and infrastructure** that can be applied to a further phase of the project, enabling the completion of Darwin's Library and his broader reading
- **A white paper and other contributions to relevant events and publications** outlining the lessons learned by the project
- **Project documentation**, as required by the JISC (e.g. plan and reports) and as necessary for the conduct of the project (minutes, internal discussion documents and reports)

4.2. Knowledge and experience

In addition to the tangible deliverables listed above, the project expects to gain some further knowledge and experience:

- **Greater knowledge of Darwin's reading practices**, through the process of checking annotations. Each time these are reviewed new information is uncovered.
- **Greater knowledge of the condition of Darwin's Library and the 19th century collection of the Natural History Museum**. The process of digitising Darwin's collection and the surrogates at the Natural History Museum will provide further information about the conditions of those collections. In both cases, conservators will be available to provide any necessary support.

- **Development of relationships.** Some of the partners have experience of working together, but this project establishes some new relationships, offering potential for further cooperation.

5. Project Outcomes

The project expects to deliver several outcomes

- **Increased research into Darwin's intellectual processes and the development of his thought.** The notes Darwin made to himself as he engaged in reading others provides important insights into the development of his thinking. Research in this area has been opened up by the work of Gill of Di Gregorio, but is currently hampered by the difficulties of accessing the original texts.
- **Availability of material for teaching.** In addition to opening up its research potential, the collection produced by this project will become available for inclusion within teaching programmes, and could enliven a range of courses (e.g. literature, history, psychology)
- **Success in fundraising to complete the project.** We expect the availability of a significant part of Darwin's Library within the BHL to assist in fundraising for its completion and enhancement
- **Access to Darwin's Library is improved, with access to originals reduced.** Once the collection is fully available, there will be less need for scholars to visit Cambridge or hunt out the often obscure texts in other libraries.
- **The BHL is enhanced through the addition of further natural history works.** While some scholars will be interested in accessing the collection because of its Darwin association; others will want access to the works themselves for other purposes. This project will serve both audiences.

6. Stakeholder Analysis

Stakeholder	Interest / stake	Importance
Darwin scholars	User group	High
Other relevant researchers (e.g. specialists in nineteenth century history, literature, science, philosophy or culture)	User group	Medium/High
FE Teachers and students in relevant disciplines	User group	Medium/High
School teachers and students	User group	Medium
JISC and NEH	Funders	Medium/High
Cambridge University Library	Project Partner	High
Natural History Museum	Project Partner	High
American Museum of Natural History	Project Partner	High
Biodiversity Heritage Library	Project Partner	Medium/High
Internet Archive	Project Associate	Low/Medium

7. Risk Analysis

The following table identifies and analyses risks associated with the project. It indicates actions undertaken to prevent or manage the risk, which may involve: avoidance (A), minimisation (M), or contingency (C).

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Staffing				
1. Loss of key staff	1	4	4	(M) Project is staffed from existing teams, with long contracts or commitments
Organisational				
2. Breakdown in consortium relationship	1	5	5	(M) Builds on existing long-term relationships (CUL and AMNH, NHM and AMNH, NHM and BHL), MOU and regular communication and visits (of US Director to UK)
3. Cost overrun	1	3	3	(M) Review deliverables and budgets
4. Schedule overrun	2	3	6	(C) There is contingency time within the schedule
Technical				
5. Delays in development	2	3	6	(M)(C) The project can be delivered with minimal development; some of the planned enhancements could be accomplished within stage 2; there is some contingency within development schedule.
6. Quality issues with images	1	4	4	(M) Both Cambridge and NHM have procedures in place to check quality and address any issues
7. Quality issues with metadata, transcriptions, indexing	1	4	4	(M) The project is checking this data and where possible data is being automatically migrated and mapped rather than re-keyed
8. Loss of data	1	5	5	(M) We will ensure appropriate back-up of data
External suppliers				
Not used in this project				
Legal				
9. Copyright issues	1	5	5	(M) Works will be assessed for copyright status and cleared or excluded as necessary. Permissions for use of Darwin annotations and DiGregorio and Gill transcriptions have been secured.
10. Licensing issues	3	4	12	(M) Where possible existing licenses will be adopted or adapted

8. Standards

Name of standard or specification	Version	Notes
BHL Imaging Requirements	03/02/09	Images produced in this project will meet and or exceed the minimum imaging requirements (600dpi bitonal, 300dpi grayscale, 300dpi colour; uncompressed

		TIFF or a JPEG2000; Adobe RGB 1998 colourspace)
BHL Metadata Requirements	03/02/09	Image produced in this project will meet or exceed the minimum metadata requirements for BHL

9. Technical Development

The technical development is being managed on the US side of the project, by Chris Freeland, Technical Director of the BHL. It will involve a modification of the standard BHL delivery interface to enable display of the transcriptions and associated index terms. An input interface will also be developed to enable easy input and editing of the additional data. BHL's development is done within the wider context of the Internet Archive's development, so the work done for this project is likely to add to the functionality of other Internet Archive services. Details of the BHL technical infrastructure can be found on the BHL blog: <http://biodiversitylibrary.blogspot.com/search/label/Technical%20Notes>

10. Intellectual Property Rights

The BHL currently includes out of copyright content, which it licenses via the Creative Commons. However, the current licence (Attribution Non-Commercial) does not offer sufficient protection for Darwin's annotations (which are still in copyright in UK law until 2039) or for Gill and Di Gregorio's work (in copyright in both US and UK law). The project is currently discussing how best to license this content in a way that maximises access and use, but minimises miss-use or the wholesale replication of this content.

Project Resources

11. Project Partners

The table below lists the project partners, their roles and main contacts. The UK side of the project is lead by Cambridge, with NHM as partner; the US side, by DMP for AMHN, with the BHL Technical Development team as partner. However, the BHL representation really spans both sides of the project, since NHM are a key participant within the BHL consortia. The key decision-making group for the project, then, consists of Grant Young, as Cambridge representative; David Kohn, as DMP representative; and Jane Smith as NHM and BHL representative.

Partner	Role	Contacts
Cambridge University Library (CUL)	UK lead, digitisation partner	Grant Young , UK Project Director and Digitisation Specialist– gy219@cam.ac.uk
American Museum of Natural History (ANHM) – Darwin Manuscripts Project (DMP)	US lead, transcription and metadata partner	David Kohn , US Project Director and Darwin Scholar – dkohn@amnh.org

Natural History Museum (NHM)	Digitisation partner and BHL representative	Jane Smith , Head of Library Collections and Services – Jane.Smith@nhm.ac.uk
Biodiversity Heritage Library (BHL)	Technical Partner and BHL representative	Chris Freeland , Technical Director of BHL – Chris.Freeland@mobot.org

12. Project Management

As noted in the previous section, the project is managed for JISC and NEH by Grant Young and David Kohn, but Jane Smith has an important role as BHL representative. The project makes use of existing, skilled staff, so no special training is required. Workpackage 1 (Appendix B) outlines project management activities.

Partner	Key team members
Cambridge University Library (CUL)	<ul style="list-style-type: none"> • Grant Young, UK Project Director and Digitisation Specialist – gy219@cam.ac.uk. Grant will undertake UK project management (approximately 0.1 FTE – 2 days a month) • Don Manning, Head of Imaging Services – dm10009@cam.ac.uk • Adam Perkins, Keeper of Scientific Manuscripts – ajp21@cam.ac.uk
American Museum of Natural History (AMNH) – Darwin Manuscripts Project (DMP)	<ul style="list-style-type: none"> • David Kohn, US Project Director and Darwin Scholar – dkohn@amnh.org. David will undertake US project management (approximately 0.2 FTE) • Nick Gill, Darwin Scholar – nwg20@cam.ac.uk • Huw Jones, Metadata Specialist – hej23@cam.ac.uk
Natural History Museum (NHM)	<ul style="list-style-type: none"> • Jane Smith, Head of Library Collections and Services, BHL representative – Jane.Smith@nhm.ac.uk • Judith Magee, Library Collection Development Manager – j.magee@nhm.ac.uk
Biodiversity Heritage Library (BHL)	<ul style="list-style-type: none"> • Chris Freeland, Technical Director of BHL – Chris.Freeland@mobot.org

13. Programme Support

The project will keep the UK programme manager well briefed on progress and any issues that arise. We do not currently anticipate the need for special support.

14. Budget

Please see Appendix A for the budget. This makes adjustment to the digitisation costs to take into account the lower than expected budget allocation (£103,000 rather than £120,000).

Detailed Project Planning

15. Workpackages

Please see Appendix B for a description of the workpackages.

16. Evaluation Plan

Timing	Factor to Evaluate	Questions to Address	Method(s)	Measure of Success
Jan-10 – Oct-10	Quantity and quality of content (images and texts)	Has project achieved quantity targets? Is data of sufficient quality?	Page image tallies. Quality plan (below).	At least 300,000 pages of content delivered. Quality targets are achieved.
Nov-09 – Feb-11	Usability of delivery interface	Is delivery interface fit for audience?	Wide input into design. Usability testing. Gather early feedback from users	Positive user feedback. Web logs indicate broad and effective use.
Sep-09 – Feb-11	Effectiveness of methodology/ infrastructure	Does chosen methodology deliver high quality resource?	Programme's evaluation. User feedback.	Project delivers quality outputs within time/budget. Team are successful in funding Phase 2.
Jan-11 – beyond project	Value of white paper	Does the white paper add to understanding of digitisation?	Review by programme manager and peers	White paper is widely cited and influences other projects
Beyond project	Impact on scholarship	Does the availability of this resource stimulate research?	User feedback. Literature reviews by Darwin scholars	Citations within publications. Demand for further content
Beyond project	Use within teaching and learning	Does the resource become used within teaching?	User feedback	Web links to the resource. Demand for further content

17. Quality Plan

Output	Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
Archival scans (CUL)	Jan-10 – Oct-10	Legibility and fidelity	Visual check by operator and post-scan sampling	Internal reporting	CUL – Imaging Services Department	Regular calibration
Archival scans (BHL)	Jan-10 – Oct-10	Legibility and fidelity	Visual check by operator and post-scan sampling	Internal reporting	NHM – Internet Archive and Imaging Studio operators	Regular calibration
Transcriptions and indexing	Feb-10 – Dec-10	Accuracy	Existing transcriptions are being re-checked	Acceptance	AMNH editors	-
Metadata	Feb-10 – Dec-10	Complies with BHL specifications and MARC standard	Automated checking	Acceptance	Metadata specialist	BHL ingest, CUL: voyager library system
Delivery dataset	Nov-10 – Feb-11	BHL standards	BHL methods	Acceptance	BHL	-
Interface	Nov-09 – Feb-11	Usability and accessibility	User feedback	Acceptance	BHL	-
White paper	Jan-11 – Feb-11	Fitness for purpose	Programme Manager and peer review	Acceptance	Authors and reviewers	-

18. Dissemination Plan

Timing	Dissemination Activity	Audience	Purpose	Key Message
Nov-09 – Feb-11 (intervals)	Project website on CUL	All interested parties	Raise awareness and inform of development	Project outline and progress
Nov-09 – Feb-11 (intervals)	News and features on other partner sites (e.g. BHL blog, DMP site)	All interested parties	Raise awareness and inform of development	Project outline and progress
Sep-09 – Feb-11 (at key milestones)	Press releases	All interested parties	Raise awareness and inform of development	Project outline and progress
Sep-09 – Feb-11 (as	Internal and external presentations	Specific, targeted groups (e.g.	Raise awareness and inform of	Project outline, progress and

opportunities arise)		colleagues, BHL members, Darwin specialists)	development	targeted
Jan-11 – Feb-11	White paper	Those undertaking similar projects	Share methodology and lessons learned	Explanation of process and learning
Feb-11	Final report	JISC community	Share methodology and lessons learned	Explanation of process and learning
Sep-09 – Feb-11 (as opportunities arise)	Funder proposals and presentations	Potential funders	Raise funding for Phase 2	Project worthy of further support

19. Exit and Sustainability Plans

Project Outputs	Action for Take-up & Embedding	Action for Exit
Archival datasets	Held by BHL and CUL (Darwin manuscripts) and DMP (annotations and metadata)	Ensure appropriate metadata and documentation
Delivery datasets	Generated and maintained by BHL and CUL (marc metadata) and DMP (other metadata)	Ensure appropriate metadata and documentation
Interface	Interface will become component of BHL infrastructure for use with other similar content	Ensure development work is documented and available for re-use
Knowledge and learning	Recorded in project documentation and white paper	Ensure key learning is recorded in project documentation and white paper

Project Outputs	Why Sustainable	Scenarios for Taking Forward	Issues to Address
Archival datasets	Required for generation of delivery data	Standards-based, maintained by BHL/Internet Archive, CUL (in repository) and DMP	
Delivery datasets	Core content	Maintained by BHL/Internet Archive	
Interface	Necessary for delivery of content	Will become core part of BHL functionality	
Project documentation	Project requirement, needed for Phase 2	Core documentation archived by lead, partners and JISC	
White paper	Project requirement	Archived by partners and JISC	

Appendixes

- Appendix A. Project Budget (UK)
- Appendix B. Workpackages

Percentage Contributions over the life of the project	JISC 41 %	Partners 59 %	Total 100%

Notes

1. We are asking for £10,000 towards these Directly Incurred Staff Costs, £7,000 in the first year and £3,000 in the second
2. The per page costs incorporate all the costs associated with the digitisation, including staff, equipment and overheads. The page numbers here are estimates and will be confirmed once the selection and checking work is completed.
3. A proportion of the Project Director's time is provided by CUL as a partner contribution
4. CUL FECs are calculated on 0.4 FTE, including the CUL Project Director, CUL Senior Curator, and CUL Conservation Assistant.
5. NHM FECs are calculated on the basis of 1.7 FTE and include all the NHM Directly Incurred and Directly Allocated staff listed in this budget.

Appendix B. Workpackages

WORKPACKAGES	<i>Month</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
		S	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D	J	F
1: Project Management		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
2: Selection work		X	X	X	X														
3: Cambridge digitisation						X	X	X	X	X	X	X	X	X	X				
4: NHM and other BHL digitisation						X	X	X	X	X	X	X	X	X	X				
5: Technical development				X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
6: Transcriptions, indexing and metadata			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X		
7: Dissemination		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
8: Evaluation		X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X

Project start date: September 1 2009

Project completion date: February 28 2011

Duration: 18 months

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Milestone	Responsibility
--------------------------	---------------------	------------------------	--------------------------------------------------------------	-----------	----------------

WORKPACKAGE 1: Project Management	Sep 09	Feb 11			
Objective: To ensure effective internal communications & success of the project					
1. Project start-up and planning	Sep 09	Oct 09	JISC Project Plan	M	GY, DK
2. Frequent meetings/calls between partners (according to project needs)	Sep 09	Feb 11	Communications and meetings between project partners, minutes of formal meetings		GY, DK
3. Financial monitoring	Sep 09	Feb 11	Budget reports to JISC		GY, DK
4. Project agreements and licenses	Sep 09	Feb 11	MOU, content licences, end-user licence		GY, JS, DK
5. Project reporting	May 10	May 10	JISC Mid-Project Report		GY
6. Project closure	Feb 11	Feb 11	JISC Final and Completion Reports	M	GY
WORKPACKAGE 2: Selection work	Sep 09	Dec 09			
Objective: To finalise scanning schedules for Cambridge and NHM/BHL					
7. Bibliographic list for checking	Sep 09	Oct 09	Darwin's Library list for checking		DK, NG
8. Checking of NHM and BHL holdings for availability and condition	Oct 09	Dec 09	Scanning schedule for NHM/BHL	M	JS, JM
9. Confirm list of original volumes and pages to scan from Cambridge	Nov 09	Dec 09	Scanning schedule for Cambridge	M	DK, NG

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Milestone	Responsibility
<p>WORKPACKAGE 3: Cambridge digitisation</p> <p><u>Objective:</u> To capture high quality images of original volumes and pages from Darwin's Library</p>	Jan 10	Oct 10			
10. Scanning of Cambridge list	Jan 10	Oct 10	Approximately 35,700 images with associated metadata	M	GY, DM
<p>WORKPACKAGE 4: NHM and other BHL digitisation</p> <p><u>Objective:</u> To capture high quality images of original volumes and pages from Darwin's Library</p>	Sep 09	Sep 09			
11. Scanning of NHM list	Jan 10	Oct 10	Approximately 12,250 images with associated metadata	M	JS, JM
12. Scanning of BHL list	Jan 10	Oct 10	Approximately 262,400 images with associated metadata	M	JS, JM
<p>WORKPACKAGE 5: Technical development</p> <p><u>Objective:</u> To develop the BHL system to enable input and display of Darwin's Library content</p>	Nov 09	Feb 11			
13. Develop input interface	Nov 09	Jan 10	Fully functional input interface	M	DK, CF
14. Develop and test end user interface	Jan 10	Dec 10			
15. Release interface to public and build up content	Jan 11	Feb 11	Fully functional end-user interface	M	DK, CF

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Milestone	Responsibility
WORKPACKAGE 6: Transcriptions, indexing and metadata <u>Objective:</u> Check, repurpose, develop, and input transcriptions, index terms and metadata	Sep 09	Sep 09			
16. Check and edit transcriptions and indexing	Oct 09	Dec 10			DK, NG
17. Transform metadata for BHL, DARBASE and CUL catalogue	Mar 10	Dec 10			DK, HJ
18. Input transcriptions, index terms and metadata	Feb 10	Dec 10		M	DK, NG, HJ
WORKPACKAGE 7: Dissemination <u>Objective:</u> Integration of Tools with DAMS	Sep 09	Feb 11			
19. Develop project webpage	Sep 09	Nov 09	Public webpage by lead institution	M	GY
20. Produce white paper outlining methodology and lessons learned	Jan 11	Feb 11	White paper	M	GY, DK
21. Contribute presentations and articles to appropriate fora/publications	Sep 09	Feb 11	Presentations, publications		GY, DK, JS
WORKPACKAGE 8: Evaluation <u>Objective:</u> Ensure the project produces high-quality outputs and captures lessons learned	Sep 09	Feb 11			
22. Quality assurance of images, text and interface	Oct 09	Dec 10	Internal records and reports		All partners

Workpackage and activity	Earliest start date	Latest completion date	Outputs (clearly indicate deliverables & reports in bold)	Milestone	Responsibility
23. Document lessons as the project proceeds	Sep 09	Feb 11	Lessons are communicated in formal reports and dissemination activities (above)		GY, DK

Project members:

GY – Grant Young, UK Project Director/Manager
DK – David Kohn, US Project Director/Manager
JS – Jane Smith, NHM lead and BHL representative
JM – Judith Magee, NHM
CF – Chris Freeland, BHL technical lead
HJ – Huw Jones, metadata specialist