

The Natural History Production Line: An Industrial Approach to the Digitization of Scientific Collections

MAARTEN HEERLIEN, JOOST VAN LEUSEN, STEPHANIE SCHNÖRR, SUZANNE DE JONG-KOLE, NIELS RAES, and KIRSTEN VAN HULSEN, Naturalis Biodiversity Center

In 2010, Naturalis Biodiversity Center started one of the largest and most diverse programs for natural history collection digitization to date. From a total collection of 37 million specimens and related objects, 7 million relevant objects are to be digitized in a 5-year period. This article provides an overview of the program and discusses the chosen industrial production line approach, the applied method for prioritization of collections that are to be digitized, and some preliminary results.

Categories and Subject Descriptors: E.1 [Data]: Records; I.3.3 [Computer Graphics]: Digitizing and Scanning

General Terms: Natural History, Collection Digitization

Additional Key Words and Phrases: Scientific collections, prioritization, industrial approach, production lines, public engagement, crowdsourcing

ACM Reference Format:

Maarten Heerlien, Joost van Leusen, Stephanie Schnörr, Suzanne de Jong-Kole, Niels Raes, and Kirsten van Hulsen. 2015. The natural history production line: An industrial approach to the digitization of scientific collections. *ACM J. Comput. Cult. Herit.* 8, 1, Article 3 (February 2015), 11 pages.
DOI: <http://dx.doi.org/10.1145/2644822>

1. INTRODUCTION

In 2010, the newly formed Naturalis Biodiversity Center, the national museum and research institute for biodiversity of the Netherlands, was awarded with a 30 million euro government grant from the Dutch Fund for Economic Structure enhancement to give shape to the new institute. From this grant, 13 million euro was allotted to a digitization program. Through this program, Naturalis aims to digitize in detail a cross section of at least 7 million relevant specimens and related objects from a total collection of 37 million specimens, whereas the remaining 30 million objects are to be digitized on a higher level in a 5-year period. Furthermore, the program focuses on developing a sustainable

This fund, *Fonds Economische Structuurversterking*, in Dutch, was established in 1995 to invest profits from the natural gas reserves from the northern regions of the Netherlands in the Dutch infrastructure and as of 2005 also in the Dutch knowledge economy. Naturalis Biodiversity Center was among the last institutions to receive a grant from the fund, which was discontinued in 2011. The other 17 million euro was used to (1) integrate the collection of the National Museum of Natural History Naturalis with those of the Zoological Museum of Amsterdam and of the National Herbarium of the Netherlands, with Naturalis Biodiversity Center being the result of a merger between the three, and (2) to establish a DNA barcoding facility.

Authors' addresses: M. Heerlien, J. van Leusen, S. Schnörr, S. de Jong-Kole, N. Raes, and K. van Hulsen, Naturalis Biodiversity Center, Darwinweg 2, 2333 CR Leiden, The Netherlands; emails: {Maarten.Heerlien, Joost.vanLeusen, Stephanie.Schnoerr, Suzanne.deJong-Kole, Niels.Raes, Kirsten.vanHulsen}@naturalis.nl.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

2015 Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM 1556-4673/2015/02-ART3 \$15.00

DOI: <http://dx.doi.org/10.1145/2644822>

infrastructure that will allow Naturalis continuing digitization. This article provides an overview of the program, which is one of the largest natural history collection digitization programs to date. The program approach with respect to prioritization of collections, to the digitization process, and to public engagement is discussed. We conclude with an overview of program results so far.

1.1 Motivation for Digitization

Natural history collections like those maintained by Naturalis are of great importance to the community at large, as they give invaluable insights to the past and present state of global biodiversity and potentially help to solve current challenges in many areas, such as the effects of environmental change, public health, and crop pollination [Baird 2010]. Digitizing these collections makes the enclosed knowledge more accessible to scientists researching these issues and facilitates more efficient collection management as well as protection of collected specimens from overhandling. Additionally, digitization makes virtual repatriation of collections possible to the countries from which they were gathered—in Naturalis’ case, mainly former Dutch colonies such as Suriname and Indonesia.

1.2 Challenge

The challenge in this respect, and with regard to the heterogeneity of collection types maintained by Naturalis, is to determine which 7 million objects out of a total of 37 million are the most relevant in relation to current scientific, social, and economic issues and how to develop a digitization process for these that would not exceed the 13 million euro budget, allowing for a maximum average price of 1.86 euro per digitized object, including overhead, permanent storage, and equipment costs. At the time Naturalis applied for the grant, the average digitization cost per object was estimated to be close to 5.00 euro, based on experience from previous project in which the traditional digitization approach was taken. In this approach, high-quality images of selected specimens are made and all available specimen data is registered—a labor-intensive method that requires expert knowledge and is therefore costly but does not result in datasets that cover complete collections. With the objective to digitize 7 million specimens for 1.86 euro per specimen, Naturalis needed to divert from this traditional method.

1.3 Approach

To deal with this challenge, an industrial approach was chosen based on the following starting points:

- To develop a framework for prioritization to determine which collections should be digitized
- To develop digitization processes based on the collection types, such as collections preserved in alcohol, dry collections, microscopic slides, and printed publications, enabling the digitization of any collection in that type category regardless of its specific content
- To divide complicated and labor-intensive processes into a series of shorter tasks, each executed by a coworker specialized in that task
- To standardize the data entry process through the use of one metadata standard and central data management systems
- To limit metadata capture to a minimum by only registering metadata needed for collection management and for basic accessibility to researchers
- To only capture photographic reproductions of specimens where this has a proven added value
- To make use of (commercial) third parties for digitization where it makes sense (price/quality-ratio driven).

Naturalis is not the first natural history institute to apply these kinds of starting points to develop a large-scale digitization program. In recent years, there has been growing concern within the scientific

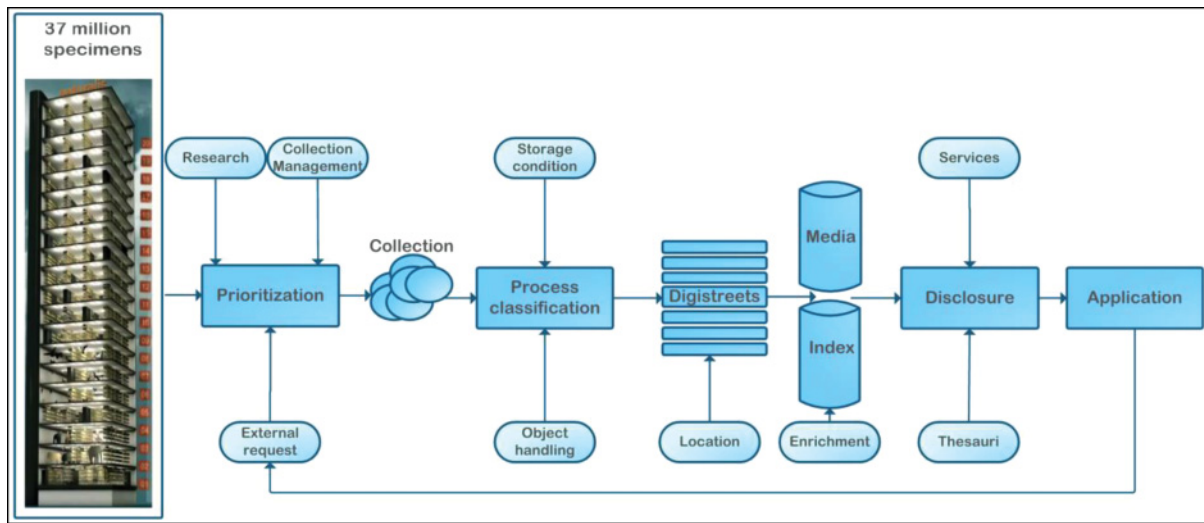


Fig. 1. Schematic overview of the Naturalis approach to digitization process organization.

heritage community about the slow pace of the digitization of the estimated three billion zoological, botanical, and geological specimens that are maintained in natural history collections worldwide [Vollmar et al. 2010]. The Natural History Museum in London applied a similar approach in the digitization of its entomological collection [Blagoderov et al. 2012], as did the Muséum National d'Histoire Naturelle in Paris to digitize its entire herbarium collection, which is the largest in the world. The Digitalium in Finland has applied a production line approach to both entomological and herbarium collections [Tegelberg et al. 2012]. However, the Naturalis program is unique in the diversity of the collection types that are being digitized. Whereas London, Paris, and Finland have focused on insect drawers and herbarium sheets, Naturalis has diversified the approach to facilitate the digitization of virtually all types of collections that natural history heritage encompasses on both the specimen and storage unit level (Figure 1).

The starting points presented earlier were worked out further in a series of 3- to 6-month pilots, each focusing on developing and optimizing the process workflow and tools needed for the digitization of a specific collection type, as well as developing a business case and success indicators to determine whether or not the chosen approach was viable. The results were presented to the program's steering committee, consisting of the institute's general and scientific management, to make a final decision on whether or not a pilot would be promoted to a full project. Most pilots were taken into production, whereas a few were deemed ineffective and have been shut down, such as a pilot for 3D capturing of dry vertebrates for which at the start of the program the investment turned out too large in relation to the added value for scientists and collection managers of the 3D images captured [Van den Oever and Gofferjé 2012]. In 2014, however, the decision not to apply 3D digitization has been reevaluated. A new pilot for experimenting with various 3D imaging techniques (CT scanning, laser scanning and photogrammetry) was started in the summer of 2014.

1.4 Digistreeets for Detailed Object Digitization

Projects that are taken into production are executed in so-called digitization streets, or Digistreeets—facilities comparable to production lines. Currently, Naturalis operates seven Digistreeets, with all

Table I. Targeted Number of Objects to Be Digitized, Currently Realized Number of Digitized Objects, and Average Cost per Object per Digistreet as of April 1, 2014

| Digistreet | Target | Realized | Average Cost per Object in Euros (including overhead) |
|------------------------------------|-----------|-----------------------------|--|
| Herbarium sheets—Leiden | 3,800,000 | 450,000 (3,500,000 scanned) | 1.29 |
| Microscopic slides | 900,000 | 350,000 (590,000 scanned) | 1.57 |
| 2D objects (books, journals, etc.) | 900,000* | 600,000* | 1.87 |
| Entomology | 850,000 | 600,000 | 1.51 |
| Herbarium sheets—Wageningen | 800,000 | 800,000 | 1.47 |
| Mollusks | 510,000 | 640,000 | 1.37 |
| Vertebrates dry | 325,000 | 180,000 | 2.37 |
| Geological specimens | 220,000 | 70,000 | 1.90 |
| Wood samples | 125,000 | 125,000 | 1.27 |
| Specimens preserved on alcohol | 100,000 | 70,000 | 4.65 |
| Total | 8,530,000 | 3,885,000 | |

*Number of pages.

except one situated within the institute itself and each equipped to handle the digitization of a specific collection type. These are herbarium sheets, microscopic slides, entomology collections, 2D objects (journals, rare books, archives, etc.), dry vertebrates, geological specimens, and specimens preserved in alcohol. Two other Digistreeets—the street for mollusks and the wood samples street—met their targeted results in 2013 (the mollusk street surpassed its target by more than 100.000 specimens; Table I) and were wound down.

For each Digistreet the targets are worked out in further detail based on the business case developed in the pilot to determine how many objects can be digitized during the production phase; how many people are needed to meet these targets; and how many objects are needed to process per day, week, and month.

1.5 Prioritization Framework

To facilitate the decision making process for which 7 million objects are to be digitized in detail, a framework for collection prioritization has been developed. This ensures the consistency of the digitized collections with current scientific, educational, or economic affairs. The framework was used to fill in 80% of the total number of targeted objects per Digistreet, with 20% being reserved for additional on-demand digitization (e.g., at the request of external research institutes).

In the first phase, scientists and collection managers are invited to submit proposals for the digitization of specific collections or parts thereof. A typical proposal contains a description of the collection that is to be digitized as well as solid arguments for digitization, such as the expected benefits from digitizing the specified collection with regard to current research programs and collection preservation. In all cases where the method was applied, the total number of objects in the collections for which digitization was proposed surpassed the target set for the Digistreet in question. To weigh the relevance of the proposed projects as objectively as possible and to rank them accordingly, the second phase consists of a two-step approach. First, a number of hard selection criteria are set, including the following (in random order):

- Relation to the institute's own research priorities
- Relation to the institute's own public and educational programs
- Relation to national and international biodiversity programs

- Economic importance of the proposed collection
- Availability of existing collection documentation and data
- Physical state of the proposed collection.

The proposals are judged and ranked according to these criteria. Second, an online survey is held among a large group of stakeholders, who are asked to rank the significance of each proposal with regard to their own fields of expertise. In the third phase of the method, the results of both these steps are processed into a recommendation on the most favorable proposals, which is submitted to the program's steering committee for a final decision.

1.6 Digistreet Production Process

After a positive decision, the collections are digitized in the Digistreet. Although the details of the digitization process vary from street to street with respect to scanning techniques, registered data, and so forth, the overall approach is the same for all Digistreeets. Here, the process is illustrated by the Digistreet for microscopic slides. This Digistreet, known as the Glass Street, is one of the largest that is operational within Naturalis at this time. The target set for this street is to digitize (i.e., to capture high-resolution images) and register all relevant label data of the entire collection of microscopic slides maintained by Naturalis, comprising approximately 900,000 objects. In this sense, the Glass Street diverges from regular practice, as in this case there is no need for prioritization. To reach this target, an innovative industrial process approach is developed where the Glass Street acts like an efficient production line.

Step one in this line (Figure 2) is to supply the Glass Street with microscopic slides. This is done by curators authorized to transport specimens from the storage facility to the Digistreet and back. Once in the Digistreet, the next step is to label every slide with a unique data matrix code, which is linked to an empty data record in the central collection registration system, after which they are placed in a scan tray. This custom-made tray can harbor 100 slides at once. The tray is scanned with the use of a SatScan[®] collection scanner, a system for capturing high-resolution images of large area objects, developed by SmartDrive. Naturalis operates two of these devices; the other is used to scan entomology drawers containing pinned insects. (For details on SatScan, see Blagoderov et al. [2012] and Mantle et al. [2012]). The result is a 600Mb high-resolution overview image containing all 100 slides. Subsequently, specially developed software cuts each individual slide from the overview image and renames each cut-out image according to the object number contained in its data matrix code. The result is a set of 100 individual slide images of 4Mb that are stored in an online repository and used for data registration in a later process stage. The image quality does not facilitate digital image-based specimen research. Producing images with a resolution high enough for such purposes would take more production time per slide as well as more digital storage capacity, thus making the process much less efficient.

Using this scanning production line, the Glass Street scans between 2,000 and 2,500 slides per day, with four to five employees operating it, each executing one step in the process. The scan tray is used for the normal size microscopic slides of 1 by 3 inches. However, 30% of the collection consists of geological slides that have a different size. These are scanned while in their storage trays. Besides the normal and geological slides, there is a collection of slides that have irregular sizes. Depending on the number and size, these slides are either photographed individually or are scanned using a specially sized scan tray. After the scanning process, the microscopic slides are replaced in their original storage container to which each individual slide is digitally linked. Subsequently, the container is digitally linked to its physical storage location.

In the last step of the Glass Street production line, the label data from the microscopic slides is registered. To do so, street workers use the high-resolution images instead of the slides themselves.



Fig. 2. Schematic representation of the Glass Street production process. Illustration: Ben van Arkel.

This ensures minimal handling of the physical objects. The images are automatically linked to their corresponding data entry record by the unique code that was assigned to the slide at the beginning of the production line. Because of this, the image is visible once the corresponding data record in the collection registration system is opened and the label data can be registered. Entering the data is a manual process. Most labels are handwritten, ruling out automatic data capture.

1.7 Outsourcing

Although six of the seven currently operational Digistreets are situated within Naturalis and are being operated by permanent and temporal staff members, operation of the seventh Digistreet, aimed at digitizing herbarium sheets from Leiden, Amsterdam, and Utrecht³ has been outsourced to Picturae, a Dutch service provider in the field of collection digitization and digital collection management. This Herbarium Street launched in July 2013 and was the last of the originally planned Digistreets to become operational. It is also the largest Digistreet within the Naturalis digitization program with a target to digitize—that is, to capture high-resolution images and register all of the relevant label

¹In addition to the herbarium sheets within the Leiden collection, there is also a (smaller) herbarium collection in the city of Wageningen. This collection is digitized in a separate, internal Digistreet in the Wageningen location of the former National Herbarium of the Netherlands, now part of Naturalis Biodiversity Center; see the Introduction and Table I.

data—3,800,000 herbarium sheets. To reach this target in a cost-effective way, an innovative process has been developed that closely resembles an actual industrial production line.

First, the Herbarium Street is provided with boxes of herbarium sheets. Because of the volume of the operation, this is done by a professional transport company. Once in the Digistreet, the box is labeled with a unique data matrix code, after which it is emptied and placed onto a conveyor belt. Subsequently, the corresponding covers and herbarium sheets are labeled with unique data matrix codes and placed onto the conveyor belt as well. All three items are photographed with a Nikon D800e camera. Built-in software checks the label and reads out color and sharpness. If an error occurs, the conveyor belt stops and automatically takes a few steps back so that the photo can be remade. In all, it takes 8 seconds to produce the final file.

The output of the process is a 300ppi TIFF image of the herbarium sheet, a 150ppi image of the box, and the cover and a comma-separated value (csv) file format. This file contains all information from the image that is needed to start the data entry. At the end of the conveyor belt, the herbarium sheets are packed into their boxes again in the exact same order as they were before digitization, after which they are transported back to Naturalis. Using this production line, the Herbarium Street is able to digitize between 22,000 and 24,000 herbarium sheets a day, with three conveyor belts and 12 employees operating it.⁴

The last step of the Herbarium Street production line concerns the metadata entry. This is done in Paramaribo, Suriname, by a team of 40 employees trained in transcribing handwritten labels. For this purpose, jpeg derivatives of the high-resolution images created in Leiden are used. As in the Glass Street, the images are automatically linked to their corresponding data entry record by the unique data matrix code that was assigned to them in the Herbarium Sheet at the beginning of the production line. Entering the data is a manual process. Most labels are handwritten, and some of them are hundreds of years old, ruling out automatic data capture.

1.8 Public Engagement

Since the digitization program is entirely financed with public funds, transparency with respect to expenditure, approach, and results is key. Here, the Glass Street plays an important role, as it is situated in one of the museum exhibition spaces where it functions as a public demonstrator of the digitization program. In this exhibition space, called *LiveScience*, museum visitors get to observe the digitization process and interact with the Digistreet workers, being only separated from them by a low, open barrier.⁵ Before the launch of the Glass Street in LiveScience in April 2013, the exhibition space was home to the Mollusk Street, which opened in May 2011 and reached (and surpassed) its target of 400,000 specimens in early 2013, after which it was wound down.

The public is also invited to participate in digitization. The Mollusk Street and Glass Street applications were developed that enable enthusiasts to transcribe object labels. The Web app developed for Mollusk transcription was primarily aimed at fostering appreciation for the digitization of scientific heritage among museum visitors by letting them try for themselves and to a lesser extent at producing high volumes of user-generated collection records.⁶ The Glass Street crowdsourcing application took this to a next level. Here, an existing Dutch online platform for transcribing handwritten heritage objects, *VeleHanden* (Many Hands), was used for a full-scale crowdsourcing project aimed at the Dutch-speaking regions, called *Glashelder!* (roughly translating to Crystal Clear).⁷ The public was

²For a visual presentation of the Herbarium Street digitization process, see <http://www.youtube.com/watch?v=hmG4twyHXkE>.

³See <http://www.naturalis.nl/en/museum/livescience>.

⁴See <http://www.naturalis.nl/en/museum/livescience/crowd-sourcing>.

⁵See <http://www.velehanden.nl/projecten/bekijk/details/project/nat`nbc> (in Dutch).

The screenshot displays the Glashelder! web application. The top navigation bar includes the logo and user profile information. The main interface is split into two panels. The left panel contains a data entry form with the following fields: 'Nectarosiphon persicae' (scientific name), 'Sulzer' (author), 'onbekend' (year), '(leeg)' (sex), 'Prunus persicae' (host species), 'Land' (gathering country), 'Wageningen' (locality), '14.VI.68' (collection date), 'J. Oldenburger' (collector), 'Aantal' (number of specimens), and 'Oude code(s)' (previous registration code). Below the form are buttons for 'Afronden', 'Te moeilijk', 'Onbruikbaar', and 'Opnieuw!'. The right panel shows a microscopic slide image of several aphids on a plant, with a label that reads: 'Nectarosiphon persicae Sulzer', 'RMNH.JNS.837032', 'N. persicae Pl. persicae Loc. Wageningen Date 14.VI.68 Leg. J. Oldenburger'. Below the image is a JavaScript viewer for the image.

Fig. 3. The transcription module of the Glashelder! project. Left: The current data entry form with (top to bottom) the field's scientific name, author, and year (of publication of the scientific name), sex, type specimen, host species (relevant in case of parasites), gathering country, locality, collection date, collector, number of specimens, and previous registration code. Right: Microscopic slide image with several Aphids of the species *Nectarosiphon persicae* Sulzer. Image: Vele Handen–Picturae. With permission of Picturae.

encouraged to sign up for the project on the VeleHanden platform and to transcribe microscopic slide labels, aided by a comprehensive manual, FAQs, and a forum for project members to discuss occurring issues among themselves and with museum staff. Glashelder! served as an experiment to determine to which extent online transcription of natural history data by volunteers can contribute, both quantitatively and qualitatively, to collection digitization. To determine this, a separate production target was set for the project—transcription of 100,000 microscopic slides in a period of 6 to 9 months—as well as a set of success indicators.

The Glashelder! project was launched on March 26, 2013, after a 1-month trial period during which members of the VeleHanden platform (i.e., people with little, if any, knowledge about natural history collections) tested the transcription module and supporting documentation. Based on their user feedback, several changes were made to the project design, most notably a simplification of the transcription form (Figure 3), the first version of which left too much room for interpretation while at the same time did not provide enough room for the recording of exceptions in, for instance, the zoological nomenclature, thus raising many questions among the test crowd. The final form was reduced to 11 fields.

The Glashelder! project is regarded a success. On December 30, 2013, 9 months after launch, the last of the 100,000 glass slide labels was transcribed, whereas the validation of the transcription was finished on January 19, 2014. During the project, a total of 511 participants had signed on, about one third of whom did so in the first project month. In part, this can be credited to a media campaign, but mostly the project was able to benefit from the community already present at VeleHanden. About 150 project members were regarded as active participants, each having transcribed up to 1,000 labels. During this project, 23 members were regarded as super participants, each having transcribed

1,000 slides or more. Although the project was not specifically aimed at biologists and did not require any prior knowledge of the domain to participate, an inquiry among the project members showed that most either had a professional background in the natural sciences or were nature enthusiasts with an above-average interest in and knowledge of the species preserved in the slides, mostly Mites, Aphids, and Springtails.

Together, the participants produced 200,000 transcriptions in 9 months, resulting in 100,000 validated label transcriptions, as the label of each microscopic slide was transcribed twice by different participants with a third, more experienced, participant validating the two transcriptions using a comparison tool and submitting a final version. The average daily number of produced transcriptions over the course of the entire project lay at 712 slide labels, whereas the highest number of slide labels transcribed on 1 day is 1,914 (August 7, 2013). On average, 334 slides were validated daily. The first of the crowd-generated datasets, the Collembola or Springtails set, has been published through the Global Biodiversity Information Facility (GBIF), with the rest of the data to come.⁸

1.9 Application of Digitized Collections

The largest part of the collections digitized during the program was chosen because of relevance to current research topics. Therefore, the digitized collections play a vital role in the national and international research programs that Naturalis leads or is otherwise involved in, such as the current research program on the decline of pollinators in Europe and the effects on crop pollination and food supply [Carvalho et al. 2013], in which the digitized bee and bumblebee collections were used to analyze changes in the occurrences of these pollinators over the past 60 years. In addition, the high-resolution images taken of the bumblebees are currently used to develop algorithms for automated species identification based on their wing venation, the patterns of which provide the only means to distinguish between some species of bumblebees.

The digitization of the herbarium sheets contributes to biological conservation research that identifies global hotspots of botanical diversity, how past (glacial) climatic conditions have shaped the spatial distribution of hotspots, and how future climate change predictions will likely impact their distribution. The digitized records are also used to determine the distribution and growth conditions of plants that are the (crop) wild relatives of human food crops [Raes et al. 2013]. The breeding of crop species with species that are evolutionary closely related and are currently found to grow under warmer, drier, or more saline conditions will result in crops that are resistant to future climate change (IPCC 2014); this type of research is known as climate smart agriculture.

1.10 Current Program Status

Currently, the digitization program is approaching the end of its fourth year—and its third year of full-scale production. In the past 3 years, all Digistreet were made operational and close to 4 million specimens and related objects have been digitized. The realized average costs of object digitization vary from street to street due to the different nature of the treated objects. However, through ongoing optimization of the production processes and tools, the overall average cost of object digitization has been reduced to 1.52 euro per object, including overhead costs. This enables Naturalis to digitize 1.5 million more collection objects during the digitization program than was originally planned. Table I provides an overview of targets and objects realized per April 1, 2014, and average cost per object per Digistreet. In addition, 70,000 entomology drawers have been digitized on a higher, less detailed level by registering each drawer, the species and quantities it contains, the geographic location from which

⁸See <http://www.gbif.org/dataset/4f8de55f-5967-46c4-b689-31de17090ed4>.

the species were gathered, and the exact depot location of the drawer.⁹ Based on that experience, a new project was started to treat the rest of the specimens that cannot be digitized in detail in a similar way. In this 30M project (where the M stands for million), the remaining specimens will be digitized on the storage unit level, with a storage unit being a unit in (or on) which objects are stored, such as boxes, drawers, and shelves. The 30 million specimens to which the project name refers are stored in approximately 150,000 of these units, scattered throughout several collection facilities. During the 30M project, the taxonomic and geographical information recorded on the exterior of each of these storage units is registered in the central collection registration system, whereas specimen-specific information will not be recorded. This way, the basic information of all 37 million collection specimens will be digitized, traceable, and virtually accessible by the end of 2015.

A series of additional projects to develop central data registration systems and workflows contributed significantly to these positive results, most notably the implementation of a central collection registration system for zoological and geological specimens and the subsequent implementation of a data model to fit all collection data—the ABCD 2.06 standard for biological collections—thus aligning data registration in all Digistreeets.¹⁰ Additionally, a central workflow for processing the images captured in the various Digistreeets has been developed in cooperation with the Dutch National Institute for Sound and Vision. Through this workflow, the captured images from each Digistreet are centrally processed into user copies of various resolutions, whereas the original images, mostly tiff files, are sent to a facility for durable storage of digital content at Sound and Vision—all overnight. This process is capable of handling 15,000 images encompassing 600Gb per day.

With regard to visibility of the digitized collections, 3 million specimen records have been made available through GBIF.¹¹ In addition, a cross section of about 100,000 digitized objects have been made available to Europeana,¹² whereas the digitized 2D objects are being made available through the European branch of the Biodiversity Heritage Library.¹³ The release of Naturalis' own Web-accessible public collection portal is planned for the end of the third quarter of 2014, in addition to a public API that enables third parties to retrieve sets of data and multimedia from several of Naturalis' core content management systems. To advance the reuse of the digitized collections further, the data and images generated in the Digistreeets are provided under the Creative Commons Zero (CC0) copyrights waiver, effectively placing them in the public domain.

1.11 Conclusion

At this time, the Naturalis program remains one of the largest and most diverse digitization programs in the natural history community. Although it may not seem like it, with little more than 3 million specimens and related objects digitized in 3 years and more than 5 million objects to be digitized in the remaining year, the program is well on schedule. The largest part of these are to be digitized by the Leiden Herbarium Street, which has been operational since July 2013 and has already produced 92% of its targeted images; the data transcription is up to speed as well. The regular Digistreeets will digitize the remaining 1.3 million objects. The digitization program will come to an end in June 2015. By then, Naturalis Biodiversity Center will have made 23% of its entire collection digitally available

⁷This is done separately from the Entomology Street, where a selection from the entomology collections, 850,000 specimens out of a total of 17 million specimens, is digitized in detail. See <http://youtu.be/TywNYCigY0k>.

⁸For the botanical collections a dedicated collection registration system (Brahms) and uniform data model were already in place at the beginning of the program.

⁹See <http://www.gbif.org/publisher/396d5f30-dea9-11db-8ab4-b8a03c50a862>.

¹⁰See http://www.europeana.eu/portal/search.html?query=DATA_PROVIDER%3A%22Naturalis+Biodiversity+Center%22.

¹¹See <http://www.bhl-europe.eu/>.

in detail, with the rest of it on a metalevel, and the institute will possess the digital infrastructure as well as the expertise to keep digitizing in the years after.

REFERENCES

- R. Baird. 2010. Leveraging the fullest potential of scientific collections through digitization. *Biodiversity Informatics* 7, 130–136.
- V. Blagoderov, I. J. Kitching, L. Livermore, T. J. Simonsen, and V. S. Smith. 2012. No specimen left behind: Industrial scale digitization of natural history collections. *Zookeys* 209, 133–146.
- L. G. Carvalheiro, W. E. Kunin, P. Keil, J. Aguirre-Gutiérrez, W. N. Ellis, R. Fox, Q. Groom, S. Hennekens, W. Van Landuyt, D. Maes, F. Van de Meutter, D. Michez, P. Rasmont, B. Ode, S. G. Potts, M. Reemer, S. P. Roberts, J. Schaminée, M. F. Wallisdevries, and J. C. Biesmeijer. 2013. Species richness declines and biotic homogenisation have slowed down for NW-European pollinators and plants. *Ecology Letters* 16, 870–878.
- B. L. Mantle, J. La Salle, and N. Fisher. 2012. Whole-drawer imaging for digital management and curation of a large entomological collection. *Zookeys* 209, 147–163.
- R. K. Pachauri and L. Meyer (ed.). Climate Change 2014, Synthesis Report. (IPCC, 2014).
- N. Raes, L. G. Saw, P. C. van Welzen, and T. Yahara. 2013. Legume diversity as indicator for botanical diversity on Sundaland, South East Asia. *South African Journal of Botany* 38, 265–272.
- R. Tegelberg, J. Haapala, T. Mononen, M. Pajari, and H. Saarenmaa. 2012. The development of a digitising service centre for natural history collections. *Zookeys* 209, 75–86.
- J. P. Van den Oever and M. Gofferjé. 2012. From pilot to production: Large scale digitisation project at Naturalis Biodiversity Center. *Zookeys* 209, 87–92.
- A. Vollmar, J. A. Macklin, and L. S. Ford. 2010. Natural history specimen digitization: Challenges and concerns. *Biodiversity Informatics* 7, 93–112.

Received February 2014; revised April 2014; accepted July 2014