

Migration Option for Citizen Science data from the UCT Virtual Museums to iNaturalist

Plan of action & FAQs Draft June 2020 (for distribution to VM users)

Compiled by: Tony Rebelo, Carol Poole & Domitilla Raimondo (SANBI); Rene Navarro & Peter Ryan (UCT)

1. Background

Citizen Science projects have demonstrated several advantages to scientists and biodiversity conservation initiatives, including that:

1. the increased number of observers allows for more data to be collected from more places than surveys carried out by professional scientists could;
2. there is an opportunity to turn these volunteers into ambassadors for biodiversity as they gain first-hand information on the biodiversity; and
3. ordinary citizens volunteer their time, and many cover their own cost of the required equipment and travel expenses, making this a highly cost effective option for monitoring and surveillance.

All three are important considerations given the limited resources available for biodiversity monitoring in South Africa.

Various entities have developed citizen science platforms to generate foundational data and to engage citizens in science projects. The University of Cape Town (UCT), through the Animal Demography Unit (ADU) and the FitzPatrick Institute of African Ornithology (Fitztitute), has a strong history of developing capacity for citizen science projects and the biodiversity informatics required to support these projects, as well as the development of systems to capture and access the data.

SANBI, with its mandate to monitor and report regularly on the status of South Africa's biodiversity, aims to holistically manage as much biodiversity information as possible to support researchers, conservationists, planners, policy and decision-makers in achieving their goals. SANBI has noted that investment in existing and future biodiversity monitoring programmes is essential to strengthen scientists' ability to detect and report on trends, plan accordingly and manage effectively. This is supported by Strategic Objective 6 of the National Biodiversity Strategy and Action Plan and the National Biodiversity Framework, which is: effective knowledge foundations, including indigenous knowledge and citizen science, support the management, conservation and sustainable use of biodiversity. SANBI started with iSpot South Africa (in collaboration with the Open University) and, after a change of platforms, now manages iNaturalist South Africa (in collaboration with the California Academy of Sciences and National Geographic) – with the express intention to support citizen scientists who wish to allow their data to be used for biodiversity monitoring and assessment purposes. SANBI has also helped to fund other citizen science projects such as the Southern African Bird Atlas Project.

Both SANBI and UCT adhere to the principle of open access of biodiversity information and both parties also wish to increase the generation of foundational biodiversity data and engage citizens in biodiversity science projects, with the understanding that such data can provide very valuable biodiversity information.

2. The current status of the UCT Virtual Museums

The UCT Virtual Museum, with its various Projects (hereafter referred to as the Virtual Museums (VMs) to recognize the individual projects), have been a very successful and popular enterprise, providing much data of scientific value and generating ongoing public interest, goodwill and effort. Originally developed by the Avian (later the Animal) Demography Unit (ADU) at UCT in 2005, the UCT VMs have a strong history of generating excellent foundational biodiversity data and engaging citizens in biodiversity science. Over 560 000 records, by nearly 2 000 citizen scientists, have been submitted to the VMs since inception. The different UCT VMs differed in popularity and success. The relative activity and key features for each VM project are shown in Table 1.

Table 1. Virtual Museum overview (as at noon 05/06/2020). Activity was determined as the percentage of record uploaded in 2019 in relation to total number of submissions divided into three (approximate) quantiles: low (<10%); medium (10–15%); and high (>15%). The number of records on iNaturalist for the same taxonomic group are presented to give a comparison of citizen science platforms – note the relative success of LepiMAP, OdonataMAP and ScorpionMAP. (VultureMAP was subsumed into BirdPix)

Project	Activity level in 2019	Managed by	Number of records	% ID in 2019	Number of observers	iNat records
BirdPix	high	Les Underhill	113,221	100.0	571	120,352
BOP	medium	Les Underhill	518	100.0	185	0
DungBeetleMAP	medium	Mervyn Mansel	2,525	23.7	187	1,929
EchinoMAP	low	Charles Griffiths	2,086	75.9	87	3,568
FishMAP	low	—	1,395	32.0	114	11,299
FrogMAP	low	Darren Pietersen	10,926	94.4	565	11,774
LacewingMAP	high	Mervyn Mansel	3,861	87.2	259	2,100
LepiMAP	high	Steve Woodhall	218,750	78.7	792	67,471
MammalMAP	low	—	34,202	100.0	587	46,617
MushroomMAP	high	Marieka Gryzenhout	10,405	0.6	247	14,752
OdonataMAP	high	Megan Lottjie-Eaton	91,881	99.3	493	11,254
OrchidMAP	low	Craig Peter	7,985	45.4	240	10,040
PHOWN	low	Dieter Oschadleus	23,772	84.0	436	278
ReptileMAP	medium	Darren Pietersen	33,785	89.4	717	27,118
ScorpionMAP	medium	Ian Engelbrecht	4,151	98.6	312	1,943
SpiderMAP	medium	Ian Engelbrecht	12,248	32.4	483	17,014
TreeMAP	low	—	28,860	0.3	195	69,272
TOTAL			600,571		1993	1,008,855

- (iNat Fungi excluding 8,066 Lichens)

The UCT VMs rely on experts for ID, and expert panels adjudicate the data. Some UCT VMs were more successful than others in recruiting experts and ensuring ongoing engagement by experts in the identification process. It must be noted that not all identifications are to species level, so identification is an ongoing and refining process.

The UCT VM was a successful tool for collecting occurrence data across many projects. The UCT VM data (photographs with locality data) were augmented by other sources of data, including checklists, museum specimens and other data. Examples of such ancillary data include: CapeNature, 154 000 sightings; SANParks, 95 000 sightings; Baboon Research on the Cape Peninsula, 41 000 sightings; together adding almost 1 million additional records to the VM databases. These data were already in the conservation authority domain, but not available to the general public prior to the development of the VMs.

Although the data collection costs of the VMs are borne by the citizen scientists, the systems that capture and store the data need professional IT support. The ADU employed two data managers to run the VMs and

the ADU's other citizen science projects: SABAP (Southern African Bird Atlas Project), CWAC (Co-ordinated Waterbird Counts), CAR (Coordinated Avifaunal Road counts), etc. Raising funds to cover the costs of these two posts each year eventually became too much for the ADU in 2017, and the Fitztitute was asked by UCT to try to find funding to continue these very valuable biodiversity monitoring tools and the data managers were transferred to report under the Fitztitute.

3. The decision to offer an option to migrate VM data to iNaturalist

There are currently several challenges with maintaining the UCT Virtual Museums on a long-term basis:

1. There is a scarcity of resources and/or funders available to support numerous platforms. SANBI provided support funding for the UCT Virtual Museums and bird atlas projects through a formal Collaboration Agreement in many phases, but the latest phase with the Fitztitute from 2018 ended on 31 March 2020. SANBI is not able to continue to support the Fitztitute VMs on a sustainable basis due to fluctuating budget availability from central government, recently exacerbated by the COVID-19 crisis. Fundraising for the Virtual Museum as a whole has not proven successful.
2. Some of the Virtual Museums have been relatively inactive, but others are highly active. This is usually a reflection of whether there is a coordinator or facilitator in place or not. Similarly, the amount of identifications added to the VMs is a reflection of the number of experts found able to devote time to the task.
3. The Fitztitute needs to focus its limited resources on bird-related data, and it is recognised that the bird-related data from the Virtual Museums may be best served by being managed alongside the Southern African Bird Atlas Project, which has collected over 18 million bird sighting records over the last 12 years, and is now supporting successful bird atlas initiatives across Africa.
4. The Fitztitute is now faced with the challenging prospect of having to lay off at least one of the two data managers. Dr Rene Navarro, who has maintained the Virtual Museum platforms at UCT since their inception, will leave UCT's employment on 5 September 2020. He will continue to be employed to manage the VMs through funding secured for the OdonataMAP project until mid-2021, but the future of the VMs beyond then remains uncertain.

There are four options (not mutually exclusive) for the Virtual Museum data:

1. The continuation of the VMs using the UCT-based servers without technical support: Thanks to the support from the OdonataMAP project, technical support will continue until mid-2021. The situation after that is unclear, but the servers are new and should be able to continue without support for at least a few years.
2. Transfer management of the individual VM projects to societies: This option would require the different VMs taking over responsibility of managing and hosting their VMs privately on a platform of their choice. The entity taking over the database needs to plan ahead, decide on software/platforms, negotiate with users, do the migration, and negotiate with data end-users (e.g. SANBI) should such end-users have interests in the data. To date only LepSoc (Lepidopterists' Society) has expressed interests in migrating the LepiMAP project into a platform of their choice. The Fitztitute has indicated to LepSoc that SANBI needs to be included in the negotiations in this regard.
3. Archive it: UCT is planning to archive the complete VM collection of images and related data as of August 2020 within the Digital Library Services, UCT Libraries, which will provide a long-term

repository for the VMs. The images will be available for anyone to view, but copyright protection will restrict downloads. The VM data will be available for download from GBIF (Global Biodiversity Information Facility) as well as from ZivaHub, UCT's institutional data repository service.

4. SANBI to negotiate for the data, in whole or in part, to be transferred to another VM (in this case, iNaturalist) for continued use. This is the focus of this document, as detailed below.

The decision as to how to move forward rests with the teams managing each VM and the many citizen scientists who contribute their valuable biodiversity records.

4. About iNaturalist: differences to the UCT VMs and other FAQs

iNaturalist is also a Virtual Museum, and most of the fields and procedures are similar to the UCT VMs. There is an active coordinator within SANBI for iNaturalist, and many of the various projects on iNaturalist have active facilitators. As at 31 May 2020, there are 1 006 210 observations of 30 507 species by 11 264 observers with 6 797 identifiers on iNaturalist South Africa; with 382 481 observations of 20 616 species submitted in between 1 April 2019 and 31 March 2020. Over 100 projects are run on iNaturalist in southern Africa with SANBI support.

iNaturalist is a joint initiative of the California Academy of Sciences and the National Geographic Society (<https://www.inaturalist.org/pages/about>). iNaturalist does not have commercial interests, and exists through donations in funding that support a small core staff, volunteers writing code and offering translation services, and through the many people who have joined the iNaturalist community.

iNaturalist has sophisticated copyright options on images, sounds and observations available for users to choose from, and one can change the default at any time. Users are also able to assign alternative copyright options on any particular image, sound or observation.

Differences between iNaturalist and the UCT VMs

1. iNaturalist summarizes data on site in a large variety of formats and scales, whereas the UCT VMs only provide data summaries at a few mapped scales. iNaturalist does not summarize data in pentads and QDS scales, but raw data can be downloaded for use in mapping and analytical packages.
2. Anyone can download iNaturalist data (subject to conditions imposed by the user and sensitive species), while access to UCT VM data is via the data manager. This makes iNaturalist data more available, but less controlled. Access to Sensitive Species data on iNaturalist is either via SANBI (still to be implemented under an agreement) or by a mutual "trust" agreement between users.
3. Identifications on the UCT VMs are by a small team of experts per taxon group, who exclusively determined the identifications without discussion. On iNaturalist identifications are "crowd-sourced", with ordinary users having the same weighting as field experts. Experts and specialists are thus not able to "ensure" identifications on iNaturalist, but may have to justify their identifications and solicit identifications and agreements from other users, in order to have their identifications accepted.
4. iNaturalist allows unlimited observations fields to be added to any observation, which can be managed by curators via "traditional projects" (there are three types of project on iNaturalist

catering for different curation needs). Among the VMs, only Phown allowed the capture of ancillary data (“nest count” and “nest site”).

5. iNaturalist manages all projects and taxa and communities under a single iNaturalist unit, whereas the VMs are each managed as separate entities. Both systems though are a single set of software packages, so that the different UCT VMs (“Projects”) are the equivalent of projects on iNaturalist. However, on iNaturalist, any user, specialist, society or organization can create and run their own projects. iNaturalist allows curators and managers of projects to provide regular updates, news, summaries and blogs via the project journal, as well as download data.

Frequently Asked Questions about iNaturalist:

a) How popular is iNaturalist South Africa?

The southern Africa iNaturalist growth rate from Jan-May 2019 (134,118 observations for 13,110 species and 3462 observers) to Jan-May 2020 (170,189 for 13,975 species and 4,743 observers – see [early 2020 iNat](#)) is 26% for observations, 6.5% for species and 37% for observers. iNaturalist southern Africa collected over a quarter of the total data in the UCT VMs, with over twice as many contributors, just within the last 5 months. As an example, the Lockdown Chameleon Atlas of Cape Town (see [Cape Town Chameleons](#)) collected 1302 observations over 3 months by 582 observers, compared to the 188 *Bradypodium pumilum* observations in the entire UCT VM. iNaturalist southern Africa is popular and growing.

b) How useful is iNaturalist for checking distribution data?

iNaturalist has a handy feature on the taxon page and observation maps, where one can display all the GBIF data for the species as a layer with the iNaturalist observations. Especially for plants, this is by far the easiest way to access SANBI’s old PRECIS and the latest BODATSA data (even easier than accessing the data on SANBI sites). iNaturalist also allows the input of “taxon ranges” as maps, and “taxon atlases” as magisterial districts, which can be displayed as additional layers on the map. This does require specialist input, but for instance the distribution maps compiled for calculating EOO in Red Listings is suitable for import as “taxon ranges”.

SANBI is currently developing a National Biodiversity Information System which will have species pages with maps showing all historic records collated from South Africa’s collection institutions and recent citizen science collections in one place. We suspect this will be available for the public within the next 6 -12 months.

c) Is taxonomy up-to-date on iNaturalist?

The iNaturalist taxonomy is reasonably up to date, but may be taxon-specific in terms of accuracy. There are automated procedures to check matches from the iNaturalist dictionary to the reference dictionaries (for instance, the plant dictionary follows POWO (<http://powo.science.kew.org/>) – although deviations are allowed, these need to be documented. Updates suggested by citizen scientists on iNaturalist are fed through to both POWO and BODATSA (SANBI’s botanical backbone database) where discrepancies arise. But the iNaturalist dictionaries are not designed to be reference works: they follow standard references: please see <https://www.inaturalist.org/pages/curator+guide#authorities>.

For animals, SANBI uses the SPECIFY backbone, and we can occasionally cross-check and ensure that the respective dictionaries are in sync.

Fungi are a problem as there is no international database, but many of the specialists on the many databases that do exist regularly use and curate iNaturalist taxonomy.

If you have any specific taxonomy-related questions for your group, please contact Dr Tony Rebelo at SANBI for more information (t.rebelo@sanbi.org.za). For specific taxonomy issues on iNaturalist, please add a flag on the taxon and curators will address the issue.

d) Are crowd-sourced identifications accurate?

Crowd-sourced identification allows for more experts from the global pool to make identifications and also faster identifications for the easily identified species. Although southern Africa only joined iNaturalist in 2018, it now has many local taxonomic experts utilising the site and therefore the quality of identifications has radically improved. In the past two years we have tested this ability to cope with over a million observations over four days during the City Nature Challenge. Experts and specialists around the world regularly use iNaturalist, and help with the identifications (and dictionary curation), and interested observers learn and become better at making identifications, and increase the knowledge pool.

There is also an ever-improving Artificial Intelligence tool that suggests identifications to the more common species. It is not so accurate with the rarer species, but the majority of users – especially less experienced beginners – seldom encounter these anyway. Furthermore, data quality checking is also open: so that any users can interact with observers when substandard data is uploaded. Any mistaken identifications are usually sorted out by more knowledgeable users within hours, often long before experts get to them.

A documented issue is when schools and universities use iNaturalist for curriculum-related field work, and scholars and students collaborate to push incorrect identifications to research grade to improve their marks. Solutions to this issue are being investigated and experimented with.

e) How secure is iNaturalist if other citizen science platforms fail in the long-term due to funding challenges?

Several virtual museums have had to close or power down due to challenges (iSpot, Bowerbird) and even iNaturalist is not immune to the possibility. SANBI is negotiating for the possible outcome that iNaturalist too might collapse, and that all southern African data will be transferred to SANBI if this were to happen. The platform and software running iNaturalist is shareware, and can be cloned off without any copyright costs at any point, which is one of the reasons why SANBI has chosen to use iNaturalist.

f) How are the data on iNaturalist being used?

iNaturalist data are heavily used, in part because it is so easy to filter and download the data.

The southern African iNaturalist data are already available to any researcher or user, directly via:

- the front end filters (<https://vimeo.com/331279497>),
- through projects (<https://www.inaturalist.org/pages/help#projects1>),
- via the APIs (<https://www.inaturalist.org/pages/developers>),
- or through GBIF (<https://www.inaturalist.org/pages/help#GBIFdata>).

We welcome the sharing of Citizen Scientist data with all users. There are also widgets to allow the display of personal and project data on your own or official websites. Here are some examples of how iNaturalist data are being used currently:

- It is the data capture tool for CREW (Custodians of Rare and Endangered Wildflowers), specifically for the monitoring and status updating of Red List plant species.
- The iNaturalist data was extensively in South Africa used for conservation planning of the Marine programme (SEAKEYS).
- iNaturalist data are being used for the current reptile Red List assessment.
- There are dozens of specific monitoring projects on iNaturalist using data (see [South African Projects](#)), including Roadkill, Alien surveys, Western Leopard Toad monitoring, Chameleon surveys in Cape Town, and so forth.
- iNaturalist is also featuring prominently in descriptions of new plant species – heavy use is being made of citizen science to collect distribution data between discovery and description. For instance there are abundant data on over 10 undescribed species of *Indigofera* (Undescribed Indigofera). Two plant species listed as Extinct have been rediscovered on iNaturalist in the last 12 months.
- Here is a direct link to some 677 scientific articles using iNaturalist data: iNat_literature@gbif.

g) How does iNaturalist protect sensitive species information?

Users can obscure their own data when and if they want to. They can also “trust” particular users to be able to see their data. iNaturalist is quite sophisticated in this regard.

Sensitive Red List species are additionally obscured by the system, but researchers can apply to access precise coordinates for these data through SANBI.

Be aware when downloading data to include those data fields which indicate if the downloaded locality data are obscured or not. Obscured data are assigned random coordinates within a 28x28km grid cell in which the actual locality resides, and may confusingly appear as precise localities to naïve users.

h) What happens to the non-image data on the VMs?

h.1 Institutional data.

The VMs included data that are not part of the “Visual Museum” (i.e. not citizen science image related observations). These are largely historical records from specimen data in museums, herbaria, private collections and NGOs, but also from other surveys and literature. These data are not suitable for migration to iNaturalist, as they don’t contain images.

Many of these datasets were digitised specifically for SANBI use in red listing activities, which are now completed. However, there is a major problem with these datasets. Namely that they are a version of databases maintained by the holding institutions. It is impossible to efficiently maintain different versions of these databases, each with different levels and sets of data updated or corrected. The only solution is that all updates and corrections need to be made by the holding institutions. All errors detected, and updates to the data, must therefore be sent to the holding institute for incorporation. Every time one needs to use such data, one should obtain the current version (which should be quoted as to the date extracted in any publications), inform the holding institution of any errors detected, and then discarded.

SANBI is aware of these issues. To this end SANBI has:

- Obtained funding and staff for collections to be digitized, some under the UCT VM programme;
- Has set up taxonomic backbones as BODATSA (plants) and SPECIFY (animals), so that all institutions are using the same latest names, and any updates can be automatically updated where possible (some VMs are still active with this); &

- Is setting up a single hub at SANBI, much as GBIF does globally, where data can be accessed live, from all participating institutions in South Africa (and where possible southern Africa).

To facilitate this, SANBI has been in contact with the original data owners, and negotiated new agreements for data sharing. The Natural Science Collections Facility, established in 2016, has the role of ensuring that digitized data in collections is available for use through data sharing agreements, and that database systems are compatible for updating from the national backbones and allow easy sharing and serving of data through the National Biodiversity Information System.

SANBI is aware that some institutions lack the capacity to regularly update their databases, and corrections returned to them may not immediately be added to their collections. SANBI has been involved in digitizing many of these collections, and is exploring options to help alleviate these problems.

In short. These extra data are not suitable for migration to iNaturalist. These extra data are historical databases that are outdated, increasingly inaccurate and should be abandoned. For research and analyses, current data should be accessed via the SANBI hub, and any issues should be fed back to the respective database curator.

SANBI acknowledges the role of the UCT VMs in collecting these indispensable databases for the initial Red List assessments, and has put in place systems to ensure that current data are available to everyone.

h.2 Personal data.

iNaturalist requires pictures or sounds for identification verification. Only observations with media will be migrated. However, many users have observation data without images or sounds, that they feel are accurate enough to submit to VMs. iNaturalist accepts such data in various ways.

1. Straight observations, without verification – e.g. [observations without media](#) (8,854 currently in southern Africa). You are welcome to upload your observations.

2. Checklists – these are associated with places: e.g. https://www.inaturalist.org/check_lists/982798-Tokai-Park-Check-List (open the “other checklists” below the filter bank for examples). You can add checklists to any places, where they may be useful, and add annotations and notes. iNaturalist keeps checklists up to date with taxonomy changes, and integrates checklists and data for a comprehensive place checklist.

iNat does not cope with specialized data such as Braun-Blanquette and plot data, but SANBI has tools for storing these.

5. Challenges with the migration of data from UCT VMs to iNaturalist

iNaturalist has already migrated data from southern Africa from iSpot, and Australia from Bowerbird onto its system. However, the migration option does require buy-in from both the VM curators and the individual users. SANBI currently has the resources and capacity to facilitate such a migration, as a service to those projects and users wishing to migrate. Note that migration will only copy the data and images and sounds over – the original UCT VM data and images and sounds will not be deleted.

It must be noted that the migration of data to iNaturalist is best done while Dr Navarro is still at UCT, and therefore we are strongly recommending that the migration takes place before the end of August 2020.

The agreement of use of the UCT Virtual Museums did not cater for data and image transfer, and did not make future provision for the data or images in the event that the systems would cease or be closed.

However, this challenge can be overcome because of the following:

- Data transfer is not an issue, as the data is open under a Creative Commons Attribution 4.0 International License and can therefore be transferred from the VMs to SANBI or iNaturalist or any third party.
- Under the terms of agreement, the UCT VM was granted permission to use the pictures (images) on the website, but copyright remains with the observers. When the VMs started in 2005, creative commons licences were in their infancy, and contributors were concerned with protecting the intellectual property rights of images uploaded: they were happy to upload if they retained copyright. The VM policies regarding image copyright was never changed. For purposes of migration this should not be an issue, as users can choose from iNaturalist's diverse copyright options a licence of their preference, during the creation of their accounts. Users who do not wish to migrate will not be affected.

6. Summary of requirements for the proposed migration

In order to migrate data from the UCT VMs to iNaturalist, should users wish to do so, the following are needed:

1. SANBI needs a Collaboration Agreement with the FitzPatrick Institute of African Ornithology *vis a vis* data provision and accessibility for migration. This is in progress.
2. UCT needs to approach the curators, champions, experts and users of the different Virtual Museums to offer the migration option. Ultimately we need an understanding of which users might want to migrate to iNaturalist. The FitzPatrick Institute of African Ornithology will therefore send out a first and follow-up calls to VM users (see Step 1 below).
3. Users who wish to make use of the migration option will need to register on both iNaturalist to assign their usernames, agree to conditions and set their copyright preferences, and provide SANBI their contact details and permission to migrate their data to iNaturalist.
4. SANBI needs an agreement with iNaturalist *vis a vis* the migration process, and to establish costs. Negotiations are also underway in this regard and will be confirmed when the total number of records to be migrated is confirmed.
5. SANBI needs to oversee the entire migration process, ensuring that the correct data is migrated timeously and completely. The FitzPatrick Institute of African Ornithology's responsibility here would be to ensure that the data are available online and accessible to iNaturalist programmers to enable migration. Ideally data managers from the VMs needs to be available to assist with snags during the process.

7. Activities for the migration

CS = Citizen Science.

iNat = iNaturalist (www.inaturalist.org) – specifically the southern African community, for which an agreement between iNaturalist and SANBI is in process

User = data contributor to the VMs, including champions (who promoted the VM), experts (who provided identifications) and curators (who maintained data integrity)

VM = the individual packages the ADU/Fitz (vmus.adu.org.za/) ran. Thus each of MammalMap, OrchidMap, etc. is a separate VM (if used collectively = VMs)

VM data – only the Citizen Science component of the ADU-VM data. The picture submissions. This does not include the museum and non-photograph portions of the databases (except for those data associated with the pictures).

The steps involved are:

STEP 1: Evaluate user and curator interest in the migration option

1.1 Assessing interest via a questionnaire. UCT and SANBI are highly respectful of the VM users' personal motivations and the background of each VM project. A questionnaire has therefore been set up on SurveyMonkey in order to collate the sentiments from all VM users and curators. It has two components:

- the champion/curator collective decision for the individual VM project
- the individual user decision.

So even though a particular VM may not decide to migrate, a particular user might request their data from that VM to be migrated (and vice versa: the collection migrates, but not a particular user).

The questionnaire consists of the following questions so that we can evaluate these issues:

(the questionnaire is available at <https://www.surveymonkey.com/r/6K8YTYW> until the end of June 2020.)

1. What is your role in the Fitzitute VMs?:
 - I am a VM observer (I have posted images)
 - I am a VM curator (I manage one of the VM projects)
 - I am a VM expert (I am an identifier on one of the VM projects)
2. Would you like to make use of this once-off opportunity to have your data copied (migrated) to iNaturalist:
 - Yes - I would like to migrate my data to iNaturalist, and I understand that this will require me to register on iNaturalist and provide SANBI with my VM username, iNaturalist username & e-mail address for communication. I agree to give SANBI permission to undertake the migration on my behalf. I will choose my copyright options for photos, sounds and observations on iNaturalist when I sign up.
 - Yes - I would like to migrate my identifications to iNaturalist, and I understand that this will require me to register on iNaturalist and provide SANBI with my VM username, iNaturalist username & e-mail address for communication. I agree to give SANBI permission to transfer my identifications on my behalf.
 - No, I do not want to migrate data. I understand that this means that my VM data will be archived on the UCT library system. I understand that there is no technical support for the VMs currently available beyond mid-2021 and that the option to migrate data automatically to

iNaturalist is a once-off offer.

[I will skip Question 3]

3. If you want to migrate your data to iNaturalist, please provide the following data:
 - Current Fitz Virtual Museum username and number:
 - Registered iNaturalist username (please first register, and complete your profile and select your copyright options)*
 - email address (so SANBI can keep me informed of progress, and address my specific issues)
 - I hereby give permission for SANBI to facilitate the migration of my VM data to iNaturalist. My first and last names.
4. Data to Migrated?
 - Please migrate all my observations
 - Please only migrate my data from these projects:
5. Do you have any further comments, questions or special requirements relating to the migration of data from the Fitz VMs to iNaturalist?

**Note: it is crucial for users who wish to migrate to register on iNaturalist (if they are not already registered). This must be done first otherwise chosen usernames may be occupied and transfer of data to the incorrect user may occur. Being registered on iNaturalist will ensure that we have the correct user name to migrate data to and also binds the user to the terms and conditions of iNat – which precludes most possible issues arising, and should address most copyright concerns.*

The link to this questionnaire will be emailed with a letter by the Fitztitute to all VM users, placed on the VM Facebook pages and on the banner page of the UCT VMs.

1.2 The Fitztitute and SANBI will attempt to locate users where their email address on the VM system is no longer valid, and will assist in deciding to contact next of kin for permission, automatically transfer images and data, or to not migrate images and data.

1.3 All responses must be received by closed of business on **Monday 30 June 2020**. If a VM user does not answer the survey, it will be assumed that the user is not interested in the migration option.

STEP 2: Contacting individuals who wish to migrate data

2.1 SANBI will evaluate user and curator interest in the migration option according to the survey results to determine final decisions on which data/VMs to migrate; and to facilitate understanding of scale of the migration for iNaturalist.

2.2 SANBI will contact users who wish to migrate to confirm the following:

- Whether the user wants all their data transferred, or just a particular VM. We will not allow specific observations not to be transferred: everything (or nothing) will be transferred – users may remove specific instances after migration if they do not want specific observations transferred. In addition, all data will be transferred – there is no option to selectively transfer particular fields. The onus is on the user to execute any actions outside of the simple migration.
- Users will need to:
 - Choose the type of copyright the user wishes as default for the images, sounds and observations on iNaturalist. The user needs to set this as default on iNaturalist before the

migration commences. Specific instances can be adjusted by the user after migration, but the onus is on the user to do so. This part of the contract does not involve SANBI.

- Manage after migration any duplication in data that will occur for observations posted on both the VMs and iNaturalist prior to migration. One of the duplicates (user to decide) should be deleted. Duplicates are not encouraged. The user shall be responsible for this.
- All identifiers (expert panel) should preferably also agree to migration. The ID is part of the observation, and therefore belongs with the identification. However, if the identifier does not migrate, then the ID becomes a comment posted by the user, and not an identification. This may mean that the identification on iNaturalist will not match the identification on the VMs as a consequence.

2.3 and 2.4 SANBI and UCT will continue to follow up with users who have not been contactable, and/or asked more questions via the SurveyMonkey system; with the understanding that all users who wish to migrate are registered with iNaturalist and that SANBI has a complete understanding of data migration needs **by 31 July 2020**.

STEP 3: Data migration preparation

3.1 Mapping of data fields. The data fields in the VM need to be matched to the iNaturalist data fields, and non-matching fields catered for so that no data is lost. These options are outlined in Table 2.

Table 2. Mapping of ADU data fields to iNaturalist fields. Fields are grouped into “batches” based on the APIs used for their transfer. The order is: Base data with Images and Sounds, Identification data, Comment data, Observation field data.

ADU data field	iNaturalist data field	Transfer batch
Record number	None – will get iNat acc #; can be stored as “observation field .e.g. “UCTVM Number”	(Observation field data)
User	User	Base data
Province	Not transferred: iNat has GIS system	-
Country	Not transferred: iNat has GIS system	-
Closest Town	Incorporate into Locality Guess	
Locality	Locality Guess	Base data
Locus	= grid cell : superfluous. Can be stored as observation field e.g. “UCTVM SA Grid System”	(Observation field data)
Altitude	Not in iNat: use Observation Field (Alt (m))	(Observation field data)
Source (of coordinates)	Not in iNat: use Observation Field e.g. UCTVM Locality Source”	(Observation field data)
Accuracy	Location accuracy (both in m as radius)	Base data
Date of record (3 fields: Yr, Mon, Day)	Date observed (concatenate date fields)	Base data
Number of photos	Superfluous: can be used as check.	-
Notes	Description	Base data
Last updated	Superfluous (iNat manages this differently)	-
Sp. Code	Superfluous (to species name)	-

Species name	Superfluous: iNat has its own “summary” ID system – only actual IDs will be transferred – add as Observation field “UCTVM ID:”	(Observation field data)
Common name	Superfluous (to species name)	-
Family	Superfluous (to species name)	-
Record Status	Irrelevant to iNat. Can be stored as observation field e.g. “UCTVM Record Status” *1	(Observation field data)
Collectors species ID (Identity)	Species	Identification data
(User)	User name	Identification data
(Date of record)	Date identified	Identification data
Additional observers	capture in Description	Base data
Institution	Irrelevant (but only ADU data to be used?)	-
Record URL	Irrelevant to iNat. Can be stored as observation field e.g. “UCTVM Record URL”	(Observation field data)
VM accession date *2	Accession date	Base data
Latitude	Latitude	Base data
Longitude	Longitude	Base data
Location error	Locality accuracy	Base data
Location name	Location guess	Base data
Images (only full size – or original)	Images	Images
Sounds	Sounds	Sounds
Comment number	Irrelevant – iNat will provide own accession number	-
Comment by:	User name	Identification data (or Comment data)
Comment:	ID description	Identification data (or Comment data)
Comment Date:	Date identified	Identification data (or Comment data)
Comment Species name:	Species	Identification data (Comment without an ID to go to Comment data)
Comment Species code:	Superfluous (to species)	-
Project (VM name)	Best done as a tag	Base data
Phown: Nest count	Create a specific Observation field “Phown Nest Count”	(Observation field data)
Phown: Nest site	Create a specific Observation field “Phown: Nest site”	(Observation field data)

*1: values: Accepted; Introduced; Historical; Questionable; Pending; Deleted; Hidden.

*2: to be added to Darwin Core API

3.2 Mapping of species list

- UCT (Rene) to provide species list (Species name, sp. code, common name) for all species for all UCT VMs
- SANBI (Tony) to curate iNat lists for a 1:1 mapping. i.e. latest scientific names, synonyms and common names.
- Issues with mapping VM to iNat to be negotiated by Tony with UCT VM curators (Rene to provide lists of curators – at previous step).

3.3 SANBI will compile lists of users who have agreed to data migration.

- Calculate proportion of data to be transferred.

3.4 SANBI will also contact iNaturalist and find dates for the migration and finalize procedures. Negotiate the following:

- All data will be transferred. iNaturalist wants to minimize observation fields with extra data.
- Date of submission will be the date of submission to the VM, and not the date of migration. (this can be done – was done for iSpot, but not for Bowerbird (Australian VM))
- Costs: iNat have waived migration costs, provided that the data and images are ready for transfer in appropriate formats. This will be re-examined when the full scale of the migration is known.
- Determine the optimum way to transfer the data (field mapping, steps, etc.)

3.5 The FitzPatrick Institute of African Ornithology will have to make sure that all data (i.e. the databases and the images) are available to the iNaturalist teams via APIs remotely over the web or other means during the period under question. The API currently available is api.adu.org.za and identification data will be provided separately.

3.6 UCT VM projects. Each of the UCT VM data can be collected in a project on iNaturalist. However, the observation fields (e.g. "UCTVM Number") will exist on iNaturalist to pull together data via filters, so these will be superfluous. iNat has advanced filters allowing filtering, viewing or extraction of data by any taxon.

- However, UCT VM champions may be interested in recreating their teams on iNaturalist. iNaturalist projects allow easy data collection (by taxa, by region – e.g. southern Africa, Namibia, Western Cape, Cape Town, Meadowridge Common) with superb tools for commenting on observations, providing feedback, news, progress reports, and other useful information for managing taxon or area-based projects. However, these projects will seamlessly collect both the UCT VM and iNaturalist data.
- Tony to provide assistance to any curators wanting to set up such projects.

STEP 4: Data migration

- To be managed by iNaturalist
- SANBI (Tony) and UCT (Rene) to be available for any issues arising.
- This will take about a month, assuming that data and images are accessible.

STEP 5: Mopping up

- SANBI to compile snag list with iNaturalist assistance
- Determine how best to resolve these. Ideally everything will go smoothly, but by definition, and problems will be unexpected (or we would have catered for them already). So it is not possible to put a time or type value on these.

8. Timeframes

The summary of the above activities and their timeframes is as follows:

Table 3. Time frame of the migration of the VMs to iNaturalist

Step in migration process	Responsible party	Timeframe for step	Timeline (Deadline / completion – by end of) All in 2020				Challenges/risks/problems with step
			June	July	August	September	
0.1 Collaboration Agreement Fitz & SANBI	SANBI & UCT	1 month	Jun				SANBI takes some time to put such agreements in place
0.2 Contract SANBI & iNaturalist	SANBI	2 month	Jun	Jul			
1.1 Users & Experts notified of questionnaire	Rene	1 day	Early Jun				
1.2 Chase up deceased and missing users	Rene	1 month	Jun				
1.3 Users complete questionnaire	VM users	1 month	30 June				
2.1 Evaluate user and curator interest in the migration option to determine final decisions on which data/VMs to migrate	Tony (in consultation with UCT)	1 week		Early Jul			
2.2 Contact users who wish to migrate to confirm outstanding information	Tony	2-3 weeks					
2.3 Chase up users not replying, esp. power and key users	Tony & Rene			Jul			
2.4 Deadline for all users to be registered on iNaturalist, confirmed the data they wish to migrate, confirmed their copyright statements.	Users who wish to migrate	1 month		31 July			
3.1 Map fields from VM to iNat	Tony & Rene	1 month	Jun				
3.2 Ensure species dictionaries match	Tony & Rene	1 month	June				
3.3 Compile lists of users and data to be migrated	Tony	Ongoing to end July	Jun	Jul			
3.4 Finalise arrangements with iNaturalist	Tony and iNat	By 31 July		End July			
3.5 Make sure all data are available to iNat	Rene	1 month	Jun				Check ID transfer
3.6 Assist with setting up of projects on iNat	Tony	Ongoing	Jun	Jul	Aug		
4. Data migration	iNaturalist + Tony + Rene	1 month			Aug		Rene leaves UCT employment on 31 August
5. Snags	iNaturalist – Tony	1 month				Sep	