Unique Identifiers in an Online Permit and Monitoring System

Paul Oldham

24 May 2016

Table of Contents

## Introduction

This section discusses the use of unique identifiers for permits and related documents and the application of the principle of triple redundancy.

Unique identifiers associated with permits and associated documents will allow for sets of documents linked to a specific application to be stored within a retrievable register or file history. As such, unique identifiers will enable internal coherence within the permit system.

Unique identifiers linked to a permit and corresponding MAT (an ABS contract) are also the key to monitoring the outcomes of research including the genetic resources. We propose a system consisting of the following unique identifiers

1. Country Codes, dates and unique numbers (e.g. BS20151234)
2. Bar codes
3. QR Codes (Quick Response codes)
4. Html embed codes.

In proposing the use of unique identifiers we also suggest the application of the principle of triple redundancy, notably in labelling. The principle of triple redundancy is a well established engineering principle that involves three different systems performing the same function. In the event that one system fails a second system takes over the same function. If the second system fails then the third system takes over. Given that the failure of all three systems is unlikely the function of a particular system within a wider system is maintained. It is important to note that triple redundancy does not guarantee that a system will not fail completely, rather it reduces the likelihood of failure.

The principle of triple redundancy is particularly relevant to an online permit and monitoring system in connection with the use of unique identifiers to maintain the link between a permit, mutually agreed terms and samples of biological material that are subsequently deposited with a collection, transferred to a collection in a third country and potentially utilized by third parties.

The problem that triple redundancy helps to address is in ensuring the maintenance of the connection between a permit and a document containing MAT (an ABS contract) and the materials that are collected and transferred. These materials may be either physical or electronic (e.g. DNA and amino acid sequence data). The scope of genetic resources is likely to be defined in domestic ABS legislation.

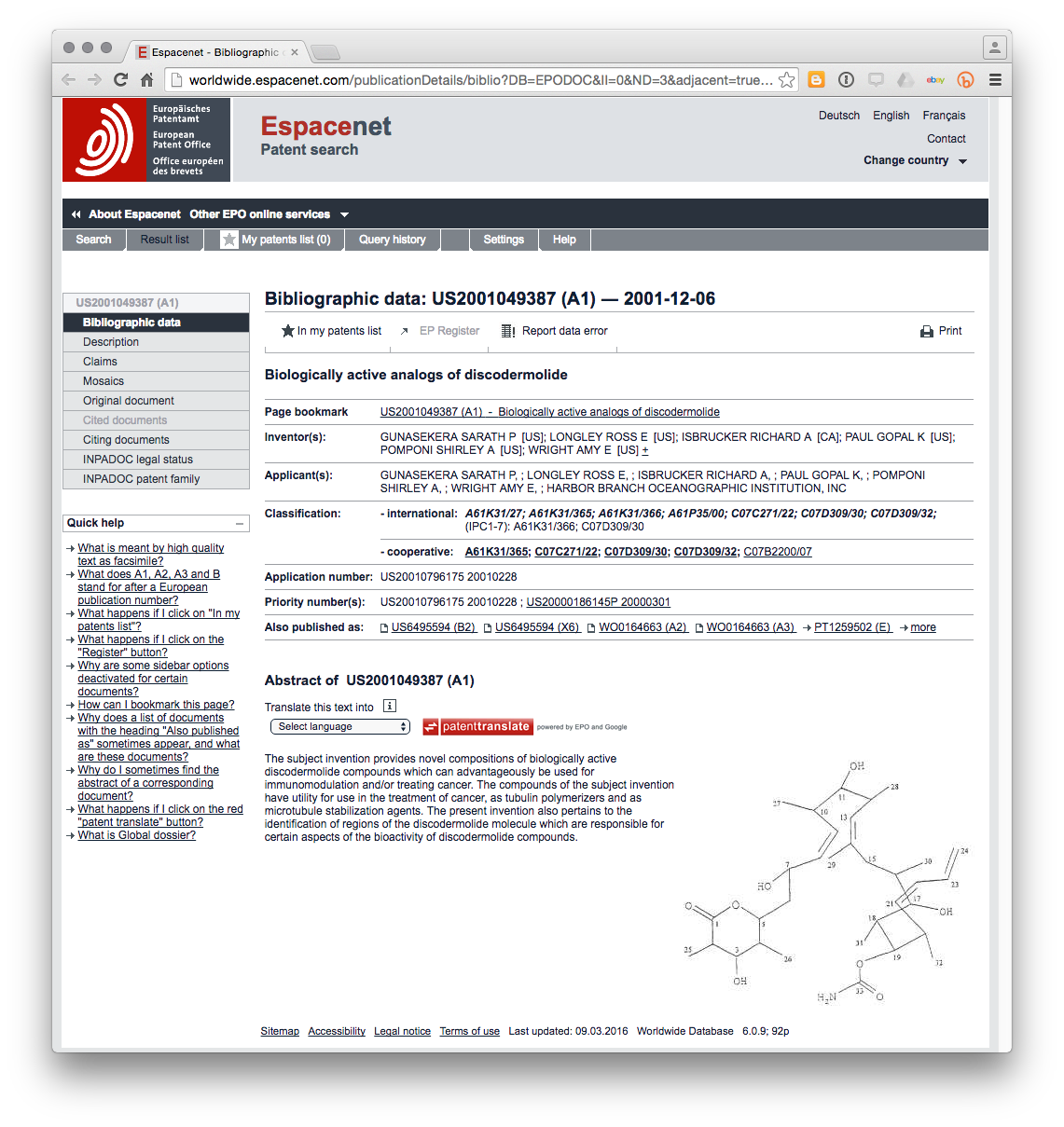
### Country Code Identifiers

All countries possess a unique two letter country code defined in international standard ISO3166-1 alpha 2 (e.g. see [this Wikipedia entry](https://en.wikipedia.org/wiki/ISO_3166-1) and the [ISO browsing platform](https://www.iso.org/obp/ui/#search) and select Country codes and search).

For example, the standard two letter country code for the Bahamas is BS, Kenya is KE, Uganda is UG and South Africa is ZA. If these country codes are combined with a date (YYYYMMDD) and a unique number (1234) a unique identifier will be generated. In this case we will simply use the year (YYYY) to generate the following identifiers.

* BS20151234
* KE20151234
* UG20151234
* ZA20151234

What is clear from this is that a single and distinctive unique identifier (country - year - number) has been created for each permit and corresponding MAT that is immediately distinguishable using the country code from similar numbers issued by other authorities. The combination of the country code, the date and a numeric identifier (country - YYYMMDD - number) is already used very successfully to keep track of approximately 90 million patent documents in countries around the world and is recommended. The example below is for a patent application from the United States that makes reference to collections from the Bahamas and can be viewed on the main worldwide patent database [espacenet](http://worldwide.espacenet.com/publicationDetails/biblio;jsessionid=J6iiTFoqvaPnW3CcNQbj9A-N.espacenet_levelx_prod_0?FT=D&date=20011206&DB=EPODOC&locale=en_EP&CC=US&NR=2001049387A1&KC=A1&ND=4) operated by the European Patent Office [here](http://worldwide.espacenet.com/publicationDetails/biblio;jsessionid=J6iiTFoqvaPnW3CcNQbj9A-N.espacenet_levelx_prod_0?FT=D&date=20011206&DB=EPODOC&locale=en_EP&CC=US&NR=2001049387A1&KC=A1&ND=4).

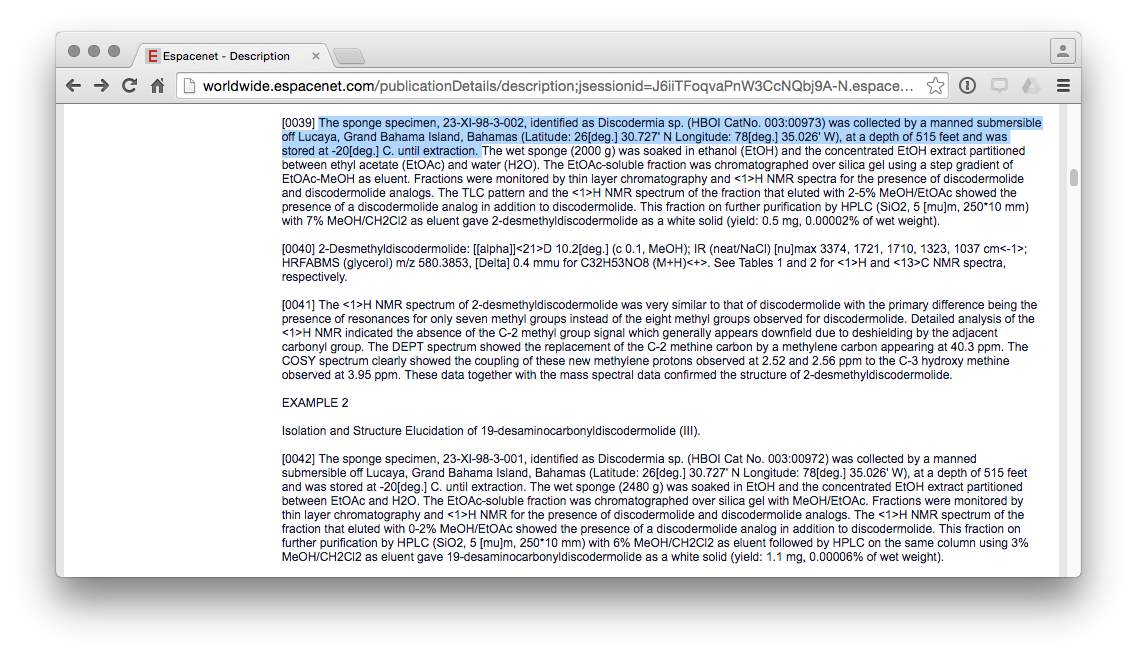


In this example we can see that the unique document identifier US2001049387A1 consists of the following [US] – the country code – [2001] the year of application – [049387] – the unique numeric identifier, and [A1] known as the “kind code” for the type of document (in this case a patent application).

The strength of this system is that the components combine into a unique identifier that is:

1. distinctive, and;
2. easy to retrieve.

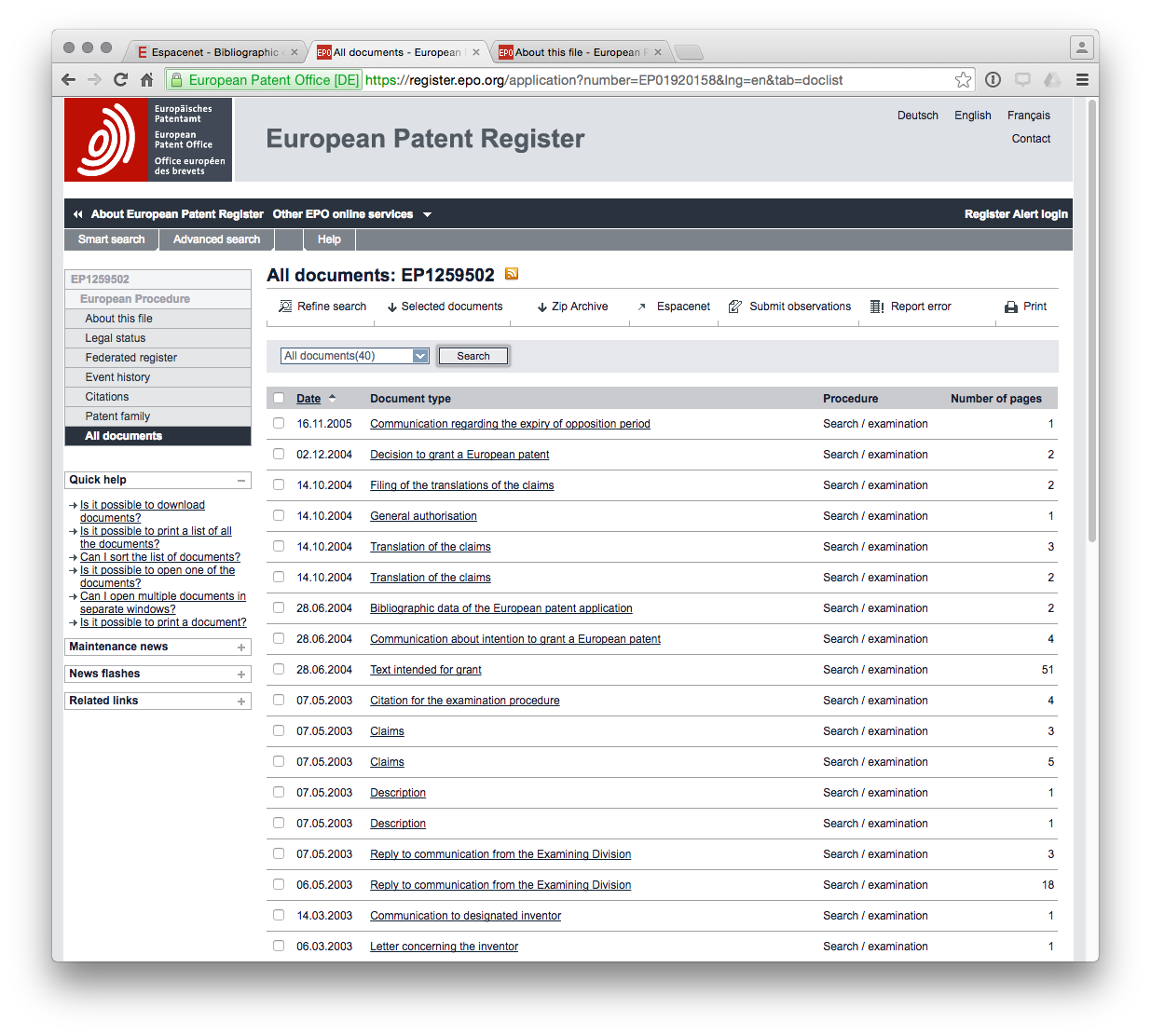
The relevance of this type of numbering system becomes apparent when we consider the section of this patent application that makes reference to biological collections in a country. In this case the example is from the Bahamas (See the first paragraph highlighted or read directly [here](http://worldwide.espacenet.com/publicationDetails/description;jsessionid=J6iiTFoqvaPnW3CcNQbj9A-N.espacenet_levelx_prod_0?CC=US&NR=2001049387A1&KC=A1&FT=D&ND=3&date=20011206&DB=EPODOC&locale=en_EP)).



In this case, the applicants make reference to a sponge specimen 23-XI-98-3-002 and a HBOI CatNo. 003:00973. Under the proposal advanced in this document the ABS domestic legislation, permit and associated MAT could require recipients of a permit to disclose the permit number (e.g. BS20151234) in any patent application arising from the research.

## Unique Identifiers and Registers

An important feature of this system as deployed within patent databases is that the unique identifier is used to store all documents that relate to a particular application over time. In this case the United States application was also filed in Europe at the European Patent Office and this document can be identified in the [patent family](http://worldwide.espacenet.com/publicationDetails/inpadocPatentFamily;jsessionid=J6iiTFoqvaPnW3CcNQbj9A-N.espacenet_levelx_prod_0?CC=US&NR=2001049387A1&KC=A1&FT=D&ND=3&date=20011206&DB=EPODOC&locale=en_EP) of the US document. European Patent Application [EP1259502A2](http://worldwide.espacenet.com/publicationDetails/biblio?DB=EPODOC&II=6&ND=4&adjacent=true&locale=en_EP&FT=D&date=20021127&CC=EP&NR=1259502A2&KC=A2) was filed prior to the introduction of the year into the identifier. However, for our purposes the document is important because in Europe it is possible to access all documents linked to that application, including formal communications between the patent office and applicants in the [European Patent Register](https://register.epo.org/application?number=EP01920158&tab=main). The figure below displays the list of documents linked to this identifier and can be accessed [here](https://register.epo.org/application?number=EP01920158&lng=en&tab=doclist).



It is immediately clear that the use of this unique identifier system allows all documents related to that identifier to be linked together into an electronic register (file history) for each application. We propose a very similar system where a unique identifier is used to link all documents arising from a permit application both internally within the system and for samples, publications and patent applications arising from the research. Specifically, the use of the unique identifier in labels for samples and specimens would facilitate the monitoring of compliance by users.

This type of unique identifier is simple, easy to use and robust over time. For that reason it is recommended as the first component of a triple system.

### Simple Barcodes

A second system is the standard bar code which can be optically scanned to reveal basic information. The bar code system was developed in the 1960s and became ubiquitous for tracking and scanning products from the 1970s onwards. A range of bar code types are available along with free bar code generators. A simple example using a free tool is provided below. This bar code could be attached to documents and samples with basic information that could be encoded into the bar code. The advantage of a simple bar code is that it can be scanned by a machine.

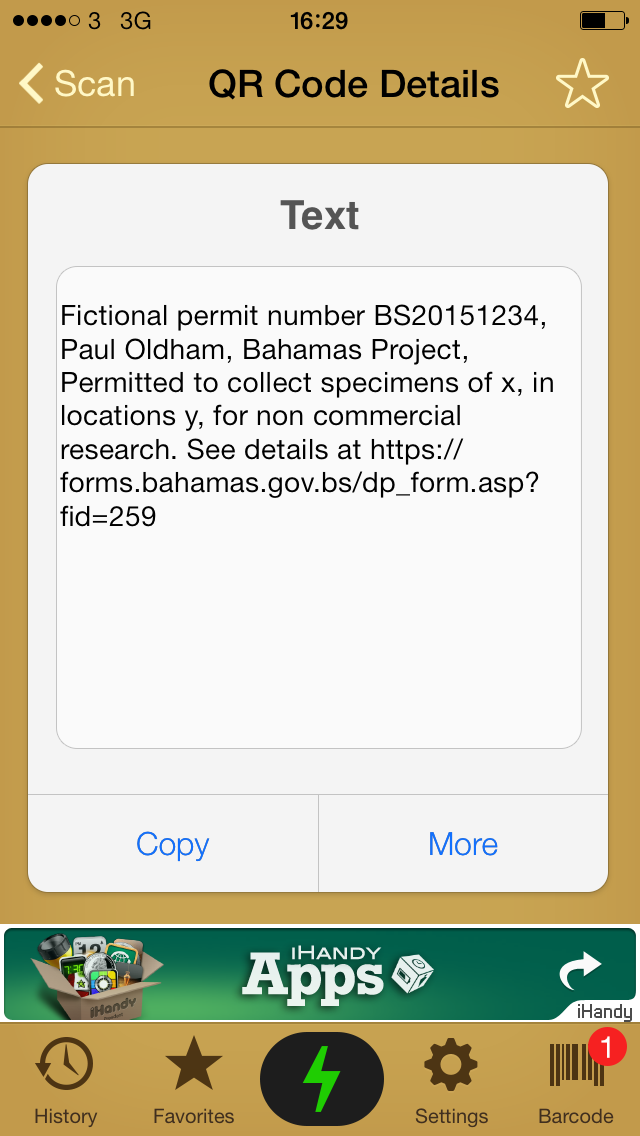


### Quick Response Codes (QR Codes)

The third system is QR (Quick Response) codes which provide a much greater level of embedded detail than bar codes and can be used to embed geographic and other information. An example is provided below.



The information in this QR code can be read using free software on a smart phone such as an iPhone, Android phone, or tablet, as can be seen in the image below (using QRReader on the iPhone). QR codes are normally open. However, encryption of data may potentially be desirable so that only authorised users (police, customs, port authorities) can scan the contents.



The combination of the three systems would meet the requirement for triple redundancy. The main issue is not likely to be the means to generate the identifiers and codes but with ensuring that both authorities and applicants consistently use the identifiers and codes in documentation (including sample documentation) linked to a permit and associated MAT.

Finally, the use of identifiers is likely to be desirable in cases where ABS domestic frameworks include DNA and amino acid sequence data arising from research under a permit and associated MAT. This requires further exploration but is briefly considered below.

## DNA and Amino Acid Sequences

Additional options include a requirement to use the basic unique identifier (BS20151234) in the documentation entered into databases for DNA and amino acid sequence data or html embed codes (for web publications).

In the case of DNA data an example of the existing use of identifiers, and the ability to monitor DNA sequence and amino acid information is provided by the [Barcode of Life Database (BOLD)](http://www.boldsystems.org).

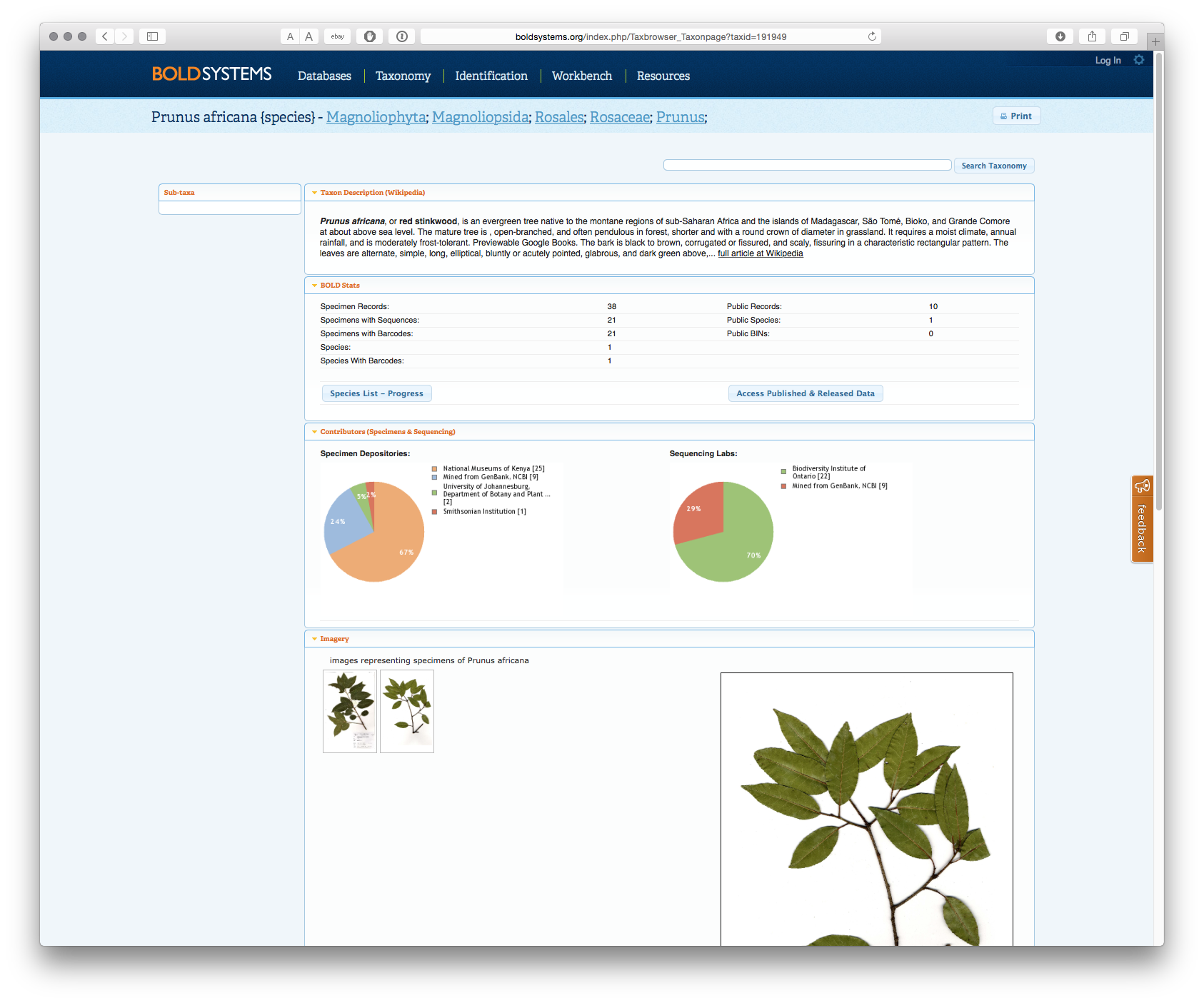
Information using species names or identifiers can be retrieved using the taxize package in R using [RStudio](https://www.rstudio.com) which generates the following link for a search for a species name:

library(taxize) # load the taxize package  
get\_boldid(searchterm = "Prunus africana") # search for the dna barcode id for a species

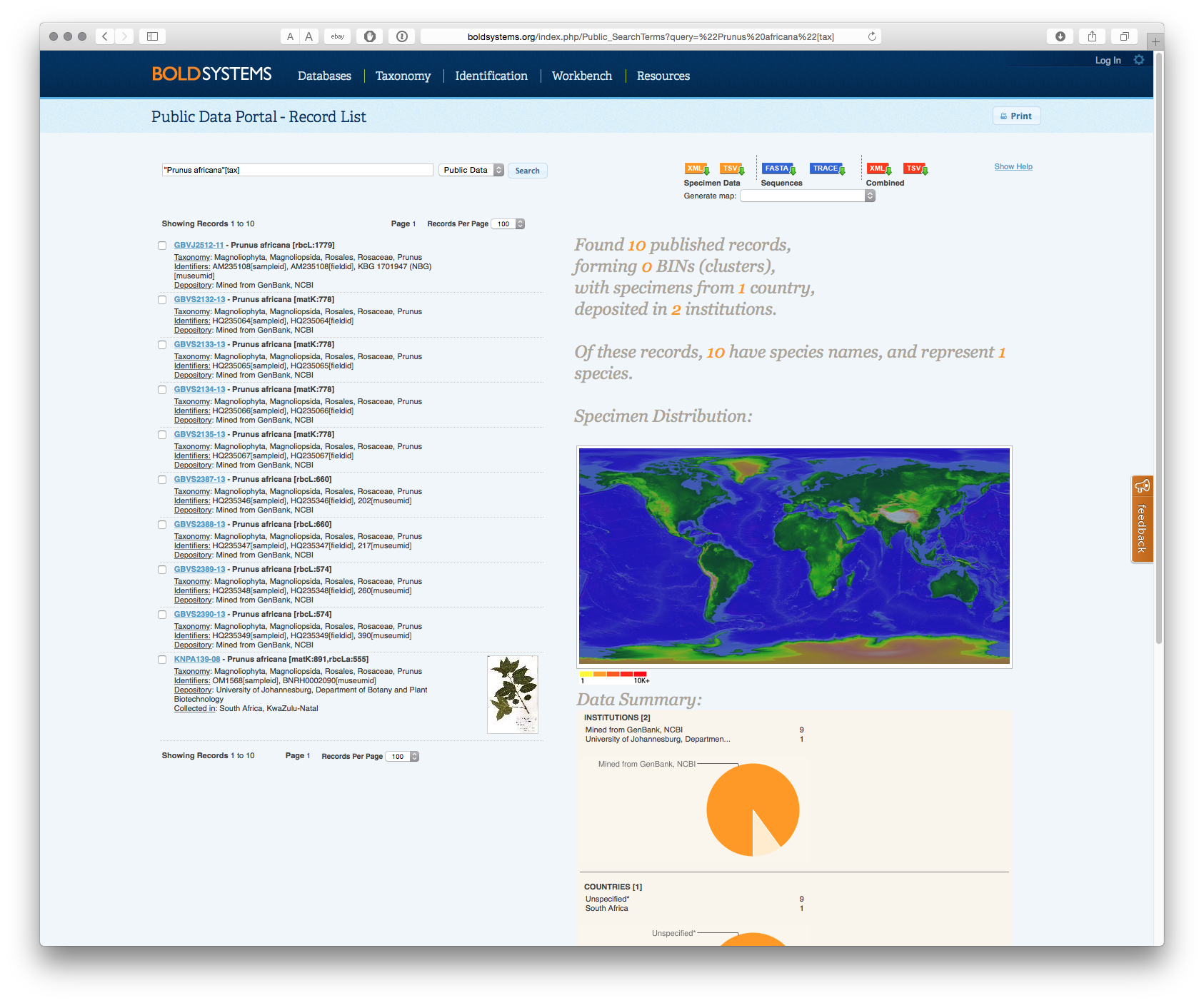
##   
## Retrieving data for taxon 'Prunus africana'

## [1] "191949"  
## attr(,"class")  
## [1] "boldid"  
## attr(,"match")  
## [1] "found"  
## attr(,"uri")  
## [1] "http://boldsystems.org/index.php/Taxbrowser\_Taxonpage?taxid=191949"

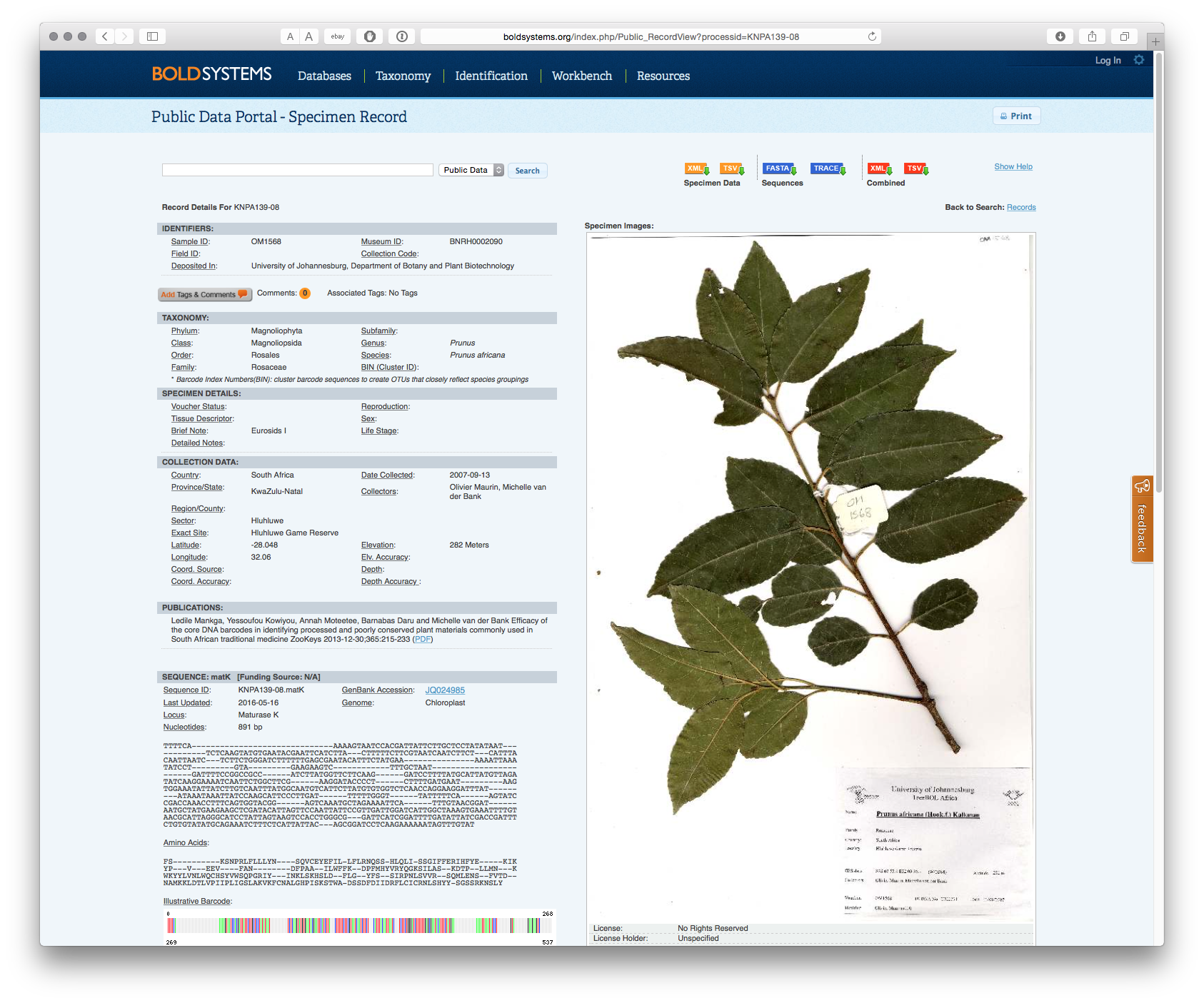
More information from the URL can then be accessed in R or through the BOLD website. On the BOLD website this produces a list of sequence related records. The image below is viewable [here](http://boldsystems.org/index.php/Taxbrowser_Taxonpage?taxid=191949)



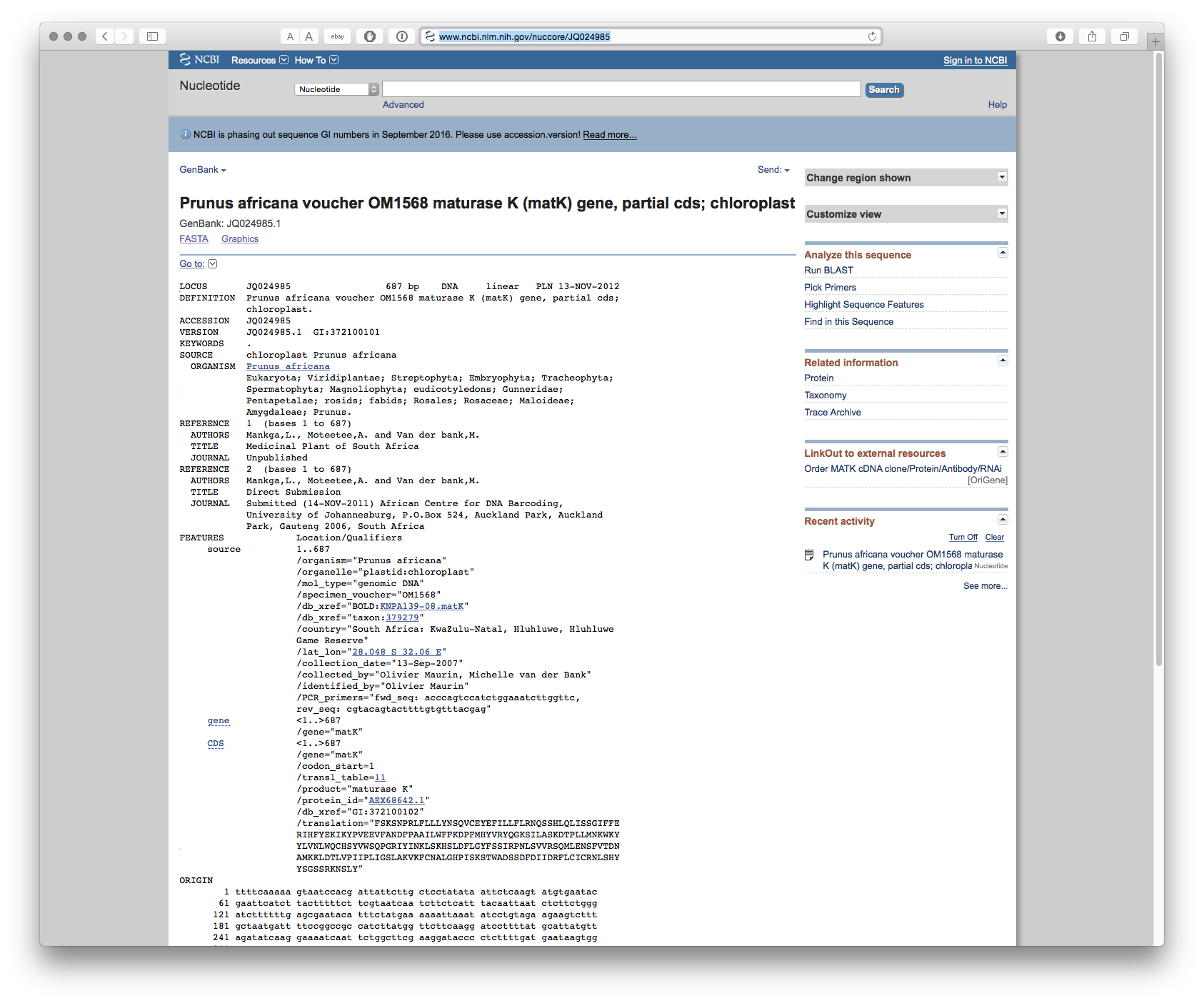
Selecting Access Published & Released Data produces the following list of [records](http://boldsystems.org/index.php/Public_SearchTerms?query=%22Prunus%20africana%22%5Btax%5D).



The final record in this list includes a specimen image and the [following information](http://boldsystems.org/index.php/Public_RecordView?processid=KNPA139-08):



A significant amount of information is contained in this record, including the record number, sample ID, Museum ID, where the specimen is located along with where the material was collected, by whom, along with the sequence listing. A link is also provided to the Sequence ID and [GenBank Accession number](http://www.ncbi.nlm.nih.gov/nuccore/JQ024985) as seen below.



As this demonstrates, it is increasingly possible to rapidly access sequence and associated record information for a particular species or list of species. Given the presence of multiple ID fields it appears reasonable to assume that the simple permit identifier (e.g. BS21051234) in sequence records arising from research could be included in the conditions of the permit and associated MAT. This could readily lead to the creation of an archive of electronic records for biodiversity in a country that contain known sequence data. Uses of such sequences could then become amenable to monitoring using the relevant IDs or sequence searching for identical or similar sequences using BLAST (Basic Local Alignment Search Tool) and associated tools.

### Conclusion

In this section we have discussed the use of unique identifiers and the application of the principle of triple redundancy as part of the design of the online permit and monitoring system. Unique identifiers using standardised country codes, dates and sequences of numbers allow for the construction of an internal permit system that establishes and maintains links between a permit application and associated documents (MAT) and communications. This system already works well for millions of patent documents. The principle of triple redundancy was then applied to the generation of labels containing identifiers that could be used to maintain links between the original permit and samples, publications, patent applications, and sequence data originating from the grant of a permit under an ABS framework. The use of free tools (such as taxize in R) allows this information to be readily retrieved from a range of different data sources. While requiring further elaboration, the use of unique identifiers combined with the principle of triple redundancy provides a route to cost effective monitoring and is recommended.