Thanks, that is very helpful! How would I create a GeoLocation table in the SQL database? Is this similar to how the project's database will be created because I would like to know that as well.

I just tried doing this by creating a Countries table in Access then deleting the text box for "AuthorCountries" and creating a combo box called "AuthorCountries" that references the Countries table in Access. When I try to enter a new record this way, I get this error Graphical user interface, application

Description automatically generated which I think is because Access thinks I am not filling in AuthorCountries since I created the combo box wrong. How should I create the combo box in Design View besides just deleting the text box and adding a combo box with the same label is not working.

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <thomas.balmat@duke.edu>  
**Sent:** Monday, December 14, 2020 6:41 PM  
**To:** Grace Llewellyn <grace.llewellyn@duke.edu>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Oooooh, exciting.  Basically, there are two important parts to a combo box:  the control source and the row source.  The control source is the table column that a combo box is bound to.  Enter data into the field and it is recorded in the current record of the table in the database that the form is bound to.  The row source is what presents available values to the user.  This can be a query (select statement against a table in the access database, a call to a saved query that is relayed to the SQL database, or a validation list.  The easiest style to configure is the validation list – all you have to do is change the row source type to “Value List” and enter a semi-colon delimited list of values in Row Source.  But easy is not always best.  The problem with this approach is that your validation list is part of the form design and anytime the list changes, you have to modify the form.  Perhaps not with this project, but another, where you have the same combo box appearing on 55 different forms (most production systems have this sort of thing).  Are you certain that all 55 combo boxes would always have the same value list?  What if they require 100 elements in the list?  A better approach is to use row source type of “Query” and row source something like “select Country from GeoLocation.”  GeoLocation can be a table in your Access database or, even better, a link to a GeoLocation table in your SQL database.  When a new country is required, it is entered into GeoLocation and, now, all combo boxes that reference it are updated.

Two methods of updating the GeoLocation table are:

1. Create a maintenance form bound to the linked table GeoLocation (this requires users to refresh their form before the new values appear)
2. Create an “On Not in List” event to update GeoLocation with user supplied values as the need arises (this is the ultimate implementation and requires considerably more planning and effort)

I suggest, for starters, creating a GeoLocation (or Country, or whatever) table in the SQL database, entering values (from within Access) directly into the table, and using a row source type of “Query” in your combo boxes.

Another important setting is “Limit to List” (on the Data tab of the combo box property sheet).  I suggest changing this to “yes.”  Otherwise, you will have both “United States” and “Untied States” in your data.  The problem with limit to list is that all of the valid values must exist in your table or list when the user wants to select them.  But, this is where your ideas on system and transaction design must be developed.

Tom

**From:** Grace Llewellyn <grace.llewellyn@duke.edu>  
**Sent:** Monday, December 14, 2020 4:13 PM  
**To:** Tom Balmat <thomas.balmat@duke.edu>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

That works, thank you!

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Monday, December 14, 2020 4:11 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

You should not have to exit the app in order to see new records.  There is a thing called form refresh.  You might need that.  I will respond to the remainder later.

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Monday, December 14, 2020 4:09 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

It works! I didn't realize that I needed to exit out of Access then enter it again in order to get the results to show up. And now I don't get the error of not being able to type directly into the form since the form has already been created and saved!

Also, if I want author countries to be a combo box instead of a text field, do I just change that on the article form directly in the design view? Will this affect how the data is being stored in the database in the backend (I know it will affect the data we have entered already)?

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Monday, December 14, 2020 2:44 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

I see data in records.  Did you key that into the table or into the form?  I am able to enter data using the form.  “Save” retains the record.  Do you receive any errors when you save?

Something I noticed with continuous layout is that you need to adjust the tab order of your fields.  You can do that in form design with the Tab Order sheet:

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Monday, December 14, 2020 12:27 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)



Reply

Forward

**TB**

Tom Balmat

Mon 12/14/2020 6:41 PM











To:

* Grace Llewellyn

Oooooh, exciting.  Basically, there are two important parts to a combo box:  the control source and the row source.  The control source is the table column that a combo box is bound to.  Enter data into the field and it is recorded in the current record of the table in the database that the form is bound to.  The row source is what presents available values to the user.  This can be a query (select statement against a table in the access database, a call to a saved query that is relayed to the SQL database, or a validation list.  The easiest style to configure is the validation list – all you have to do is change the row source type to “Value List” and enter a semi-colon delimited list of values in Row Source.  But easy is not always best.  The problem with this approach is that your validation list is part of the form design and anytime the list changes, you have to modify the form.  Perhaps not with this project, but another, where you have the same combo box appearing on 55 different forms (most production systems have this sort of thing).  Are you certain that all 55 combo boxes would always have the same value list?  What if they require 100 elements in the list?  A better approach is to use row source type of “Query” and row source something like “select Country from GeoLocation.”  GeoLocation can be a table in your Access database or, even better, a link to a GeoLocation table in your SQL database.  When a new country is required, it is entered into GeoLocation and, now, all combo boxes that reference it are updated.

Two methods of updating the GeoLocation table are:

1. Create a maintenance form bound to the linked table GeoLocation (this requires users to refresh their form before the new values appear)
2. Create an “On Not in List” event to update GeoLocation with user supplied values as the need arises (this is the ultimate implementation and requires considerably more planning and effort)

I suggest, for starters, creating a GeoLocation (or Country, or whatever) table in the SQL database, entering values (from within Access) directly into the table, and using a row source type of “Query” in your combo boxes.

Another important setting is “Limit to List” (on the Data tab of the combo box property sheet).  I suggest changing this to “yes.”  Otherwise, you will have both “United States” and “Untied States” in your data.  The problem with limit to list is that all of the valid values must exist in your table or list when the user wants to select them.  But, this is where your ideas on system and transaction design must be developed.

Tom

**From:** Grace Llewellyn <grace.llewellyn@duke.edu>  
**Sent:** Monday, December 14, 2020 4:13 PM  
**To:** Tom Balmat <thomas.balmat@duke.edu>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

That works, thank you!

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Monday, December 14, 2020 4:11 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

You should not have to exit the app in order to see new records.  There is a thing called form refresh.  You might need that.  I will respond to the remainder later.

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Monday, December 14, 2020 4:09 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

It works! I didn't realize that I needed to exit out of Access then enter it again in order to get the results to show up. And now I don't get the error of not being able to type directly into the form since the form has already been created and saved!

Also, if I want author countries to be a combo box instead of a text field, do I just change that on the article form directly in the design view? Will this affect how the data is being stored in the database in the backend (I know it will affect the data we have entered already)?

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Monday, December 14, 2020 2:44 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

I see data in records.  Did you key that into the table or into the form?  I am able to enter data using the form.  “Save” retains the record.  Do you receive any errors when you save?

Something I noticed with continuous layout is that you need to adjust the tab order of your fields.  You can do that in form design with the Tab Order sheet:

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Monday, December 14, 2020 12:27 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)



They were all already enabled. However, when I went to Design View then back to Form View, I could then type in the form (even though I made no changes in design view). However, when I fill out the form and press save, the entries I added are not reflected in the table.

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Monday, December 14, 2020 12:12 AM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

By “key” I mean “manually input.”  So, your SQL authority looks good.  Try this:

Open the form in design view (right click, design).  Display the form properties:

Allow Additions, Allow Deletions, Allow Edits, and Allow Filters should all be set to “Yes.”  If not, then correct them and save the form.  Open in normal view (double-click).  Can you enter data?

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Sunday, December 13, 2020 9:45 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

What do you mean by key data? I can manually input data into the table (not in form)

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Sunday, December 13, 2020 9:40 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

If you open the table (double-click), are you able to key data?

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Sunday, December 13, 2020 8:35 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

This is how I have been trying to do it and it won't let me type in the text boxes of the form

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Sunday, December 13, 2020 6:17 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

In the Object panel, click on the Article table.  Then, in the toolbar, under Create, click the Form icon:

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Sunday, December 13, 2020 5:16 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

How did you create that form so that its data was sent to the table after being saved? Like created the form from the table?

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Sunday, December 13, 2020 4:39 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Sunday, December 13, 2020 4:35 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Where is the Objects panel?

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Sunday, December 13, 2020 4:28 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

The form properties related to inserts, updates, and deletes appear on the “Data” tab of the form property sheet:

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Sunday, December 13, 2020 4:13 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

When you say "open the form," does that mean having the table you want to create a form for open, going to the create tab, then clicking form? I feel as though I am doing this incorrectly and have tried researching it for the last couple days and am still confused (trying to add new records via a form in a database that I created).  When I try it this way and begin typing in any field, the fields are still grayed out and will not let me type.

Also, I am familiar with ACID from my Intro to Database Systems class which will be very helpful!

Thank you this was all very helpful and informative!!  
Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Sunday, December 13, 2020 3:13 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

A few points:

When you open the form, it reports which record is being viewed (record i of n).  at first, this is 1 of 1 and, since this record is empty, we know that there are actually no records in the table.

Also, beside the i of n indicator, there are record navigators and a “new record” indicator (triangle with a \*).

Initially, when viewing a new record, the new record indicator is inactive (gray), but if you begin to type in any field, it becomes active:

This informs that you are inserting a new record into the table.  The table happens to be linked to one in your SQL database, so you are about to create a new SQL record.  But the SQL record is not created until you initiate a save operation from within Access.  This can be done with the “Save” button of the toolbar:

or by moving the “focus” away from your form while editing.  I suggest studying the concept of focus and events that you can intercept during transition of focus.  Notice, also, the pencil in the selector bar beside the record.  That indicates editing.  You can edit an existing record, of course.

To view table properties, use your right-button on the table in the objects panel and select  “Design View”:

You will see the table design but, since the design actually belongs to a SQL table, you cannot modify it from within Access.  That has to be done from within MySQL on the VM.  The design should appear something like:

Your column names and types should be apparent.  The key beside ArticleID indicates that it is a primary key field.  Values in this field cannot be duplicated.  Each column has its own properties and they appear in the column property section when a column is selected:

You will note that I made each column in your table required.  You do this, while creating a table, by declaring “not null” for a given column.  I also created an index on each column.  These will make query execution much more efficient than if columns are not indexed.  SQL where and join clauses use them.

You asked about viewing table properties, but your images were for forms.  To view form properties, use right-button, Design View on the form from within the Objects panel:

In the gray area beside the form, use right-button, Properties:

This will display a property sheet for whatever you click on (click a text box to view that element’s properties, click the gray area outside of the from to view form properties):

Note that the Record Source for the form is Article.  That indicates the Article table in the object viewer.  We know that the Article table is linked to the actual table in the SQL database.  Neat, huh?  The user will open a from and interact with records, never knowing where they actually reside.  But you will.

Click on a form field to view its properties:

In the case of ArticleID, we see that it is named “ArticleID” (you can rename it, but then, within the context of this form, you must reference it by its new name), has a label of “ArticleID” (this only affects the text that appears beside the field on the form, and likely should be modified, at least to “Article ID”), and it is bound to (Control Source) the table column “ArticleID.”  Entering a value into the ArticleID field of this form attempts to place that value into the ArticleID column of the Article table.  When you save a record (explicitly with the “Save” button or by moving the focus from a record being edited (click on a different record or close the form), Access will attempt to apply changes (by composing actual SQL insert/update statements behind the scenes – no mystery there) to the SQL database.  Neat, huh?  A very important aspect of this is that all rules regarding security, data integrity, foreign keys, concurrency, and transactions, as defined within the SQL database, are enforced by the SQL engine.  Access merely acts as an intermediary on behalf of the user – but strictly through your client/server design.  If you are not familiar with the concept of ACID in database execution, I suggest you explore it as you design and implement your app.  ACID is what gives us confidence in data integrity in a dynamic environment.  Very important.

You are making good progress.  Keep it up.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Sunday, December 13, 2020 1:46 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Thank you! That worked! I am now having trouble getting to this part of what you described in your email and being able to add new records via a form. I think it's just a small button I am missing/ can't find. For example, how do I open up the property sheet of a table? I tried googling this and cannot find out how.

Re: Design considerations

Having sub-table for authors- This seems like a good strategy that we will want to implement, but I am just worried about "once an author exists, he/she never has to be recreated." Sometimes an author might have different affiliations at the time they wrote different articles. I do not think that retyping the same author is a concern for us (not that many repeating authors at least in our last project). Also, if we create this sub-table for authors, then does that mean when users are entering information about their article's author(s), they would be able to input information for an unlimited number of authors in a sub-form?

Having fixed data element tables- I believe I already did this for author countries with a combo box that only lets users select countries from that list. Did I do this incorrectly?

Also, that makes sense about not having to do this each time, that's for informing me!

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Sunday, December 13, 2020 12:53 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

The link I supplied to the password file was incorrect (so was the actual password – I see you got to the actual file).  Go to <https://duke.app.box.com/notes/697450437159?s=2ewmau8t62njtnshb09de9xs3dlfrdu1> and use the updated ArticleWriter password in this step:

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Sunday, December 13, 2020 11:02 AM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

Below is what I enter into the table. I get the following error (even after establishing a VPN connection). I tried retyping everything 3 times and changing the cases of Article and ArticleWriter.

Also, will this need to be done every time that a new table is created with data elements?

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Friday, December 11, 2020 2:43 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

I created a new table with your data elements.  To access it, you need to modify your ODBC DSN.  Do the following:

Start, Windows Administrative Tools, 64 bit ODBC

Go to the System DSN tab:

Double-click OEGM:

Change the user name to ArticleWriter, password to the ArticleWriter entry of <https://duke.box.com/s/2ewmau8t62njtnshb09de9xs3dlfrdu1@rticleR3der>, and Database to Article

Click Test.  You should get:

Open your Access app.  Delete the table named “a”

Under External Data, click New Data Source

Under From Other Sources, click ODBC Database

Select “Link to the data source by creating a linked table” and click OK

On the Machine Data Source, double-click OEGM:

Select Article and click OK

Under Tables, you should have a linked table (plus sign and globe) named “Article”

Double click the Article table:

It is empty, of course, but if you specify it as the record source of a form, you can create and update Article records using that form:

Note the \* preceding the initial record in the table view above.  That indicates that records can be inserted, updated, and deleted.  If you ever have a problem creating or modifying data using a form, check to make sure that the star appears in the table view.  If missing, ODBC or SQL permissions are generally the culprit.

You should now be able bind form elements (combo boxes, text fields, etc.) to your linked table elements:

.

Any records created or updated (form, not design, view) are saved in the SQL database, so that one user’s changes are visible to all users.  The SQL engine can handle dozens of simultaneous users, so throughput should not be a problem.

DESIGN CONSIDERATION:

Some of your fields appear designed to contain composite data, such as Authors and AuthorCountries.  This will make searching for individual authors or countries problematic.  I suggest creating separate tables for composite data, so that an each element (author) appears in a single sub-table record.  The sub-tables would be linked to the Article table by a key field and, for instance, author searches would begin in the Authors table then join to corresponding articles.  This approach does require entering authors into their table prior to creating an article attached to them, but once an author exists, he/she never has to be recreated.  Also, having fixed data element tables, as opposed to free form entry fields on the Article form, avoids the ever-present problem of multiple spellings.  Once “United States” is available in the Countries table, it is never typed again.  If there is a misspelling (Untied States), the Country record is modified and all Article records that reference that Country record are corrected.  It is generally a good idea to use this strategy with all data elements with invariant codes (names, here) that may appear on multiple referencing records (Articles).

Although the above statement is good advice, you might consider beginning with your current design, get basic record creation, management, and retrieval (from within R?) working, but implement the reference table method (I would call it a foreign key method) before entering too much actual research data.

Let me know how this goes.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Thursday, December 10, 2020 10:46 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

I create my own Access form for testing and this is the information about it:

Article ID- integer

Title- character (max 100)

Journal- character (max 50)

Year of publication- integer

Author(s)- character (max 100)

Countries of author(s): character (max 100)- combo box

Funding source- character (max 100)

DOI- character (max 50)

Publication type- character (max 50)- combo box

Affiliation(s) of the first author- character (max 100)- combo box

Screenshot of the form I created:

Thank you,

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Wednesday, December 9, 2020 11:48 AM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

I know you are busy with your approaching deadline, so would you be able to tell me how to create the data elements in the test database or give me some references? That way I can play around with this more this week.

Best,

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Tuesday, December 8, 2020 1:49 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

The schema will actually be in the MySQL database on the server.  With that in place, we will connect to it through ODBC.  The ODBC connections are what we wrestled with earlier in the year.  If you will prepare a list of data elements (names and types – integer, float, character along with max positions, binary, etc.) then I will create them in our test database.  That should be enough for you to continue developing a meaningful form.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Tuesday, December 8, 2020 1:16 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

No worries! I do have one question: Do I need to explicitly create the database schema myself for an Access form that has already been created (i.e. STE review). Based on the select SQL statements that we discussed before (shown below) that are a way of viewing entries in an Access form SQL table, I would think that there needs to be something in the backend of Access having the database schema to be able to create those select statements. Can/ how would I access such database schema?

Also, will you be able to meet with me anytime next Monday (12/14)? I am free all day.

Good luck with your approaching deadline,

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Saturday, December 5, 2020 1:22 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hello Grace,

Yes, let’s meet soon.  I am busy at present completing something with a deadline, but will have some time after the upcoming week.  Your plan on connecting an Access form to the database is a good one.  I suggest experimenting with creating forms to insert, update, and delete records in the experimental database we used earlier to verify connections between Access and the server.  Next, if you will begin to configure forms with the text, selection, and control elements you will need (perhaps borrowing from professor Gill’s STE example) then we can, later, include corresponding data elements in the database to bind the new visual elements to.

Send any questions you have as they arise and I will answer as time permits.  Then let’s plan to meet in about one week.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Wednesday, December 2, 2020 8:42 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

I hope all is well with you! I am starting to look toward this project during winter break, and I was wondering if you would be able to meet with me to get started on this project? I am available all day tomorrow and Friday and anytime next week since classes are over. I am planning on starting with connecting the SQL database to the Access form as first steps in this process.

Hope to talk soon,

Grace

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Monday, October 26, 2020 11:31 PM  
**To:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>; Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Looking forward to it.

Tom

**From:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Sent:** Monday, October 26, 2020 7:43 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>; Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

Thanks for the help!!

David

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Monday, October 26, 2020 3:21 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Cc:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Yes, I plan to have a database ready to be interfaced to your Access forms when they are ready.  Please keep me informed of your progress and feel free to ask any questions that may arise.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Monday, October 26, 2020 11:58 AM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Cc:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

I am planning on working on this over winter break as the primary thing I am doing for 3-4 weeks! Will you still be available during this time in case I have any issues I would like to talk about?

Thank you for reaching out,

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science major | Environmental Sciences & Policy minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Monday, October 26, 2020 11:50 AM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Cc:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

I hope you are doing well.  With the end of semester nearing, I thought I would check in and ask if you have a plan for continuing development of the OEGM project over the upcoming break.  As I recall, when we left off, we had configured and tested the SQL database and Access components needed to implement an article cataloging app.  I think the next  steps would be to develop a database and forms for data entry.  Do you have any specific plans?

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Thursday, September 10, 2020 1:58 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

I hope all is well with you! Thank you so much again for helping me with everything with this project! I have decided though that I am going to put this project on hold for a little while I am recruiting for an internship this summer. I am hoping to pick up the project again later in the semester, but more likely I will work on it a lot over winter break. Are there any problems with this?

Thank you again for all your help!

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science Major | Environmental Sciences & Policy Minor | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Friday, August 28, 2020 10:37 AM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

Very good.  The sooner we have an actual project, the sooner our effort will be productive, in respect to supporting research.  Examples are fine to begin with, but developing around a real world scenario can be both experimental and lead to a functioning product for analysis.  See what you can do about getting definition on an initial project.

I hope that this process is stimulating thoughts on the role a support analyst plays in research projects.  By being thorough and truly understanding client objectives, you will go beyond simply providing tools for them to access data.  I encourage you to understand as much as possible regarding the models, interpretation of data elements and their associations, provenance, statistical methods, and potential for applying new techniques for visualization and human interaction in the domain of those you support.  If you try to see your applications through the lens of your user, as opposed to glitzy, whiz-bang architecture, then you will become a partner to them and much treasured, much in demand.  Using this approach, I have become a co-author (or at least credited for support) on papers in a variety of disciplines, including genetics, political science, economics, law, statistics, and rare disease research.  The only one of these that I can claim any (modest) expertise in is statistics.  So, you see, there is a wealth of opportunity awaiting, if you are interested.

Keep up the good work!

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Thursday, August 27, 2020 4:25 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Oh Web of Science is what I thought you meant and I now understand you meant the query to get the potential articles.

That also is what I thought you meant by data elements as well. Since that is the case, we need a specific project first before creating the data elements (except for examples).

And sounds good, I will get started on understanding Access more and trying to submit a form/ make my own Access form.

And the rest makes sense thank you!

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Wednesday, August 26, 2020 5:51 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace.  WOS = Web of Science, that Professor Gill mentioned.  I believe that is where the articles to be reviewed are available.  I picture a user reviewing articles in one window and keying in key identifiers and text into the Access app, which is connected to the SQL DB (through ODBC – that’s why we had to nail it down).

Data elements are the actual data fields that we will store.  Whatever fields are needed to fully describe articles must be present in our SQL tables, so that they will answer all possible research questions.  We will also have corresponding fields on various Access forms so that the database fields can be populated.  Later, we develop queries to be executed in R to retrieve sets of observations so that stats, analyses, model fitting, and map generation can be accomplished.

I will create the initial version of the SQL database and we are able to connect to the server through ODBC.  Next we need a form with fields corresponding to table elements (fields) in the SQL DB.  How about if you start on that.  I suggest using Professor Gill’s example Access form.  With these items, we can begin to experiment with data entry and modification.  That would be a good time, also, to begin to ask in-depth questions on research objectives so that we capture all necessary data from WOS and other sources.  There also has been mention of the possibility of executing queries against the WOS data server.  We can explore that once we have a functioning manual data entry system.

R and Shiny interfaces can be developed after the data maintenance system is complete.  Doing both simultaneously, I think, would distract one from one another and result in inefficiency and possible poor design.  This is why we must do a good job up front in specifying all data elements necessary for analysis.  If done well, then developing analysis scripts will go smoothly, at least from a data sourcing point of view.

Does this make sense?

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Wednesday, August 26, 2020 5:32 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Sorry I am getting back to you late! I have included my answers to your questions below (and walked through the steps of what you suggested on Access). Also, I am willing to mess around in Access and get to know it better, but where are we going with next steps of getting Access to connect with SQL (am I right is this what we want to do)? I am just not sure where to be sending the information once a form is submitted.

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Friday, August 21, 2020 2:20 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Yes, David’s template is a good place to start, since (I suspect) it is in the neighborhood of what you need.  Fortunately, it does not contain too many objects.  We will replace the internal tables with ODBC connected versions to be created in the MySQL database (that is why ODBC was the first thing for us to tackle).  To simplify things for you, I will create the MySQL tables and an Access app with ODBC links to them.  I suggest that you study the forms in David’s template.  Fortunately, there are only two, and one is a sub-form (Outcome), which appears in a window on the Extraction Questionnaire form.  To review the design of a form, right click on it in the Objects window:

Then right-button, click “Form Properties”

Notice the record source.  It is a select statement with the from clause:

FROM [Data extraction questionnaire] LEFT JOIN Outcomes ON [Data extraction questionnaire].[01 Article ID] = Outcomes.[01 Article ID];

So, you might want to brush up on from clauses.  The columns returned in the row set resulting from this select statement are specified as control sources for gui elements on the form (text boxes, combo boxes, etc.).  Right-click on a control and click “Properties” (I use Publication type, here)

Publication type’s control source is “11 Publication type” so that when the Extraction questionnaire form is displaying contents of a given record, the value of 11 Publication type for that record appears in the Publication type field.  It can also be modified while displayed.

So, the Extraction questionnaire form displays individual records taken from its record source (the select statement).  To view all records in that source, you can execute a query.  Click the ellipses (…) on the form’s data source property:

then right-click, SQL View:

This displays the entire select statement:

To execute the query, click Run!

The 269 records returned are what source the records in form view.

To return to SQL editor mode, right-button on the “Extraction questionnaire” query tab then click “SQL View.”

Note that you can modify the select statement in the query review window.  Go ahead and experiment with your SQL skills.  You will be prompted whether or not you want to save changes, so be sure of results before answering “yes.”  I suggest making an archive copy of any app prior to making changes.

Back on the form, note the “Outcomes” object:

It is a subform containing the independent object “Outcome form” which is the second form in the app.  Generally, a subform displays a subset of records from its source that is related to some key field from the main form’s current record.  The subset filter is applied using the “Link Master Fields” and “Link Child Fields” properties, which are empty in our case, so that, effectively, no filter is applied.  This explains why all outcome records are displayed for all extraction records (review the record for article 67201 and notice that the outcome for article 68040 is displayed in the outcomes subform).  I suspect that the outcome records should be filtered to the article being displayed and, if that is correct, we will implement filtering.

Important next steps, I believe, are to answer questions resulting from your project objectives:

*After selecting an article for inclusion, the team members extract relevant data from the article such as information about the author, the restoration, and the outcome of the project.*

Does “selecting an article” imply WOS query or review?  If so, which data elements available there will be used, or which data elements will be generated by the reviewer as a result of using WOS?  What are their types, lengths, and validation lists?  What is your definition of “Outcome” in a data element framework?

​What do you mean by WOS? I am not really sure what you mean by these questions, but I think it is possible that we will need to have an actual project with what to extract from an article outlined to answer them. Do you think that is right?

*reduce the likelihood of errors*

Specifically, which errors must we check for?  Validation lists for combo boxes are useful in elimination typos and enforcing consistent article data descriptors.

​The errors really depend on what information is being extracted from an article. Validation lists are a good start as is changing the user's choices after they type in certain information. For example, if the user chooses to reject an article, stopping them from entering more information about the article would be ideal. I also think a way to save the user's article extraction information would be helpful for them as they are typing in information... not sure if this is already a setting in Access

*After the data is submitted and validated …*

What validation rules are to be implemented?

​This will really depend on the project, but there shouldn't be that many validation rules and the ones there are would be simple checks of drop down options most likely.

*For example, research project A could use information about the restoration while research project B uses information about the outcome.*

What are the various types of projects that are to be supported?  What are the data elements associated with each?

​Really just any project that has to do with data extraction should be supported. What do you mean by data elements?

*how to implement the backend, maybe with R and Shiny*

This, generally, comes after the database and data entry components are in place.  R and Shiny will query the database using (guess what?) ODBC.  However, we want to explore the various analyses to be implemented while the database is being constructed so that all known, necessary data elements are identified and included.  Otherwise, we will have to repeat the cycle of database development, Access query and form redesign, and R analysis, which is inefficient and lacking in results directed design.  We want to consider your project from the direction of desired results toward necessary data components.  Doing so improves your chances of arriving at stated project goals.

Yes, I agree I want to deal with analyzing the data and R and Shiny after we have the data extraction complete and keep them in mind going through the project so I am able to add them later. Is the project being set up now so that they can be added later? Or have we not gotten that far yet?

*modeling and creating gap maps*

These sound like your stated objectives, so put some thought into these and discuss them with Professor Gill to stimulate new ideas.  You might be surprised at how a few well focused questions from you generates new paths of inquiry in the minds of those you support.

Well, that was a real eye-full.  I think you are correct, in that you must focus on the data collection component first.  Read all you can about SQL and Access form development and experiment with example queries and forms.  As you develop new designs, always consider how your users will accomplish their goals with them.  In the case of the analysis components, always ask yourself what new knowledge the analyst will develop by using your tool.  In an academic setting, the most important result is to generate new knowledge – to take data and present a new finding.

Good luck.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Friday, August 21, 2020 12:25 AM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Professor Gill did share his copy of an Access project with me [here](https://www.dropbox.com/s/cy1qwvnvbdopqfl/STE%20review.accdb?dl=0)! I am not sure what you are suggesting for next steps- is this something you would be willing/ we need to set up a meeting to discuss? I am not really sure how to use Access but can look into it. Can we use David's template for now?

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Thursday, August 20, 2020 5:07 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

OK.  We are on our way.  I think that the next step is to design table structures, basically the columns, or variables, that will be stored.  Did Professor Gill share a copy of his Access DB with you?  That might be a good starting point.  An alternative would be to review the analysis scripts and create a list of variables that they use.  We need to consider what other useful info might be available on WOS also.  Our list of variables should include names (pay attention to case), type (alpha, real number, integer, logical), and maximum length required.  We also need to organize variables into related groups.  The groups will become tables to contain columns to represent identified variables.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Thursday, August 20, 2020 4:49 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

I am 64 bit and I just figured out how to get 64 bit Office on my laptop! Clicking table a works now yayy!!

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy | Energy & Environment certificate

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Thursday, August 20, 2020 1:03 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

OK.  First, let’s verify that your OS is 64-bit.  Go to Start, Windows, System, Control Panel, System.  Does System Type report “64-bit Operating System?”  When you downloaded and installed Office, were there any options for choosing 32-bit or 64-bit versions?  According to this article, <https://www.cnet.com/how-to/how-to-install-the-64-bit-version-of-office-365/>, 32-bit is the default version and you must select the 64-bit installer.  The example it gives downloads Office from a Microsoft site, but the Duke site may have similar behavior.  It’s been a while since I installed Office on my system, but I do recall selecting 64-bit.  Try deleting Office then reinstall, making sure you select the 64-bit version.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Wednesday, August 19, 2020 11:36 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

It does say the 32 bit version so I worked with OIT trying to get the 64 bit version and they told me that redownloading Office 365 should automatically give me the 64 bit version. I deleted office from my computer, restarted my computer then redownloaded Office and Access was still the 32 bit version

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Tuesday, August 18, 2020 6:41 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

I created a VM at vcm.duke.edu and selected the Office 2016 configuration.  On it, Access and ODBC behave as they do on your system.  After installing the MySQL 64-bit ODBC driver, I noticed that, while creating an ODBC connection from within Access, no drivers appeared for selection.  I then installed the 32-bit ODBC driver and they appeared.  Next, I checked the version of Access that is installed on the VM.  It is 32-bit.  Let’s check yours.  From within Access, go to “Account,” “About Access” and read the red text just beneath “About Microsoft Access.”  Does it say “blah, blah, … 32-bit?”  If so, then re-install Office from the Duke software site, but specify the 64-bit version and try opening the “a” table in our test app.  I think that it will work.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Monday, August 17, 2020 6:33 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Oh, sorry I didn't realize that! I also realize now that while I was getting to that error page, Office was still being added to my cart so I was able to checkout and get office again. I restarted my computer and Access is up to date on my computer now, but I am still getting the same error

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Monday, August 17, 2020 8:47 AM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

I suggest that you contact OIT assistance regarding the “but when I tried to add it to my cart I get a "**software.duke.edu** is currently unable to handle this request" error which is the same error I was getting last night when trying to download office again" problem you are having.  They certainly have responsibility for that.  Do not mention Access or ODBC.  Just get MS Office installed.  Then we can assess whether or not it affects the ODBC problem.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Sunday, August 16, 2020 7:50 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

I just called OIT and they said they couldn't help me so I am not entirely sure what to do. They also said they weren't sure who I should ask instead since they don't really use Microsoft Access

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Saturday, August 15, 2020 12:21 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

You might try submitting an OIT assist request.  Be sure to let them know that you are specifically having a problem with ODBC from within Access, while DSN creation and connections from within R function properly.

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Saturday, August 15, 2020 11:37 AM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Mine says “Microsoft 365 Apps for enterprise" as well. I tried getting office from this [Duke webpage](https://oit.duke.edu/help/articles/kb0024448), but when I tried to add it to my cart I get a "**software.duke.edu** is currently unable to handle this request" error which is the same error I was getting last night when trying to download office again

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Saturday, August 15, 2020 10:48 AM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

In the Access Account section, under “Product Information” my system reports “Microsoft 365 Apps for enterprise.”  What does yours report?  Yes, you might consider reinstalling Office.  Be sure to select 64 bit enterprise.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Friday, August 14, 2020 8:23 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

I ran the Rscript but still get this error when I try to open the Access app you sent me before:

I ran the Rscript again and it loads the content in table a so I'm pretty sure there is a connection. I enabled macros then closed Access then opened it, checked again that macros were still enabled, and tried opening a and got the same error message as above (I also restarted my computer and did it again but it is still not connecting).

I am running version 2007 of Access, should I try upgrading?

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Thursday, August 13, 2020 8:33 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

This means that your MySQL ODBC driver is working and that your computer makes a successful connection to the database.  So, why doesn’t Access access the DB.  Perhaps we should try Microsoft NoAccess.  Run the R script once and make sure you retrieve records, then open the Access app I sent you earlier and try to open the “a” table.  If you are unable to retrieve records then run the R script again to verify that you have a connection.  If that works then return to Access and click File, Options, Trust Center, Trust Center Settings, Macro Settings, and make sure that “Enable all macros” is selected.  If you changes the Macro setting then close Access, restart the app, and try to open the “a” table.

Tell me what happens.

Also, tell me which version of Access you are using (File, Account, under Product Info).

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Tuesday, August 11, 2020 7:25 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

That worked! This is my output:

What are the next steps?

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Monday, August 10, 2020 1:34 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Instead of upgrading, I have always simply installed the latest version.  You will have multiple versions installed (multiple directories under your Documents\R directory), but it doesn’t seem to be a problem.  You do have to reinstall packages, but that, to me, takes less time than wrestling with an upgrade.  In your notes, below, I see only warning messages.  What happens when you execute library(RODBC)?

As an alternative to upgrading R, you can copy the RODBC folder of the attached zip file to c:\users\gllew\R\win-library\3.5\.  Replace the one there, if it exists.  This folder contains RODBC package contents that is compatible with R 3.5.  I don’t think that there are any binaries that require compilation, so it should function.  Further, it was installed on a Windows 10 machine, so any binaries should be compatible with yours.  It might interest you to know that install.packages() really just copies files to you user\R directory.  The library() function searches that directory for associated R scripts and binaries.  Binaries resulting from compiled languages (C, FORTRAN) generally must be compiled on your computer, or have been compiled on a compatible device.

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Sunday, August 9, 2020 12:38 AM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

I actually needed to retype the ". Now I am getting this error:

I am trying to update my R but it is proving more difficult than I thought. I tried using the installr approach, but I get this error:

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Saturday, August 8, 2020 9:17 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Did you change your password when you used ssh?  If so, then that is the password you should use.  If not then ssh to the VM and change your password then try the RStudio URL again.

It appears that ODBC is functioning, but not from within Access.  Let’s try it from within R.  Open RStudio on your computer and execute the following instructions (execute one at a time):

install.packages(“RODBC”)

library(RODBC)

db <- odbcDriverConnect(connection="driver={MySQL ODBC 8.0 ANSI Driver}; server=oegm.rc.duke.edu; port=3306; database=article; uid=articlereader; pwd=articleread;")

sqlQuery(db, "show tables")

sqlQuery(db, "select \* from a")

Are you able to connect to the DB?  The final query should return rows of text.

What do you see?

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Saturday, August 8, 2020 12:04 AM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

Sorry, I should have caught that! Now, when I put oegm.rc.duke.edu:8787 in my search bar and use the log in from the box account I just get an invalid login error.

Also, this is what my drivers tab looks like so I do have MySQL ODBC 8.0 ANSI driver

Also, when I click on OGEM in the user DSN tab it looks like your picture and pressing test yields the connection successful message as well.

I tried your suggested approach at the bottom of your last email and double clicking the "a" table gives me this error:

I also see this information later:

But I am never prompted to enter a user name and password

Thanks,

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Friday, August 7, 2020 4:28 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

RStudio:  I misspelled duke.  Try oegm.rc.duke.edu:8787

ODBC:  Go to the Start menu, Windows Administrative Tools, ODBC Data Sources (64 bit), then display the Drivers tab.  Mine looks like this:

Do you have MySQL ODBC 8.0 ANSI Driver?  That is the driver I am using.  On the User DSN tab, double-click OEGM.  You should see something like:

What happens when you click “Test?”  I get:

Let’s also try this:

Open the attached Access app.  Double-click the “a” table.  Do records appear?  This linked table has the ODBC connection information saved in its definition, which is not as secure as a system created DSN, because anyone who views the database can view the connection string, which includes user name and password fields.  To view the connection string, go to External Data, Linked Table Manager, check ODBC, then click Edit.  The connection string contains everything needed to make a connection to the database.  If the user name (uid) and password (pwd) are omitted then the user is prompted for them the first time a linked file is accessed.  However, with my computer, they are not accepted and I am continually prompted without ever gaining access to a file.  If the link works for you then we can use this method temporarily until we devise a more secure solution.  Further, if you plan to distribute an Access database tailored to a user’s need (read, write, etc.) then saving uid and pwd on ODBC connection strings may not be a security problem, as long as users do not share Access files.

Let me know what happens with the above tests.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Thursday, August 6, 2020 11:40 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

Should I just put "oegm.rc.deuk.edu:8787" into my browser without launching anything first? Because that didn't work- I got a "This site can't be reached" error

Also, I am the only one using my computer and the user dsn showed up, but after I clicked on it I got this error:

I also tried adding the system DSN again after deleting it but that didn't work either

Thanks,

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Thursday, August 6, 2020 6:36 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

For RStudio, user your web browser to target oegm.rc.deuk.edu:8787.  RStudio Server listens on port 8787, so that this URL creates an R session on the server for you.  From these sessions, we will create database connections.  Let me know when you successfully launch R.  I will forward a script to execute a simple query.

For the ODBC connection, try this.  Delete the existing OEGM system DSN then recreate it on the User DSN tab.  System DSNs are available to any user on your machine, but a user DSN will also work, since you are likely the only person using your computer, correct?

Let me know how it goes.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Wednesday, August 5, 2020 11:53 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

When I try ssh [gal16@oegm.rc.duke.edu](mailto:gal16@oegm.rc.duke.edu) in my terminal to check the Rstudio connection, I get the connection timed out error again- is it possible the VM is down?

I also installed MySQL (which I thought I did before because it had been downloaded to my laptop already) and now I could find the MySQL ODBC ANSI Driver. I worked through the steps and then got stuck when I needed to add OGEM DSN. The DSN shows up in the image below and I tried turning off my laptop and turning it back on but other than that I am not sure how to address this problem. My only options in Access are in the second image below

Thanks,

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Wednesday, August 5, 2020 11:40 AM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Good.  We are making progress.  Try downloading and installing the MySQL driver from [https://dev.mysql.com/downloads/connector/odbc/.](https://dev.mysql.com/downloads/connector/odbc/)  Choose

After clicking “Download,” select

then execute the mysql-connector-odbc.msi file that is downloaded.  Return to ODBC connections in Windows Administrative Tools and see if MySQL is now a driver option.  If not, then schedule a Zoom meeting and we can look at it.  I have Visual Studio, SQL Server, MySQL, RTools, and other system software installed on my laptop, which is most likely how the MySQL driver was installed.  Have you made an RStudio connection (oegm.rc.duke.edu:8787)?

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Tuesday, August 4, 2020 11:58 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

I attempted the instructions in the first link I sent, but I'm not sure what I am doing and I still can't get the MySQL ODBC ANSI Driver to show up when I press add. Would you be able to meet sometime this week to figure it out?

Also, I made a VPN connection and made the ssh connection into the VM!

Thanks,

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Tuesday, August 4, 2020 6:01 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Yes, to both.

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Tuesday, August 4, 2020 12:47 AM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

I was attempting to follow your directions from your first email, but when I click add on the Systems tab, "MySQL ODBC ANSI Driver" isn't an option for the driver. Do you think following the directions on this link would solve this problem <https://dev.mysql.com/doc/connector-odbc/en/connector-odbc-installation-binary-windows.html>

Also, in your most previous email you say to make a VPN connection first, should I follow these instructions: <https://oit.duke.edu/help/articles/kb0016403>

|  |  |
| --- | --- |
| [Graphical user interface  Description automatically generated with low confidence](https://oit.duke.edu/help/articles/kb0016403) | [VPN: University VPN Install and Connection Guide for Windows | Duke University OIT](https://oit.duke.edu/help/articles/kb0016403)  Contact Us: 334 Blackwell Street Suite 1100 Durham, NC 27701 Internal: Duke Box 104100 Phone: (919) 684-2200  oit.duke.edu |

Sorry, I am very new to this and would appreciate the help!  
Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Monday, August 3, 2020 2:19 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

I am now able to make ssh connections to the VM.  Give it a try (ssh [gal16@oegm.rc.duke.edu](mailto:gal16@oegm.rc.duke.edu) or use putty).  Be sure to make a VPN connection first.  Let me know how it goes.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Saturday, August 1, 2020 11:10 AM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

I am using a Windows device. Did you want me to put "oegm.rc.duke.edu:8787" in my url bar? It asks for a login to R studio, but my netid and the password in the Box doc don't work

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Saturday, August 1, 2020 10:38 AM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>; David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

For some reason ssh is not accepting connections.  Other services (RStudio Server and MySQL) are.  I restarted the VM, but the problem persists, so I have sent a help request to the administrators.  I will update you when they respond.  In the meanwhile, try connecting to RStudio Server, oegm.rc.duke.edu:8787.

Are you using a Mac or Windows device?  Our next test will be to connect to the database and there is a difference in how this is done on different platforms.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Friday, July 31, 2020 10:41 AM  
**To:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>; Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

I get the same "ssh: connect to host rapid-1316.vm.duke.edu port 22: Connection timed out" error when I try doing "ssh [gal16@rapid-1316.vm.duke.edu](mailto:gal16@rapid-1316.vm.duke.edu)" in my terminal

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Sent:** Thursday, July 30, 2020 9:16 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>; Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Thanks Tom,

Ocean Evidence Gap Map  
<https://bassconnections.duke.edu/project-teams/ocean-evidence-gap-map-2020-2021>

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Thursday, July 30, 2020 8:24 PM  
**To:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>; Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi David,

I will check out the Access example and Colandr.  Grace, please do the same then we can plan on testing these on the new VM.  I should have stated, earlier, that Research Computing aliases end in rc.duke.edu, so the alias for our VM is oegm.rc.duke.edu.  BTW, what does OEGM stand for?

Tom

**From:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Sent:** Thursday, July 30, 2020 2:12 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>; Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

Thanks so much for this effort!

* I have an example Access review [here](https://www.dropbox.com/s/cy1qwvnvbdopqfl/STE%20review.accdb?dl=0) that you can use for testing. It has two simple tables connected by a primary key.

* The SQL database I’m referring to is called [Colandr](https://www.colandrapp.com/signin" \t "_blank). Users download citation records from Web Of Science (for example) and Colandr uses machine learning to help reviewers sort and screen articles that they want. I can loop in Dr. Sam Cheng who manages Colandr if you have any questions.

* A good alias for the VM could be oegm.rapid.duke.edu (Ocean Evidence Gap Map)?

David

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Wednesday, July 29, 2020 6:17 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>; David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

I have a VM configured with R, RStudio Server, and MySQL.  I have tested database connections from within Access and R on a client and everything seems fine.  A few comments:

* I created two DB accounts:  one for reading data and one for creating records.  The only security at the moment are the user names and passwords on these accounts.  With a name and password, anyone can access the database, even from outside of Duke’s VPN.  Is this acceptable?  We can strengthen passwords and also create individual accounts to be distributed to users for access accounting.  We can place the VM behind the VPN, if necessary, but that will delay things a bit.
* I am able to modify records in the server DB from my client Access session.  This means that simultaneous updates are available to us.  Do you have a database schema (basically, record layouts) and a functioning Access app that we can use to begin more thorough testing?  Placing these in a new directory on David’s git repo would be helpful.
* David mentioned the possibility of querying the World of Science (did I get that correct?) directly to viewed by users while cataloging articles.  If SQL queries can be relayed to a WOS server then we can prepare them within Access (using Visual Basic), send them for execution, then display the results in a separate query result window for review.  The returned records can be saved in a table within the MySQL DB for later use, if needed.  One approach may be to:
  + compose a select statement from user supplied query parameter values
  + deliver query to WOS for execution
  + save return set in MySQL DB with a unique session ID
  + display returned records for the current session in a separate window
  + create a query form in Access to retrieve records from past WOS record sets retrieved

For starters, let’s make certain that you can access the VM.  Try to ssh rapid-1316.vm.duke.edu.  Use your Duke netID (gal16) and the password in the Box file I sent you a link to.  You will have to change your password  once you log in.  Next, try something in R using RStudio Server (this behaves like RStudio Cloud, but is running on your VM, and behaves just like any other RStudio session) by targeting [http://rapid-1316.vm.duke.edu:8787](http://rapid-1316.vm.duke.edu:8787/) from your browser.  Run the following R commands:

db <- dbConnect(MySQL(), host="127.0.0.1", dbname="article", user="tjb48", password="Art1cle")

dbGetQuery(db, "show tables")

dbGetQuery(db, "select \* from a")

dbDisconnect(db)

The second parameter is a SQL statement to be executed by the server specified in the dbConnect function, which happens to be our MySQL server.

Once we are successful with the above then we can have you test an ODBC connection from within Access.

Let me know how this goes.

Tom

**From:** Tom Balmat  
**Sent:** Wednesday, July 29, 2020 2:11 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>; David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

I used the official pricing schedule to estimate cost, so $210.00 should be correct.  What would you like to use as an alias for the VM?  The alias is used to access the VM using ssh and is the URL for web traffic.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Sunday, July 26, 2020 10:44 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>; David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Thanks Tom! Also, the fund code for my project is 2910345

The VM is $210 right?

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Sunday, July 26, 2020 3:17 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>; David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

I should have a VM up and running sometime this week.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Sunday, July 26, 2020 2:43 PM  
**To:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>; Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

I have the fund code already!

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Sent:** Saturday, July 25, 2020 10:18 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>; Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

Sounds good. I believe we will need the fund code. If we have that then we should be good.

David

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Saturday, July 25, 2020 9:49 PM  
**To:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>; Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Sounds great David, we can do that! Also, Laura Howes just approved my budget change so Tom I can work with you on Monday or early next week to purchase the VM if that works for you!

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Sent:** Saturday, July 25, 2020 8:21 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>; Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

WE can use my Github a/c or the Gitlab one, whichever works best for you. We may have the occasional non-Duke collaborator, but very few folks will likely be messing around with the code.

David

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Thursday, July 23, 2020 11:15 PM  
**To:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>; Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi,

Thanks again for meeting with us today, Tom! I can set up a git repo on Duke's gitlab site- David is it okay that everyone will need Duke credentials to use it?

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Sent:** Thursday, July 23, 2020 4:58 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>; Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

Thanks for our conversation today. Here are my responses below:

* One more item.  Do you have a git repo configured for the project?  If not then we should create one, either on Duke’s gitlab site or the public github.  I think that using Duke’s gitlab requires Duke credentials, so that access would be restricted.  If you plan to share resources with people outside of Duke then github makes sense.

* + <https://github.com/davidgillmarine/oegm.git>, which only has analysis scripts right now.
* Are there any access restrictions for the data that we will be storing?  I suspect that there are no privacy considerations to be addressed, correct?  Is there any need to restrict certain users (people entering data or those retrieving it) from certain data, say project X users from project Y data?
  + Not that we know of (other than being research data and not wanting to get scooped!). For safety reasons we may not want to have students hopping across projects
  + I think we like the option with having a mega citation database which can be accessed project by project (e.g. with a Project ID filed). Turns out that Colandr is built on an SQL database and we could technically query directly from their software. Would be good to chat with the data manager.

David

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Thursday, July 23, 2020 12:13 PM  
**To:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>; Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

One more item.  Do you have a git repo configured for the project?  If not then we should create one, either on Duke’s gitlab site or the public github.  I think that using Duke’s gitlab requires Duke credentials, so that access would be restricted.  If you plan to share resources with people outside of Duke then github makes sense.

Tom

**From:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Sent:** Thursday, July 23, 2020 11:08 AM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>; Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

Thanks so much for your advice on getting us set up with a robust workflow for this project. I look forward to working with you to get this set up!

Cheers

David

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Saturday, July 18, 2020 10:04 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>; David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi David and Tom,

10am on Thursday 7/23 sounds great! Here is the zoom information:

<https://duke.zoom.us/j/92446242978>

Meeting ID: 924 4624 2978

See you Thursday and thank you both for meeting with me!

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Thursday, July 16, 2020 5:31 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Cc:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

Next Thursday is fine.  How about 10:00?  Do you want to schedule a Zoom meeting?

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Thursday, July 16, 2020 11:30 AM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Cc:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

Thank you for all this helpful information again! David and I are interested in purchasing the VM from Research Computing, but we would like to meet first sometime next week. Would you be able to meet anytime between 9-11am next Thursday (7/23)?

If this doesn't work for you we can definitely find other times!

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Wednesday, July 15, 2020 2:26 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Cc:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

A few additional points:

You initially indicated interest in R and Shiny and, since Shiny Server executes on Linux only, I have assumed that your VM would ultimately be Linux.  I wanted to mention this after reading David’s comments on Windows and Linux, below.  My assumed specification of Linux is not arbitrary or necessarily a preference – it is more a requirement when using Shiny Server.  However, I do think that Shiny Server is a great platform for delivering results of R analyses over the web.

David also asked about the Access to database server arrangement.  Essentially, with this approach, all of the data are hosted and queries executed on the server and you can configure as many databases as desired (given storage constraints).  Access executes on your personal device and connects to the server databases using an ODBC (open database connector) connection.  There are several important benefits to using the client/server method (Access on personal device with SQL database on Linux, in our case):

·       The database server (containing your most precious resource - data!) is maintained within OITs domain, giving you a high level of reliability, backups, and support

·       SQL databases offer an independent level of data update and access security (for reproducibility of results, this is important)

·       A server based SQL database supports simultaneous users and is scalable to support any conceivable number of concurrent processes (more users, bigger VM)

·       For data integrity, sophisticated transactions can be programmed, so that if any part of an update fails, no data are updated (you never have partially accurate data)

·       Query performance is generally far better than that of corresponding operations in R (using which(), match(), or dplyr like functions) and Access

·       Queries saved within databases can be executed from a wide variety of client applications (R, Access, Excel, SAS, Stata, ArcGIS, so on and so on), requiring data collection, preparation, and aggregation instructions to be designed once only then used by all inquiring clients (instead of redundant, synonymous instructions within each client)

·       Computational resources can be distributed, for scalability, between independent, but connected, database, web, and analysis servers

I hope that this information is useful.  Let me know if you have any questions.

Tom

**From:** Tom Balmat  
**Sent:** Monday, July 13, 2020 2:03 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

Using the OIT cost schedule (<https://oit.duke.edu/help/articles/kb0025194>), with a specification of four cores (VMs), 10g of mem, and 50g of storage (bronze), I estimate an annual cost of $210.00.  This is a minimal configuration.  I would recommend adding another 10g of mem, which brings annual cost to about $360.00.  As for support, similar to other research projects, Research Computing will want to see your project through to completion.  We don’t have policies, per say, regarding numbers of meetings or consultations, but are a resource available to you to make sure your project objectives are met.  As long as you are investing in your work, we will also.  The configuration above (20g version) should support a few concurrent web visitors along with one or two concurrent R users.  This assumes modest compute demand, no thousand-level fixed effects problems, neural nets, or that sort of thing (if you do have large problems, then great!, but we then need to equip you for that).  As for using your VM for R and other apps while serving data requests, unless you expect dozens of concurrent web visitors, then it makes sense to have your data server also serve web requests and handle interactive R sessions.  With a large number of web and database requests (hundreds), you want individual servers dedicated to database, web, and interactive sessions.  But, you are likely not there yet, correct?

Let me know if you have further questions.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Sunday, July 12, 2020 2:52 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Fw: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

I reached out to my project PI about your last email and he had the following questions below.

Also, I had one question:  If I pay for the VM option, does this mean unlimited help from OIT for the year that I am paying for (2020-21) or is there a limit to the number of meetings?

Thanks,

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Sent:** Friday, July 10, 2020 9:54 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

Thanks for the update and sharing the conversation with Tom. Really interesting approach! Faculty have a provision for VM resources from Resource Computing ( 4 cores, 40gb Ram and ~100 gb). I use mine for long R processes right now, but it is a windows machine. I’ve never used a linux machine, but happy to explore the possibility of using that allocation to at least test this out.  I would have a few questions:

1. Can the VM be used for this database as well as other side jobs (e.g. running R processes on the side requiring quite a bit of RAM and 2 cores). If not,
2. What is the cost of a VM with the resources Tom suggested?

I would like to understand the Access -> server set up a bit more, particularly to see if we could set up the server on the back end that could support multiple Access databases (for future projects).

Exciting!

David

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Thursday, July 9, 2020 3:49 PM  
**To:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Subject:** Fw: Bass Project Getting Started Guidance (Research Computing)

Hi David,

I hope all is well! Below is the conversation Tom and I have been having about starting my research project. It appears I may have to pay for a VM, which I would be able to do with part of my budget for this year, but I am worried about how future years will use my project. We can also use the free option for a VM, but ads will be on it and I won't be able to receive help creating it. I think the benefits of paying for a VM for this year with Duke's Research Computing would be very helpful to me as I will likely need their services to help me with the project. Maybe I can buy the VM for this year and then try to work out what future years will do as the project progresses? Or maybe even Duke will keep the VM up if it proves helpful?

What are your thoughts?

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Thursday, July 9, 2020 2:22 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

Yes, there is an annual fee, in the low hundreds of dollars per year for what we are discussing.  But then, nothing is truly free.  AWS and other services are, of course, available for a fee.  Free hosting sites exist, but will likely come with a requirement to display advertisements.  Hosting-only services may not give you the database and R services you need.  The .duke.edu suffix on your URL may also be of value, lending an official, academic air to the site.  To me, the real benefit of an in-house solution is the support that you receive.  OIT has experts in just about every field imaginable, including security, performance, programming, visualization, and statistics.  I draw on their knowledge all the time.  If your project is worthy (disseminating knowledge to the public), you should have no problem renewing funding in the future.  You might even consider arranging that now.  Two of the links I sent yesterday were the result of Data+ projects from years past.  They remain on-line and current.  The university generally has an interest in maintaining useful and relevant resources.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Thursday, July 9, 2020 12:01 AM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

Thank you for this information! Yes, my Bass Connections Student Research Award has funding for resources, but only for 2020-2021 and I would like the framework I create to be able to continue being utilized after next year. Therefore, I am worried about using a VM that has an annual cost. Would the approach that you have recommended to researchers in the past, "have Research Computing configure a VM, with all necessary software components" cost money/ have an annual cost?

Thank you so much again,

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Wednesday, July 8, 2020 11:11 AM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

Being part of Bass connections, I assumed that your project has funding for resources.  The cost of a VM on, say, Research Toolkits is in the low hundreds of dollars annually and it comes with support from Research Computing and gives you Duke affiliated URLs for your web site(s).  This has worked well for the many projects I have been involved in.  The approach that I have recommended to researchers is to have Research Computing configure a VM, with all necessary software components, and assist with database development, R, and Shiny development as needed.  One benefit to you is that your time is spent populating the database, implementing scripts for modeling, and developing the system for rendering results (user interface, graphs, tables, etc.) as opposed to system configuration.  Following are a few examples of platforms we have developed for research teams.  They involve various combinations of database, R, and Shiny resources.

[http://h2p2.oit.duke.edu](http://h2p2.oit.duke.edu/)

[http://bd4rh.rc.duke.edu:3838](http://bd4rh.rc.duke.edu:3838/)

<http://shindellgroup.rc.duke.edu/>

[http://waterqualityexplorer.rc.duke.edu:3838](http://waterqualityexplorer.rc.duke.edu:3838/)

I have used VirtualBox to run Linux on a Windows machine and it functions fine, but you still need a machine of some sort to install it on.  If you have a server available, that is an option, but since your project is Duke affiliated, it seems like a Duke server makes sense.  Back to a VM somewhere on Duke’s network.  This also restricts database access to internal and VPN connected devices.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Tuesday, July 7, 2020 11:54 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

Thank you so much! I agree, it seems like using Access for data collection will be the easiest, and I will start attempting that approach this week. In terms of getting a VM, why do you suggest that I get a VM through my project PI or a professor? Can I get one myself through something like VirtualBox, or should I not do this for security reasons?

Also, thank you for taking another look at the rest of my questions!

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Tuesday, July 7, 2020 2:41 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

Yes, I read your last message.  One of the main points I took is that all data will be entered from within Duke’s network.  This makes the Access/SQL design for entry with a Shiny Server design for on-line querying a possibility.  Using a database server solves your simultaneous entry problem and the Shiny Server process can also run on the same VM as your database, so everything is accessible from a single server.  If all data entry users have Access installed on a personal device (it is available to all faculty, staff, and students at Duke, but may only run on Windows) then you might consider this design.  I have implemented many data entry and analysis systems using this approach.  However, the HTML forms approach is also valid, but I have always found it to be a bit more difficult to implement.  Access is easy and if you need it only for data entry then it is a simple solution that gives you important features like error handling, forms with tables from which rows can be selected for editing, etc.

Let me take another look at your questions, then I will send an additional response.  In the meanwhile, I suggest that you request resources for a VM (from your project PI or professor).  I recommend Linux, four cores and about 20g of mem and 50g of storage.  That should suffice for starters.  I can also configure a VM for experimentation that can be transferred to your project later, if you would like to try that.

More later.

Tom

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Monday, July 6, 2020 10:13 AM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

I hope you had a great weekend! I was wondering if you got my last email with my questions- I have already received a lot of helpful information from you for this project, but am still not quite sure where to start (especially for data collection). If you have the chance to answer my questions below, that would be amazing!

Thanks,

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Sent:** Thursday, June 25, 2020 2:25 PM  
**To:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Cc:** David Gill <[david.gill@duke.edu](mailto:david.gill@duke.edu)>  
**Subject:** Re: Bass Project Getting Started Guidance (Research Computing)

Hi Tom,

I’m so sorry for the late reply, I have been looking into the helpful information you sent in order to form my response and coordinating with my research supervisor to get his thoughts. I really appreciate your help- I think I have a much better idea of how I would like to start this project! However, I do have the following questions:

1. One of your bullet points says “Access is a good choice for data entry, but distributing an Access app across universities is challenging (VPNs and inter-institution accounts are required).” Everyone entering data for this project will most likely be affiliated with Duke University, by different campuses I meant Duke’s main campus and Duke’s marine lab. However, for my research supervisor's last project, Access didn’t work for simultaneous data entry. Is there a way to get Access to work for simultaneous data entry so that the team members can work at the same time?

2. Also, you said “searching for and modifying existing records is a bit challenging without some sort of user interface query layer.” Could I use PostgreSQL as a user interface query layer to search for and modify existing records from my SQL database?

3. A bullet point says: “Many other options exist, such as HTML forms and Visual Studio, but these add additional layers of complexity that are beneficial only if continued and extensive record management are required.” What exactly do you mean by continued and extensive record management? Data entry would be concentrated in this project with the goal of the team members to entering the data and then doing analysis with the results after. Would record management mean changing the data entry form to use for other research projects? I’m just trying to figure out if I might want to use HTML forms, Visual Studio, etc.

4. Could you give an example of when “data management will continue as results are analyzed” would happen? Would this be when, for example, I had an entry field that was City, State and later I decided to split it up into City and State for analysis?

5. I may need help with configuring a VM sized according to concurrent user and data volume requirements - can I do this in anaconda? And this VM would hold the data in the database? I recently learned how to use Docker and was wondering if this might be helpful for storing data?

Sorry for all the questions, but already your help has given me much more direction! If any of my questions don’t make sense, please let me know as I am very new to this!

Thank you so much again!

Grace

Grace Llewellyn

Duke University | Class of 2022

Computer Science | Environmental Sciences & Policy

**From:** Tom Balmat <[thomas.balmat@duke.edu](mailto:thomas.balmat@duke.edu)>  
**Sent:** Thursday, June 18, 2020 1:53 PM  
**To:** Grace Llewellyn <[grace.llewellyn@duke.edu](mailto:grace.llewellyn@duke.edu)>  
**Subject:** RE: Bass Project Getting Started Guidance (Research Computing)

Hi Grace,

Here are a few observations:

* For robust data recording and secure data housing, a SQL database is a good design choice (google “SQL” and “ACID”)
* R and Shiny are great for analysis, model fitting, and visualization, but they have limited out of the box features for implementing a true data collection system
* Access is a good choice for data entry, but distributing an Access app across universities is challenging (VPNs and inter-institution accounts are required)
* A web app implemented in HTML can create new database records using posts, but searching for and modifying existing records is a bit challenging without some sort of user interface query layer
* Security in an HTML environment is not as stringent as with a VPN approach
* Shiny implements HTML (no java required), so that data collection can be implemented over the web and, because it is layered on R, various querying functions are available, but these would have to be wrapped into a data review and maintenance system, tailored to your project (the result might be good, but with much work – basically repeating what is already in Access)
* R, Shiny, Access, and various other user interfaces can be interfaced to SQL
* If your modeling and analysis system will be implemented in R and Shiny then using them for the data collection phase might make sense, to promote consistency of functions and appearance, especially if analysts who conduct studies will also review and maintain data
* Many other options exist, such as HTML forms and Visual Studio, but these add additional layers of complexity that are beneficial only if continued and extensive record management are required

If you intend to conduct a single, concentrated data entry phase, especially if it involves only a few people, then you might consider using something like Access for that, then use R and Shiny Server for analysis.  If data management will continue as results are analyzed, then R and Shiny Server might make sense.  Research Computing has implemented many server environments using MySQL and Shiny Server, primarily for analysis, model fitting, and visualization, with very good results.  Typically, we configure a VM sized according to concurrent user and data volume requirements, then a project leader (you) takes over with database and application design.  We have significant experience with SQL, R, Shiny, Access, and HTML and remain available to you to answer questions and address specific scripting, querying, and statistical modeling questions.  This has resulted in on-line apps that make the work of researchers available to their community in a collaborative, engaging, and effective manner.  Please consider the above points and let me know if you are interested in exploring any of these ideas.

Tom