# A whirlwind tour of the NSF databases and data analysis in Excel

## Jack Snoeyink, [snoeyink@cs.unc.edu](mailto:snoeyink@cs.unc.edu), 18 Nov 2017

This document focusses primarily on using data from the report server to populate Excel spreadsheets

## Databases: Sybase ReportServer and Oracle Enterprise Reporting

### What is the ReportServer?

The ReportServer copies some of NSF’s production databases, including rptdb and FLflpdb, to support rapid response to queries that may come from other parts of the government, making it a fantastic resource for those who want to build automation tools. It is a Sybase ASE 16 database, for documentation, see [the Sybase Infocenter](http://infocenter.sybase.com/help/index.jsp?topic=/com.sybase.infocenter.help.ase.16.0/doc/html/title.html). Either search, or start with the Reference Manuals and TransACT SQL. In DIS Weber Lauh is an expert on creating queries, and I have learned a lot from queries he has created for me.

Note: [Wikipedia](https://en.wikipedia.org/wiki/Adaptive_Server_Enterprise) reports that Sybase 15.5 was released in early 2010, then SAP bought Sybase. 16 was released in April 2014. NSF upgraded from 15.5 in 2017. This was encouraging, because I thought that the reportserver would be going away in favor of the Oracle DB behind Enterprise reporting. m

### What programs talk to ReportServer?

ReportServer listens to port 5000 on rptsql.nsf.gov. Only PCs have the ODBC connection by default; Macs can use 3rd party openlink

At NSF, all PCs have the app [DBI SQL](http://infocenter.sybase.com/help/index.jsp?topic=/com.sybase.infocenter.dc00168.1603/doc/html/san1276279412126.html)[[1]](#footnote-1) which is useful for creating and optimizing queries.   
Excel, on both PC and Mac, contains Microsoft query, a visual query editor in Excel ’97, based on technology from [Zloof 1977](https://en.wikipedia.org/wiki/Microsoft_Query).

### What is Enterprise Reporting?

The team of Priya Jayaraman [pjayaram@nsf.gov](mailto:pjayaram@nsf.gov) is actively extending the Oracle Dimension database behind Enterprise Reporting, which is the clear successor to the ReportServer. Excellent team, but has to deliver tools that work for all NSF, so one can more quickly prototype something that works for yourself (and maybe share it with them to show them what is useful at least to you, and guide their implementation.)

The limitation of a Dimension database is that you can make connections only if the DB designer has built them in. E.g., suppose you have a list of highly rated proposals that you are considering for award, and want to know, for all the PIs & co-PIs, what are their current active awards and spending rates. You need to link from proposals to PIs to awards to obligations. When I first asked, none of those links was there, but a month later they could link proposals to co-PIs, and I think 9 months later I could link Proposal PIs to Award PIs. By prototyping in ReportServer,

### Why use database query tool?

When a spreadsheet on someone’s computer is the sole authority on, say, which projects are to be funded, then nothing can be done if that person is sick or that computer fails.

A properly designed tool will have the computer do as many of the repetitive and tedious steps as possible, because those are where human inattention leads to errors.

, allowing you to focus on work that requires additional knowledge, insight, and/or experience.

As you get familiar with the tools, they may

While many see automation as a way to get the computer to do my work for

### Where is the data dictionary?

The most reliable source for information about the data is the database itself.

SELECT TOP 20 p.\* FROM csd.prop p WHERE p.prop\_id LIKE ‘18\_0\_0\_’

You cannot do with with csd.pi\_vw, because it contains a protected column with social security numbers, but you can name fields you want to see.

SELECT TOP 20 p.lead\_prop\_id, p.prop\_id, pi.pi\_last\_name, pi.pi\_frst\_name, pi.pi\_gend\_code  
FROM csd.prop p   
JOIN csd.pi\_vw pi ON pi.pi\_id = p.pi\_id  
WHERE p.prop\_id LIKE ‘18\_0\_0\_’

Query to find which tables contain a given field or table

-- column names & types for table

SELECT sc.colid, sc.name, t.name, sc.length FROM sysobjects so

JOIN syscolumns sc ON sc.id = so.id

JOIN systypes t on t.usertype = sc.usertype

WHERE so.name like '#myTmp%' -- table name

ORDER BY colid

-- table & column names from fields

select so.name, sc.name, t.name, sc.length

from syscolumns sc

JOIN sysobjects so ON sc.id = so.id

JOIN systypes t on t.usertype = sc.usertype

where sc.name like '%ibm\_logn%'

order by so.name, sc.name

-- with protection information, too

select so.name, sc.name, t.name, sc.length, su.name, sp.uid, sp.action, sp.protecttype, sp.columns

from syscolumns sc

JOIN sysobjects so ON so.id = sc.id

JOIN systypes t on t.usertype = sc.usertype

JOIN sysprotects sp ON sp.id = sc.id

JOIN sysusers su ON su.uid = sp.uid

where sc.name like '%logn%' -- field name

AND sp.uid = 0 -- skip to see non-public tables, too

order by so.name, sc.name, su.name

Don’t overlook the power you get from being able to query all submitters, say, and deciding

## Mac section

### Openlink installation

Thanks to , the NSF security team has approved openlink as a 3rd party driver for connecting to ReportServer and making queries

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 4/11/2017 5:00 PM | 1-178045149 | [ODBC Driver for Sybase Database](https://collaboration.inside.nsf.gov/sites/it_specialists/_layouts/15/listform.aspx?PageType=4&ListId=%7B8EA2266E%2DE31C%2D4E0E%2D94E7%2DFDA993D90D87%7D&ID=1110&ContentTypeID=0x0100405069017ABFF64B83AC6B4BD9325826) |  |  | Accepted | allow MacOS / MS Excel 2016 users to connect to Report Server database |

### Connection string:

Password form:

VBE environment:

Change sheet/module names:

Tables & table references:

ListObjects and QueryTables:

Keeping sheets uncorrupted:

SQL:

Concatenation

Temporary tables and tempdb:

Declare @variable

Parameters

Tables

csd

prop

addl\_pi\_invl

awd

budg\_splt

eps\_blip

pi\_vw

revr

rev\_prop

panl

panl\_prop

panl\_revr

cgi

SELECT \* FROM csd.pgm\_ele WHERE pgm\_ele\_code LIKE ‘779%’

SELECT \* FROM csd.pgm\_ele WHERE pgm\_ele\_long\_name LIKE ‘%brain%’

SELECT count(\*) FROM csd.abst WHERE prop\_id LIKE ‘17%’ AND abst\_txt like ‘%?[a-z]%’ – # of abstracts with ?letter

FLflpdb

COVR

Oracle Smart View

NSF’s installation instructions for Smart View:

<https://collaboration.inside.nsf.gov/oirm/dis/edw/BI%20Training/BI%20Training/SmartView%20for%20Office%20-%20Installation%20Guide.pdf#search=smart%20view>

Excel Get & Transform (aka Power Query)

Note that this is included in Excel 2016 as Data>>Get & Transform. It is one way of getting around the linkage limitations of Enterprise Reporting – you can use the Oracle SmartView to get all relevant data from ER into your spreadsheet and then you can do the linking and query computations in Excel that the database should do for you...

<https://www.excelcampus.com/install-power-query/>

**Power Query** is included with **Excel 2016**. It has been renamed and is now on the Data tab of the Ribbon in the Get & Transform section. This means there is nothing to install. If you are using **Excel 2016**, go to the Data tab on the ribbon and press the New **Query** button to create a **query** and open the **Power Query** editor.

I’m going to send emails as I learn things while porting, like these first two:

1. Maria, I see that your Settings page is using embedded ActiveX forms for the userid and password.  ActiveX doesn’t work on Mac, so I’m going back to named ranges on the sheet.  Since the sheet will be hidden, we don’t need to have a pretty form there.  (The userform will still work on both mac & PC.)  
   Are there any considerations that I am missing by doing this?
2. I also realized that there is a better way of keeping the long SQL query strings:   
   Paste the relevant pieces straight from my SQL code into named ranges on a hidden sheet, rather than dealing with the quotes, vbNewLines, and continuation characters needed to make them into vba strings.  (Input SQL is still good as strings because that is more parsing than static strings.)

I wrote the attached macro that looks in my sql code for comments like

--[name

SELECT ....  --sql code

--]name

It will take that code and put it into a table with that name (and last update timestamp).  Other stuff in the file (like the sql code for testing) is ignored.  This makes it really easy to maintain the sql with its test cases and copy into pd-3po or the tracker...

1. Quoting the Sybase documentation... Note: For backward compatibility, SAP Sybase IQ includes the older Interactive SQL Classic (dbisqlc) utility. Use Interactive SQL rather than Interactive SQL Classic. Interactive SQL Classic is deprecated and will be removed in a future release of SAP Sybase IQ. [↑](#footnote-ref-1)