



Dennis Wightman



**Immune  
Tolerance  
Network**

**Globus File Transfer Module and Other Useful Tools**

# Agenda

- **Background on the new File Transfer Module**
- **What's a Globus?**
- **Why did we need to integrate it with TrialShare?**
- **Extending what LabKey wrote for us**
- **Other tools we've written for working with LabKey servers**

# Background

- **The Immune Tolerance Network has more and more large files such as RNASeq data and FCS files**
- **Many of these are appropriate to be shared with the community and outside collaborators**
- **Adam Asare evaluated options and chose the Globus Endpoint system**
- **The Globus Connect system was created by the University of Chicago as a way to create a network to share large files**

# What is Globus Connect?

## What is Globus Connect?

Globus Connect is easy-to-install, pre-configured software that turns your laptop, server, cluster or other local resource into a Globus [endpoint](#).

There are two versions of Globus Connect, one for use with personal machines such as your laptop, and another for use with server-class machines such as campus computing clusters and lab servers. Use [Globus Connect Personal](#) to enable file transfer to and from your personal machine (laptop or desktop.) A Globus Connect Personal endpoint is intended to be used only by a single user. [Globus Connect Server](#) enables system administrators to turn shared resources such as campus clusters and lab servers into a Globus endpoint. If it is made publicly visible, a Globus Connect Server endpoint can be used by multiple Globus users.

From What is Globus Connect:  
<https://www.globus.org/globus-connect>

# What is a Globus Endpoint?

## What is an "endpoint"?

An "endpoint" is one of the two file transfer locations – either the source or the destination – between which files can move. Once a resource (server, cluster, storage system, laptop, or other system) is defined as an endpoint, it will be available to authorized users who can transfer files to or from this endpoint.

From Globus endpoints explained:

<https://docs.globus.org/faq/globus-connect-endpoints/>

# “Integrate Globus with TrialShare”

- Adam Asare, the former Director of ITN Bioinformatics and TrialShare, asked Noha Lim and me to “Integrate our new Globus Endpoint with TrialShare”
- What does integration mean?



Noha



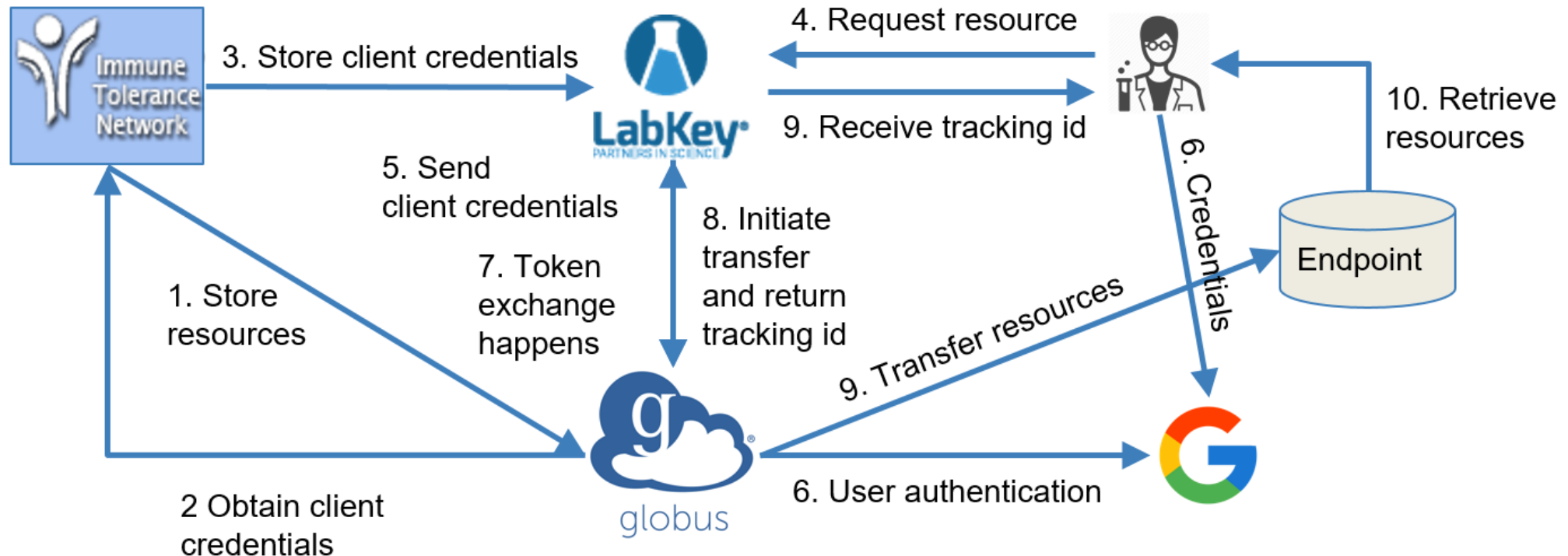
Dennis

# What is Integration?

- **Maintain a list of endpoint files inside a TrialShare project, and allow users to initiate file transfer from within TrialShare/LabKey**
- **Display relevant scientific data to associate files with information**
- **Pressure from other projects... so let's have LabKey do the initial development**
- **Start with a List as the interface for the user... LabKey Lists can go in any container and can join to other data**

# What is Integration?

- Susan Hert from LabKey developed the negotiation system and initial version of the File Transfer Module (diagram courtesy of Susan):





# Development Process

- **LabKey delivered on time, and the File Transfer Module did everything we had asked for**
- **During the development process, much IT configuration had to be done**
- **A file server needed to be set up and configured as a Globus Connect Server**
- **Decision was made to only ever put publicly available files on the ITN Endpoint**
- **This simplified security, and eased the development/configuration process**
- **Needed to move our Dev server outside the domain for testing**
- **Users can download files to their machine with Personal Connect, or if their organization allows them to, they can initiate a file transfer from the ITN Endpoint to their org's Endpoint.**

# Lists don't really join with Datasets

- Realization that we needed a Dataset, not a List for our use case
- Be careful what you ask for when driving software development



Noha



Dennis

# Lists don't really join with Datasets

- We want the users to be able to create their own joined views between endpoint files and any study data that will help them choose files
- Admins and editors should be able to create joined views to present to users, without duplicating data from a dataset into a list
- We build from source, and have developed several modules in-house
- One week of coding let us use Datasets instead of Lists



# The Final Product

**Immune Tolerance Network Clinical Trials Research Portal** INNOVATION · COLLABORATION  
A clinical research consortium funded by NIDDK

AbATE ITN027A: Autoimmunity-Blocking Antibody for Tolerance in Recently Diagnosed Type 1 Diabetes

Overview Data & Reports Participants Download CD8 Fastq Files Download W5 Fastq Files Manage

TrialShare File Transfer Module Help ▾

1) Create a Globus endpoint

You will need a Globus endpoint to transfer these files.

A Globus endpoint is a place where you can transfer data to (or from) using the Globus Connect client. You can set up a personal endpoint to transfer data from TrialShare to your computer. Go to <https://www.globus.org/globus-connect-personal> to set up your endpoint.

You will also need to log in with a Globus ID. Your organization may already have a login you can use; this will be visible on the login page when you attempt to transfer files. Otherwise, you can create a Globus ID at <https://www.globusid.org/create>.

2) Make sure the Globus client is installed

3) Select files and click "Transfer"

4) Optional: Select an endpoint

Troubleshooting (Click to expand)

GSE85531 Whole Blood


**GSE85531 Whole Blood RNA-Seq ▾**


FILTER	GRID VIEWS ▾	REPORTS ▾	CHARTS ▾	GROUPS ▾	EXPORT ▾	PRINT	PAGING ▾	MANAGE	INSERT ▾	DELETE	VIEW SPECIMENS	TRANSFER	1 - 100 of 190 Next >
Participant ID	Visit	Run	Filename	Months	Sample Name	Spots	Bases	Avg Length	Library Selection	Platform	Model	Treatment Group (Char)	Age (years) Rounded
<input type="checkbox"/> AbATE_106306	19	SRR4026931	SRR4026931.fastq	6mo	GSM2275079	12712228	648323628	51	cDNA	ILLUMINA	Illumina HiSeq 2000	Sort Ascending	12.0
<input type="checkbox"/> AbATE_106306	42	SRR4026979	SRR4026979.fastq	24mo	GSM2275127	11505927	575296350	50	cDNA	ILLUMINA	Illumina HiSeq 2000	Sort Descending	12.0
<input type="checkbox"/> AbATE_122083	0	SRR4026884	SRR4026884.fastq	0mo	GSM2275032	6740772	337038600	50	cDNA	ILLUMINA	Illumina HiScanSQ	Clear Sort	14.0
<input type="checkbox"/> AbATE_122083	19	SRR4026929	SRR4026929.fastq	6mo	GSM2275077	11904201	607114251	51	cDNA	ILLUMINA	Illumina HiSeq 2000	Filter	14.0
<input type="checkbox"/> AbATE_122083	21	SRR4026837	SRR4026837.fastq	12mo	GSM2274985	4552222	227611100	50	cDNA	ILLUMINA	Illumina HiScanSQ	Clear Filter	14.0
<input type="checkbox"/> AbATE_122083	42	SRR4026976	SRR4026976.fastq	24mo	GSM2275124	11268443	563422150	50	cDNA	ILLUMINA	Illumina HiSeq 2000	Remove Column	10.0
<input type="checkbox"/> AbATE_137567	0	SRR4026874	SRR4026874.fastq	0mo	GSM2275022	5290266	264513300	50	cDNA	ILLUMINA	Illumina HiScanSQ		10.0
<input type="checkbox"/> AbATE_137567	19	SRR4026919	SRR4026919.fastq	6mo	GSM2275067	11633165	593291415	51	cDNA	ILLUMINA	Illumina HiSeq 2000		10.0
<input type="checkbox"/> AbATE_137567	42	SRR4026967	SRR4026967.fastq	24mo									
<input type="checkbox"/> AbATE_156104	0	SRR4026892	SRR4026892.fastq	0mo									
<input type="checkbox"/> AbATE_156104	19	SRR4026940	SRR4026940.fastq	6mo									
<input type="checkbox"/> AbATE_156104	21	SRR4026866	SRR4026866.fastq	12mo									
<input type="checkbox"/> AbATE_156104	42	SRR4026988	SRR4026988.fastq	24mo									
<input type="checkbox"/> AbATE_161201	0	SRR4026909	SRR4026909.fastq	0mo									
<input type="checkbox"/> AbATE_161201	19	SRR4026956	SRR4026956.fastq	6mo									
<input type="checkbox"/> AbATE_161201	21	SRR4026849	SRR4026849.fastq	12mo									
<input type="checkbox"/> AbATE_161201	42	SRR4027007	SRR4027007.fastq	24mo									
<input type="checkbox"/> AbATE_166103	42	SRR4027016	SRR4027016.fastq	24mo									


**GSE85531 Whole Blood RNA-Seq ▾**

FILTER	GRID VIEWS ▾	REPORTS ▾	CHARTS ▾	GROUPS ▾	EXPORT ▾	PRINT	PAGING ▾	MANAGE	INSERT ▾	DELETE	VIEW SPECIMENS	TRANSFER	1 - 100 of 127 Next >
Filter: (TRTC = hOKT3) CLEAR ALL													
View: default													
Participant ID	Visit	Run	Filename	Months	Sample Name	Spots	Bases	Avg Length	Library Selection	Platform	Model	Treatment Group (Char)	Age (years) Rounded
<input checked="" type="checkbox"/> AbATE_106306	19	SRR4026931	SRR4026931.fastq	6mo	GSM2275079	12712228	648323628	51	cDNA	ILLUMINA	Illumina HiSeq 2000	hOKT3	12.0
<input checked="" type="checkbox"/> AbATE_106306	42	SRR4026979	SRR4026979.fastq	24mo	GSM2275127	11505927	575296350	50	cDNA	ILLUMINA	Illumina HiSeq 2000	hOKT3	12.0
<input type="checkbox"/> AbATE_122083	0	SRR4026884	SRR4026884.fastq	0mo	GSM2275032	6740772	337038600	50	cDNA	ILLUMINA	Illumina HiScanSQ	hOKT3	14.0
<input type="checkbox"/> AbATE_122083	19	SRR4026929	SRR4026929.fastq	6mo	GSM2275077	11904201	607114251	51	cDNA	ILLUMINA	Illumina HiSeq 2000	hOKT3	14.0
<input type="checkbox"/> AbATE_122083	21	SRR4026837	SRR4026837.fastq	12mo	GSM2274985	4552222	227611100	50	cDNA	ILLUMINA	Illumina HiScanSQ	hOKT3	14.0
<input type="checkbox"/> AbATE_122083	42	SRR4026976	SRR4026976.fastq	24mo	GSM2275124	11268443	563422150	50	cDNA	ILLUMINA	Illumina HiSeq 2000	hOKT3	14.0
<input type="checkbox"/> AbATE_137567	0	SRR4026874	SRR4026874.fastq	0mo	GSM2275022	5290266	264513300	50	cDNA	ILLUMINA	Illumina HiScanSQ	hOKT3	10.0

# The Final Product

**Immune Tolerance Network**  
Clinical Trials Research Portal

**TrialShare**  
INNOVATION · COLLABORATION



Manage Data · Publish · Groups · Support · Account

AbATE ITN027AI: Autoimmunity-Blocking Antibody for Tolerance in Recently Diagnosed Type 1 Diabetes

### Globus File Transfer

You are about to transfer the following files:

- SRR4026979.fastq
- SRR4026931.fastq

First, **SELECT AN ENDPOINT**

[BACK](#) [TRANSFER](#)

### Browse Endpoint

Endpoint:  [☆](#)

Path:  [Go](#)

select all up one folder refresh list

178EE20B.PNG

1913E5B8.PNG

2ACD41E4.PNG

3726FFD0.PNG

37313BCB.PNG

3FF76823.PNG

4D0DD56D.PNG

5F38728C.PNG

6415DB70.PNG

7813C1AC.PNG

7BD05C59.PNG

AB60046B.PNG

B28A3356.PNG

B6B701C8.PNG

B9610D87.PNG

DAC761B2.PNG

E2AF358C.PNG


FC9F8524.PNG

FTM\_BOPE\_snippet.txt

L\_9E02.tmp.PNG

Label This Transfer  This will be displayed in your transfer activity.

[Submit](#)

**Immune Tolerance Network**  
Clinical Trials Research Portal

**TrialShare**  
INNOVATION · COLLABORATION  
A clinical research consortium funded by NIAID

AbATE ITN027AI: Autoimmunity-Blocking Antibody for Tolerance in Recently Diagnosed Type 1 Diabetes


### Globus File Transfer

You are about to transfer the following files from directory /ITN027AI/ITN027AIS1/GSE85531 on source endpoint 'ITN\_Public'.

- SRR4026979.fastq
- SRR4026931.fastq

These files will be downloaded to the endpoint 'DW\_DELL\_LAPTOP'.  
Click "Transfer" to start the download, or [SELECT A DIFFERENT DOWNLOAD ENDPOINT](#)

[BACK](#) [TRANSFER](#)





Manage Data · Publish · Groups · Support · Account

[Transfer Files](#) | [Activity](#) | [Endpoints](#) | [Bookmarks](#) | [Console](#)

### Activity

[Recent Activity](#) [History](#)

 **Test transfer**  
warning: connection reset; see event log

 **IE 3**  
transfer completed 2 days ago

# Other ITN tool programs for LabKey

- Querycat: a small Java program that allows TrialShare users to use the LabKey Remote API to write LabKey SQL queries to a database of their choice (Open source)
- Dangermouse: a small but dangerous Java program that exports studies to file pipelines, deletes the existing files, and places the new export in the pipeline. (Requires a new action in the Pipeline module source code. Contact [dwightman@immunetolerance.org](mailto:dwightman@immunetolerance.org) for information about this program)
- Biocat: a scheduling program for loading specimens and datasets into TrialShare/LabKey (very ITN specific, contact Dennis for information)
- Servicecat: A microservice patterned Java web application that provides a REST API to import specimens from our DBoR and load them into TrialShare (in testing, contact Dennis for information)
- Hiapo: A small Java program that can 'diff' two LabKey servers (i.e. TrialShare Production and TrialShare Data Staging) and write DB tables with information about participants, specimen counts and types, dataset row counts and column information (good candidate for open sourcing, contact Dennis if you are interested).
- Xplort: (Xplan Export) A small Java program that works with the ITN Xplan (Experimental Plan) module to programmatically pull files that are 'blobbed' into a LabKey server database and export them to the server's local file system (specific to ITN, but code may be useful for others).

# Useful Links

- ITN TrialShare: <https://www.itntrialshare.org>
- A File Transfer Module in The AbATE Scientific Immunology study:  
<https://www.itntrialshare.org/project/Studies/ITN027AISI/Study%20Data/begin.view?pagelId=Download%20WB%20Fastq%20Files>
- File Transfer Module on LabKey Github: <https://github.com/LabKey/fileTransfer>
- Globus Connect: <https://www.globus.org/globus-connect>
- Globus endpoints explained: <https://docs.globus.org/faq/globus-connect-endpoints/>
- Querycat on Github: <https://github.com/DenWJ/querycat>