Data Federation Training Working Group

Updates and Vision for the Future
Members

Abbas Saka  
*Institutional Support Manager*
Adediran Daniel Adewole  
Alberto Camara Bellesteros  
Bob Cottingham  
Can Vuran  
**Lincoln**
Ghulam Sarwar  
*Faisalabad Pakistan*
Jennifer Clarke  
Jinha Jung
Marcos Paulo da Silva  
Mark Wilkinson  
Monica Poelchau  
Paola Pesantez  
Peter Selby

Sectoral Policies and Institutional Support Manager

Helix Biogen Institute  
CBGP UPM/INIA-CSIC  
Oak Ridge National Laboratory  
University of Nebraska-
Cotton Research Station, AARI,  
University of Nebraska-Lincoln  
Purdue University  
University of Arkansas  
CBGP UPM/INIA-CSIC  
USDA-ARS  
Washington &  
Cori
Objectives

From the Working Group Proposal

“... This working group will provide training resources on data sharing technologies, either via a collection of existing, vetted training materials; generation of new, written training materials; and/or other materials…”

“...Roughly one third of data federation survey respondents indicated that they would benefit from learning more about Discoverable APIs; Linked-Data; Client-side integration of results from multiple data sources; Index-driven search technologies; Data Management Systems; and Data Sharing via services (e.g. Globus)...”
Results – Technology awareness

What data sharing technologies are you familiar with?

What data sharing technologies would you appreciate learning more about?
Updates

- Finding a meeting schedule
  - Every other Friday at 12 EST

- Developed a short term strategy to begin working
  - Create a list of technologies to focus on
  - Learn about these technologies through a series of guest presentations

- Two presentations down, more scheduled
  - March 17 - Peter Selby - BrAPI (https://youtu.be/K8NQehzBTDA)
  - March 31 - Mark Wilkinson - RDF (https://youtu.be/c8fZaGU3d0A)
  - May 26 - Nirav Merchant - iRODS
  - June ?? - Cyril Pommier - Index driven search, FAIDARE
  - ?? - Linked-data (JSON-LD)
  - ?? - GraphQL
  - ?? - Globus
Future Vision

● Repository of training material
  ○ Recordings
  ○ Written material for example use cases
  ○ Links to further training material and documentation

● Demo of at least one of the technologies
  ○ Simple use case, low cost demo
  ○ Illustrate what can be done, inspire further development

● Assessment of matching technology to use case
  ○ Document examples of when some tech is better suited than others
  ○ Advise the community on where to focus their efforts

● Touch base with Education WG
Discussion Questions

● As a database maintainer, what types of data do you have available that you think is in common with other groups/systems/databases?
  ○ What data do you have that is unique, but generally useful for the community?
  ○ What community standards are you using?

● As a scientific user, what types of data do you want but are having trouble accessing?

● What medium would be best for the training?
  ○ Would a repository of training materials be helpful?
  ○ Would a demo of technology be helpful?

● Who is the audience for the training?
  ○ Who needs this training from your group and what is their role (PI, student, developer, scientist, etc)?