

SC_2012.06.05

iPG2P Steering Committee June 5, 2012

Participants

Matt Vaughn, Steve Welch, Steve Goff, Martha Narro, Ruth Grene, Chris Myers, Dan Kliebenstein, Jeff White, Dan Stanzone

Agenda

- Update on development of iPlant's infrastructure
- New features in DE 1.2 (Verdant Prime) and apps of interest to G2P
- Updates from Steve Goff
- Possible collaboration on climate modeling
- Security breach at UA

Notes

Update on development of iPlant's infrastructure

- iPlant's infrastructure continuing to evolve.
- At end of 2012 prior to PAG. All services were 1.0 (beta). Polished since then (Atmosphere, DE, and others).
- Also made a lot of progress on unified authentication.

User services and authentication.

- There is no longer an array of places to login.
 - Initially this was the case since we'd taken a number of different technologies and put them together.
- Now we have a unified user management and authentication service (Trellis).
 - <http://user.iplantcollaborative.org/>
- All the different iPlant services are listed.
- Example of the power this offers:
 - In CIPRES, can now log in with iPlant credentials.
 - Paves the way for full integration between CIPRES and iPlant.
- Encouraging other portals to do the same. They can offload managing authentication and gain unified storage, access to HPC resources.
- Eventually CIPRES will show up on the services list on Trellis webpage.

Atmosphere

- A fantastic new addition is that now have clients (desktop or shell) directly accessible in the VM instance.
 - Previously would have had to get address, go outside the VM and ssh into shell.
 - Now can click shell icon within the VM instance and login using iPlant credentials.
 - Can launch a VNC session (remote desk top session) within the VM instance too.
- People are responding very well to it since it's easier to use.

DE's new features

- UI hasn't changed which is a sign of maturity.
- Data window
 - Now supports high speed threaded uploads from desktop.
 - Web Browsers limit upload to 1.9 GB
 - We can take advantage of iRODS' iDROP Lite.
 - iDROP loads a small Java applet.
 - Can select multiple files for upload.
 - Can specify upload destination.
 - Moves large files pretty much at the limit of the users network connection.
 - Uploading from URL, can upload many as long as each is less than 1.9 GB.
- iPlant Data Store feature
 - Now have the beginning of being able to add metadata to files.
 - Eventually will be able to do bulk editing and intelligent searching on the metadata tags.
 - Will make it easier to keep track of files and their analysis.
 - The metadata services are mostly in place, but the UI development is lagging, but catching up.
 - Can filter for file names.
- SW: UC Davis will move 330,000 images so have repositories at UC Davis, iPlant, Kansas. Can these files be batch tagged using a script?
 - MV: Even better than that, iDROP provides functionality to do that. So yes.

- DS: If the Kansas or UC Davis file storage systems don't support metadata tagging, the metadata will be lost. Just be aware of that and plan for it.
- MN: Will metadata tags be inherited?
 - MV: They can be and usually that would be the desired behavior, but not always. In some cases, especially with a lot of metadata, may not want to inherit all of it. So, some careful thought needs to be put into implementing inheritance.
- Analyses Window
 - Can view and save a job's parameters, which is part of provenance tracking.
- Applications catalog
 - Looks similar
 - Thinking about ways to present customized views so only see the tools relevant to subdomains of biology.
 - Full text search on apps has been implemented. Works on both the app name and written description.
 - Now 212 public applications
 - Notable new entries:
 - 50% of EMBOSS suite
 - Assemblers: Trinity for denovo, runs on TACC resources and is much easier to do through DE rather than through command line.
 - Newbler (454 assembler)
 - In development: Soap denovo,
 - Peptide analysis (pepMOD)
 - QTL and GWAS mapping algorithms (by Christos and Liya)
 - BED
 - A new RNASeq app in experimental category
 - Micro RNAseq
 - Autodoc vina (for molecular docking), publishing for public use.
 - Automatically generate the template for the documentation page, so it's easy and consistent for all documentation pages.
 - All these apps are being integrated since we've made it easier.
 - Also, to highlight another feature, how do we know when an application is working and when there's a problem with it – to capture community feedback?
 - Now when a user rates a tool they get a comment box and the comment is attached to the bottom of the document page for the app. We'll have to have someone police this.
 - Pipelines allow convergence not just linear pipelines.

Updates from Steve Goff

- Steve G has been working with Joellen Russell, who has some of the best climate models.
 - 5-10 KM resolution instead of 100 KM climate resolution.
 - <http://www.geo.arizona.edu/~jrussell/>
 - Steve wrote to Syngenta, Monsanto and Pioneer and made an initial proposal to see if they'd be interested in discussing modeling GxE since we have access to high resolution climate model data.
 - All three have responded positively already.
 - Will need to discuss if they would be interested in letting their yield-related data be used to train the climate models.
 - Could start a modeling effort using corporate data for yield. Can use their data to train the models without while keeping their data private.
 - Joellen has high resolution historic and predicted climate data.
 - JW: One is doing some of this but genotype and exact gps location data were not included.
 - SW: exact location is important for soils.
 - JW: Ag mip person will be visiting
 - JW: What about high throughput phenotyping?
 - SG: Should write something into the renewal.
 - Phytobisque
 - Metabolite profiling Brett Tyler at Oregon State.
 - A person will be hired part time by iPlant and part time by Brett to see if Brett's tools fit will in iPlant's infrastructure.
 - SG also wants to tap into what CIAT in Colombia is developing.
 - JW: What is the status of the IBP portal?
 - Being released at the GCP conference in Beijing today
 - Portal has been translated into Chinese
 - Martha will send around IBP portal URL so people can try it out. OK if it is just this small group within iPlant: <http://www.integratedbreeding.net/>
 - Steve noted the National Academy's "A New Biology for the 21st Century is worth reading. What iPlant is doing is consistent with its projected direction for biology research.
 - http://www.nap.edu/catalog.php?record_id=12764
 - iPlant's renewal options
 - Karen Cone is trying to get us a message about renewal this week. Steve hasn't heard yet, but expects to in the next few weeks.

Security Breach at UA

- SG: Steve Welch found out there was a security breach at UA. Transferring documents in financing. 14 had sensitive information (names and ss numbers), 5 were accessed. 7700 names were accessed. It's vendor info. Steve Welch's info was in it and some other iPlant participants since the way some subcontracts are written causes the to be considered as vendors. UA sent letters to each person impacted and offered free identity theft protection for a year and insurance.
- Steve G is working with UA to cross-check iPlant participant list with the names that are vulnerable.

- The UA letter seemed poorly done but was real.
- The breach occurred in March. And notifications just went out.
 - DS: Although seems slow, time is needed to confirm breach, understand it and get legal authorization to offer the free coverage for identity theft protection.