VA_20091217

iPG2P Visual Analytics Minutes
December 17, 2009; 12pm to 1pm EST

*Present: Eric Lyons, Chris Jordan, Lenny Heath, Bjoern Usadel, Bernice Rogowitz, Ruth Grene, Adam Kubach, Greg Abram, Karla Gendler

Notes/Agenda:

Item 1: Discuss Workflows

Bernice and Ruth presented their workflow strategies:

- Workflow Strategy
- SimpleWorkflow

Discussion followed:

Abram commented that with a more sophisticated demo, more science could be present. Lyons asked if metadate and provenance call all be tracked? Abram answered that one can annotate works as it is happening; there are two types of metadata: where data originally comes from (has to be propagated through entire workflow) but along the way, can add metadata, your observations, provenance coming from vistrails, need to provide very open form of metadata. Jordan added that what is needed is some sore of standardized way to come in to the DE and perhaps some sort of transformation will need to take place internally and thus will need original source and the transformation. Abram suggested that another tool that would be added to Vistrails would have the role of being able to move and transform data to work with other tools and also to keep the provenance of the data. Jordan stated that from the DI group, they will have to figure out how to bring in and convert data for "blue bubbles". There are enough orange bubbles defined to create new workflows.

Abram said that the most fundamental thing to be done is to define the data model, or the structure in which all data can be described to be handed to each operation. The ultimate goal of the whole workflow porcess is to map out the boundaries of what that data model has to be able to support. The data model will constrain what will need to be done in this whole workflow. Jordan asked what are the boundaries and added that DI willh ave to define some boundaries also. It is not possible to deal with every possible way that data comes in but can find ways to work with it. Abram would like to bounce this off the StatInf group to also help define the data model.

Grene asked if CoGe has a way of dealing with groups of genes. Lyons said that it needs higher level of organization to track groups of genes. It can track a list of genes for a given user that they can annotate and share. Jordan said that what is needed is the notion of arbitrary groups that any user could treat as a whole. Abram stated that Vistrails doesn't necessarily have data model underlying it and thus it is up to us to define it. Lyons suggested a user space component: will there be a way to share data? Rogowitz added that this gets to the idea of casual user versus advanced user (canned workflow versus changing workflow). Abram said that it will be necessary to share Vistrails applications. Jordan asked if the data model will then have to incorporate operations?

Abram said that releases will be defined both internally and externally with two goals: 1) infrastructure flexible enough to track into the future; and 2) initial set of capabilities that provides something that is instantly useful (so we will have to wrap some of the very useful tools tout there and will make point-by-point decisions as we move forward). Grene is concerned with the first two points in the build or buy continuum. Rogowitz added that the ideas that percolate out of G2P will be incorporated into iPToL and Abram added that this will have be decided on a much higher level.