

## **FEB 5 – Report on DANA OPTIMIZATION.**

Missing METADATA and possible design choice in Feedback System:

Labels could be more explanatory.

### **EXECUTION VIEW:**

In Variant calling resequencing example of DANA, the input dataset `ref_genomeseq` had a query string rather than a link to workflow explorer, it was linked to wings workflow. The name of the dataset was missing.

#### **Data narratives for result: `filt Data File1496101618223`**

**Data Narrative 1: Execution view** Hide

See more See less i

The Variant Calling\_ Resequencing method was run on dataset `MAF annot.params` (input `varAnnotParams`), dataset `polyphred.params` (input `SNP_params`), dataset `consed.params` (input `assemblyParams`), dataset `canary test.bam` (input `seq_dataFile`), dataset `fetch?data id=http%3A%2F%2Fwww.wings-workflows.org%2Fwings-omics-portal%2Fexport%2Fusers%2Favalli%2Fspellbook%2Fdata%2Flibrary.owl%23hg19.fa` (input `ref_genomeseq`), dataset `polyscan.params` (input `polyscan_params`) and dataset `phred.params` (input `basecaller_params`), with `FILTVALUESCUTOFF1496101618223` set to 0, `JOINBYDIMENSION1496101618223` set to col and `FILTVALUESCOLNAME1496101618223` set to col. The `filt Data File1496101618223` results are stored online.

Text automatically generated by DANA.

Not all the data variable are specified (importance of other ignored variables not known), as a design choice in feedback system, more details on the missing metadata of data variables(blue data variables in workflow).

Information about which stage of execution a particular data variable is set to some value could be specified for better understanding.

### **DATA VIEW:**

For the input data specified, metadata briefly describing the parameters of the dataset could be included for better understanding of the dataset.

(example: metadata could include, size of dataset, number of attributes, data format information, small description of the dataset.)

A sample data from the dataset could be provided as an extra information for data visualization.

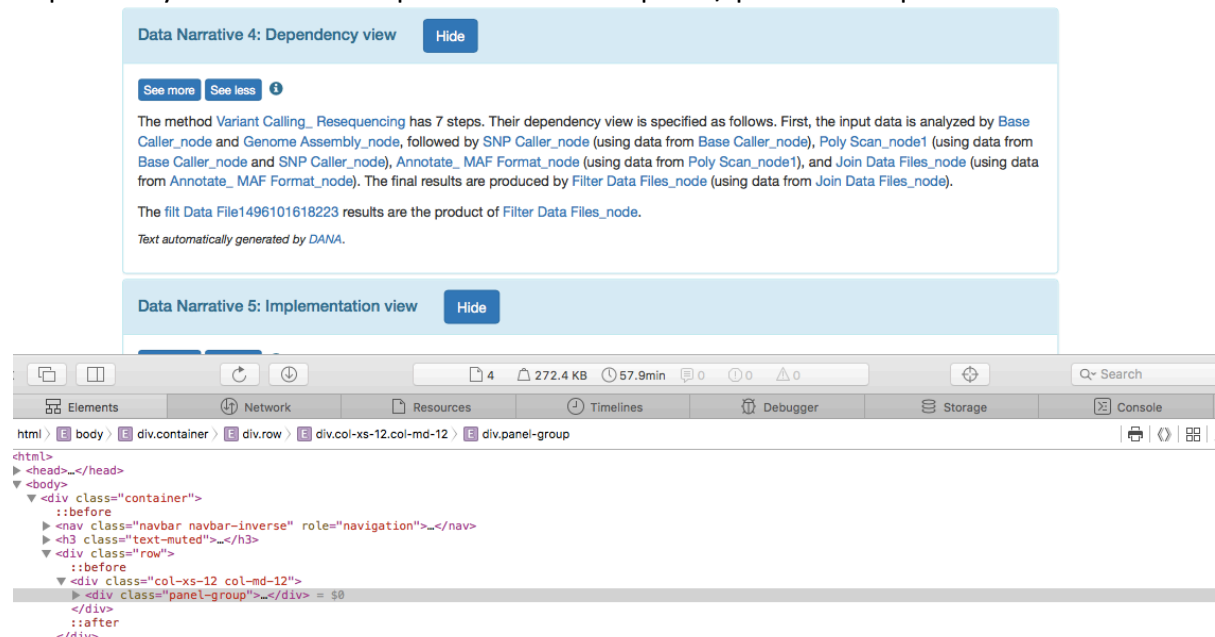
## FUNCTIONALITY VIEW:

The functionality view could be represented as preprocessing step (cleaning of dataset), main/ important functionalities of software and the post-processing step(like merging data).

Functionality view and Dependency view could be possibly merged explaining on a whole, rather than the redundancy. (Possible design option for merging different views.)

## DEPENDENCY VIEW:

Dependency view could be represented as a stepwise/ pointwise explanation.



It says the input data is analyzed by a method, does not specify which dataset, sometimes leads to ambiguity without the workflow.

## SOFTWARE VIEW:

Links to the website of the software could be provided.

Brief description about the software functionality as extra information (Design choice in Feedback system).

Metadata about the software like, how to set it up, hardware and software requirements to set it up, license if any. Such details could be provided in DANA as extra information.