SUMMARY - Optimisation of DANA

The reproducibility, understanding and reuse of the scientific papers for computation by a beginner, novice researcher, developer is hard due to the fact that textual data creates ambiguity which is due to the author ignorance towards detailed description or high level abstraction of concepts, software, data.

To help avoid the process, notebooks and workflows are in use.

Workflows are created to represent experiments that later people use to write papers. The problem right now with such papers is that it's very difficult to obtain the workflow from them, as it is not explicit or not present at all.

Workflows also have another disadvantage where the software implementation is described rather than the conceptual idea of the work done. Problem also arises when focused on the data format/ representation rather than its content in the workflow.

To overcome these problems, workflow is changed to have new workflow abstractions like step abstraction, Data abstraction, criticality abstraction, sub-work flow abstraction.Abstractions could help to improve communication

Now again to understand the workflow, there is a textual description which is manual and leads to problem of understanding the workflow.

DANA (Data Narrative) is the automated text description of the workflow and the scientific paper, which is expressed in six views, each view is obtained by filling out the explanation pattern by querying the data (OPMV format) by SPARQL.

With DANA, it should be accurate and easy to understand the workflow from the produced text.

The idea of the research project is to improve the accuracy and the context of DANA by filling out the important missing data or metadata in the DAta NArrative which is present in scientific paper and the workflow.

Later as the work progresses, to build a feedback system for the users and clients of DANA, where they can provide the accuracy, also tell about the missing information, Data which had to be focused more. These feedbacks can be again used to improvise the performance of DANA and also to automate the process in future by neural networks.

In summary,
a) Create narratives from the workflows described in the K-CAP papers
b) Identify what key metadata is missing and try filling the gaps.
c) Think about a mechanism to save the feedback received from users reusing the narratives produced.