

Integrating GALAXY workflows in a metadata management environment

Francois MOREEWS¹, Yvan LE BRAS², Olivier DAMERON³, Cyril MONJEAUD² and Olivier COLLIN²

¹ Genscale team, IRISA / INRA, Campus de Beaulieu, 35042 Rennes Cedex, France

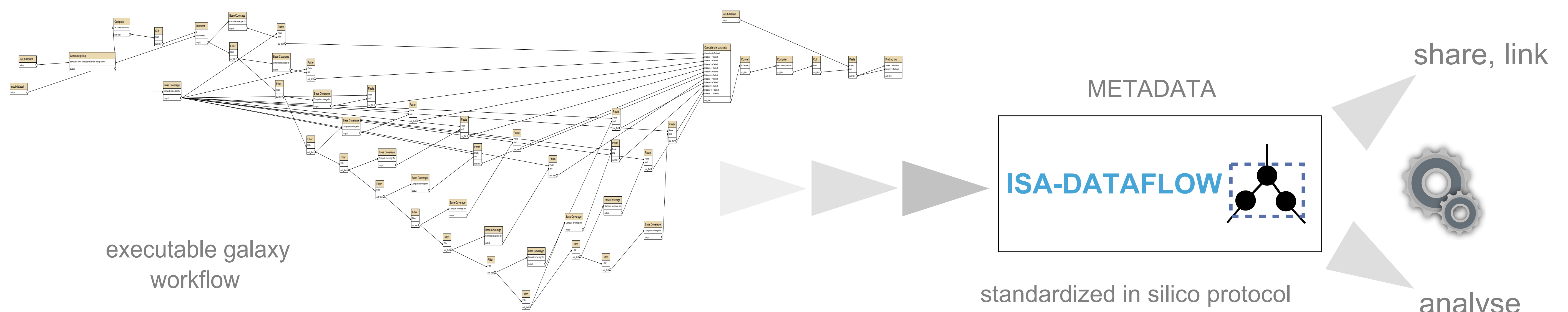
² GenOuest Core Facility, UMR6074 IRISA CNRS/INRIA/Université de Rennes 1, Campus de Beaulieu, 35042, Rennes Cedex, France

³ Dyliss team, IRISA / INRIA Rennes-Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, France

The Galaxy platform offers repositories of user data and related analysis processes (data histories and workflows).

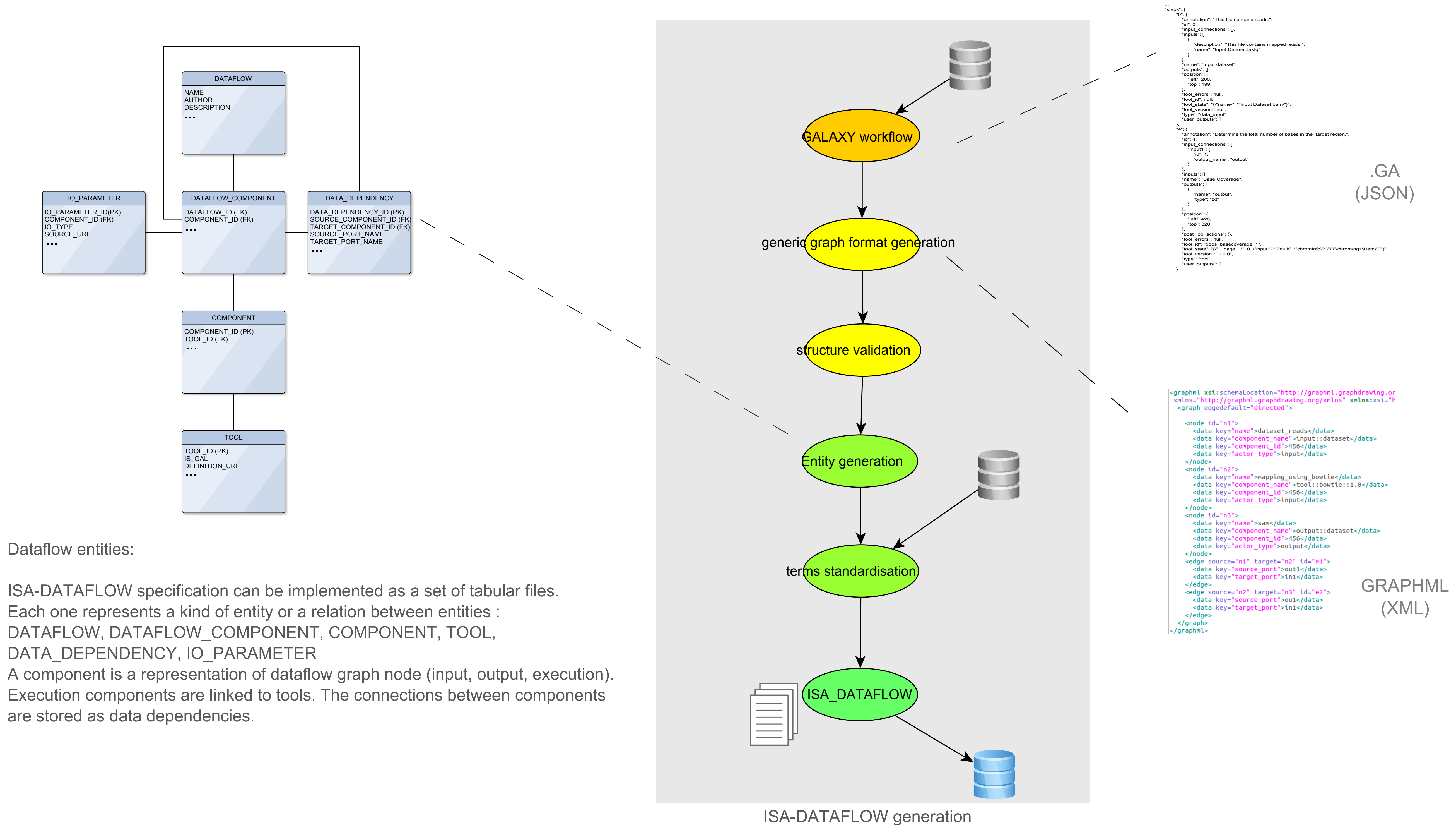
These repertoires enable traceability and reproducibility of the processes within the platform.

At a larger scale, to answer questions like "What protocol was used to analyze my data?" or "how were these data generated?", we could consider any protocol as a metadata set that annotates inputs and results.



We present here a preliminary approach for integrating the GALAXY workflows in an extensible meta-data management environment.

We have developed a formalism to describe an abstraction of data processing workflows. This specification is named ISA-DATAFLOW.



From GALAXY workflow GA format, a conversion tool extracts a structured dataflow representation in GRAPHML. This intermediary format is used to generate the dataflow entities. The entities attributes values can then be normalized using controlled vocabularies. We finally export the entities as a set of files formatted following our ISA-DATAFLOW specification.

We plan to integrate this work to propose advanced research functionalities within a virtual research environment (VRE) deployed on a geographically and thematically distributed infrastructure already using multiple Galaxy instances.

Future developments will concern workflow meta-analysis and workflow composition assistance.