**D4.2: Design, implementation and deployment of Workflow Integrity and Authenticity Maintenance components – Phase II**

**Summary**

This deliverable presents the updated prototypes of the different integrity and authenticity (I&A) components of the project. This is the last of two deliverables regarding the design, implementation, and deployment of Workflow Integrity and Authenticity.

These components use the Research Object resources allocated in RODL (Research Object Digital Library) for evaluating the RO overall quality and providing some meaningful information for a better understanding of its current status. Among other useful information provided by these tools is worth to highlight the importance of collecting the provenance of the different quality dimension scores to provide a historical point of view of these measurements.

We also provide here a summary of the updated implementation of the I&A evaluation tool during phase II. In this second version of the deliverable we have focused on updating the two specific quality dimensions which were used to drive the design and implementation of our I&A evaluation tool, namely completeness and stability, and we also have contributed for the definition of a new dimension so called reliability which uses provenance of quality information for providing a more user oriented and meaningful information regarding the quality of an RO.

Regarding the provenance work, the models wfprov and wfdesc that we developed during the phase I did not need any further updates. Taking advantage of this fact, we have created a big corpus of provenance, so called BigProv, by using the wfprov ontology and the plugins implemented during the Phase I making it accessible to the community for benchmarking purposes.

Our updates on the current design and implementation of these three dimensions can be summarized in our new checklists designs, our new Minim model based on new specified requirements, an update on the evaluation software to use SPARQL1.1, accessing to RODL for retrieving the ROs which want to be evaluated, implementation of the new reliability dimension, and new visualization tools for these three dimensions. We furthermore started the evaluation process which will be also finished before M36 and fully included in the deliverable D4.3. “Final evaluation report of the workflow integrity and authenticity maintenance”. A glimpse of this evaluation process is also presented in this document by showing specific individual completeness’ dimension evaluations and also a simulated scenario based on real parameters for the validation of the reliability dimension.

**1. Introduction (Esteban)**

This document provides a precise description of the software components produced during phase II of Wf4Ever in the context of WP4 (workflow integrity and authenticity maintenance).

To be included:

* How the new implementations improve the previous quality end-user experience.
* How the new dimension reliability improves the previous quality end-user experience.
* How these new dimensions makes use of RODL and other wf4ever platform components for accessing to the ROs and needed functionalities (e.g. wfprov, OAI-ORE aggregates).
* We have focused on the main purposes of reuse and availability.

According to the DoW, this prototype will include the following functionalities: an updated Research Object provenance model that is the basis of the standardisation process in existing international initiatives (this was done for Y2), and extended methods for computing integrity and authenticity (we have improvement of stability and completeness and a new dimension reliability), taking into account different granularities (we use different minim models and different ways of evaluating reliability), and visualisation tools for them (we got new visualizations both for completeness, stability + reliability).

We have identified the following main contributions:

* Due to the advance state of the provenance vocabularies and the fact that they were early available to the project we have been able to create a PROV-Corpus based on Taverna and Wings workflow repositories so called BigProv. The main purpose of this corpus is to provide a suitable number of provenance of workflow results for benchmarking (e.g. extraction of macros, or identification of similar workflows based on their provenance of workflow results).
* Extending the current checklist-based approach for computing the completeness and the stability by: i) including a redefinition of the Minim model, ii) by adding a third dimension (reliability) based on the previous two mentioned, iii) by providing deeper granularity by extending the purpose criteria definition, and iv) storing and providing accessibility to the provenance of the quality results as a resource of the RO.

**2. BigProv: Big Provenance Corpus Generation (Jun)**

BigProv collects 120 real provenance of workflows from the world scientific community. These workflows belongs to two different systems: Taverna and Wings and are associated to 12 different applications domains. The provenance traces have been specified by using the PROV-O ontology. The terms from other vocabularies as RO model and OPMW have been also used for the association between the provenance with their corresponding workflow description.

This dataset has been created for supporting the following scientific community interests and applications::

* Discovery of common “motifs” for annotation of workflows subgraphs by identifying the most frequent in-use patterns. This work can be consulted in the D2.2v2 “Design, implementation and deployment of workflow lifecycle management components– Phase II“.
* Discovery of pattern similarities and linking similar scientific experiments.
* Identification of patterns of use for obtaining dependencies recognition.
* Verification of replicability of previously certified results.
* Include other uses....

This corpus has been generated by automatic capture of provenance from Taverna by using the provenance plug-in[[1]](#footnote-0) which provides PROV-O output format. This plug-in was already implemented in its early stage at M20 and has been improved and tested for the generation of the BigProv corpus[[2]](#footnote-1).

This can be

**3. Evaluating the Quality of a RO (Esteban)**

**3.1 Ontological Framework (Jose?, Jun?, Esteban?)**

This section introduces the models designed and implemented in Wf4Ever which have provided the needed information for the establishment of a quantitative measure of the different dimensions identified as very important for the definition of an overall quality RO criteria. Despite these models have been defined either previously or are part of other documents to be deliver at the same time of this one we want to introduce them briefly here for the creation of a self-contain and understandable document. We also want to highlight the dependencies between the different layers that ranges from a qualitative perspective to a quantitative one.

**3.1.1 Models**

* RO model (for each one it should be highlighted if there is any improvemente vs. the M20 delivearable and also which parts do we use for the evaluation of the quality of the RO)
* AO
* wfprov
* wfdesc
* roevo
* prov-o
* OAI-ORE

**3.1.2 Knowledge generation framework (Jose)**

This subsection explains how based on the current implemented models and the data generated by using this models in the project a quantitative measure of completeness, stability and reliability can be constructed (knowledge generation pyramid ISWC2013 in-use).

**3.2 Calculating Quality of the RO by using Completeness + Stability + Reliability(Esteban)**

**3.2.1 Completeness: definition of the completeness score (Esteban)**

**3.2.2 Stability: definition of the stability score (Esteban)**

**3.2.3 Reliability: definition of the reliability score (Esteban)**

**4. Monitoring the Quality of a RO (Esteban)**

This section shows the toosl used for the monitorization and for the improvement of the user experience by providing to them visualization tools of the above introduced quality dimensions. This monitoring tool is integrated in the Wf4Ever Sandbox and is available at[[3]](#footnote-2). We also explain the implementation done towards the generation of that live demo and the APIs that we have specified for allowing the reuse of the services for any other purposes inside or outside of the project.

**4.1 Implementation of the RO Monitoring Tool (Graham + Aleix)**

**4.1.1 Completeness (Graham)**

It should include the new checklist designs, refactoring of Minim model (ro manager updated[[4]](#footnote-3), and new checklist ontology[[5]](#footnote-4) showcase128), updated evaluation software using SPARQL1.1, new visualizations and checklist services.

**4.1.2 Stability (Aleix)**

it should include the storing data for ROs and how this values are stored or access, how to use this data for evaluating the stability, stability updated service126. Generate RO Monitoring traces for existing ROs.

**In order to get the stability values of a RO we have to check their completeness values periodically over time. Once we have a completeness value for each point in time we are able to calculate the stability trace of the RO. This process is formed by various steps. First of all we have to identify all the existing ROs in RODL. To do so we perform a SPARQL query on the RODL endpoint which allows us to retrieve the URI that identifies every RO and we store them in a list. Iterating over the list of URIs gives us the target of the Checklist Evaluation service. The other parameters needed for the call have been previously defined such as the minim file and the purpose. After getting the results of the checklist evaluation service in JSON format we parse them in order to synthesize the content to the minimum needed for our calculations. We store all the rules together with their “pass” or “not pass” value and their level (must, should or may). Each RO has its own xml file that gathers the summarized evaluations. If the RO is evaluated for the first time we generate the xml file and store the summarized data. However, if the xml file had been created before then we edit it by adding the new evaluation. Most of times the evaluations do not change from day to another we only store the evaluations that are different to their preceding so we save a lot of memory. When the complete trace is requested by the service it will be reconstructed by filling in the empty days by copying the previous stored evaluation. Once we have the full trace of completeness values we can proceed to calculate the stability trace (one stability value each day).**

**4.1.3 Reliability (Aleix)**

It should include the reliability API and the sequence diagram of its execution, accessing to the stability and checklist service. The list of the parameters stored (provenance of RO quality) or how they are accessed through the service and specifying the parameter of the RO.

**5. First Evaluation (Esteban)**

Despite the overall evaluation of the tools implemented in this workpackage will be fully included them in the deliverable D4.3. “Final evaluation report of the workflow integrity and authenticity maintenance” due to M36, we have already started the individual evaluation process of the different dimensions described in this deliverable.

**5.1 Evaluate Checklist Tool (Graham)**

We have done different experiments related with the evaluation of the completeness assessment. More specifically we have evaluated the following scenarios:

* Evaluate Minim w.r.t. Chembox (showcase128)
* Evaluate Minim w.r.t. GSKOS
* Evalutate checklist w.r.t. KEEG

**5.2 Evaluate Stability+Reliability Tool (Esteban)**

Due to we the final global Wf4Ever prototype is being finalized during these months, we still don’t have access to the full functionalities needed for collecting enough historical data on a real scenario which would be needed for the validation of the improvements obtained by using the proposed dimensions. However, until the needed historical data can be collected (a few months would be needed) and for an early evaluation of these dimensions, we have created a simulated scenario which is based on heuristics obtained in previous works (include ref to e-science2012). It is worth to highlight that the results that we obtained are very encouraging and in 120 cases the decision of reusing a workflow was better vs. 41 worse whenever we used the RO Quality monitoring tool vs. using a simple completeness score.

**6. Conclusions (Esteban)**

**7. Publications (ALL)**

* José Manuel Gómez-Pérez, Esteban García-Cuesta, Jun Zhao, Aleix Garrido and José Enrique Ruiz, “How Reliable is Your workflow: Monitoring Decay in Scholarly Publications” Sepublica’2013 Workshop held jointly with the ESWC’2013 (Best Paper Award).
* Belhajjame, Khalid, Jun Zhao, Daniel Garijo, Aleix Garrido, Stian Soiland-Reyes, Pinar Alper, and Oscar Corcho. "A workflow PROV-corpus based on taverna and wings." In Proceedings of the Joint EDBT/ICDT 2013 Workshops, pp. 331-332. ACM, 2013.
* Submitted: José Manuel Gómez-Pérez, Esteban García-Cuesta, Aleix Garrido and José Enrique Ruiz, ISWC2013 in-use track: "When History Matters - Assessing Reliability for the Reuse of Scientific Workflows".

-Include any other publications here

1. <http://wf4ever.github.com/taverna-prov/> [↑](#footnote-ref-0)
2. <http://www.wf4ever-project.org/wiki/display/docs/Provenance+corpus> [↑](#footnote-ref-1)
3. http://sandbox.wf4ever-project.org/decayMonitoring/monitor.html [↑](#footnote-ref-2)
4. https://github.com/wf4ever/ro-manager/tree/develop/Minim [↑](#footnote-ref-3)
5. <https://github.com/wf4ever/ro-manager/blob/develop/Minim/minim-revised.omn> [↑](#footnote-ref-4)