CoRR — REPRODUCIBLE RECORDS CAPTURE AND DISSEMINATION IN THE CLOUD

Faïçal Yannick P. Congo[a,b], Jonathan E. Guyer[a], Carelyn Campbell[a], David Hill[b], Shengyen Li[a]

[a] *Material Science and Engineering Division, Material Measurement Laboratory*

[b] *LIMOS, Blaise Pascal University, France*

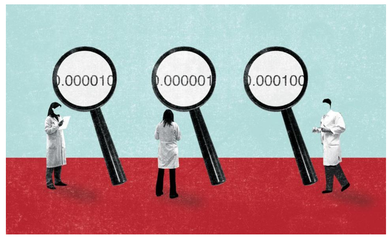
The Problem: Experimental & Computational (Stodden et al 2013)

❖**The untracked changes (libraries, compilers, scheduler, …) in the user session that lead to non-repeatable simulation results.**

❖**The untracked changes of a valve, pump or any other specific sensing or measurement part that lead to non-repeatable experiment results.**

❖**The variety of operating systems, computer architectures and paradigms complicate the repatability and dissemination of the scientific results.**

❖**The multitude of machines, calibration techniques and procedures harden the repeatability, collaboration and verification of the research results claims.**



Current Limitations

Variety of record structure. The tools do not agree on a common way of representing a record which creates an Interoperability drawback.

Record are in textual format: JSON, XML, DOCS. Text based records as mentioned in (Congo 2015) are not enough to ease the path back to controlling the environment.

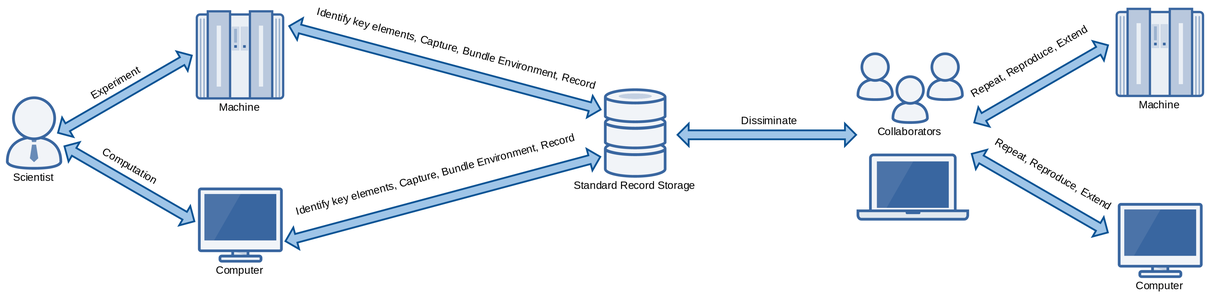
No central space for intuitive dissemination and referencing. Not enough collaboration scheme to enhance reproducibility: Cloud need.

Research Solutions: Record and Share all the investigation's relevant data

❖**Identify the key parameters of the scientific investigation environment: Hardware specs, Operating System, Machine and Calibration specs.**

❖**Identify the key parameters of the investigation itself: Source code, Dependencies, Compiling options (Jézequel et al 2014), Execution parameters, Experimental procedure, the execution settings, Experiment Requirements.**

❖**Identify the investigation input and output: Input, Output data for Computational; Sample specs, Results for Experimental.**



By recording all the identified keys components (cited), one is able to assess the reproducibility of a scientific experiment :

- Different results should point to key elements variations that can be corelated: Non-reproducibility source extraction.
- After Sharing, Enforce investigation parameters before execution to be able to repeat or use previous results foundation.

Existing Approaches

Currently in the computational side, three approaches exist:

- Work-flow based tools: Taverna, Galaxy
- Event Base control tools: Sumatra, Dexy
- Library based tools: Sumatra, OOF GTKLogger

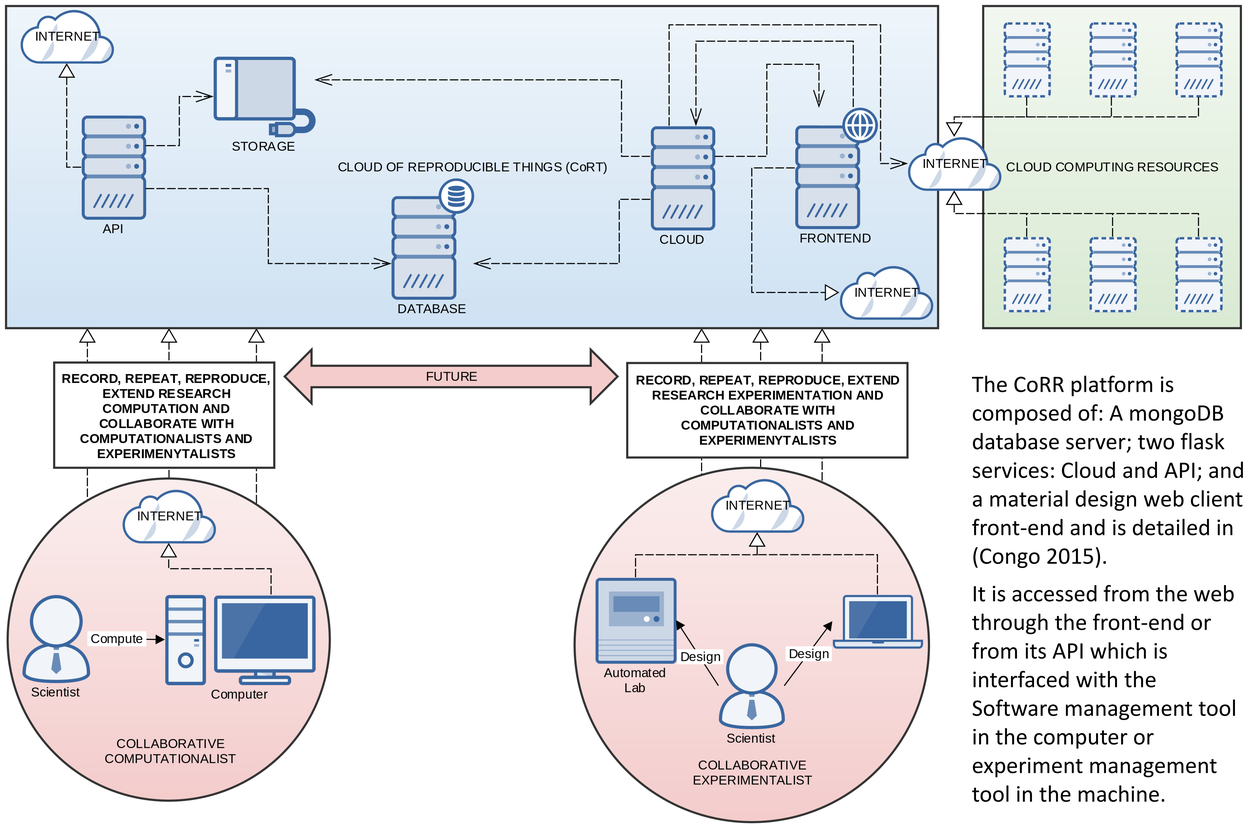
Sumatra dexy

For the experimental side, there are three techniques that can also be used:

- Application Programming Interfaces: Automated labs.
- File system watcher: No API but open files format.
- Custom approaches: No API and proprietary files.

Galaxy

Cloud Platform



The CoRR platform is composed of: A mongoDB database server; two flask services: Cloud and API; and a material design web client front-end and is detailed in (Congo 2015).

It is accessed from the web through the front-end or from its API which is interfaced with the Software management tool in the computer or experiment management tool in the machine.

Conclusion & Perspectives

❖CoRR provides a centralized way for investigation management tools to record the investigation execution.

❖CoRR is a cloud platform built to help scientists collaborate and disseminate their Reproducible Records.

❖CoRR advocates the necessity of experimental investigation to enforce the design of more automated lab machines for full control of the key parameters of the whole experiment.

❖CoRR will be linked to a pool of computing resources to allow direct computing on the cloud. The execution of a computation and the record captured will then all be executed in the cloud and collaborators will be able to repeat, reproduce, extend and run a record in the cloud as well.

Please contact 'Yannick' faical.congo@nist.gov for further information

References

[1] Congo F.Y.P., "Building a Cloud Service for Reproducible Simulation Management", In Proc. Of The 14th Python In Science Conf. (Scipy), 2015, pp. 1-6.

[2] Fomel S. and Claerbout J.F., "Reproducible Research", Computing in Science & Eng., vol. 11, no. 5, 2009, pp. 5–7.

[4] Jézéquel F., Langlois P., and Revol, N. "First steps towards more numerical reproducibility". In ESAIM: Proceedings and Surveys, vol. 45, 2014, pp. 229-238.

[5] Stodden V., Bailey D.H., Borwein J., LeVeque R.J., Rider W. and Stein W., "Setting the default to reproducible: Reproducibility in computational and experimental mathematics", 2013, unpublished ICERM workshop synthesis; https://www.carma.newcastle.edu.au/ jon/icerm12.pdf