Lightning talk: Software Impact Measurement at the Netherlands eScience Center

Willem van Hage, Jason Maassen, Rob van Nieuwpoort Netherlands eScience Center Amsterdam, The Netherlands <u>http://estep.esciencecenter.nl</u> {w.vanhage | j.maassen | r.vannieuwpoort}@esciencecenter.nl

Abstract—In this lightning talk we present a new initiative at the Netherlands eScience Center (NLeSC) that aims to measure software impact, recognizing software as research output.

Index Terms-eScience, software impact, eStep

I. INTRODUCTION

The NLeSC is a Dutch funding agency and scientific software development expertise center that supplies specialized manpower along with research funding for innovative eScience projects to all of research. The main goal of the NLeSC is to enhance science with information technological innovation. This is accomplished by alleviating a number of problems that come with the adaptation of the scientific process as we know it to include new IT. Examples of these issues include missing scientific software tooling; the lack of skilled scientific software developers; missing specific technological expertise such as big data analytics, optimized data handling, and efficient computing; missing publication for afor scientific software; and the lack academic professional assessment by scientific software impact. Of all these urgent problems, the last might be the most important. If academics are assessed by their software, then the academic world could eventually reorganize itself so that all the other issues are solved.

II. ESTEP: THE ESCIENCE TECHNOLOGY PLATFORM

To stimulate scientific software development, the NLeSC has set up the eScience Technology Platform (eStep), which consists of three parts: a software catalogue of eScience software, interfaces, libraries, tools workflows, applications, models and data sets; a knowledge base consisting of guides, reports and recommendations for scientific software development; and documents describing eScience research such as scientific publications and demos.

Currently, eStep is used by the NLeSC itself for internal dissemination of knowledge and for self-evaluation, by NLeSC partners like the funded research institutes to learn about the portfolio of the NLeSC and new trends in eScience, and by other users of the eScience software produced by the NLeSC to find out how to contribute or whom to contact for more information.

To help all of these users to access the information they need NLeSC launched a new Web site that presents all the interrelated topics in eStep, such as software packages, people that were involved in the development and use of these software packages, projects in which they were used, and organizations that participated in these projects. Users can browse, search, and filter the website. The analytics site that allows visitors to filter the contents of eStep is illustrated by the screenshot shown in Figure 1. An example software project page is shown in Figure 2.

The website is automatically generated from metadata records that can be submitted added to the website by the developers themselves, putting the responsibility at the scientific developers themselves. Currently, the only contributors to the website are NLeSC employees, but in the near future the website will become a shared effort by an increasing number of eScience research institutes. *Increased visibility and software impact measurements are incentives for partners to join.* We foresee that the infrastructure will become more federated in the future. We will make the impact measurement technology available to other software catalogs as well.

III. SOFTWARE IMPACT TESTBED

This year, the NLeSC will start with an experimental addition to the eStep website to aid scientific developers with the assessment of their software impact. The goal is that each software package described on the website, and added by the developers themselves, will get an automatically generated dashboard presenting various aspects of the software impact.

There are many different approaches to measure software impact based on different measurable quantities: direct and indirect software (report) citations, software dependency structure, software contribution, forking of the code base, usage in projects, endorsements or reviews, and software quality metrics.

The goal of the software impact measurement website is to serve as a test bed to assess the usefulness and reliability of all of these different approaches in practice. The NLeSC hopes to learn which aspects of software projects are best used to measure software impact, and to use this empirical knowledge to influence scientific organizations to start assessing researchers by these measures. Hopefully this will lead to an increased awareness at both funding agencies and research institutes that software drives innovative research, that better software means better research, and that appropriate key performance indicators can be used to improve the scientific performance.

Software			
Click on the bars to find software projects.	Name	Description	
53 selected out of 53 records Reset All	3D-e-Chem Virtual machine	Virtual machine with all software and sample data to run 3D-e- Chem Knime workflows	
Disciplines	AHN2 pointcloud	WebGL point cloud visualization of AHN2	
eScience Methodology	viewer		
Environment & Sustainability	AMUSE	The Astrophysical Multipurpose Simulation Environment	
Humanities & Social Sciences	0.01		
Physics & Beyond	CClusTera	A 3D web tool for interactive visualization of hierarchically clustered big data	
Life Sciences & eHealth			
	Cesium-ncWMS	3D Globe Visualization of NetCDF data.	
Competence areas	Chemical Analytics	Packer template to create Vagrant box with Knime inside	
Big Data Analytics	Virtual Machine		
Optimized Data Handling	Common Sense	User-friendly web application for showing (GIS) data on a map.	
Efficient Computing	Common Sense	Oser-mendiy web application for showing (GIS) data on a map.	
	Cross-perspective	A Gibbs sampler that implements Cross-Perspective Topic	
Technical expertises	Topic Modeling	Modeling	
Scientific Visualization	DataVaults	Technology of Attachment to a DBMS of large file repositories.	
Distributed Computing			
Information Visualization	Differential Evolution	Differential Evolution global optimization algorithm, with Metropolis for uncertainty estimation	
Databases	Lyonadon	or uncertainty estimation	
Text Mining	eAstroViz	This tool can convert and visualize radio astronomy measurement	
Handling Sensor Data		sets, as well as most LOFAR intermediate data producs. It also does RFI mitigation.	
High Performance Computing		does KFI mugation.	
Information Retrieval	eEcology Annotation Tool	Visualize & annotate GPS measurements of bird movements	

Figure 1: The eStep software project analytics webpage (screenshot August 2016).

eScience Technolog	eScience Technology Platform	
MAGMa		
Links	Expertise	Contact Person
Website, Source code	High Performance Computing	Lars Ridder
Start Date	Scientific Visualization	Owner(s)
2011-05-10		Netherlands eScience Center, Wageningen University and Research Centre (WUR)
Status	Programming Language	
active	Python JavaScript	Contributors Lars Ridder, Stefan Verhoeven, Marijn
		Sanders
Discipline	License	Involved Organizations
Life Sciences & eHealth	Apache 2.0	Netherlands eScience Center, Wageningen
Competence	Tags	University and Research Centre (WUR)
Big Data Analytics	Distributed Webservice	Users
		Lars Ridder, Marijn Sanders
Efficient Computing	Dependencies	Used in Projects
	Osmium	eMetabolomics

Description

MAGMa is an online application for the automatic chemical annotation of accurate multistage MSn spectral data.

MSn data can be uploaded as a hierarchical tree of fragment peaks, either based on m/z values or elemental formulas, or as an mzXML file of the raw data., Candidate molecules are automatically retrieved from PubChem, from a subset of PubChem compounds present in Kegg, or from the Human Metabolome Database., Candidate molecules can be predicted based on in silico reaction rules describing microbiotic and human biotransformations, For each candidate molecule, substructures are generated and matched with the observed fragment peaks., The web browser enables efficient mining of the automatically annotated data., Open Source, source code available at

Figure 2: An example software project description. The screenshot shows the metadata of the MAGMa chemistry tool and a small part of the description of the tool. (screenshot August 2016).