EPIC
Earth and Planetary Innovation Challenge
http://www.elsevier-epic.com/ - submission draft,

Edzer Pebesma

October 2, 2013

Overview of questions you will need to answer to tell us about your innovation idea

1 Opening questions

Title: One-Click-Reproduce.

Description (50 words max):

Researchers face an increasing need to share the input data, data created, and analysis steps along with published papers, in order to allow readers to reproduce their analysis. This submission explains how Elsevier journals can enable this, with focus on the R environment (http://www.r-project.org/).

2 Application Type

What does your application support?

- New content types (first)
- Data repository linking (second)
- Reader use/understanding of content that already exists on ScienceDirect
- **Other (please specify): Enabling reproducible research**

3 Innovation/Vision

How will your application support Geoscience research and dissemination? (Max 1000 characters)

Reproducibility is an important aspect of geoscientific research, because the credibility of science is at stake when research is not reproducible. A mature and growing community relies on the R software environment for carrying out geoscientific research, and numerous R extension packages have been published for geoscientific analysis. Geoscientific data often have complex structures (variety in reference systems for space and time, high dimensionality, complex phenomena need be represented by appropriate data structures), and concensus on data file formats is lacking. R Data files can represent data of arbitrary complexity in a direct-to-use form. To reproduce the work presented in scientific publications, the open source R environment only requires R Data files and R analysis scripts. The One-Click-Reproduce button makes reproducing research simple.

What issue(s) will your application solve? (Max 1000 characters)

Most papers describe analysis procedures but do not allow readers to reproduce the results (numbers, tables, figures) presented exactly the way the researchers did this. Data repositories such as PANGAEA encourage users to publish data in simple form (ascii, table), which makes it time-consuming to import and analyse – analysis scripts or software are usually not posted. By publishing data and procedures in a simple-to-reproduce form alongside the paper, readers are more motivated to carry out reproduction, and are more inclined to adopt a similar approach and/or cite the paper. Besides transparency, increasing citation is an incentive for researchers to provide reproducibility.

How will the application solve this issue(s)? (Max 1000 characters)

The One-Click-Reproduce application enables readers of the paper to reproduce the analysis done in the paper by a single mouse click, and see the results, tables and figures being generated. In addition, readers get access to the R Data file and R script needed for the reproduction. Initial output as generated by the authors documents the software versions used, permitting differences arising as the underlying software is updated to be highlighted. A solution that requires no software installation from the reader runs the reproduction on the server side or in a cloud, and returns an html document. Readers that have R installed can opt for reproducing on their own computer, making it easier to study and modify the analysis and data, and checking the robustness of the results. The application contains a link to a document explaining how all this works. Author instructions explain researchers how to write readable scripts that work on different operating systems.

Is your application a new innovation or is it an improvement on a previous innovation?

- **New innovation**
- Improvement on a previous innovation, please describe

4 Application Design

Describe your ideas for the visual layout of the application. (Max 1000 characters)

At the right hand side of the paper web site, a box is added with a button called "Click-To-Reproduce", which visually hints at the R logo (http://www.r-project.org/Rlogo.jpg). Clicking this button gives access to the options: "One-Click-Reproduce" reproduces the analysis in the cloud and returns an html page (see e.g. http://rpubs.com/edzer/ for examples), "Reproduce locally" gives access to the R Data file and R script that allow reproduction on a local computer (Windows, Mac, Linux, other). For those unfamiliar with R, a link is added to a document explaining how remote and local reproduction work. A link for authors explains how R Data files and R Scripts are created cleanly.

Is the application intuitive and easy to use i.e. requiring no learning curve for the user? Please explain your reasoning. (Max 1000 characters)

The procedure is obvious and extremely simple. It is fully clear and transparent for those familiar with R. Those unfamiliar with R who want to reproduce locally need instructions for downloading and installing R and RStudio, and instruction how to install add-on R packages, but all this is very easy. For those unfamiliar with R, the learning curve to understand the script is similar to that of other scripting environments that allow publishing reproducible research (e.g. matlab, octave, python, grass GIS). Excellent documentation for this exists in the form of on-line help, manuals, tutorials, and books.

5 Anything Else

Please describe any further details relating to your application idea. (Max 1000 characters)

Providing reproducibility intensifies scientific communication, citation, and speeds up innovation, re-use and further development of ideas. This application can be reused for many other areas of science where data are analysed and R is being used, as well as serve as a template for other computing environments that allow research to be reproduced. When the full input data set is too large or cannot be disclosed, the analysis procedure can use a small excerpt of the total data set, or a modified/anonimized version of the data. R scripts may remotely access large geoscientific data repositories through data services or data bases such as SciDB. I offer help to develop author instructions for submitting data and scripts, indicating for instance how to write readable scripts that work on all operating systems, and how to avoid modifying a user's computer. Licenses for data and scripts need to allow reproduction, modification and redistribution (ODC-ODbL, CC BY 2.0).