Some observations based on the case studies of research tools

Research tools as accelerators for science

Authors:

Maurits van der Graaf; orcid.org/ 0000-0002-2296-7568; m.vdgraaf@pleiade.nl

Leo Waaijers; orcid.org/0000-0003-1433-2543; leowaa@xs4all.nl

August 2014

This work is made available under a Creative Commons attribution 4.0 licence

For details see: http://creativecommons.org/licenses/by/4.0/

Introduction

Research tools that are freely available and accessible via the Internet cover an emergent field in the worldwide research infrastructure. Clearly, research tools have increasing value for researchers in their research activities. How are these research tools developed? What are their possibilities? How many researchers use them? In addition, what does this relatively new phenomenon mean for the research infrastructure?

Knowledge Exchange and its partners express a common vision based on its five national strategies: 'To make a layer of scholarly and scientific content openly available on the Internet.' One of the focal points of Knowledge Exchange is to stimulate the development, usage and funding of research tools that exploit this open layer of scholarly and scientific content. For this purpose, KE and its partner organisations explore developments in research tool and identify challenges in setting up such tools. The Knowledge Exchange partners are:

- CSC-IT Centre for Science in Finland
- Denmark's electronic research library (DEFF) in Denmark
- German research foundation (DFG) in Germany
- Jisc in the United Kingdom
- SURF in the Netherlands

Knowledge Exchange recently commissioned a project for 11 case studies of research tools in order to show their potential and their relevance for the present research landscape. For this project, we interviewed the makers of those research tools. In this paper, we like to offer some observations with the eye on possible improvements of the research infrastructure.

Nowadays, research funders, e-infrastructure providers, universities, research institutes, learned societies and others invest significant funds in developing tools to support research. Without mature support systems in place for research tools, it is all the more important to share learning experiences and to show effective approaches. With these notes on the case studies, we hope to contribute to the developments of infrastructures for research tools.

The limitless potential of research tools as showed by ten case studies

We are not aware of any estimate on the number of presently available research tools on the Internet. However, there must be a lot: the Finnish IT Centre for Science hosts more than 100 tools, while the Bamboo DiRT wiki lists over 600. The limitless potential of research tools, however, is already showed by the 11 research tools that participated in the Knowledge Exchange case study project:

- LARM: this project opened up more than 1 million hours of radio broadcasting in Denmark for researchers, who responded by using the broadcasts for an amazing range of creative research lines: from the development of street sounds over time in Copenhagen to a map of dialects in Denmark and many more.
- MERDES, CoMerDa, QuaMeRDES and Trove: four tools to support the exploration and contextualization phase of the media research cycle. An important aspect of these tools was that the makers also investigated the effect on the media researchers themselves. This showed that media researchers using these tools explored more materials before moving to the next phase of their research.

- SOMA: a tool for biomolecular scientists to facilitate the supercomputing calculations on three-dimensional molecular reactions. The most important outcome is an increased efficiency for its users.
- FuD2015: FuD is a combined working, publishing and archiving environment for researchers in the humanities. Its main aspect is that it facilitates collaboration on primary data and provides numerous analysis tools: tools for text analysis, text interpretation, text annotation, semantic analysis, etc.
- CMME: Renaissance music from the 13th century until about 1600 was written in a different notational system than modern music. The unique CMME tool provides transcription and editing options for this music, so that musicologists can see in one view the original notation and the edited notation(s).
- GWAP: Artemis, the art collection of 40.000 images of the Institute of Art History of the Munich University, lacked metadata. The concept of Games With A Purpose, competitive games in which two players attempt to assign identical labels to an image, was successfully applied. The approach was extended to other image collections and to Italian expressions (social and regional tagging).
- Elmer: Solving partial differential equations by numerical simulations has long been a matter of specialized software applications. Elmer has developed these applications in user friendly 'mass' tools, firstly for fluid mechanics, later for other areas as electromagnetics, heat transfer, acoustics and even forecasting of glacier behaviour. Elmer is now one of the most popular multi-physics simulation software. It is open source code and gratis.
- DARE: Philosopher Averroes wrote in Arabic. Over the centuries his works were translated in Latin and Hebrew, original texts were lost and numerous annotated editions saw the light. DARE provides researchers with structured access to the complete oeuvre. DARE technology will also be applied in similar projects on Avicenna and the medieval book Schedula diversarum artium.
- Salzwiki: brings together and makes accessible the spread results of studies into the deteriorating effects of salt on buildings and statues in two wiki's (English and German) and a repository for research data and images.
- Chipster: Biogenetics is a data driven science. Chipster frees scientists in this field from the need to gain considerable computer skills by developing readily applicable open source tools. Firstly, for the field of micro arrays and later complemented for next generation sequencing. Chipster is being used now in many research projects, both in Finland and abroad.
- COSMOS project: Social media such as Twitter or Facebook offer a completely new window on community life and thus exciting opportunities for sociological research: it makes it possible to study society's reaction (almost) real-time, will give a much richer picture that traditional statistics and can offer insight into phenomena that might remain below the threshold of official datasets. The COSMOS project is developing a tool for researchers to analyse datasets from Twitter. The aim is to expand the tool to other social media such as Facebook in a later stage.

Four benefits of research tools

These examples of research tools show clearly the value of research tools for the research process:

1. **Efficiency:** some tools increase the efficiency of research process - the researcher can focus on the research (instead of data conversion, data formatting or even programming, etc.).
2. **Quality:** some tools increase the quality of the research process – for example because the researcher has more options for analysis and comparison of different sources.
3. **Collaboration:** some tools facilitate collaboration and inter-operability between researchers, often at different places and of different disciplines.
4. **New research approaches:** some tools open up completely new research approaches, bring together research disciplines or inspire serendipitous applications.

Four issues with research tools

However, we also observed in our interviews a number of areas for special attention:

Awareness and findability

Most research tools are developed by researchers or in collaboration with researchers for a specific research project. However, most researchers focus on research results and not on research tools. After the research project for which the research tool was developed, researchers generally move on to the next research project. The toolmakers themselves continue using the tool but in many cases will not have time and capacity to find a broader audience for their tool as this is simply not their mission and their task. This is why a number of research tools in the case studies appear to be underused by researchers in the field. How to make other researchers aware of the existence of this research tool? This puts awareness and findability of research tools firmly on the agenda.

The Scientist's User interface (SUI)

The Finnish IT Center for Science provides modelling, computing, and information services for the Finnish research community on a wide range of disciplines: from Biosciences, Physics and Chemistry to Earth science and Language research. In all more than 100 software tools and databases, in combination with supercomputing services and data storage services. All these services are available via a portal. Nevertheless, that leaves the question: how to lower the threshold to make use of these facilities? The solution is the scientist user interface (SUI).

SUI has been launched in a beta version in 2010 and from 2011 onwards in the official version. It is estimated that between 1000 and 1500 scientists make nowadays use of the interface. What does it do for them? It facilitates the use of the services for them:

- After registration, an account is opened with a user profile that can be maintained by the user himself
- Rights are assigned to each user account, defining the access rights of this particular user to the various services
- SUI provides a graphical interface to all services: a common look and feel and easy access to very different services.
- SUI also creates visibility for services that a particular user might not be able to access. If the user does want access, the interface provides a form to request it. Examples are requesting computer time on the supercomputer.

CSC actively works on the further development of SUI. The ambition is to create 'a one-stop click service' for all applications and services within the CSC research infrastructure. This will mean that users will be able to apply and control the access to different applications - using the numerous applications, handling batch jobs for the supercomputer, handling data transfers etc. The development team of four people also aims to make SUI suitable for usage via mobile apps. By this, SUI will remove many thresholds of using applications with different interfaces and make life for scientists easier.

www.csc.fi/english/research

Accessibility

Generally, researchers will use a number of research tools during a research project. This will mean that a researcher has to understand the application of the tools and has to learn to work with the interface. In addition, the various research tools might need different data formats as input or have different data formats as output. In other words, the threshold in using a number of research tools becomes rather high and possible need for conversion of data formats requires a technical knowledge and handiness of the researcher that is not directly relevant for the research process itself. Thus, (the lack of) accessibility might form an important threshold for many researchers. In the text box, the approach of the Finnish IT Centre for Science to increase accessibility is highlighted with the development of their SUI interface. This interface aims to serve all research tools of the CSC centre.

Modularity and interoperability

Almost all makers of research tools, who we interviewed, mentioned modularity and interoperability when asked for lessons learned. If they could restart the entire process of building the research tool, they would focus more on a modular set-up for their research tool. This would make it easier to keep the software up to date with the eye on technological developments: one could start with one module at a time. In the same breath, the interviewees mentioned that they would a next time look more carefully into interoperability issues with related research tools. The makers of the FuD tool mentioned a nice example of the advantages of such interoperability: the FuD tool facilitated a semantic network of over 10,000 newspaper articles on the particular topic. However, for the visualisation of this semantic network another, existing tool could be used (GEPHI).

Sustainability

Of all the issues mentioned by the interviewees, clearly the most pressing issue is the sustainability of research tools. We have heard the following examples:

- A success story: the MeRDES tool for audio-visual archives is taken over by the Netherlands Institute for Sound and Vision who will further develop this tool for its academic user group.
- A 'depending-on-project-subsidies' situation: the development of the CMME music-editing tool is completely dependent on project subsidies. Several times CMME received funding and its development could make a growth spurt with that money. Between project subsidies, the research tool can be kept alive, as part of the university infrastructure, but significant further development is not possible.
- A business model has to be developed: the FuD tool is part of the infrastructure of the University of Trier but its main project subsidies will end in 2014. The supporting team is now developing a business model, including a financial and organisational model.
- Chipster, Elmer and SOMA2, three tools developed by the Finnish Centre for Science (CSC) all apply the same business model. They were developed by public financing in response to a broadly felt need in the Finnish academic research community. Once operational, CSC guarantees a small amount for

maintenance of the tools. New developments however have to take place under the umbrella of greater (European) research projects. The tools are freely available and circulated as Open Source software.

Three key points in developing an infrastructure for research tools

Knowledge Exchange has already taken a number of steps to stimulate the thinking on and actions taken in developing infrastructures for research tools[i]. Based on the case studies, we would like to highlight three key points in the further development of infrastructures for research tools:

1. **Support staff:** although the development of a research tool is in many cases carried out by researchers themselves, making the research tool available for a wider audience of researchers and the maintenance of the research tool with the eye on software updates and other technological developments will generally need to be carried out by (specialised) support staff.
2. **Research infrastructures:** Research infrastructures are being set up to collect, maintain and develop research tools. There is a wide range of possibilities for these infrastructures: from a wiki, that simply lists research tools (Bamboo DiRT) to a specialised institute for research services (Finnish IT Centre for Science) or a specialised portal on the specific area of research (CLARIN). The last one is an example of the ESFRI initiative by the European commission with the aim to build a wide diversity of (specialised) research infrastructures. However, the case report demonstrates that not all research tools might fit in these research infrastructures or that they are part of an institutional infrastructure. Clearly, *one size will not fit all* might be especially true for infrastructures for research tools.
3. **Recognition of research tools**: most research tools in our case studies could show a list of publications that were facilitated by their use. However, these publication lists were not necessarily complete and depended completely on the willingness of their authors to mention the use of the research tool. When research infrastructures host a growing number of research tools and when research funds are increasingly used to support and maintain these tools, the question of how to select and deselect research tools for further maintenance and development will become crucial. We think that impact of a tool may be an important criterion for this. For this reason, we propose a further discussion within the research community on possible methods to measure this reliably. Identifiers for research tools could be the starting point. These identifiers would then become reference points for citations, downloads and other measures ('altmetrics'). On this structure, metrics can be based, indicating the impact of a tool.

Research tools available via the Internet are a relatively new phenomenon, but with exciting possibilities. During the interviews, we were struck by the limitless potential of research tools, making the research process more effective and efficient and bringing together different disciplines by opening up completely new avenues for new types of research. This potential calls for an adaption of the present research infrastructures and thus for actions by the various actors that

are responsible for the research infrastructure. We hope that the case studies and these notes on the case studies will contribute to this.

Acknowledgements

The authors like to thank the interviewees for their time and patience while explaining us the ins and outs of their work, Jan van Odijk (director CLARIN-NL) for giving his valuable insights during an extensive interview and Tapani Kinnunen (product manager at CSC) for his enthusing explanation of the SUI interface.

---

[i] - Virtual Research environments – the next steps; Knowledge Exchange workshop; 2010;
- Exploring research tools; Knowledge Exchange meeting 2012