The Past, Present and Future of the ELKH Cloud

This review article summarizes the history the Hungarian Scientific Cloud Infrastructure project. This research infrastructure was launched officially on 1 October 2016, funded by the Hungarian Academy of Sciences. With the support of ELKH, the infrastructure's capacity has been substantially boosted; the features and workflows that it offers to scientists were significantly expanded to celebrate the arrival of the year 2022. The article reviews the types of work Hungarian researchers implemented on the infrastructure, thereby providing an overview of the state of cloud-computing enabled science in Hungary.

Keywords: *IT History, IaaS Cloud, MTA Cloud, ELKH Cloud*

Author Information

Mihály Héder, Department of Network Security and Internet Technologies, SZTAKI
https://orcid.org/0000-0002-9979-9101

Ernő Rigó, Department of Network Security and Internet Technologies, SZTAKI
https://orcid.org/0000-0003-1044-7167

Dorottya Medgyesi, Department of Network Security and Internet Technologies, SZTAKI
https://www.sztaki.hu/megyesi-dora

Róbert Lovas, Laboratory of Parallel and Distributed Systems, SZTAKI
https://orcid.org/0000-0001-9409-2855

Szabolcs Tenczer, Department of Network Security and Internet Technologies, SZTAKI
https://orcid.org/0000-0002-7646-1116

Ferenc Török, Datacenter, Wigner Research Centre for Physics,
https://orcid.org/0000-0003-3810-6698

Attila Farkas, Laboratory of Parallel and Distributed Systems, SZTAKI
https://orcid.org/0000-0003-0579-3516

Márk Emődi, Laboratory of Parallel and Distributed Systems, SZTAKI
https://orcid.org/0000-0002-8787-7993

József Kadlecsik, Computer Networking Center, ELKH Wigner Research Centre for Physics
https://wigner.hu/infopages/kadlecsik.jozsef

György Mező, Wigner Research Centre for Physics
https://orcid.org/0000-0002-0686-7479

Ádám Pintér, Datacenter, Wigner Research Centre for Physics
https://orcid.org/0000-0002-6071-2822

Péter Kacsuk, Laboratory of Parallel and Distributed Systems, SZTAKI
https://orcid.org/0000-0001-6926-6622

How to cite this article:

Héder, Mihály, et al. "The Past, Present and Future of the ELKH Clouds".
Információs Társadalom XXII, no. 2 (2022): 128–137.
https://dx.doi.org/10.22503/inftars.XXII.2022.2.8

Introduction

This review article enumerates some of the achievements of the Hungarian Scientific Cloud Infrastructure project, which was formerly called MTA (Hungarian Academy of Sciences) Cloud but has now been rebranded to ELKH (Eötvös Lóránd Research Network) Cloud. This research infrastructure was launched officially on 1 October 2016, funded by the Hungarian Academy of Sciences. With the support of ELKH, the infrastructure's capacity has been substantially boosted; the features and workflows that it offers to scientists were significantly expanded to celebrate the arrival of the year 2022.

In their seminal paper, Wilson et al. (2014) explored the landscape of scientific computing and its relationship to science in general. By that time it was evident to everybody that no field of scientific investigation can progress without the use of computers. Further, other than in a few fields, reliance on computers goes way beyond office functions: they must be viewed as scientific instruments and approached accordingly. Already by the end of the first decade of this century a survey showed that a significant number of scientists were implementing scientific software and systems, 90% without prior training for such tasks (Hannay et al. 2009; Prabhu et al. 2011).

As on several occasions before in the history of science, it became clear that there needs to be a division of labour when making these tools. While Galileo was revered for making his own telescopes, his successors in astronomy needed to realise very quickly that this practice is not simply unmaintainable but, except for during the field's initial revolutionary period, presents as an outright disadvantage against their peers in the competition for knowledge. When it comes to the computing infrastructure of science, this change arrived even more rapidly.

Due to the complexity of scientific software and the computational hunger of the infrastructure running it, specialisation of work and consolidation of resources are inevitable. This happened sooner in fields where the needs were most intensive, such as twenty-first-century physics; then, with the applicability of artificial intelligence (AI), it spread to virtually all fields, even those previously thought to be rather untouched, such as philosophy and other fields of the humanities.

It is then crucial for any nation or political or commercial organisation to create the infrastructure of scientific computation in order to remain scientifically competitive; being competitive in science is one of the values accepted virtually everywhere on the planet.

History of the ELKH Cloud

The inevitable consolidation of scientific computing was acknowledged on all institutional levels from individual universities and research centres to large political entities like the EU and beyond. On the national level, in the mid-2010s in Hungary it was natural that the burden of responsibility for establishing this infrastructure fell on the national research network. The organisation not only oversaw the most important collection of research institutes but also involved itself to some degree in university research through supported research groups (TK – 'Támogatott Kutatócsoportok') as well as through issuing Doctor of Science (DSc) titles to senior scientists, often required for full professor positions at Hungarian universities.

So, the work began to establish the research infrastructure, but, of course, it did not come about without prior know-how. Two institutes – SZTAKI (the Institute for Computer Science and Control) and the Wigner Research Centre for Physics – had already built their own institutional cloud infrastructures, so they had acquired the necessary expertise. It was this knowledge capital, together with generous funding, that made the MTA Cloud possible.

There are several types of cloud that could have been built – Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) or Software-as-a-Service (SaaS). This series starts at the most flexible offering and ends with the most ready-to-use one. An IaaS cloud offers virtual machines access to CPU (central processing unit) and GPU (graphics processing unit) power, complete with the necessary virtualised network and storage, on top of which scientists can build their own platform without the need to interact directly with hardware; this saves all the concern over server room maintenance, electrical supply and so on.

However, the logic of increasing the complexity of scientific computing made it necessary for the infrastructure to be maintained by a specialist team, and not the scientists themselves; increasingly, too, this does not stop at the infrastructure level. As scientific computing platforms are becoming more and more elaborate, maintenance is becoming a profession of its own. As certain tools that implement the workflows of a scientific field reach a certain maturity and share of uptake by the scientist end-users – that is, they become quasi-standard tools of a field – the incentives become aligned for offering them as a ready-made platform (PaaS), so as to save even more effort for the scientists. The ultimate endpoint of this evolution, just like in any other domain of software, is SaaS, where the offering needs only to be configured and loaded with data.

The MTA Cloud provided an IaaS for its users, but this may serve as a basis for possible extensions to PaaS or SaaS. It was an inevitable choice as, technically, an IaaS is a perfect starting point for building PaaS solutions; in turn, a PaaS is required for SaaS. But a shift towards PaaS and SaaS is also inevitable for the aforementioned reasons.

Evolution of the user base over time

Natural science was very well served by the MTA Cloud in its first five years of existence. The cloud's user base was initially provided by the two founding institutions, SZTAKI and the Wigner Research Centre for Physics. This soon grew to include the partners of these institutes, both from Hungary and from abroad. After it became clear that interest in this system was increasing, the entire Hungarian research institute network was onboarded. Currently, the user base is being expanded further, towards small and medium-sized enterprises (SMEs) and the government. Universities also sometimes use this cloud as partners of core institutes, although it needs to be said that for educational institutions there is another cloud available in Hungary.

Life science (broadly construed) has been represented over the past five years by users from the Institute of Evolution (Evolúciótudományi Intézet), the Institute of Ecology and Botany (Ökológiai és Botanikai Intézet), the Balaton Limnological Research Institute (Balatoni Limnológiai Intézet), the Danube Research Institute (Duna-kutató Intézet), the Institute for Soil Sciences (Talajtani és Agrokémiai Intézet), the Biological Research Centre, Szeged (Szegedi Biológiai Kutatóközpont), the Institute of Organic Chemistry (Szerves Kémiai Intézet), the Agricultural Institute (Mezőgazdasági Intézet), the Institute of Enzymology (Enzimológiai Intézet) and the Institute of Experimental Medicine (Kísérleti Orvostudományi Kutatóintézet).

The sciences dealing with inanimate nature were on a par in terms of usage with the life sciences, represented by researchers from Konkoly Thege Miklós Astronomical Institute (Konkoly Thege Miklós Csillagászati Intézet), the Research Centre for Astronomy and Earth Sciences (Csillagászati és Földtudományi Kutatóközpont), the Institute for Geological and Geochemical Research (Földtani és Geokémiai Intézet), the Institute for Particle and Nuclear Physics (Részecske- és Magfizikai Intézet), the Institute for Solid State Physics and Optics (Szilárdtest-fizikai és Optikai Intézet), the Atomic Energy Research Institute (Atomenergia-kutató Intézet), the Institute of Technical Physics and Materials Science (Műszaki Fizikai és Anyagtudományi Intézet), the Institute for Nuclear Research (Atommagkutató Intézet), the Wigner Research Centre for Physics (Wigner Fizikai Kutatóközpont) and the Geographical Institute (Földrajztudományi Intézet).

The social sciences were also represented quite well, by the Institute for Sociology (Szociológiai Intézet) and other parts of the Centre for Social Sciences (Társadalomtudományi Kutatóközpont) such as the Institute for Political Science (Politikatudományi Intézet), the Research Institute for Linguistics (Nyelvtudományi Intézet) and the Institute of Economics (Közgazdaság-tudományi Intézet).

Of course, the indispensable mathematics also made use of the cloud, represented by places like the Alfréd Rényi Institute of Mathematics (Rényi Alfréd Matematikai Kutatóintézet). It was also the *subject* of research for SZTAKI, which further conducted a large amount of applied research on the infrastructure.

In 2021 the original MTA Cloud was substantially upgraded to cater for a larger uptake, more varied use cases and more GPU. The current deployment, the ELKH Cloud, enables researchers to use an IaaS cloud that is dynamically adjustable so that its size and type fit perfectly with the projects they have ongoing at the time, without having to go through complicated procurement and deployment procedures. These infrastructures range from simple desktop machines (e.g. MS Windows, Linux) to high-performance computing (HPC) clusters (e.g. SLURM); the flexibility of the ELKH Cloud makes both – and anything in between – possible, including the usage of pre-configured PaaS solutions that users can instantiate. Currently, PaaS and SaaS patterns are provided to cloud users in the form of reference architectures that are then implemented using infrastructure as code (IaC).

The present

Cloud services continue to spread rapidly for both academic and business uses. The factors behind the cloud's popularity have never faded: cloud computing is scalable and flexible, enabling users to utilise it in varying fields and at different scales, and it also represents a consolidation of know-how, alleviating scientists of the need to build infrastructures. In recent years, ELKH has taken over the mantle as caterer of cloud resources to Hungarian researchers; its continuance of operations and the upgrades were made possible by the joint efforts of SZTAKI and the Wigner Research Centre for Physics.

The current (as of the beginning of 2022) infrastructure offers 5904 virtual CPUs, 28 terabytes of RAM and 1248 terabytes of hard disk drive (HDD) capacity complemented by 228 terabytes of solid-state drive (SSD), both three times replicated. Its internal network offers 100 Gb/s.

The time frame of the existence of the current research infrastructure and its institutional predecessors at SZTAKI and Wigner saw an exponential growth in GPU-based machine learning as well as GPU-based simulation. Therefore, the upgraded ELKH Cloud offers substantial GPU resources. The 68 GPUs hosted by the system offer 2400 GB of GPU RAM, yielding 1174 teraflops of single-precision or 584 teraflops of double-precision computing capacity; or, for half-precision (FP16) TensorFlow workloads, 13736 teraflops.

Scientific use cases

The aforementioned shift of emphasis from IaaS to PaaS and SaaS is expressed by the growing number of reference architectures and platforms offered by the ELKH Cloud.

A reference architecture is a pre-arranged set of solutions to a typical scientific workflow – in essence, they are templates that help to kick-start a scientific project while alleviating the complexity of the initial configuration, a very taxing task for the users.

Some of the reference architectures are commonly used cloud stacks all over the world. These include:

- JupyterLab (Granger and Grout 2016), which allows for running JupyterLab notebooks (Bisong 2019) with or without GPU. A JupyterLab notebook is a widely used tool in machine learning research and education;
- Apache Spark (Salloum et al. 2016), which can be coupled with RStudio (Racine 2012) as well as with Python, and which enables big data research and, ultimately, thanks to the versatility of RStudio and Python, any other kind of research;
- Docker Swarm cluster, which is a distributed system of systems solution for engineers (Naik 2016);
- Kubernetes (Luksa 2017), which is a clustered distributed service deployment system;

- TensorFlow (Pang, Nijkamp and Wu 2020), which is a generic machine learning framework; and
- Keras, which is a deep learning framework (Ketkar 2017).

However, the ELKH Cloud also features more advanced solutions that are developed locally by the cloud team. This set of tools includes the Occopus cloud orchestrator, which '*provides a language to specify infrastructure descriptions and node definitions based on which Occopus can automatically deploy and maintain the specified virtual infrastructures in the target clouds*' (Kovács and Kacsuk 2018; Kovács et al. 2018; Lovas, Nagy and Kovács 2018; Lovas et al. 2018; Kacsuk, Kovács and Farkas 2018). An example usage of this is the Apache Hadoop architecture orchestrated by Occopus (Lovas, Nagy and Kovács 2017; Nagy, Kovács and Lovas 2016).

Some typical scientific research requires workflows and pipelines of processing. These are facilitated by Flowbster (Kacsuk, Kovács and Farkas 2018), which is a cloud-oriented workflow system. It was designed to create data pipelines in clouds that could efficiently process very large data sets. The Flowbster workflow can be deployed on the ELKH Cloud, on-demand with the help of the underlying Occopus cloud deployment and orchestration tool.

Moving on to storage, DataAvenue is on offer (Hajnal, Farkas and Kacsuk 2014; Hajnal et al. 2014, 2015, 2018). This data storage management service provides access to different types of storage resource (including S3, sftp, etc.) using a uniform interface.

And, of course, the authentication method matches the user base; with the help of the eduGAIN-enabled Hungarian SAML federation (eduID.hu) and a module developed for OpenStack by the MTA Cloud team (Héder, Tenczer and Biancini 2019), there is no need for local passwords for any users. The authorisation is performed by HEXAA (Tétényi et al. 2015).

International outlook

Science is international and so are the ambitions of the user base of the ELKH Cloud. From the very beginning, it was typical to have international collaborators working on the project. This was achieved by leveraging the eduGAIN worldwide login federation.

However, enabling user mobility alone soon became insufficient. Researchers also expect collaboration between the facilitating institutions. A major vehicle for that collaboration is the European Open Science Cloud (EOSC); another is the European Strategy Forum on Research Infrastructures (ESFRI). In order to integrate with these international efforts, SZTAKI participates in two EU-funded projects. In EGI-ACE the compatibility with other EOSC clouds is ensured, while in the SLICES-SC project we participate in an ESFRI effort that prepares the European usage of the Scientific LargeScale Infrastructure for Computing/Communication Experimental Studies.

Security challenges

As data has become the oil of the twenty-first century, recognition of scientific data as a bearer of value is also increasing in national researcher communities. Data collection and analysis techniques are applied in more fields of study than ever before. The ELKH Cloud and its sibling project, ELKH Data Repository, are among the leading infrastructures supporting the Open Science and Open Data movements in Hungary. The inclusive attitude of the project attracts many researchers as well as an increasing number of private entities interested in applied research.

As a result of its central role and ease of access, the infrastructure is experiencing an unparalleled volume in terms of ingress (and egress) of heterogeneous research data. Depending on the field of science, this may include data originating from 'clearly' scientific equipment, like astronomical sensors, but also from medical imaging devices, together with patient records necessary for meaningful analysis. Besides similarly clear data protection issues, more subtle considerations are also present in cooperative and applied research scenarios where special intellectual property issues can often arise, for example during scientific analysis of partially anonymised financial transactions or sensor data originating from industrial environments holding trade secrets.

This renders the ELKH Cloud not only an indispensable platform for efficient collaboration but also a cybersecurity target with considerable value. Involvement of national research communities in modern research techniques also implies raising awareness of various underlying issues, introducing individuals to the ever-changing balance between acceptable levels of usability and information security.

For these reasons and in order to keep the legal constituency clear in most cases, a strategic decision was made. The ELKH Cloud is committed to keeping its infrastructure 100% within national borders, utilising the holistic technical competency of the two research institutes that host its data centres. This is achieved while still maintaining support for scale-out operations with different research infrastructures, including well-known public cloud providers.

In order to improve resilience against malicious activities, the infrastructure is engineered considering security requirements from the first step of its planning. All internal and external components are isolated by various containment techniques to the point where they can still provide services as requested. Continuous security monitoring of internal infrastructure components and perimeter defence appliances is complemented by periodic active security scans of the hosted scientific projects. Security incident prevention and management processes are maintained in order to keep the overall attack surface at a minimal level.

User policies are also defined so that the shared responsibility of the infrastructure and its users is clear from the beginning; further, security-related consulting services are included in the primary user support package.

Future of the ELKH Cloud

As with any contemporary scientific infrastructure, both the policy-related and the technical issues need to be addressed in an integrated manner. This is especially important when it comes to protection of the research data. Policy decisions and data governance may be facilitated with built-in templates and tools, which constitute a future research area.

Another important area of improvement is the implementation of even more PaaS and SaaS functionality and reference architectures. Also, not only are new user-facing features planned but integration of the ELKH Data Repository is also on the agenda. This component allows for long-term storage of data, instead of handling working data as IaaS clouds usually do. Long-term data retention that happens in such a way as to obviate the need for the researcher to worry about maintenance is an important and often requested feature.

Another area of future work is introducing specialised hardware: it is a current trend in AI, for example, that in order to gain performance advantages certain algorithms are implemented not in software but in special-purpose hardware (such as ASICs, which stands for application-specific integrated circuits). Yet another area is the hybrid cloud – in this set-up, on-premise cloud computing can combine its workflows with a public cloud, such as Microsoft Azure or Amazon AWS.

The ELKH Cloud has during the first five years of its existence proven itself to be a popular platform among researchers. This was achieved by the cloud maintainers being responsive to trends in use of the tools of internationally competitive science. This trajectory is a century-long professionalisation and division of labour in the instruments of science. The trend is unlikely to stop; therefore, we expect the ELKH Cloud to grow and become even more extensively used in the future.

References

Bisong, Ekaba. *Building Machine Learning and Deep Learning Models on Google Cloud Platform.* New York City: Apress, 2019.

Granger, Brian E., and Jason Grout. *Jupyterlab: Building blocks for interactive computing.* Slides of presentation made at SciPy, 2016.

Hajnal, Ákos, Zoltán Farkas, and Péter Kacsuk. 'Data avenue: Remote storage resource management.' In *WS-PGRADE/gUSE, 6th IEEE International Workshop on Science Gateways (IWSG),* 1–5. Dublin, Ireland: International Workshop on Science Gateways (IWSG), 2014.

Hajnal, Ákos, Zoltán Farkas, Péter Kacsuk, and Tamás Pintér. 'Remote storage resource management in WS-PGRADE/gUSE.' In *Science Gateways for Distributed Computing Infrastructures: Development Framework and Exploitation by Scientific User Communities*, edited by Péter Kacsuk, 69–81. New York City: Springer, 2014.

Hajnal, Ákos, István Márton, Zoltán Farkas, and Péter Kacsuk. 'Remote storage management in science gateways via data bridging.' *Concurrency and Computation: Practice and Experience* 27, no. 16 (2015): 4398–4411.

Hajnal, Ákos, Enikő Nagy, Péter Kacsuk, and István Márton. 'Data migration for large scientific datasets in clouds.' *Azerbaijan Journal of High Performance Computing* vol. 1, issue 1(2018): 66–86. https://doi.org/10.32010/26166127.2018.1.1.66.86

Hannay, Jo Erskine, Hans Petter Langtangen, Carolyn MacLeod, Dietmar Pfahl, Janice Singer, Greg Wilson. 'How do scientists develop and use scientific software?' In *Proceedings of the Second International Workshop on Software Engineering for Computational Science and Engineering,* 1–8. Piscataway, New Jersey: IEEE - Institute of Electrical and Electronics Engineers Inc., 2009. https://doi.org/10.1109/SECSE.2009.5069155

Héder, Mihály, Szabolcs Tenczer, and Andrea Biancini. 'Collaboration between SAML federations and OpenStack clouds.' *SoftwareX 9* (2019): 44–48. https://doi.org/10.1016/j.softx.2018.12.003

Kacsuk, Péter, Zoltán Farkas, Miklós Kozlovszky, Gábor Hermann, Ákos Balasko, Krisztián Karóczkai, and István Márton. 'WS-PGRADE/gUSE generic DCI gateway framework for a large variety of user communities.' *Journal of Grid Computing* 10, no. 4 (2012): 601–630.

Kacsuk, Péter (ed.). *Science Gateways for Distributed Computing Infrastructures: Development Framework and Exploitation by Scientific User Communities*. Springer International Publishing, 2014.

Kacsuk, Péter, József Kovács, and Zoltán Farkas. 'The Flowbster cloud-oriented workflow system to process large scientific data sets.' *Journal of Grid Computing* 16 (2018): 55–83. https://doi.org/10.1007/s10723-017-9420-4

Ketkar, Nikhil. *Introduction to Keras*. New York City: Apress, 2017.

Kovács, József, and Péter Kacsuk. 'Occopus: A multi-cloud orchestrator to deploy and manage complex scientific infrastructures.' *Journal of Grid Computing 16*, no. 19 (March 2018): 1-19. https://doi.org/10.1007/s10723-017-9421-3

Kovács, József, Péter Kacsuk, and Márk Emődi. 'Deploying Docker Swarm cluster on hybrid clouds using Occopus.' *Advances in Engineering Software 125* (November 2018): 136–145. https://doi.org/10.1016/j.advengsoft.2018.08.001

Lovas, Róbert, Enikő Nagy, and József Kovács. 'Cloud agnostic orchestration for big data research platforms.' In *Proceedings of the Fifth International Conference on Parallel, Distributed, Grid and Cloud Computing for Engineering*, Paper 15., edited by Péter Iványi, Barry H. V. Topping, and Géza Várady, Dun Eaglais Station Brae. United Kingdom: Civil-Comp Press, 2017. https://doi.org/10.4203/ccp.111.15

Lovas, Róbert, Enikő Nagy, and József Kovács. 'Cloud agnostic big data platform focusing on scalability and cost-efficiency.' *Advances in Engineering Software 125* (2018): 167–177. http://dx.doi.org/10.1016%2Fj.advengsoft.2018.05.002

Lovas, Róbert, Attila Farkas, Attila Csaba Marosi, Sándor Ács, József Kovács, Ádám Szalóki, and Botond Kádár. 'Orchestrated platform for cyber-physical systems.' *Complexity* (2018):1–16. Paper: 8281079. http://dx.doi.org/10.1155%2F2018%2F8281079.

Luksa, Marko. *Kubernetes in Action*. New York, USA: Simon & Schuster, 2017.

Nagy, Enikő, József Kovács, and Róbert Lovas. 'Automated and portable hadoop cluster orchestration on clouds with Occopus for big data applications.' In *Proceedings of the Cracow Grid Workshop '16 CGW 2016*, edited by Marian Bubak, Michał Turala and Kazimierz Wiatr, 47-48. Kraków, Poland: Academic Computer Centre, 2016. ISBN: 978-83-61433-20-0.

Naik, Nitin. 'Building a virtual system of systems using Docker Swarm in multiple clouds.' In *2016 IEEE International Symposium on Systems Engineering (ISSE)*, 1–3. Edinburgh, UK: IEEE, 2016.

Pang, Bo, Erik Nijkamp, and Ying Nian Wu. 'Deep learning with TensorFlow: A review.' *Journal of Educational and Behavioral Statistics* 45, no. 2 (2020): 227–248.

Prabhu, Prakash, Thomas B. Jablin, Arun Raman, Yun Zhang,Jialu Huang, Hanjun Kim, Nick P. Johnson, Feng Liu, Soumyadeep Ghosh, Stephen R. Beard, Taewook Oh, Matthew Zoufaly, David Walker, and David I. August. 'A survey of the practice of computational science.' In *Proceedings of the 24th ACM/IEEE Conference on High Performance Computing, Networking, Storage and Analysis,* 19:1–19:12.New York (USA): Association for Computing Machinery, 2011.
https://doi.org/10.1145/2063348.2063374

Racine, Jeffrey Scott. '*RStudio: A Platform-Independent IDE for R and Sweave*.'
Journal of Applied Econometrics, John Wiley & Sons, Ltd. 27, no. 1 (2012): 167-172.

Salloum, Salman, Ruslan Dautov, Xiaojun Chen, Patrick Xiaogang Peng, and Joshua Zhexue Huang. 'Big data analytics on Apache Spark.' *International Journal of Data Science and Analytics* 1, no. 3 (2016): 145–164.

Tétényi, István, Mihály Héder, Zsuzsanna Magyar, and Kristóf Bajnok. 'Open Call Deliverable OCK-DS1.1 Final Report (HEXAA) Paper: OCK-DS1.1.' 2015.
https://geant3plus.archive.geant.net/Documents/HEXAA_final_report.pdf

Wilson, Greg, Dhavide A. Aruliah, C. Titus Brown, Neil Philippe Chue Hong, Matt Davis, Richard Guy, Steven H.D. Haddock, Kathryn Dorsey Huff, Ian M. Mitchell, Mark D. Plumbley, Ben Waugh, Ethan P. White, and Paul Wilson. 'Best practices for scientific computing.' *PLoS Biology 12, no. 1* (2014). e1001745.
https://doi.org/10.1371/journal.pbio.1001745.