

# CESDIS and BEOWulf

**How USRA's creative research environment delivered unexpected computing technologies for science-based applications.**

At the request of NASA in 1972, USRA established the Institute for Computer Applications in Science and Engineering (ICASE) at NASA's Langley Research Center. The subsequent success of ICASE prompted NASA to request USRA to create the Research Institute for Advanced Computer Science (RIACS) at NASA's Ames Research Center in 1983. In turn, the accomplishments of ICASE and RIACS led NASA to request a third computer-oriented institute from USRA in 1987, this time at NASA's Goddard Space Flight Center (GSFC). The new institute was called the Center of Excellence in Space Data and Information Sciences (CESDIS).

The original cooperative agreement between NASA and USRA for the establishment of ICASE stated that its first purpose would be to:

*(P)rovide a focal point with a university atmosphere to serve as a center of the academic community for activities related to applied mathematics, computer science, and the application of the computer to the solution of scientific and engineering problems.<sup>1</sup>*

As an association of universities, USRA has sought to keep the spirit of this language in the management of all its programs, because "a university atmosphere" connotes freedom for creativity in research and often results in surprising benefits. CESDIS represents a case study of this principle.

CESDIS was formed, at the urging of Dr. Milton Halem of GSFC, to carry out advanced computer science research in areas of potential, long-term interest to NASA programs, placing special focus on the processing and managing of data from space-borne, Earth-observing systems.

USRA asked Professor John Hopcroft of Cornell University to serve as the Interim Director of CESDIS while a search was conducted for the first permanent director. That search resulted in the appointment of Dr. Raymond E. Miller. The University of Maryland had collaborated with USRA in the formation of CESDIS by agreeing to provide a joint appointment in its Computer Science Department at the College Park campus for the CESDIS Director. Miller had been at Georgia Tech, but accepted a joint appointment with USRA and the University of Maryland to become the director.

Research areas of primary interest at CESDIS included:

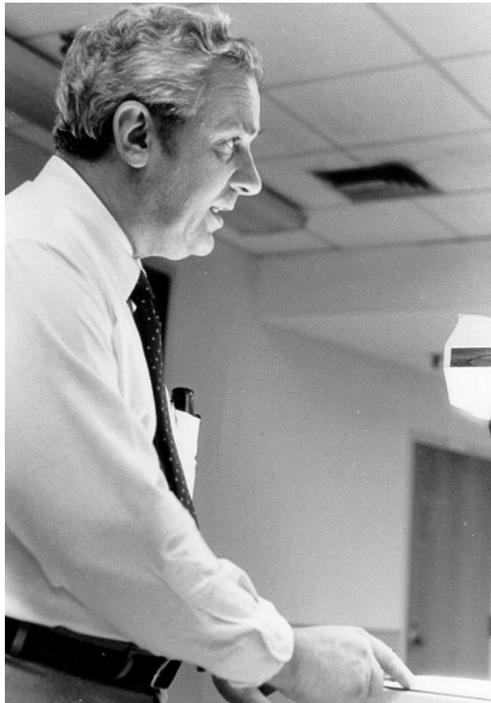
- High performance computing, especially software design and performance evaluation for massively parallel machines
- Parallel input/output and data storage systems for high performance parallel computers
- Database and intelligent data management systems for parallel computers
- Image processing
- Digital libraries
- Data compression

The motivation for these research areas arose out of the use of satellites for Earth science studies, e.g., climate change, crop yields, Earth resources, pollution studies and, in general, understanding the impact of human activities on major Earth ecosystems.<sup>2</sup>

As CESDIS had a small permanent staff, the Center issued calls for proposals to the computer science community and funded the top proposals on a competitive basis. Thus, a major part of the CESDIS research effort was carried out by associated scientists at universities who won the bids for multi-year funding.

In response to the first call for proposals issued in December 1987, CESDIS received 86 proposals. The first four awards from CESDIS were for:

- *Parallel Compression of Space and Earth Data* – Duke University
- *A Knowledge-based Advisory System for General Scientific Data Visualization* – George Washington University and Georgia Institute of Technology
- *Computer Assisted Analysis of Auroral Images Obtained from High Altitude Polar Satellites* – Stanford University and the University of Michigan
- *Image Pattern Recognition Supporting Interactive Analysis and Graphical Visualization* – The University of North Carolina at Chapel Hill

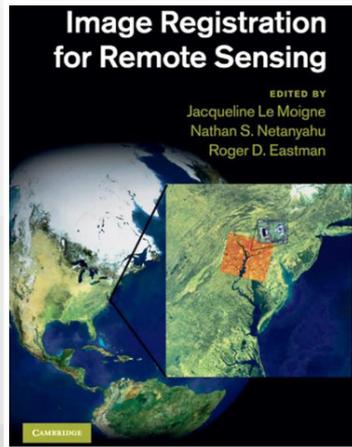


**Raymond Miller**  
USRA Director of CESDIS



**The motivation for these research areas arose out of the use of satellites for Earth science studies. CESDIS issued calls for proposals to the computer science community and funded the top proposals.**

The small core staff of CESDIS generally worked on enabling technologies, such as the task of finding efficient methods of registration of remotely sensed satellite imagery. Registration, which is the process of determining the best match of one scene with another, is a fundamental first step in the processing of satellite images of the Earth. For the large data flows produced by instruments on the Landsat and Earth Observing System satellites, image registration by hand is not feasible. The process must be automated, but automatic registration of satellite images is a difficult challenge for Earth science applications. The data often come from different instruments, on different satellites, with different resolutions, taken at different times of day and seasons.



## Jacqueline Le Moigne

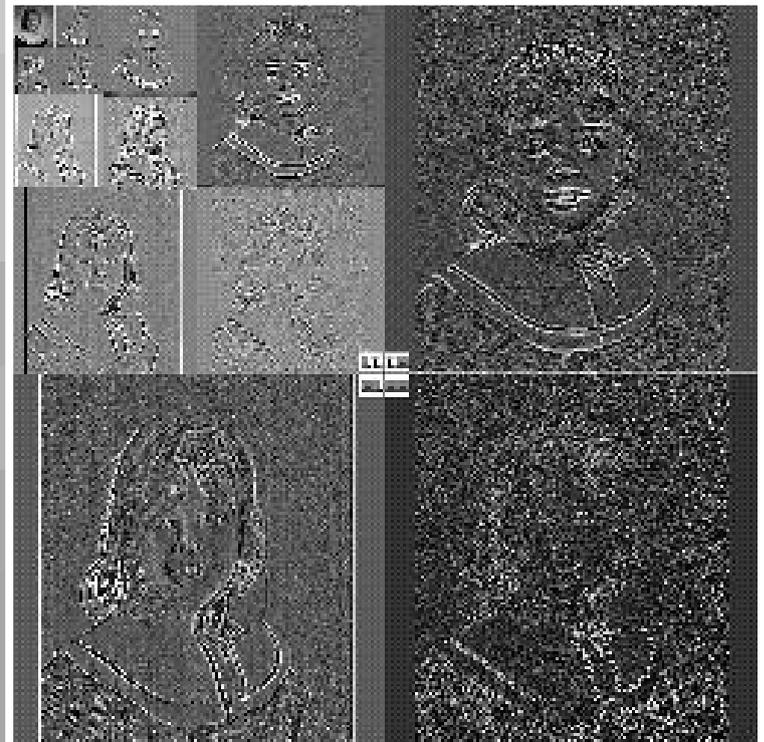
USRA CESDIS,  
Computational Sciences

Dr. Jacqueline Le Moigne, who joined CESDIS in 1992, became a world leader in automated image registration techniques that use wavelet transforms. Dr. Le Moigne received her Ph.D. in Computer Vision from the Pierre and Marie Curie University in Paris before coming to the U.S.

Le Moigne's work was critical in the analysis of a stream of data from satellite imagery because one often wants to know when a sudden change occurred in the data stream, e.g., when an edge in the scene was encountered. Fourier transforms, which are based on sine and cosine functions, can give the overall frequency components of a signal, but are less useful in space localization. Transforms that use sets of wave functions of limited duration ("wavelets") are better at space localization and have found widespread application in image processing, including image registration.<sup>3</sup>

In a 1994 paper, Le Moigne used an image of her daughter to illustrate how:

*Wavelet analysis can be implemented in a separable fashion by filtering the original image by a high-pass and a low-pass filter, iteratively in a multiresolution fashion, and separately in rows (vertical filter) and in columns (horizontal filter). At each level of decomposition, four new images are computed. Each of these images has one quarter of the number of pixels of the original image at the previous level, and it represents the low-frequency or high-frequency information of the image in the horizontal or/and the vertical directions; images LL (Low/Low), LH (Low/High), HL (High/Low), and HH (High/High).... The Low/Low image appears to be the same as the original image, except that it has a lower resolution. The other three images contain edge information, which when added to the Low/Low image can recreate the original image at full resolution.<sup>4</sup>*





**Yelena Yesha**

USRA Director of CESDIS and Applied Information Technology



**Thomas Sterling**

USRA CESDIS, Scalable Systems and Beowulf Project



**Donald Becker**

USRA CESDIS, Beowulf Project

CESDIS undertook many other tasks beyond the problem of image registration. In June 1991, USRA was asked by NASA to initiate a search for a Senior Technical Consultant within CESDIS to provide technical advice for NASA's High Performance Computing and Communications program. A USRA search committee consisting of the directors of its three computer-oriented institutes selected Dr. Thomas L. Sterling for the position. Sterling received his PhD from MIT's Electrical Engineering and Computer Science Department in 1984. Prior to joining USRA, he was a member of the research staff at the Supercomputing Research Center (SRC) of the Institute for Defense Analysis.

Raymond Miller retired as the Director of CESDIS in the fall of 1993. A search for his successor resulted in the appointment of Dr. Yelena Yesha as CESDIS Director in 1994. Born in the Ukraine, Dr. Yesha received her Ph.D. in Computer and Information Science from the Ohio State University. Following her appointment, Dr. Yesha organized CESDIS into three branches:

- Applied Information Technology, headed by Dr. Yesha
- Computational Sciences, headed by Dr. Le Moigne
- Scalable Systems, headed by Dr. Sterling

In April 1994, Donald J. Becker joined CESDIS to explore the potential of a high performance parallel workstation built from inexpensive hardware and software. Becker was a graduate of MIT, and, like Sterling, had worked at the SRC before coming to CESDIS. While at the SRC, Becker wrote a substantial portion of the low-level Linux<sup>5</sup> networking code, including over a dozen device drivers for network adapters.<sup>6</sup> His experience on these drivers would be essential for the next

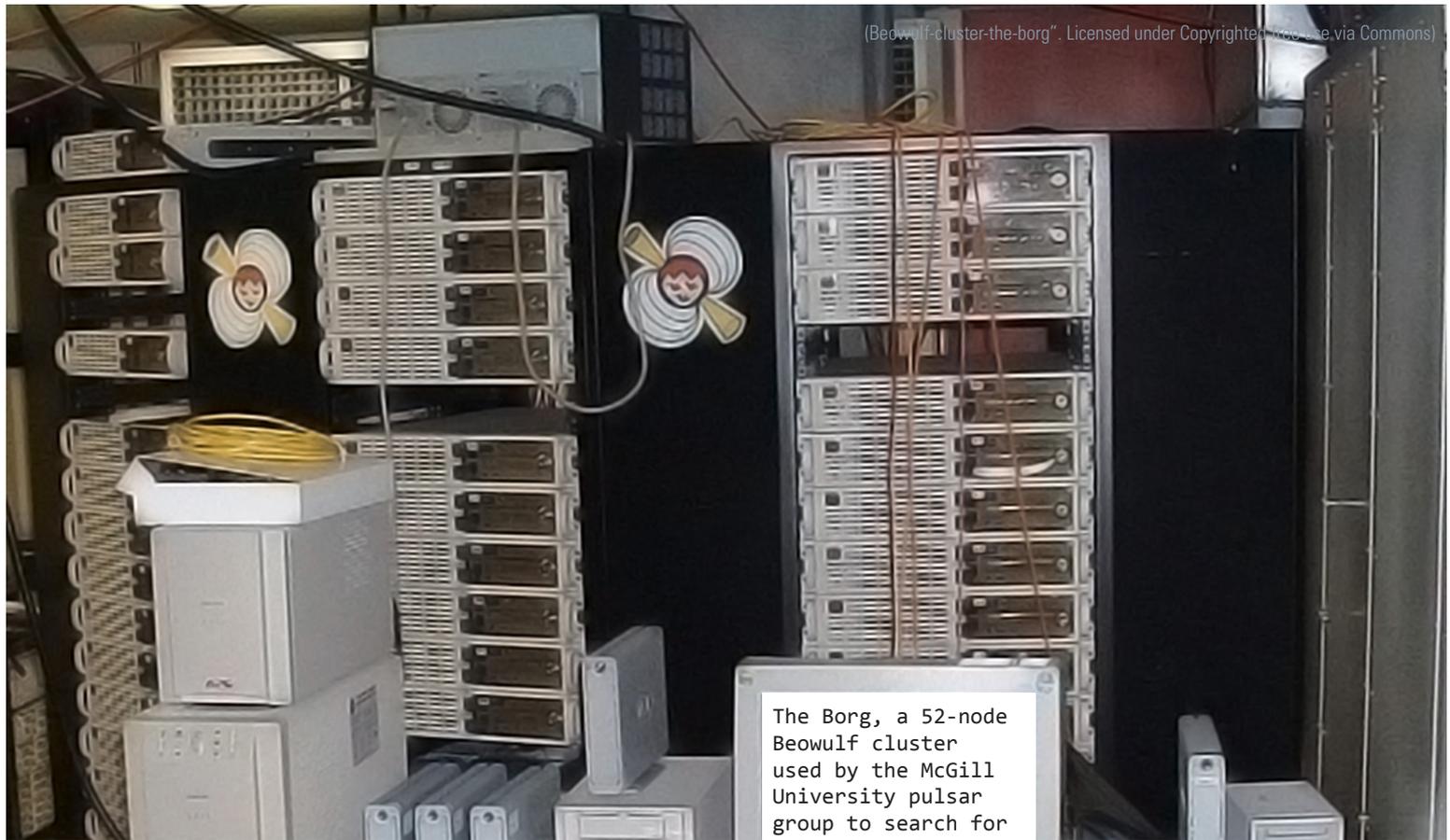
**As a metaphor for a legendary hero, 'Beowulf' has been applied to a new strategy in high performance computing that exploits mass-market technologies to overcome the oppressive costs in time and money of supercomputing.**

phase of Becker's work. With the support of James R. Fischer, a NASA project manager at GSFC, Becker and Sterling teamed to work on the "Beowulf Project." As described by Sterling, et al:

*Beowulf was the legendary sixth-century hero from a distant realm who freed the Danes of Heorot by destroying the oppressive monster Grendel. As a metaphor, 'Beowulf' has been applied to a new strategy in high performance computing that exploits mass-market technologies to overcome the oppressive costs in time and money of supercomputing.<sup>7</sup>*

The history of the project was recounted nicely in the citation of the Dr. Dobb's Excellence in Programming Award, which Becker was awarded in 1999 for his efforts in the development of the Beowulf.

*One of the challenges in the realm of scientific computing is to efficiently and affordably handle large data sets. This is precisely the problem faced by researchers participating in the Earth and Space Sciences Project at the Goddard Space Flight Center. To tackle the problem, Donald Becker and Thomas Sterling launched*



The Borg, a 52-node Beowulf cluster used by the McGill University pulsar group to search for pulsations from binary pulsars.

# **B**eowulf clusters turned out to have a surprisingly wide application. The use of Beowulfs is so widespread that “Beowulf Cluster” now refers to a class of supercomputers.

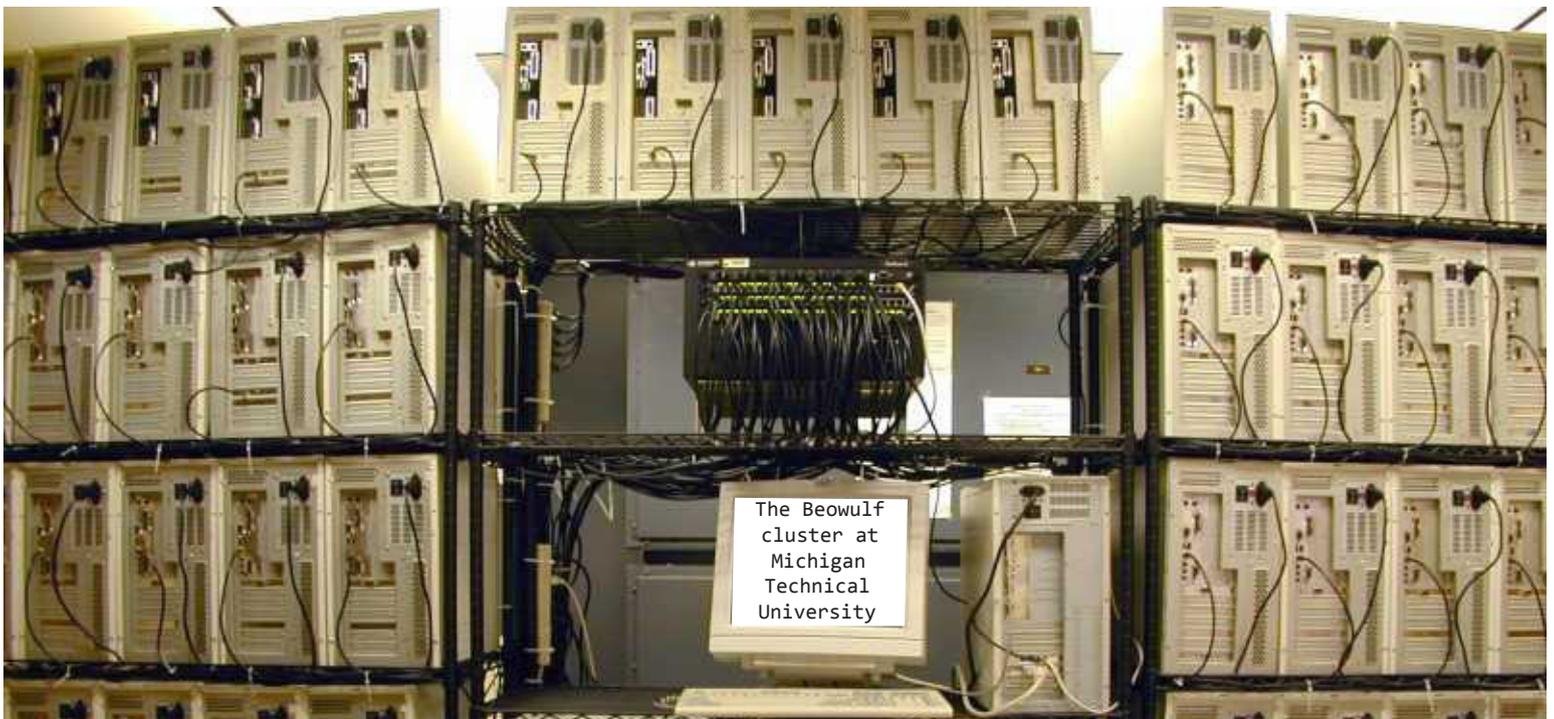
*the Beowulf Project, a cluster computer consisting of high-performance PCs built from off-the-shelf components, connected via Ethernet, and running under Linux. Ultimately, the goal of the Beowulf approach was to achieve supercomputer (gigaflop) performance at PC prices.*

*To implement such a system, however, Becker, who is a staff scientist with the Center of Excellence in Space Data and Information Sciences (or CESDIS, part of the Universities Space Research Association, a nonprofit consortium of universities that sponsors space-related research), had to come to grips with Linux’s unstable networking capabilities, and the lack of Linux support for off-the-shelf network cards. Consequently, Becker ended up writing enhancements to the kernel network subsystem to support faster I/O on high-speed networks, device drivers for countless Ethernet cards, and a distributed shared memory package. ...*

*Although much of his initial work [at CESDIS] was in support of Beowulf, the entire computing community ultimately benefited from Becker’s efforts. Linux would*

*not have achieved the level of success and acceptance it has today had it not been for Becker’s work, which resulted in a Linux with robust, stable networking and support for “every shipping Fast Ethernet chipset.” As for Beowulf, dozens of university and research groups have now built their own Beowulf clusters, ranging from the original 16-node cluster running on Intel DX4 processors connected by channel-bonded 10-Mbits/sec Ethernet, to Avalon, a 19-gigaflop cluster of 140 Alpha processors that was built by the Los Alamos National Laboratory and that cost only \$150,000.*

*Along with other members of his team at the Center of Excellence in Space Data and Information Sciences, Becker was the recipient of the ...1997 Gordon Bell Prize for Price/Performance “in recognition of their superior effort in practical parallel-processing research.”<sup>8</sup>*



The motivation for the development of the first Beowulf cluster computer at CESDIS was the desire to attack the problems faced by Earth and space science research communities, which are usually associated with large data sets. Beowulf clusters turned out to have a surprisingly wider application, so widespread that “Beowulf Cluster” now refers to a class of supercomputers. Dr. Phillip Merkey, a former CESDIS researcher, described the wider application:

*Beowulf class cluster computers range from several node clusters to several hundred node clusters. Some systems have been built by computational scientists and are used in an operational setting, others have been built as test-beds for system*

*research and others serve as inexpensive platforms to learn about parallel programming.<sup>9</sup>*

In 1999, Sterling and Becker were co-authors (along with John Salmon and Daniel Savarese) of a popular book titled *How to Build a Beowulf: A Guide to the Implementation and Application of PC Clusters*. This book helped to make possible the construction of Beowulf clusters around the world, including at USRA institutes. As an example, at the request of NASA, Brian Fessler and others in the computer staff at USRA’s Lunar and Planetary Institute constructed a 96-processor Beowulf cluster so that USRA’s Division of Space Life Sciences (DSLS) could use it in NASA’s Space Radiation Program. The DSLS Beowulf cluster was used to perform research on the health risks to astronauts from space

radiation. Systems biology models of cancer, central nervous system effects, heart disease, and acute radiation sickness were developed by DSLS and NASA staff using the Beowulf cluster with models of particle track structure and the space environment. In addition, USRA’s ICASE built a 64-processor Beowulf cluster to serve as a test bed for the institute’s research activities in parallel and distributed computing, and CESDIS used the original 16-processor Beowulf cluster for its own research projects.

As noted above, USRA has interpreted “provide a focal point with a *university atmosphere*” to mean that USRA’s institutes and programs should give researchers like Jacqueline Le Moigne, Thomas Sterling and Donald Becker as much freedom as possible to develop their innovative ideas. The results are often surprising and far reaching.

<sup>1</sup> Hussaini, M. Y. (1993). ICASE twenty-year anniversary publication. Columbia, MD: Universities Space Research Association; p. 112.  
<sup>2</sup> Le Moigne, J., Netanyahu, N. S., and Eastman, R. D. (Eds.). (2011). Image registration for remote sensing. Cambridge University Press.  
<sup>3</sup> El-Ghazawi, T. A., Chalermwat, P., and Le Moigne, J. (1997). Wavelet-based image registration on parallel computers. In Proceedings of the 1997 ACM/IEEE conference on Supercomputing (pp. 1-9). ACM; p. 2.  
<sup>4</sup> Le Moigne, J. (1994). Parallel registration of multisensor remotely sensed imagery using wavelet coefficients. In *SPIE’s International Symposium on Optical Engineering and Photonics in Aerospace Sensing* (pp. 432-443). International Society for Optics and Photonics; p. 243.  
<sup>5</sup> Linux is an open-source computer operating system that was developed in 1991 by Linus Torvalds, who was a graduate student at the University of Helsinki at the time.  
<sup>6</sup> Becker, D. (1994). HPCC/ESS GOPS workstation system software environment. In Pratt, T. W. CESDIS Annual Report, Year 6, July 1993 – June 1994. Columbia, MD: Universities Space Research Association; p. 30.  
<sup>7</sup> Sterling, T., Salmon, J., Becker, D. J., and Savarese, D. V. (1999). *How to Build a Beowulf: A Guide to the Implementation and Application of PC Clusters*. Cambridge, MA: MIT Press; p. 1.  
<sup>8</sup> Erickson, J. (1999). Excellence in programming awards. Dr. Dobb’s Journal. Retrieved from <http://www.drdoobbs.com/open-source/dr-dobbs-journal-1999-excellence-in-prog/184410960>  
<sup>9</sup> Merkey, P. (2015). Beowulf history. Beowulf Overview for Beowulf.org. Retrieved from <http://www.beowulf.org/overview/history.html>.