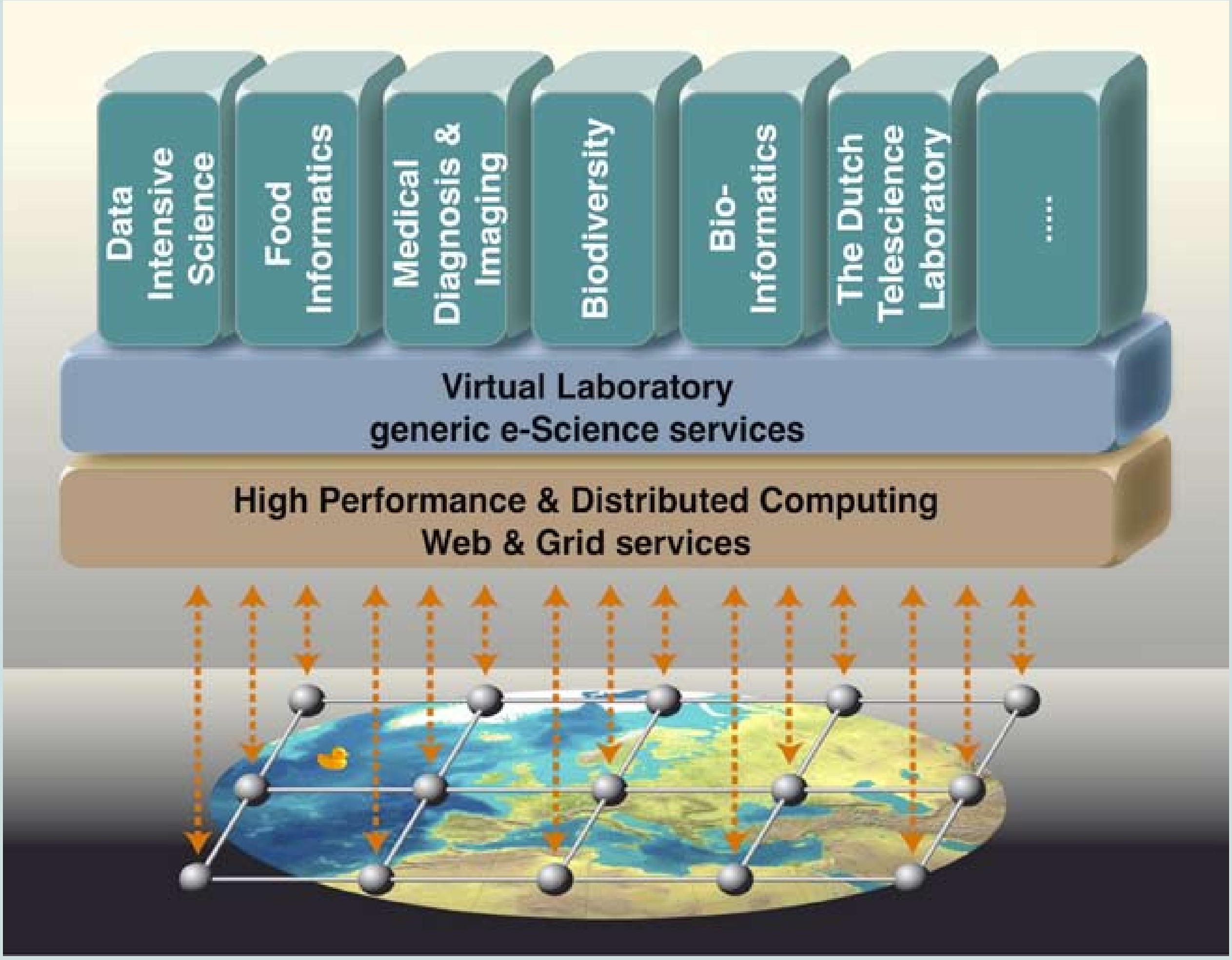
VL-e
Virtual Laboratory for e-Science

2004-2009

Final report



Project number:	Bsik 03019
Project title:	Virtual Laboratory e-Science (VL-e)
Start of the project:	1 January 2004
Planned final date:	31 December 2009*

* *By letter of October 4 2007 the Minister of OCW agreed to the request of the VL-e directorate to extend the project with one year until December 31, 2009.*

Content

Executive Summary

"VL-e: a structured approach to e-Science enabling system level research"

The mission of the VL-e project has been 'to boost e-Science by creating an e-Science environment and performing research on methodologies'. e-Science is crucial for modern science and society. It is a new paradigm that enhances science by enabling system-level science: the integration of diverse sources of knowledge about the constituent parts of a complex system with the goal of obtaining an understanding of the system's properties as a whole system.

If one considers data, information and knowl-edge metaphorically as raw materials, semi-finished products and end products respectively, then e-Science is the ICT-based production chain. An e-Science infrastructure enables researchers to handle the ever-increasing size and complexity of their data, for example, to gain insight in the fundamentals of matter in physics or in the environmental factors determining the survival of species or to unravel the functioning of the living cell. Also, in terms of socio-economical impact, e-Science has unprecedented potential. Food safety (Unilever and others), health systems (Philips) or a bird avoidance system for flight safety (commercial aviation and air force) are examples, where VL-e has enabled enormous progress. However, e-Science does not come of its own accord. It requires a new, multidisciplinary approach towards science. The VL-e project has been instrumental in showing the importance of this change. Its successful results have led many to call VL-e the flagship Dutch e-Science project. To safeguard the results and to continue the e-Science mission, it is crucial to establish an e-Science research center.

To accomplish its mission, the VL-e project formulated the following strategy: '*To carry out concerted research along the complete e-Science technology chain, ranging from applications to networking, focused on new methodologies and reusable components.*'
This is reflected in the structuring of the project along three different program layers: the application layer, the virtual laboratory layer (containing generic services for applications) and the infrastructure oriented layer (in which the resources are harnessed and exploited, such as web, grid and high performance distributed computing). To underline the importance of addressing the complete technology chain, two experimentation environments have been set up as a fourth program line: a rapid prototyping (RP) environment and a proof of concept (PoC) environment. A dedicated team (the VL-e Integration Team, VLeIT) has coordinated the integrating effort between all VL-e components. The structuring (reflected in the well-known model depicted below) greatly contributed to the coherence in the project. This model has been widely adopted in national and international communications on the notion of e-Science.

The VL-e project has delivered distinctive scientific results within these various layers. We learned that it is useful to discern an extra layer, the domain-generic layer, in which e-Science elements are developed that can be reused by multiple end-users or applications within a same domain. Furthermore, all application domains have succeeded in creating one or more problem-solving environments (PSEs) covering (a part of) their e-Science technology chain. Many of them have even reached the stage of effective production, such as the bird avoidance flight safety environment (Flysafe). For some

PSEs, such as KnowEx (information management) and ViroLab (a virtual laboratory for infectious diseases), efforts are underway to create a (commercial) spin-off company. Many concepts and products from the VL layer have found their way outside the VL-e project, such as the vBrowser, JavaGAT, VL-e Toolkit, AIDA and VLAM, showing that reusability has been brought into practice. In terms of economic impact this reusability may well have saved in the order of M€ 30 compared to development from scratch, not even including the extra spillover from EcoGrid estimated at M€ 20. We note that the project's PoC environment is the most likely candidate to evolve into the *de facto* standard for the future e-Science infrastructure in the Netherlands.

Other — more classic and easily measurable — output from the project has far surpassed our own prior expectations: 463 refereed publications, 251 presentations, 22 completed PhD theses and a wealth of spin-off projects, from which the BiG Grid project, rolling out a national grid based e-Science infrastructure, is the largest, with a funding of M€ 29. VL-e has also been an important contributor to the ICT Regie report '*Towards a competitive ICT infrastructure for scientific research in the Netherlands*' (December 2008), adopted by the Dutch government in Spring 2009, containing a strong recommendation to sustainably fund this infrastructure and include an e-Science Research Center to support the innovation. This is fully consistent with the strong appeal from the 'Commissie van Wijzen' to the VL-e directorate, expressed after the project's midterm review, to establish such a center.

Finally, via the consortium partners, the VL-e efforts have been well integrated with many other national, European and international projects, notably with the emerging collaborations around the ESFRI projects (such as LifeWatch and ESS). The ESFRI projects are important focal points for the formation of an integrated vision on realizing a European e-Science infrastructure (or e-Infrastructure). Again, modern system-level science is unimaginable without such an infrastructure.

Financially the VL-e project closed the books per December 31, 2009. Compared to a budgeted amount of M€ 41.3, the total expenditure during the project period (01.01.2004 - 31.12.2009) has been M€ 42.8 (104%), meaning that all subsidy has been spent. Industrial consortium partners have spent less in the project (M€ 3.5, i.e. 8% compared to a budgeted 19%), be it for well understandable reasons: for instance FEI and DSM have early in the project shifted their efforts to the adjoining Collaboratory.nl project, whilst Philips Research has directed part of its efforts to the BiG Grid project. The universities and research institutions have more than compensated for this by increasing their effort in the project.

The VL-e project has been the second in a series of projects that started in the mid nineties and have been funded from national 'gas funds' to strengthen the knowledge infrastructure. e-Science research requires time, patience and hard work. Scientific disciplines vary greatly in the speed at which they take up e-Science methodologies. The special funding has enabled the formation of multidisciplinary teams in an environment that does primarily judge and reward researchers on the basis of their mono-disciplinary output. With the project finished, it is important to reflect on how the unique and valuable expertise built up during more than a decade will be secured and expanded. Part of this expertise has been transferred to adjoining projects, such as BiG Grid, and part will be passed on to succeeding projects, such as COMMIT. But, all these projects have a predefined lifespan.

"e-Science Research Center for embedding VL-e results"

Following the ICT Regie recommendations, NWO and SURF are now jointly investigating the establishment of a permanent e-Science Research Center. The experience of VL-e shows that a truly multidisciplinary approach along the complete technology chain, well embedded within international developments, is the only viable way in which such an enterprise will succeed.

Chapter 1
e-Science and VL-e

By the time the VL-e project had started in early 2004, it was clear that many science domains were already rapidly moving to become *system-level sciences*, mainly because modern science, engineering and society were increasingly faced with complex problems that can only be understood in the context of the full overall system they belong to. During the course of the VL-e project, this trend has further intensified, in that most science domains have introduced, or are introducing a system-level science approach. This is even increasingly true for the alpha and gamma sciences.

There are several reasons and enablers for this shift to system-level science:

Technology push
In many domains, technology innovations on, for instance, detectors with ever increasing resolution, allow deep observations of scientific phenomena important for the better understanding of a whole system. In addition, information technology innovations, such as digitalization of collections of written documents, also unlock resources at a systems-level.

Globalization
Like in economics and society, science is experiencing an up-scaling due to globalization. Establishing and managing big data and information repositories often demand an international effort. This can also be read from the ever-increasing aggregation of research funding, such as ESFRI.

System-level science:

The integration of diverse sources of knowledge about the constituent parts of a complex system with the goal of obtaining an understanding of the system's properties as a whole.
Ian Foster (Argonne National Lab), November 2006 issue of scientific journal IEEE Computer

Resource integration
Never before did researchers of so many domains have such a wealth of resources at their disposal. The integration of these worldwide available resources has further fueled system-level research. An important contribution of e-Science as a system-level science is its potential for integration of information.

"System level science enabled by technology push, globalization, and resource integration"

Examples of system-level science are:

- Study of black matter for understanding the origin of the universe *(Data Intensive Science)*
- Study of the fundamentals of matter in physics (CERN) *(Data Intensive Science)*
- Improving the quality, safety, and sustainability of food and food production *(Food Informatics)*
- Biobanking-based cohort studies in medicine *(Medical Diagnostics & Imaging)*
- Environmental studies into the role of biodiversity and species behavior in ecosystems *(Biodiversity)*
- Systems biology to unravel the functioning of the living cells *(Bio Informatics and Telescience)*
- Engineering study in network congestion behavior for the next phase Internet *(High Performance & Distributed Computing)*

- Huge sociology studies based on large digital collections of annotated language data
- Large-scale archiving and data-access services in the arts and humanities.

However, system-level science could not be realized with the available concepts, methods, tools, and infrastructure that were designed and developed to meet the needs of traditional pursuit of science. System-level science demands new innovative versions of all these elements.

e-Science is the emerging science paradigm that enables modern system-level studies. e-Science uses computer techniques to handle and harness the exponentially increasing amount of data, as well as the complexity arising from these system-level studies. It also addresses the important aspects of multidisciplinary approach and allows remote collaboration. e-Science develops the required concepts and methodologies, implemented via software tools which can be applied as part of an ICT based research infrastructure for system-level science in virtually all science domains. The strength of the e-Science approach is that it supports *real world* and *in-silico* experimentation plus the powerful combination of both.

At the same time, however, e-Science enabled system-level research and the corresponding ICT based research infrastructures have become so complex that they can only be realized by multidisciplinary collaborations, often dispersed all over the globe as, for example, in high energy physics and astronomy. Hence, developing the necessary integrated infrastructure is a challenging research problem in itself that can only be tackled by multidisciplinary research teams.

The success of e-Science will heavily depend on the ability and willingness of researchers to accept a cultural shift in the way they undertake science. VL-e has made them aware that such a change is necessary.

VL-e realized early onwards that it is imperative to structure the e-Science process along the full technology chain in order to clarify the role of the various activities and stakeholders involved (physical instrument engineers, computer and computational scientist and engineers, application domain end-users).

"System level science demands an integrated approach to instruments computing storage visualization and networking"

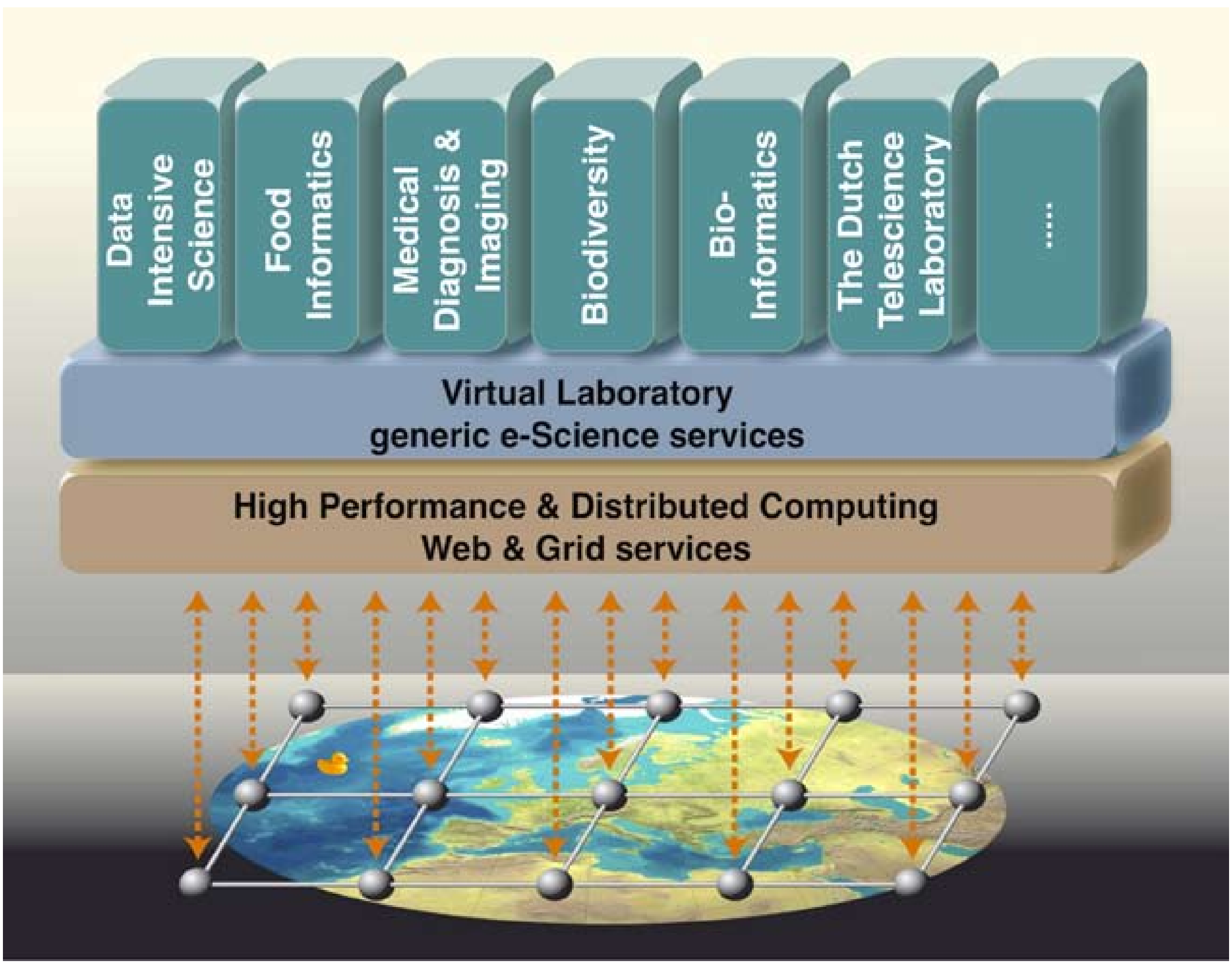
To realize e-Science and system-level science, an integrated approach to instrumentation, computing, storage, and networking is essential. As a result, two distinct types of research innovation have been abundantly achieved by the VL-e project: firstly, innovations of e-Science itself and secondly, innovations of the employment of e-Science in the various applications domains.
Consequently, VL-e has resulted in direct output in the innovation of science methodology by the new results it produced. However, it had a far larger indirect output by the creation of the awareness of the impact of e-Science via, among others, education, publication in various journals, and presentations at various national and international events.
Moreover, the creation of the scaling and validation program and the close collaboration with M€ 29 research-infrastructure program BiG Grid, of which VL-e was one of the most important contributors to the proposal and the implementation, has realized a multiplier effect for VL-e's direct as well as indirect output.
For example more than 34 user groups are using VL-e e-Science methodology and software and influence others to do the same.

"more than 34 Dutch user groups deploy VL-e methodology and software"

The dissemination strategy of the VL-e project has always been to first exploit results on the national (Dutch) level (for instance through BiG Grid) and then on the international level. During the last phase of the project, much attention was therefore given to international embedding. During the whole project, VL-e researchers of course collaborated intensively in the international arena, but especially during the last phase the VL-e ideas, expertise, methodology, and software (i.e. know how) were adopted by many international researchers. Consequently, all these activities helped to strenghthen the international position. Moreover, the indirect outflow will be further augmented by the many new international projects that were started and will spread the VL-e concepts even further. Examples of outflow of VL-e tools in such international projects as HealthGrid, German D Grid, the French CREAT, the US TeraGrid, TextGrid (a grid for the humanities in Germany and Europe), AeroGrid (aerospace research community) and AstroGrid-D (for astronomy), and many more, are presented in Chapter 5.

On the technical level we further structured the process by developing a functional model.

Following the functional model, VL-e was organized in four categories of subprograms that reflect the three abstraction levels in the function model: application, generic e-Science, high performance and distributed computing. In addition, the fourth category scaling and validation created the necessary infrastructure. Table 1 in the Appendix presents the various subprograms in more detail.



Functional model for e-Science developed in the VL-e project
The green columns represent the system-level studies performed by multidisciplinary collaboration of the domain scientists with scientists from the lower layers.
The Virtual Laboratory layer provides the generic e-Science services that are necessary to carry out system-level studies.
The web, grid, high-performance distributed computing layer takes care of harnessing the resources and exploiting the parallelism of the distributed computer system.
The bottom level contains the network + connected infrastructure that provide the computing, data storage, processing and visualisation.

Chapter 2
Milestones and Outflow

All the research results and knowledge transfer as realized by VL-e can be presented in terms of facts and figures (deliverables). In this report we distinguish five categories: scientific output, economic output, societal output, innovation, and knowledge transfer by outflow of staff.

Scientific output

Within the project itself, the number of scientific publications (and presentations) is impressive and far exceeds what could reasonably be expected at the beginning of the project. This is even more remarkable, taken into account the considerable amount of methodologies, software, and ICT infrastructure that had to be developed before scientific results could be obtained. Moreover, many of our VL-e results have been published in top-quality journals or presented at high-ranking conferences. The high number of completed PhD theses illustrates the project's success in educating young people, in the difficult multidisciplinary e-Science domain.

Publications	463
Presentations	251
PhD theses	22
New projects	> 27
Other new research lines	21
Researchers involved	150
Locations	33
POCs	37

"463 publications, 251 presentations, 22 PhD theses, more than 21 new projects"

Detailed information is presented in the Appendix, Table 2.

Economic output

Although we did not quantify the economic milestones in our original planning, we are convinced that the economic impact, for a science-oriented project, is quite substantial. VL-e started a major collaboration with industry: the COS group at the Philips Research Laboratories has actively participated in the project.

"Unilever incorporated company-wide VL-e knowledge discovery methods"

Unilever has company-wide incorporated the knowledge discovery methodologies that were developed in VL-e. Also Logica/CMG has integrated grid and e-Science processes in their business. VL-e is also involved in several techno starters. Of special interest is the growing number of end-users in various domains of the life/earth sciences, in various universities, academic hospitals, and non-for-profit organizations such as the RIVM and Natuurmonumenten, which increasingly are integrating e-Science and VL-e concepts in their research. A conservative estimate, as also described in the recent Cook report, showed that a generic e-Science approach based on reuse of components reduces the development costs of methods with ~ 65%. This estimate did not include the spill-over revenues from involved projects such as EcoGRID (M€ 20).

"EcoGRID spill-over effect M€ 20"

Detailed information is presented in the Appendix, Table 3.

Societal output

VL-e has generated a large number of societal activities. New ones are emerging such as, biomedical applications that are actually used in daily medical practice, for example, supporting first-aid care at the AMC and neurosurgery planning for oncology patients. The construction of an e-BioLab with a high-resolution tiled display to support e-BioScience project teams proved to offer an actual gateway to the know-how of the Virtual Lab for multidisciplinary teams in life sciences. These examples show that the VL-e concept makes e-Science work in many different societal application areas.
The details of societal output are presented in the Appendix, Table 4.

Innovation

As became evident during the course of the project, the e-Science concept was absorbed by the various application domains at a different speed and in a different manner. The six application areas differed considerably in their experience with e-Science prior to the VL-e project, which had a distinct effect on the innovation achieved via this project:

- The *Data-intensive sciences (SP1.1)* and *DUTELLA (SP1.6)* already had substantial experience with e-Science; these areas (in particular the latter) were considerably strengthened in their e-Science methodology and infrastructure.
- For *Medical diagnosis & imaging (SP1.3)* and *Biodiversity (SP1.4)*, e-Science was relatively new; the project caused a strong boost in their application of e-Science, resulting in many new system-level research activities in these areas.
- For *Food informatics (SP1.2)* and *Bioinformatics (SP1.5)*, e-Science was completely new; here the adoption of the e-Science concepts and the initial e-Science results are the most important achievements.

The described differences had an impact on the way in which e-Science was realized. For instance, in high-energy physics and astronomy, historically data-intensive sciences, the e-Science developments were oriented at the technical development in the VL-generic e-Science services and VL-High Performance & Distributed Computing Web & Grid services layer. In the life sciences, not much experience with data-intensive experimentation was present. The life sciences encompass a multitude of different subdomains and have to deal with extremely heterogeneous data, information and knowledge resources. Therefore, the e-Science focus in this domain has been mainly on e-Science developments in the VL-generic e-Science services layer and in the VL-domain specific services layer. In addition, activities had to be carried out at the conceptual level, which resulted in a new layer in our model: the domain-generic layer.

To summarize the dissemination activities, which have been the most relevant to our innovation milestones:

- VL-e helped to start several techno platforms, in particular GridForum Nederland.
- VL-e members contributed to grid standardization groups in several Open Grid Forum areas.
- Strong international embedding led to both 'import' and 'export' of knowledge and tools, especially with the help of the scaling and validation program.
- VL-e earned a prominent and positive place in an internationally renowned COOK report, thanks to an extensive interview with the VL-e director, among other things.

Finally, we feel that an important, yet difficult to measure milestone VL-e achieved is the fact that there is now an active Dutch e-Science community. This can be concluded from the many new initiatives (cf. Table 2a) in which VL-e partners play an important role. At the same time, this project has had an invaluable impact on the development of the e-Research domains. There are currently many e-Research application groups that completely depend on e-Science expertise, methodology, and infrastructure for many of their research and development activities.

The detailed innovation results are presented in the Appendix, Table 5.

Knowledge transfer by outflow of staff

Within the consortium 135 researchers were involved in VL-e.
Where are they, what do they do, now the project has ended?

	% of VL-e	% of category	number of persons
Stayed at their knowledge institution	60%	63%	81
Stayed at their industrial institute	4%	100%	5
Moved from knowledge institution to industry	25%	27%	34
Moved from knowledge institution to 'other'	8%	9%	11
Moved or stayed from 'other' to 'other'	1%	100%	2

Knowledge institutions: universities including academic hospitals, Nikhef, Amolf, CWI
Industry: Philips, IBM, Unilever etc.
Other: SARA or other non-research companies, retirement, unknown

More details in the Appendix, Table 6.

e-Science Research Center (e-SRC)

On the advice of the Commissie van Wijzen, a possible manner was sought during the last two years of how to embed the knowledge acquired in the VL-e project in the Dutch research environment. The VL-e board of directors proposed a plan to set up an e-Science Research Center (e-SRC). We are very pleased to find that this proposal has now been adopted by influential bodies such as NWO and SURF, so that also other parties – that were not represented within VL-e – will be able to play a role. For the realization of this initiative it is of the utmost importance that the knowledge acquired within VL-e will be given its due and that it will be used productively. Therefore, the traditionally monodiscipline-oriented funding and assessment of research must be broken through to enable the multidisciplinary approach (application, generic e-Science and infrastructure) that is essential to e-Science.

Chapter 3
Results of the VL-e Project

Following the VL-e functional model, we have classified the results of the VL-e project in two categories: the results from the virtual laboratory & web, grid, high performance computing layers being the generic reusable components, and the results enabling system level studies via the use of problem-solving environments.
After listing these two types of results we have placed them in the context of their contribution to innovation, dissemination & valorization, and/or embedding, based on their main impact on research, economics and/or society as indicated in the diagram presented later in this paragraph. The intersecting aspects of each result are thereafter described as examples in the associated paragraph.

Reusable e-Science implementations

AIDA Toolbox is a toolkit that enables interoperable access to knowledge resources with a set of web services. It has been used to implement browsing and searching of knowledge resources in a web application. It has a Taverna plugin, and a plugin for the vBrowser. Its services have also been used in the Food PSE 'Bitterbase' (SP2.2).

VLAM is a prototype grid-enabled workflow-management system, which aims to cover the entire lifecycle of scientific application workflows. The workflow engine of the WS-VLAM is implemented as a WSRF compliant web service, thus following the Execution Management Services described in the OGF documents (SP2.5).

WorkflowBus is a generic workflow execution layer in which a user can execute a workflow using a suitable (remote) engine, as well as execute different workflows plus connect them using different specific logic (SP2.5).

Ibis offers high-level communication primitives and programming environments for large-scale distributed systems like clusters, grids, desktop grids, and clouds (SP3.1).

JavaGAT offers transparent access to commonly used operations such as file I/O and job submissions, independent of the underlying middleware (SP3.1).

KOALA is a grid scheduler that includes innovative techniques such as co-allocation. Ibis and JavaGAT are used extensively in VL-e, for example, by VUmc, UvA and Amolf, and they are used by large international projects such as D-Grid, TeraGrid and ProActive (SP3.1).

vBrowser, the virtual resource browser, is a user-oriented exploration environment for browsing distributed grid infrastructures. Its generic e-Science approach has led to the adoption of this tool in several affiliated and external projects, such as NBIC and BiG Grid in the Netherlands, but also within EGEE NA4 (SP4.1).

VL-e Proof-of-Concept environment (VL-e PoC) is the environment where the developed software is applied in the PSEs of the application domains. Essentially, the PoC contains generic grid and networking software, as well as some generic application tools with a focus on high-performance grid computing. Here 'real-life' experimentation is done to evaluate the developed e-Science infrastructure (SP4.1, SP1.1, SP2.1, SP2.3, and SP2.4).

Problem-Solving Environments (PSEs)

Bitterbase predicts the bitter-response which is important in food production. The success of Bitterbase convinced Unilever to adopt the VL-e e-Science methodology throughout the company (SP1.2).

e-Science for food is an ontology-based toolkit with many software components oriented towards food production and safety. Elements are Tiffany, Emerging Risk Detection, and Experience Box (SP1.2).

VL for Medical Imaging is an e-Science environment for research in medical image analysis at academic hospitals that has been used autonomously by medical and image processing scientists and bioinformaticians at the AMC and at CREATIS (Lyon, FR) (SP1.3).

MIRAGE (Medical Image Analysis on Grid Enabler) is a grid job farming application for Medical Imaging, and allows easy access to the grid using JavaGat and runs on a PoC enabled 'staging computer' (SP1.3).

EcoGRID shows how species-observation records from a wide range of organizations are made available in the Netherlands and how this data is processed into scientific knowledge and predictions (SP1.4).

FlySafe is a bird-avoidance system for military aircrafts. A system of systems integrates radar measurements and models that predict high bird migration intensities to issue air safety warnings (SP1.4).

Virtual Lab for Bird-Migration Modeling allows data exploration with real-time access to military radars and weather radars, the weather forecasts, and GPS-loggers on individual birds (SP1.4).

RNA-GENIUS is a set of functional e-BioScience PSEs containing bioinformatics, semantic-web, workflow, plus grid-based methods and tools developed and used by the UvA transcriptome core facility for standardized and explorative transcriptome analysis in many life-science prokaryotic and eukaryotic domains (SP1.5).

KnowEx is a collaborative environment in which multidisciplinary groups can work together on a common project. It enables knowledge exchange and metadata management, supporting cross organization collaborations around large scale datasets (SP1.6).

ViroLab is a PSE consisting of a set of integrated components that, used together, form a distributed and collaborative space to support virologists, epidemiologists and clinicians investigating the HIV virus and the possibilities of treating HIV-positive patients (SP2.1).

e-BioLab is the connection to the VL-e of e-Bioscience PSEs via a real collaborative laboratory with high-resolution tiled display to support multidisciplinary e-Science project teams in their data visualization and interactions (SP1.5).

	Research	Economic	Societal
Innovation	vBrowser AIDA VLAM/WorkflowBus Ibis/JavaGAT/KOALA VL for bird migration RNA-GENIUS KnowEx	Bitterbase e-Science for Food	VL for medical imaging MIRAGE EcoGRID FlySafe
Dissemination & Valorization	vBrowser AIDA VLAM Ibis/JavaGAT/KOALA VL-e PoC RNA-GENIUS KnowEx Virolab e-BioLab	Bitterbase e-Science for Food	VL for medical imaging MIRAGE EcoGRID FlySafe Virolab
Embedding	AIDA JavaGAT VL-e PoC RNA-GENIUS KnowEx e-BioLab	Bitterbase e-Science for Food EcoGrid RNA-GENIUS KnowEx	VL for medical imaging EcoGRID FlySafe Virolab

This table presents the VL-e results placed in the context of their contribution to innovation, dissemination & valorization, and/or embedding, based on their main impact on research, economics, and/or society. The intersecting aspects of each result in this table will be described as examples in the associated paragraph.

Innovation

Research Innovation

The research innovation of the VL-e project can be deduced from the impressive number of 463 scientific publications, 251 international presentations, and 22 completed PhD theses (see Appendix, table 2). There are still five theses in the pipeline.
In addition to these traditional indicators for research and innovation success, the impact of the VL-e project is best described along the following lines: *structuring the field*, *lessons learned*, and *examples*.

Structuring the e-Science approach

The decisive success factor in our VL-e approach proved to be our strategy of development along the total-technology chain, because the weakest link determines the overall result.
As e-Science is an enabler for system-level research, integration at the various levels of the e-Science technology chain from e-Science applications via generic e-Science middleware to the e-Science infrastructure is a key issue. Therefore, research in all these domains is required with the specific objective to realize this integration. In the VL-e project we have approached e-Science via co-ordinated research in the development of e-Science concepts, methodology and technology. This has been accomplished by stimulating devel-

opments along the total-technology chain, from research in new applications in the various scientific domains, to new methods and techniques in computer science, as well as by breakthroughs in design and realization of the underlying infrastructures. We further structured the process by designing a functional model.

We started the VL-e project with a strong emphasis on end-user involvement, with almost half of the sub-programs application-oriented. In order to anchor all sub-projects in the project as a whole, each researcher's role in the multidisciplinary sub-projects had to be clearly defined. Also, the Virtual Lab e-Science Integration Team (VLeIT) was formed consisting of representatives of each VL-e layer. This team operated via a model of use cases (or user scenario's). This approach allowed high-level e-Science research to evolve to truly translational research where basic e-Science concepts, e-Science middleware, experiment design and instrumentation research are integrated.

"Dedicated team responsible for integration"

In addition, three other important e-Science research innovations were developed and tested:

- Reusability of e-Science components from generic and/or toolbox solutions
- The concept of coherent problem solving environments (PSEs) for each domain
- Functional workflows plus information and workflow-management systems.

These innovations are also a form of research dissemination and will be extensively discussed under Dissemination & Valorization.

"Re-usability via generic libraries, toolbox solutions and Problem Solving Environments"

e-Science lessons learned

As the project progressed, several important adaptations were made to the original structure and strategy of the VL-e project. The adaptive capability originating from the e-Science functional model allowed us to adjust our project based on the e-Science lessons learned. This adaptibility strongly added to the success of the overall project and led to a deep understanding of the e-Science concept as a whole.

Initially, there was ambiguity between the fundamental e-Science terms 'reusable' and 'generic'. Although reusable components are often considered generic, generic components are not necessarily reused. However, e-Science is only interesting and functional if it results in concepts, methods, tools and infrastructure that are functional to many users in their domains, i.e. the concept of reusable modules.

This is why during the project an additional layer was introduced in the overall VL-e structure: the domain-generic layer, in which e-Science elements are developed that (at least) can be reused by multiple end-users from different domains or the same domain in the case of different domain applications. The thorough embedding of such a domain generic layer in the VL-e middleware layer is a *conditio sine qua non* for system-level research. The embedding allows e-Science concepts to cross the boundaries of scientific domains resulting in new paradigms of natural research.

The total-technology chain approach applies to any application domain that aspires to an ambitious e-Science approach. It means that the whole chain of an experiment or study has to be embedded in an e-Science environment. This has led to the adaption of the concept of Problem-Solving Environments (PSEs). PSEs are coherent computer environments in which all e-Science functionality is present and functional in the context of a predefined experiment, study, or scientific subdomain. A method can be a standard data-analysis approach associated with a detector. In such a PSE, similar experiments using the same type of detector can easily be analyzed. PSEs can also be defined as e-Science environments for virtual communities where the relevant generic components are gathered and interconnected by some very domain-specific tools into a coherent environment for flexible experimentation. In both methods, reusability of components is optimally achieved, and integrating system level research is readily enabled in such a manner that different data, information and knowledge resources are efficiently utilized.

The standard data-analysis PSE concept led to the introduction and development of the e-Science workflow concept. These workflows can remove barriers in multi-disciplinary collaborations resulting in translational research that has the potential to drive the advancement in science. The consequence is that in any e-Science PSE, parts that deal with a standard data-acquisition or data-analysis chain are collected in an automated workflow which considerably enhances the reusability of the incorporated parts. Other highly functional workflows are those that are used to retrieve and integrate information from heterogeneous, worldwide resources. The flexible experimentation PSE concept required an additional concept: the workflow-management systems (WFMSs) which handle the assembly of workflows from generic modules and their employment in experimentation. Until now, the reusability promise of WFMS has been hampered by the lack of standardization, resulting in the situation that different scientific domains have adopted different WFMSs.

A new concept based on a type of meta-execution framework was able to overcome these standardization problems. This concept became more feasible after most WFMSs had adopted a service-oriented architecture. This led to the introduction and development of the workflow bus, an interactive-workflow environment which allows wrapping a number of popular and relative mature legacy SWMSs as federated components, and loosely couples them into one meta-workflow system. From the system level point of view, the partial functionality from different systems can be aggregated and complemented as one meta-system. The introduction of the workflow bus has dramatically increased the potential of VL-e to incorporate third party tools and software.

"Workflow based metadata management to capture data handling and analysis steps"

The development of a workflow based meta-data management system additionally facilitated cross-disciplinary collaborations around data intensive experimental sciences.

The adoption of a service-oriented architecture (SOA) has been an important advancement. A SOA is a flexible set of design principles used during the phases of system development and integration. At the technical level, it became possible to integrate grid services and web services which allowed access for web services to the grid, thus relieving the burden of designing specific grid services.
The integration of web-services in e-Science PSEs has proven to be successful as it allows a convenient interface to external resources. Furthermore, web services offer also an appealing concept for developing scientific applications, by providing interoperability and flexibility in a large scale distributed environment. The combination of concepts like PSE, workflow systems and the workflow bus together with SOA and web services were very useful to support applications in a distributed environment.

An interesting, yet unexpected innovative twist relates to the effect e-Science has on the research approach of a domain. For instance, in life sciences, an e-Science approach could in practice be applied to many system-level applications. Interestingly, by doing so, it became clear that the requirements for system-level e-Science experimentation demanded a thorough reconsideration of the design for experimentation in many life-sciences domains. So, e-Science requested for a reassessment of life-science system-level experimentation. Now, design for experimentation is construed as an important conceptual part of the e-Science approach. A successful system level experiment has demonstrated a requirement to be fully anchored in an e-Science environment from innovative experiment design, execution and analysis.

Examples

vBrowser
is an novel user-oriented exploration environment for research on a distributed grid infrastructure. It is an intuitive explorer-like grid browser for end-users with a Graphical User Interface to access their grid resources and it supports core grid file systems like Grid-FTP, SSH-FTP, SRM, LFC and SRB. Its logical tree-like structure can be customized by creating virtual folders and links to remote (grid) resources grouping relevant grid files and other resources together into a single browseable (vir-

tual) environment. More than a file browser, the vBrowser also serves as an extensible framework to build application-specific front-ends. Plug-ins for workflow management and grid job monitoring are the basis of the virtual lab for medical imaging adopted in the Netherlands and France.

AIDA
is a pioneering e-Science toolkit intended for groups of knowledge workers that co-operatively search, annotate, interpret, and enrich large collections of heterogeneous documents from diverse locations. AIDA entails a generic set of components that can perform a variety of tasks such as learning new pattern recognition models, performing specialized search on resource collections, and storing knowledge in a repository. W3C standards are used to make data accessible and manageable with semantic-web technologies such as OWL, RDF(S) and SKOS. AIDA is also based on Lucene and Sesame. Most components are available as web services and open source.

VLAM
is a workflow-management system for scientific applications designed to take advantage of the state of the art in grid-based systems. It uses an innovative service-oriented methodology to improve the management of workflow on geographically distributed resources. From usability point of view, VLAM allows a high abstraction from the complex underlying infrastructure. It enables scientists with various backgrounds to easily design, execute and monitor complex scientific workflows using functionalities like a Semantic Annotation of Workflows (SAW), a Hybrid-bAsed Match-Maker for E-science Resources (HAMMER) and an alternative data transport for web services.

WorkflowBus
is designed as a generic workflow execution layer in which a user can execute a workflow description using a suitable engine (which might be installed remotely), execute different workflows, and connect them using specific logic.

Ibis/JavaGAT/KOALA
is a novel integrated programming environment that eases development and deployment of high-performance grid applications. The Ibis system offers high-level communication primitives for large-scale distributed systems like clusters, grids, desktop grids and clouds. Also, it addresses the problems of heterogeneity, fault tolerance and connectivity which these systems often have. The JavaGAT offers transparent access to commonly used operations such as file I/O and job submissions, independent of the underlying middleware. The KOALA grid scheduler includes innovative techniques such as co-allocation.

VL for bird migration
is a landmark virtual lab that facilitates integrated use of measurement systems and models. It has a real time connection to military and weather radars in the Netherlands and Belgium, local ship radars and antenna systems used for the data acquisition of GPS-loggers on individual birds. In addition, dynamic data of the weather forecast and landscape properties are also available. The VL facilitates data post-processing and data storage, visualization and data exploration, statistical analysis and spatially explicit modeling of bird migration.

RNA-GENIUS
is a set of e-BioScience PSEs that is quite innovative in the life sciences, because most life-science solutions were isolated bioinformatics tools. The development focussed on reusable components for gene-expression analysis oriented at design for experimentation, data handling, data analysis and data interpretation. About fifteen e-BioScience tools were developed and used by at least eight PSEs for array design, result management etc. e-Science methods like semantic web, workflows and grid computing were used to make the PSEs coherent and end-user friendly.

KnowEx
is a modern collaborative environment that enables an effective collaboration between research groups from various disciplines in virtual organizations (VOs). It can be used as a powerful information manager and information organizer to help researchers keep track of their data, provide relevant retrieval, capture and store experiment data, implement dynamic raw data files, individual bookmark collections for personalized access to user's recurring topics. It is used to design and start simulation oriented workflows on distributed computer systems. It also provides data integrity and security of sensitive experimental data. The crucial difference from existing information man-

agement systems is that the stored meta-data is based on actual experimental workflows.

Economic Innovation

For an enabling-science research project such as VL-e, the direct economic innovation is rather difficult to quantify. In essence, by enabling 'better' science via a pioneering e-Science approach, an indirect economic benefit is clearly noticeable. The direct benefit is achieved through the advancements in the application domains. As such, the economic innovation is different for each application. We feel that our project certainly has had a noticeable impact on economic innovation in the ICT industry. Examples are IBM, Phillips Medical Research, LogicaCMG. In the end-user industry, there is a clear economic innovation at the companies that participated in VL-e and adopted concepts and tools in their R&D. An example is Unilever. Finally, for the applications without direct industry involvement, such as academic organizations, economic innovation has led to the start of several e-Science based spin-off companies.
Examples of VL-e results that have contributed to economic innovation are:

Bitterbase
Bitter is a complex taste. The human tongue has 23 different receptors to recognize bitter substances. For food designers, this is a complicating factor. Most people do not like bitter components in their food. It is therefore important to be able to predict which molecules induce a bitter taste, and which do not. The Bitterbase is a Unilever database that can be accessed using web services. With components from the AIDA toolbox, a food researcher anywhere in the Unilever company can identify whether a specific molecule is bitter, how it can be masked, or by what component it may be replaced. The Bitterbase-services have reduced the number of expensive food trials.

e-Science for Food
This toolbox has led to several economic innovations. Tiffany, for example, is a web-based research management system that enables the systematic sharing and discovery of knowledge in experimental food research. It allows researchers to formulate research questions and progressively add information about related research using an ontology and organized research questions. Faceted browsing and tag-based search are used to better disclose the captured knowledge. Tiffany enables a better sharing of knowledge, thus creating a more efficient research process for academic and industrial participants, e.g. Unilever, Royal Friesland Campina, Vion, CSM, DSM.

Societal Innovation

The same arguments that apply to economic innovation go for societal innovation. The primary difference is the character of the application domain. Whilst the focus in economic innovation is most often industrial oriented innovation, societal innovation aims at not-for-profit organizations such as, (academic) medical hospitals and biodiversity foundations. Several of such organizations were directly involved in VL-e, so the impact of the results on societal innovation was considerable as can be read in the examples below.

Examples of VL-e results that have contributed to societal innovation are:

VL for medical imaging
This is a PSE for the logistics problems experienced by neuroscientists sharing the 3.0 Tesla Philips MRI scanning facilities of the UvA Academic Medical Centre (AMC). It provides a computational infrastructure to facilitate the storage, analysis and sharing of fMRI and DTI data. The basis of this infrastructure is the VL-e PoC environment. Its main components are: acquisition devices located at the AMC; computational resources (data storage and computing) provided by the Dutch grid; low-level services provided by standard grid middleware; generic VL services (workflow and front-end); and application-specific services for data acquisition, storage, analysis and access.

MIRAGE
Mirage is a PSE for the logistics problems experienced by neuroscientists who produce large numbers of similar data sets, such as MRI scans, that require hundreds, thousands, or tens of thousands of hours of computational time. Mirage has already been used to analyze data sets, which has resulted in several publications. With many more research projects underway, Mirage has be-

come an essential part of the VUmc's research infrastructure. Mirage is built on JavaGAT which was also developed as part of the VL-e project. Both Mirage and JavaGAT are built on the VL-e PoC. Mirage's strength lies in hiding most of the details of the grid usage from the neuroscientists thus making the interface very simple and suited to the set of skills of neuroscientists working on computational projects.

"VL-e knowledge EcoGRID is important for ESFRI type projects like LifeWatch"

EcoGRID
Although the Netherlands has the densest network of biodiversity observations (citizen science), ecologists generally work with small data sets. EcoGRID is the first PSE in the world that combines distributed data (now almost 40M records) and visualization and analysis tools in a truly operational system. It is now facilitating innovative ecological research: e-Ecology. EcoGRID is one of the examples that play an important role in the development of the ESFRI-LifeWatch infrastructure.

FlySafe
FlySafe is an ecological application that enhances the safety of military flights by warning for high bird migration activities. In ecological research, typically hundreds or thousands of data points are used. In FlySafe, however, distributed measurement systems produce continuously over 2MB of data per second. Data are processed on site, post-processed and stored in a central facility and made available for thirteen international research partners through a virtual lab. Although comparable systems are used in the field of meteorology, FlySafe is unique in ecology.

Dissemination & Valorization

In the final stage of the project, dissemination and valorization became increasingly important and numerous valorization options emerged. Valorization can be achieved via commercial options or by generating societal value as in the Flysafe example. This example shows that the most direct valorization potential lies in the application domains. Even so, these achievements would not have been possible without the efforts of all the other scientists giving contributions to numerous other elements of the project: the VL layer, the grid/networking layer, and the e-Science experimentation environments.

As the VL-e project encompassed a wide range of partners, from universities to industry, all along the e-Science technology chain, the dissemination and valorization has been quite effective and extensive. Dissemination of the e-Science concepts and research results has been achieved via the impressive numbers of international publications (463), presentations (251), and communication to the general public. Dissemination was also furthered by the huge groups of directly involved researchers, industrial partners, societal users, and students. Valorization has been accomplished via the implementation of the e-Science results in the research of academic organizations, the daily practice of academic hospitals, the R&D of industry, and via the start of several spin-off companies. Internationally, valorization has been realized by participation in standardization platforms and other international activities. Of course, direct involvement of VL-e partners in numerous new national and international initiatives is an important indicator for the considerable dissemination and valorization success of the project. In addition to these familiar paths of dissemination and valorization we will here describe some novel ways that we specifically defined for e-Science.

Research Dissemination & Valorization

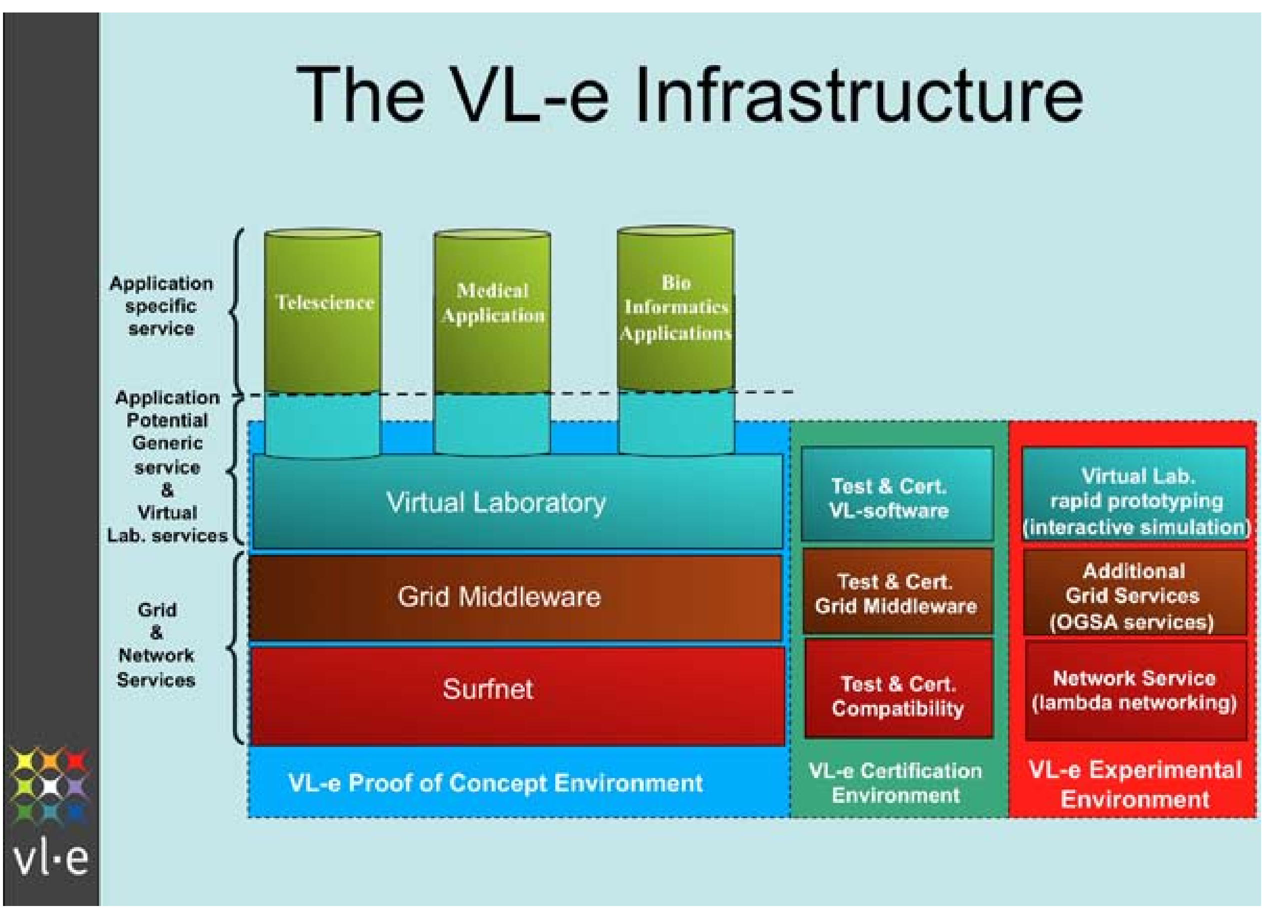
For effective dissemination and valorization of our VL-e results, a high-quality and dependable e-Science environment is an absolute prerequisite. No end-user in any application domain will base its research, R&D, or production approach on an environment that is not reliable. This is especially true for end-users that are new to the e-Science approach. That is why we constructed two complementary experimentation environments:

- The VL-e Rapid-prototyping environment (RP), in which scientists can freely develop their software without disturbing the functionality needed by other scientists. It is more or less a playpen for VL-e lower and middle layer e-Science research.
- The VL-e Proof-of-Concept environment (PoC) in which the developed software is applied in the PSEs of the application domains. Here 're-al-life' experimentation is done to evaluate the e-Science infrastructure.

"Rapid Prototyping for software development"

In addition, a transition environment between these two experimentation environments was installed: the Certification environment. This environment has been created to facilitate the transition of developed software from the RP environment to the PoC environment. Testing and certification are the main functionalities of the certification environment, before a new software tool may become an operational part of a larger software environment. The certification process is necessary to determine how a tool fits with all the other pieces of the PoC environment. Together, these environments form a natural flow from the development to the production-ready stage. Production itself is not part of the VL-e project.

"In Proof of Concept, software applied into various domains"



The VL-e Integration Team (VLeIT) consisting of representatives from the RP, PoC, Certification and Application environments and chaired by the VL-e scientific director, controlled this flow. The VLeIT took care of the PoC software distributions, coordinated the necessary support and guarded the quality and certification of new software releases. For planning and prioritizing, VLeIT took into account the wishes of the VL-e application groups and the available programming manpower.
The VL-e Proof-of-Concept environment, including its diversity of resources, has also led to the development of innovative methods for assessing reliability, stability and fault-tolerance of infrastructures. Automated, agent-based systems can now analyze complex heterogeneous systems of autonomously operating resources, a very common element in today's industry, centered on 'software as a service' and service-oriented architectures.

Dissemination and valorization aspects of examples of successful PoC tools are:

vBrowser
vBrowser and VL-e Toolkit combined form a user-oriented exploration environment in the VL-e PoC for research on a distributed grid infrastructure. This open source product is extensively used also for scientific and societal applications, such as in neurology and biomedicine. The generic e-Science approach has led to the adoption of this tool in several affiliated and external projects and provides a way of introducing e-Science in an understandable way to third parties through the VL-e series of grid tutorials. Since it is written in Java no additional (grid) software is needed and it can be easily downloaded and deployed. The vBrowser has also been rolled out to the EU community through the EU-COMPUTIS project, where it is used to browse imzML data, a new imaging standard developed with the Human Proteome Organization.

AIDA
AIDA web services were deployed in 2006 and an ontology enrichment application of AIDA services was demonstrated at ISMB/ECCB in Vienna in 2007. Currently, knowledge resources such as (medical) ontologies and terminologies are available from configurable applications built on several different platforms, including web browsers and plug-ins for Taverna and vBrowser. This has resulted in influence of AIDA on the uptake of semantic web technologies like Simple Knowledge Organization System (SKOS), as well as on the development of applications for the Concept Web Alliance (CWA). Specific terminologies that have been made available through AIDA services such as SNOMED-CT, MeSH, and UMLS have acquired new impetus through accessibility on the Semantic Web. Increased interoperability will also make all ontologies available from the U.S. National Center for Biomedical Ontologies accessible via the AIDA Repository Browser in 2010.

VLAM
VLAM is being used in a number of applications within the VL-e consortium. Other projects and groups throughout the Netherlands consider using WS-VLAM to develop their applications, such as Leiden UMC, Biomedical engineering Cardiovascular Biomechanics TUE, and in the NBIC project BioRange. VLAM has been adopted by the Technische Universität München as a platform to develop a new framework to facilitate the execution, monitoring and management of CSE simulations in computational grids.

Ibis/JavaGAT/KOALA
Ibis and JavaGAT are used extensively in VL-e, for example by VUmc, UvA and Amolf, and also by large international projects such as D-Grid, Tera Grid, and ProActive. KOALA is used in a number of real use cases (e.g., in medicine, physics, and bio-informatics). JavaGAT which hides the different underlying grid middleware is used by AstroGrid (the German Astronomy Community Grid), TextGrid (for humanities), and AeroGrid (a grid for the aerospace research community). In February 2010, the Max Planck Institute organized a JavaGAT workshop in Potsdam.

VL-e PoC
The 34 scientific communities that now conduct their research on the VL-e PoC infrastructure include not only traditional domains such as physics and computational science but also life sciences, arts and humanities that are newcomers to e-Science. The diverse and extensive use of the BiG Grid infrastructure (see later) is a direct result of the foundational VL-e project.

The formation of domain-generic PSEs that connected the applications to the PoC environment was of utmost importance to end-user communities because the PSEs play an important role in dissemination of research results. Project members in the VL-e applications were the first to learn about and adopt the parts of the e-Science technology chain that they consider important for their activities in domain-generic PSEs. They were therefore in a perfect position to demonstrate and advocate the potential of this technology to enhance their specific research discipline. The general outcome for all e-Science user groups was that they got 'hooked'. Every improvement regarding their PSE allowed them to do better experiments or offered a better service to end-users. Of course, this meant that the methods and tools they used to employ had become 'old-fashioned'. In science, researchers want to do the best experiments they can, so they will never use outdated know-how. This is why the e-Science paradigm spreads like an ink-stain, be it with different pace, in the various application domains. This also explains the ongoing and growing commitment of the application domains to the development of VL-e and maintenance of its results, because they become increasingly aware of its indispensable value to their own research and services. Because of the close collaboration within the VL-e consortium between R&D scientists from the private and the public domain, we thus have created a natural flow of e-Science know-how.

Dissemination and valorization aspects of examples of successful PSEs are:

RNA-GENIUS
Because these PSEs were developed at the MicroArray Department & Integrative Bioinformatics Unit (MAD-IBU) of the UvA, dissemination & valorization are quite well organized. The MAD-IBU acts as a national transcriptomics service provider. All PSE parts were immediately applied into the daily bioinformatics service & support of IBU. Not only IBU distributed these methods directly to life-sciences end-users, but also made them aware and, by now, dependent on these VL-e PSEs.

KnowEx
KnowEx has been evaluated in two EU-projects to share round robin data. It facilitates a bilateral collaboration between John Hopkins medical school and the FOM-institute for Atomic and Molecular Physics targeting the investigation of the molecular basis of breast cancer. Its use is also evaluated in biomarker research where from KnowEx various grid tasks are initiated with the purpose to speed up the processing of large-scale proteomics datasets. Pharmaceutical companies utilize KnowEx for collaborative pharmacokinetic studies with FOM-Amolf. A business plan for a spin-off company based on KnowEx has been developed and a FOM valorization grant has recently been awarded to further develop KnowEx into a market-ready product.

ViroLab
The ViroLab PSE is a set of integrated components that, used together, form a distributed and collaborative space to support virologists, epidemiologists and clinicians investigating the HIV virus and the possibilities of treating HIV-positive patients. Virolab consists of textmining tools to discover drug-resistance patterns and a suite of simulations tools to relate virus mutations to disease progression and spreading (drug-drug interaction, sexual network simulators). Virolab is in use in seven hospitals and currently under evaluation for commercialization.

e-BioLab
Last but not least in the list of successful dissemination and valorization tools is the new e-Science Laboratory e-BioLab. Many e-BioLab demonstrations of VL-e tools and PSEs were given to a wide range of audiences, such as students, industry representatives, life-science research groups, university boards, etc. The many life scientists that have used the prototype e-BioLab were very enthusiastic about the positive effects on the exploring and creative parts of collaborative e-BioScience experimentation. Several new e-BioLabs at other universities are being set up thanks to our VL-e prototype e-BioLab.

Economic impact		
	Estimated costs in M€	
	From VL-e	From scratch
Problem-Solving Environment: Tiffany, VLeMED EcoGRID, FlySafe, RNA-GENIUS, DUTELLA	8	20
Virtual Lab tools: AID, vBrowser, VLAM, IBIS, VL-e toolkit	3	10
PoC and RP: Contributions to BiG Grid and DAS	5	15
total	16	45
Extra spill-over from EcoGRID	20	

Economic Dissemination & Valorization

The key concept of reusable component in the VL-e project has proven to be an important economic factor in the dissemination and valorization of the results of this project. Moreover, the fact that VL-e developed a large number of research prototypes helped considerably in disseminating our results and their economic uptake.

The table illustrates that in some of the cases we built applications and tools, which, if you had to build them from scratch, would already have cost more than our total VL-e project budget. One example is the biodiversity group that built in VL-e a prototype PSE for ecological analysis, *EcoGRID*. Municipalities, project developers and building companies in the Netherlands, are now actively using this PSE. Before they can start construction activities, they have to make an inventory of rare plants and animals living on the land where they wish to build. They can use shared data and methodologies or tools to find out what species there are and whether counter measures need to be taken. Currently about 50 people are working on this project with a total budget of M€ 20 (presented as the spill-over explicitly named in this economic impact table because the roots all originate in VL-e.)

"Building some applications from scratch would have cost more than total VL-e budget"

The economic dissemination and valorization aspects of some specific VL-e examples are:

Unilever is – as a result of the VL-e project – currently using *Bitterbase* worldwide to predict the for food production important bitter-response. This will reduce the screening and evaluation processes dramatically and enhance the chances of finding successful food components. Given the size of Unilever as one of the biggest food companies worldwide, the economic dissemination and valorization impact is considerable. The resounding success of the Bitterbase prototype will surely lead to the development of other e-Science information integrating applications in the food industry.

The *e-Science for Food* toolbox has been applied in many important and economically relevant projects in the food domain. Examples are Tiffany, the research management system at Top Institute Food & Nutrition; Emerging Risks Detection System, a web-based rule-based system for detecting world-wide incidents that may have an impact on food safety of a product in the Netherlands; Database Healthy Components, a prototype web-based knowledge system set up for the South-Netherlands in which health claims of vegetables and fruit are combined with preparation influences; Experience Box, an ontology-based search system in which lessons learned from projects are disclosed for the general public; MARVIN, a knowledge-based computer vision system, in which ontologies and formal rules capture expert knowledge to automatically inspect tomato seedlings; and so on.

Finally, VL-e members are actively involved in several techno starters that are in various stages of maturity, including Virology, Gridwise, Foldyne, Personal-Space Technologies (PS-tech), Life Sciences Pioneers, Culios, KnowEx, and Treparel.

Societal Dissemination & Valorization

We believe that the results presented here and above speak for themselves with respect to societal dissemination and valorization. It is a further illustration of the fact that e-Science is an important enabler for system-level science. Because of the complexity of society nowadays, it presents numerous system-level problems for which e-Science solutions are a key factor as we illustrate here. As with economic, societal dissemination and valorization depends heavily on the success of functional prototypes. The VL-e project has produced some very important societal relevant e-Science prototypes, which resulted in an excellent societal dissemination and valorization.

The societal dissemination & valorization aspects of some specific VL-e examples are:

VL for medical imaging/MIRAGE has achieved that medical and image processing scientists and bioinformaticians of the AMC regularly use the Dutch Life Science Grid. This VL is currently adopted in France by the CREATIS Lyon and NeuroLOG projects. Also, with Philips and the university hospitals AMC and VUmc, medical imaging PSEs were realized for: enhanced visualization of CT Angiography of the head supporting first-aid care in the AMC; epilepsy treatment planning by MEG data processing at the VUmc; and assessment of brain tissue damage in naive ecstasy users by MRI techniques.

EcoGRID is a distributed system for biodiversity information. The first step in VL-e was to setup a virtual lab and facilitate 10 organizations by redesigning their data models such that they could co-operate over the boundaries of their organizations. After two year, the concept was embraced by the Dutch Data-authority for Nature (Gegevensautoriteit Natuur), and further developed to facilitate municipalities, provinces, and (non-) governmental organizations like Staatsbosbeheer, Rijkswaterstaat, Natuurmonumenten, and so on.

FlySafe is a Bird Avoidance System that enhances the safety of military flights. It started as a national initiative with the Dutch Air Force and builds upon the expertise gained in the 'Virtual Lab for Bird Migration Modeling' and 'EcoGRID'. Thereupon, the concept was embraced by ESA and was further extended to an international System of Systems, being selected as the first precursor after launching the ESA Integrated Application Program.

Virolab is an infectious disease PSE that support medical doctors in deciding on what drug regime to apply to a HIV patient at a certain point in time. A patent for the rule-based component has been granted in the USA and efforts are underway to commercialize parts of ViroLab.

Embedding

Embedding of the obtained results will be achieved by different means, depending on the area it concerns. But in al cases, system-level research or applications heavily rely on the quality and stability of the infrastructure. With every newly involved application, the (financial) responsibility grows. Therefore, the support of the ever-expanding e-Science technology chain must be institutionalized. Here we will illustrate the way in which we have contributed towards solving this problem in the Netherlands and at the same time making our international position far more competitive.

To safeguard and maximally capitalize the investments in e-Science, we feel that finding structural funding for e-Science research and the associated ICT infrastructure, as well as setting up an adequate organization that ensures a successful integration of all necessary components of the e-Science chain have to be the next steps. Therefore, already during the VL-e project, VL-e members have made major contributions to several initiatives regarding e-Science and its necessary infrastructure.

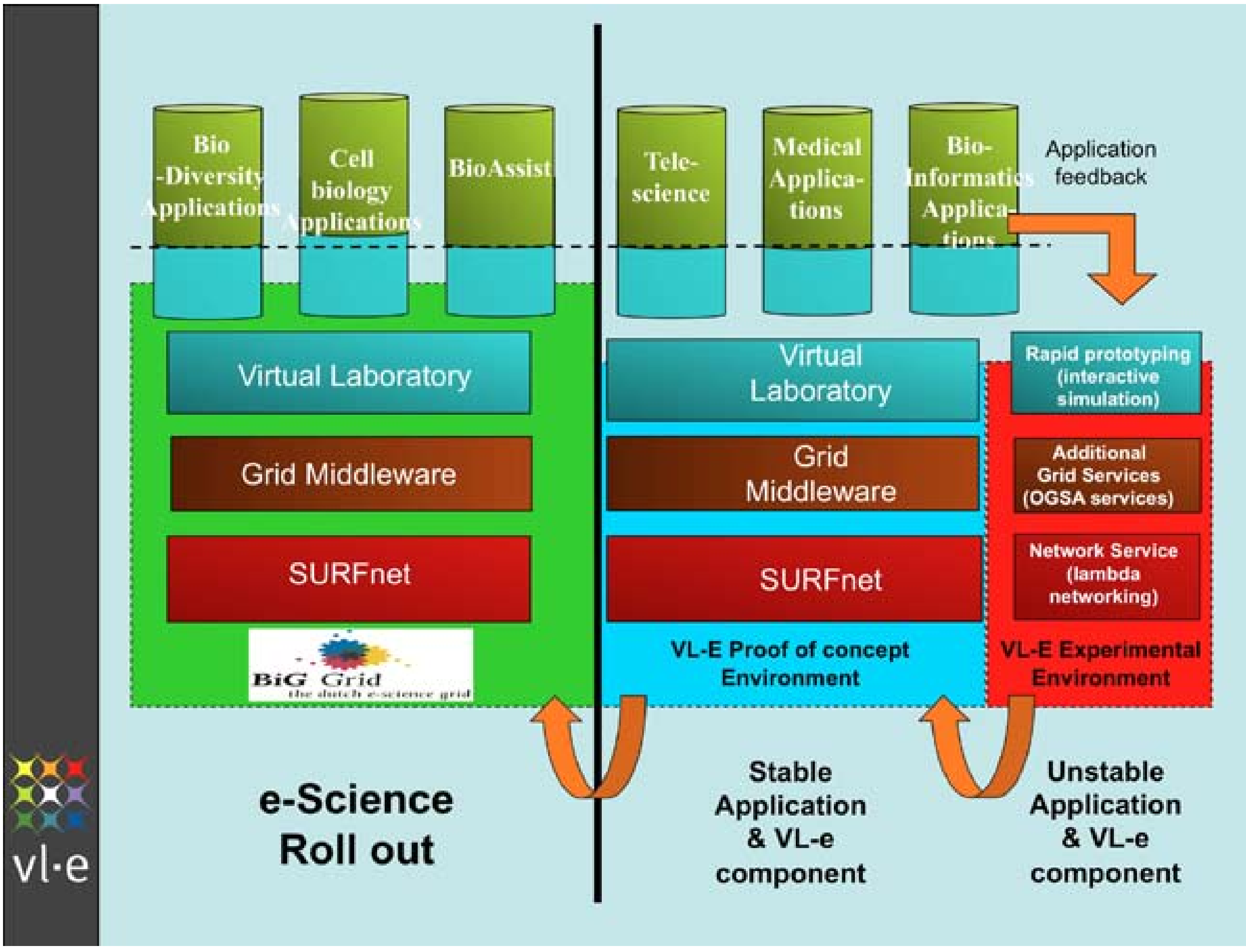
SURF has coordinated a proposal entitled 'Towards a national ICT research infrastructure'' for the national roadmap committee for large-scale research facilities (committee van Velzen), which was approved by this committee and put on the so-called ESFRI list. VL-e has played an important role in the design of this proposal. ICTregie has written a report entitled 'Towards a competitive ICT infrastructure for scientific research in the Netherlands', which was submitted to minister Plasterk in December 2008. The VL-e directors have made a significant contribution to this report, especially in relation to future research in the field of e-Science. In this report the minister of OCW, the secretary of state of EZ, and the general director of NWO were advised to create structural funds for the described nationwide ICT infrastructure for scientific research. Recently, VL-e members have been the main architects of a new e-Science project proposal M€ 32 (2010-2015) via the ICTregie program COMMIT, which has been submitted for funding in the context of FES 2009. Now that we – via the realization of various prototypes – have demonstrated the potential of the e-Science, we want to create real-life experimentation environments for some carefully selected domains, such as food, and biobanking in combination with whole human genome sequencing.

The Netherlands has an excellent infrastructure position along the total e-Science technology chain for system-level studies. This chain is based on networking (SURF) and the grid. As a first step towards structural funding, VL-e considerably contributed to the proposal that resulted in a four year funding of a nation-wide grid based ICT research infrastructure called BiG Grid as illustrated below. BiG Grid aims to be the nationwide production grid for many data-intensive sciences, including particle physics, astronomy, and the life sciences.

"BiG Grid rolls out VL-e products and methodology"

BiG Grid is a separate project organization based on a proposal that was submitted by Nikhef (the Dutch institute for subatomic physics), the Netherlands Bio-Informatics Center (NBIC) and the National Computing Facilities foundation (NCF). The Dutch National Computer Center SARA was brought on board as an operational partner providing grid operations and support to the BiG Grid communities. One major part of the BiG Grid activities is the running of the Dutch Tier I facility for LHC, which is operated jointly by Nikhef and SARA.

The 34 scientific communities that now conduct their research on the VL-e PoC infrastructure include not only traditional domains such as physics and computational science but also life sciences, arts and humanities that are newcomers to e-Science. The diverse and extensive use of the BiG Grid infrastructure is a direct result of the foundational VL-e project. The VL-e partners Philips Research, SARA and Nikhef, joined in BiG Grid by the computing centre from the RUG. The PoC convinced the Philips ICT support group (COS) to become heavily involved in the VL-e and BiG Grid projects. In the future, we also envision dissemination to other projects, like Parelsnoer (via NBIC) and LOFAR.



New tools are first tested in an experimentation environment for computer scientists (on the right). Successful tools can be migrated to the proof-of-concept environment (middle) after thorough testing and validation. Both the RP and PoC are experimentation environments in VL-e. Finally, software can be rolled out on the BiG Grid infrastructure which supports production-like deployment.

After the midterm review and the discussions with – among others – the Commissie van Wijzen, the VL-e directors came to the conclusion that the foundation of a national e-Science Research Center (e-SRC) as an important part of a national ICT infrastructure for scientific research was necessary. The e-SRC will be, among others, a national capacity and expertise center for e-Science, and should concentrate on research in e-Science, e-Science applications and e-Science infrastructures. Although this initiative is still in an early stage, initial financial commitment from SURF and NWO has already been obtained.

Research Embedding

Besides the above mentioned large embedding initiatives, individual VL-e partners have been active to realize research embedding for their specific VL-e results. For instance, many other NWO projects have been initiated by VL-e in programs such as GLANCE, VIEW and STARE. Several of these projects are collaborations with other Bsik programs, such as the Netherlands Bioinformatics Centre (NBIC), MultimediaN, and Gigaport-NG. Other examples include collaborative projects with university hospitals that will be based on VL-e PoC technology, the e-BioLab, and several European projects. Virolab, for example, is an important eHealth project that was funded by the EU as a result of experience gained in VL-e. In all these cases, the know-how, i.e. knowledge, expertise, tools and infrastructure, developed within VL-e, plays a central role. Also, VL-e members, together with scientists from the University of Leiden and ASTRON in Dwingeloo have obtained NWO funding for the DAS-4 proposal 'DAS-4: prototyping future computing infrastructures'. Within this project, a hybrid system will be set up in 2010 consisting of six clusters with various add-ons (GPUs, FPGAs, Cells, etc). The result will be an excellent computer science test bed that can be used to investigate how e-Science applications can be mapped efficiently onto future hybrid systems. The DAS-4 project enables comparison of different architectures to find the most energy saving handling of an application, thus contributing to prototyping a greener infrastructure.

The research embedding aspects of some specific VL-e examples are:

AIDA
AIDA researchers will continue to participate in the international standardization efforts, such as the Open Grid Forum, the World Wide Web Consortium (W3C) with the BioRDF taskforce and Terminology taskforce of the W3C HCLS IG, where the AIDA Toolkit has been used to construct technology demonstrations and prototypes. This was the case for NBIC/CWA as well. The LUMC has adopted AIDA technologies in their Bioinformatics Center of Expertise. AIDA researchers have also joined the Swiss Institute of Bioinformatics in a project to make Uniprot, one of the most important bioinformatics resources, available on the Semantic Web.

JavaGAT
The JavaGAT is now open-source software with an extensive user community that maintains the software collectively, in open-source fashion. The JavaGAT is also used as a foundation for the VU implementation of the OGF standard SAGA. The Ibis software is embedded in a large (M€ 1) VU-ERC project and in many other PhD projects. The group at the VU will therefore also continue supporting the Ibis and JavaGAT software for the foreseeable future.

VL-e PoC
First and foremost, the PoC has been solidified as part of the BiG Grid infrastructure. BiG Grid, implementing the Dutch e-Science grid, has adopted not only the common PoC software distribution as the basis for application support in the Netherlands, but also the validation methodology and the lessons learned during the execution of VL-e are an integral part in the design and execution of this four-year project.

RNA-GENIUS
Because this subprogram was done at the MicroArray Department & Integrative Bioinformatics Unit (MAD-IBU) of the UvA, embedding is automatic, also because the MAD-IBU acts as a national transcriptomics service provider. This e-BioScience research will be continued in approved grants from the NBIC BioRange and BioAssist programs, the pending ICTregie proposal COMMIT, and the future e-Science Research Center. The MAD-IBU strives to become the National Transcriptomics Center, which would ensure nationwide embedding of the VL-e results in the majority of life-science domains.

KnowEx
Several visualization approaches developed in VL-e have become an integral part of Amolf's molecular imaging facility. The data storage and metadata management approaches have been embedded in the core facility. The vBrowser is used to browse molecular imaging data. Embedding for this knowledge exchange system is anticipated via a potential spin-off company that has been developed and for which market development a FOM valorization grant has been awarded.

e-BioLab
The prototype e-BioLab has now its permanent place in the new UvA-FNWI building at Amsterdam Science Park. Additionally, the NBIC BioRange and BioAssist programs have granted continuation funding for further research and exploration of the e-BioLab concept. It is also a topic in the ICTregie proposal COMMIT and will certainly be a part of the future e-Science Research Center. The growing number of e-BioLabs at other institutes will also contribute to the embedding of this concept as more end-users will start to integrate e-BioLab meetings in their multidisciplinary research.

Economic Embedding

An important form of economic embedding of the VL-e results has been the infrastructure established via some of the large projects mentioned above. An example is the embedding of the VL-e research at Logica. During the VL-e project an expertise team called 'Collaborative Network Solutions' (CNS) was created inside Logica. The team focused on the subject of monitoring volatile networks using intelligent agents. The system administrators who wanted to get supported in their daily task of operational management, often administrators from multiple organizational domains, had to collaborate extensively in order to perform their analysis and had to combine information from monitor applications and system log-files. The VL-e distributed collaboration approach allowed smooth service problem resolution and presentation of the total (grid) infrastructure overview to end-users.
Inside Logica, the VL-e project contributed to shared interest with the Java software practice, which brought practical steps towards solutions for customer problems that can be expected in the upcoming area of cloud computing.
Furthermore, the economic embedding of several VL-e applications has been quite promising, as can be read from the collaborations with industry, the number of techno starters and the active participation in techno platforms. The sheer size of some end-user groups also holds a good perspective for successful economic embedding.

The economic embedding aspects of some specific VL-e examples are:

Bitterbase
Bitterbase is one of the best VL-e examples of economic embedding. The e-Science prototype for integration of food information has been so successful that it quickly has been adopted worldwide by Unilever R&D. The prototype is currently being developed towards a production version for the whole company. Moreover, this also implies that the e-Science concept has been adopted by Unilever, which means constant involvement of Unilever in new e-Science initiatives.

e-Science for Food
In addition to the hardware infrastructure, a knowledge infrastructure in the food domain is extremely important. Therefore, a collection of frequently used food vocabularies and a library of ontology modules has been created in the food domain. With these resources available from one location, food research can optimally reuse existing knowledge and save a lot of time, effort and money by not having to reinvent the wheel over and over again.

EcoGRID
The embedding of EcoGRID is long-time guaranteed by the European and Dutch legislation. In the course of the VL-e project, the foundations 'National Database for Flora and Fauna (NDFF)' and 'Gegevensautoriteit Natuur (GAN)' were established. The two foundations are responsible for keeping the available information up to date. EcoGRID is the prototype technical infrastructure that is now being rolled out by Geodan for use in science and society.

RNA-GENIUS
These PSEs are developed by the MAD-IBU, which acts as a semi-commercial national transcriptomics service provider. Thus, all PSEs are used to provide paid support for transcriptomics analyses. Currently there is an ongoing initiative to spin out the bioinformatics analyses that are based on these VL-e PSEs into a techno starter named Life Science Pioneers.

KnowEx
Economic embedding of KnowEx is undertaken by creating a spin-off company. The FOM valorization grant that has been awarded to further develop KnowEx into a market ready product will certainly have a positive effect on this effort.

Societal Embedding

Also, societal embedding of the VL-e results has been accomplished by the infrastructure established via some of the large projects mentioned above.
Furthermore, strong progress has been made for societal embedding of the VL-e results as can be read from the substantial number of non-commercial applications, as well as from the substantial number and size of the various societal user groups.

The societal embedding aspects of some specific VL-e examples are:

VL for medical imaging/MIRAGE
In 2008 the fMRI user group was consolidated at the AMC. It meets on a monthly basis to discuss fMRI data processing in neuroscience and cognition research. The IT-infrastructure at the AMC is now firmly established, with computing resources and support. Also, an e-bioscience research group has been formed at the Bioinformatics Laboratory of the AMC. This group has recently received a grant of over k€ 500 from the AMC ICT innovation fund and the BiG Grid project to further develop the VL for medical imaging and adapt it for other (genomics) applications.

VUmc's involvement in VL-e and its work on Mirage has led to VUmc's involvement in the EU FP7 neuGRID project. neuGRID (neuGRID.eu) is a co-operation between several universities and companies spread across six EU countries tasked with implementing a GRID based infrastructure for EU neuroscience researchers. Two years into the three-year project, a database intended to hold MRI and clinical data from a wide range of neuroscience studies is nearing its final form. A sophisticated user interface, designed after extensive consultation with users, is also nearing completion. In addition, three interoperating clusters have been installed across Europe for the neuGRID software to run on. Initial projects are being run on the infrastructure to assess its performance and usability. A post project plan, to ensure neuGRID's wide spread availability after the completion of the project, is in the works.

"VL-e partners involved in FP7 projects provide international dissemination"

EcoGRID
On the one hand EcoGRID plays an important role in facilitating citizen science with respect to field observations of biodiversity. On the other hand it facilitates decision making in the domain of city planning with respect to flora and fauna legislation. EcoGRID also provides the information needed for improving the effectiveness of nature conservation measures.

FlySafe
FlySafe increases the safety of military flights. It contributes to saving airplanes and human lives. A preliminary study is currently being carried out to identify the possibilities to use comparable information systems for increasing flight safety in civil aviation. Beyond the safety aspect, FlySafe has also facilitated multidisciplinary scientific co-operation, which has brought the Netherlands to the forefront of bird migration research.

Virolab
Virolab is in use in seven European hospitals as a drug-ranking PSE for HIV-1 infected patients.

Chapter 4
Synergy with other Bsik Projects

We have intensively collaborated with a considerable number of other Bsik projects.

- We have an extensive formal collaboration with NBIC, as described at several places in this document. In particular, NBIC uses the VL-e PoC. Currently an M€ 4 project is installed which specifically aims at ensuring the transfer of VL-e e-Science expertise to the life-sciences domain via a partnership of NBIC and BigGrid.
- We collaborate with Gigaport-NG in the DAS-3 and DAS-4 projects and in the NWO/GLANCE funded StarPlane project. We investigated how light path technology could be used for data-intensive applications like video or image processing and distributed model checking. The KOALA scheduler has been used extensively on DAS-3 in the I-Share BSIK project for measurements and simulations of peer-to-peer systems.
- We collaborated with MultimediaN in an NWO/GLANCE funded project about multimedia content analysis on a grid. The board (CvB) of the VU Amsterdam has provided a research grant of over M€ 1 to Dr. Seinstra for new research in the area of multimedia grid computing, giving a very strong boost to this collaboration.
- We collaborated with BRICKS in the visualization sub-program, studying occlusion and measuring in augmented reality.
- With Collaboratory.nl we developed an authorization and security system that allows virtualisation and remote control of instrumentation. Partly due to this contribution FEI later succeeded in winning a bid for delivery of electron microscopes to the King Abdullah Science and Technology (KAUST) University in Saudi Arabia. This was due to the fact that these microscopes could be remotely controlled and serviced using some of the concepts developed in VL-e.
- There has been a successful collaboration with the Netherlands Proteomics Center through the Telescience program.
- We collaborated with LOFAR in the area of sensor networks for applications outside astronomy. In general, sensor networks will become very important, as shown by examples like Ijkdijk. Also, we will strengthen our collaboration with LOFAR through the DAS-4 project, in which ASTRON is a new partner. This effort will focus on the usage of accelerators (e.g., Graphics Processing Units) to speed up fundamental algorithms of the LOFAR pipeline (e.g., correlator algorithms).

Altogether, it can be concluded that the multidisciplinary nature of e-Science has resulted in many intense collaborations with several Bsik initiatives and other informal networks. These collaborations will certainly continue and strengthen in the future.

Chapter 5
International Position

The strategy of the VL-e project has always been to exploit results first on the national (Dutch) level (for instance through BiG Grid) and then on the international level. During the last phase of the project, much attention was therefore given to international embedding. During the whole project, VL-e researchers collaborated intensively in the international arena, but especially during the last phase the VL-e ideas, expertise, methodology, and software (i.e. know-how) have been adopted by many international researchers. In addition, many new international projects have been started that will spread the VL-e achievements even further. Below, we describe the collaborations and usage of the VL-e know-how.

International collaborations

Several applications are internationally organized, for example, at the European level. A good example is the HEP (High Energy Physics) participation through the LHC Computing Grid project and through the EGEE project. The methodology used in VL-e for software verification and its deployment in the operational e-Science environment has led to a prominent position in the certification process used for the European e-Infrastructure, in particular in EGEE. As a consequence the sustained results from VL-e (and BiG Grid) have placed the Netherlands in a key position in Europe, resulting in the awarding to Amsterdam of the head office of the European Grid Infrastructure (EGI.eu), the persistent organization fostering grid e-infrastructures.

Many other VL-e groups have a similar international orientation:

- The UvA-IBED group coordinates a new program working towards an international system for warning pilots in case of active bird migration, using both radar networks and models. It has also established a new collaboration with Movebank (USA).
- Amolf is involved in a number of international collaborations through the Knowledge Exchange (KnowEx).
- Partners of the medical (AMC, VUmc) and life-sciences (UvA) sub-programs are involved in European projects addressing e-infrastructures for health, biomedical research and research on Alzheimer disease and MRSA. The AMC also participates actively in the HealthGrid association and in organizational efforts of the European Life Science virtual community (BIOMED VO) in the new EGI structure.
- Other examples are collaborations such as TREC, CoreGrid, and Taverna and Kepler involving workflow systems.

The two experimental environments, PoC and RP, cannot do without a good international embedding. The importance of this embedding is emphasized by the role the Rapid Prototyping environment plays in the collaboration with the French Grid'5000 project (see Cappello's keynote presentation at CCGrid'07). The DAS-project and the French Grid'5000 project collaborate intensively, including dissemination activities. Researchers from the PoC and RP environments further participate in international standardization efforts, such as OGF (see later), the World Wide Web Consortium (W3C) with the BioRDF taskforce and Terminology taskforce of the W3C HCLS IG.

VL-e project members were among the founding fathers of the association 'Gridforum Nederland', which is one of the first official 'affiliates' of the Open Grid Forum aimed at transferring know-how on grid technology by organizing tutorials, master classes, business days, and evening lectures. LogicaCMG and IBM (the multinational partners participating in the PoC) established a grid 'technology crossing' course and have started delivering consultancy services to their client base on the basis of the experiences in VL-e. VL-e organized several international workshops on e-Science infrastructure for animal tracking, workflow-management systems, and biomedical applications. VL-e researchers also gave numerous presentations and demos at international meetings.

As the Scaling and Validation program (P4) matured, its international impact grew. The methodology used in VL-e for software verification and its deployment in the operational e-Science environment has led to a prominent position in the certification process used for the European e-Infrastructure, in particular in EGEE. The team executing the VL-e software validation program is scheduled to contribute to the next generation European Middleware Initiative (EMI), playing a role in the Quality Assurance activities therein.
Also, the software hardening and support as part of the P4 program, in particular the VL-e Toolkit (VLeT) and vBrowser, have met international acclaim. Users from diverse projects, including European projects such as HealthGrid, the EGEE biomed community, GateLAB, and D-Grid, appreciate the integrated environment offered by VLeT.

The project has also played a driving role in moving e-Science web portals from ad-hoc elements of the grid to an integral part of the European infrastructure. Through its security development activity, VL-e P4 and BiG Grid ensured that the method used to securely embed portals in the infrastructure has been adopted at the European level, resulting in the adoption of the VL-e and BiG Grid-developed 'VO Portal Policy' by the European Grid projects DEISA and EGEE. In this way we have also ensured that Dutch e-Science communities can seamlessly extend to the European scale. This was convincingly demonstrated by the pan-European e-NMR project (an EU integrated infrastructure initiative funded under FP7), which received support from the VL-e Scaling and Validation program and whose approach was subsequently used at the international scale.

As a result of this international visibility, many new international projects have been started that involve VL-e groups. An important initiative during an earlier phase of the project was the leadership over the EU-GridCoord project (an SSA under the FP6 program), to positively influence the EU-wide thinking on bridging the gap between the Grid R&D world and the e-Science community. Also, several new international programs were initiated, such as EGEE-II (SP1.1, SP4.1), NeuGrid and SHIWA (SP1.3), EcoGRID, ESA-SoS, and LifeWatch (SP1.4), e-BioLab and CONCORD (SP1.5), EU-Computis, EU-MEDITRANS, Virolab, ACGT, QosCoSGrid, COAST (SP2.1), EGI and LifeScience SSC (SP1.3, SP4.1), XtreemOS, and (under contract negotiation) Contrail (SP 3.1), Geysers (SP3.2) and NOVI (SP3.3).

Usage of VL-e know-how

The VL-e Toolkit (VLeT) is the most important software engineering result of VL-e. It is hardened and supported by the P4 program and has met international acclaim. The vBrowser is used by several projects, including European projects such as HealthGrid, the EGEE biomed community, and GateLAB (CREATIS, FR). The Ibis/JavaGAT software is used extensively in the international grid community. The JavaGAT (Java Gridlab Application Toolkit) is in the basic software of the German D-Grid initiative. The Max Planck Institute even organized a workshop about JavaGAT in Potsdam in February 2010. The JavaGAT is also used by the GridChem gateway for TeraGrid, TextGrid (a grid for the humanities in Germany and Europe), AeroGrid (aerospace research community) and AstroGrid-D (for astronomy).

The application oriented PSEs developed in VL-e are currently moving from a national to a more international focus. As all groups involved were able to capitalize on their VL-e efforts and results by obtaining one or more international projects and/or collaborations, this shift is realized at a good pace. The success of each PSE will eventually determine its own international positioning, but also that of VL-e as a whole.

To ensure long-term adoption of the VL-e methods, VL-e researchers have been actively involved in international standards and working groups. An example is the Open Grid Forum where a member of the VL-e staff was responsible for supervision of the standard area for security as its area director. VL-e contributed to the standardization of operational security and identity management through the CAOPS working group. In this way the software and methods developed in VL-e for secure storage of identities were disseminated. VL-e also contributed to other groups related to its mission, in particular 'Grid Interoperability Now' (GIN-WG), the Production Grids Infrastructure (PGI-WG) working groups, the SAGA and SAGA-Core groups (in Application Standards), the Grid High-Performance Networking and Network Measurements (Infrastructure Standards), and the OGSA Authorization and Authentication, Firewall Issues and Levels-of-Assurance groups (Security Standards). The software engineering and integration efforts are aligned with corresponding efforts elsewhere, such as the Build-Test-and-Certification community group in OGF or World Wide Web Consortium (W3C). The team executing the VL-e software validation program is scheduled to contribute to the next generation European Middleware Initiative (EMI), playing a role in the Quality Assurance activities therein.

The VL-e partners have accepted the challenge that was posed in the midterm review, to operate more in the international arena by: i) increasingly organizing international workshops, ii) other dissemination activities in order to promote the e-Science concept as well as VL-e know-how, iii) involvement in new international initiatives and projects, and iv) participating more frequently (and successfully) in international competitions.
The positive appraisal from the COOK report proofs that we succeeded in positioning the VL-e concept high up the international ladder.

Chapter 6
Organization and management

The management structure as formulated in the original VL-e proposal has been preserved. Mid-2006 the project leader decided to reinforce the directorate of VL-e by increasing the number of directors from four to five and over time there have been some minor personal changes in the team. The team has worked in close co-operation with the sub-project (SP) leaders. The consultation structure between the project leader and the supervisory board worked well. The same goes for the relationship between the directorate and the SP leaders. The directors met once a month.

Early 2007, the management team in consultation with the 'Commissie van Wijzen' organized an international review committee consisting of five members, whose task it was to carry out the midterm-evaluation of the progress of VL-e. The review was facilitated by a comprehensive self study made by the management team in collaboration with the SP leaders during the first half of 2007. This report is titled *VL-e Mid-Term Progress Report*.

VL-e contributes to the innovation of the ICT based e-Science research infrastructure in particular, and, by means of BiG Grid, to the entire e-Science infrastructure in the Netherlands. As noted in previous reports there is an bottleneck at the national level. A better organizational integration of the two parties which are responsible for the national computers and network infrastructure in the Netherlands is needed.

At the end of 2007/beginning of 2008, the managing team decided, in consultation with the Supervisory Board, how the last phase of the VL-e project should be filled in and how the embedding of the project should be accomplished. Judgments and recommendations of the review committee and the Commissie van Wijzen were taken into consideration. The expectation is that an *e-Science Research Center (eSCR)* will be set up (see also Embedding) in 2010, from where the knowledge of the VL-e project not only can be exploited and disseminated, but also can be developed further in collaboration with other e-Science projects.

Supervisory Board (= Supervisory Board WTCW N.V.)

Drs. P.W. Doop, chair	Universiteit van Amsterdam
Drs. L. Steenbergen	ING Bank
Dr. K. Rutten	Vrije Universiteit Amsterdam
Dr. A.W.M. Rijnders	Akzo Nobel
Dr. L. Wesdorp	Unilever R&D
Drs. E.E.C. Hupkens van der Elst	VNU N.V.
B.J. Marttin MBA	Rabobank Amsterdam
Ir. H. van Dorenmalen	IBM Nederland

Board of directors of WTCW N.V.

Prof.dr. L.D. Noordam, chair	UvA
Prof.dr. J.K. Lenstra	CWI
Prof.dr. F.L. Linde	Nikhef
Prof.dr. J. van Mill	VU
Drs. J.A.A. Vos, administrative director	WTCW

Board of Directors VL-e

Prof.dr. L.O. Hertzberger, chair	UvA
Prof.dr.ir. H.E. Bal	VU
Dr. G. Meijer	Logica
Prof.dr. P. Adriaans	UvA (2005 - june-2008)
Prof.dr.ir. C.T.A.M. de Laat	UvA (from june-2008)
Drs. A. van Rijn	Nikhef
Drs. J.A.A. Vos, administrative director	WTCW

Sub-program leaders

Dr. Jeff Templon	SP1.1	Nikhef
Dr. Dirk Out	SP1.2	Unilever R&D
Dr. Sílvia Delgado Olabarriaga	SP1.3	AMC
Prof.dr.ir. Willem Bouten	SP1.4	UvA
Dr. Timo M. Breit	SP1.5	UvA
Prof.dr.ing. Ron Heeren	SP1.6	FOM-Amolf
Prof.dr. Peter Sloot	SP2.1	UvA (till June 2007)
Drs. Breanndán Ó Nualláin	SP2.1	UvA (from June 2007)
Prof.dr. Pieter Adriaans	SP2.2	UvA
Prof.dr.ir. Robert van Liere	SP2.3	CWI, TU/e
Dr. Hakan Yakali	SP2.4a	UvA (till June 2007)
Drs. Guido van ‘t Noordende	SP2.4b	UvA (from June 2007)
Dr. Adam Belloum	SP2.5	UvA
Prof.dr.ir. Henri Bal	SP3.1	VU
Dr. David Groep	SP4.1	Nikhef
Dr. Geleyn Meijer	P5	LogicaCMG, UvA

Consortium partners

A&F	FCDF	LogicaCMG	TNO – I&T	UvA-IvI	WTCW
ATO	FEI	NBIC	TNO-voeding	UvA-SILS	
AMC	FOM Amolf	Philips Research	TU Delft	VU	
CWI	FOM Nikhef	Philips Medical	Unilever	VUmc	
DSM	IBM	SARA	UvA-IBED	WCFS	

Chapter 7

Project Funding in Relation with the Estimate

Comments

- The partners of VL-e have invested more than expected: € 42,823,938 in stead of the anticipated € 41,300,000.
- At the end of the project, after all financial statements have been received including approval of accountants, the remaining € 1,195,623 of subsidy minus cost for management has to be re-divided over the partners who have invested (substantially) more than expected at the beginning in 2004.

		Begroot (in €)	Werkelijk (in €)
Totale projectkosten		41.300.000	42.823.938
Toegekende Bsik subsidie		20.000.000	20.000.000
Nog te financieren		**21.300.000**	**22.823.938**
Eigen bijdrage voor:			
Penvoerder	<<naam/ soort organisatie>>		
Deelnemer 1	SARA	250.000	422.587
Deelnemer 2	TNO I & T	250.000	482.334
Deelnemer 3	TNO KvL	250.000	770.557
Deelnemer 4	Unilever	500.000	450.394
Deelnemer 5	Friesland Foods	500.000	315.090
Deelnemer 6	IBM	1.100.000	16.778
Deelnemer 7	LogicaCMG	1.000.000	534.227
Deelnemer 8	Philips	2.000.000	505.460
Deelnemer 9	FEI	800.000	0
Deelnemer 10	DSM	1.300.000	0
Deelnemer 11	Nikhef	1.500.000	2.397.430
Deelnemer 12	Amolf	800.000	945.049
Deelnemer 13	CWI	400.000	576.777
Deelnemer 14	TIFN (WCFS)	500.000	761.267
Deelnemer 15	A & F	500.000	840.405
Deelnemer 16	UvA	5.000.000	9.559.545
Deelnemer 17	AMC	700.000	1.472.463
Deelnemer 18	VU-CS	1.000.000	2.300.975
Deelnemer 19	VU-MC	400.000	548.616
Deelnemer 20	TUD	650.000	1.119.604
Deelnemer	WTCW	1.900.000	-1.195.623
Deelnemer	NBIC		
		0	**3**
Ontvangen subsidies van:			
Subsidie 1	<<subsidienaam/ subsidiegever>>		
	enz.		
		0	**3**
Andere financieringsbronnen:			
Bron 1	<<financieringsbron/ financier>>		
	enz.		

Chapter 8
Expenditure per cost category

Comments

- Most of the cost is labor-related. In the overhead cost of existing desk equipment, such as computers and software, is included. Though we expected more investments in hardware in 2004, with the start of Big Grid and the intense cooperation afterwards most of the investment in hardware has been done by Big Grid, enabling VL-e investing more in manpower.
- The cost for dissemination and technology transfer as specified below is only out-of-pocket; invested manpower is part of the labor cost.

Uitgaven	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	***Totaal***
Loonkosten		2.517.016	3.910.090	3.850.820	3.803.166	3.442.797	3.729.075	**21.252.964**
Kosten van arbeid		700.505	729.836	825.095	670.849	1.157.521	1.329.285	**5.413.091**
Opslag algemene kosten		1.555.860	2.093.272	2.188.528	2.144.901	2.175.268	2.359.189	**12.517.018**
Kosten van machines en apparatuur		345.817	319.972	487.336	389.626	965.820	552.850	**3.061.421**
Kosten van te verbruiken materialen en hulpmiddelen		14.446	64.931	62.023	43.394	7.071	39.711	**231.576**
Kosten voor verspreiding en overdracht van kennis		37.402	81.676	30.105	60.336	84.942	53.406	**347.867**
Totaal	**0**	**5.171.046**	**7.199.777**	**7.443.907**	**7.112.272**	**7.833.419**	**8.063.516**	**42.823.937**

Chapter 9
Funding of the VL-e project

Comments

- Because of the change of investments among the partners - see also Chapter 11 - in 2005-2007 less than average was invested; this was compensated with more than average investments in 2008 and 2009.

	2003	2004	2005	2006	2007	2008	2009	Totaal
Totale projectkosten		5.171.046	7.199.777	7.443.907	7.112.272	7.833.419	8.063.516	42.823.937
Toegekende Bsik subsidie		2.790.285	3.666.727	3.750.784	3.800.761	3.205.048	2.786.395	20.000.000
Eigen bijdragen	0	2.380.761	3.533.050	3.693.123	3.311.511	4.628.371	5.277.121	22.823.937

- Most net (investments minus Bsik subsidy) partner contributions come from the University of Amsterdam, Nikhef, VU University, AMC and Delft University.
- The amounts of these five add up to almost 70%.
- There has been no direct funding from third parties.

Eigen bijdragen		2003	2004	2005	2006	2007	2008	2009	Totaal
Bedrijven	SARA		63.326	67.321	58.848	46.170	90.770	96.153	422.588
	TNO I&T		37.835	98.726	66.328	131.556	147.888	0	482.333
	TNO KvL		3.299	83.018	51.951	51.086	296.396	284.808	770.558
	Unilever		86.485	106.735	73.114	82.258	101.802	0	450.394
	Friesland Foods		16.006	83.145	180.050	35.889	0	0	315.090
	IBM		16.778	8.713	-8.713	0	0	0	16.778
	Logica		92.660	129.255	77.979	26.506	29.056	178.772	534.228
	Philips		98.649	255.009	217.938	18.512	-84.649	0	505.459
	FEI		0	0	0	0	0	0	0
	DSM		0	0	0	0	0	0	0
Universiteiten	UvA		1.057.714	1.166.608	1.309.303	1.329.883	1.766.040	2.929.997	9.559.545
	AMC		213.538	240.147	288.286	259.767	226.916	243.808	1.472.462
	VU-CS		227.850	268.575	273.782	289.189	513.455	728.125	2.300.976
	VU-MC		0	45.407	61.709	67.731	218.892	154.877	548.616
	TUD		59.659	137.861	151.090	185.350	294.148	291.496	1.119.604
Onderzoek-instellingen	Nikhef		213.508	340.057	344.733	362.775	404.795	731.562	2.397.430
	Amolf		79.111	166.377	172.856	119.090	180.009	227.606	945.049
	CWI		77.797	91.721	94.652	78.855	127.400	106.351	576.776
	TIFN (WCFS)		0	109.977	131.408	108.635	176.611	234.637	761.268
	A&F		36.545	134.397	147.809	118.259	138.841	264.555	840.406
Overigen								-1.195.623	-1.195.623
Bijdragen derden		0	1	1	0	0	1	-3	0

Chapter 10

Distribution VL-e subsidy over the consortium partners

Comments

- In relation to the net investment of the five largest investing partners (70%, see box before) only 60% of the Bsik subsidy goes to the University of Amsterdam, Nikhef, VU University, AMC and Delft University.
- In the last line of this box 'Anderen' is WTCW N.V. the managing partner of the project. The rather large figure in 2009 consists for almost € 1,200,000 of subsidy which still has to be re-divided among the partners of VL-e and WTCW for management. This re-division (some partners have invested less than expected at the beginning of the project and some more) can only be done after all final financial information and figures have been received and approved by the accountant of WTCW N.V.

Partijen		**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	***Totaal***
Bedrijven	SARA		76.005	80.800	70.630	55.414	60.740	30.259	**373.848**
	TNO I&T		45.411	118.494	76.096	60.000	60.000	0	**360.001**
	TNO KvL		3.299	83.018	51.951	51.086	100.000	0	**289.354**
	Unilever		34.642	42.754	29.287	32.949	40.778	0	**180.410**
	Friesland Foods		6.402	33.256	51.049	7.143	0	0	**97.850**
	IBM		6.863	3.564	-3.564	0	0	0	**6.863**
	Logica		18.576	25.913	15.633	14.272	19.419	105.581	**199.394**
	Philips		0	0	0	0	0	0	**0**
	FEI		0	0	0	0	0	0	**0**
	DSM		0	0	0	0	0	0	**0**
Universiteiten	UvA		1.373.812	1.515.250	1.700.589	1.727.319	1.116.238	450.239	**7.883.447**
	AMC		44.392	83.293	110.844	176.700	171.028	88.621	**674.878**
	VU-CS		318.553	402.472	437.021	463.140	410.244	28.290	**2.059.720**
	VU-MC		0	22.700	30.850	33.861	109.430	27.160	**224.001**
	TUD		73.419	169.658	185.939	228.101	201.468	50.532	**909.117**
Onderzoek-instellingen	Nikhef		170.775	271.996	275.737	290.168	292.166	57.834	**1.358.676**
	Amolf		79.111	166.377	172.856	119.090	120.106	151.864	**809.404**
	CWI		116.696	137.582	141.978	118.283	85.217	71.137	**670.893**
	TIFN (WCFS)		0	39.003	24.593	6.350	24.759	58.546	**153.251**
	A&F		29.231	107.498	118.226	94.590	77.122	130.000	**556.667**
Anderen			393.096	363.100	261.071	322.296	316.333	1.536.330	**3.192.226**
Totaal		**0**	**2.790.283**	**3.666.728**	**3.750.786**	**3.800.762**	**3.205.048**	**2.786.393**	**20.000.000**

Chapter 11

Amplifying on the progress of expenditures and funding

The total investments in VL-e have been larger than anticipated at the beginning. At the same time we see a shift among the partners who have actually made these investments.

Two companies, DSM and FEI, which – at the beginning of the project – projected their investments at M€ 1.3 and M€ 0.8 during the VL-e period did not realize these figures. It turned out rather quickly after the agreement with VL-e consortium was signed that DSM and FEI decided to spend their funds in an adjacent project: collaboratory.nl. Collaboratory.nl was – just as VL-e – a form of continuation of the Virtual Lab Amsterdam, part of the ICES-KIS WTCW project. DSM and FEI joined forces in Collaboratory.nl with Corus, Philips, Telematica Insitituut (now Novay) and University of Amsterdam.

Something similar happened with Philips. Though Philips has been participating during the whole lifetime of the project it invested less than expected. The forecast was Philips would invest M€ 2, but it realized only M€ 0.6. The reason is that halfway the project time of VL-e Big Grid started and Philips decided to spend more time and effort in Big Grid. For VL-e Big Grid has from the beginning been an important sister-project working towards the same goals and objectives as VL-e, but more focused on the e-Science infrastructure than VL-e. So, Philips spent less in VL-e, but this was more than fully compensated with its investments in Big Grid.

In general we have seen that during the project period the focus of the work in VL-e has been led more to scientific methodologies. Though industry is very aware of the importance, impact and opportunities in e-Science in practice it was hard to get the right connections. IBM for example expected to be able to invest more hard- and software for VL-e, which turned out not to be the case, also because of Big Grid where large funds were available for basic e-Science infrastructure.

Logica, Friesland Foods and Unilever have been able to get the right connections and results. TNO, both the divisions of Information & Communication Technology as Quality for Life and Sara have invested even double or triple the amount they expected in the beginning.

Overall we have seen reluctance in the participation of industry, because of other opportunities (DSM, FEI and to a certain extent Philips) or because of the (basic) research that did not directly lead to concrete products or services within months. At the same time all the industrial partners have been very interested in the results of the VL-e project. This has been proven by the letter of support with financial commitments that have been signed by industrial partners for the COMMIT proposal, in which – also – continuation of e-Science research is being described.

Fortunately, the other partners have been able to compensate for the lacking investments from industry. Nikhef, the University of Amsterdam, the VU University of Amsterdam and Delft Technology University have all doubled their initial investments. These extra investments and the efforts of all partners have led to a total investment of € 42,823,937 in VL-e.

In box 8 the cost is split over several items; one of these being cost for dissemination. To prevent any misunderstanding: this figure only represents the cost for out-of-pocket expenses; all the manpower connected to technology transfer and dissemination is part of the labour cost.

The financial information as presented in the boxes 7-10 has been based on the financial reports of each individual partner. From 2004 until 2008 these reports have been checked and approved by both the accountant of the partner and the accountant of WTCW N.V. The figures of 2009 are the best estimate of each partner, as made in the beginning of 2010. We expect to have received all the final financial reports including the approval of the accountants by the summer of 2010. So the figures as presented above can slightly be different from the final figures as we expect to have later on this year.

APPENDIX

Table 1. VL-e subprograms

Sub-project	Title and description sub-project	Date		Executors: Institute(s), Project leader
		Start	End	
SP 1.1	**Data Intensive Science** *Design and development of a specific common layer to enable data-intensive sciences with a focus on high-energy physics. Important elements are; metadata management, data-replica management and virtual organizations*	Jan 2004	Dec 2009	FOM/Nikhef Dr. Jeff Templon
SP 1.2	**Food Informatics** *Design and development of a decision support problem-solving environment for the Dutch food research community by organizing the relevant information via disclosure techniques such as, ontologies and semantic web*	Jan 2004	Dec 2009	Wageningen UR, Unilever, Friesland Foods, TNO, TIFN Dr. Dirk Out
SP 1.3	**Medical Diagnosis & Imaging** *Sub-projects: Service-oriented workflow management for image analysis and logistics (FARP); Tracking of white matter tracts in DTI (HAFT); Virtual Lab for Medical Imaging (functional MRI, DTI), MIRAGE (medical imaging analysis on grids); visualization in medical image analysis*	Jan 2004	Dec 2009	UvA-AMC, IBM, Philips Research Eindhoven, VUMC, SARA, UvA-IvI, Nikhef, Logica Dr. Sílvia Olabarriaga
SP 1.4	**Biodiversity** *Development and application of Virtual Lab and grid technology based information systems for identifica-tion, distribution, integration, and analysis of data and model results of the dynamics of flora and fauna*	Jan 2004	Dec 2009	UvA-IBED Prof. Willem Bouten
SP 1.5	**Bioinformatics-ASP** *Development of bioinformatics problem-solving environ-ments in which life-science researchers are able to integrate biological experiment data for* in-silico *experimentation thus allowing new research approaches for system-level biology*	May 2004	Dec 2009	UvA-SILS Dr. Timo Breit
SP 1.6	**The Dutch Telescience Laboratory** *Development and implementation of telescience infra-structures that enable scientific virtual collaborations using e-science environments and application-specific tools for advanced protein identification and chemical imaging*	Jan 2004	Dec 2009	FOM-Amolf Prof. Ron Heeren
SP 2.1	**Interactive Problem-Solving Environments** *Design and development of a generic problem-solving environment architecture that supports high-performance computation on the grid, including mechanisms for dynamic interaction for* in-silico *experimentation in life sciences*	Jan 2004	Dec 2009	UvA-IvI Drs. Breanndán Ó Nualláin

SP 2.2	**Adaptive Information Disclosure** *Design and development of a suite of semantically-oriented, dynamic model-driven information and knowledge extraction tools on top of the Virtual Lab architecture for grid-based distributed data analysis*	May 2004	Dec 2009	UvA-IvI, VU, TNO Prof. Pieter Adriaans
SP 2.3	**User Interfaces and Virtual Reality Based Visualization** *Design and development of novel scientific visualization methods for the analysis of large and complex data sets arising from VL-e applications*	Feb 2004	Dec 2009	CWI, TUD, Amolf, UvA-IvI Prof. Robert van Liere
SP 2.4	**Collaborative Information Management** *Development of infrastructure and tools to support the management and dynamic modeling of data and information, plus their collaborative management for co-operative problem solving among various VL-e partners*	Jan 2004	Jun 2009	UvA-IvI Dr. Hakan Yakali
	Information Security and Privacy *Development of models and architecture that allows secure handling of privacy sensitive data in distributed processing environments as found in grid and cloud systems*	July 2007	Dec 2009	UvA-IvI Drs. Guido van 't Noordende
SP 2.5	**Virtual Laboratory and System Integration** *Development of e-science infrastructure for robust access to various grid-enabled facilities, reliable resource management techniques, process and data workflow support for applications and a collaborative environment*	April 2004	Dec 2009	UvA-IvI Dr. Adam Belloum
SP 3.1	**High Performance Distributed Computing** *Research into suitable programming models and communication mechanisms for the development of efficient and robust distributed supercomputing applications suitable for grid that can schedule the many grid resources*	Jan 2004	Dec 2009	VU-CS, TUD-ITS, UvA-IvI Prof. Henri Bal
SP 3.2	**Security and AAA *** *Authentication, Authorization and Accounting models and architectures for VL-E environments and research its security related implications and considerations. This SP was transferred to and successfully completed in GigaPortNG as noted in the SAC evaluation of that project.*	Jan 2004	Dec 2009	UvA-SNE, SURFnet and partners Prof. Cees de Laat
SP 3.3	**Optical Networking *** *Optical network transport models for e-science applications. This SP was transferred to and successfully completed in GigaPortNG as noted in the SAC evaluation of that project.*	Jan 2004	Dec 2009	UvA-SNE, SURFnet and partners Prof. Cees de Laat
SP 4.1	**Scaling and Validation in real-life applications** *Validation of the VL-e results by evaluating in real-life environments and applications, with a focus on analysis of several comprehensive soft- and hardware e-science environments for scientific experimentation*	Jan 2004	Dec 2009	FOM-Nikhef, SARA, UvA-IvI, PCC, Logica, IBM, Philips, VU Dr. David Groep

Please note that SP3.2 and SP3.3 overlapped completely with subjects in GigaPortNG. As was noted in the 'aanbiedingsbrief' of both projects the research of these sub-programs would default to GigaPortNG in case both projects would succeed. Since both projects were approved, the research in SP 3.2 and SP 3.3 were completely done and reported in GigaPortNG. During the project there was a strong collaboration and joint work, specifically on the DAS-3, StarPlane and SCARIe projects amongst others. One result of the strong collaboration between the authorisation security group in GigaPort and the VL-e middleware groups is the recognition that the topic of information security and privacy was becoming essential to make distributed handling of medical and industrial data acceptable. This research was included in SP 2.4 in July 2007. With regard to the work that was transferred to GigaPortNG, the Scientific Advisory Committee of that project specifically notes in its end report:

> *"The scientific quality of the sub-project is demonstrated by the Research on Networks sub-project with a total of 54 peer reviewed papers, 196 presentations and 31 software deliverables. A significant number of these results came from the System and Network Engineering group at the University of Amsterdam (UvA), with 36 peer-reviewed papers, 86 presentations and 15 software deliverables. The research performed by the UvA within the GigaPort NG project is an indispensible part of the network innovation model and receives world-wide recognition."*

Table 2. Milestones scientific output

	Milestones for scientific output: status and progress regarding deliverables **Start project – December 2009**
1.	Number of publications in international journals and conference proceedings
	463. The number of publications in international journals and conference proceedings far exceeds our original expectations of 25 per year. This number represents accepted publications at the end of 2009. Obviously, many papers are currently in preparation or in the reviewing process. Thus, we expect the total number of publications to grow still.
2.	Number of presentations at international conferences
	251. The number of presentations at international conferences is higher than the anticipated 25 per year.
3.	Number of completed PhD theses
	22. This number of completed PhD theses exceeds our original estimate of 10 theses during the total project. Moreover, there are still several theses in progress.
4.	New research activities initiated by VL-e
	Major new initiatives: - BiG Grid The largest initiative is the M€ 29 NWO funded BiG Grid project. - COMMIT A new project proposal for an e-science project of M€ 32 of which M€ 14.3 subsidy was written within the ICTregie program COMMIT. This FES application has been submitted September 2009. A final decision is expected in 2010. - eSRC A plan for a new e-science research centre is being drawn up. An initial startup funding is already granted. The decision for full funding is expected in 2010. - LifeWatch LifeWatch is an ESFRI project with 19 participating countries. It concerns an e-science and technology infrastructure for biodiversity data and observatories. *Other EU and NWO funded projects include (sub-projects within):* - EGEE-II Enabling Grids for E-sciencE - CREATIS/NeuroLOG French institute and project that adopt the VL-e medical image analysis platform - NeuGrid FP7 project to set-up a European grid for imaging in Alzheimer's disease - SHIWA recently approved FP7 project on workflow interoperability - BAMBAS Bird Avoidance Model / Bird Avoidance System - EcoGRID A system that discloses the Dutch nature observations for nature management, conservation and legislation - CONCORD CONtrol of COmmunity-acquired MRSA - BioRange-II NBIC bioinformatics research program - BioAssist-II NBIC bioinformatics support program - COMPUTIS to develop new and improved technologies for Molecular Imaging Mass Spectrometry enabling innovative methods functional genomics. - MEDITRANS a multidisciplinary integrated project about targeted nanomedicines. - WISDOM Pan-European studies on in-silico anti-malaria drug discovery. - H5N1 Pan-European studies on the H5N1 virus. - Virolab A virtual laboratory for decision support in viral diseases treatment. - ACGT Advancing Clinico-Genomic Trials on cancer, developing open-source, semantic and grid-based technologies for post genomic clinical trials. - NIH-RO1 On breast cancer research - ESA-SoS Space Observatories in School for science awareness of young people - COAST A project on Complex Automata Simulation Technique. - CWA Concept-Web-Alliance for data, information and knowledge integration - NWO-RBFR NWO grant on Dutch-Russian Research Co-operation - NWO-VIEW Visual Interactive Effective Worlds; several projects (a.o. VEARD) - NWO-GLANCE GLobAl computer scieNCE; StarPlane, JADE-MM,GUARD-G - QosCoSGrid Quasi-opportunistic supercomputing for complex systems in grid environments.

	Other new research lines: - Reuse-based Ontology Construction Kit (ROC) - Ontologies of Units and Measures (OUM) and Qantitative Knowledge (OQR) - e-Bioscience group at the AMC, created for embedding and continuation of part of the VL-e Medical research and team. - Usage of DTI and functional MRI data in clinical neuroimaging and neuroscience research - e-BioInfra: e-infrastructure for bioscience research (at the AMC) - Advanced image analysis as a service - Prediction of ingredient-receptor response - Virtual Lab for functional MRI - Usage of DTI data in clinical neuroimaging and neuroscience research - Architectures for medical image analysis applications - MEG based measures for monitoring MS patients, - Image reconstruction algorithms. - Linking species abundance to environmental properties. - New methodologies for inverse modeling and data-assimilation - An actual laboratory for e-science (e-BioLab) - Usage of MDL on large datasets using the grid - New methods for automatic workflow composition and analysis using IO automata. - New routines for part-whole learning and subclass mining. - Focus on data privacy and security. - Workflow interoperability. - Extending food informatics: Tiffany, Emerging Risks Detection System, DB healthy food components, Experience Box, MARVIN, Webshop plantconnect, and KODA, Gateway2Food.
5.	'Zwaartepuntvorming': the number of researchers involved in VL-e and the number of locations where research takes place. User groups for VL-e.
	In total **150** researchers (100 fte) were involved in VL-e, working at **33** locations. In addition, there are many diverse user groups for VL-e methods software and infrastructure, including: LOFAR, DANS, the humanities, historical archiving, social sciences, Top Institute Food and Nutrition (TIFN, formerly known as WCFS), the Netherlands Bioinformatics Centre (NBIC), Unilever, neuroscientists (AMC, VU and UvA,) clinicians and medical imaging researchers (AMC, VUMC, UvA), human geneticists (LUMC), microbiologists (UMCU), molecular biologists (UvA and RIVM), bioinformaticians (LUMC, UvA and AMC), eNMR groups (UU), students an researchers using the e-BioLab (UvA, TU), EcoGRID, NLgrid and RIDgrid (a Russian-Dutch grid), the national VRlab user group, and EU-SAFE consortium. The VL-e PoC supports in total over 37 user communities in the Netherlands.

Table 3. Milestones economic output

	Milestones for economic output: status and progress regarding deliverables **Start project – December 2009**
1.	Collaborations with industry:
	- The COS group at Philips Research Laboratories is responsible for building and supporting a grid infrastructure within Philips. Collaborations also exist through Philips Medical Systems. - The industrial partners of the TIFN (formerly WCFS) (AVEBE, COSUN, CSM, DSM, Unilever, NZO) have access to TIFN output through RMS, which is partly based on VL-e results. - Unilever has shaped its search environment based on developments and lessons from VL-e. - EcoGRID is extended in the context of GAN (Data Authority Nature) in co-operation with (non-) commercial parties; ministry LNV, foundation VOFF, municipality Haaglanden, Provincie Flevoland, Natuurmonumenten, Waarnemingen.nl. - The ESA System of Systems Bird Migration is a new industrial co-operation with the KNMI. - Several spin-off ontology activities; ‘plant ontology’ in horticulture, 'food ontology' in Tiffany, 'risk ontology' in emerging food safety hazards. - There are transcriptomics PSEs; PROGENIUS with UMCU and EUGENIUS with LU. - Logica is exploring collaborative networking technologies and business models together with academic partners (AMC, Nikhef, and RUG). - TI Food and Nutrition, Unilever, TNO Quality for Life are involved in the Food Informatics sub-project of VL-e.
2.	Other industries using know-how of the project are:
	- *Technology companies*: IBM, Philips Research Eindhoven, Philips Healthcare, Unilever, Logica, the industrial partners of TIFN, TNO quality of life, Man systems, PS-Tech, Medigon and Bruker Daltonics, and ten companies participating in VOFF. - *Academic hospitals*: A growing number of biological users from RIVM, AMC, LUMC and UMCU. - *Life-science companies*: De Ruiter Seeds, Syngenta, ENZA Zaden, Incotec, Seminis, Bejo Zaden, Nunhems Zaden, Nickerson-Zwaan, Beekenkamp plantenkweker, Vreugdenhil plantenkweker, Leo Ammerlaan plantenkweker, WPK plantenkweker, Innovatiecentrum Gezonde Voeding, Q-ray, Flora Holland, and ABL Luxembourg.
3.	VL-e involvement in techno starters:
	VL-e is involved in several techno starters that are in various stages of maturity, including: Virology, Gridwise, Foldyne, Personal-Space Technologies (PS-tech), Life Sciences Pioneers, Culios, KnowEx services inc., and Treparel.

Table 4. Milestones societal output

	Milestones for societal output: status and progress regarding deliverables **Start project – December 2009**
1.	The number of non-commercial applications in different sectors of society
	- An Emerging Risk Detection Support System for food safety for the Ministry of Agriculture. - Bird Avoidance Models to enhance flight safety (https://ecogrid.sara.nl/bambas/). - A national biodiversity portal (http://www.natuurloket.nl/). - The EcoGRID portal (used by 18.000 volunteering field observers to upload their data). - Contribution to anti-malaria drug design through molecular docking (WISDOM). - Research on Huntington's disease, epilepsy treatment, and ecstasy use. - A HIV drug resistance decision support system (in clinical use). - Visualization of CT Angiography of the head (supporting first-aid care in the AMC). - Neurosurgery planning for oncology patients - An OUM/OQR-ontology web service oriented at units of measurement and quantities - Collaborations with Movebank (http://www.movebank.org/about/index.html) - The construction of an e-BioLab with a high-resolution tiled display to support multidisciplinary e-science project teams. - Experience Box for sharing experiences from public-private collaborations with a broad audience (http://www.experiencebox.nl/). - Gateway2Food, a guide for finding experts in the food technology domain (http://www.gateway2food.nl/). - Hosting services for medical imaging workflow and adaptive information/ontology (AIDA). - UvA W3C active membership in the W3C Semantic Web HCLS Interest Group. - Active 'Project or Area Liaison' (PAL) for OMII-UK. - The VL-e PoC supports directly 34 external user communities in astronomy, astroparticle physics, atmospheric research, humanities, archive (DANS), neurolinguistics, electron tomography, accelerator design, and other biomedical work including drug design.
2.	Societal user groups
	The purpose of VL-e is to develop *generic* tools and the broad range of the social and economic applications illustrate the advantage of our approach. Many societal user groups were created and/or are supported with significant VL-e involvement: DANS (archaeologists and historians), FAO (UN Food and Agriculture Organization), Nature and Food Quality, Royal Netherlands Air Force, ICEAGE (International Collaboration to Extend and Advance Grid Education), neuroscientists (AMC, VU, UvA), clinicians (AMC Radiology/ Neurosurgery, VUmc), medical imaging researchers (VUmc, AMC, UvA), human geneticists and bioinformaticians (LUMC), and omics researchers (RIVM, UMCU, LU). Furthermore, VL-e know-how is used in higher education at several universities. Strong co-operation with NBIC for e-science in transcriptomics and proteomics exists together with the Netherlands Proteomics as well as Toxicogenomics Centres (NPC and NTC). The KnowEx research information management tool developed by Amolf in the context of VL-e is used by EU-COMPUTIS and EU-MEDITRANS. The Virtual School Lab is used for education purposes. A Food Informatics seminar for the general public, reported in a dedicated issue of the professional journal; Agro Informatica. The VL-e Toolkit in the PoC is ported to large-scale immersive visualization environments.

Table 5. Innovation milestones

	Milestones for innovation: status and progress regarding deliverables **Start project – December 2009**
1.	Establishment of participation in techno platforms:
	VL-e co-initiated the Dutch Grid Forum, the steering group NCF Grid Relating Policy, Virolab (STREP), accession to W3C, SKOS, VLAM, myGrid, Taverna, Kepler, myExperiment.org, BioCatalogue.org, and BioAssist (NBIC).
2.	Participation in user groups and standardization platforms
	- Global Grid Forum; Steering Group, Application Developers Users Research Group (APPS-RG), Authentication (CAOPS-WG), Security (OGSAAuthZ), GridCoord: coordination of national and Eur grid projects, Int. Grid Trust Federation, Grid Forum NL, Systems biology - 'GRID and Virtual Laboratory Innovation Consultation' between the partners IBM en LogicaCMG, in which new application areas and user groups are identified and contacted. - OUM (Ontology of Units of Measure) has been submitted to W3C; - W3C Healthcare and Life Sciences Interest Group (HCLS IG), - Active 'Project or Area Liaison' (PAL) for OMII-UK. - Representation in the UK e-science strategic advice committee. - International professional organizations on visualization (ACM, Eurographics, IEEE CS). - OGF working Group on Workflow Issues and currently, the security area directorship in OGF. - Use of web-services and AIDA in BioRDF task force and Terminology task force of HCLS IG. - VL for medical imaging is adopted by various French research laboratories. - A vocabulary for the EU-project Hitech Europe. - Application of semantic web technology for food safety standards in EU INPLISTA-project.
3.	Education and training
	Bachelors Systems Biology, Masters Grid, Virtual School Lab; IBM class, tutorials Ibis + GAT at INRIA, Talks at grid tutorials, Inverse modeling and parallel computing in MSc program Earth sciences and Biological Sciences; Organization of tutorials after each PoC release (mainly via ISOC and GridForum Nederland); Organization of ASCI course 'Advanced Grid Programming Models', including Ibis, JavaGAT, SAGA, KOALA and ProActive (INRIA) , NBIC Information Management course for Bioinformatics PhD's, training of Bioinformatics developers in the NBIC BioAssist program, contributed lectures in SIKS Advanced Semantic Web courses. Nikhef organized several GRID tutorials. VL-e worked with GridForum.nl and ISOC.nl to organize masterclasses on various grid and Internet subjects. Summer schools in mass spectrometry, Nordic-QP course at Amolf. Osei NUFFIC-project to support the Osei Tutu II Institute for Advanced ICT Studies in Ghana. Project Information Science at VU, several Internships of MSc. students.
4.	International embedding of (research) activities, global evaluation of results and optimal deployment of knowledge developed elsewhere
	Participation in: Parmenides, Ecolead, EGEE, DEISA, CrossGrid, EUGridPMA, CoreGrid (SSA) collaboration food safety, TeraGrid, OptIPuter, BIRN network (USA), NeuGrid, evaluation of the workflow tool Taverna from the MyGrid project (UK), Healthgrid, COMPUTIS (EU), OGF steering committee members (de Laat, Groep, Kielmann); CoreGrid SAB (Strategic Advisory Board) member; intensive collaboration with Grid '5000, K-WF. The establishment of global interoperable identity management for distributed resource access and co-founded the Shared Names initiative. Convergence with other projects is organized via participations in OGF, TERENA task forces, via the e-IRG. Collaboration on SAGE with University of Chicago. The DUTELLA infrastructure rolled out in several national and international research projects. Virolab, ACGT, QosCoSGrid and COAST are international spin-off projects. Participation in TREC challenges, OAEI workshop, BioCreative shared task, collaboration with the e-Research Centre at Monash University in Australia is emerging from a recent visit. Collaborations with the Kepler project in the US and the MyGrid project in the UK. Development of ESFRI research infrastructure LifeWatch (http://www.lifewatch.eu/).

	Awards: 2 first prizes at DACH (First IEEE International Data Challenge for finding Supernovae) 2008 (Japan), first prize at SCALE (IEEE International Scalable Computing Challenge) at CCGrid 2008 (France). Best paper award (HealthGrid 2008, Chicago); third place on best demo contest (EGEE User Forum 2008, Clermont-Ferrand), Nomination for best paper award (third, HealthGrid 2009, Berlin; HPDC 2007, Monterey), Leadership award at OGF20 2007. *Co-organized:* - Dagstuhl seminar Distributed Verification & Grid Computing (2008, Germany); - 4th International workshop on Workflow Systems in e-science (2008, China); - International workshop on Applications of workflows in Computational Science (2008,Poland); - 4th Workshop on Challenging Issues in Workflow Applications, (2008, USA); - F2F meeting W3C Semantic Web HCLS Interest Group (2008, France, 2009, Boston, MA, USA; 2009, Santa Clara, CA, USA); - Food Knowledge Symposium: AIDA in Food Informatics (2008, Amsterdam); - 1st and 2nd international workshop e-science infrastructure for animal tracking (2008, 2009, Amsterdam); - GEOSTAT international Summer School: Spatio-temporal data analysis: (2008, Amsterdam); - 1st workshop about medical imaging on Grids: MICCAI-Grid (2008, New York); - Euro-Par 2009 (Delft). - Semantic Web Applications and Tools for the Life Sciences Workshop (SWAT4LS). - Shared Names meeting (2009, Cambridge, MA, USA) - Scientific Discourse Workshop at Int. Semantic Web Conference (2009, Atlanta, USA) - Workshops in conjunction with Int. conferences ICCS, e-science, CCGrid; SWES06, SWBES06, SWES07, SWBES07, SWES08, WACS08, SWES09, and PC3 challenge 2009 - The Future of the Web for Collaborative Science Workshop (WWW2010, Raleigh, NC, USA) - 1st +2nd workshop about grids in health care: CCGrid-Health (2009 Shanghai, 2010 Melbourne)
5.	The level at which and the way in which communication of the results to the general public and non-scientific target groups is organized
	Interviews and publications via SenterNovem, AT5 interview, learning to experiment on distance, general public demonstration days, articles in popular press, lectures at semi-scientific conferences, lectures at companies, and many events organized by ICTregie. Talks and demonstrations have been presented at ICT Delta Kenniscongres 2006-2010, Surfnet ZorgSymposium 2006 and Siemens User's Day 2007. Biodiversity has given many interviews in news papers (from Intl. Herald Tribune to local journals) and several magazines (Quest). Radio interviews and television science program. Documentary on Cellular Automata, on Amsterdam local and national television. Interview in GenomeWeb. Interview and column in (NBIC) Interface magazine. Medical imaging interview published as an article in the AMC Magazine. Technocrossings event. Medical imaging article published on ISGTW newsletter. 8 publications in the professional journal Agro Informatica. Several news items published at TIFN extranet. Several publications in VoedingsmiddelenTechnologie (VMT). The VL-e project is extensively discussed and praised in the recently published COOK Report On Internet Protocol: 'Building a National Knowledge Infrastructure; How Dutch pragmatism nurtures a 21st century economy'.

Table 6. Outflow of staff

KI = Knowledge Institution, IN = Industry, OT = Other

sub-project	person	in VL-e	after VL-e	additional information
SP1.1	Willem van Leeuwen	KI	OT	Retired
	Jeff Templon	KI	KI	Now leader of grid computing project at Nikhef
	Sander Klous	KI	KI	Researcher in ATLAS experiment
	Kors Bos	KI	KI	Head ATLAS computing, located at CERN Geneva
	Rutger Kramer	KI	KI	Software Development Coördinator
SP1.2	TNO: not applicable			
	Diego Faneyte	KI	IN	WebApplications company, applied research
	Lars Hulzebos	KI	IN	Culios, technostarter food advisory for consumers
	Jelle Wisse	KI	IN	Meteoconsult
	Lobke van Oorschot	KI	OT	ZLTO belangenbehartiging landbouw Zuid-Nederland
SP1.3	Matthan Caan	KI	KI	Post-doc AMC, same organization, other project
	Ketan Maheshwari	KI	KI	PhD programme at CNRS Sophia Antipolis
	Michiel Scarpa	KI	OT	unknown
	Tristan Glatard	KI	KI	Researcher at CNRS Lyon
	Kamel Boulebiar	KI	IN	Sc.programmer Software company (FR)
	Jeroen Snel	KI	OT	unknown
	Johan Alkemade	KI	IN	Scientific programmer Software company (NL)
	Martin Stam	KI	IN	Scientific programmer Software company (NL)
	Sílvia Olabarriaga	KI	KI	Assistant professor, leader e-Bioscience group
	Robert Belleman	KI	KI	UvA: Computational Science
SP1.4	Willem Bouten	KI	KI	Project leader, same organization
	Emiel van Loon	KI	KI	UD, same organization
	Guido van Reenen	KI	KI	Same organization
	Edwin Baaij	KI	KI	Software developer, same organization
	Jasper Vrugt	KI	KI	Post-doc at Los Alamos National Laboratory
	Ammar Benabdelkader	KI	IN	Post-doc PCC
	Danny Cleary	KI	KI	CESAM Dept. Biology, Universidad Aveiro (Portugal)
	Klaske Grimmerink	KI	OT	Gemeente Almere, afdeling Statistiek
	Anders Bouwer	KI	IN	Company (UK)
	Scott Davis	KI	IN	Industry (USA)
	Floris Sluiter	KI	OT	Advisor e-science support HPC&V, SARA
	Tom Visser	KI	OT	Advisor e-science support HPC&V, SARA
	Judy Shamoun	KI	KI	UD, same organization
	Diana Gorea	KI	KI	Same organization
SP1.5	Dr. T.M. Breit	KI	KI	Project leader same organization
	Dr. S. Marshall	KI	KI	Post-doc same organization, VL-e sub-project SP2.2
	Dr. F. Wittink	KI	KI	Post-doc same organization
	Dr. X. Yung	KI	KI	Post-doc same organization
	Dr. M. Roos	KI	KI	Post-doc same organization, VL-e sub-project SP2.2
	Dr. M. De Jong	KI	IN	Post-doc spin-off company

	Dr. A. de Inda	KI	IN	Post-doc Industry
	Dr. S. Salamanca	KI	IN	Post-doc Industry
	Dr. R. Stad	KI	IN	Post-doc Industry
	Dr. C. Henkel	KI	IN	Post-doc Industry
	Dr. J. Dyczkowski	KI	IN	Post-doc Industry
	Dr. T. Pronk	KI	OT	Post-doc RIVM
	Dr. P. Sterk	KI	KI	Post-doc EMBL (UK)
	Dr. W. De Leeuw	KI	KI	Scientific Programmer same organization
	Drs. H. Rauwerda	KI	KI	Senior Scientist same organization
	Drs. O. Bruning	KI	KI	Junior Scientist same organization
	Drs. M. Brink	KI	KI	PhD ? post-doc VIB (Belgium)
	Drs. M. Itterson	KI	KI	Junior Scientist same organization
	Drs. R. Monajemi	KI	IN	Post-doc small self owned enterprise
	Drs. L. Post	KI	IN	AIO? programmer Industry
	Drs. Y. Tan	KI	OT	Junior Scientist Other (childcare)
	Ing. F. Verster	KI	IN	Scientist Programmer Industry
	Ing. M. de Haan	KI	KI	Technician same organization
	Ing. J. Verkooijen	KI	KI	Technician same organization
	Ing. W. Ensink	KI	IN	Technician spin-off company
	J. Batson	KI	KI	OBP same organization
SP1.6	Aleksey Merkulov	KI	IN	Post-doc company
	Ioana Taban	KI	KI	PhD
	Andrey Kharchenko	KI	KI	Post-doc same organization
	Basak Kaletas	KI	KI	Post-doc same organization
	Gert Eijkel	KI	KI	Software developer same organization
	Ivo Klinkert	KI	KI	Software developer same organization
	Marco Konijnenburg	KI	KI	Software developer same organization
	Erika Amstalden	KI	KI	PhD
	Ron M.A. Heeren	KI	KI	Group leader same organization
	Martin Froesch	KI	IN	Post-doc, company
SP2.1	Prof. Dr. P.M.A. Sloot	KI	KI	Project leader same organization
	Drs. Breanndán Ó Nualláin	KI	KI	Project co-leader same organization
	Dr. V. Krzhizhanovskaya	KI	IN	PhD ? Postdoc Industry
	Drs. G. Qiu	KI	KI	PhD Same organization
	Dr. R. Belleman	KI	KI	Same organization
	Dr. Z. Zhao	KI	KI	PhD: Same organization
	Dr. K.A. Iskra	KI	KI	PhD ? PD USA
	Dr. K.J. Rycerz	KI	IN	PhD ? Industry
	Dr. A. Tirado-Ramos	KI	KI	PhD ? PD USA
	Dr. V.V. Korkhov	KI	KI	PhD ? PD Russia
	Drs. G. Qiu	KI	KI	PhD same organization
SP2.2	Prof. Dr. P. Adriaans	KI	KI	Project leader same organization
	Dr. S. Marshall	KI	KI	Project manager LUMC
	Prof. Dr. F. van Harmelen	KI	KI	Same organization
	Prof. Dr. M. de Rijke	KI	KI	Same organization
	Prof. Dr. G. Schreiber	KI	KI	Same organization
	Dr. M. Roos	KI	KI	Post-doc LUMC
	Ing. F. Verster	KI	IN	Scientific Programmer Industry

	S. Katrenko	KI	KI	PhD? post-doc same organization
	W. van Hage	KI	KI	PhD? Post doc same organization
	E. Meij	KI	KI	PhD same organization
	A. Wibisono	KI	KI	Scientific Programmer same organization
	K. van den Berg	KI	KI	Scientific Programmer other organization
	K. Krommydas	KI	KI	Scientific Programmer other organization
	Frank Terpstra	KI	IN	PhD? Post-doc CapGemini
SP2.3	Robert van Liere	KI	KI	Project leader same organization
	Alexander Broersen	KI	KI	Post-doc LUMC
	Chris Kruszynski	KI	KI	Post-doc same organization
	Michal Koutek	KI	OT	RIVM
	Charl Botha	KI	KI	TUD
SP2.4	Cesar Garita	KI	KI	Post-doc Prof. Instituto Tecnologico de Costa Rica, Consultant at Innova Technology SA
	Hakan Yakali	KI	IN	Scientific Programmer Developer TomTom
	Ersin Kaletas	KI	IN	Post-doc Oracle e-Business Suite Technical Consultant Canon
	Djamel Abtroun	KI	IN	Scientific Programmer Company (France)
	Nader Mirzadeh	KI	IN	Scientific Programmer Software Architect Bell ID
	Ozgul Unal	KI	OT	PhD unknown
	Victor Guevara	KI	IN	Post-doc PCC
	Ammar Benabdelkader	KI	IN	Post-doc PCC
SP2.5	Vladimir Korkhov	KI	KI	PhD ? Post-doc State Un. of St. Petersburg (Russia)
	Dmitry Vasunin	KI	IN	Programmer started his own company in St. Petersburg
	Adianto Wibisono	KI	KI	Scientific Programmer same organization
	Zhiming Zhao	KI	KI	Post-doc same organization
	Philip Jonkergouw	KI	KI	Information Analyst at MWH (UK)
SP3.1	Rob van Nieuwpoort	KI	KI	Same organization and ASTRON
	Kees van Reeuwijk	KI	KI	TUD
	Gosia Wrzesinska	KI	IN	Vectorwise
	Niels Drost	KI	KI	Same organization
	Ceriel Jacobs	KI	KI	Same organization
	Kees Verstoep	KI	KI	Same organization
	Hashim Mohamed	KI	IN	financial industry
	Alexandru Iosup	KI	KI	Assistant professor same organization
	Ozan Sonmez	KI	KI	Post-doc same organization
SP4.1	Dennis van Dok	KI	KI	Same organization
	Jan Just Keijser	KI	KI	Same organization
	Ronald Starink	KI	KI	Same organization
	David Groep	KI	KI	Same organization
	Piter de Boer	IN	IN	Same organization
	Spyros Koulouzis	IN	IN	Same organization
	Ronald van Driel	IN	IN	Same organization
	Serge Vrijaldenhove	IN	IN	Same organization
	Wico Mulder	IN	IN	Same organization
	Maurice Bouwhuis	OT	OT	Same organization
	Axel Berg	OT	OT	Same organization