On-demand virtual research environments and the changing roles of librarians

Leonardo Candela, Donatella Castelli and Pasquale Pagano
Istituto di Scienza e Technologie dell'Informazione "A. Faedo", Consiglio Nazionale delle Ricerche, Pisa, Italy

Received 13 March 2009
Revised 14 March 2009
Accepted 14 March 2009

Abstract

Purpose – The aim of this paper is to discuss how new technologies for supporting scientific research will possibly influence the librarians' work.

Design/methodology/approach – The discussion is conducted in a context that takes into account the emergence of e-infrastructures as means to realise a new model of producing, using and sharing information resources and even to change the concept of information resource itself. At the core of this innovation there are virtual research environments, i.e. evolved versions of the current "research libraries".

Findings – The environments provide scientists with collaborative and customised environments supporting results production and exchange around the globe in a cost-efficient manner. The experiences made with these innovative research environments within the D4Science project is reported.

Originality/value – On the basis of this experience, possible professional profiles are suggested for librarians working in these new evolved "research libraries".

Keywords Librarians, Digital libraries, Communication technologies, Research, Scientists

Paper type Viewpoint

Introduction

In the recent past, digital technologies and communication networks have greatly influenced the way research activities have being carried out. Nowadays, the rapid evolution of technologies is preparing a considerable and wide-reaching change in the model research activities are organised, and knowledge is acquired, communicated and exploited. Not only areas like biodiversity, climate change, environmental monitoring, are affected by this change (GRL2020, 2008) but, as demonstrated recently, more traditional ones, like humanities and social sciences (Kuster *et al.*, 2007; DARIAH, 2009; Váradi *et al.*, 2008), are beginning to look at the new technologies as means to largely improve the performance of their researchers.

Virtual research environments (VREs) lay at the core of this digital innovation. They provide collaborative frameworks enabling scientists to produce and exchange results with peers around the globe in cost-efficient manner. They not only offer capabilities for accessing cross-disciplinary data and knowledge, but also provide

The D4Science project illustrated above is partially financially supported by the European Commission, Research Infrastructures, Grant Agreement No. 212488.

Special thanks go to Maria Bruna Baldacci for her valuable help in improving the readability of the paper.

0737-8831
DOI 10.1108/07378830910968191

services for exploiting a multiplicity of other disparate tools and computing resources enabling innovative analysis, simulation and domain specific knowledge generation processes.

Realising operational VREs able to satisfy the needs of the very diverse scientific user communities imposes a paradigm shift in the way in which research libraries operate nowadays. This change involves at least three major areas:

(1) the technology;

(2) the organisational model of the resources; and

(3) the human processes.

These three areas are not independent, but any choice in one of them strongly influences and constrains the others. So any progress in one area is expected to stimulate modifications and adaptations in the others, and also to start possible further changes in that area itself.

In this paper we analyse the implications related to the introduction of VREs starting from the technological perspective. Then we discuss the consequent impact on the other areas.

From the technological point-of-view, realising VRE applications is complex and usually quite demanding in terms of effort and required resources. These applications are expected to provide powerful and easy to use functionality for supporting scientific collaboration and in-silico experiments. VREs include services capable of performing selection, retrieval and usage of a multiplicity of cross-domain resources, each having its own interface and usage policies. These services must be orchestrated, monitored and managed in order to realise the domain specific workflows governing the generation of the expected results and to guarantee the necessary quality of service. *Ad-hoc* implementation of VREs is hardly sustainable due to their complexity. In order to reduce the cost and thus enable their diffusion, new technological solutions based on appropriate e-Infrastructures have emerged recently. These e-infrastructures provide basic functionality that facilitate the creation and operation of VREs. This functionality can range from the simple registry of shared resources and the monitoring of their status, to the on-demand creation of VREs. e-Infrastructures are usually operated by institutions that maintain them operational and guarantee their necessary quality of service. They operate in a conceptual framework that distinguishes among:

- external resources, managed by third-party resource providers which maintain and make the resources available to the e-Infrastructure under certain policies; and
- VREs, implemented as applications enabled by the e-Infrastructure, specifically designed to meet the needs of the scenarios addressed (see Figure 1).

The above technological solutions rely on a resource organisational model in which resource providers, which locally maintain and curate their own resources, agree on sharing them under certain policies. The shared resources may range from publications, multimedia material, sensor and experimental data, to tools that manipulate this data, and computing and storage resources. In this organisational model, the resource providers are required to adhere to the policies and guidelines

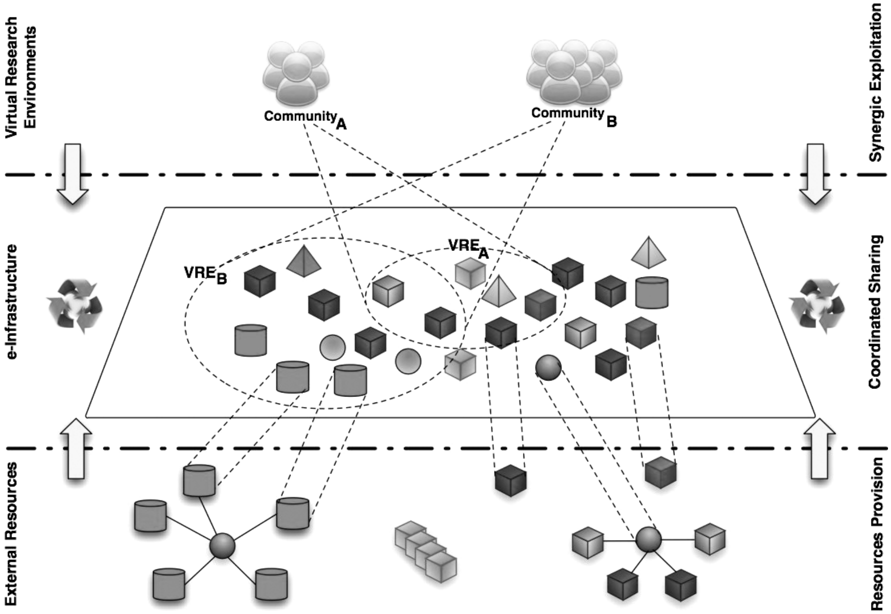

Figure 1.
E-infrastructure framework

established by the e-infrastructure in order to make these resources exploitable by third-parties and particularly by the supported VREs. The guidelines may require to specify not only information about the protocol to be used when accessing the resource but also semantic information about their role and capabilities. This semantic information turns out to be extremely important for people selecting the resources to be included in a specific VRE.

Naturally, the technological and organisational solutions just described impose an in-depth rethinking and revision of the typical processes of a library and of the roles played by library actors in these processes. In a world populated by VREs the ideal librarians must be qualified for satisfying the relevant requirements of the community they are serving. They must not only have an in-dept knowledge about existing information sources on the web, but also about the different tools and techniques to access them. Their role is expected to shift from organisers of stable collections to VREs designers defining transient communication channels among sparse people and resources. On the other hand, in the context given by the new technological and organisational model librarians are also required to play an increasingly more important role in curating and publishing resources on the side of resource providers so that these resources can easily be shared, discovered and exploited. This again requires a shift in the role of librarians – which must be prepared to operate in a global context characterised by rules and governing policies – that may differ from those of their own organisation.

The exact kind and modality of this shift is still unclear at the moment. However, initial experiments with VREs are starting to provide useful insights. In this paper, we describe the experience done with the VREs built in the framework of the D4Science project (D4Science, 2009). This is an EU co-funded project that has developed a VRE enabling e-infrastructure of the type described above. Currently, it runs four VREs serving scientific application scenarios in the Environmental Monitoring and Fisheries and Aquaculture Resource Management application areas. The global and cross-domain nature of these VREs make them particularly interesting cases from which initial insights on the impact and consequences of VREs can be drawn.

The rest of the paper is organised as follows: Section 2 introduces the D4Science e-infrastructure; Section 3 illustrates how resources are published in this e-infrastructure; Section 4 described how VREs are created; Section 5 discusses the role that librarians are expected to play the new D4Science envisaged framework; and finally, section 6 concludes.

The D4Science e-infrastructure

The D4Science e-infrastructure is the result of a large effort played in the last four years by 15 European and international research organisations, universities and software companies[1].

This e-infrastructure provides a rich framework for supporting the on-demand creation and operation of VREs. By exploiting this infrastructure a number of VREs have been created, serving large multidisciplinary scientific communities operating in challenging scientific contexts, like Environmental Monitoring (EM) and Fisheries and Aquaculture Resource Management (FARM).

D4Science assumes the organisational model depicted in Figure 1. It operates as a "broker" in a market of resources accommodating the needs of resource providers and consumers. Resources here are intended as shareable generic entities, physical (e.g. storage and computing resources) or digital (e.g. software, processes, data), that can interact with other resources to synergistically provide some functions serving their clients, either humans or automatic systems. In the current version D4Science supports resource providers in "selling" their resources, and resource consumers, i.e. the scientific communities, in "buying" and orchestrating such resources to build their VRE applications. Selling occurs through the publishing of these resources according to the policies established by their owners. By exploiting the e-infrastructure services and facilities, the proprietary formats and protocols used by these resources can be transformed into common ones to enable their cross-consumption.

The pool of resources shared by third-party providers is enriched by a set of service resources, i.e. software units which deliver generic digital library functions, like retrieval, access, annotation of content and creation of new one. This functionality, which constitutes a core part in the majority of the VREs, can be used as any other public resource by exploiting available physical resources, implemented and made available by the e-infrastructure itself. A unique feature offered by D4Science is its capability to support the on-demand set up of VREs. VREs tailored to specific needs of a scientific scenario can be created and maintained alive for the time they are required and dismissed when the community does not need them anymore (e.g. when a user community project comes to its end). The e-infrastructure provides communities with logistic and technical aids for VREs building, maintenance, and monitoring in order to reduce as much as possible the human intervention and facilitate these tasks. As it will be described more in depth in section 4, interactive tools have been made available to support the selection of the resources to be included in these environments from the pool of the available assets. Once selected, these resources are organised and manipulated by the e-Infrastructure in order to make the VRE operational, e.g. the services are deployed on specific servers, monitoring of these services is activated, reallocation is executed when needed. All these tasks are performed transparently to the user.

The D4Science e-infrastructure is in production mode since June 2008. By then it has been populated with several different resources. Particularly, the information resources published so far are quite heterogeneous. They range from multidisciplinary fisheries data sources, such as Fishery Country Profiles, National Aquaculture Legislation Overviews, Capture Time series graphs, species distribution maps, to very different Earth Observation products. By exploiting these resources, four VREs have been created at the moment serving application scenarios addressed by the EM and FARM communities.

The following two sections present more in details the functionality that D4Science offers to support the activities of resource providers and VRE designers.

Populating the e-infrastructure with resources

An e-Infrastructure can offer different degree of support to the resource providers. This ranges from common tools for advertising resources availability and facilitating their discovery to more complex ones for curating the resources and guaranteeing their quality, for abstracting over resources specificities and guaranteeing seamless access to heterogeneous resources, for equipping the resource management with autonomic capabilities.

At the current stage the D4Science infrastructure includes a number of services including:

- *Resources registration* – allowing resource providers to instruct the infrastructure on the specificities characterising a resource including the policies governing its usage. The information is used by the D4Science to properly handle that resource. Different kinds of resources require that different information is specified, e.g. if the resource is a web service implementing a specific functionality its URL has to be provided, if the resource is a data source both a characterisation of its content and the protocol governing the access to it must be given.
- *Resources ingestion* – allowing the e-infrastructure to automatically enrich the resource description explicitly specified at registration time and complement the resource with additional resources facilitating the exploitation of the initial one. For instance, in the case of data sources, metadata collections in specific schemas can be generated, new collections of information objects resulting from original data aggregation and manipulation can be produced, different indices supporting data discovery can be automatically generated.
- *Resources validation and approval* – allowing e-infrastructure administrators to analyse the characteristics of the external resources the e-infrastructure has been requested to manage and decide whether these resources are entitled to partake the infrastructure or not.
- *Resource monitoring* – allowing e-infrastructure managers (whether human or not) to be aware of the current status of a specific resource. The data needed to monitor the resource status are per resource, i.e. the status of different resources is characterised by different aspects (Figure 2). For instance, the status of a web service includes its workload, the status of a data source includes the number of information objects it contains.

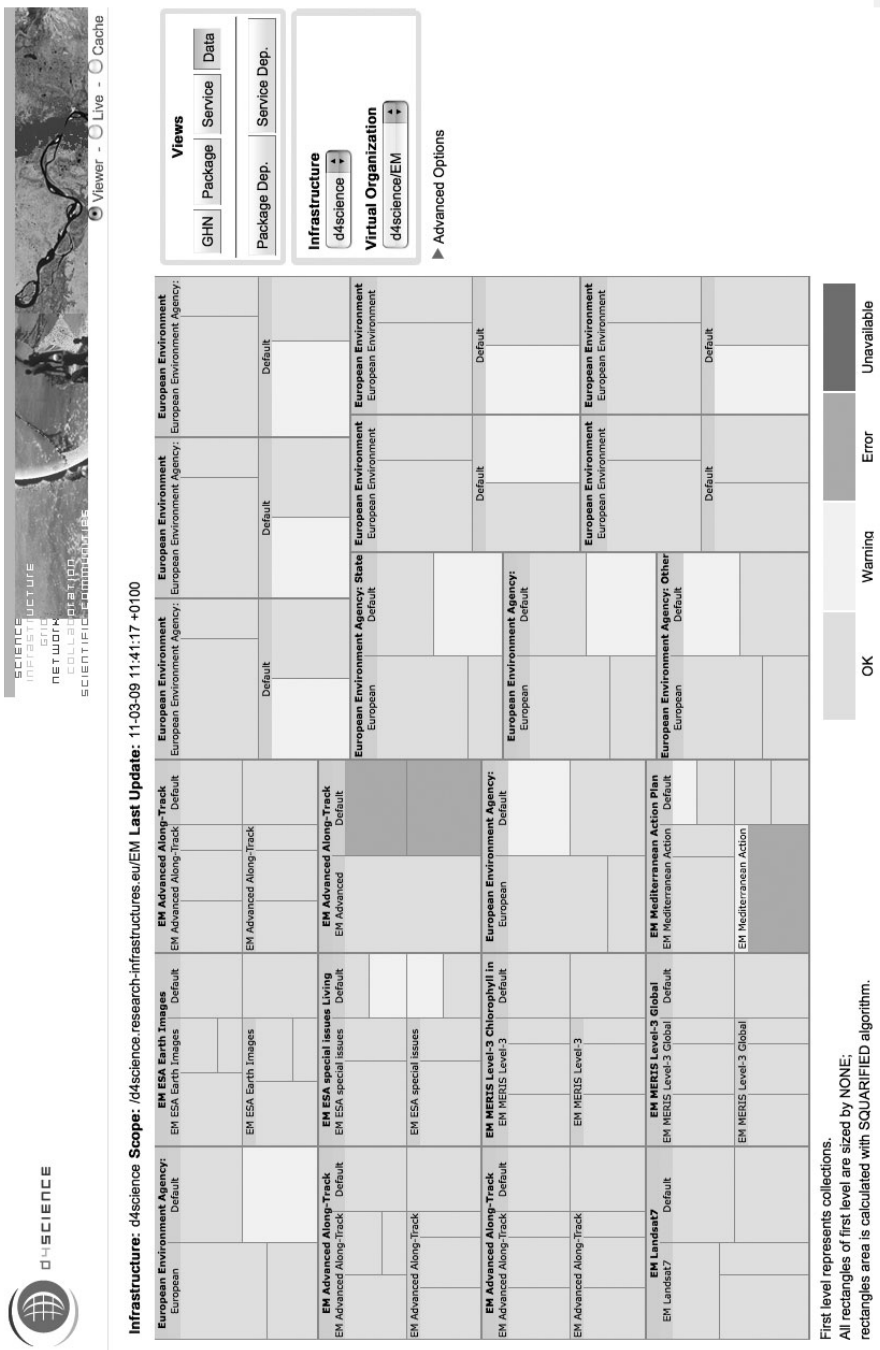

Figure 2.
E-infrastructure monitoring: data sources resources

Notice that in developing the D4Science e-infrastructure enabling system, i.e. *gCube* (Candela *et al.*, 2008; Pagano *et al.*, 2009), the realisation of the services above has been prioritised with respect to those supporting the digital curation. While the former lay at the core of any e-infrastructure supporting resource sharing, the latter can also be obtained by integrating in the current infrastructure services already developed in other contexts, e.g. those developed in the CASPAR project (Bonardi and Barthélemy, 2008). Actually, the D4Science project plans to move in this direction in a near future. The Service-oriented Architectural paradigm adopted in gCube and the intensive use of standards, such as XML and Web Service Resources Framework (Banks, 2006) facilitate such an integration. This will again have an impact on the digital curation processes that will evolve towards more standards methodologies and workflows, while being sustained by automatic services.

Building VREs

The D4Science e-infrastructure supports the creation and management of VREs by offering mechanisms for VRE definition, deployment and operation (Assante *et al.*, 2008).

The definition process is organised in steps, each allowing the designer to characterise different aspects of the expected VRE. These steps make it possible to collect the semantic information needed by the D4Science enabling system in order to automatically deploy the resources that are necessary to operate the VRE. The identification of these steps – and the dependences between them – has been strongly influenced by the digital library model presented in the DELOS Digital Library Reference Model (Candela *et al.*, 2007b). In particular, these steps aim at capturing VRE constituent elements belonging to the Content, Functionality, Users, and Architecture dimensions.

A wizard-based user interface guides the VRE definition process. Figure 3 contain a screenshot of the wizard service, namely the one related to collection selection.

The whole process consists in steps allowing the designer to provide the system with information on the VRE under creation like the VRE's name, a human oriented description of its role and the motivations leading to its creation, the name of the designer and the people entitled to manage it, the expected duration period. Subsequently, the designer is requested to specify the characteristics of the VRE information space, i.e. to choose which of the data sources available in the e-infrastructure have to be made available through the VRE, which metadata formats the objects of these collections should be described with. These metadata can either exist or be dynamically generated by relying on transformation programs, i.e. another resource that describe how to map data in a certain source format to a target format. In the subsequent step the designer is requested to select, among the available functions, the ones the VRE should be equipped with. In fact, thanks to the third party resources made available through the infrastructure, different implementations of the same function might occur and the designer should select the one appropriate to its application scenario. The final step consists in the selection of the servers that are needed to operate the so specified VRE in order to guarantee the expected quality of service. These servers are selected among the ones managed by the e-infrastructure.

Once the specification is completed, the VRE generation logic implemented by the D4Science infrastructure analyses it and derives an optimal deployment plan aiming at maximising the usage of existing resources and eventually including dynamic resource

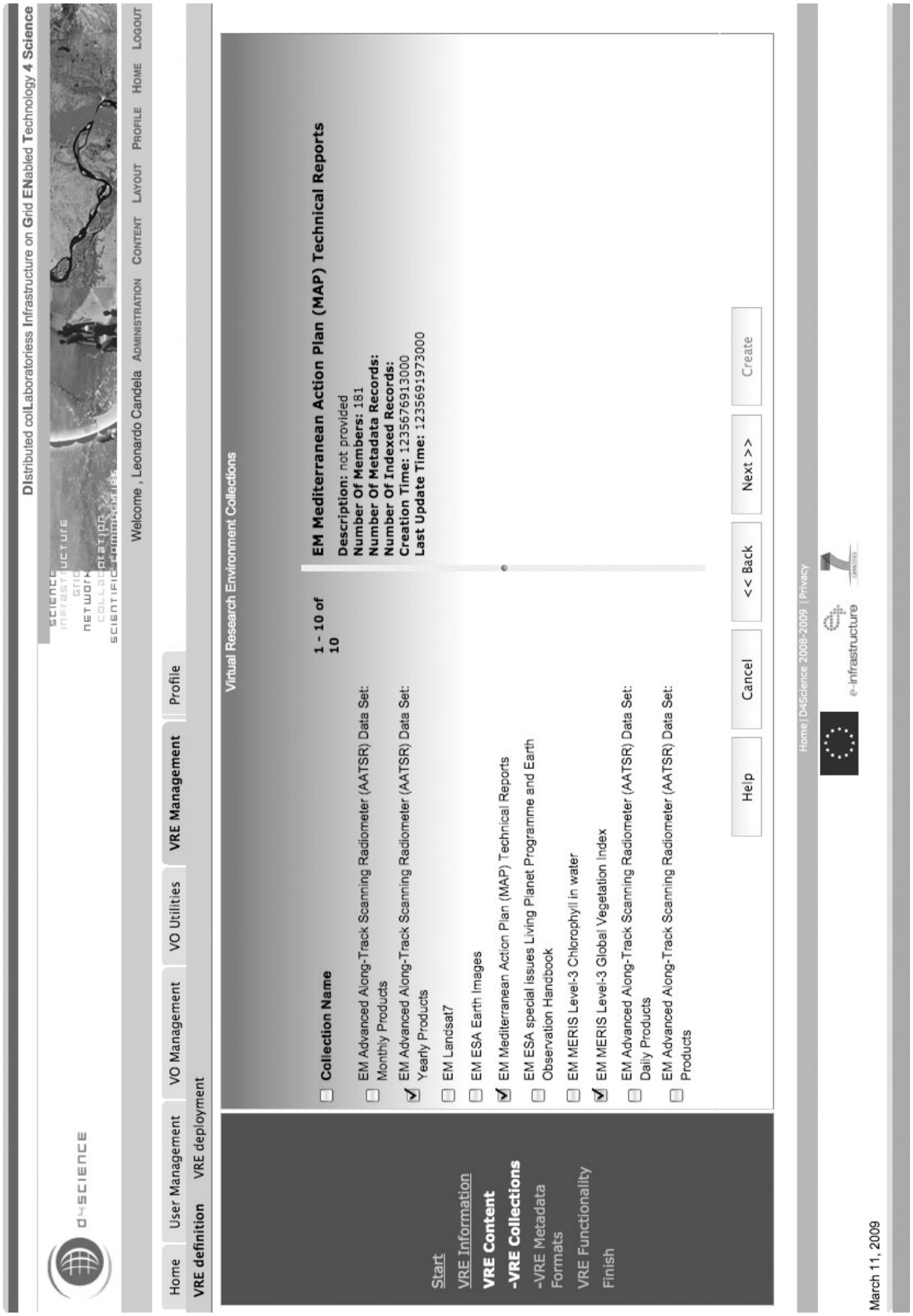

Figure 3.
VRE definition: the Wizard Service

generation. This generation is particularly critical since the quality of the VRE strongly depends on the choices done at this level. The infrastructure guarantees an optimal consumption of the available resources by selecting the minimal amount of them sufficient to meet its established performance and robustness criteria.

The heterogeneity of the resources and their dynamic nature makes this task particularly complex from the systemic point-of-view.

By using these mechanisms until now four VREs have been created serving very different application domains:

(1) *Fishery Country Profiles Production System (FCPPS):* supports scientists in the generation of fisheries and aquaculture reports. The production of country profiles requires complex aggregation and editing of continuously evolving multi-lingual data from a large number of heterogeneous data sources. Availability of the FCPPS VRE permits the scientists producing them to update and web-publish these vital reports as frequently as the community requires, while also having access to additional resources when needed.

(2) *Integrated Capture Information System (ICIS):* supports scientists in integrating regional and global capture and distribution information of aquatic species, from a number of Regional Fishery Management Organisations (RFMOs) and international organisations (FAO, WorldFish Center) into a common system. The VRE provides not only access to the necessary data but also a number of services for providing a harmonised view of catch statistics and allowing the community to merge/distribute such statistics over world areas, according to pre-defined reallocation rules.

(3) *Global Ocean Chlorophyll Monitoring (GCM):* provide scientists with an environment that integrates satellite data of microscopic marine plants and sea surface temperature. This environment supports research on biodiversity, by facilitating processes like the measuring of the distribution, the monitoring and modelling of phytoplankton (microscopic marine plants), the provision of forecasts of sea state and currents, the monitoring of algal blooms and marine pollution and the measuring of changes in the ocean productivity.

(4) *Global Land Vegetation Monitoring (GVM):* provides a virtual environment that integrates satellite images of vegetative land cover. It facilitates specific research on how climate changes and land cover influence environmental resources. By having access to the data and tools of this VRE, scientists can determine important measures like the total green leaf area for a given ground area, how much water will be stored and released by an ecosystem, how much leaf litter it will generate, and how much photosynthesis is going on.

As it should emerge from the brief description above, each of these four VREs offers an innovative collaboration environment. In this environment scientists addressing a specific problem can access a number of geographically disperse cross-domain resources of different nature and operate with them as if these resources were belonging to their own organisation (although in the limits imposed by the resources regulating policies).

The experience done so far in populating the D4Science infrastructure with resources provided by third-parties and in designing, creating, using and maintaining the above VREs has provided us with concrete feedback on the impact that VREs have on the processes and workflows performed by the scientists and by the librarians supporting them.

The following section reports some observations on the expected evolution of the librarians' roles derived by the analysis of this feedback.

The changing roles of librarians

In the past, information technology opened the way to shrewd librarians for designing how managing library services so that librarians could do their work more rationally and efficiently, and users could exploit such services at best. Today, instead, digital technology and communication networks are influencing scientists, suggesting new models for conducting and organising research work, and library changes are driven by research communities needs rather than by librarians' skill. VREs will reinforce this trend. The need for cross-border and cross-disciplines scientific collaboration between scientists and the availability of new technologies that make VREs feasible, demands for a change and a diversification of the library and librarians roles.

Many use cases and experiences will be necessary to fully define how these have to evolve to operate in the new technological and organisational frameworks operating VREs. Nevertheless, the experience gained in operating D4Science and its VREs already provide us with some useful insights on the expected changes.

As it should be clear from the description of the D4Science operational mode, all librarians will necessitate both IT skills and qualification in knowledge organisation. Also, linguistic competence will be very important. However, librarians engaged in providing resources and VRE designers will need different specialised expertises. This prevision is better motivated by making reference to Figure 1.

The basic elements of the new organisational model are information sources as shown at the bottom of the figure. In the digital framework surrounding the e-infrastructure, libraries, domain specific repositories and data centers will progressively become similar entities. It is also expected that research will become more and more data driven, and information sources will contain extremely large amounts of data produced by observations, simulations or similar sources besides the information objects traditionally managed by libraries. By exploiting their expertise, librarians will have to maintain the necessary domain specificities when needed but at the same time they will have to widely extend the scope of their work and become able to deal with different resources and different processes.

In the knowledge infrastructure devised in the Figure 1, the key element allowing research communities to create the basis for their working is the information service, where all the published resources are described so that these resources can be searchable and selectable. Managing such a service could seem nothing news for librarians, since associating metadata to information objects as well as describing the subject of such objects are fundamentals of the library profession. However, here "resources" include objects very different from those librarians are usually dealing with, as they span from textual documents to row data and a multitude of processing tools. A new specific profile for librarians might emerge from the need of aggregating and rearranging knowledge in different subject areas, as well as creating the tools and interfaces that allow VREs designers, be they librarians or scientists, to find and use the resources recorded in the information service. This profile should certainly include semantic content management, e.g. ontology management and development, and semantic provenance tracking issues (Norton, 2009).

In publishing resources, “new” librarians will have also to connotate such a content with the context in which it originated, so that it can correctly be interpreted by users. Further, they will have to drive the harmonisation of solutions and common standards that will progressively emerge to facilitate sharing, both at the level of metadata/information object models and access functionality. For example, they will have to become expert in accommodating the content of their library/repository/data center, so that data mining or similar services can succeed.

It is still questionable whether the resource curation functions will still be performed by the single institution or whether it will be assigned to third-party service providers. Whatever the solution, librarians will have a key role in dealing with knowledge organisation and digital curation. This implies the curation, preservation, maintenance, collection and archiving of digital assets. Librarians, in particular, will have an important role to play in helping scientists in knowing what to preserve and how to protect it while also providing open access. The experience acquired in the library context will be extremely valuable in this process, but the background of “librarians” will have to be widely extended, as they will have to deal with all the type of resources involved in the sharing. New librarians, for example, will have also to face the curation of the software repositories. All the typical digital curation functions will have to be applied in this case, but the different nature of the digital objects will require new skills.

VRE design and creation are other activities new librarians have to perform. These activities can be seen as an evolution of the more traditional one played by librarians when supporting the library users in accessing the library content and services. Given the heterogeneity of the available resources and the complexity of the scientific processes that VREs may be called to support, these activities certainly require multiple expertises. In particular, the new librarians must fully understand the needs of the specific research communities asking for the VREs and the characteristics of the available resources. They must be capable of selecting the resources to be included, deciding their most appropriate configuration, the functionality workflow, and so on. This means that they must have domain knowledge in the specific user community discipline, knowledge in information management, as well be trained in IT.

It is certainly not yet clear whether the entire VRE design, creation and maintenance process can be covered by a single professional. Despite that, an e-infrastructure like D4Science transparently supports the more systemic steps involved in creating and maintaining a VRE and in ensuring consistency in its design choices, the competences required are still wide. Certainly, however, innovative “librarians” profiles will have to emerge with complementary expertise from many disciplines (Lawton, 2009).

Concluding remarks

This paper has introduced the D4Science e-infrastructure as a technological instrument for supporting the on-demand creation of Virtual Research Environments. It has also discussed the effects and constraints it imposes on the other major aspects involved in the realisation of innovative VREs. In particular, the paper has focused its attention on the new processes that library and librarians have to play in the new context defined by e-infrastructures.

The new e-science world librarians have definitively to rethink their profession. The experience drawn from D4Science gives us some insights on the demand, it does not allow us yet to envisage how this evolution will happen. What seems to be apparent is that librarians' professional profiles will go toward becoming further and further specialised ones, corresponding to the requirements of the new organisational models research is going to be conducted. This is well summarised by Cathy Norton in (Norton, 2009):

> Research libraries will continue to work more closely with researchers so they may continue the role of guardians of the institutional intellectual capital to preserve and disseminate the information thus increasing its value. This value involves moving the intellectual capital beyond its traditional institutional or project walls so that it can fully operate on a global level.

Note

1. D4Science is a continuation of the DILIGENT project (Candela *et al.*, 2007a).

References

Assante, M., Candela, L., Castelli, D., Frosini, L., Lelii, L., Manghi, P., Manzi, A., Pagano, P. and Simi, M. (2008), "An extensible virtual digital libraries generator", paper presented at the 12th European Conference on Research and Advanced Technology for Digital Libraries, ECDL 2008, Aarhus, September 14-19.

Banks, T. (2006), "Web Services Resource Framework (WSRF) – Primer v1.2. OASIS Committee Draft", available at: http://docs.oasis-open.org/wsrf/wsrf-primer-1.2-primer-cd-02.pdf (accessed March 10, 2009).

Bonardi, A. and Barthélemy, J. (2008), "The preservation, emulation, migration, and virtualization of live electronics for performing arts: an overview of musical and technical issues", *Journal on Computing and Cultural Heritage (JOCCH)*, Vol. 1 No. 1, pp. 1-16.

Candela, L., Castelli, D. and Pagano, P. (2008), "gCube: a service-oriented application framework on the grid", *ERCIM News*, Vol. 72, pp. 48-9.

Candela, L., Akal, F., Avancini, H., Castelli, D., Fusco, L., Guidetti, V., Langguth, C., Manzi, A., Pagano, P., Schuldt, H., Simi, M., Springmann, M. and Voicu, L. (2007), "DILIGENT: integrating digital library and grid technologies for a new earth observation research infrastructure", *International Journal on Digital Libraries*, Vol. 7 Nos 1-2, pp. 59-80.

D4Science (2009), *DIstributed colLaboratories Infrastructure on Grid ENabled Technology for Science*, EC Contract 212488, available at: www.d4science.eu (accessed March 10, 2009).

DARIAH (2009), *Digital Research Infrastructure for the Arts and Humanities*, available at: www.dariah.eu (accessed March 10, 2009).

GRL2020 Europe (2008), "Outcomes of GRL2020 Europe", Tirrenia, Pisa, 27-28 March, available at: www.grl2020.net/index.php/grl2020-europe (accessed March 10, 2009).

Kuster, M.W., Ludwig, C. and Aschenbrenner, A. (2007), "TextGrid as a digital ecosystem", *Proceedings of the Digital EcoSystems and Technologies Conference, DEST 2007, 21-23 February 2007*, pp. 506-511.

Lawton, F. (2009), "Position Paper 4.12", paper presented at the GRL2020 Asia, 24-25 February, Taipei, available at: www.grl2020.net (accessed Mach 7, 2009).

Norton, C. (2009), "Position Paper 4.17", paper presented at the GRL2020 Asia, 24-25 February, Taipei, available at: www.grl2020.net (accessed Mach 7, 2009).

Pagano, P., Simeoni, F., Simi, M. and Candela, L. (2009), "Taming development complexity in service-oriented e-infrastructures", *Zero-In*, Vol. 1 No. 1, pp. 19-21.

Váradi, T., Krauwer, S., Wittenburg, P., Wynne, M. and Koskenniemi, K. (2008), "CLARIN: Common Language Resources and Technology Infrastructure", in European Language Resources Association (ELRA) (Ed.), *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08), Marrakech, Morocco, 28-30 May 2008*, available at: www.lrec-conf.org/proceedings/lrec2008/summaries/317.html (accessed March 2, 2009).

Further reading

Candela, L., Castelli, D., Ferro, N., Koutrika, G., Meghini, C., Ioannidis, Y., Pagano, P., Ross, S. and Soergel, D. (2007), *The DELOS Digital Library Reference Model – Foundations for Digital Libraries*, DELOS Network of Excellence on Digital Libraries, Pisa.

Theng, Y.L., Foo, S., Goh, D. and Cheon, J. (Eds.) (2009), *Handbook of Research on Digital Libraries: Design, Development, and Impact*, IGI Global, Hershey, PA.

Corresponding author

Leonardo Candela can be contacted at: candela@isti.cnr.it