

---

## D-Lib Magazine

January/February 2015  
Volume 21, Number 1/2

---

### Science 2.0 Repositories: Time for a Change in Scholarly Communication

Massimiliano Assante, Leonardo Candela, Donatella Castelli, Paolo Manghi and Pasquale Pagano  
Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Italy  
{assante, candela, castelli, manghi, pagano}@isti.cnr.it

DOI: 10.1045/january2015-assante

---

#### Abstract

Information and communication technology (ICT) advances in research infrastructures are continuously changing the way research and scientific communication are performed. Scientists, funders, and organizations are moving the paradigm of "research publishing" well beyond traditional articles. The aim is to pursue an holistic approach where publishing includes any product (e.g. publications, datasets, experiments, software, web sites, blogs) resulting from a research activity and relevant to the interpretation, evaluation, and reuse of the activity or part of it. The implementation of this vision is today mainly inspired by literature scientific communication workflows, which separate the "where" research is conducted from the "where" research is published and shared. In this paper we claim that this model cannot fit well with scientific communication practice envisaged in Science 2.0 settings. We present the idea of Science 2.0 Repositories (SciRepos), which meet publishing requirements arising in Science 2.0 by blurring the distinction between research life-cycle and research publishing. SciRepos interface with the ICT services of research infrastructures to intercept and publish research products while providing researchers with social networking tools for discovery, notification, sharing, discussion, and assessment of research products.

---

#### 1 Introduction

In the last decade, information and communication technology (ICT) advances have deeply changed the way research is conducted within research infrastructures (RIs). A Research Infrastructure is intended as the compound of elements regarding the organization (roles, procedures, etc.), the structure (buildings, laboratories, etc.), and the technology (microscopes, telescopes, sensors, computers, Internet, applications, etc.) underpinning the implementation of scientific research. Research is based on digital research products, such as datasets, software, services, and generates further digital products. Along the same line scientific communication has mutated in order to adapt the underlying business models and mission to such new scenarios. Indeed, the traditional paradigm of research publishing by articles cannot cope with the increasing demands of immediate access and effective reuse of research results. Scientists, funders, and organizations are therefore pushing for innovative scientific communication workflows (deposition, quality assessment and dissemination), marrying an holistic approach where publishing includes in principle any product (e.g. publications, datasets, experiments, software, web sites, blogs) resulting from a research activity, that is relevant to the interpretation, evaluation, and reuse of the activity or part of it.

The implementation of this vision is today mainly inspired by literature scientific communication workflows, which separate the place where research is conducted, i.e. RIs, from the place where research is published and shared. In particular, research products are published "elsewhere" and "on date", i.e. when the scientists believes the products obtained so far are mature enough. In our opinion, this model cannot fit well when other kinds of research products are involved, for which effective interpretation, evaluation, and reuse can be ensured only if publishing has the properties of "within" the RIs and "during" the research activity.

In this paper we present the notion of Science 2.0 Repository (SciRepo). Living in synergy with RIs, SciRepos meet research publishing requirements arising in Science 2.0 settings by blurring the distinction between research life-cycle and research publishing. In particular, by relying on social networking practices they provide researchers with collaboration oriented facilities enabling a seamless and complete access to any research product in the context leading to it. Finally, we present the idea of a SciRepos platform, a system facilitating the realization of SciRepos on top of existing RIs.

---

#### 2 Research infrastructures and modern scientific communication workflows

Research Infrastructures are the setting supporting scientists at performing their research activities, which generally consist in running experiments relying on existing research products (e.g. publications, datasets, software, manuals, services, processes, web sites, blogs) in order to yield new research products. In such scenarios, ICT services are becoming increasingly essential to perform research activities. They may range from simple computers and connection to the Internet (e.g. web and email) to data centres offering computational resources (e.g. web servers), services for data management (e.g. document stores, column stores) and processing (e.g. workflow management).

ICT services are intended not only for supporting scientific investigation, but also for publishing and re-using the resulting research products. Today's scientific communication workflows are based on the availability of Internet connection and devices, which make drafting, publishing, and accessing scientific publications in digital form the norm for the average scientists. Moreover, ICT services have been playing a central role in shaping up modern forms of scientific communication, which are today reaching beyond publishing articles in digital format. For example, many RIs provide scientists with ICT tools for the elaboration of large quantities of data, and the community invest energies into collecting, curating, and creating research data. Such trends, stimulated funding agencies, organizations, and researchers to find ways to publish research data [8][10]. Evidence of this is provided by the diffusion of data repositories (e.g., GigaDB, Dryad, FigShare, Pangaea) and by the establishment of initiatives studying data citation format and data citation indexing [13]. Recent investigations are reinforcing such new paradigms by studying the problem of publishing research experiments, intended as the methodological processes or ICT-based workflows necessary to achieve given scientific conclusions [14]. The objective is to offer researchers all the elements to repeat ("same experiment, same lab"), replicate ("same experiment, different lab"), reproduce ("same experiment, different configuration"), or reuse ("include part of the experiment into another experiment") [3][5]. Finally, ICT services offer scientists tools through which they can create and share alternative forms of research products, which are not generally intended as valuable for publishing. Examples are software, web sites, blogs, notes, chats discussions, electronic notebooks, etc. Several studies on "altmetrics" are today being conducted to understand how to enable certificated evaluation and citation methodologies for such products [16].

In summary, the advent of ICT facilities are today paving the way towards modern scientific communication workflows, where the act of "publishing" is invested of a newer holistic interpretation. Researchers should be able to publish literature, datasets, experiments, any form of research outcome they perceive to be important for the interpretation and reuse of their scientific results. The benefits are clear:

1. Better interpretation of scientific results.
2. More rigorous, possibly automated, evaluation of the research outcomes.
3. Omni comprehensive scientific reward practices.
4. Maximization of research reuse, thereby reducing the costs of research.

---

### 3 Methodological barriers to modern scientific communication: research product de-contextualization

When referring to the action of "publishing", most people would refer to the scientific communication practices that are typical of research literature. These practices are (a) supported by policies and services of a "research marketplace", intended as the set of online services thanks to which publications can be shared (e.g. discovered, accessed, cited, referred, interlinked, tagged) by scientists, (b) applied to selected research products while research activity is still ongoing, i.e. it is up to the scientists involved in a research activity to decide "what" is a candidate research product and "when" to publish it. Publishing consists of the following phases:

1. **Deposition:** scientists deposit research products into accredited repositories (e.g. Institutional repositories, Journal's repositories);
2. **Quality assessment:** scientists submit their candidate products to a peer-review process of some kind (e.g. single/double blind, open peer review);
3. **Dissemination:** besides repository-provided dissemination tasks, there are a number of web applications (e.g. Google Scholar, DBLP) taking care of aggregating, indexing, and cataloguing publication metadata, in order to provide advanced publication discovery mechanisms and citation indexes.

Publishing is usually conceived as the concluding step of the research activity lifecycle (cf. Figure 1). It comes conceptually after the research activity step – i.e. the phase leading to the production of research results – although this does not imply that this step is complete. It is expected that a new research activity lifecycle starts by using the results of previous lifecycles manifested in published products.

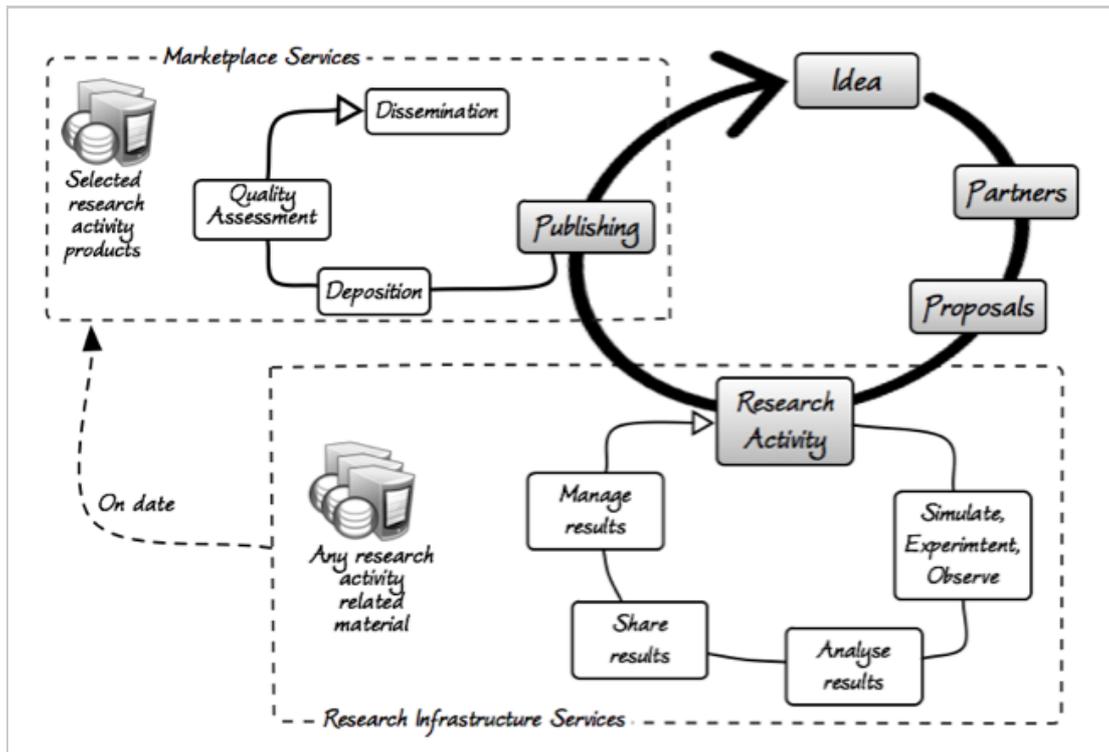


Figure 1: Publishing in a Research Lifecycle

As highlighted by many, e.g. [6][7], publishing research products different from scientific literature is hindered by several "cultural barriers", intended as the partial or full absence of well-established communication workflows defining what it means to publish, peer-review, citing, and guarantee scientific reward for products that are different from the traditional scientific publication [19]. Such lack of common understanding impedes and in some cases discourages scientists willing to share their results. As a result, research products of several kinds still sits in "the researchers desk drawer" or in the ICT services of RIs.

However, in our opinion, the reason behind the slow growth and establishment of new scientific communication workflows is not only cultural, but also methodological. In fact, most practices and technologies proposed today to renovate scientific communication tend to reflect the literature publishing workflows. According to this model, research activities are conducted into the RI, while publishing of the relative publications takes place "elsewhere", on the research marketplace, and "on date", when the researcher believes the publication is mature enough. The same methodology and attitude is applied to research data publishing, where products are eventually stored into data repositories, and to experiments, for which dedicated experiment repositories are being devised (e.g. myexperiment.org [14]). The simple idea is that of extending the marketplace with new sources dedicated to publish new kinds of digital products, and to implement submission and peer-review tools similar to those existing for literature.

Although we reckon this approach to be pragmatically correct, since researchers are accustomed to this way of thinking, we also believe that its immediate side-effects are counter productive and hindering the implementation of proper scientific communication in Science 2.0. This is because Science 2.0 research activity, being supported by RI-oriented ICT services, is (i) strongly contextualized and (ii) intrinsically dynamic; these features and requirements conflict with the "elsewhere" and "on date" philosophy of literature scientific communication workflows. According to these, products leave the RI ICT services to be transferred and deposited in marketplace repositories of specific kind (e.g. repositories for publications, datasets, and experiments). As such they are subject to the relative metadata/file deposition idiosyncrasies and management policies. As a consequence, published products suffer from the drawbacks discussed in Table 1 in addition to the rest of potential problems usually associated with scientific communication, i.e. *no communication*, *slow communication*, *incomplete communication*, *inaccurate communication*, or *unmodifiable communication* [18].

<i>Publishing Phase</i>	<i>Drawbacks</i>
Deposition	<b>Decontextualisation:</b> Although once published research products are annotated with metadata referring to the context leading to them (e.g. provenance), in the reality these products are deprived of any relationship to the original research activity, i.e. the notion of research activity does not survive in an effective way in marketplace repositories.

	<p><b>Staticity:</b> Published products are frozen to their publishing status, i.e. marketplace repositories often contain snapshots at the time of publishing of the products and they are not concerned with their evolution over time.</p> <p><b>Extra Cost:</b> Expensive to transfer and maintain when copies of the products need preparation before being transferred (e.g. anonymization) or entail hardware and administration cost for their management (e.g. disks, synchronization, IPR issues).</p>
Quality Assessment	<p><b>Ineffective peer-review:</b> The real evaluation of research products other than papers can be hardly performed out of the scope of the research activity or RI and in most cases without the support of dedicated ICT services (e.g. evaluate quality of large datasets or alternative products).</p>
Dissemination	<p><b>Fragmentation:</b> Research products are scattered across several marketplace repositories, i.e. scientists willing to re-use products published by others must interact with several end-points to find what they want (e.g. Google Scholar, DataCite, Google, repositories); for some products (e.g. blogs, websites), such sources are only search engines, since there is no dedicated marketplace repository for their publishing.</p> <p><b>Lack of semantic linking:</b> There is no guarantee that published products contain relationships between them, since such links have to be specified and maintained overtime by the authors across several marketplace repositories (e.g. dataset and publication repositories) or are not even maintained by such repositories.</p>

Table 1: Drawbacks affecting Publishing practices

Drawbacks discussed in Table 1 limit the effective interpretation of research results, hence their correct evaluation and reuse, and reduces the number of products eligible for publishing. For example, the staticity issue is important when the RI production chain is characterized by high velocity and dynamicity, it is not-trivial to decide which products and when are worth publishing; e.g. datasets can be dynamic (e.g. versioned, staged, query results), and deciding which stage/version of the data should be published implies some form of selection. As a consequence, some products may live in their RIs but never be published due to the implicit drawbacks of publishing.

#### 4 Science 2.0 Repositories

To enable effective scientific communication workflows, research product creation and publishing should both occur "within" the RI (as opposed to "elsewhere") and "during" the research activities (as opposed to "on date"). To make this possible, research infrastructures ICT services should not only be devised to provide scientists with facilities for carrying out their research activities, but also to support marketplace like facilities, enabling RI scientists to publish products created by research activities and other scientists to discover and reuse them. In other words, RIs should not rely on third-party marketplace sources to publish their products, rather should integrate them into the RI.

Such a merge between research infrastructure and research marketplace would overcome known "cultural barriers" and the "methodological barriers" mentioned above. In the RI scope, research products publishing would:

1. be facilitated by the very ICT services that are generating them;
2. take advantage of research activity awareness and product interlinking;
3. support access rights issues that fit the need of the community; and
4. subsume the major costs of storing and curating products. In other words, by bringing marketplace features within the RI, researchers can finally achieve their best effort in terms of scholarly communication.

Unfortunately, current repository platforms are not apt to implement this vision, as they are designed not to integrate with existing RI ICT services but to support instead today's notion of "elsewhere" and "on date" research marketplace. In this paper, we propose an innovative class of repositories, named *Science 2.0 Repositories (SciRepos)*. SciRepos are characterized by the following features:

- o Integrate with RI ICT services in order to intercept the generation of products within **research activities** and publish such products, that is making them discoverable and accessible to other researchers;
- o Provide scientists with repository-like tools for accessing and sharing research products generated during their **research activities**;

- o Rely on *social networking* practices [15][20] thus to modernise (scientific) communication both intra-RI and inter-RI, e.g. posting rather than deposition, "like" and "open discussions" for quality assessment, sharing rather than dissemination.

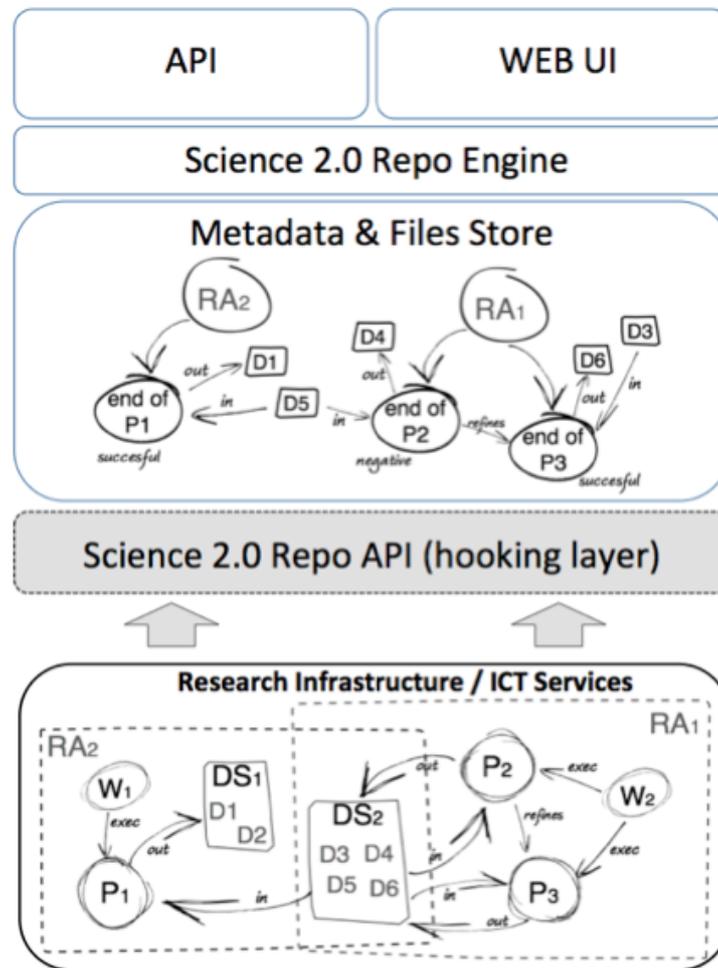


Figure 2: Example of SciRepos integrated with RI ICT services

Figure 2 illustrates a SciRepo running on an hypothetical RI. The RI enables two research activities RA<sub>1</sub> and RA<sub>2</sub> and keeps track of the executed experiments, their input and output data, and their final status (success or unsuccess). For example, in RA<sub>1</sub> researchers run experiments by executing workflow W<sub>2</sub>. Each execution of the workflow, e.g. processes P<sub>2</sub> and P<sub>3</sub>, collects input data and deposits output data from and into the local store DS<sub>2</sub>. In particular, process P<sub>3</sub> has refined the unsuccessful execution P<sub>2</sub> of W<sub>2</sub>, and improved the experiment to make it successful. The SciRepo sits on top of the RI which interfaces its ICT services with the repository in order to disclose the outcome of research activities to SciRepo users. The figure shows that the repository consists of a metadata store and a file store. The former stores an information graph representing research activities, their related products, and the relationships between them, while the latter can store product payloads originally residing out of the RI (e.g. publications, alternative products). Products can be of different typologies, e.g. workflows, executed workflows, datasets; their metadata can include different information, e.g. descriptive, attribution, provenance, rights, versioning, execution status, execution parameters, quality; their relationships may represent different associative semantics, e.g. input to process, output to process, refines process. It is important to note that the graph is populated automatically by the hooking layer during the research life-cycle and without scientists being directly involved in the actual action of publishing. The SciRepo supports scientists with two kinds of end-user functionalities:

- o **Repository-oriented facilities:** they offer typical repository facilities on the information graph such as search and browse allowing search by product typology, but also to navigate from research activities to products and related products. It offers ingestion facilities, allowing scientists to manually or semi-automatically upload "external" products into the repository and associate them to a research activity, thus including them in the information graph. Examples are publications, but also alternative science products, such as web sites, blogs, slides, documentation, manuals, etc. Ingestion allows scientists to complete the action of publishing a research activity with all products

that are connected to it but generated out of the boundaries of the RI. The way scientists or groups of scientists can interact with products (access and reuse them) is ruled by clear rights management functionalities. Rights are typically assigned when products are generated in the RI or ingested by scientists, but can vary overtime.

- o **Collaboration-oriented facilities:** they offer typical social networking facilities such as the possibility to subscribe to happenings relative to research activities and products and be promptly notified, e.g. the completion of a workflow execution, the generation of datasets obeying to some criteria. Users can reply to posts and, most importantly, can express opinions on the quality of products, e.g. "like" actions or similars. More sophisticated assessment/review functionalities (single/double blind) can be supported, in order to provide more traditional notions of quality. Interestingly, posts are themselves a special typology of products of the research activity and are searchable and browsable in the information graph.

In order to implement a SciRepo, RIs should develop their own software, thereby investing in a direction that requires different kinds of skills and dedicated funds. In order to facilitate this process we are designing and developing a SciRepo platform, devised to support the implementation of SciRepos at minimum development cost for the RIs, to be described in the next section.

The Science 2.0 Repository platform implements the conceptual model described below out-of-the-box. RI developers can freely instantiate this model to match the publishing expectations of their scientists by implementing what is envisaged in the hooking layer. Developers can also customize their SciRepo by enriching the data model specification with directives regarding how the different functionalities should be instantiated with respect to it. For example, directives may specify how end-user interfaces should enable discover and browse of the information graph, e.g. which product typology and metadata fields should be displayed, browsable, post-able, assessable. Similarly, directives can be used to configure export APIs, e.g. protocol, subset of information graph to be exported. Most importantly, given the data model, the platform generates the APIs required by RI developers to write the "hooks" needed to interconnect their ICT services with the platform and enable "during" publishing workflows. In the following we give more details of the platform data model and on the publishing functionalities it will offer.

#### 4.1 SciRepos Conceptual Model

A SciRepo is called to support users by providing them with a set of facilities for managing Research Activities with respect to scholarly communication practices. Like any other system, all of this is regulated by policies, e.g. who can do what. This very basic model is in Figure 3 below.

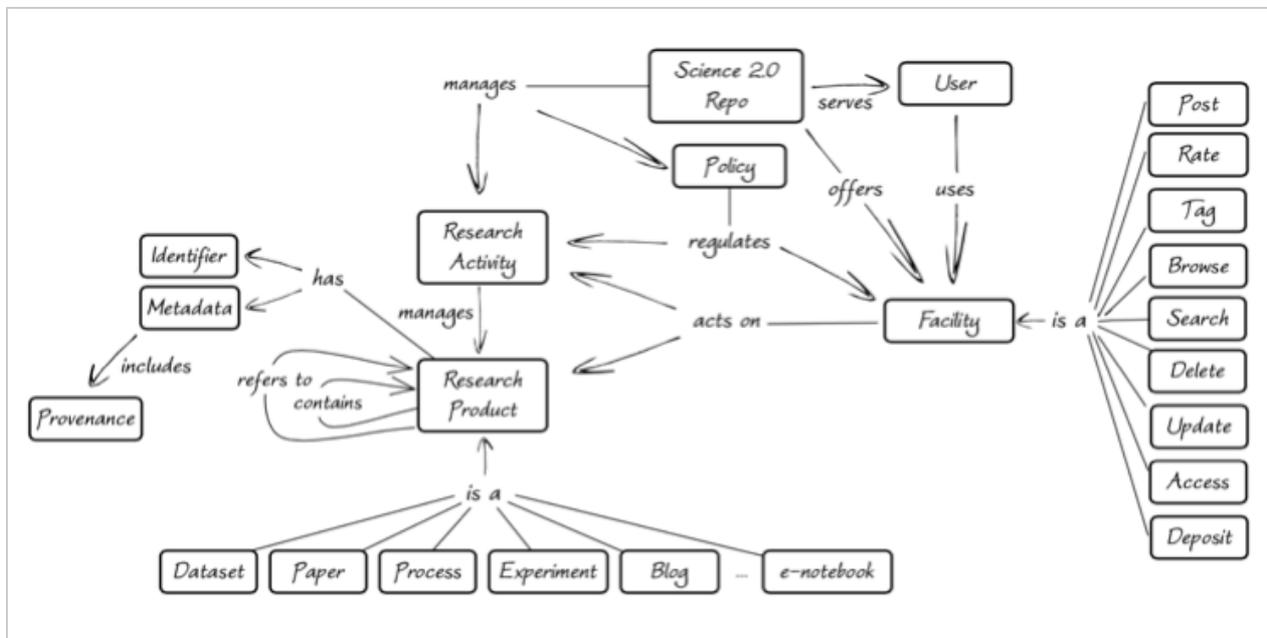


Figure 3: Science2.0 Repos Conceptual Model

A "research activity" is here intended as a set of actions, with a start date and an end date, carried out by researchers to achieve scientific results. As such they are generally performed by using RI ICT services, can use existing products, execute experiments, and create further products. Implicitly, products generated within the scope of a research activity are related with it and can have semantic relationships between them (e.g. citedBy, versionOf, inputDataset). SciRepos vary in the nature of the products they support, depending on the specific research context. Any output of the research process is potentially a relevant research product, as such it may be subject to publishing and be related with each other in time and

semantics. Examples include papers, datasets, blogs, web sites, electronic notebooks, external discussion threads, experiments, services, but also SciRepo research activities and SciRepo discussion threads. Both research activities and research products are provided with an identifier and with comprehensive metadata (potentially in multiple formats) automatically generated via the hooking facility.

The functionalities a SciRepo is expected to offer are oriented to Science 2.0 setting and with the scholarly communication. They include repository-oriented (Deposit, Access, Update, Delete, Search, Browse) and collaboration-oriented (Post, Rate, Tag). How these facilities can be realised to overcome the issues identified in [Section 3](#) is described in the following section.

---

## 4.2 Publishing cycle in SciRepos

The combination of integration with RI, rights and quality information about products introduces a novel publishing paradigm where "publishing" is intended as making a product *online available, discoverable, peer-reviewable, re-usable* according to given rights, *real-time accessible, citable*, and *interlinked* with its research activity and associated products. Possibly according to the [FAIR principles](#).

SciRepos can be considered as RI-oriented sources in the research marketplace. They offer functionalities allowing scientists to publish products automatically or manually, discover and access products according to their metadata descriptions and end-user access rights, peer-review products according to several evaluation models. In addition, they can also integrate interoperability mechanisms to move products in and out the boundaries of the RI.

In this section, we shall present the publishing phases (Deposition, Quality Assessment, Dissemination) as realised by a SciRepo, accompanied by the relative benefits in terms of overcoming the aforementioned methodological drawbacks (cf. Table 2).

### 4.2.1 Deposition

SciRepo offer both *automatic deposition*, i.e. in the style of RI services usage, and *manual deposition*, i.e. in the style of marketplace repositories.

Automated publishing is achieved by connecting the SciRepo with the underlying ICT services of the RI, in order to intercept the creation of products, publish them, and notify interested scientists of this. Such integration is part of the design of the SciRepo, is based on the publishing needs of the given RI community, and manifests in the implementation of the hooking layer envisaged. For example, the community may be willing to publish as a product new datasets generated by experiment into ICT storage services, or publish the execution of an experiment and the relative results (i.e. application of an algorithms over given input data, together with resulting data). To make this possible ICT services should communicate with the SciRepo to notify that products have been produced within a research activity context, with a given unique web resolvable identifier, and metadata description. Notification of products creation translates in SciRepo events, highlighted to scientists via social tools as posts of type "publishing". Scientists find in the post the link required to access the original data, can start a thread of discussion about the post, can forward the post according to standard practices.

Publishing can also occur manually, typically for all products relative to a research activity that are not automatically produced by the RI ICT services during experimentation. For example, interesting web sites created by scientists, threads of discussions in online blogs, technical documentations, software, scientific publications, datasets produced out of the RI boundaries, etc. In all such cases, scientists access the SciRepo and deposit under a given research activity a product of a given type, together with descriptive metadata. The product can be deposited locally or just be referred to from where it is online accessible. As such, this action resembles deposition operations typical of publications and datasets repositories, with the important difference that products are ingested in the context of a given research activity and are notified to scientists with a post. As such, they are implicitly linked to all products of the research activity and also associated to one or more discussion threads in the SciRepo.

### 4.2.2 Quality Assessment

SciRepos should support both traditional forms of single/double blind peer-review and alternative forms of peer-review, counting on social tools (e.g. likes, discussion threads, marks) and underlying RI ICT services that can (i) automatically verify and rank quality or compliance of products to agreed community quality indicators (e.g. dataset conforming to standard formats, within given size or value ranges), or (ii) record access logs, analytics and accounting (e.g. altmetrics).

An example of the personal web page every SciRepo user is provided with is in Figure 4. This is a sort of console where the user is acquainted with the happenings occurring in its community (e.g. Research Activities outcomes) as well as have easy access to the Research Activities he / she is involved in. Through it the user can express his / her position with respect to the ongoing activities.

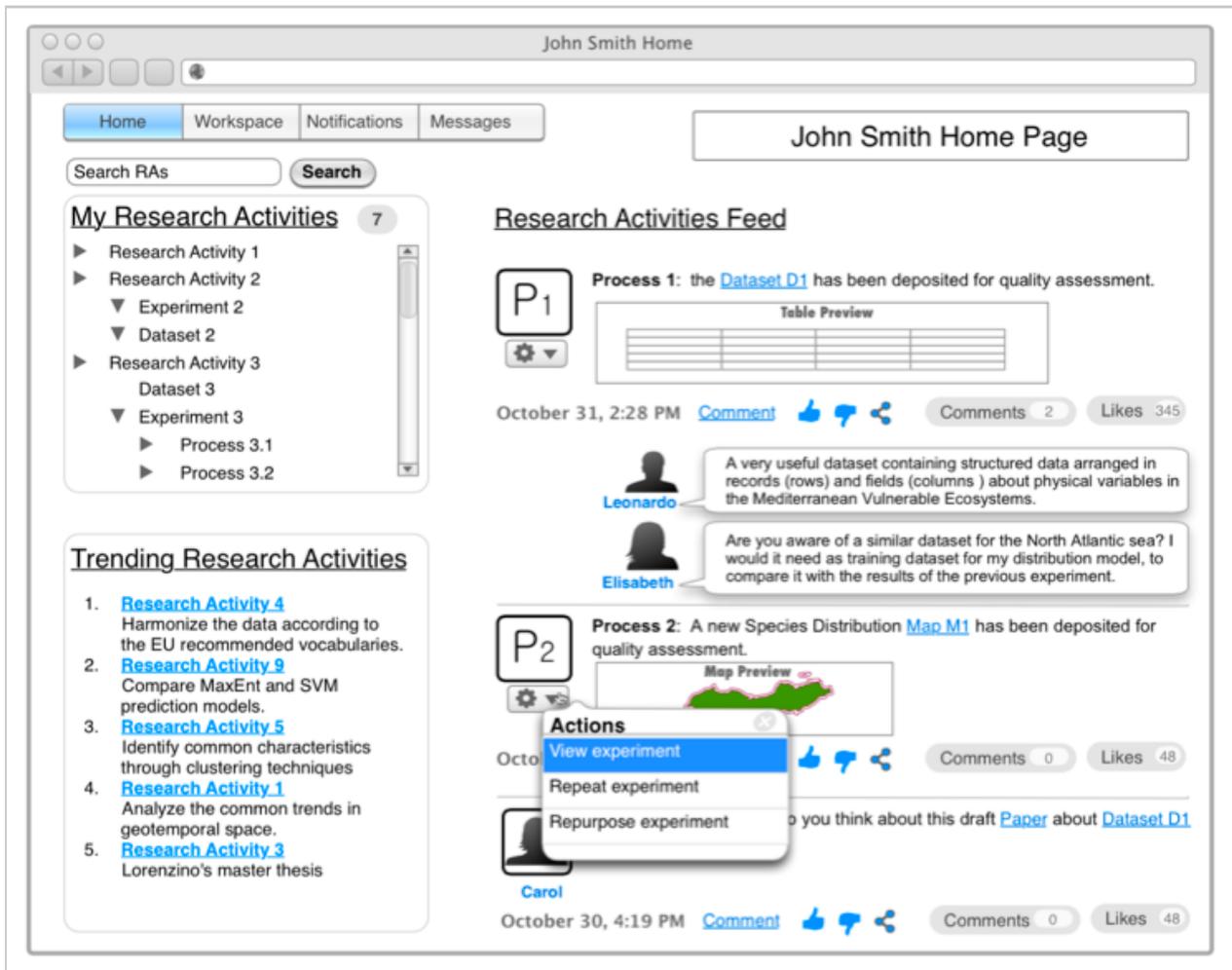


Figure 4: An Example of a SciRepo User Web Page

#### 4.2.3 Dissemination

When deposited, products are assigned a unique web resolvable identifier and may be associated with different kinds of metadata descriptions. Such descriptions may be gathered by underlying ICT services (e.g. provenance, authorship) or be specified by scientists, and discerns products by their typology (to be decided by RI scientists). Their nature depends on the RI at hand and typically enables discovery and reuse of products (e.g. interpretation, citation, access rights) at different level of scope (e.g. experiment, research activity, RI, outside the RI). Similarly, relationships between products can be collected by ICT services while performing experiments and creating products (e.g. versionOf, relatedWith, likedByScientist, discussedByScientist) or be specified by scientists via SciRepo user interface. As specified above, access to products should be ruled by proper right management tools, which may authorize scientists to access products based on their role in the RI (e.g. groups of scientists), the research activity at hand, the typology of products, and the quality of products (e.g. products of low quality are not "published" to given groups of scientists). Finally, once scientists discover the products they are interested in, the SciRepo enables the set of actions they are authorized to fire, based on product typology and user rights. The complexity of such actions depends on the SciRepo implementation and its embedding within ICT services. For example scientists may be authorized to visualize or download a product (e.g. a publication PDF), re-execute a product (e.g. an experiment), link a product to another, interact with the history of versions of a product, open a discussion about a product, or review a product.

An example of the web page of a Research Activity that is automatically generated by the SciRepo is in Figure 5. This page contains links and actions for all the research products associated with it as well as to provide its users with the discussions related to the activity, the people actually contributing to it, and impact indicators resulting from the "use" of the activity as a whole.

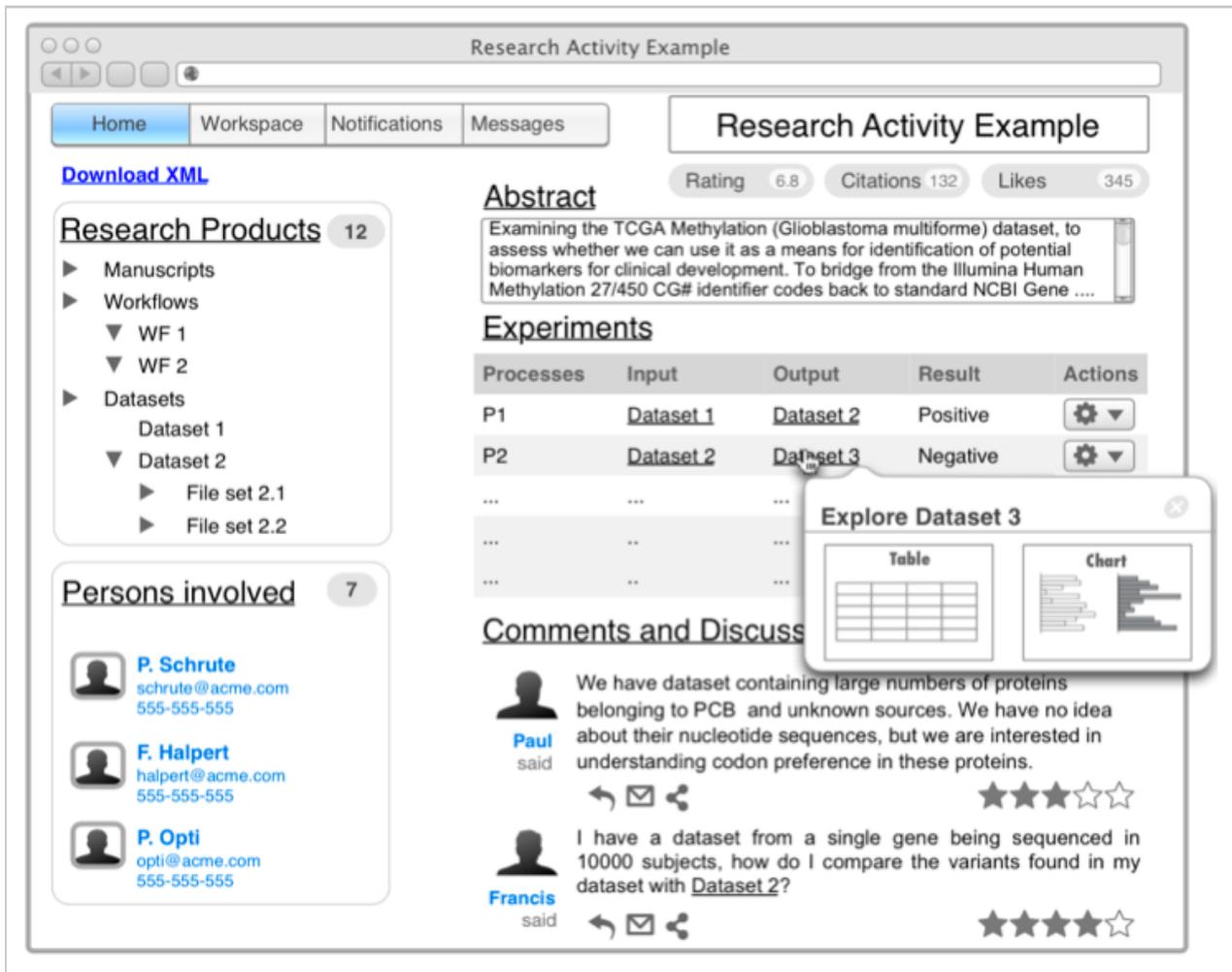


Figure 5: An Example of a SciRepos Research Activity Web Page

#### 4.2.4 Publishing in SciRepos: The Benefits

Table 2 describes the benefits resulting from SciRepos. In particular, the table describes how the SciRepo deposition, quality assessment and dissemination phases overcome the drawbacks of current publishing practices introduced in Section 3.

Publishing Phase	Benefits resulting from SciRepos
Deposition	<p><b>In context:</b> products are fully fledged, i.e. they are linked to the entire setting leading to them. Products that are published in marketplace sources, such as datasets, software, or scientific publications, can be manually re-connected to their research activities hence be discovered in-context together with links to related products.</p> <p><b>Products remain "alive":</b> products may change after they have been published, scientists are actually using a reference to have access to them. Moreover, they are expected to be dynamically versioned so that it is always possible to have access to the instance of a resource at a certain point in time.</p> <p><b>No extra cost:</b> products are stored in the underlying ICT services, where they are created and managed, hence costs and risks of moving products outside RI boundaries are dropped.</p> <p><b>Alternative products:</b> support to alternative products, which can find in SciRepos a place where they can be manually deposited, evaluated (e.g. social "like" tools, discussion threads), discovered, and linked to a research activity or other products.</p>

Quality Assessment	<p><b>Continuous and in context:</b> published products can be continuously assessed from a qualitative perspective, e.g. it is always possible to annotate any research product with a comment (also a process) aiming at demonstrating either the outstanding nature or the mediocrity of the product. The scientific context where the product is created and possibly reused is the best qualified to assess the quality of the product since it represents the primary target domain to be served.</p> <p><b>Self-assessment:</b> the RI community has the ability and interest to define "certificates of quality" which are crucial to enable scientific reward mechanisms pertaining non-traditional products. For example, a web blog kept by a scientist and considered a reference for other RI scientists may be published in the SciRepo, be peer-reviewed and certificated as high-quality research, and such awards be spent to enrich the scientist's CV.</p>
Dissemination	<p><b>Unified:</b> Scientists can be offered marketplace facilities to discover and access products that subsume those typically offered by publication and dataset repositories; they can discover objects by typology, cross-typology, navigate their relationships, configure their access rights, and profit from advanced online re-use functionalities.</p> <p><b>Automatic and complete:</b> product authors are less burdened by tedious activities of metadata information and relationship curation.</p>

Table 2. Benefits resulting from Publishing practices in SciRepos

---

## 5 Conclusions and Future Works

The initiatives aiming at enlarging and strengthening scientific communication thus to meet the expectations of modern science are hindered by two major factors: cultural barriers (e.g., lack of reward, additional effort) and methodological barriers (e.g., many repositories to deal with).

In this work we introduce the notion of Science 2.0 Repository that aims at overcoming the methodological barriers by providing scientists with an integrated and innovative environment that supports "within" and "during" scholarly communication workflows. This repository is conceived to nicely integrate and complement the offering of RIs towards holistic scholarly communication practices. The notion is for the moment intuitive and in the process of being concretized into a reference architecture, but its implementation in the context of the D4Science infrastructure proves its benefits [1][2]. This repository is expected to be offered as a platform that every RI can use, configure and deploy to extend the working environment of its community. Future steps in this direction will be to define a SciRepo formal data model and architecture, in order to realize a general purpose platform facilitating the realization of a SciRepo over any ICT-based research infrastructure environment with limited costs and efforts if compared with from-scratch approaches.

---

## Acknowledgements

The work reported has been partially supported by the FP7 European Commission projects iMarine FP7-INFRASTRUCTURES-2011-2, Contract No. 283644) and OpenAIREplus (FP7-INFRA-2011-2, Contract No. 283595).

---

## References

- [1] Assante, M., Candela, L. & Pagano, P. "An Environment Supporting the Production of Live Research Objects," *The Grey Journal*, Vol. 9, 2013.
- [2] Assante, M., Candela, L., Castelli, D., Mangiacrapa, F. & Pagano, P. "A Social Networking Research Environment for Scientific Data Sharing: The D4Science Offering," *The Grey Journal*, Vol. 10, Number 2, 2014.
- [3] Bardi, A. & Manghi, P. "[A rationale for enhanced publications](#)," *LIBER Quarterly*.
- [4] Bartling, S. & Friesike, S. "Towards Another Scientific Revolution". *Opening Science*. Springer International Publishing, 2014. [http://doi.org/10.1007/978-3-319-00026-8\\_1](http://doi.org/10.1007/978-3-319-00026-8_1)
- [5] Bechhofer, S., De Roure, D., Gamble, M., Goble, C. & Buchan, I. (2010). "Research Objects: Towards Exchange and Reuse of Digital Knowledge". *Nature Precedings*. <http://doi.org/10.1038/npre.2010.4626.1>
- [6] Borgman, C. (2011). The Conundrum of Sharing Research Data. *Journal of the Association for Information Science and*

Technology, 63 (6), 1059–1078. <http://doi.org/10.1002/asi.22634>

- [7] Bourne, P. E., Clark, T., Dale, R., de Waard, A., Herman, I., Hovy, E. H., & Shotton, D. (2012). [Improving the future of research communication and e-scholarship \(Force11 White Paper\)](#). Force11.
- [8] Callaghan, S., Donegan, S., Pepler, S., Thorley, M., Cunningham, N., Kirsch, P., Ault, L., Bell, P., Bowie, R., Leadbetter, A., Lowry, R., Moncoiffé, G., Harrison, K., Smith-Haddon, B., Weatherby, A., & Wright, D. (2012). Making data a first class scientific output: Data citation and publication by NERCs environmental data centres. *International Journal of Digital Curation*, 7 (1), 107–113. <http://doi.org/10.2218/ijdc.v7i1.218>
- [9] Candela, L., Castelli, D., Coro, G., Pagano, P. & Sinibaldi, F. "Species distribution modeling in the cloud," *Concurrency and Computation: Practice and Experience*, 2013. <http://doi.org/10.1002/cpe.3030>
- [10] Candela, L., Castelli, D., Manghi, P. & Tani, A. "[Data Journals: A Survey](#)," *Journal of the Association for Information Science and Technology*, 2014.
- [11] Candela, L., Castelli, D. & Pagano, P. "Virtual research environments: an overview and a research agenda," *CODATA Data Science Journal*, vol. 12, pp. GRDI75-GRDI81, 2013. <http://doi.org/10.2481/dsj.GRDI-013>
- [12] Castelli, D., Manghi, P. & Thanos, C. "A vision towards scientific communication infrastructures," *International Journal on Digital Libraries*, 13 (3-4), 155–169, 2013. <http://doi.org/10.1007/s00799-013-0106-7>
- [13] CODATA-ICSTI Task Group on Data Citation Standards and Practices "Out of Cite, Out of Mind: The Current State of Practice, Policy, and Technology for the Citation of Data", *CODATA Data Science Journal*, vol. 12, pp. CIDCR1-CIDCR75, 2013. <http://doi.org/10.2481/dsj.OSOM13-043>
- [14] De Roure, D.; Goble, C. A. & Stevens, R. The design and realisation of the myExperiment Virtual Research Environment for social sharing of workflows. *Future Generation Comp. Syst.*, 2009, 25, 561–567
- [15] Doan, A., Ramakrishnan, R. & Halevy, A. Y. "Crowdsourcing systems on the World-Wide Web," *Commun. ACM*, vol. 54, no. 4, pp. 86–96, Apr. 2011. <http://doi.org/10.1145/1924421.1924442>
- [16] Moed, H. & Halevi, G. "[Research assessment: Review of methodologies and approaches](#)". *Research Trends*, Issue 36, 2014.
- [17] Nosek, B. A., & Bar-Anan, Y. (2012a). Scientific communication is changing and scientists should lead the way. *Psychological Inquiry*, 23 (3), 308–314. <http://doi.org/10.1080/1047840X.2012.717907>
- [18] Nosek, B. A., & Bar-Anan, Y. (2012b). Scientific utopia: I. opening scientific communication. *Psychological Inquiry*, 23 (3), 217–243. <http://doi.org/10.1080/1047840X.2012.692215>
- [19] Parsons, M., & Fox, P. (2013). Is data publication the right metaphor? *Data Science Journal*, 12, WDS31-WDS46.
- [20] Wang, F.-Y., Carley, K., Zeng, D. & Mao, W. "Social computing: From social informatics to social intelligence," *Intelligent Systems, IEEE*, vol. 22, no. 2, pp. 79–83, 2007.

---

## About the Authors



**Massimiliano Assante** is Research Staff at the Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy. His scientific and professional activity involves the Research and Development on (Hybrid) Data Infrastructures and NoSQL Data Stores. Dr. Assante is currently member of the iMarine EU Project. In the past he has been member of and EUBrazilOpenBio, D4Science II, D4Science, DILIGENT and DRIVER European Projects. Within these projects, he progressively covered different positions, ranging from software engineer (gCube System web services and front-end web applications) to researcher (analyst, system designer, system integrator).



**Leonardo Candela** is a researcher at Networked Multimedia Information Systems Laboratory of the Italian National Research Council – Institute of Information Science and Technologies. Dr. Candela graduated in Computer Science in 2001 at University of Pisa and completed a PhD in Information Engineering in 2006 at University of Pisa. He has been engaged in several EU co-funded research projects concerned with digital libraries and data infrastructures. His research interests include Data Infrastructures, Virtual Research Environments, Data Publication, Open Science, Digital Library [Management] Systems and Architectures, Digital Libraries Models, Distributed Information Retrieval,

and Grid and Cloud Computing.

---



**Donatella Castelli** is a Senior Researcher working at Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy. She has been the principal investigator of several European and National funded projects on digital libraries and data e-Infrastructure acquiring considerable experience in these domains. Currently, she is acting as technical director of the EU OpenAIREPlus project. She is also the scientific director of the EU iMarine project and of the corresponding data infrastructure. Her scientific interests are centered on data modeling, data interoperability and data infrastructures. She is a member of the RDA Europe Expert Group that promotes research and cross-infrastructure coordination at global level.

---



**Paolo Manghi** is a researcher at Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy. He received his PhD in Computer Science at the University of Pisa (2001). Today he is member of the InfraScience research group, part of the Multimedia Networked Information System Laboratory (NeMIS). His current research interests include data ICT infrastructures for science and technologies supporting modern scholarly communication. He is technical manager of the OpenAIRE infrastructure ([www.openaire.eu](http://www.openaire.eu)).

---



**Pasquale Pagano** is Senior Researcher at the Networked Multimedia Information Systems Laboratory of the Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, Italy. Dr. Pagano received his M.Sc. in Information Systems Technologies from the Department of Computer Science of the University of Pisa (1998), and the Ph.D. degree in Information Engineering from the Department of Information Engineering: Electronics, Information Theory, Telecommunications of the same university (2006). The aim of his research is the study and experimentation of models, methodologies and techniques for the design and development of distributed virtual research environments (VREs) which require the handling of heterogeneous computational and storage resources, provided by Grid and Cloud based e-Infrastructures, for the management of heterogeneous data sources. Dr. Pagano has a strong background on distributed architectures. He participated to the design of the most relevant distributed systems and e-Infrastructure enabling middleware developed by ISTI–CNR.

---

Copyright © 2015 Massimiliano Assante, Leonardo Candela, Donatella Castelli, Paolo Manghi and Pasquale Pagano

---