diid

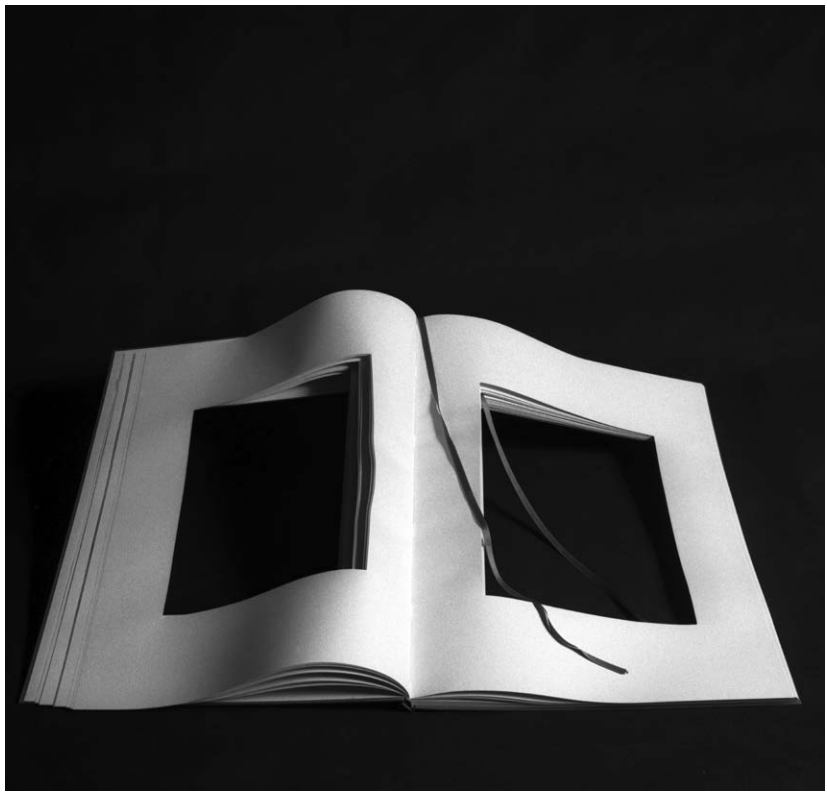
78

disegno industriale
industrial design

theorìa
pòiesis
praxis

diid.it

Kim Frederic Albrecht, Alberto Calleo, Mauro Attilio Ceconello, Maria Claudia Coppola, Derrick De Kerckhove, Carmelo Di Bartolo, Sascha Friesike, Daniel Garber, Miaosen Gong, Meixiang He, Amilton J. V. Arruda, César Kieling, Carla Langella, Silvio Lorusso, Stefano Luca, Eleonora Lupo, Amalia Magril, Paolo Manghi, Kieran McDonald, Lorela Mehmeti, Nicolas Misdariis, Sara Radice, Mágda Rodrigues da Cunha, Sigrid Schmeisser, Ambra Trotto, Louise Valentine, Rosa van der Veen, Paula Visoná



Colophon

diid
disegno industriale
industrial design
No. 78 — 2022

Year
XX

diid is an open access
peer-reviewed scientific
design journal

diid is published
three times a year

Registration at Tribunale
di Roma 86/2002
(March 6, 2002)

www.diid.it

Print subscription
(3 issues)
Euro 60,00
Subscription office
ordini@buponline.it

Publisher
Fondazione
Bologna University Press
Via Saragozza 10
40123 Bologna
Tel. (+39) 051 232 882
Fax (+39) 051 221 019
www.buponline.com
info@buponline.com

ISSN
1594-8528

ISSN Online
2785-2245

DOI
10.30682/diid7722

ISBN
979-12-5477-246-1

ISBN Online
979-12-5477-247-8

Cover image
Vincenzo Agnetti, Libro
dimenticato a memoria,
Vitalità del negativo
nell'Arte Italiana, Palazzo
elle Esposizioni, Roma, 1970.
Fotografie Ugo Mulas

Contents

Editorial

Open Debate

Designrama

Forum and Reviews

Policies and Infrastructures for Innovation and Impact in Open Science

An interview with Paolo Manghi

Paolo Manghi
Università degli Studi di Pisa
paolo.manghi@isti.cnr.it
ORCID 0000-0001-7291-3210

Eleonora Lupo
Politecnico di Milano
eleonora.lupo@polimi.it
ORCID 0000-0003-3677-0245

Abstract
In this interview Paolo Manghi discusses how policymaking, e-infrastructures and funding mechanisms are trying to bring into everyday routine the difficult shift of mindset towards Open Science: a scenario of publishing service systems that supports communities with methodologies to produce concrete, testable and reusable scientific results.

Keywords
Open science
Publishing service systems
Policies
Infrastructures
Scientific results

Paolo Manghi is a Full-time Researcher in computer science at the Istituto di Scienza e Tecnologie dell'Informazione (ISTI) of the Consiglio Nazionale delle Ricerche (CNR), in Pisa, Italy. His research areas of interest today are data e-infrastructures for science and scholarly communication infrastructures, with a focus on technologies supporting open science publishing within and across different disciplines, i.e. computational reproducibility and transparent evaluation of science. He is the Chief Technology Officer of the OpenAIRE infrastructure.

EL Paolo, could you tell us something about your career? You have specific expertise in scholarly communication and especially in e-infrastructures and technology for supporting open science publishing and transparent evaluation. I'm curious about how this influenced your way of publishing scientifically and conversely, how your role as a technical officer of the EU is in some way addressing new forms of publication beyond the more traditional forms.

PM I'm a computer scientist. I'm 52 now, so I started quite a long time ago in 1997. The publishing ecosystem at that time was "simple": scientists published and were evaluated based on articles and the whole mechanism of publishing was managed by companies, the publishers. Over the years, this became a quite perverse mechanisms, where you, as a scientist, are publishing your results into journals, your colleagues are evaluating your work for free, and publishers sell subscriptions to your own universites or libraries. If a journal is good enough, researchers tend to publish there, and a journal is good enough when articles are highly cited by other authors. We have built a whole system for evaluating science around this, which has affected the way scientists are performing and publishing science, and implicitly nourishes this mechanism.

When I started my career, I was publishing within this mechanism precisely, and was feeding the system. Many scientists are still victims of this system today, including computer scientists and engineers in general.

I later moved to CNR and spent a couple of years in Glasgow for my PhD. When I came back, I started working at CNR in a group called InfraScience, led by Dr. Donatella Castelli in the digital libraries context, focusing on scientific publication management. Over the years, digital libraries extended their scope to include research data and software, to address community needs for reproducibility of science and policy-makers needs of reusability of science.

We started doing research in this extremely challenging domain that was, at that time, miles away from reality. What we are seeing today is the gap between reality and vision getting smaller and narrower. Things are changing in many domains, going towards the idea of open science, both in the way of publishing, and in the way of doing science. Open science is basically the two sides of the same coin: on the one hand, you want to share what you're doing (so while you're doing science, you're sharing your process, making it open). And on the other hand, you also publish openly, which

diid No. 78 — 2022
Doi: 10.30682/diid7822c

means open access, but in a transparent and resusable manner. FAIRness is the buzzword. In such a way, others can understand the products of science that you're producing and can understand where they come from, and what the theory behind them is. They can easily compare it to other theories and combine and reassemble your results to perform further science. And it's clear that scientists want this to happen, but this will never happen unless the policymakers and research organization are there to support it.
This is what is happening in Europe. Roughly since 2009, the [European] Commission actually broke the rules, imposing degrees of open access to the publication of articles funded with public money. It has not been easy, as this choice affected the business models of the "giant" publishers, like Elsevier and Springer, but things have been changing so far; in Horizon Europe there are open data mandates and much of the funding, about 350 million, has been spent around the European Open Science Cloud. The European Open Science Cloud is there to enable this change of mindset and understanding, towards an open and reproducible way of publishing and sharing resources.
So even in my domain, I changed my way of publishing, and it became harder. You have to spend much more time because when you explain the results, you have to support them with other kinds of output, such as data software; you have to explain how this can be used and in which context, and possibly using online tools that are available to others. Enabling the repeatability of an experiment through publication complicates the way of writing, including more roles in the process, such as the data producers, the software producers, etc. While science is changing, some of the practices are not supporting authors: for example, how do we distinguish between those who wrote the paper and those who produced the software in the dataset. There are a lot of studies, and a lot of forums where you can discuss structure methodology and share it with your communities. The Resource Data Alliance is one of those, another is Force11[1].

1

EL You were talking about disciplinary paths dependency, with domains that are more open and experimental than others. The design sector, unfortunately, still tends to stick with traditional writings: we can attach supplementary data, but the common scientific product remains the traditional article. There are more experimental examples coming from domains like computer science, digital humanities, art science, and this context really broadens and evolves the scenario. In any case, we may observe a contradiction: on the one hand, some scholars push for this very open and collaborative scenario. On the other, there are a lot of boundaries, especially in Italy's regulatory system for evaluating the quality of the research: early career researchers must make credit in order to achieve the best possible evaluation and the highest level of citation to support their career development. Basically, in Italy we still have to count the articles we write.

Leaving policymaking in the background, how can we best learn from each other, and transfer some transdisciplinary encounters to make this transition smoother?

PM Changing publishing practices is incredibly hard. First, there is no one-fits-all solution. Communities and nations behave differently. Even in the design domain, there will be countries where given concerns are irrelevant and others where the same concerns are in fact crucial.

What I suggest to my PhD students is to publish in open access. There are some good open access journals free of charge. Then we have the hybrid journals (those that are not in principle open, but if you pay the APC article processing charges your paper is open). The European commission is providing funds, for example, through which you can pay the APCs. However, open science and open access practices state that this hybrid approach is not viable, for reasons of double dipping (libraries are paying for the subscriptions and, at the same time, the authors are paying for the individual licences for the articles).

So, how to change the culture of publishing? The one that is most effective is funding mandates. If your funder imposes an open access mandate, you have no way out. So, the community needs to adapt. This is what the Commission has been doing in many fields. Those who get funds must publish in open access, both publications and research data. If somebody doesn't want to provide data, he can opt-out, but must explain why and have a good reason. Such an imposition is useful to the scientists, because they call for the community to do something. And the community reacts as a whole, mut find a solution.

There is also the push from the scientists themselves, because they understand that science becomes much easier when you find easy access to somebody else's content. And science as a whole can be faster, more useful, less redundant, and less boring in many ways because it evolves quicker.

The other thing that the [European] Commission has done, which has been of real impact, is to build Research Infrastructures. We have more than a hundred Research Infrastructures now in Europe, which operate and support specific domains. Their aim is to build digital tools or devices that a community can use to perform science, followed up by methodologies, common metadata, and common understanding practices, agreed within the community. On top of Research Infrastructures, the European commission has built Clusters, grouping research infrastructures in the same domain, trying to understand how they can perform cross-discipline science and share experiences, practices, content, data, services, implementing economy of scale methodologies. Above the Clusters and the Research Infrastructures, the European Open Science Cloud is trying to cross-link all sciences, trying to maximize synergies across all sciences. All these levels leverage a culture of open science and pave the foundation for a common vision to its implementation.

diid No. 78 — 2022
Doi: 10.30682/diid7822c

EL I would like to go back to another thing you said: that it's really more difficult to publish research results to make them really open and reausable. It's a kind of mental shift for which authors should be trained and maybe this is also a reason that there is some resistance in terms of the cognitive change that we have to make; because it's non-linear knowledge. In a recent article you authored, the concept of publishing "packages of artefacts" emerges, a kind of package of contents that are more "actionable", so these features are really different from traditional scientific writings. My question is ontological: how are these new forms of publishing changing the way we reason? We are moving from linear writings to new forms of scientific outcomes, clustering and packaging them to make them as reusable as possible. In this new "publishing for repurposing" slogan, we are replacing the "publish or perish" motto by shifting the scholar's mindset about the way we think and consolidate our science towards a kind of granularity of knowledge. What do you think?

PM Today publishing is for the sake of publishing. Following the *publish or perish* motto researchers tend to maximise the number of papers they can write about an idea. Authors are not really concerned with proving that what they're saying is true, in most cases no reviewer will be able to check or repeat their experiment, science claim are based purely on trust. We are literally witnessing this in the last decades, with an increase in the number of publications that is unbearable. But when you introduce reproducibility as a requirement, publishing become more complex; the obligation of publishing concrete, testable scientific results and elements that are expected to be testable, increases the chance of soft-pruning the part of science that is irrelevant and that is, in many cases, just there for the sake of being cited.

In my domain, when as part of an investigation we produce and publish research software or data, we must describe in the paper the whys and the hows of the software and data, a non-trivial narrative and metadata-ization process.

An ideal scenario is one where services used by scientists support and facilitate the publishing process: for example, a service that processes data under specific conditions or criteria to produce an output approved by the authors, and publishes the data on behalf of the authors. In addition, the service takes care of the metadata around the publication, of the storing conditions according to the practices of the community, of packaging it and producing an object that is published in a repository, for example Zenodo.org. This is the ideal scenario, and many are moving in this direction. Of course, it's complicated, so technology must support these workflows and provide the communities with methodologies, policies, and processes. Librarians who were key in the past for producing readable, interpretable, and valuable information about scientific literature, cannot also ensure that the data from different communities is FAIR or correct. And here, as you mentioned before, we need the communities to be in charge. The evaluation process must come from the community with specific solutions that depend on the

community context, since each science has its own specific packages whose elements have their own specific names, their own ontologies.
Open science implementation is complex for this reason. Even in discovery: discovery across publications on literature is an obvious thing: I can use standard technologies to do that, basically semantics AI and language processing. For the data, discovery is a different thing, as is assessment, because each community has its own ontology, a local understanding of different kinds of objects and the way they're connected, so they need different kinds of services.

EL I continue from this point that you just mentioned: the discovery, connected to the knowledge representation. How can we support the discovery of this kind of fragmented but connected pieces? OpenAIRE is ahead in this, with OpenAIRE Graph. Can you tell us about this "flexible context-sensitive, fine grain and machine-actionable representation of scholarly knowledge"?

PM The general concept of a knowledge graph is a set of nodes connected by edges. This is a known concept and has been used in scholarly communication for ages, such as Scopus for example. The Web of Science already produces these graphs of different types, for example, linking a publication to a journal, or a conference, to the author and the organisations behind them and the different affiliations.
What we have done in OpenAIRE: first of all, we wanted to offer a knowledge graph openly and transparently to researchers. Today, known operational graphs, the ones that are used to produce statistics and for assessment all belong to companies, such as Elsevier, and the scientific communitu has spent millions to use them. There have been many initiatives seeking to provide similar graphs in a way that was transparent and reusable by others, for example Dimensions, Microsoft Academic Graph (now deprecated), OpenAlex. OpenAIRE supports, collaborates with, and shares with such initiatives skills and data. Unlike these we add to our graph, beyond scientific articles, the research data, the research software, the communities and the funders, together with the projects. We built a graph that goes well beyond the publications, because we wanted to identify the connections between the publications and the data, the software, and eventually the funders and how these affect the scholarly communication, in order to enable the data analysis that may identify better lines of investment or return of investment for the funders. We wanted to build this in a way that was open, transparent and reusable by others, to which others could contribute as well. Again, the Commission here has been extremely brave and visionary, challenging the big enterprises in this domain, mandating open science and open access practices, and making the graph one of the most important components of the European Open Science Cloud. Others in the world are using the OpenAIRE graph to complete their metadata collections. For example, Scopus is doing it by taking the data from our graph because they are missing the links to the data and to the software.

diid No. 78 — 2022
Doi: 10.30682/diid7822c

OpenAIRE is a community driven initiative. People and organization can join the discussions or become members, to steer the discussion and address scientific needs. We have subgroups and we are delivering, for example, interoperability solutions across graphs, and ways to share this information in such a way that we can build a network of graphs and learn from each other. I think this is key for open science as a whole: to think of a system that is sustainable, low cost and offers public services. It's important to have it in your hands, basically, and to control it, to understand what's going on and possibly fix things. And this is not possible with Scholar or Scopus or Web of Science, with all the tools that we are using today. With one tenth of the budget that we spend every year in Europe to have access to such products, we could build it for free, and we are indeed doing it.

EL The next question refers to this idea of the discoverability and reusability of knowledge that you were mentioning. One of the concerns is about the credits, recognition and the attribution of authorship. I think this is something the community is discussing. I have found a very interesting model promoted by a scholar from the UK, which proposes moving from authorship to contributorship. That is a very well-suited idea for a different kind of contribution. Our articles still rely on a traditional model in which all the authors are listed without exactly specifying any contribution, except, for instance, in credits at the end, but not in a very specific and detailed way. How can open research infrastructures pave the way for this new method of attribution?

PM I think it's reasonable. It's something that is in the ideas, but it has not been methodologically structured enough to be effectively used. You may have a very nice contribution framework, but then who is selecting the effective roles, and who is ensuring they are correct? It's not obvious, especially when there is no control over the metadata, or nobody is in charge of the quality of the metadata that is being produced for an article, research data, or research software. And this is often the case in the open science settings, where publishing behaviour is not ruled and validated via agreed-on practices. Still a long way to go.

More generally, assessment in general is very hard to capture under one number: economists claim that if you have a rewarding system or an evaluation system that is based on a couple of numbers, then these numbers will inevitably change the system and it will adapt to the numbers. It's something that we've done for too long, without even thinking about the side effects. In the open science scenario, more numbers can be extracted and used differently. From the number of papers, or the number of citations, the average per paper, or the total number of citations you have had, looking at how your citations are increasing over the years, may indicate something more interesting, such as your ability to identify new fields of science that then become trendy. Or the ability to capture funding, to work with a group, grow

a group, and at the same time, work with other groups, so the interconnections, all these are numbers that identify the quality of researcher that you are.
It's the impact of your science in science, which is not determined by citations alone, but indirectly by how many years after are indirectly referring to your root article. Therefore, reducing all that to one number, the impact factor with three numbers after the comma, that makes a distinction and the quarters, is simply crazy.
For instance, evaluating an organisation is also important. How much is the organisation investing in open science? To evaluate the impact of one's organisation, you should look at how it is delivering infrastructure and enabling science.
I am in a project today, EOSC-Future, that is trying to capture data on the usage of services to measure how scientists are using them. And again, you will not need one number but probably many to have an idea and monitor what's happening, because open science is multi-faceted. For research data, it is different. In a rencent work (Mannocci, Irrera & Manghi, 2022), we were trying to identify the trends of an author, analysing how the distribution of a researcher's contribution across his career has developed over time. For example, whenever you have a publication linked to a data set, you may notice that some authors appear only in the data set. Some authors appear only in the paper, and some appear in both. So, you can say something about the nature of the contribution. You will see that some always appear in the data and the software and rarely in the paper, for example. And in some scientific disciplines, this is clear cut. Where can you identify the kind of contribution, at least in percentage or rates? What's the "style" of the scientist?

EL You have anticipated a lot about this question because my next point was about the impact. if I have understood correctly, such a system can track this kind of contribution for a long time, and in some way build a profile of a scholar. This is not about quantity alone anymore, but it is a different way of understanding quantity. It is a kind of merit or as some authors would call it, reputation. It is suggesting how much for example an institution is working on providing open services, or funding. This is a question of reputation. You can use numbers, but in the end numbers build reputation. I'm trying to scale it down anyway by returning to journals, and for now this is difficult to apply to a traditional article, or a traditional research output. How can we try to establish this alternative way of building and measuring our impact? Because in the end, we still rely on traditional metrics, and above us, there is a world we are trying to climb to get there.

PM Most of what I said is, again, at the level of research investigation; practice is different. The changes that are now taking place in Italy and in many other countries are suggesting that the change will take place soon. It has been in the process, in the mandates at the National and European levels for a long time. Somehow, we will need to adapt to a common understanding of what science assessment is. To the point

diid No. 78 — 2022
Doi: 10.30682/diid7822c

that one of the messages that we deliver to the PhD students in Pisa, is to start changing the ways of publishing because the assessment will evolve. Soon, you will be evaluated on your ability of publishing in an open science manner. Open Science roadmaps are planned and implemented differently, at community and country level. Policymakers play a key role and have the carrot and the stick to steer and evolve the system. For instance, the Nordic countries have a five-country system, based on a sort of green list of journals (similar to what we have in Italy): journals classified A, journals B, journals C, etc. The list may include brand-new journals, based on the fact that the programme committee is high quality, hence guarantee that its published science is good. New open-access journals, typically funded by the universities, can therefore become part of the list bypassing the Impact Factor indicator eligibility, which would otherwise take years. This strategy opens a crack in the traditional Impact Factor and closed publishing system.

EL I hope your optimistic vision becomes real. We need to promote this vision in some way. So, the last point I want to focus on, is this idea that you have just mentioned, of evolving knowledge. I think the infrastructures of open science and the open access scenario are working on this idea of evolving knowledge, because you can build upon other knowledge, you can build upon other people's data. I have also mentioned the idea of tracking content. What would be the best way to ensure the traceability of datasets or any other content when the content is reused and passed on from one to another? Do you think blockchain technology can be relevant to this infrastructure?

PM That is a technical question: how can we implement policies? The technical solution that you're going to adopt will bias whatever you're going to produce. The choice really depends on how cumbersome you want it to be, how safe you want it to be, how complex to adopt it will be, how expensive and sustainable it will be, and so on. Blockchains bring interesting properties but they may not be essential to or easily find a way within open science infrastructures, which are built bottom up as "systems of systems" and not as monolithic or highly integrated cross-disciplinary environments.

EL Moving on from the technical side tothe intellectual: when it comes to a traditional written paper, maybe you won't make it actionable in the same way. In other words, if people could build and write on it, who is the author in terms of intellectual property at the end? I'm not talking about open access, but of authorship in the traditional sense.

PM I recently wrote a text (Tennant et al., 2020) using a collaborative platform. A colleague started writing the paper and called out for co-authors. I joined in the writing. At the end there were about 20 authors, actually it evolved into a book; the idea was that we would all be authors at the same level, except for one who was leading the exercise, trying to make sense out of the input that we were producing. Again, the question here finds answers with policies; authorship is based on policies. If we imagine a a system where a text that reaches a level of maturity for publication can be reopened by other authors to start building a new one, then policies become central. For example, policies may establish that all authors of the original work are de facto authors of the new version; or that they are out if less than 20% of the original work is used. The idea is interesting, although for a proper application it should cope with the IPR and copyright aspects imposed by the publishers.

References

Mannocci, A., Irrera, O., & Manghi, P. (2022). Will open science change authorship for good? Towards a quantitative analysis. *Proceedings of the 18th Italian Research Conference on Digital Libraries (IRCDL 2022)*. https://arxiv.org/abs/2207.03121

Tennant, J. P., Agrawal, R., Baždarić, K., Brassard, D., Crick, T., Dunleavy, D. J., Evans, T. R., Gardner, N., Gonzalez-Marquez, M., Graziotin, D., Greshake Tzovaras, B., Gunnarson, D., Havemann, J., Hosseini, M., Katz, D. S., Knöchelmann, M., Lahti, L., Madan, C. R., Manghi, P., … Yarkoni, T. (2020, March 6). A tale of two 'opens': intersections between Free and Open Source Software and Open Scholarship. *SocArXiv*. https://doi.org/10.31235/osf.io/2kxq8