Fabio Ciotti

A landscape of Digital Humanities in Italy: backgrounds, institutions and infrastructures

The main objective of this paper is to give an overall sketch of the situation in the Italian digital scholarship in the humanities and cultural heritage and to stress the relevance of research infrastructural initiatives to facilitate its future development.

Scientific collaboration and the development of shared practices, resources and tools, in fact, is undoubtedly one of the major driver of consolidation and progress for a scientific domain. It is natural, therefore, that the establishment of *Digital Research Infrastructure for the Arts and Humanities* (DARIAH) as a formal European research infrastructure, and the Italian participation in it, have to be seen as substantial step forward.

The tradition of the digital in Italian Humanities

The "tradition of the digital" in Italian humanities and cultural heritage dates back to the half of the last century and is indeed very rich of projects both in the academic and the institutional context. In general we can trace two branches in that tradition, that had systematic intersections only in recent times:

1) the tradition of research and experimentations of computational and digital methods and tools in linguistic, literary, historical and artistic studies, which have taken place in research and higher education institutions;

2) the tradition of digital activities led by the institutions for the management and conservation of cultural heritage such as libraries, archives and museums and the related coordinating institutions at local and central level.

The former, usually, have been academic research oriented, but as is typical of this world, especially in Italy, have suffered a greater dispersion and lack of coordination (with the result of reducing their reach and visibility and of replicating methodological mistakes and bad implementation choices). The latter on the other hand, have been more often service oriented, aimed at a more socially extended and not necessarily professional community of users. From the very beginning, they assigned great importance to the cooperation and to the need to produce shared resources and services. To cite only the most relevant: SBN project[1] – the union catalogue of Italian Libraries - and the more recent projects of portals for the digitization and dissemination of bibliographic and cultural heritage as Internet Culturale[2] and Culturaitalia[3].

In this paper we will concentrate exclusively on the first of these two traditions, that in Italy have been identified under the label of "Humanistic Informatics" (in contrast to the "Humanities Computing" label adopted in the Anglosaxon context), the sort of scholarship that in the last decades has been widely known as Digital Humanities[4].

---

[1] http://www.sbn.it/.

[2] http://www.internetculturale.it/.

[3] http://www.culturaitalia.it/.

[4] These different nomenclature are not at all irrelevant as they convey different notions and perception of the field, and bring different metaphorical burden, but we cannot delve into this issue here. A good historical account of the evolution

Of course, it is not possible to draw an exhaustive diachronic and synchronic picture of this crowded and lively field here. Therefore my point of view will be partial and necessarily bound to my personal experience.

Let's start with a first observation: Italian tradition in Humanistic Informatics has been developing for long time. The reference to Father Busa, universally acknowledged as founder of this domain, and to his lemmatization and indicization of Tommaso D'Aquino's works dating back to the 1940s of last century, is rather obvious[5]. But I want to stress that Busa's work was not at all isolated in Italy. It suffices to remember that in 1962 the prestigious annual *Almanacco Bompiani* was devoted to "Le Applicazioni dei Calcolatori Elettronici alle Scienze Morali e alla Letteratura", showing how since those early times Italy produced cutting-edge research at the crossroad between computing and humanities[6].

Within this tradition the *ante litteram* attention to what we call today infrastructure has always played a central role and several projects (some of those with a long and prestigious history) perfectly fit this broad definition. I would like just to mention two examples, which are part of my personal experience.

Since the Eighties of the last century University of Rome Sapienza has been the most important Italian center of studies in the field of Digital Humanities thanks to the work of Professor Tito Orlandi, one of pioneers and international scholars of this discipline. He is not only the author of fundamental scientific publications[7], which had great impact both at the theoretical and methodological level on the Digital Humanities, but he is also the founder, in 1991, of CISADU (Centro Interdipartimentale di Servizi per l'Automazione nelle Discipline Umanistiche) which was probably the first proper DH center in Italy[8].

In the group of scholars that grew up with Orlandi, there was Giuseppe Gigliozzi[9] – who left us prematurely in 2001 – who in turn founded in the late 90s the **CRILET** (Centro Ricerche Informatica e Letteratura). There he gathered a group of young researchers which started to digitalize and encode literary texts using formalisms and standards like the Text Encoding Initiative (TEI) SGML/XML based markup language[10], and to use computational methods for text analysis. This group played a leading role in

of this research field, though centered on the Anglo-Saxon world, is EDWARD VANHOUTTE, *The Gates of Hell. History and Definition of Digital / Humanities / Computing*, in MELISSA TERRAS, JULIANNE NYHAN & EDWARD VANHOUTTE (a cura di), *Defining Digital Humanities. A Reader*, Farnham, Ashgate Publishing 2013, pp. 119-156. More centered on the Italian landscape is TITO ORLANDI, *Per una storia dell'informatica umanistica*, in FABIO CIOTTI E GIANFRANCO CRUPI (a cura di), *Dall'Informatica umanistica alle culture digitali*, Atti del convegno di studi (Roma, 27-28 ottobre 2011) in memoria di Giuseppe Gigliozzi, Roma, Sapienza University Press - Digilab 2012, p. 49-102.

[5] ROBERTO BUSA, *The Annals of Humanities Computing: The Index Thomisticus. Computers and the Humanities*, "Computers and the Humanities", 14, 1980, pp. 83–90.

[6] SERGIO MORANDO, *Almanacco letterario Bompiani:1962*, Milano, V. Bompiani & C. 1962.

[7] TITO ORLANDI, *Informatica Umanistica*, Roma, La Nuova Italia Scientifica 1990 e *Informatica testuale. Teoria e prassi*, Roma, Laterza 2010.

[8] See TITO ORLANDI, *Per una storia dell'informatica umanistica*, cit.

[9] Cfr. GIUSEPPE GIGLIOZZI, *Saggi di informatica umanistica*, Milano, UNICOPLI 2008.

[10] http://www.tei.c.org.

the dissemination of TEI in Italy (the translation of the TEI Lite *Guidelines* dates back to that time[11]) and in the end most digitization programs in Italy adopted XML and TEI.

In the same years professor Pasquale Stoppelli created LIZ (Letteratura Italiana Zanichelli[12]) a CDROM based corpus of Italian literature texts supported by a textual analyzer. The corpus was later extended by CIBIT (Centro Interuniversitario Biblioteca italiana Telematica), headed by Amedeo Quondam, and made available through the site of BibIT Biblioteca Italiana[13] after the conversion of the texts into TEI/XML.

This long tradition of initiatives in the DH field at the Sapienza University culminate in 2012 with the establishment of the DigiLab, an interdepartmental center of research and services[14]. DigiLab takes advantage of technical personnel and various research facilities, as well as of the scientific work of scholars and researchers from many different disciplines, which are developing projects regarding knowledge and communication of cultural heritage contents in a digital environment, digital archives and digital libraries, human computer interaction and computer graphics and digital cultures analysis. Among various projects I must cite the participation in the international network DiXiT (Digital Scholarly Editions Initial Training Network) and in many in FP6 and FP7 projects like DELOS, ECLAP-The E-Library of Performing Arts, Linked Heritage, and EAGLE -Electronic Archive of Greek and Latin Epigraph. Perhaps the most relevant achievement of DigiLab in our context has been the development of *Sapienza Digital Library*[15]. SDL is a typical digital infrastructure tool and service, a Digital Library Management System based on OAIS reference model offered as a service to the academic community to create manage and disseminate digital collections.

My second example is the Institute for the European Intellectual Lexicon and History of Ideas (ILIESI)[16], an institute of the Italian National Research Council (CNR). The CNR has a solid tradition of research in the field of Digital Humanities. It dates to the sixties and to the pioneering work of Antonio Zampolli, at first animator of the CNUCE's Linguistics Section, then founder and Director of the Institute of Computational Linguistics of CNR (Pisa). In the following years, the ILC quickly became a reference point of excellence for the automatic processing of language at an international level.

In close connection with the experiences of the ILC, the Center of Study for the European Intellectual Lexicon (that in 2001 became ILIESI), since the seventies has devoted many activities and research projects to the development of linguistic and textual data with specific reference to the area of the history of ideas of the early modern age. The preparation of indexes, concordances and lexicons of philosophical and scientific texts was for several years one of the main aspects that distinguished the Institute. In the eighties, the creation of a large database was added, containing dozens of classics of

---

[11] FABIO CIOTTI (a cura di), *Il manuale TEI Lite: introduzione alla codifica elettronica dei testi*, Milano, Edizioni Sylvestre Bonnard 2005.

[12] PASQUALE STOPPELLI E EUGENIO PICCHI, *Liz 4.0: Letteratura Italiana Zanichelli : Cd-Rom Dei Testi Della Letteratura Italiana*, Bologna, Zanichelli, 2007.

[13] http://www.bibliotecaitaliana.it/.

[14] http://digilab.uniroma1.it/.

[15] http://sapienzadigitallibrary.uniroma1.it/.

[16] http://www.iliesi.cnr.it/.

modern philosophical-scientific in various languages. It remained for years an *unicum* in this field in an international scholarly level of research.

Over the years, many other textual archives have been added to this initial data base. More recently, the ILIESI is strongly engaged in Research projects funded by the EC. In particular, thanks to the Discovery and Agora Projects, the Institute has set up a portal called Daphnet which includes a platform of texts of Greek Classical philosophy (Ancient Philosophy), a platform about philosophy of early modern European (Modern Philosophy), as well as a platform dedicated to secondary sources (The Daphnet Digital Library)[17]. In all of those platforms, texts are encoded according to the TEI-XML standard. Moreover, the Institute has developed systems for semantic annotations link to an explicitly elaborated ontological domain.

I could keep going, mentioning many other institutions and centers, some of them not any longer active, some still working and carrying excellent scholarly research and fundamental digital resources creation. The activity of such centers and projects has produced a significant amount of primary and secondary digital resources, of computational tools and methods, of scientific and theoretical elaboration. Even if we cannot access an amount of funding as in US or even other EU countries, Italian Digital Humanities researchers have been able to create:

1) a number of scholarly projects and service centers with high level human and technological resources;

2) some huge and mid text and document archives in XML/TEI with search facilities like Biblioteca Italiana, Musisque Deoque18, digilibLT19 or Bibliotheca Iuris Antiqui20;

3) state of the art services and platforms for digital asset management and digital libraries and archives systems;

4) a lot of language resources and corpora, NLP and text mining tools.

The lively and qualified level of the Italian digital scholarship in the humanities is epitomized by the recent establishment of a national scholarly association. It is difficult to define the Digital Humanities as a discipline according to the traditional canons (based on the sharing of methods and objects of scholarly research) since it has become over the years so broad and multifaceted. This dispersion and eclectic nature makes pressing the need for places and institutions of mutual exchange and recognition. Hence, after several unsuccessful attempts made since the 90s of last century, in 2011 a large group of Italian scholars and researchers decided to found the Associazione Italiana per l'Informatica Umanistica e la Cultura Digitale (AIUCD)[21]. The aim of the organisation is "to promote the methodological and theoretical reflection, the scientific collaboration and the

---

[17] http://www.daphnet.org/.
[18] http://www.mqdq.it/.
[19] http://www.digiliblt.unipmn.it/. Cfr. MAURIZIO LANA, *Metodologie e problematiche per una biblioteca digitale. Il caso di digilibLT*, "Digitalia", VII, 1, 2012. http://digitalia.sbn.it/article/view/516/355.
[20] http://bia.lex.unict.it/.
[21] http://www.aiucd.it/.

development of practical tools and shared resources in the field of humanities computing and the use of digital applications in all areas of the humanities".

Despite some peculiar concerns and difficulties of the research in DH (and not only) in our country, AIUCD has achieved excellent results (for instance, the organization of the TEI Conference 2013 in Rome). It has been the first of the diverse national DH associations established recently in several European countries and has become an associated organization to the European Association of Digital Humanities (EADH). Among its many activities, AIUCD regularly organizes an annual conference whose the program during the last years has offered a high level of papers and project presentations. This is just a demonstration of the vitality of the Italian landscape in the field.

DARIAH: an European effort to build a DH infrastructure

The quick picture I have depicted so fare is not of course without spots and gaps. Some of those shortcomings are extrinsic, determined by the general situation of the Italian university system and by the social and economical situation of the country. Nonetheless, there are also some intrinsic weaknesses.

The very first thing missing, in fact, is the coordination between experiences and projects. As I have already pointed out, for anthropological, cultural, social, institutional, and economic reasons, those research initiatives have seldom had the ability to coordinate, to share tools approaches and knowledge, except in rare cases and on the basis of personal liaisons. The second main issue is proper funding, of course. In the economic phase Europe and our country in particular is experiencing, the investment in research funding is undergoing drastic downsizing. The only way to continue developing high-quality research is network building, pooling resources and technologies, on the one hand and supporting researchers and research institutions to access the few available funding sources. I refer of course to the European level program Horizon 2020 but also to the local and regional level funding initiatives. Creating a virtual reference and counseling service in this area could really be a strategic infrastructure. One last issue is the lack of a proper institutional recognition of digital scholarship in the Humanities, especially in order to facilitate the enrollment and career progress of DH scholars in the academic context.

I believe that the main reason to foster the development of a large scale research infrastructure for digital humanities resides in facing and possibly solving these kind of basic level issues and limitations, in order to create the condition for "high quality" research to stem out of the collective intelligence of the community. The establishment and development of research environments, in fact, do not rely only upon theoretical and methodological groundings, but also need a solid social economical and institutional context[22]. This is why we must welcome the recent formal establishment of a European *Digital Research Infrastructure for the Arts and Humanities* (DARIAH-EU[23]), in the context of the European Strategy Forum on Research Infrastructures (ESFRI) effort.

---

[22] Cfr. TOBIAS BLANKE, *Digital Asset Ecosystems: Rethinking crowds and cloud*, Elsevier 2014.
[23] https://dariah.eu/.

The debate about what research infrastructures are has been quite lively in the last years. ESFRI itself suggest a wide definition[24]:

Research Infrastructures are facilities, resources or services of a unique nature that have been identified by European research communities to conduct top-level activities in all fields.

This definition of Research Infrastructures, including the associated human resources, covers major equipment or sets of instruments, in addition to knowledge-containing resources such as collections, archives and data banks. Research Infrastructures may be "single-sited", "distributed", or "virtual" (the service being provided electronically). They often require structured information systems related to data management, enabling information and communication. These include technology-based infrastructures such as Grid, computing, software and middleware.

In this context UE has started a Roadmap for the development of continent wide research infrastructures consortia[25]. The DARIAH infrastructure is legally established as such a European Research Infrastructure Consortium (ERIC). The participation to the ERIC is based on a national level structure: each country is represented by a national research institution and is expected to pay an annual fee and to assure a certain amount of in-kind contribution. The consortium is hosted by France and has Austria, Belgium, Croatia, Cyprus, Denmark, Germany, Greece, Ireland, Italy, Luxembourg, Malta, Netherlands, Serbia and Slovenia as partner members (and others are expected to join in the next years). At the national level each member can make the organizational choices he prefers (some national members have for example established local consortium) and select the scientific partners that contribute to the infrastructure. DARIAH-IT, the Italian branch of the EU infrastructure has started its work in 2013, and a lot of details are to be decided.

DARIAH's mission is to enhance and support digitally-enabled research and teaching across the humanities and arts, to develop, maintain and operate an infrastructure in support of ICT-based research practices, to provide access to and disseminate research that stems from collaborations between its members and to ensure that best practices, methodological and technical standards are followed. To achieve the results of implementing a truly coordinated bottom-up participation in DARIAH and ensure that services can be deployed in an efficient way DARIAH operates through a double level of community scientific cooperation:

1) Virtual Competency Centres (VCC) level

2) Working Groups (WG) level

The VCC level defines the strategic areas and topics that need to be worked on according to the DARIAH high-level principles. Each VCC is a kind of hook, which the WGs can latch on to. The VCCs provide sustainability, incorporate the outcomes of the WGs, and are responsible for the communication. The VCCs also help to coordinate and assess working groups, the in-kind contributions of the DARIAH-EU members and participation to EU research grants (e.g. Horizon2020). DARIAH has defined actually four VCC, each centered in a strategic domain of intervention:

1) E-infrastructure : establish a shared technology platform for A+H research

---

[24] ESFRI, ESFRI Roadmap, 2010. https://ec.europa.eu/research/infrastructures/pdf/esfristrategy_report_and_roadmap.pdf.
[25] http://ec.europa.eu/research/infrastructures/.

2) Research and Education: expose and share researcher's knowledge, methodologies and expertise

3) Content Management: expose and share scholarly content

4) Advocacy and outreach: interface with key influencers and constituency in/for Arts and Humanities

Infrastructure efforts are very important, but their success and their impact on the quality of the research, especially in the Humanities domain, is not assured a priori (and the USA Project Bamboo case is epitomizing in this respect[26]). To avoid the risk of failure or of limited acceptance by the wider community of scholars, it is of the maximum importance adopting a bottom up and grassroots approach. The recent kick-off of the Working Group level in the DARIAH enables an organizational structure, which is not just flexible and dynamic, but also driven by feedback and helps DARIAH to be sustainable. The Work Groups (WGs) are self-organized structures with a name, definition, objectives/goals, in-kind budget, operating mode and communication means. They are the embodiment for all joint activities in the DARIAH community[27]:

DARIAH working groups (WG) are based on a three-step approach:

1. Conceptualisation: Members of DARIAH develop a concept for the WG, aggregate participants and means as well as define an action plan. The primary outcome is a proposal to the DARIAH community that meets the WG requirements.

2. Implementation: At this stage, efforts are integrated and the connection to further communities outside DARIAH is established.

3. Service: Finally, the service is enabled with guaranteed hosting and sustainable funding through DARIAH. The service is published to the community.

Working groups are the real "scientific engine" of DARIAH, where services and resources for the community (be they scholarly, technical, editorial and/or organizational) must be planned, developed and maintained.

A mission for DARIAH (Italy): some hints

As we have seen, bottom-up coordination, resource and tools sharing and network building are of course general objective of a research infrastructure like DARIAH. But it remains to be decided which are the main assets of this infrastructure. There are dozens of polls and surveys that tries to identify emerging needs of the research communities and their expectations towards a shared research infrastructure. Often in these investigations, some complex questions come up. Yet not always complex answers, ambitious platforms which pre-determine methodologies, methods and tools of research, get the desired results. I think in general it is much better to focus on low level enabling technologies and basic resources that scholars can adapt to their needs and preferences.

[26] QUINN DOMBROWSKI, *What Ever Happened to Project Bamboo?*, "Literary and Linguistic Computing", 29, 2014, pp. 326-339; JORIS VAN ZUNDERT, *If you build it, will we come? Large scale digital infrastructures as a dead end for digital humanities*, "Historical Social Research/Historische Sozialforschung", 2012, pp. 165-186.

[27] Tobias Blanke, Conny Kristel, Laurent Romary, *Crowds for Clouds: Recent Trends in Humanities Research Infrastructures*, in Agiati Benardou; Erik Champion; Costis Dallas; Lorna Hughes. *Cultural Heritage Digital Tools and Infrastructures*, 2016, 978-1-4724-4712-8. https://hal.inria.fr/hal-01248562.

Starting from the bottom level and from the feasible achievements let me indicate at least two priorities that rise from the specific Italian situation, although I'm sure they can be common to the entire DARIAH community.

The first is the need of a common infrastructure for scientific publishing, based on the principles of open content and open access, without excluding anyway profit business models for some particular products or services. Actually, some platforms that comply with these requirements are already in activity, the most famous of which is Open Edition (which has recently opened a branch in Italy)[28]. I think it would be appropriate to foster the encounter between this emerging reality and the dispersed landscape of University Press and academic publishers.

A parallel, though not less important aspect of scientific communication is the multilingualism. We are all aware that English has become the lingua franca of scientific communication, and it would be pointless and anachronistic to resist this trend. And yet, especially in the humanities, subtracting a scholar the full control of the rhetorical, linguistic and argumentative tools of its native language is likely to result in a threat to the quality of its research papers. I hope that DARIAH can support the experimentation of scientific communication frameworks and workflows that seriously take into account multilingualism.

The second priority asset I envision is the need of a common platform for the assessment of digital products. Gathering and long term preservation are also central issues strictly connected with it. But I really think that the actual acceptance of digital methods and products within the larger community of scholars in the humanities (with all the consequences it can have) must cope with the problem of evaluation.

When I talk about digital research products obviously I'm not referring to the classic article or monograph (although in digital form). Anyone who has really been involved in a research project in Digital Humanities knows that most of the research work consists in the modeling and creation of the data sets, computational analysis methods, tools and frameworks. If the goal of a research project is the creation of an archive of documents, a digital scholarly edition or a visualization tools to display historical data, those are the real products of research, not the articles describing them (often tedious and report style). The assessment of Digital Humanities research must apply to these objects.

The creation of a platform of this kind requires an ontology of the roles and functions performed by the researchers involved in a project; a set of minimum quality requirements (standards compliance, interoperability, scientific control, coverage of the domain, etc.) to assign a digital resource the status of research product; a metric based on these requirements to enable its evaluation. On this basis it would be possible to build an assessment platform based on peer review (be it classic and/or open) that would have a dual function:

1) provide a repertoire of research-quality digital products certified by the community of scholars, where every researcher can have access and find resources and tools for new projects;

---

[28] https://www.openedition.org/.

2) provide a set of data for the assessment of individual researchers involved in Digital Humanities and help their recruitment and career progression in academic and research institutions.

Some building bricks and approximations to this system are already there: I'm referring to projects such as DH Commons[29] and NINES[30]. It is time to start the creation of this platform and I think that DARIAH is the appropriate environment where this can happen.

---

[29] http://dhcommons.org/.
[30] http://www.nines.org/.