Research data and humanities: a European context

Maura Funari

1 Defining research data

The subject of research data[1] has handled in scientific research's contexts, both in theorizations on research's different topics, and in management policies of research's results and in international organizations policy's documents. To bypass the necessity of defining research data, in the strict sense of the word, is a common tendency. But virtually to meet with conceptual ambiguities is frequent: depending on the field of application, peculiar aspects are identified so they, even though in some cases almost imperceptibly, vary. So far one definition with a univocal and a universal kind, does not exist.

UNESCO (Swan) describes research data as a type of "research output" together with journals, peer-reviewed conference proceed-

---

[1]The italian translation has not found a clear definition: the italian version of European Commission's documents refers both to "dati della ricerca" (*Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca*) and to "dati di ricerca" (*Raccomandazione della Commissione, del 17 luglio 2012, sull'accesso all'informazione scientifica e sulla sua conservazione*). In this treatise the Anglo-Saxon term has been kept because of the common acceptance.

ings, and books. This categorization seems to individualize in this typology of data, an instrument to externalize research's products, also acknowledging to research data a general increase in the attention that open access policies attach them.

A set of definitions comes from Australian National Data Service (ANDS) which, specifying in the introductory the fact that any definition is likely to depend on the context in which the question is asked, collects definitions enunciated in data management policies of some Australian universities: University of Melbourne, Monash University, Griffith University (Australian National Data Service). The first two definitions quote common elements in research data's characterization, identifying a variety of shapes and contents and so leaving out a previous determination based on these criteria. The founding role of research data, that is the datum used as primary source, or on which the research's theory is based, is an additional aspect just in the first of the three quoted definitions: "Research Data means data [...] on which an argument, theory, test or hypothesis, or another research output is based" (University Of Melbourne). The Griffith University's definition verbatim refers, in the first part, to the OECD Organisation for Economic Cooperation and Development (OECD) one (13) in which research data are defined as "factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research and that are commonly accepted, in the scientific community, as necessary to validate research findings". Therefore this definition determines that the condition of commonly acknowledged, as necessary material in order to validate research findings by scientific community, is determining for research data. A similar formulation comes from definition of scientific data reported in the Memorandum for the Heads of Executive Departments and Agencies[2] of the Executive Office of the Presi-

---

[2]http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_

dent, Office of Science and Technology policy, of the White House. These data are defined as: "the digital recorded factual material, commonly accepted in the scientific community, as necessary to validate research findings including data sets used to support scholarly publications". Also in this case the condition of necessary material to validate research findings, not defined as "factual record" but as "factual material (digital recorded)", is connotative.

The ANDS is, moreover, the RDA's founder[3] (Research Data Alliance), with the European Commission's support (through the iCordi project, treated in the paragraph assigned to infrastructures) and the United States one too (through the National Science Foundation). The international organization aims to accelerate and to improve innovation and data-driven research, encouraging the actions connected to research data (such as exchange, sharing, uses and re-uses, standards and visibility) and achieving the development and the adoption of infrastructures, policies, practices, standards e services.

Getting back to the research data's definitions, the Communication of the European Commission (*Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca* 3), emphasizing the increasing attention to the research data's access improvement, characterizes them as "experimental results, observations and computer-generated information which form the basis for the quantitative analysis underpinning many scientific publications". Holding into consideration the heterogeneity coming out of the above-said definitions, however, it is possible to establish that research data can be meant to data, in different forms and contents,

---

public_access_memo_2013.pdf.

[3] http://rd-alliance.org/.

which constitute the basis for a scientific research, as a primary resource and foundation of the research's findings. Since the intrinsic value and the necessity of collection, preservation, sharing, are qualities varying according to different factors, from the nature of the research to the disciplinary field, the individuation of pre-established criteria is essential to set what kind of data, produced into a specific area, could be included in research data's category. This task should mainly belong to research's findings treatment policies developed by centers, agencies and institutions involved.

From both a qualitative and a quantitative point of view, during the research's phases, a lot of data could be produced but certainly their potential value constitutes the essential assumption for the interest toward this typology, in the fields of information science and research's findings treatment policies. This potential value can vary depending on datum's form, nature, origin (National Science Foundation. National Science Board12-13). This relativity is sharpened by differences that emerge both in natural sciences/human sciences macro-areas, and in the single disciplines, inside the two areas.

Furthermore it must be specify that research's datum can has got the double role of product (as result or resultant of a specific research) and of source (as a datum already produced by someone else and re-used as the basis of a new research): a circumstance that has been appointed referring to the antithetical ideas of output and input.[4] This doubleness brings out the pattern of a circulation and knowledge's sharing system which, whereas an open level of sharing is looming, founded on the action of re-using (Murray-Rust)(Murray-Rust). Murray himself, quoted in the Italian studies (De Robbio and Giacomazzi), notices one difference in practices of data's publication

---

[4]"Data are outputs of research, inputs to scholarly publications, and inputs to subsequent research and learning" (Borgman, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*115).

and use between the typologies of "Large Science" and "Small Science".[5] Furthermore, from a terminological point of view,[6] the open level of sharing is the discriminating condition between research data and open research data: the last one refers just to open data, while the first does not leave out them.

2   Research data in humanities

Natural sciences differ from humanities, besides for field of study and methodologies, also for a greater quantity of data produced (as a consequence of the technical level and the objectivity founding natural sciences), for their typology (which affects also their level of elaboration) and also for the degree of necessity and practice in their sharing and re-using. These circumstances have made scientific-natural research's data protagonists, not just in theoretical studies, but also in the executions of systems for their collection, management and sharing, and in the policies concerning to research. The disadvantage characterizing humanities in this field, such as in the open access one (Suber), derives both from scientific-humanistic research's features and so from their results and sources, and from correlated economic and cultural explanations, especially concerning spread and timing (in the scientific-natural research the need of sharing, both in results 's storage and in the access to them, is characterized by a quickness that is reduced in humanities.)[7]

---

[5]The distinction made by Murray between "Large science" and "Small science" is based on the research's unit dimension that, in the first case, is vast and narrow (individual or laboratory) in the second one.

[6]The requirements satisfying the attribute of Open, referring to data are, various. See the definition of "open" proposed by the Open Knowledge Foundation: http://opendefinition.org/okd/.

[7]Peter Suber ("Promoting open access in the humanities") in his analysis on the open access slow moving in the humanities compared to natural sciences, identifies

If according to natural sciences the individualization of what can correspond to a research data is rather intuitive, for when it comes humanities, the question is more complex.[8] The National Science Foundation (12-13) identifies different data categories ("observational, computational, or experimental") as a result of their giving origin to an observation, a computation, or an experiment. If in natural sciences these actions belong to the standard researching method during the interpretation and the study of phenomena, understood as observational events, in humanities, which have as subject-matter no physical entities, the correspondence is not so obvious.

The overlapping between data concept and "primary source" (Burrows) constitutes the key point of the matter.[9] The connection between the two concepts is clear in the OECD's quoted definition (*Principles and Guidelines for Access to Research Data from Public Funding*), in which research data are the "factual records" used as "primary sources". This role comes true in the research's process: the datum considered as factual record becomes the primary source of the research. And yet, considering the data products and their

---

nine differences concerning the research in the two different areas. Although the analysis's subject-matter are the research's findings in the form of journal articles, many observed circumstances are valid also for when it comes research data, outlining a contest in which open access in humanities appears as "less urgent and harder to subsidize, than in the sciences".

[8]In Borgman ("The digital future is now: A call to action for the humanities") the question "What constitute data in the humanities?" is investigated but it does not found a clear answer. The author concludes, referring to the quoted question and to other four questions concerning digital humanities: "Answering these questions will enable the digital humanities community to be more articulate about its scope and its goals, and better positioned to identify their requirements for infrastructure". (The question is broached also in Borgman, *Scholarship in the Digital Age: Information, Infrastructure, and the Internet* 215-217).

[9]Burrows ("Sharing humanities data for e-research: conceptual and technical issues") claims that to not discern "primary source" from "data" in humanities "would be analogous to describing the stars and galaxies as an astronomer's 'data'".

utilization in the scientific humanistic research as factual records is hard, even if this circumstance can not be excluded at all, referring for example to the findings of a text mining in textual analysis, or to archeographic data in the archaeological field. However, it is beyond doubt that, from a quantitative and a qualitative point of view, this data typology does not represents the pre-eminent one, also taking into consideration the nature of investigation's subject, often constituted by abstracts entities in the form of representation. Instead of simple data, in most cases, they are data sources (data and information's sources), in different forms, varying from texts to objects.

Briefly said, to identify what to mean for research data in humanities, taking into consideration in a unavoidable way, the discretionary power conferred from the "primary" founding label which characterizes research data's definitions and role, it is necessary to make a distinction between:

- Data intended as immediately knowable elements, as a result of observations of phenomena, reality, experiments, computations. (Examples: findings of a text mining in textual analysis, archeographic data, questionnaires, audio gatherings in field researches, etc.).

- Data sources as investigated elements providing the datum, or on which the datum itself is based. (Examples: painting, literary work, musical manuscript, archaeological find, etc.).

This last typology, being on the basis of the datum, represents its source, where the research leads to findings belonging to the first category (which constitute the proper research data). For what concerns the production and the use in the scientific-humanistic research, the first typology, has said before, is not quantitative and qualitative substantial as in scientific-natural one, in which the study

of physical phenomena generates a computational or not production
of raw data, directly recorded. Otherwise the using of resources,
which are constituted by both data and sources (ex. documents,
imagines, texts, etc.), is relevant so much so that, as explained in the
dedicated paragraph, the European infrastructures expressly collect
digital resources. In the humanities area can not be said that the
factual records are equivalent to the primary sources and, therefore,
that these constitute the research data, but it is possible to detect that
they have got the same role, as basis of a research. Furthermore, in
humanities, technology offers in many cases the direct possibility
of recording, reproduction, graphic representation, accessing and
linking[10] (circumstance less likely in the natural sciences area). So,
digitalization, if integrated with accessible infrastructures, makes
possible the collection, the sharing and the use of resources 's collec-
tions, even if leaving out their materiality.

3   Research data in European policies

The attention to research data, among European Community, has
been realized through a series of European Commission's Commu-
nications (*Comunicazione della Commissione al Parlamento Europeo, al
Consiglio e al Comitato economico e sociale europeo. Sull'informazione
scientifica nell'era digitale: accesso, diffusione e conservazione comuni-
cazione; Comunicazione della Commissione al Parlamento Europeo, al
Consiglio, al Comitato economico e sociale europeo e al Comitato delle
Regioni. Le Infrastrutture TIC per la e-scienza; Comunicazione della Com-
missione al Parlamento Europeo, al Consiglio, al Comitato economico e
sociale europeo e al Comitato delle regioni. Un'agenda digitale europea;
Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al*

---

[10]An example of digitalized data sources collection, linked to data, is Europeana
www.europeana.eu.

Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca), till it has reached the form of a Recommendation (*Raccomandazione della Commissione, del 17 luglio 2012, sull'accesso all'informazione scientifica e sulla sua conservazione*) to member States. This interest is aim to research data and, more in general, to scientific information, produced in all research's fields, among which humanities are quoted.[11] Research data, as a parallel form but also as associated[12] to the proper publications, constitute the scientific information, of which wide and quick diffusion has a central role in terms of research's innovation, progress, efficiency and excellence. The desirability of this diffusion, however valid on principle, is necessary especially in the case of findings, both journal articles and data, resulting from publicly funded researches.[13]

Already in 2009, in the field of ICT Infrastructures (Information and Communication Technologies) for e-science, the assumption of the emergence of new research methods that exploit advanced computational resources and data collections, as well as the awareness

---

[11]"The emergence of 'big data science' has a global dimension, as it reflects the increasing value of raw observational and experimental data in virtually all fields of science (humanities, biodiversity, high-energy physics, astronomy, etc.)" (Commissione Europea, *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Le Infrastrutture TIC per la e-scienza*9).

[12]The Communication deals with a "'continuum' of the scientific information space from raw data to publications across different communities and countries". Internet and the new information and communication instruments allow, indeed, to use research data coming from experiments and observations, associating them to other information's sources, to the aim of taking out meanings (*Comunicazione della Commissione al Parlamento Europeo, al Consiglio e al Comitato economico e sociale europeo. Sull'informazione scientifica nell'era digitale: accesso, diffusione e conservazione comunicazione*3).

[13]Ibidem However some delay for the first use by researchers or for commercial purposes can be considered as justifiable (3).

of "the increasing value of raw observational and experimental data in virtually all fields of science" (*Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Le Infrastrutture TIC per la e-scienza*9), had identified as objective, the adoption by Europe of a "coherent and managed eco-system of repositories of scientific information" (11). Member States and scientific communities had been asked to step up investments in scientific data infrastructures, as also provided for by the Commission into the Seventh Framework Programme, with the aim of "support accessibility and preservation policies". In this ambit of interest toward research data takes place the Communication (*Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca*) followed by the Recommendation (*Raccomandazione della Commissione, del 17 luglio 2012, sull'accesso all'informazione scientifica e sulla sua conservazione*), in which the attention to them becomes more detailed. The traditional debate focused just on publications and, at the same time, the increasing importance of the improvement of access to research data, defined as already quoted, are indeed acknowledged. The inefficiency of public research investments reveals itself where findings in the form of data, for the verification and the possible use, are made not available to a wide public of users. In the Communication are identified different obstacles to the development of this new clutch of knowledge sharing and are stated the initiatives already carried out by the Commission (OpenAIRE) and those planned (financial supports to data infrastructures and to research on digital preservation). The identified obstacles related to the development of research data access and to their use and re-use are:

- The lack of organization and clarity about responsibilities.

- The lack of financing models to ensure long-term access.

- Persistence of interoperability problem among countries and disciplines.

- Researchers and innovative enterprises reluctance due to different reasons (data perceived as their property, time to invest in the practicalities of depositing, absence of reward and recognition mechanisms, such as citation mechanisms and measurement of the data citation impact (*Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca*7).

Although these problems are real, other and more specific (here omitted) factors affect negatively on the process. It's about barriers[14] closely related to the access and to the use of research data, in the field of dedicated infrastructures. These barriers are legal (copyright, restrictive licenses, limitative editorial policies), financial (subscription to datasets 's access, payment for the use of materials) and technical (restricted visibility, restricted length, impossibility of accessing/using/re-using). Their existence and the weight held in the difficulty of realization and in the efficiency of systems for free and open research's findings sharing in the form of data, are due to the commercial value which often many data have, to the acknowledgment of creative work qualification, and so susceptible to the copyright, to the absence of sustainability plannings aimed to maximize the investment and to guarantee long term effects.

The importance assigned to research data comes out also into the new Framework Programme for Research and Innovation (2012-

---

[14]On the definition of the different kind of barriers see Murray-Rust ("Open Data in Science") and Suber ("Promoting open access in the humanities").

2020) called Horizon 2020.[15] The Proposal for decision (*Proposta di decisione del Consiglio che stabilisce il programma specifico recante attuazione del programma quadro di ricerca e innovazione (2014-2020) - Orizzonte 2020*) of European Commission Council dealing with the realization of the quoted programme, identifies in research's findings dissemination and communication on a continent wide, a "key added value" in order to enhance their impact (22). In line with this awareness are collocated the actions aimed to support the creation, the development and the operation of the TIC infrastructures with the goal to " achieve, by 2020, a single and open European space for online research where researchers enjoy leading-edge, ubiquitous and reliable services for networking and computing, and seamless and open access to e-Science environments and global data resources" (36). Research in social sciences and humanities is declaredly fully integrated in this specific objective concerned research infrastructures and in those generals of the programme (21). So, in Horizon 2020, are provided actions aimed to realize open access to research data. Indeed, the Communication provides for the launching of a "pilot scheme on open access to and re-use of research data generated by projects in selected areas of Horizon 2020." (*Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca*9). The Recommendation exposes the urgency of adopting political actions on access to data and, therefore, recommends to member States to define clear policies providing for objectives and indicators to measure progress, implementation plans and financial plannings, to guarantee that "research data that result from publicly funded research become publicly accessible, usable and reusable through digital e-infrastructures" (*Raccomandazione della*

---

[15]http://ec.europa.eu/research/horizon2020/index_en.cfm..

*Commissione, del 17 luglio 2012, sull'accesso all'informazione scientifica e sulla sua conservazione*3).

The European Commission approach in defining research data policies is characterized, as already seen, by an attention to scientific information on the whole. Research data, in case open and accessible, constitute one of the instruments (with the other results) to strengthen the research system, in an interoperable and cooperative European context, both inside and in the extra European ambit. This setting out, which practically leaves out investigations and analysis on research data as an independent element, derives both from the nature of Commission's documents (they are not guides or technical texts) and from the propulsive role that the Commission has got, as a coordination center between the members States, for when it comes the development and the improvement of the scientific research system on the whole.

A different setting out is presented by the OECD's contribution (*Principles and Guidelines for Access to Research Data from Public Funding*), in a guide form, and addressed to the single States to encourage an international and efficient research data sharing and use, overcoming variety of laws, policies and national practices. The text offers an analysis of the different aspects concerning the definition of policies by research's institutions and founding agencies. Furthermore, as seen, it offers a clear and complete definition of this data category, placing it in a separated dimension, however not subordinate or necessarily parallel to the other research's results. After an introduction specifying that the principle of opening and ideas, information and knowledge free sharing is on the basis of Organization member States scientific public systems, is recognized that new technologies have created a "new fields of application for not only the results of research, but the sources of research: the base material of research data" (9) and that an "effective" access to these data, giving enunci-

ated benefits and advantages, should improve profits coming from
public investments. The subject of this access are data resulting
from public founding researches. The known advantages resulting
from these data open sharing have a character both universal and
individual for the singles members States. The principles on which
data research have to be founded (openness, flexibility, transparency,
legal conformity, protection of intellectual property, formal responsi-
bility, professionalism, interoperability, quality, security, efficiency,
accountability, sustainability) are, as said, enunciated and analyzed
in their different aspects.

Also the UNESCO (Swan) stated the research data value, defining
them, as seen, as a "research output" category which is increasingly
getting a central role in the open access policies. The text, which
has the aim to promote the open access in the members States, facili-
tating the understanding of the connected aspects, is concentrated
on the relevant issues. Even though the research data belonging
to research output group (so to the research information on the
whole) and the inclusion of this research's results typology in the
open access concept are recognized, open access main and original
"target" is determined in the "journal literature" (10). It's just this
category the main subject of the work. In the document, however,
is underlined the centrality of data-intensive sciences in the open
sharing process of finding research and the differences of rules and
data management between disciplines. For when it comes strate-
gies to promote open access is admitted an increasing difficult in
separate open access to the "literature" from open data (referred to
research data), and the consequent need to include in future strate-
gies supporting open access, those concerning data. Although this
accepted tie[16] the need to develop diversified policies for open data,
that considering problems connected to privacy and circumstances

---

[16]Swan (27) talks about an "ecosystem of 'open' issues".

preventing the diffusion because of other reasons, is highlighted.

A policy on the research results access is modelled on the subject it regulates: in defining contents, storage methodologies, standards and everything else, subject's characteristics of the collection, preservation, sharing and re-using process can not be neglected. The product, in the datum form, takes a different shape from a journal article. Some aspects related to the quoted category need, in the policies 's definition, a particular attention: for example, data can be contained confidential information (such as data collected in social sciences field) or can have a commercial value. Therefore, in these cases it is necessary to identify provided actions aim to establish conditions and specific limits. The characteristics of research data differ not only from those of scientific articles (for this reason a separated policy for the two research output categories would avoid omissions and gaps), but also from the single research ambits. Indeed it has been gathered that "a generic approach to data curation will not be sufficient to cope with the different data-related needs and expectations of researchers working in different disciplines other than at a superficial level" (Key Perspectives2).

Although the variety of contexts in which research data theme is broached, in European policies, involving a difference of purposes and approaches, the research data value (related both to scientific-humanistic research and to scientific-natural one), guided towards the improvement of the research's process and of everything else related to, can be said unanimously recognized. This value materializes in the open access to data through TIC infrastructures, with the aim to make them widely accessible, usable and re-usable.

The presuppositions to the necessity of making data widely available are valid on principle, since results from public funding projects and researches should return to the funding community. In this way, not only citizens 's right to potentially take advantage of final findings is

observed, but also the investment itself is optimized: the researches 's repetition is limited (if data derived from previous researches are not available, it is necessary to repeat the research), times for research are increased thanks to the speed in information's finding (also at an interdisciplinary level), costs for the access are removed (for example, the access to some datasets accessible for payment), the research system on the whole becomes more transparent (data quality can be verified and the datasets 's use is measurable). All that positively influences, in a more or less direct way, the growth, the progress and the development at an economic and social level (For a detailed list of the advantages in terms of public investment optimization in the scientific research, see Organisation for Economic Cooperation and Development10).

4   The Italian position

To make research data value concrete in a practical form, need systems aimed to collection, management, preservation, sharing, exploitation of research's findings (and everything connected to these operations) and States 's action to support and to promote these initiatives. Since they are real "systems", by definition corresponding to sets of instruments, mechanisms and elements, the coordinative action is essential, both to an internal level and to an external one. As it can be intuitively understood, also the financial aspect is included in the actions for these systems, together with the spread of a cultural education about research's datum. Currently, Italy faces two circumstances which constitute the starting point of the implementation of policies, focused on the quoted actions: the first one is a political constraint deriving from European Commission Recommendations and Communications, for which, as a part of an aggregative organism, should conform to common policies; the

second circumstance is a practical aspect constituted by advantages in terms of cultural, economical and social progress, coming from development, improvement, growth, exploitation, efficiency of the scientific research process. In this condition are fitted in the participations to European intra and extra initiatives in favour of the development of research data collection, management, sharing systems. (ex. CLARIN, DARIAH, ARIADNE, MAPPA, RDA and iCordi, seen in the next paragraph). Compared to other European countries, which are in the lead for when it comes initiatives to support research data projects, such as the United Kingdom one, it is clear that Italy is in a backwardness position. First of all, national policies aimed to discipline the matter, in terms of definition of the different aspects connected to the spread of research's results (from the contents identification to the responsibilities) lack; there is not a national reference point as a coordination centre; financial and organizational models, assuring the infrastructures 's long term sustainability, have not been outlined.[17] With regard to existent infrastructures, besides the limit constituted by the lack of a multiannual planning aimed to guarantee the long term sustainability, there are a limited capability to exploit the social-economical benefits connected to the realization and to the operation of a research's infrastructure and an unfit presence of European infrastructures into the national territory. The quoted participation to European infrastructures projects is real, but properly defined as "assicurata da gruppi di ricerca di massa subcritica" (Ministero dell'Istruzione, dell'Università e della Ricerca66): so it is not sufficient to set off a widespread phenomenon. The Horizon

---

[17]Conservations' structures are often created to specific projects and the fundings are limited to a certain period (For when it comes the obstacles to research data access and to their use and re-use and to the long term preservation, see Commissione Europea, *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca*7).

2020 Italia document, drawn up by the MIUR (8), on communication of the research and its results deals with italian "limitata capacità" of "trasferimento, diffusione e valorizzazione" depending on both italian researchers, and still insufficient support services of universities and national Public Research Agencies. Furthermore, the same document underlines the existence of "criticità importanti per il posizionamento del sistema europeo della ricerca e dell'innovazione e, al suo interno, in maniera più accentuata, di quello italiano" (5). In accordance with the comunitary programme, Italy intends to realize "un sistema della ricerca sano ed efficiente, non frammentato e privo di duplicazioni, forte, coeso e strategicamente orientato" (26). To pursue this scope, the circulation and the sharing of scientific research's results have a relevant position. Indeed, the document states the importance that the access (open, free and with an interoperable format) to data and information resulting from public funded activities has, with regard to the connection between science and society and to the optimization of the financial investment itself. The Researchitaly portal[18] is identified as the aggregating gateway of the initiatives on open access of the national research system and as the platform to list local repositories of universities and research's centers and to contain a national repository. For the Italian system of research's infrastructures in which, as quoted, are highlighted a series of lacks, is provided a national plan (PNIR), aimed to improving them. Moreover, is planned the strengthening of the existent infrastructures and the realization of new ones, according to the European Strategy Forum on Research Infrastructures (ESFRI) and the legal instrument European Research Infrastructure Consortium (ERIC)[19] is identified as valid to take part in pan European infrastructures projects.

---

[18]https://www.researchitaly.it/.
[19]http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric1.

The perspectives enunciated in Horizon 2020 Italia programme, reveals a new scenario for the country: if it becomes reality, we will be a European country that eventually will takes part actively in knowledge's free movement.[20]

5   European infrastructures

The increasing attention to research data, also in humanities, besides theories, has been carried out in European projects and infrastructures, both with an international participation and with individual country or centres initiatives, collecting and making available to access these data, or providing the instruments to support and to promote these initiatives. In many cases the process is integrated in a context of interest toward digital resources concerning the research, in a general and widened meaning. In the practice, indeed, the distinction between research data and data is not clean: the firsts are a subcategory of the seconds, and being the infrastructures, as said, primarily aimed to collect digital data or resources resulting from research, often also data sources are included in the data collections, together with research's results intended as raw. This circumstance causes a difficult individuation of resources 's nature really preserved and accessible in the different infrastructures and the almost impossibility of a clear and certain classification.

Repositories and research data list published by DataCite,[21] re-

---

[20]Referring to the knowledge's free movement, Janez Potocnik, European commissioner for the Science and the Research, in 2007, in the occasion of the Green Book presentation «The European Research Area: New Perspectives», talked about a "Fifth freedom". The other four freedoms are those of common market(free movement of people, services, products and capital) enunciated in the CEE Treaty (1957). For more information, see: http://cordis.europa.eu/fetch?CALLER=NEWSLINK_IT_C&RCN=27454&ACTION=D.

[21]http://datacite.org/repolist, the list is continuously updated.

alized through the DataBib initiative[22] (the instrument for the individualization and location of these infrastructures), shows that, between European countries, most of humanities repositories is located in the United Kingdom. For when it comes other countries such as France, Sweden and the Netherlands, the result is just one for each country.

The Registry of research data repositories re3data.org[23] founded by the DFG (German Research Foundation) allows to search by repositories 's subject and to the *Humanities and social sciences* category are linked nine results. Most of these (with the exception of two cases) are European repositories. Between the individual States, a well-advanced position in the realization of digital resources infrastructure, including research data, is owed, as said, to the United Kingdom.

The Arts and Humanities Data Service (AHDS) service[24], founded by JISC Joint Information Systems Committee (JISC) and JISC, was born in 1996 as a national service, with the aim to collect, preserve and promote electronic resources resulting from research and teaching in the arts and humanities. Funded until the end of March 2008, is now decentralized between host institutions. Divided into five disciplinary areas (archaeology, history, literature languages and linguistics, performing arts, visual arts), it incorporates the infrastructures sharing the same scopes, collecting and making accessible, those defined digital resources trough data archives. The Archaeology disciplinary area is hosted by Archaeology Data Service (ADS),[25] founded by a consortium constituted by the Council for British Archaeology and the Universities of Birmingham, Bradford, Glasgow, Kent at Canterbury, Leicester, Newcastle, Oxford and York. The

---

[22]http://databib.org.
[23]http://www.re3data.org.
[24]http://www.ahds.ac.uk.
[25]http://archaeologydataservice.ac.uk.

aim of the service is to collect, describe, catalogue, preserve and provide user support for "digital resources that are created as a product of archaeological research". Moreover, the ADS promotes standards and guidelines for best practice of creation, description, preservation and use of archaeological information. In collaboration with national and local agencies, involved in the funding of archaeological research or in the funding of archaeological research (Arts and Humanities Research Council, Natural Environment Research Council, British Academy, Council for British Archaeology, English Heritage, Society of Antiquaries of London), it collects datasets of different origin and typology and resources from maps to text report. The research's mechanism, for the user, is structured in two different systems, both much framed. Archsearch for the records allow interrogations by key words, preset categories (What, Where, When) and resources (intended as collections). In the system called Archive the research is set by archives classified into subject, programme and region. Furthermore, there are additional features which not only allow definite research's narrowing, but also offer advanced instruments, such as the map function and the external research (at the moment in an experimental phase).

A French national platform, launched in December 2010, funded by the government, which collects, enriches and provides a unified structure of access to digital data in humanities and social sciences is Isidore.[26] Created by TGE ADONIS (nowadays merged in the TGIR Huma-Num) and realized by the Centre pour la Communication Scientifique Directe (CCSD) with the participation of the Antidot, Sword e Mondeca societies, it is defined research platform and currently collects 80 collections, 2026 sources and 2.271.736 resources.[27] The quick interface allows a prompt research by key word,

---

[26]http://www.rechercheisidore.fr/index.
[27]Figures recorded in May 2013.

by categories (type of resource, category, historical period, discipline, collection etc.), by sources and by repositories.

The TGIR Huma-Num, in a section dedicated to the project in its web site,[28] tackles the question of the peculiarity of humanities and social sciences in relation with the need of data produced sharing and collection. Acknowledging a central role to the information's sources ("sources d'informations"), in particular, is highlighted that, between these, the text is basic in the knowledge production process in humanities and social sciences. The time, that often makes a scientific text rapidly obsolete, does not make the same for when it comes a medieval manuscript, which can preserves the same importance and topicality as the last article published in an international journal.

Some international collaboration infrastructures offer support and services with the aim of improve research's results access and use in humanities, realizing real networks: an example is Digital Research Infrastructure for the Arts and Humanities (DARIAH)[29] that currently counts fourteen member States, including Italy. The project's origins date back to 2005 but the preparatory phase, aimed to identify and define the elements (physical, strategical and human) and financial and legal aspects, concluded in February 2011. The declared aim of DARIAH is enhance and support digital research in humanities and in the arts, realizing a cooperative infrastructure, that puts together national, regional and local efforts, in an interconnected network of instruments, people, information, methodologies etc. DARIAH's has been created in the ERIC form and this normative frame facilitates the long term sustainability of the project: financial and technological needs are confronted in collaboration between members, in a coordinated and uniform atmosphere. The project's grand vision is aimed to facilitate long-term access and use

---

[28]http://www.huma-num.fr.
[29]http://www.dariah.eu.

for all the "European Arts and Humanities (A+H) digital research data". DARIAH operates through four virtual competency centres, the Virtual Competency Centres (VCCs), each of them operating in a specific area (e-Infrastructure, Research and Education, Scholarly Content Management, Advocacy).

Common Language Resources and Technology Infrastructure (CLARIN),[30] exclusively dedicated to linguistic resources, is another pan European infrastructure with members and institutions from thirty-three countries, which offers services to provide a easy access to the resources, through an integrated and interoperable system. Specifically, the scope of the project is to encourage the research's progress in humanities and in the social sciences realizing a unified single sign-on access platform which integrates, at a European level, language-based resources and advanced tools, creating a shared and distributed infrastructure. The preparatory phase of the project started in 2008 and ended in 2011; nowadays CLARIN is still under construction, but a set of services are available and accessible. These are divided into two typologies and the first is constituted by the services for users relating to research, transformation, resources 's archiving: the depositing service, the Virtual Language Observatory (VLO), web services and consulting services. The second typology concerns technical infrastructure services, for CLARIN centers: CLARIN IdP (Identity Provider), CLARIN Discovery Service, Component registry for the Metadata Infrastructure (CMDI), ISOcat concept registry, Relation Registry.

With regard to infrastructures dedicated to scientific data, the iCordi project[31] (which through the European Commission, as said, supports the Research Data Alliance) represents an important goal, most of all in terms of practical attention to interoperability. Started

---

[30]http://www.clarin.eu.
[31]https://www.icordi.eu.

in September 2012 and funded by European Commission, into the Seventh Framework Programme, the project pursues the aim to realize a coordination platform between Europe and the USA, to discuss and improve the interoperability of scientific data infrastructures and to extend this to the global level. In the specific instance, the declared strategic vision is to give an important contribute to the development of the policy for the management and the curation of scientific data, leading to a common policy addressed to the development of a wider global infrastructure. ICordi, which currently counts fourteen partners, will be guided by an High Level Scientific Forum composed by experts in the fields of management and curation policy, coming from both the involved continents, which will realize strategic recommendations aimed to improve the concurrence of data integration, interoperability and infrastructures. The action will be based on three programmes (analysis, prototype and workshop): the first will be devote to analyze data organizations and solutions, as they emerge from the various scientific communities; the prototype one will coordinate activities between Europe and United States important projects, supporting cross-infrastructure experiments on EU-USA interoperability; the workshop programme will investigate the infrastructures 's convergence, paying a particular attention to a wide set of scientific disciplines.

For when it comes individual disciplines, important developments (and almost just italian initiatives)[32] has been realized in archaeological field, as a consequence of the subject's characteristics which, because of needs and methodologies, make research data sharing an exigency, besides a value added. The Advanced

---

[32]Italy, through institutions and centers, takes part in European infrastructures projects, such as CLARIN and DARIAH. In the archaeological field the Metodologie Applicate alla Predittività del Potenziale Archeologico (MAPPA) has realized MOD (MAPPA Open Data, an archaeological digital repository, available at http://mappaproject.arch.unipi.it/mod/Index.php.

Research Infrastructure for Archaeological Dataset Networking in Europe (ARIADNE) project[33] launched in February 2013, with the expected duration of four years and funded by the European Commission, into the Seventh Framework Programme, aims to the creation of an archaeological data infrastructure. The scope is to realize an infrastructure for a transnational access to data, to instruments and to guidelines through a collection of several databases, offering a unified access point and instruments to place new technologies at research's disposal. Coordinated by the PIN (Polo Universitario Città di Prato) of Università degli studi di Firenze, with the collaboration of the Ministero per i Beni e le Attività Culturali and other italian institutes, it gathers partners coming from sixteen European countries.

This brief roundup of some main examples of research data infrastructures in humanities, shows that Europe proceeds in two different but analogous and linked ways. Indeed, on the one hand, there is the creation of international level infrastructures that collect, support, address, answering to a need of coordination and collaboration; on the other hand, the subject of this coordinative action seems to be the process of creation and need of research data sharing, collected in other infrastructures through independent or not initiatives. The single States role in defining policies to promote, support, improve, enhance this process in progress, is crucial not just to make the individual enterprises efficient, but also to restrict lacks of homogeneity, favoring the realization of systems easily identifiable and usable singly and integrated.

---

[33]http://www.ariadne-infrastructure.eu.

6   Conclusions

Research data constitute a typology of material which have a value that, in the field of research and in the more general sharing of knowledge, enables it to have a place in the European policies and in the international organizations 's documents. The production and the existence of this kind of data, integrated with the possibilities offered by information technologies (systems of collection, management, preservation, interoperable languages and formats, the internet), have led to the awareness of the advantages, at different levels, coming from their free sharing (and from their free use and possible re-use). In particular, for research data resulting from public funding researches, to the reason of advantages given to science, progress, economy, society, is added the one concerning the necessity (if not the duty) of making research's results available to the same funding users. If in theory (how it can be inferred also by documents of the European Commission, the OECD and the UNESCO) all that is recognized and, in the case of the European Commission, the awareness is addressed to real initiatives, in practice difficulties exist and persist. The obstacles to the realization, to the development and to the efficiency of systems for free and open sharing of research data are of different kinds: financial, organizational, technical, legal, cultural. To identify them and to find out answers and solutions is crucial to realize valid and solid models and infrastructures aimed to guarantee the correct collection, diffusion, preservation and sharing of research data. However, until then, data will continue to be produced but many of them will not be available, other will be available for restricted lengths of time, other more will not be available for everybody (because of circumstances depending on, for example, visibility, formats or access conditions). This will make investments inefficient (or efficient in a reduced manner): many researches will be repeated, times for the finding of material might be

long, resources invested in projects without adequate financial and risk planning could be lost. However, the European situation, on the whole, seems to move toward the realization of policies, systems and infrastructures dedicated to research data, paying attention to what guarantees the efficiency (financial and technical sustainability, interoperability, etc.); see the examples of DARIAH and CLARIN projects.

In Italy the purpose of tackling the research data question emerges, for the first time, in the document Horizon 2020 Italia. So that this declared intention would not remain just a purpose but would become a concrete reality, some State's actions will be essential:

- In the scope of responsibilities 's individualization, the realization of a national coordination center in the field of open access to resources resulting from italian research.

- Financial investments into national initiatives and participation to pan European ones, such as for example the realization of a national infrastructure, integrated with European systems.

- Promotion of an open access culture, that would remove all fear (often springed from the lack of knowledge of new diffusion models and practices, such as the Creative Commons licenses), proposing new solutions.

- Complete and clear policies 's definition, that maintain the separation between open access to scientific articles, and to research data. Moreover it is important to maintain a conceptual separation between the open data in a generic sense (which include, for example, data from public administration) and research data which have proper and specific characteristics and questions and which need focused actions. These policies should draw up multi-year financial plannings, but also strate-

gic ones (such as risk action plans), with the aim to guarantee the long term sustainability.

Finally, it would be suitable and desirable that the question of research data in humanities, which, as said, suffer from the subordination as regards attention they have in scientific-natural sciences, be considered with the peculiarities which characterized the research and its results in this ambit. To look at the reality of humanities paying attention to the peculiarity that characterizes it, would stimulate a greater interest and development, keeping the value of the comparison and the sharing of solutions, practices, questions and attributes between the two sectors.

References

Australian National Data Service. *What is research data?* ANDS, 2011. http://ands.org.au/guides/what-is-research-data.pdf. (Cit. on p. 210). Print.

Borgman, Christine L. *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge: MIT, 2007. (Cit. on pp. 212, 214). Print.

—. "The digital future is now: A call to action for the humanities". *Digital humanities quarterly* 3.4. (2009). (Cit. on p. 214). Web.

Burrows, Toby. "Sharing humanities data for e-research: conceptual and technical issues". (2011). (Cit. on p. 214). Web.

Commissione Europea. *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Le Infrastrutture TIC per la e-scienza*. Bruxelles, 2009. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0108:FIN:IT:PDF. (Cit. on pp. 216–218). Print.

—. *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle regioni. Un'agenda digitale europea*. Bruxelles, 2010. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:IT:HTML. (Cit. on p. 216). Print.

—. *Comunicazione della Commissione al Parlamento Europeo, al Consiglio, al Comitato economico e sociale europeo e al Comitato delle Regioni. Verso un accesso migliore alle informazioni scientifiche: aumentare i benefici dell'investimento pubblico nella ricerca*. Bruxelles, 2012. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0401:FIN:IT:PDF. (Cit. on pp. 209, 211, 216, 218–220, 225). Print.

—. *Comunicazione della Commissione al Parlamento Europeo, al Consiglio e al Comitato economico e sociale europeo. Sull'informazione scientifica nell'era digitale: accesso, diffusione e conservazione comunicazione*. Bruxelles, 2007. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2007:0056:FIN:IT:PDF. (Cit. on pp. 216, 217). Print.

—. *Proposta di decisione del Consiglio che stabilisce il programma specifico recante attuazione del programma quadro di ricerca e innovazione (2014-2020) - Orizzonte 2020*. Bruxelles, 2011. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0811:FIN:it:PDF. (Cit. on p. 220). Print.

—. *Raccomandazione della Commissione, del 17 luglio 2012, sull'accesso all'informazione scientifica e sulla sua conservazione*. Bruxelles, 2012. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:194:0039:0043:IT:PDF. (Cit. on pp. 209, 217, 218, 220). Print.

De Robbio, Antonella and Silvia Giacomazzi. "Dati aperti con LODe". *Bibliotime* 14.2. (2011). (Cit. on p. 212). Web.

Key Perspectives. *Data dimensions: disciplinary differences in research data sharing, reuse and long term viability. SCARP Synthesis Study*. Digital Curation Centre, 2007. http://www.dcc.ac.uk/sites/default/files/documents/publications/SCARP%20SYNTHESIS.pdf. (Cit. on p. 223). Print.

Ministero dell'Istruzione, dell'Università e della Ricerca. *Horizon 2020 Italia*. 2013. https://www.researchitaly.it/uploads/50/HIT2020.pdf. (Cit. on pp. 225, 226). Print.

Murray-Rust, Peter. "Open Data in Science". *Serials Review* 34.1. (2008): 52 –64. (Cit. on pp. 212, 219). Web.

National Science Foundation. National Science Board. *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. Washington: National Science Board, 2005. http://www.nsf.gov/nsb/meetings/2005/LLDDC_draftreport.pdf. (Cit. on pp. 212, 214). Print.

Organisation for Economic Cooperation and Development. *Principles and Guidelines for Access to Research Data from Public Funding*. Paris: OECD Publications, 2007. http://www.oecd.org/science/scienceandtechnologypolicy/38500813.pdf. (Cit. on pp. 210, 214, 221, 224). Print.

Suber, Peter. "Promoting open access in the humanities". *Syllecta Classica* 16. (2005): 231–246. (Cit. on pp. 213, 219). Web.

Swan, Alma. *Policy Guidelines For The Development And Promotion Of Open Access*. Paris: UNESCO, 2012. http://unesdoc.unesco.org/images/0021/002158/215863e.pdf. (Cit. on pp. 209, 222). Print.

MAURA FUNARI, laureata in Storia e Scienze della Documentazione, Università degli Studi di Perugia.
funmaura@libero.it

ABSTRACT: The article holds an inquiry on research data in the field of humanities, into the European zone. Defining these kind of data as primary source and basis of the scientific research's results, specificities characterizing them, in humanities, have been individualized. The attention paid to research data in European policies confirms their strategic role to the development and optimization of the scientific research. The analysis of a few research infrastructures and projects focused on research data in humanities, shows the state's policy central role to improve and develop them, making individual activities efficient and restricting a lack of homogeneity, supporting the presence of easily identifiable, usable and integrated systems. In this context, up until today, Italian rear position referring to systems of collection, management, preservation and sharing of the research data, seems to place itself in development's prospects.