

Pedro Ruiz Pérez

Little Big Data: The Poem Against The Database

Digitalization entails processes of uniformization that are not always desirable. In the Aristotelian model of science (to know by universals), the neutralization of specific differences is necessary and is the basis for intellection, and this guides the formulation of laws of general application. In the field of aesthetics, and even in that of history, being able to generalize is inescapable and valuable. This is so in the field of theory, in *longue durée* periodization or in the defining of genres (literary, artistic . . .). However, ultimately the value of singularity endures: the singularity of the artistic object and the receiver, be they student, researcher, or one who does so for pleasure. The notion and development of the “digital humanities” offers up an initial solution to this duality – or at least a space for dialogue of the double perspective. The potential and speed of growth of the technological component of the polarity multiplies the modalities of application and at the same time colonizes the supposedly shared terrain. Without going too deeply into the difference between information and knowledge, the play of tensions runs the risk of reduction in the face of a situation of hegemony when numerical logic,¹ and its extreme dimension, the realm of Big Data, imposes itself. What follows is a reflection undertaken from a position of one working in research connected to the Digital Humanities, and with the distrust of one who is wary of its risks and limitations.²

1 Obviously, I am not referring to the meaning that prevails in the use of *numérique* in French, for which the term *digital* is used in Spanish and English. Neither am I evoking the idea that the superseding of the analogical is produced through the 0/1 binary code, numbers that are also coming to replace letters, beginning with those of the alphabets.

2 The reflections that follow have arisen while working on the ongoing project, *Biografías de autor e institución literaria en la edad moderna* [Author and Literary-Institution Biographies in the Modern Age] (SILEM) RTI2018-095664-B-C21 of the National R&D&I Programme (<<http://www.uco.es/servicios/ucopress/silem>>). It is a continuation of *Sujeto e institución literaria en la edad moderna* [Subject and Literary Institution in the Modern Age], a National R&D&I Programme coordinated project, FFI2014-54367-C2-1-R. Prior to this, there were two projects (PHEBO) on late baroque poetry (<<http://www.uco.es/phebo/es>>). In this period I also coordinated the setting up of the Aracne Network (<<https://www.red-aracne.es/presentacion>>), along with the complementary *Humanidades digitales y letras hispánicas* [Digital Humanities and Spanish Literature], National R&D&I Programme FFI2011-15606-E.

Forms of Reading (And of Research)

If the steps I follow and the end I seek had to be reduced to a formula, it might be found around a dialectic between the quantitative and the qualitative, one in which balance was not imposed by an overwhelming power and which did not end up being diluted by the traditional relationship between means and ends. In other words, one in which mere technology does not end up dominating science, again with reference to Aristotle. The aim is to preserve and cast the essential factor of the aesthetic experience in a renewed space, while finding, through experience, a balance between certain bases in tradition and the expectations of new technologies and epistemologies, specified in research possibilities. I begin with a re-affirmation of the subject.

Yo nací (perdonadme) / en la edad de la pérgola y el tenis” (“I was born (forgive me) / in the age of the pergola and tennis”), as go the well-known lines by Gil de Biedma. I had no *château* with grass courts; I only belong to a time, or, at least, come from one. More accurate, both in general and for this case, would be Rafael Alberti’s words, which influenced de Biedma: “Yo nací – ¡respetadme! – con el cine” (“I was born – respect me! – with cinema”). And I say this while witnessing the growing hegemony of series on television platforms – that is, I am aware that the basis of my scientific knowledge has shifted to another dimension, something that has happened with audiovisual culture ever since the Lumière brothers.

In a space that is so propitious to uniformity, it is worth stressing the statements that give character to a position, which is what I now maintain. In essence, this is the consideration of the nature and specificity of the literary text and of philology, as well as an active assumption of the need for qualitative advances in research projects and professional networks, with the role they play in this framework of Digital Humanities. My starting point is the experience of this journey with intellectual conviction, not blind faith. From this conviction, in which there should always be a seed of doubt, I contemplate the changes that are taking place.

We should not lose sight of the lines of continuity in the paradigms. I wrote my thesis on a typewriter and using index cards. It was common practice, as it was to incorporate statistics of verses, including rhythmic variants. We were not, therefore, adverse to working with data, and that included an aspiration to globality (aimed for by the habitual title of “Life and Work of . . .”), for an all-encompassing view (which, by the way, the new thesis model has fragmented), and for the consequent development of management protocols.

Although now is not the time to historicize or fall into the habitual narcissisms of the academic field, it may be a good time for a backward glance to have a clear awareness of the place I am speaking from. Since 2005, the research pro-

jects I have been involved in with some responsibility have tackled objects in which the quantitative factor was foremost, and with a not insignificant scope in the terms of our field.³ In fact, a methodological premise of mapping the field was a constant in all of them, resulting in the producing of a more or less extensive corpus or repertory, although in the usual terms in the research of our field, it could be considered relatively *big*. This was the case in the study of fifteenth- to seventeenth-century poetic texts with lists of authors, in the analysis of the poetry that we ended up calling the Late Baroque, and, finally, in the research on discourses in which the concept and image of the author and the literary institution was forged – now in a longer period that stretched to the mid-nineteenth century. We thus addressed the work that could be considered a poetic subgenre, in a defined chronology and in a modality of discourse, with aspirations of attaining exhaustiveness. The instrumental aims were the drawing up of repertories or catalogues, databases and digital libraries. The differences denote a process that was both technological and conceptual: in the first project we worked with paper index cards and Word files, resulting in the publication of a book (Ruiz Pérez 2010), in which the traditional indices tried to maintain a certain flexibility of use. The second project included the design and creation of a database with an advanced search system and a library of static texts, all available on a website. Lastly, in recent years, digital libraries with treatments of texts and the capacity for conceptual searches have shown their potential for examining the compositional mechanisms of a discourse.

This experience can explain my perspective. Its timing, coincidental with what can be considered the widespread development of the Digital Humanities, gives it a value of some significance. Computerization has exponentially increased both pace and possibility, even turning them into a qualitative change, similar to that brought about by the printing press when it overcame the incunabulum phase (reproduction of manuscripts) and configured another model of volume; or, as Guglielmo Cavallo (1975) and Armando Petrucci (1979) have shown, in the previous shift from the scroll to the codex: changes in formats, changes in models of reading. We are immersed in an equivalent paradigm shift.⁴ We need to seek suitable adaptation, and this involves a review of the problems and limits of re-

3 Please see the references and information given on the websites mentioned in the previous footnote.

4 On the speed of change, see the advances in information and reflection given in the monograph coordinated by Morrás and Rojas Castro (2015) and in the volume edited by González and Bermúdez Sabel (2019). Since the start of the century there has been a proliferation of associations, conferences and journals on Digital Humanities, which record the extent and progress of this discipline. References and links can be found on the Red Aracne website, cited above.

search, a reconsideration of the object (susceptible to being tackled in broader terms, also quantitatively), and an appropriate design of the tools, hierarchically subordinated to the previous principles.

In the shifts and turns undergone in the Digital Humanities, the initial stage multiplied the possibilities of the individual reading, making a greater volume of information available in the form of texts. Globalizing models in the consolidation of the practice and the refinement of its tools will soon be established, if they have not already. The development of the technique involves multiplication and acceleration; its results, in exponential growth, first affect the object and end up imposing uniformization, which affects the perception of the subject. The process, logically, is in perfect harmony with the discourse of globalization. The problem makes itself felt in its application to poetry. I use the strict meaning, which has been central to my work, but the term can be applied to all textual reality that cannot be reduced to strict categorization. Even without positioning myself at the extreme end of the paradigm of the singularity of the literary text, its resistance to the processes of neutralization is clear here – or, at least, the necessity for nuances and particularizations.

The methods of quantitative analysis, and their application with tools that exponentially increase the numerical dimension, open up exciting paths for a renewed study of literature. However, many of these paths are unexplored; down them lies the risk that the hope or impression of oases is but a mirage. In order for the numbers of digitalization not to be imposed upon the discourse of words in the humanities, we need to reflect, based on the premise that all technology implies ideology, and prepare for incursions into wild territories without due calculation of risk. In this perspective, the aspects considered in this reflection are those relative to the specificity of literature and its discourses, the role of digital tools and the dimension of big data in research in this field, the redefinition of the object of study, and the creation of corpora to scale. I offer this as a trial run of the possibilities of dialogue between the critical and philological tradition and the epistemology derived from a technology that is still expanding.

The dilemma is this: will we master the technology and put it to the service of certain criteria, or will we sacrifice the criteria to the omnipotence of the god of technology? In less Manichean terms, up to what point should we refresh our criteria with new perspectives?

Some Aspects for Reflection

Let's take the following postulate: the relationship between data and base, the published edition of a text and the digital library, the index card and big data, reproduces a comparable text-macrotext relationship, conceptually and functionally, which operates with particular intensity in a collection of poems. As a part of the whole the poem does not lose its autonomy, but it is enriched in the dynamic of the book. And it is updated in every reading. Can we establish and maintain such a relationship in the design of our digital tools?

To walk this road, we need to establish a perspective, in its double meaning of a depth of field and a line of sight. We can do so with a question that requires an answer: what differences are there between humanist concordances, of the Bible or the oeuvre of Virgil, and the possibilities of a database, aside from making it possible to endlessly expand the corpus? The advantages produced by technology, where we must focus, enable us to surpass the possibilities of the alphabetical search to find the use of a word or its recurrence in the corpus of a work or author. Its repercussions in the design and execution of a research project in its widest sense are found in several shifts: 1) from the static to the dynamic; 2) from the linear to the relational; and 3) from the strictly lexical to the conceptual.

1) From the static to the dynamic. This is the relationship that takes the economy of a book of poems as parallel: that is, how its components work with the established *dispositio* –talking in the terms of classical poetics – or, in the semantic field of computers, with its architecture, with the functionality that entails. One lesson with regard to overcoming the static nature of text is that of the linguistics of the text, with its concepts of coherence and cohesion, its notion of discourse and the turn to the syntagmatic. Its model is an invitation to abandon the consideration of isolated and static elements, typical of an outlook with its basis in the paradigm. The model requires a syntagmatic consideration, in which the elements are conceived relationally, by their functions and their capacity to establish nexuses and dialogues with the rest of the elements *in praesentia* – to continue with established terminology in Philology and structuralist linguistics. Upon introducing data, we imitate the constitution of a paradigm, that is, a more or less systematic set in which elements that maintain some kind of relationship between them are arranged linearly. This is not the only line possible but one established with a specific objective. In principle, these data can enter the series with the independence appropriate to their real situation. Looking at the framework we are interested in, a piece of writing is materialized in a book (on paper or digitalized) with autonomous existence, whether on the shelf of a bookshop or of a library, no matter their organizing principle (by author, genre, chronology . . . even by size or colour). Only in the activity of the reader who

feels moved by the message of the text they read is there the possibility of jumping from one book to another with productivity. The conception and mechanics of a database or a digital library make it possible to surpass that reading dynamic, enhancing and intensifying it, multiplying its potentiality. The quantitative difference (in the number of texts and possibilities of interconnection) can be turned into a qualitative turn. When conceiving it grouped together, the datum/text ceases to be an autonomous element and a value that literary studies emphasized some time ago comes into play: all literary work is a weaving of intertextual relationships, where we can find the beginnings of a principle of dynamic relationships that transcend the strict frontiers of the text in its strict immanence. Its semantic and aesthetic richness comes alive to differing degrees, depending on the culture and capacity of each reader. In a computerized repository, at least as far as research interests are concerned, such a circumstance should be maintained, while making the most of the potentiality provided by technology to increase this dynamic of relationships and materialize them on a screen.

A paradox emerges, however. In literary philology, for centuries we have been working not with a language (limitless productive force, *dynamis*, *energeia*), but with texts (paradigm of perfect, complete *ergon*) – Text with a capital T, moreover, emblemizing a clear paradigm: the reduction in criticism of all the variants generated by a process of dynamic and complex transmission.⁵ The supposed reconstruction of the archetype or ideal text is, for a non-specialist reader, situated on a line of exclusion opposite to the virtuality of hypertext. Yet the researcher or reader moved by philological curiosity can make the most of the critical apparatus to rebuild the trajectory of the text's different avatars and contemplate it in its dynamism, whether to consider the process as a whole or whether to pause in consideration of one of the episodes of their communicative course, of their different historical realizations. As with the concordances, in the philological tradition we discover – with all the nuances brought about by technological changes⁶ – a princi-

5 I am thinking here about the essential aim, established according to principles related to faith in the ideal text and, therefore, the consideration of errors for all its variants, so that critical work consists in reverting disfigurements to return to the point of correction, the original (whether existent or not). This is what should be offered to the reader. The utility that the traces of previous work leave for the specialist, and that can be recorded in the same volume where the archetypal text is published, with the critical apparatus, is something else.

6 It is worth recalling that the development of concordances, such as indices and other paratextual elements of reading orientation, was only possible with the shift from scroll to codex – that is, from linear arrangement to a system of folding that enabled page numbering and therefore the ease of finding an element in them, which before was only relatively possible in highly articulated works, such as in books, chapters and verses, as can be seen in the Bible. The Christian holy book, with its fideist conception, also has an influence on the methodological configuration

ple of reading (and even of textual arrangement) that very much takes into account the dynamics of the text, its influences and internal recurrences and its material transformations. Perhaps we can continue to learn from what has been opted for in a centuries-old philological task in which the setting of the ideal text is presented along with the testimonies of a tale of transmission that keeps a dimension of dynamic hypertext alive.

The digitalization of the literary humanities can be adapted to a mode of work and the principles that it is based on, as long as this is done by tipping the scales toward the opposite end of idealizing staticism. Into the black box of computer tools,⁷ we introduce finished, closed texts; their very number can confine us to a merely quantitative consideration (such as counting an author's Sapphic hendecasyllables, as per the aforementioned model of theses of the past century). The challenge is to convert this new rhizomatic syntagm into an energy in motion, in which to maintain or generate qualitative reflection, preserving and planning a similar functioning to that of poems – as mentioned above – in the framework of a well-constructed book. In its volume, above all when this reaches extraordinary – or at least uncommon – dimensions, the data threaten to become an answer, the answer, with totalitarian temptations. How can we avoid this? Let us now turn our gaze to the experimental sciences. Set to collide, texts, like atoms, generate energy, but they also open vacuums, and in them we must find the questions – the questions that truly advance research. In physics, technological developments have enabled the corroboration of previous scientific theories that were born out of analysis, reflection and judgement. The Geneva Large Hadron Collider finally proved Higgs right in his theory of the boson, which pre-existed its experimental proof. The case seems a model for the situation I am setting forth, and we cannot stay on the sidelines of this situation and its implications in terms of the model of relationship between science and technique. When it is not a theory that precedes the research or data accumulation, the latter in their quantitative dimension should become the source of new research questions. In short, the amount of data is not sufficient by

of a certain model of ecdotic work, that of Lachmann and his followers, such as how the assessment of the *auctoritas* of Virgil or Horace shaped the practices of Alexandrine philology, until both models and perspectives converged in the Renaissance *studia humanitatis*.

7 I use the notion of a “black box” taken from the scientific paradigm to highlight the relative independence of philological research approaches regarding technology, in the sense of keeping the relationship between *episteme* and *techne*, between research and its tools. For the humanities researcher, it is not essential to be an expert in computing, however much an effective familiarization can produce better results. The dialogue at the heart of a multi-disciplinary team brings about the most propitious situation, since it prevents the autonomy or primacy of philological principles from becoming impermeability. As I have been stressing, technology invites the re-thinking not only of our procedures but also our concepts and objectives.

itself; what we need are modes of treatment to convert them into a dynamic element of investigation.

2) From the linear to the relational. The digitalization of a library catalogue, for example, or the creation of an archive, are not in themselves digital humanities, because they stay on the horizon of the paradigm and of linearity, in a very similar way to the text of traditional writing. The management computing tool facilitates searching in cases such as those mentioned, but does not generate knowledge by itself. The information is quicker, but it is wholly outside the questions of research – at least, if the data is kept on the level of a schematic file, such as in an OPAC. This is due to the lack or the weakness of relational models and tools. The possibility of making connections is a requirement, being essential to give a specific humanist meaning to digitalization, beyond the generic usefulness that streamlining produces in the handling of existing information. The objective of the application processes of computing to research in the humanities is the (well-oriented) creation of new discourses and knowledge. The key is the formulation of the questions. Thus it is possible to go from mere information (that of more or less numerous data) to knowledge (based on judgement/reason).

Computer memory, with its differences – above all in the order of quantity – reproduces that of the human mind, which constructs it to its image and semblance. In both cases, we can distinguish two planes. First, there is the plane of a merely receptive arrangement, in which memory is converted into a deposit for the storage of information, articulated in data with differing degrees of refinement or extent. Like material warehouses or the buried remains of an old civilization, they lack value as long as they are not put to use. This is the function of the other component of memory, as capacity to activate, to put to work and connect the available elements. This dimension is vital not only for the advance of scientific knowledge but also for the very survival of the individual or the justification of a tool.

Let us return to the mechanism of reading, from the most conventional to, in a little more than one and a half millennia, the reading fomented by hypertext. In the former, the updating of the text when received, by means of the memory (increasing with the degree of culture and capability of the receiver), calls upon other texts read, stored in the memory, but updated with a reading that, through intertextuality, activates the weave of relations. The reader has the materiality of a volume of paper and lines of ink in their hand, while in their memory they hold the vestiges of previous readings from similar objects. Intelligent reading adds to comprehension, and delight in the text revitalizes those vestiges that are only seemingly absent. The essence of this mechanism should be preserved in computerized procedures. At the same time, it should be enriched with its possibilities, since its technology makes almost instantaneous retrieval possible and from a considerably larger volume of

data, but, above all – and here we must progress – it makes retrieval possible in simultaneity. In the framework of research interests, this seems to me to be a more productive condition of hypertext than that of substituting linear reading for the attractions of browsing, and much more evocative and fruitful in the processes of active reception of artistic creation in the virtual space.

In the scientific sphere – with the humanities included within it – the outlook of objectivity means prioritizing denotation over connotation. Thus the content of big data resources does not place its value so much on the plane of the possibilities of browsing as on that of the handling of data with a certain simultaneity, that which gives the relational weave, of the available data. Unlike a reader who reads for pleasure, open to the suggestions of the (hyper-)text, whoever accesses the tool with aspirations for scientific knowledge tends to do so guided by prior questions, research questions that have emerged from earlier readings and the resulting curiosities. However, the condition of the Digital Humanities makes it possible to offer a repertoire of dimensions that was unmanageable until fairly recently, and this reality could become, in turn, fertile soil for new questions – research questions that would be tricky to consider and resolve without the existence of this tool.

In order for the answers to questions to emerge, there is the inescapable requirement of designing a conceptual and digital treatment to enrich the data and give them meaning. An image of this challenge is provided by the distance between how social networks typically function and the creation of a community framework with a bundle of relations. It is not enough to provide possibilities to generate a positive discourse. One route in our field is to incorporate not merely denotative information in the materials that are included in the database and go on to form part of big data. I don't believe it necessary to return to the example of a computerized library register in which only the mere data of bibliographical identification are recorded, and this case is transferrable to the formation of databases of critical bibliography without distinction between their data and their metadata. I propose broadening the dimension of texts in an opposite sense to that of "augmented reality": to deepen rather than adorn.

3) From the strictly lexical to the conceptual. The transcription or edition of a text for its addition to a repository only mechanizes and facilitates access, but it does not alter our knowledge, beyond the loss that is entailed when passing from the material to the virtual.⁸ A search tool using words or sign sequences makes some

⁸ Material bibliography has shown us how to discover the semiotic potential of all components of a book, particularly one coming from manual printing, where it is meaningful from its format, lettering and paper type to the paratext or constitution of its folds. Much of this information is lost in digitalization and most of the transcriptions available online, and it is not always recov-

processes possible, but does not open the way to readings and new questions. For a new stage of knowledge in the application of digital tools to literary studies, we need to advance in the hypertextual, with added levels of information, and to design suitable search tools for a relational, not linear, scenario. In more philological – rather than computer – terms, this is about creating new reading tools that positively meet the possibilities of the technology. One possible reference here is the work of Franco Moretti, with his notion of *distant reading* (2013) and its application in the activity of the Stanford Literary Lab: tools for new readings.

Some of the proposals regarding artificial intelligence are oriented in this direction. However, they do not avoid but rather increase what for us is a risk of partitioning, bias and orientation in the image offered. The algorithms of the large search engines serve as reminders in this respect. When we use them, beyond the craftsmanship of the criterion and the appropriate technique for their application, what degree of control do we have over how the software functions? Moreover, can we clearly discern whether we are in control or are being controlled? Volume of information and speed of access to it are in principle positive values; but they are not always synonymous with knowledge – and I'll let the evidence speak for itself. Indeed, they frequently replace it. From the drive that moves knowledge, the accelerators provided by technology only make sense insofar as they enable us to construct lines of access to reality (in our case of texts) with a greater chance of comprehension, free of distortions. Quantitative and technological intervention in the entity of the object of knowledge entails, at the speed we are moving, a metamorphosis of the subject's structures of thought – or, more radically, a mutation of the thinking subject.

In a process of decontextualization and neutralization brought about by its integration into the mass of data, a text – particularly a literary text – suffers a reduction in communicative efficacy. Like splashes of ink on paper, the binary code of computing cannot replace the active mechanism of reading, but neither should it limit it. In the workshops of the heirs of Gutenberg, in a process of coding reached by consensus, they fixed rules, elements and patterns of composition that regulated and conditioned what at the beginning of the modern age was considered mass diffusion and reception. A similar process took place in the computing laboratories and in the management offices of corporations that direct their work. The result is that secular textual models can become denatured when rendered into new formats and, above all, when subjected to consumer processes

ered when editions aimed at this circuit are made. Along with accessibility, other possibilities of virtualization should be considered in the treatment of texts that would make it possible to compensate for this loss.

that are far more massive than those seen back in 1500. And the phenomenon multiplies exponentially when it is the texts themselves that are diluted in the dimensions of big data. The linearity of plain text of the printed page is converted into a relic and is denatured with digitalization and insertion into a macro-corpus. Either it runs a serious risk of doing so if the particularities of the literary text are not taken into account, or its reception submits to a capitalist logic of the number. Consideration as a datum places it on the plane of pure information, which it does not belong to, in the strict sense, or which at least does not prove specific to it. Its profitability lies in the possibilities of massification. Not only must they be reduced to data; there must be many, vast numbers, like the profits on an income statement. Thus the *data* have to be *big* to have productivity. As a counterweight to this logic, what is moved by other interests is what must act. The digital humanities must not lose the first of their components, the substantive, but reserve the second for its auxiliary, instrumental function.

The condition of the datum, the minimum element of information, enables its insertion in a quantitative paradigm, even on a mass scale, where it functions like the lexical units in dictionaries – those graveyards of words, according to Julio Cortázar. Only when words become concepts is judgement possible, and we again turn to Aristotle. Thus, it is the conceptual treatment, the dimension of the humanities, that can prevent the great catalogues from becoming the macro-necropolis of postmodernity and the new versions of a liquid capitalism with its uses and values. Philological editing has carved out a well-defined space to give this value to the printed word. The footnote made it possible to incorporate a deeper level of reading, or gave it a dimension that was not entirely explicit in the text. The explanation of a term, a reference or a rhetorical schema maintained in the (para-)textual *dispositio* the etymological meaning of *explicare*, ‘unfold’, since it opened a widened – one might say hypertextual – horizon in the reading. When the footnote established the appropriate connection with the text, it ceased to be a simple datum and created a space of knowledge. If the printed page allows this duplicity of planes, expandable with subheads and double or triple apparatus of notes, the flexibility of the volume gives the chance to add the informative or analytical appendices as required. The screen, in turn, exponentially multiplies these possibilities and takes the consideration and treatment of the text into an essentially different dimension.

The notion of data and the illusion created by the implicit and latent consciousness of their availability in a database generates unconscious mechanisms similar to those of the market and consumption. In the push to purchase, the supply from large data warehouses – now the great online retail platforms – creates the generally false impression of availability, which undoubtedly facilitates where the spending is directed, shoving it closer to wastefulness translated into profits. In the

marketing sphere, the effect is no less irresistible even though it is known. In the field of social engineering and political manipulation, it constitutes one of the biggest threats to the present and the immediate future, that of the next elections. In the sphere of the humanities, these can be diluted by its new travelling companion, digitalization.

Should our discipline remain in the state it was found in when Lorenzo Valla revealed the forgery of the Donation of Constantine with which the Church sustained its earthly power? It is obvious that it should not, as obvious as the need to think about what can be changed and what is worth addressing, without losing sight of certain values and the objectives they entail.

Towards a Liberation from the Fetishism of Data

Not entirely unintentionally, I continue with an image that is as schematic in its formulation as it is rough in its visual resolution (Figure 1):

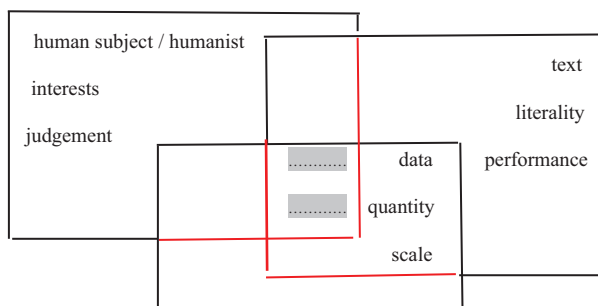


Figure 1: Reading and research processes. By the author.

Playing with graphic reduction serves to show the confluence of three elements of a disparate nature and the establishment of a space of intersection in the processes of reading and research. In a triangular relationship, we find the focal point represented by the interests of a subject, who cannot relinquish judgement; some objects of specific nature, based on an irreducible specificity that is revitalized in every reading; and, lastly, some procedures (conceptual and technological) that operate with intrinsic values unconnected to those of the other two elements – a philologist searching for a text in a digital library or a reference in a database. But let us return to the schematization.

Venn diagrams, used in my time at school to develop set theory, and the parallel of structuralist formalization, continue to serve, constrained by the limitations of Word and my computing competence. Thus its representation emblematically illustrates the dialectical triangle that I base my analysis upon and which shows the following: the opportunities and threats of digitalization in terms of big data, the specificity of texts such as literary texts, the unavoidable guiding presence of a critical subject and, overall, the need to map and colonize the intersecting space between these vectors. The limitations of the representation illustrate the spaces of perplexity and disorientation that we face in the endeavour. And in this we must distinguish between the constraints of technology and our limitations with that technology.

I shall put to one side – but not negate – reflection on the dependence of hardware, software and memory, or the potential depredation of their human correlates – operational until a couple of decades ago. Sticking to what I have outlined, and to the occasion, I shall confine myself to questions on the strategies for the positive resolution of the dialectic I have set forth. The orientation of these questions points, in my opinion, to the introduction of dynamisms in catalogues, which cannot be inert. Thus the potentiality of knowledge derived from numbers can enhance the experience – singular in nature – that arises from a more traditional reading. And this dynamism in the big data archive should not be entrusted to the typical algorithm. On the contrary, it must be established with a perspective that does not exclude the sense of the literary experience. The nature of this experience and the research regarding it should adapt to and fit the territory to be colonized, represented by the large catalogues; and, above all, the intervention must be regulated, through the investigation processes, in the definition, establishment and management procedures of these tools. Their purposes are too important to relegate their access to algorithms that are not controlled, or left to chance. It is not only a question of handling the existing tools and catalogues, those that are already given (or imposed upon us). It is imperative that, through our field of study, we take the lead in the design and making of the digital tools and the big data catalogues that not yet to be accomplished, so that in this task we keep the best remnant of a critical tradition that cannot be abandoned.

One line of work, developed by our project, is the use of the possibilities of hypertextuality, establishing a double plane – the textual and subtextual – in the materials, and adjusting the tools for the relational management of the data obtained in searches. Semantic labelling makes it possible to revitalize the text, by adding a plane to the mere sequence of signs and, above all, by introducing a semantic architecture beneath the texts, as finished accomplishments. This, through its logic, enables access to the planes of reading in which quantity becomes quality. In our case, the conceptual base is the construction of authorial notion and

image. Around this, the establishment of a systematic ontology, with the arborescent structure on the three planes defined by the TEI system (classes, attributes and labels), becomes the route for introducing a historical and philological criterion in the consideration of a good number of texts and a dynamic relationship between them, beyond their location in a shared repository. This model makes it possible to explore, for example, the recurrence of a value such as social status in the configuration of the image of the writer or the references to the networks of sociability in which they are registered.



Figure 2: Search results in the SILEM Biographies digital library. [http://www.uco.es/servicios/ucopress/silem/buscador/busqueda-pro-final.php?query=%20//note\[@type=%27network%27%20and%20@subtype=%27literary%27%20and%20%20%20%20%20soccerStatus\[@role=%27writer%27\]&cadena=Sociabilidad%20-%20Redes%20-%20Literaria%20and%20Rol%20-%20Escritor&biblio=BIO](http://www.uco.es/servicios/ucopress/silem/buscador/busqueda-pro-final.php?query=%20//note[@type=%27network%27%20and%20@subtype=%27literary%27%20and%20%20%20%20%20soccerStatus[@role=%27writer%27]&cadena=Sociabilidad%20-%20Redes%20-%20Literaria%20and%20Rol%20-%20Escritor&biblio=BIO) (June 3, 2023).

This image of search results (Figure 2) serves as an example. The search was done in the library of author biographies (currently with just over two hundred documents that have been referenced, transcribed and labelled), and the aforementioned coding parameters were introduced to locate the passages in which there is mention of the writerly status of the biography subject and also of their inclusion in literary social networks. We can see how in 1622 a man of letters such as Tamayo de Vargas makes these traits clear when outlining the biography of Garcilaso in the introduction to the annotated edition of his works, as well as the language or resources he uses to express that information. The procedure makes it possible to gather not only the passages where express use of such concepts are used in their most recognizable lexicalization, but also those in which the allusions appear obliquely, even before the critical establishment of a notion. At the

same time, one can locate the other testimonies of this subgenre, in which one can extract the same connection and establish parallelisms and differences, as well as observe the density of its use in the established repertory. It is striking, for example, that less than 25% of the texts collected – even though they are biographies of writers in very different historical and publishing contexts – refer to these circumstances, which are so decisive in our consideration of a literary author.

Regardless of the details of a specific proposal, I think we can make use of the qualitative element, which can become prevention against the fetishism of big data. If you'll forgive me the wordplay, this is going from the virtual as dematerialization to the notion of virtuality as potentiality, which refers to a space of freedom and judgement that is still the patrimony of the subject. In the notional field of data mining, it is imperative to bear in mind that the extractive engineering that makes it possible to mine a seam is as important or more so than the actual wealth of that seam.

To be precise, in the field of literary discourse, one cannot abandon the qualitative dimension, with a certain relativization of the productivism of the *big*, imposed upon the vitality of the materials that have to be reduced to *data*. One factor in this consideration is the fact that we can work with a finite corpus, more or less extensive, but stable and established. Think of the complete works of an author or the texts that make up a genre, once there has been consensus on their definition and scope. It is essential, therefore, for there to be a specific adjustment of the statistical and projective models typical of the usual mining of big data, paying attention to the qualitative that resides in the singularity. As with human beings, texts, no matter how much they are digitized and added to databases, should not have their distinctive traits annulled – or if at all, only in a methodological and functional way, and this journey requires an end as much as a starting point. And at both points we must find the text, that small redoubt of reality.

Coda

Neither apocalyptic nor integrated – we must be travellers who should always be somewhat suspicious in order to stay alert; we survey a battlefield with unequal forces, the little philological David and the giant Goliath of computing. In terms of epic battles, at Little Big Horn Colonel Custer's cavalry perished with their boots on. Epic propaganda refused to let the example given by the military disaster show its tragic importance. It was so not because of the quantity of men lost, but because of a prestigious officer's sin of hubris, which led him to enter unfamiliar

territory without sufficient precaution. Today, the role of the warriors of Crazy Horse could be undertaken by the scientists of the *quadrivium*, the tribes of the number, for their ability to lay waste. But we can and should turn this around, and, if big data advances with the capability of an army and with its uniformity, we can reclaim the value of savagery and seize a space of our own. Like the poem, small and singular,⁹ we can start to operate in a kind of *little big data*, almost craftlike, humanizing the concept without rejecting the tools that form it, taking possession of the tools of big data and adjusting them to the dimension of our needs, which is also a way of maintaining consideration for the singular nature of our purpose.

Bibliography

- Cavallo, Guglielmo (1975): *Libri, Editori e pubblico nel Mondo antico: Guida storica e critica*. Laterza: Roma.
- González, Déborah and Helena Bermúdez Sabel (eds. (2019): *Humanidades Digitales: Miradas hacia la Edad Media*. Berlin/Boston: De Gruyter.
- Jiménez Heffernan, Julián (2002): “Pequeño, claro y libre: una poética para el poema lírico en la España del siglo XVI”, in *Studi Ispanici*, 5, pp. 61–79.
- Moretti, Franco (2013): *Distant reading*. London: Verso.
- Morras, María, and Antonio Rojas Castro (eds.) (2015): *Humanidades digitales y literaturas hispánicas*, in *Ínsula*, 822.
- Petrucchi, Armando (ed.) (1979): *Libri, scrittura e pubblico nel Rinascimento. Guida storica e critica*. Roma: Laterza.
- Ruiz Pérez, Pedro (coord.) (2010): *El parnaso versificado. La construcción de la república de los poetas en los siglos de Oro*. Madrid: Abada.

⁹ This is the condition given it by the first Renaissance precepts on the genre (Jiménez Heffernan 2002), and it still survives in an underground way within the panorama of literature dominated by greater-length genres.