

Designing databases for historical research With special reference to Fichoz

Jean Pierre Dedieu, Directeur de recherche CNRS (émérite), Framespa (Université Toulouse Jean Jaurés) / IAO (ENS Lyon, Université de Lyon)

Fichoz is a database for historical research which has been working for almost thirty years. First elaborated by the PAPE group (Personal Administrativo y Político de España) at the Maison des Pays Ibériques in Bordeaux for the study of the agents of the Spanish Monarchy in the XVIIIth century (1988-2005), it later extended its scope to a new set of problems when its supervisor, the author of the present paper, was assigned to the LARHRA (Laboratoire de Recherche Historique Rhône Alpes), in Lyon (2005-2013). It reached its full development when it had to cope with entirely new questions and sources in the IAO (Institut de l'Asie Orientale), a joint institute of the CNRS and of the University of Lyon, located at the ENS Lyon, which presently hosts the database¹.

We are not interested here in the data Fichoz holds, but in the way the database itself has been structured and the model of integrated data processing it proposes which, as far as we know, makes it unique in its kind².

Fichoz was consciously planned for historical research and researchers, not for the needs of general public. Its purpose is to provide ready-for-use historical data, which can be mustered without any fundamental structural transformation to answer the needs of any research. Its underlying philosophy does not consist in elaborating different solutions for different problems and particular cases, but to handle every possible kind of historical data inside a unique space; not to store pieces of information in different compartments organized in different ways depending on their nature, but to reduce, as far as possible, all information items to a same structure so as to store them together and make them all equally accessible within a same boundless space; and allow in this way an immediate extraction and assembling from the database of any subset of items, composite as far as their nature is different, but homogeneous as far as form is concerned, without changing, at that stage, their form and structure.

Building a unique global model which account for every possible information and allow a global handling of the same demanded an in-depth analyse of the structure of historical information, to unveil its most intimate fabric and find a common factor. We reached a solution which does not match day-to-day pedestrian intuition and may look unnecessarily complex. It just reflects the *de facto* complexity of social facts: men historically delighted in creating an astonishing variety of social relationships with other men, animals, objects and abstract beings such as God, saints, prophets and heroes. It

1 The author is Fichoz's main manager and an associate scholar of the IAO (jean-pierre.dedieu@ens-lyon.fr). Huma-num, a CNRS organization for e-documentation, hosts and puts on-line the database proper, under the supervision of Gérald Foliot (gerald.foliot@huma-num.fr).

2 Most works on e-humanities surprisingly do not mention databases, and less still databases of the kind we suggest. One of the most astonishing examples in recent literature: Burdick (Anne), Drucker (Johanna), Lunefeld (Peter), Presner (Todd), Schapp (Jeffrey), *Digital_humanities*, Cambridge (Massachusetts), Massachusetts Institute of Technology, 2012, 141 p. I confess I simply do not understand the purpose of works which ignore such a basic side of the question.

is the social scientist's duty wholly to account for all that. The necessity of reducing this complexity to a unique conceptual frame makes necessary a degree of abstraction, the same way as an algebra is abstract. Fichoz is not a natural language, and operators need specific training, not only in basic computing operations, but also in the management of underlying concepts and assumptions, and in the operation of reducing historical information to a common module, what we call atomised historical data. This is the price for efficiency. The fact that, without being aware that analogous researches were conducted in other quarters, we reached conclusions similar to those of social scientists working in other fields, such as Latour's sociology³, is a strong argument in favour of our approach.

Fichoz does not mirror raw documents. It stores data. Data, in our view, mean ready-for-use pre-formatted pieces of information, equipped with the necessary "handles" to be handled, that is to be selected on demand and directly inserted into demonstrative data sets. The first points of this paper (chap. 1 to 5) address the conversion of historical information into data. In many respects this operation is not specific of databases, but common practice of any kind of historical research. We largely draw on a consensus existing among historians as to the rules which govern this process. Moreover, Fichoz does not provide narratives, but sets of unconnected data which researchers have to organize into a coherent account to make them intelligible. Fichoz users must be able to carry out such a narrative-building under their own responsibility. They must be able to assume controlled creative freedom in lieu of reproducing blindly the narrative frame sources provide. Breaking the interpretative gangue in which sources transmit data is, in fact, the only way of making data free, to be included into alternative narrative sets, in search of the most coherent one from the point of view of the problem under discussion.

By handling 'atomised' data pieces, all of them cut after the same model, Fichoz is able to manage them by means of a fairly simple database model, which we describe under the name of Fichoz core (Points 6 and 7).

Data atomisation frequently means handling huge quantities of complex raw data. In some cases complexity reaches such a degree as to make them unmanageable without the help of specific processing devices. For every class of information of this kind, we build, when necessary, a peripheral device itself devised after Fichoz's assumptions, which we link to Fichoz core. We input there raw data and get an output of atomised data which we then pass to the core for processing. At the other end of the process, the operation of restructuring atomized data drawn from the database into a coherent narrative must be regulated by contextual pieces of knowledge about the historical universe concerned. Such information we store into specific peripheral parts of the database, and we inject into the output of the database to produce the desired results (Point 8). At this point, data must be characterized, that is made significant for the problem the researcher has in mind. The potential dimensions they contain must be made explicit by means of their assignment to significant classes. Atomized and characterise data must then be exported from the database for publication, in most cases after being previously analysed with the help of specific data processing tools (statistical packages, network building packages, etc.). Fichoz users must also master these data processing packages. We comment this topic in Point 10.

Point 11 underscores a too often neglected requisite, which Fichoz ambition and complexity makes more necessary than ever, the ergonomics of the

3 Latour (Bruno), *Changer la société. Refaire la sociologie*, Paris, La Découverte, 2006, 401 p.; Latour (Bruno), *Reassembling the social. An introduction to actor-network-theory*, Oxford, Oxford University Press, 2005, X + 301 p.

database.

Conclusive remarks draw some implications of introducing this kind of databases into the process of historical research, and assess the price to be paid for such changes compared to the benefits to be expected⁴.

(1) Sources do not deliver historical data

Historians extract historically valid knowledge from sources. Historical sources are any object, of whatever kind and species, which provides informations as to past states of the universe or of any of its components.

Never was a source produced to answer the historian's demands. Sources produced by human beings always aimed at fulfilling the needs of their makers, and such needs were not the historian's. Carolingian monastic charters were legal proofs of ownership. Monastic chronicles aimed at shaping a community's memory. Louis XIV did not build Versailles to provide subjects for dissertations, nor to account for French contemporaneous political system, but to change that system by magnifying the king's part. Memoirs left by important politicians with the explicit purpose of helping to write history, usually coat an ingenuous self-glorification with a layer of supposedly rare information acquired in the author's exalted functions, in the hope that historians would inconsiderately swallow the former for the sake of the latter. As for pieces of furniture, amphora, coins, archaeological remains, jewels, philosophical writings, novels, sermons, poems, diaries, accounts, hedges, dendrochronological circles, fiscal listings, cadasters, sketches of longitudinal sections of roads, estimates for building projects, pay slips, frescoes and grafitti, their makers - Lady Nature for those which are not artefacts - would have been quite surprised to see them endowed of any other value than the purely instrumental or aesthetic one they themselves gave them.

So that:

- a) Every preserved remain of the past is a source for the historian, which opens an infinite span of potentialities.
- b) The utility which their makers and past users derived from the objects we manage as historical sources, was fundamentally practical. Their function never was to provide historical data. No raw source ever delivers by itself historical knowledge. Historians must extract this knowledge from the source.
- c) The content and structure of the source is determined by the function with which his maker endows it. Historical data, that is pieces of information from which historians build historical knowledge, are obviously contained in the source; but the source supplies them coated in a narrative gangue which polarizes them to the point of sometimes making them unrecognizable. Historians must strip them of this gangue and

⁴ Fichoz database can be freely accessed under FileMaker (v. 12 and further), by means of the Remote access command to the address actoz.db.huma-num.fr. More information available on demand from jean-pierre.dedieu@ens-lyon.fr. We strongly recommend readers to have a look at it in order to better understand this paper. The Help file develops most of the concepts we expose here.

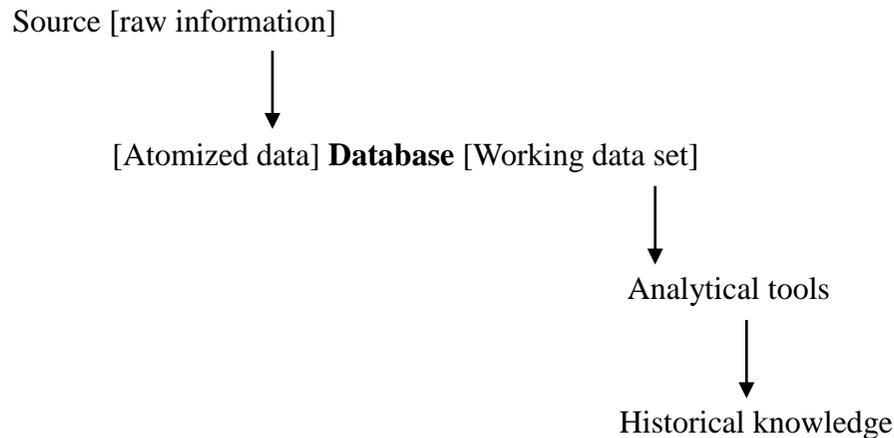
restore as far as possible their original shape and properties, to include them in new specifically oriented narratives they build in function of scientific criteria; and as far as possible they must extract and preserve the formatting gangue which, as a construction of past actors, is in itself historical information.

(2) Extracting knowledge: historical hermeneutics and historical conclusions.

Extracting historical knowledge from sources is a twofold operation. The first stage consists in extracting data from the information provided by the source and in shaping this information in accordance to rules commonly accepted by the historians' community in such a way as to make it serviceable for the second stage. This is what we call data atomization, a point we further explain in the next paragraph. The second stage consists in elaborating historical conclusions. This second stage entirely depends on the historian's personal creativity, guided and contained by the general rules of scientific methodology. Based upon data provided by the first stage, researchers try and answer a set of questions, test and validate a set of hypothesis previously designed as possible models to build a new narrative, in order to restore a degree coherency at least equal to that of the previously scientific accepted narrative, which the data provided by the database seriously challenged. The new narrative must obviously take all available data into account.

A research database, in our view, articulates both stages. It helps elaborating and stores data prepared in the first stage in view of the second one; the biggest possible volume of relevant data, and of the most varied kind possible, so as to challenge as strongly as possible the previous narrative as well as to make the new working hypothesis as embracing as can be.

Fig. 01: The database as part of the research process



An efficient database must meet various criteria. A variety of points of view in its constitution is an important factor, probably more important than sheer volume.

Example:

An impacting experience we had at the time of our first trials with Fichoz. The first batch of data we processed were administrative careers of agents of the Spanish Monarchy at the end of the XVIIIth century. They were the basis of a research essay by a student of ours. A member of the dissertation committee was François Lopez, a famous specialist of Spanish literature. He discovered, to his astonishment, that most of the poets he studied, who so earnestly praised nature, love and good spirits, were in fact bureaucrats who had spent most of their life in gloomy secretarial offices. We were surprised for our part by the artistic side of people whom official files described as merely practical personalities. Obviously, future views of both parties on these agents changed in depth.

The database must also provide an even access to all data and leave none out of scope on technical grounds.

Example:

A research group had created an interesting database on colonial civil servants of a country the name of which does not matter. They took great care to input all available data about family connexions. They succeeded in including vertical relationships (fathers, sons, grandfathers and the like) into the same structure as all other data (births, positions, deaths and so on). But not so the horizontal family relationships (brothers, spouses, brothers in law and so on), which they stored into a Miscellaneous field which mentioned the fact without providing a direct access to the same. Their conclusion was that vertical relationships explained much, horizontal ones little. Which went against all previous observations and which, for many strong reasons, looks unlikely. The fact that both sets of information were not equally accessible had probably much to do with this surprising result.

The database must also provide visual access to huge global sets of similar data displayed together on the screen. A tendency to provide access to database entries one by one, as individual separate items, is in our view one of the most frustrating features in current database building. This way of doing things may be unavoidable on small e-phones screens; it even may work when processing present day administrative matters, that is a limited set of perfectly identified facts and actors, built to answer a limited set of questions and leaving aside all extra information. It cannot work when managing fuzzy historical information to answer, as is always the case when research is concerned, an unlimited and unpredictable set of questions. The only practical way to retrieve efficient information from such databases consists in displaying all entries which may possibly have something to do with the topic and let the researcher pick what looks relevant, thus creating a working set on the basis of a common relation with the problem, to extract from the same, on visual inspection, what he needs to feed the research process. This need of global displays makes severe demands on users, and set limits to the use of the database. A large screen is in practice necessary; which precludes the use of portable devices smaller than a good-size laptop, except for the most basic queries. Even when all material requirements have been met, the sheer bulk of displayed data demands a high degree of expertise to locate the relevant bits and mark them for further use. We are confronted here with intrinsic characters of scientific information. Researchers need to be able to move among their data with the lowest possible degree of friction, without being hindered by previous arrangements. This liberty has a price: the lack of internal borders. Borders hinder movement, but they also provide guidelines. Free access means that the information which data are

carrying along must be structured by the researcher's mind at the time he is acquiring it. Only advanced professional training allows to do that⁵.

(3) Data Atomization

Reducing the cognitive load on the researcher's mind is one of the main functions of the database. How to achieve it? Reducing the complexity of contents would contradict the scientific purposes of the instrument. The only other possible way consists in standardizing the form so as to make it perfectly predictable, and to allow users to concentrate on contents. Data must be formatted in the most uniform possible way all over the database. The problem is that standardization means a loss of information, and that such a loss is unacceptable. This point was, from the beginning, one of the main checks on the development of historical research databases. Either you standardize in a minimal way, and you provide in fact roughly ordered raw sources which cannot be directly marshalled to extract historical conclusions, and keep short of the requirements we expect databases to fulfil; either you categorize every part the available information - by storing it into specific fields or labelling it - and you run the risk of getting so complex a universe of markers as to turn your files unmanageable⁶.

To solve this dilemma we had to uncover a single common factor which would account for the largest possible part of our information. We found it in the concept of action. Most of historical basic information can be described as sets of actions carried out by actors. Any historical "event" can be split into a set of individual actions, linked to one another by the relationships generated by the event.

The concept did not raise any issue as far as individuals actors were concerned. One is born, one marries, one dies, one gets a job, later another, one raises a child, another child, becomes a member of a religious community, the partner of a firm, expresses his views on various topics, builds a house, makes friend with somebody else, buys a car, writes a dissertation, and so on. Every lifecourse⁷ can be "atomized" in this way into a chronologically ordered string of actions. We quickly grew aware that many of these actions consisted in creating a relationship between couples of actors. Getting a position in a firm, for instance, means a relation with the firm. We decided to enlarge our model to include up to two actors in the same action and to describe in the relevant action entry the relationship they were creating by this action⁸. If none, to leave the second actor's position empty.

5 From a methodological point of view we fundamentally partake in the ideas of Lucien Febvre (*Combats pour l'histoire*, Paris, Armand Colin, 1953, IX + 456 p.) and Gaston Bachelard, specially his concept of spaces of configuration (*La formation de l'esprit scientifique. Contribution à une psychanalyse de la connaissance objective*, Paris, Vrin, 1977, p. 6). Simply because they match our own experience as a researcher. We reject views derived from the so-called linguistic turn. They point to obvious issues, which classical epistemology had previously perfectly identified; but contrary to previous practice they refuse to try and overcome them, thus abolishing the frontier between history (a memory constructed on the basis of specific and consciously codified rules) and sheer common-experience memory.

6 This happened to the Kleio historical database of Manfred Thaller. See: Thaller (Manfred), *Kleio. A data base system for historical research. Version 1.1.1, b-test Version*, Göttingen, Max-Planck-Institut für Geschichte, 1987, 127 p.

7 On the concept of lifecourse, an important underlying basis of our approach, see: Mortimer (Jeylan T.), Shanahan (Michael J.), ed. , *Handbook of the Life Course*, Berlin / Heidelberg, Springer, 2003, 728 p.

8 In fact, four actors. We quickly found actions in which an actor acted on behalf of another. A classical example is that of the bank manager making a loan to a shop owner. Four actors are implied: two firms and two executives acting on behalf of their respective firms.

We also grew aware that the common-sense concept of actor as individuals had to be revised. Corporations, in the example we just quoted, had the same functional status as individuals, and had to be processed in the same way, as actors. This step was easy to take: we had precedents, as law makes firms persons. The same functional point of view demanded an extension of the concept of actor to all inanimate objects and artefacts which served as anchors to relationship networks. This looked so unnatural that we were reluctant to do so, but Fichoz demanded it to process such objects in a same data space as individuals and corporations. We first decided to process in that way books, which obviously structured specific sets of relationships, had a recognizable lifecourse, exactly as living persons, from editions to quotations, from reviews to readings; books which many readers consider as something more than inert beings. Then came all other "cultural" artefacts, those endowed by public estimation of a specific value far above their intrinsic material utility. At last we also raised to the rank of actors the paraphernalia of miscellaneous properties for sale, commodities, and every kind of objects which human actors insert into relational sets, which appeared in sources as supports of the same. When extending in such a way the concept of actor, we were moved by the sheer technical necessity to make Fichoz work within its predefined limits and requisites.

We were finally able to process all the information we were getting from sources into a same simple model of action, by distributing the content of each action under five headings:

- . Who
- . When
- . Where
- . What
- . With whom

So that our database is made of records, each of them accounting for an action and composed of five fields which answer these five questions. Who, When and What are mandatory. Where remains empty when it does not make sense (being a Knight of the Bath, for instance, has no location). With whom remains empty when the action does not describe a link with some other actor. Every time one of these parameters changes, we create a new action.

Experience validated this model. It coped with surprising efficiency with any kind of data we met till now, with one exception which we shall explain in the conclusion. As we said before, we were comforted in this approach when we grew aware that French sociologists, such as Latour, were developing similar views from a more conceptual point of view⁹. The database, in its minimal form and most basic layout, looks so:

⁹ See note (2).

Fig. 02. Actions generated by Ana Pérez Camino's will, 1699, as described in Fichoz (03-23-2013)

		Otorgante testa	00098020	PEREZ CAMINO, ANA	TESTAMENTO EN SELAYA	1699-06-08
PEREZ CAMINO, ANA	00098020	Notario	00098119	MUNOZ CONCHA, JUAN	Juan Muñoz Concha, notario en Selaya, recibe el	1699-06-08
PEREZ CAMINO, ANA	00098020	Herederero universal	00098015	ARCE REBOLLAR, MATEO	Mateo Arce Rebollar es herederero universal de su madre,	1699-06-08
PEREZ CAMINO, ANA	00098020	Herederero universal	00098039	ARCE REBOLLAR, JUAN	Juan Arce Rebollar es herederero universal de su madre,	1699-06-08
PEREZ CAMINO, ANA	00098020	Herederero universal	00038334	ARCE REBOLLAR, MARIA	María Arce Rebollar es herederero universal de su madre,	1699-06-08

[With whom]

[Who]

[What] [Where]

[When]

Remark: "Where" is mentioned in the "What" field when needed. It is repeated in a specific "Where" field which we do not display in the present basic layout. "Herederero" means heir. "Testamento" means will.

(4) Handles to handle data

The next step does not raise any problem: it consists in making sure that none of the newly defined action already exists in the database.

The following one is more complicated. Each of the five headings under which we describe the action must be equipped with a set of "handles" in such a way as to make its handling easy by means of standardized procedures. Here is a list of the main ones.

- Dates are always written in the same way, eight positions in three blocks, the year coming first, yyyy=mm=dd, so as to get a chronological order by means of an alphabetical sorting order; changes in the separators (= / > / < / == / ++ / :) express relative or dubious dates (before and after such a date, between such and such a date, around such a date), so usual in historical sources. The date is compulsory and, if the source does not provide it, must be assessed in accordance to the rules of historical hermeneutics¹⁰.

- The same as dates must be set although they do not feature in the source, data concerning the What heading must be developed and made explicit. Sources leave much information implicit in the form of contextual data. If you find in the archives of the Dicken's Fellowship a "List of members", you assume it is a list of members of the Dicken's Fellowship without any more ado. If you load such data to a database of the kind we just described, you have to make explicit in every entry the name of the Club, because it is probable that you will use each entry separately and that at that time it would have to carry along all what is necessary to be correctly understood. Allusions to facts which sources do not explain because everybody was supposed at the time to know what they were, must be explained. Important steps in legal proceedings, sometimes reduced to scant barely noticeable marginal

¹⁰ See further detail at: http://actoz.db.huma-num.fr/fmi/webd#Fichoz_help; query: Fichoz date standard

mentions, because the source addressed professionals aware of legal proceedings, able to expand them to their true meaning, must be restored to their real importance and full wording. And so on. The need of such a rewriting of the source to create efficient action records is one of the main setbacks of the system: it demands from operators sufficient knowledge to be able to deploy the implicit content of the source, and at the same time to master historical hermeneutics well enough to do so in accordance to the rules the corporation established for such cases.

- Identifiers and coding fields: names of Who and With whom are set in accordance to the onomastic system in use in the group which produced the source:

Example:

Pérez Camino, Ana for a XVIIth century Spanish actor; but Muhammad bin Driss al Karaoui for a XXth century Tunisian. The first case, surname and name, European civil system; the second one: name, father, surname, Islamic system.

A degree of standardization governs the writing of personal titles, without ever making them unrecognizable:

Example:

Muhammad bin Driss al Karaoui [Si]; Pérez Camino, Ana [Doña]; Leclerc, Anne [Dame]; Téllez Girón Beaufort, María, Osuna [Duque]

A same person may feature under various names. When the source is an original document, the database always formulates the name as written in the source, given that this formulation is, in itself, an information of historical value.

Example:

Téllez Girón Beaufort, Mariano, Peñafiel [Marqués] and Téllez Girón Beaufort, Mariano, Osuna [Duque] are one and a same individual, but from a legal point of view two different persons. A same identifier, two names.

To make identification possible in spite of onomastic variations and homonyms, a personal identifier must be assigned to each actor. It can be seen at the side of the name in Fig. 02. Queries are based on this identifier. All identifiers are stored in a dictionary of actors¹¹.

Exactly as for actors' names, we add geographical identifiers to every place-name of the Where heading. These geographical identifiers remit to a geographical dictionary in which all places have been listed with the necessary basic facts for their location and identification.

A codification field also describes the What heading. It is a rather complex hierarchical coding of all the actions we found during our research. See paragraph (09) for more details.

All these coding items and identifiers are stored into independent fields, and the database works in their absence, although not to full efficiency. In such a way that it is possible to input data without immediately setting them, and to reserve this operation for a later stage; and that they can be changed, if need be, without tampering with raw data, thus preserving one of the most absolute rules of database building, that is clearly to separate raw data from coding.

11 See further, section 06. Only database managers can get a full view of this dictionary, so that it is impossible to access it by the means we mentioned in note (4).

A last important point is the existence of various program scripts which automatically execute on demand the most repetitive tasks, such as usual queries or changes of layouts, thus saving time and allowing operators to concentrate on the interpretation of data. A colour code set to the script triggers indicates their function. Far from being a decorative feature, they are an essential part in the working of the database.

(5) Grouping data

As the name tells it, atomization splits information into independent actions. The question is how to keep together sets of actions which sources define as part of a same item of information? This problem has two sides.

- First, each actor, seen as a whole or how a dictionary works. We briefly mentioned above a specific dictionary table. It holds actors' identifiers, one entry for each identifier, which is the same as one entry for each actor. Every entry of the Dictionary is linked to all database items in which an actor can be mentioned, so that from the Dictionary it is possible to call all actions in which a specific actor has been implied; and conversely to display along with the action anything mentioned in the Dictionary entry about the actor. The Dictionary entry thus allows to process automatically as a block all data referred to a same actor. Among other things (see paragraph n° 9), it allows selecting actors on the basis of details in their biography (see note 15).

- Second, each event, seen as a whole. Each action is an independent entry; but sources describe events, which are sets of connected actions. How should we preserve such connections? Exactly as for actors, through an external linked table. We add to the Actions table, which is the core of the core of our database, a table we call "Grouping" because its function is precisely to keep actions together. Every entry of the Grouping table is linked to a set of Actions records which, together, form an event described as such by a source. This entry of the Grouping table may remain empty - its sole existence is enough to fulfil its function -; or describe at length the relevant event.

Example:

We consider a legal writ as the description of a set of relations generated by the writ between all concerned parties. The Grouping entry describes the writ (date, class, summary, if needed full text). A set of entries in the Actors table stores the list of all relations thus created. Each record of the same is linked to the grouping unit. The database displays the Grouping entry, and under it all relevant actions (Fig. 03). Making changes to both tables from this same layout is quite easy.

Example:

The battle of Waterloo, if processed in the database, would be one Grouping entry, to which we should link Napoleon's decision not to attack before noon, Ney's decision to launch the charge of his heavy cavalry, Cambrone's desperate resistance after the rout of the army, Blücher's embrace to Wellington, etc. Readers should notice that a grouping entry may be indifferently shaped by a source (as in fig. 03) or by a free decision of the operator who decides to merge a block of actions into one event.

Fig. 03. Pedro Martin Cermeño's will (1787)

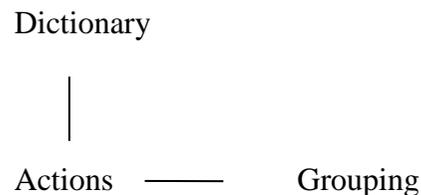
1787=02=18				Testamento		Notario - Testamento	
AMS, Martine	00000487	MARTIN CERMENO, PEDRO				La Coruña	
D0010002							
Testamento en La Coruña de Pedro Martín Cermeño, capitán general de Galicia						xxx	
						Milgrad xxx	Milgrad xxx
<p>[Preamble] "Estando como estoy sano y bueno a pie en todo mi juicio y entendimiento natural cual Nuestro Señor tuvo por bien de darme, recelándome de la muerte, que es cosa cierta y natural a toda criatura viviente, otorgo y conozco por esta presente carta que hago y ordeno en la forma siguiente: [Invocation] Ante todas las cosas creo como verdadera y firmemente he creído como bueno y fiel católico, apostólico romano, en el misterio de la Santísima individua Trinidad, Padre, Hijo y Espíritu Santo, tres personas distintas y un solo Dios verdadero, y en todo aquello que cree, enseña y confiesa la Santa Madre Iglesia Romana, y en esta fe y creencia quiero, deseo y profeso vivir y morir, tomando por mi interesora y abogada a la Inmaculada Virgen María, madre de Mi Señor Jesu-Cristo, Redemptor mío, reina de los ángeles y protectora de los mortales, para que tenga por bien ser mi interesora y abogada delante de mi Dios y Señor. Ruego igualmente a los arcángeles San Miguel y Rafael, a los santos apóstoles San Pedro y San Pablo, y a los más santos y bienaventurados de la corte celestial también lo sean, implorando desde ahora para siempre su patrocinio y amparo. Y para cuando Su Majestad quiera pase de esta vida a la eterna, continuando esta mi disposición, Mando en primer lugar mi ánima a mi soberano señor y redemptor Jesucristo, pues es suya, que la redimió con su preciosa sangre, Pasión y muerte en el santo arbol de la cruz... [Corpse] - "Quiero y es mi voluntad de a mi cadaver no se le de sepultura hasta que avise de corrupción, porque siempre tuve horror a las resultas de lo contrario por los ejemplares que he observado, encargando en este particular la consciencia a mis testamentarios, que dispondrán se verifique sin incomodidad, pues para todo hay medio - Que mi mortaja sea el pequeño uniforme y mi manto capitular de mi sagrada religión de Alcántara.</p>							
			Otorgante	00000487	MARTIN CERMENO PAREDES,	TESTAMENTO EN LA CORUNA	
			Legatario	0003877C	FAMILIA Y BIENHECHORES	Pedro Martín Cermeño Paredes lega a su familia y bienhechores se apliquen todos los sufragios que se	
			Legatario	00010398	MARTIN CERMENO, JUAN	Pedro Martín Cermeño Paredes lega a Juan Martín Cermeño se realice una misa al año, ya fundadada	
			Legatario	00038337	CORTI, MARCOS	Pedro Martín Cermeño Paredes lega a Marcos Corti, mayordomo del testador, quien fue durante muchos	
			Legatario	00008457	FERNANDEZ IDARGA, RAMON	Pedro Martín Cermeño Paredes lega a Ramón Fernández Idarga una pequeña ventera de diamantes,	
			Legatario	00098088	IDARGA, ZZ	Pedro Martín Cermeño Paredes lega a un hermana de Ramón Idarga, "se le dará por mi esposa una	
			Legatario	00000462	LOPEZ SOPENA, ANTONIO	Pedro Martín Cermeño Paredes lega a Antonio López Sopena un " estuche grande de matematicas	
			Legatario	00034780	RIAL VARELA, ANTONIO	Pedro Martín Cermeño Paredes lega a Antonio Rial Valera "la escribanía de plata de que me sirvo"	
			Legatario	00038337	CORTI, MARCOS	Pedro Martín Cermeño Paredes lega a Marcos Corti 500 escudos de vellón	
			Legatario	00098129	RADIZ MATAS, DIONISIO	Pedro Martín Cermeño Paredes lega a Dionisio Radiz Matas 100 escudos de vellón "por hallarse ya	
			Legatario	00098072	GALLAR, FRANCISCO	Pedro Martín Cermeño Paredes lega a Francisco Gallar una Ayuda de cámara de 400 escudos de vellón	
			Legatario	00000487	0000151C	LOS DEMAS CRIADOS DE LA	
			Legatario	00098058	CISNEROS ULLOA, MARIA	Pedro Martín Cermeño Paredes lega a Maria Cisneros Ulloa un usufructo de todos los bienes del testador	
			Curador ajeno	00098058	CISNEROS ULLOA, MARIA	Pedro Martín Cermeño Paredes solicita para Maria Cisneros Ulloa, esposa del otorgante y curadora de	
			Deudor	00008289	ESTHER MONTSERRAT, GABINO	Gabino Esther Montserrat, difunto comisario ordenador, debe a Pedro Martín Cermeño 3000 rv, como	
			Deudor	00034781	RIOS, JOSE	Jose Ríos debe 5000 rv a Pedro Martín Paredes, el testador, de cuando se embarcó en La Coruña para	
			Deudor	00000726	RUIZ CERMENO, JOSE	Jose Ruiz Cermeño, pariente del testador y su apoderado en Melilla para el cobro de una renta de la	
			Deudor	00098044	ARRENDATARIOS DE LA	La Encomienda de la orden de Montesa, que el reo obtiene y tiene arrendada con cobro de los	
			Deudor	00098057	CERMENO, BERNARDO	Bernardo Cermeño, pariente del testador y apoderado suyo para el cobro de los frutos de la encomienda	
			Agente negocios	00034780	RIAL VARELA, ANTONIO	Pedro Martín Cermeño tiene como agente de negocios a Antonio Rial Varela quien dará cuenta del	
			Apoderado	00098124	PILOTO, DIONISIO	Pedro Martín Cermeño tiene como apoderado a Dionisio Piloto quien fue criado de su padre y es su	
			Censualista	00098111	MILLARES, JUAN ANTONIO	Juan Antonio Millares cobra un censo recognitivo sobre la casa que tiene el reo en Barcelona	

Remarks: the Grouping entry makes the upper part of the figure. The matching Actions entries feature in the lower part.

(6) Ficho core

We are now in condition to define Ficho core, which in our view, should be the basic design of any efficient database for historical research:

Fig. 04. Ficho core



In the middle an Actions table, in which every action by any actor or pair of actors makes an entry, being the word actor taken in the broader sense we explained above; and action defined as the creation of a link between actors. This Actions table is linked to a Dictionary, in which each actor, and not each action, is an entry. Various action entries obviously match every dictionary entry. From any action entry, the dictionary entry of the corresponding actor is fully accessible; and from the dictionary entry, all actions entries in which the relevant actor takes part can also be readily accessed. Actions entries are linked, when necessary, to grouping entries which allow handling blocks of actions as if they were one object and create between all these actions a link derived from their contribution to a same event¹². This structure, coupled with the extensive conditions we formulated as to the definition of the concept of actor and with the consideration of actions as relational items linking together pairs of actors, would be, and in fact is, enough to process any kind of historical source which can be atomized into actions without noticeable loss of information.

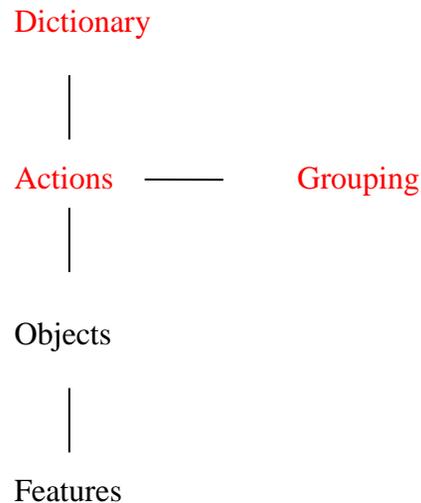
(7) Objects, an addition to Ficho core

We considered till now that an actor was described by a set of chronologically ordered actions. At any point of his lifecourse the actor is the sum of all his previous actions. We shape in this way the actor as a highly dynamic entity. Although this concept, as far as principles are concerned, holds true in any case, many sources describe in fact actors as stable entities, characterized by permanent features. A police description of a wanted person is probably the best possible example of such an object; or the characters of a house for sale described in the sale writ. Any actor can thus be described in

¹² The concept of event has been amply discussed from an epistemic point of view. We define it as any set of actions of which any observer makes a whole. Events have no proper entity; they are designed by the observer's consideration. That is, from a Ficho perspective, by the decision to create a grouping item to aggregate them as a same entity.

two ways: as an ever and fast changing entity; as a slowly changing one, which can be regarded, if change is slow enough, as practically stable. When described as a stable entity, we call it an object (in lieu of actor), described by features the properties of which condition all the actions the object carries on or suffers as long as the feature remains active. A chronological set of actions cannot easily account for objects, for if actions individually cover a little or practically null chronological span, a feature by nature embraces an appreciable time span. To do so, we add to Fichoz core two tables. The first one details the features which the source assigns to the object (one feature, one entry). The second table's only function consists in grouping all features belonging to a same object and in linking this object to the entries of the Actions table which elicited from the source the current description.

Fig. 05. Objects, as part of Fichoz core



(8) Peripheral ad hoc tables

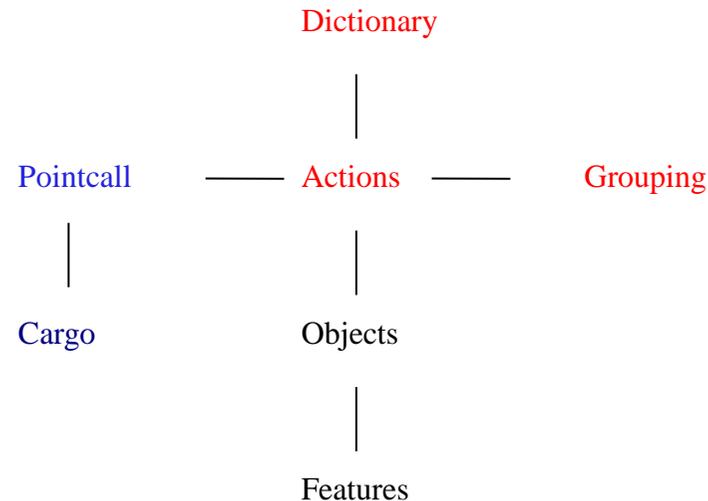
- Many sources provide an information which cannot easily be reduced to actions.

Example:

Port registers provide lists of in-coming and out-sailing ships. If they had been correctly held and preserved, every travel should be described by two entries: as an out-sailing entry in the books of the port of departure; an in-coming entry in the port of arrival. Reality is a bit more complex. Many registers are no longer available. Moreover, many others are not easy to access: while working on Bordeaux out-registers preserved in

Paris or Bordeaux archives, checking data from Bergen, London, Mumbai or Callao in-registers may be difficult. Many ships do not declare in the out-register their next call, but their final destination and the ship you are expecting in Edimburgh appears in Bergen, Trondheim and Inverness before. Some ships never reach their declared destination: don't look for the Titanic in New York in-registers... Finally, a same ship may simply change name and description from one port to another: L'Arbalète, captain Leclerc, sailing from Bordeaux to Barcelona, may appear in Barcelona in-register as La Belette, captain Claire. Navigocorpus processes ship travels as a set of chronologically-ordered crossings of geographical points, in which each crossing is considered as an action, that is a record of the database main table¹³. Before loading Navigocorpus Actions table, you must extract from raw information really sound and unique data on this basis. You need an intermediary table in which to store raw data copied from the source as they come, to subject them later in this same intermediary file to a first elaboration to rebuild identification and eliminate repeated values. In Navigocorpus we call this intermediary table Pointcall.

Fig. 06. Navigocorpus configuration



As port registers also describe cargoes, we link an extra "Cargo" table to the Pointcall table. Once you have clearly defined the Actions table as the core of your database, you can easily add peripherals of any kind. Some of them are almost systematically implemented, such as a Genealogy table, which

13 Dedieu (Jean Pierre), Marzagalli (Silvia), Pourchasse (Pierrick), Scheltens (Werner), "Navigocorpus: A Database for Shipping Information. A methodological and Technical Introduction", *International Journal of Maritime History*, 2011, XXIII, ° 2 (12/2011), p. 241-262.

makes possible processing family relationships, without any limit to their complexity; an Array table which allows storing to arrays quantitative data which appear in any kind of sources and to link them to actions. Some are more specific, such as Pointcall (Fig. 06) or Census, which allows an easy loading of data from census and taxpayers lists. This capacity of aggregating on demand auxiliary tables to the core, to create specific Fichoz implementations, allows processing highly complex sets of data without making the core itself more complex, thanks to the universal character of the categories of "action" and "actor" which support it. This universality of a simple central model is in our view the main distinctive character of Fichoz.

The convergence of all these tables toward the Actions table, the core of the core of the system, as a consequence of the set of links which link the different tables, in conjunction with the uniqueness of the identifier which identifies each actor all over the database, warrants the continuity of the data universe, even when extra tables have been added. It is always possible to reach one table from another, to link one table to another, although they stay far away in the linking diagram, to pass data from one to another and query data in various tables at the same time. So that the database always complies with the conditions of continuity which we defined above as one of our main requisites.

(9) Making sense of the data

We shall expound this problem for actors. Similar remarks and solutions can be easily extended to other objects, such as features, grouping, cargo items, points of call, etc.

The atomisation of data into an ever-similar and neutral set of fields makes the atomised item neutral too. This is a huge asset for a research database. The main benefit of Fichoz in relation to other databases is that it does not demand any previous interpretation of the data beyond the five axis we defined above (Paragraph 3). This point we consider as fundamental for two reasons:

- 1) a same data item belongs to various classes depending on the researcher's point of view; to make these classes part of the database structure means that the point of view which commanded this first inscription cannot be changed without making changes to this same database structure; which in turn means that such changes will never be made, giving this first point of view a unbeatable edge over all others; a trait that can be positively considered in an administrative universe, but which is inconsistent with research.
- 2) operators cannot assign classes to data items just on the fly, without considering quietly what they are doing. In such a way that the flow of loading material to the database would be unacceptably slowed; moreover, this immediate assignment would probably be erroneous, given that many historical documents are self-interpretative, which means that they cannot be understood piecemeal but only when considered as a whole, something impossible while they have not been fully loaded to the database, which is a blatant contradiction.

The drawback is that, beyond the neutral and elemental splitting of original data into five main fields, which allows their formal handling, but little more, all we get is an inconsistent "alphabetic soup", which makes "Conseiller de Castille" a different item from "Consejero de Castilla" and from "Conseiller du Conseil de Castille". Meaning must be introduced from outside, by researchers, on demand, in a way that will not affect the basic

structure of the database. Much is implicitly done by the reader's mind when simply reading the data¹⁴. Sticking to the reader's mind would nevertheless mean limiting the database potentiality to the cognitive capacity of the researcher, a short benefit over the pen and paper technology. Fichoz goes beyond, and accepts two kinds of meaning-making labels, to be injected by researchers. None of them changes anything either to the letter of the data or to the structure of the database, as they are stored in specific independent fields, the place of which has been marked from the beginning.

The first of them is the on-the-fly label. To put it, select first by formal queries on alphabetical strings all the entries of the database which denote that the actor concerned possesses such and such a feature, which make him a member of a specific class. Then input a marker into the on-the-fly marker field. As this marker field is not part of the Actors table, but of the Dictionary table (see paragraph (4)), it is automatically displayed on every entry of any table in which the actor is mentioned. This process makes the actor significant for analysis. As the on-the-fly marking process is very fast and flexible, marker can be changed at any moment, if needed, or preserved on the long term. As many markers as desired can be aggregated, one after the other, allowing the actor to be seen as a multidimensional object. Each marker can even be marked as belonging to such or such a researcher, thus making possible collective work¹⁵.

The second marker is the permanent coding field. It stores a label which characterises the action not from the point of view of the researcher, but from the point of view of the actor, as he himself and his contemporaries saw things. This complex marker inserts the action within the context of the time it happened. Some example will make things clearer¹⁶:

Example: birth VNxxx-Nxxxxx-xx, in which VN defines an event of natural law, -Nx a birth
Example: illegitimate birth VNxxx-NIxxxx-xx
Example: appointment as a councillor of Castile: FFEAx-AKxxxD-xx, in which the first "F" means Old regime, the second "F" royal institution, EA the Spanish Monarchy, -A a Council, AK the Council of Castille, xD a position as councillor.

14 An observation which, by the way, makes a strong argument against any inconsiderate publication of scientific databases.

15 This is what makes possible to retrieve actors on the basis of details in their biography, a problem we mentioned above (Paragraph 5). We describe actors as chronologically ordered sets of actions. Their lifecourse can easily be retrieved through a query on their identifier. But creating for every action a different record makes impossible to intersect the content of various records to locate individuals characterized by various given actions.

Example: creating a list of Councilors of Castille born in Madrid demands a selection of all action records which mention appointments to the Council; but such records do not mention birthplaces. To get information on this point, we must select all action records which mention people born in Madrid. They do not hold any information as to future appointments. Obtaining an intersection is impossible.

The "on-the-fly" coding field of the Dictionary table provides a solution. We mark in this field all actors corresponding to the first criteria; then all actors corresponding to the second one, and so on with as many criteria as we chose. All markers being stored in the same field, intersections can easily be calculated.

Example: in the example just above, we should first query all action records mentioning appointments to the Council; mark all relevant entries of the Actors dictionary with, let us say, the characters string "CCast" (a short program script makes it easy; the choice of the marking string is free). Then, we should query all records mentioning births in Madrid, and mark the dictionary entries with the string "BMadrid". All dictionary entries holding at the same time "BMadrid" and "CCast" would point to Councilors of Castile born in Madrid.

16 "x" marks empty positions

Example: appointment as a president of the Council of Castile, declined by the nominee: FFEAx-AKxxxA-xF

(10) Exporting data to working data sets

We defined the database as a mere intermediary tool between raw information and a final analysis of data to contribute new historical knowledge. Extracting data from the database to execute this last stage is as important as inputting and storing them.

The first and most usual way of exporting data is visual inspection, a point too many databases ignore or neglect. That means that the database must graphically be constructed in such a way as to make easy visual inspection and selection of relevant data. Colours and visual effects have a very practical function, far above their aesthetic side: they highly contribute to the global efficiency of the system.

Far less usual, although strategically all-important, is the export of tabulated data¹⁷ to other packages. In fact, visual inspection is limited to cases in which data are simple enough. Above a minimal degree of complexity, they must be processed by specific analysis packages. The exportation itself is easy if the analysis package accepts lists of tabulated data without any specific formatting. It must be necessary to transform data by means of an intermediary file if the analysis package demands special formatting, such as Pajek, a much-used application for network analysis¹⁸.

Anyway, whatever be the exporting strategy, a previous step consists in selecting, inside the database, the relevant data to be exported. The selection process must be able to aggregate inhomogeneous items selected through different queries from different tables of the database. The database itself must have been designed to allow such aggregations. In Fichoz, the convergence of all tables on the Actions table and the use of Dictionaries, as explained in Paragraph (5/1), provide an efficient tool for such a task.

- Transforming information into data (Paragraphs 2 and 3), as well as selecting data for exportation, means analysing and interpreting them before; that is drawing information from sources such as institutional dictionaries, catalogues of sources, chronologies, or other sources independent from the data and often difficult to access in public libraries and archives. Fichoz stores such information into special permanent auxiliary files, the most important of which is a huge institutional dictionary which operators feed with all kind of information they get during their research for their personal benefit and for that of others. The entries of the institutional dictionary can be linked to the actions they contribute to explain¹⁹.

17 We do not consider here the proper way in which the exportation is carried on, either as .tab files or .csv files. This point is practically irrelevant as far as databases are concerned.

18 De Nooy (Wouter), Mrvar (Andrej) et Batagelj (Vladimir), *Exploratory social analysis with Pajek*, Cambridge, Cambridge UP, 2005 [2002], *Structural Analysis in the Social Sciences*, 27, 334 p.

19 Fichoz obviously comprises a table for the description of sources, the content of which can be accessed from any other table.

(11) Ergonomics first

We constantly alluded in this paper to visual inspection of the data. We insisted on the fact that selecting data means visual choice among huge pre-selected sets. We insisted on the fact that most data exportation is in fact visual exportation, without any computing intermediary. We shall underline in our conclusion the benefits of huge databases, which mean that enormous sets of data may be displayed²⁰. We must remind, on the other side, that the main function of a database consists in removing cognitive load from the operators' mind, and that visual comfort and clarity is a fundamental part of this function. All that makes absolutely necessary the ergonomics of the database to be perfect. The underlying package must obviously provide dynamic fields and Wysiwyg technology, tooltips and the like²¹. Screen layouts must be carefully designed. Colours must be used as a code to point similar parts all over the database. Users must work on professional computers, with broad high definition screens. If the access to the database is through the Net, the network must be efficient. We shall develop this last point in our conclusion. Such requisites, let us insist, are not options. They are obligations.

The structure itself of the database must be arranged so as to make markers and identifiers unnecessary for a minimal working of the same. Inputting actor identifiers and identifying places can take a long time. Doing a full dressing of the data every time you create or change a record is awfully time-consuming²², apart from the fact that it obliges operators to take premature decisions as to the meaning of the data themselves; decisions which it would be far better to postpone until acquiring a better global view of the information concerned. Our model in fact means a progressive reshaping of raw information into data. That means a permanent dialogue with the database, hundreds of queries and of small changes. A high quality of the operator's interaction with the database and of the ergonomics of the same is a basic requirement, and one of the main bottlenecks we identified in practice.

Conclusion. Benefits and drawbacks

Hundreds of publications have been based on Fichoz in the last twenty years, most of them, but not exclusively, about Spanish early modern history. Some of them were important contributions to a change of paradigm in that field²³. The impact of the database on these publications is not always immediately apparent to the reader. Fichoz is not an analytical tool placed at the end of the research process which, for coming last, conditions the

20 Fichoz implementation on political actors of the Spanish Monarchy in the XVIIIth and XIXth centuries currently holds more than 0,55M records in the sole Actions table.

21 Which, by the way, bars, at least for the time being, some of the most usual database packages, among others Access.

22 This was a main drawback of an alternative database model developed from 2008 onwards in our own laboratory, the LARHRA (Laboratoire de Recherche Historique Rhone Alpes), called SyMoGHIH: Beretta (Francesco), Vernus (Pierre), "Le projet SyMoGIH et la modélisation de l'information", *Les Carnets du LARHRA*, 2012, 1 - p. 81-108.

23 Among others: Bertrand (Michel), *Grandeur et misère de l'office: les officiers de finances de Nouvelle-Espagne, XVIe-XVIIIe siècles*, Paris, Publications de la Sorbonne, 1999, 460 p.; Andújar Castillo (Francisco), *El sonido del dinero. Monarquía, ejército y venalidad en la España del siglo XVIII*, Madrid, Marcial Pons, 2004, 486 p.; Enriquez Agrazar (Lucrecia), *De colonial a nacional: la carrera eclesiástica del clero secular chileno entre 1650 y 1810*, Méjico, Institutio Panamericano de Geografía e Historia, 2006, 364 p.; Artola Renedo (Andoni), *De Madrid a Roma. La fidelidad del episcopado en España (1760-1833)*, Somonte, Trea / Universidad País Vasco, 2013, 383 p.

formulation of results and makes itself conspicuous. Everything Fichoz does could have been done manually and was in fact manually done before. The difference lies in the bulk of the data Fichoz allows to manage at a same time, far beyond the capacity of unaided human cognitive capacities (note 23, Enriquez, Andújar), in the variety of points of view the sheer presence of unexpected data left by other users in the database forces upon the researcher (note 23, Bertrand), the possibility to undertake experimental inroads into unexplored documentation which the slowness of manual research processes would have made impossible (note 23, Artola). All that breaks restrictions which the limited cognitive capacity of researchers imposed on historical research, and changes in depth the way of making history. Nevertheless, the database does not claim protagonism in the published results. Fichoz is a use-and-forget tool; and is so because it does not distort data nor force upon researchers any specific point of view. It just makes easier compliance with what our elders, at the end of the XIXth century, established as scientific norms for historical research.

Fichoz-model databases introduce another fundamental innovation. Loading data to this kind of repository is a slow task. Faster - or so fast - than taking pen-and-paper notes, but more energy-consuming because of the necessity to fully convert information into data on input, with no possibility to leave aside for a better occasion the tougher pieces, as the pen-and-paper approach allowed. The benefit lays in the fact that once fully inserted into a uniform and known structure, data become accessible to a whole scientific community, while hand notes were, with very few exceptions, notoriously of no use to any but their author. Data being all formalized in the same way all over the database, a data created as part of a specific research program can be recycled as part of another researcher's work on another topic; which means that one of the main obstacles on the way to collective research has been removed. The more so because the requirements of an atomization which all users must execute in the same way, and the use of the same analytical tools at the end of the process, force upon researchers the need to share experience; and to share experience they must clearly explain their purpose to others, describe their tools and their hypothesis, thus creating the conditions for a true collaboration far beyond the technical side of the business which triggered the first contact. A database of this kind is a magnificent tool for generating collective work. We suspected this property when we began Fichoz. Experience made clear that our most fantastic hopes in this respect felt short of reality.

Experience also made apparent two bottlenecks which we simply did not imagine at first, another we expected, and a structural limit we fully assume. The first stumbling block is an obvious difficulty in finding operators with a full capacity to atomize data. The problem does not derive from a supposed complexity of digital objects. We identified as the main drawback a difficulty to truly understand the source, to fully deploy its meaning and making transparent its structure. These abilities are supposed to be part of the historian's current training. They are not, in fact, and we developed a painful sensation that this evil is getting worse every day. Digital databases contribute to make it more apparent, because of their more demanding requirements, and because they oblige operators to execute two tasks at the same time: interpretation and input. This is by far the most worrying point.

The second unexpected drawback is the poor quality of the international Net links. To make people work together, that is to take advantage of one of Fichoz's most interesting features, a unique on-line database is practically necessary. Databases, and specially databases of the Fichoz-model, exchange huge volumes of data when working on-line. We were happily surprised by the quality of some networks: the French scientific network, as implemented in L'Ecole Normale Supérieure de Lyon, the Italian one as implemented in the Scuola Normale Superiore di Pisa, even current public networks in France or Argentina, although they do not reach the level of the former. We were strongly disappointed and worried by the poor quality of network links in famous French research institutions, such as the Ecole Française de Rome or the Casa de Velazquez. Less surprisingly we had problems to work from Chile or Mexico, and far more from China. And last but not least, access to a Lyon based database from French Antilla was

simply unserviceable. One may hope that these problems will disappear in a near future. For the time being, they make impossible any kind of true cooperation. The managers of many research institution do not look fully aware of the problem.

The drawback we expected is the ignorance of historians as to digital tools. We were nevertheless surprised by the fact that such an ignorance did not affect so much their capacity to handle the database itself; rather to handle the analytical tools downstream and the process of exportation of data to the same. We believed Excel to be shared common knowledge. It is not. We were equally surprised by the fact that young generations do not do significantly better in that respect than older ones. The heart of the matter, in our view, is that analytical tools for research demand a global view of the question at stake, of the possible alternative solutions digital tools offer and the elaboration of an implicit algorithm on how to cross from data to conclusions; while commercial e-tools offer ready-for-use and just-click-and-stop-worrying solutions, quite efficient, but closely limited to a short range of known demands. They do not prepare to an open management of digital data and applications. Just the contrary: they create a false sense of comfort, the idea that digital tools take automatically charge of all demands. Precisely what does not work and will never work as far as research is concerned.

The structural limit comes from the fact that Fichoz splits sources into 'atomized' information pieces. The source in its materiality disappears. Four mentions of a same appointment in four different sources, for instance, become one and only one action entry. The first pages of Chateaubriand's *De Buonaparte et des Bourbons*²⁴, reduced to a Fichoz actions set, would only inform us that Chateaubriand strongly and publicly disapproved of Napoleon in 1814. The amazing literary quality of the text, which gave it an impact on public opinion and made it an important factor of Bourbon's restoration, would be lost. An entry of the Grouping subsystem in which the original text would have been copied, could of course partially make up for this loss, but only partially. When the form of the source itself carries specific information Fichoz grows inefficient²⁵. This kind of sources must be processed in a different way, with the help of specific content-analysis packages²⁶; the results of which can be later re-injected into Fichoz as feature by way of the Object subsystem.

24 See the English translation published in *The pamphleteer*, London, by Abraham John Valpy, 1814, vol. III, n° V, p. 435-436 (available in Google Books, March 2012), which gives a fair account of the literary magnificence of the text.

25 Specific information, not generic information. The form of a legal writ is all-important, but for generic reasons: all writs of the same class have a similar form, and this form does not add anything to the information conveyed by the specific documents. It only backs its legal validity, a fundamental point from the legal point of view, but only a basic and generic assumption for the historian.

26 We personally use with full satisfaction NVivo, by QSR International. See: <http://www.qsrinternational.com>, the most usual package of the kind among social scientists.

Historical research and database: a global view

