Knowledge Acquisition at the Time of Big Data

Francis Rousseaux, Stéphane Cormier
Reims Champagne Ardenne University, France
Email: {first name.name}@univ-reims.fr

Abstract—**What is exactly 'Big Data', and for what purpose and application is it really efficient? Between the commercial promises made by the industrial actors and the Cassandra's cautions from some whistle-blowers, we propose a singular Big Data field to investigate with Inductive Data-Driven Algorithms: developing collections. Last but not least, we investigate the innovative possibility to curate 'figural' collections, characterized by Jean Piaget as follows:** *"A figural collection composes a figure, through the spatial relationships between its elements, whereas non-figural collections and classes are free of any figure".* **Thus, we incidentally disclose some important Abstract Truth of Big Data.**

Index Term—**predictive analysis; inductive data-driven algorithms; big data; figural collections; digital societies; search by similarity.**

I. INTRODUCTION

THE classical processes of Information Technology (IT) – including Artificial Intelligence approaches – were traditionally based on deductive knowledge modelling and simulation, so that explanation and theory, based on validations or refutations [13], were never far from engineering.

Things turn different within Inductive Data-Driven Algorithms and Big Data, because practitioners do not need anymore domain theories and explanation-based processes to succeed in providing useful results.

Those new features and paradigmatic evolutions introduce some important epistemological breakthrough in the IT classical environment, with important consequences.

At least, our digital societies are about to revisit an old utopia a human dream, close to our day live: the management of our overabundant and oversized collections, as lighten by contemporary art.

II. PREDICTIVE ANALYSIS THROUGH BIG DATA: A TSUNAMI IN IT

A. What changes has the arrival of Inductive Data-Driven Algorithms and Big Data made in the data ecosystem?

The title of the document is placed in appropriate frame on the top of the document and is formated with style `Title`. Do not capitalize short words (a, and, the, and so on).

In the world of computing, we are currently living through a period in which the epistemological nature of data is being thoroughly changed, as Big Data seems to be revolutionizing most of the methods and approaches of digital decision-making tools and processes. Indeed it is as if the Curse of Dimensionality and the pitfall of the Combinatorial Explosion that limited computer science in the 20th century have been left behind us. Natural allies of Data Mining and Machine Learning, the tendency is now towards Big Data and Data-Driven Intelligent Predictive Systems (DDIPS).

These systems are called Data-Driven because they mobilize Big Data to make connections or analogies, aiming to create certain configurations or to anticipate situations that might cause delays or predictive challenges. They are called Predictive because, unlike 20th century computer systems, their performances are measured more by their ability to predict or to discover rather than to understand, explain or theorize.

They are called Intelligent because they frequently make use of input from supervised or unsupervised automatic learning techniques, data mining, and even Deep Learning based on Convolutional Neural Networks, which contribute to performances that far outstrip the preceding generation's.

Because of this epistemological turnaround, and with the help of DDIPS, a flood of predictive data are being added to the usual data, although this new type of data has never been directly input or calculated by determinist or classically deductive methods [6], [19]. Thus Predictive-Data – to be distinguished henceforward from other types of data – are considered an additional raw material to be exploited, even if their reliability must be closely scrutinized.

B. What about the current positions and statements made by main actors?

IT professional and industrial companies, but also start-ups working in that domain, are putting forward on promising business and marketing opportunities, furnishing and putting tool and technical solutions on market places[1].

But in parallel observers and researchers, often coming from social sciences fields, forewarn potential users from a new calculus order that could mean their end of autonomy. To mention only a few of them:

1) The end of Theory

Following George Box saying: *"All models are wrong, and increasingly you can succeed without them",* Chris Anderson [1], after he has been the editor-in-chief of WIRED magazine, puts forward that: "The Data Deluge Makes the Scientific Method Obsolete".

He claimed that: *"Petabytes allow us to say: Correlation is enough. We can stop looking for models. We can analyse*

---

[1]See for example http://www.bigdataparis.com/guide/BD14-15_Guide_ BD_14136_2.pdf

the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot".

2) "Technological Solutionism"

In one of his recent bestseller books Evgeny Morozov [6], the Byelorussian expert analyst Evgeny Morozov explained: *"There is something, to me, that is very worrying about the idea of replacing causality with correlation, because if you do want to engage in reform, you do need to understand the causal factors that you will be reforming. If you just focus on correlations, all you'll be doing is basically adjusting the behaviour of the system without understanding the root causes that are driving it".*

And also: *"So the proliferation of big data and the ability to track things that we do is good only if we can actually understand why we engage in those behaviours. The ability to understand why I think is fundamental to understanding what it is that needs to be changed".*

3) "Algorithmic Governementality"

The European researcher Antoinette Rouvroy says [17]: "La rationalité post-moderne [engendrée par le Big Data] est fondée sur la découverte de corrélations entre des données recueillies avec des intentions et dans des contextes extrêmement divers, hétérogènes les uns aux autres, et qui ne sont reliés entre eux par aucun lien de causalité. C'est le renoncement au savoir causal, la dévaluation de l'expérience sensible elle-même au profit du calcul".

And also [16]: *"Operations of collection, processing and structuration of data for purposes of data mining and profiling, helping individuals and organizations to cope with circumstances of uncertainty or relieving them from the burden of interpreting events and taking decision in routine, trivial situations have become crucial to public and private sectors' activities in domains as various as crime prevention, health management, marketing or even entertainment. The availability of new ICT interfaces running on algorithmically produced and refined profiles, indiscriminately allowing for both personalization (and the useful, safe and comfortable immersion of users in the digital world) and preemption (rather than regulation) of individual and collective behaviours and trajectories appears providential to cope with the complexities of a world of massive flows of persons, objects and information, and to compensate for the difficulties of governing by the law in a complex, globalized world. The implicit belief accompanying the growth of 'big data' is that, provided one has access to massive amounts of raw data (and the world is actually submersed by astronomical amounts of digital data), one might become able to anticipate most phenomena (including human behaviours) of the physical and the digital worlds, thanks to relatively simple algorithms allowing, on a purely inductive statistic basis, to build models of behaviours or patterns, without having to consider either causes or intentions. I will call 'data behaviourism' this new way of producing knowledge about future preferences attitudes, behaviours or events without considering the subject's psychological motivations, speeches or narratives, but rather relying on data. The 'real time op-*

erationality' of devices functioning on such algorithmic logic spares human actors the burden and responsibility to transcribe, interpret and evaluate the events of world. It spares them the meaning-making processes of transcription or representation, institutionalization, convention and symbolization".

C. *Our point: Towards a scientific study of opportunity*

Many of our modern computerized activities, may they be personal, professional or even artistic, involve searching, classifying and browsing large numbers of digital objects.

Until recently, the usual tools we had at hand, however, were poorly adapted as they were often too formal, because the current models for information search often assume that the function and variables defining the categorization are known in advance.

In practice, however, when searching for information, experimentation plays a good part in the activity, not due to technological limits, but because the searcher does not know all the parameters of the class he wants to create. He has hints, but these evolve as he sees the results of his search. The procedure is dynamic, but not totally random.

The collector's experimentation is always carried out by placing objects in temporary and metastable space/time. Here, the intension of the future category has an extensive figure in space/time. And this system of extension gives as many ideas as it does constraints. What is remarkable is that when we collect something, we always have the choice between two systems of constraints, irreducible one to the other. This artificial 'undifferentiation' for similarity/contiguity is the only possible kind of freedom allowing us to categorize by experimentation.

Nowadays, our software design could strongly become backed up by both artistic and psychological knowledge concerning the ancient human activity of collecting, which can be described as a metaphor for categorizis in which two irreducible cognitive modes are at play: aspectual similarity and spatiotemporal proximity.

Inductive Data-Driven Algorithms and Big Data could help, definitely, to allow the creation of a new operational space, in between formal classes/categories and radical singularity.

III. THE NEED FOR COMPUTER-AIDED COLLECTIONS MANAGEMENT TOOLS

A. *An illustrative example*

Let us illustrate this situation. First, let us suggest that looking for new material and classifying are two important processes involved in collecting. Indeed, when someone decides to start building a collection he usually already possesses a few items. Then, to extend this collection, new items must be added. In order to do so, the collector goes into the world and looks for these new items. Then as the collection builds up, the need to arrange the items into categories will become clearer, as the collection cannot simply remain a messy stack of unordered items [14].

Let us describe a particular example: the music collector. This collector will surely possess some initial items; these

may be some CDs or vinyl records. His first action involved in extending his collection could be a visit to the record shop for example. Here, the music is classified conformingly to the record companies' desires, which can sometimes be confusing for our collector, who is a fan of Jimi Hendrix, and just does not know where to look for his albums: in the blues section? Rock section? Is there a 'sixties' section? Anyway, despite finding them rather practical at first sight, our collector didn't create these labels, and finds it difficult adapting to them. However, as he browses through the shop, he also notices some nicely illustrated records, and discovers new artists he is interested in because their records are sitting next to Jimi's. Finally, when he has bought enough music records, and come back home, he will be able to start arranging his collection in a very personal and satisfying manner, which will be pleasing to the eyes, and also allow him to retrieve items quickly.

If he had decided to collect digital music, and go online to find new items for his collection, the process would have been rather similar. Commercial music download sites allow the user to browse through predefined music categories, thus implementing a kind of virtual record shop with the same problems mentioned earlier. The search tool however can come in handy, and allow the user to search for the name of an artist, a song, an album or even musical genre. All these are still editorial information, which aren't necessarily the most useful to the collector. Then, when the music is downloaded, the album consists of a group of compressed audio files, containing preset meta-tags, again storing editorial information. When browsing these files in his audio player, the songs are defined and classified automatically, not always according to the collector's desires. His final attempt is then to create a set of folders on his disk, and arrange his items in these folders. But how does he name these folders? What if he wants to arrange and browse the items in multiple ways? What if a particular item doesn't fit in any folder, or could be placed in two or three different categories?

As we see from this example, the tools that the everyday user has at hand are too formal, and are poorly adapted to the growing activity of collecting multimedia contents. Indeed, what we have said for music can also be said for the other kinds of media, and can also be said for information research, file sharing, etc.

Attempts have been made at putting the human user back in control of the collecting process, rather than relying purely on predefined categories and automated research algorithms.

However, it has become obvious that the other extreme of handing complete control over to the user isn't optimal either. Let us take a look at online content sharing sites, such as the famous FlickR™. There is no categorization here, but there are three main strategies when looking for photos: date, location, and tags. The first two are self-explanatory, but the tags are more interesting here. When someone uploads a photo to the website, they can link a certain number of keywords, called tags, to this photo. Then, we can either browse through the most popular tags, or type a tag into a textbox for a more precise search. The users then have complete freedom on the way they choose to define their photos. But the problem is that many photos aren't tagged, and the photos that are, often have poorly named tags, making them difficult to retrieve. Therefore, we believe that an optimal solution to the problem of digital collections could lie somewhere between these two polarities: predefined categories and total user creativity.

B. Artists and philosophers' fascination for collection regimes

As a matter of fact, artists and philosophers have always been fascinated by the rebellious nature of collections and have demonstrated this in their own way ([2], [20]).

Here, for example, is the analysis of [20] on the status of excess in a collection: *"Excess in a collection does not mean disordered accumulation; it is a fundamental principle: for a collection to exist as such-in the collector's eyes the number of objects must exceed the physical possibilities of exposing and storing the entire collection at home. Therefore, someone who lives in a studio can have a collection: it is only necessary for him to have at least one work he cannot hang in his studio. That is why the reserve is an integral part of collections. Excess also applies to the capacity of memorization: for the collection to exist, it is necessary for the collector not to be able to remember all the works he owns. In fact, the number of objects he owns must be so important that it becomes too important, so that the collector can forget one of them or leave a part of his collection outside of his home. To sayd it ifferently, for a collection to exist, the collector must not have full control over his collection anymore"*.

And also: *"The scene of a collector is not his own apartment, it's the world. The main part of his collection is not at his place — his collection is to be, still scattered across the world, and every gallery and every fair is a way for him to go and find his future collection"*

As far as [2] was concerned, he had a very personal view on the subject: *"The art of collecting is a form of practical recollection, and, of all the profane manifestations of proximity, it is the most convincing. Everything that is present to memory, to thought, to consciousness, becomes a base, a frame, a pedestal, a casket for the object possessed"*.

And also: *"What is decisive, in the art of collecting, is to free each object from its primitive functions, in order to establish a relationship as close as possible with similar objects. This relationship is diametrically opposed to usefulness, and belongs to the remarkable category of completeness. What is that completeness? An imposing attempt to go beyond the absolutely irrational nature of the simple presence of the object in the world, by integrating it in a new historical system, especially created, that is the collection"*.

Thus, collections strongly differ from class, series, set, group, heaps, cluster, juxtaposition, accumulation, but also from organic whole/family.

C. Collections, between order and disorder

Until recently, a trend was mobilizing computers for the organization of our collections, considered like a group of

objects waiting to be organized in ad hoc classes that must be created simultaneously ([4], [9], [10]).

Because our collections seem to be nearer to order than disorder, attempting to assimilate them in classes is not so surprising. At least, collections look like they are waiting for their completion within a classification order, with the aim of turning into canonic achieved structures made of objects and classes. But something is also resisting this assimilation, as artists and philosophers have always noticed.

Undoubtedly impressed by artists and philosophers who considered the strange status of collections, computer program designers understood that computer modeling of object collections would necessarily involve the creation of hybrid structures including private characteristics – by which the collected objects are usually referred to – but also including characteristics that come from the activities in which these objects collectively engage.

Often, the approach implicitly chosen to characterize a collection is parsimonious and consists of over-determining the private referencing of the collected objects through a minimal description of the collective activity's context, even if it means predicting that the collection shall become a class or set of classes.

This practice presents the unquestionable advantage of not fundamentally opposing the traditional modeling of objects. However, it does not always live up to the collectors' high standards.

Here it is important to distinguish between figural and non-figural collections. This subtle distinction, introduced in the 1970s by Piaget and his research teams of child psychologists [12], brings more light to the situation. If it is certain that (non-figural) collections that adapt well to the aforementioned parsimonious approach exist, it is because they are completely independent of their spatial configuration. In that, they are already close to classification, of which they can only envy the formal completeness. On the other hand, there are collections we can label as *figural* because both their arrangement in space and the private properties of the collected objects determine their meaning.

D. Collections versus classes

In their book *La genèse des structures logiques élémentaires,* Jean Piaget and Bärbel Inhelder ([11], page 25) provide a precise distinction between figural and non-figural collections, which are still called classes or categorical collections. For the authors, a class requires only two categories or relationships, both necessary and sufficient, for its actual definition as a class:

1. The qualities common to its members and to those of the classes it belongs to, as well as the specific differences that distinguish its own members from the members of other classes (comprehension);

2. The relationship of a part to the whole (membership and inclusion) determined by the quantifiers "all", "some" (including "one") and "none" applied to the members of the class in question and to other members of the classes it belongs to, defined as extensions of that class.

For example, cats share in common several qualities owned by all cats, some being specific and some others belonging also to other animals. But no spatial considerations ever enter into such a definition: cats may be grouped or not in the space without any change concerning their class definition and properties (1) and (2).

Piaget then introduces figural collections, in which meaning defined by properties (1) and (2) is linked to the spatial arrangement of its elements. He claimed that: *"A figural collection composes a figure, through the spatial relationships between its elements, whereas non-figural collections and classes are free of any figure"*.

E. Figural versus non-figural collections

It is precisely these figural collections that computing is promising more and more an effective modeling of, pushed by an ever-growing social demand for on-line digital media browsing and information research amongst multiple sources.

But as we now understand, figural collections adapt poorly to their assimilation into non-figural collections or classes. Although according to Piaget, collections are destined to become classes, in the same way as subjects will grow psychologically so as to improve their cognitive capacity to classify. Still referring to [11]: *"It is a radical lack of differentiation that nudges figural collections out of the classical modeling field"*.

So classical IT approaches were unable to address and tackle the target of figural collections modeling.

On that particular point, we do not agree with Piaget, considering that, with the support of Inductive Data-Driven Algorithms and Big Data, even non-figural collections are about to be computerized.

IV. How Could IDDA & Big Data Help Supporting Our Figural Collections Deployment?

A. ReCollection: an experimental software for the creation of multimedia collections

ReCollection is a computer program for searching, arranging and browsing digital content, developed by Francis Rousseaux, Alain Bonardi and Benjamin Roadley [15].

As our collecting activities vary from one context to another, it is too ambitious to seek a general solution to the problem. Rather, particular application areas must be defined and isolated, in order for a specific answer to be given, however always relying on a set of basic principles. Here, we shall discuss the software prototype we have created for the digital opera/open form opera Alma Sola (designed by Alain Bonardi [3], and first performed at *Le Cube,* Issy les Moulineaux, October 2005).

1) The reserve

The ReCollection software has two main modes: reserve and gallery. The reserve allows us to store our objects that aren't exposed in the gallery. There are many objects in the reserve, and these are not always labeled; also they are rarely arranged in an orderly and tidy manner. So when we visit the reserve, we have no choice but to wander around,

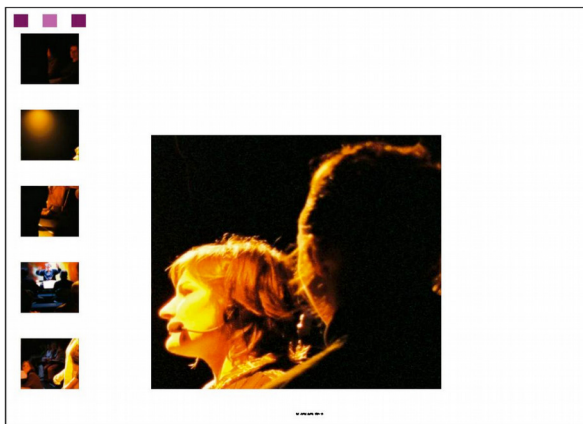picking up objects, inspecting and identifying them one at a time.

Fig 1. The reserve

The reserve can also be compared to the attic, in which our family possessions are stored similarly. As we explore our attic, we can happen to pick up an old photo album, which we had completely forgotten about. This item will surely bring back memories and emotions. We can then choose to keep this album under our arm, as we continue to explore the attic, or we can leave straight away, and put it on our fireplace, for example, making it visible to visitors. It is all these pleasant and familiar experiences, which we believe can be recreated thanks to the modeling of the reserve in our computer program.

The user can create any number of reserves. However, he must create at least one, and store at least one object in this reserve. When he is in reserve mode, he can only view one object at a time. When he decides to view another object, it is chosen randomly from the remaining items in reserve. During a visit, each object is viewed only once. If the user wants to view an item he has already visited, he may go through the history of items on the left side of the screen, as shown in figure 2. When he finds an object of interest, he can move it to the gallery. It will then be removed from the reserve, and saved in memory, with a group of objects waiting to be imported in the gallery. Then, in gallery mode, the user will see this heap of objects, and will be able to import it in the desired gallery, at the desired location.

2) The objects

The items in the Alma Sola collection are made up of three components:

1. A photo of the performance;
2. A sound recording of a few seconds of the singing;
3. A text: the line that is sang in the corresponding sound file.

These are all regular files stored on disk (bitmap, wave and .txt formats). Each item also has a name. In a more general context, the objects can be made up of any one of these types of media, a video (though not implemented in this version), or any combination of these.

Also, each object has a set of descriptors attached. There is a specific set of descriptors for each type of media, which describe the contents of the object, for example the average volume of the sound, the brightness of the photo, the number of words, etc. Depending on the application, we could also include editorial information, such as date, author, etc.

These descriptors may be assimilated to the private properties of traditional computer objects. But in the context of collecting objects, we also need to account for other properties that come from the activities in which these objects collectively engage.

3) The gallery

A collective activity involving a number of objects at a time is their relative arrangement in the gallery space. To the location of objects in this space, we have added their color; these two properties make up an extra conceptual layer, which is the framework for the creation and management of our collections.

In ReCollection, there is always at least one gallery, and the user can create as many as he wishes. There is always at least one item in a gallery, some basic content that the user can interact with, a starting point for his collection.

The objects can be placed and arranged manually in the gallery space, using click and move, just as in common user interfaces. The user can also rely on two algorithms to automatically dispose the objects. The first one, inspired by cataRT software [18], calculates the objects' positions and colors according to descriptors chosen by the user. The second calculates the positions depending on a sample of objects selected by the user. A Principal Components Analysis (PCA) finds out which descriptors vary most amongst the objects of the sample, the system can then rearrange the whole gallery according to these descriptors, as in the first method.

The arrangements resulting from the algorithmic calculations can always be modified manually in order to correct them (in the eventuality of rather subjective descriptors), to build up a global figure, or to bring items together. This way, through creative human-computer feedback loops, meaningful global figures can emerge through the arrangement in space of collected items, as well as local figures, soft pseudo-categories which are heaps of objects brought together by the system and/or the human user. These pseudo-categories are the building blocks for the classes the collection is implicitly aiming for. They are easily and constantly updated; items are added and removed instantly by being moved in space. They are loosely defined and never completely closed off from others, allowing some objects to be lost somewhere in between several heaps, when they cannot be placed in any one category. In a nutshell, this system allows for the creation of collections in which classes are in constant evolution, and are built by exploiting not only the objects' degree of similarity, but also their relative location in space and time.

Furthermore, the user may wish to search for objects in the gallery or in the reserve, in order to build on these categories, look for new kinds, or even fill in gaps in the gallery space. For this, the ReCollection system has two search tools
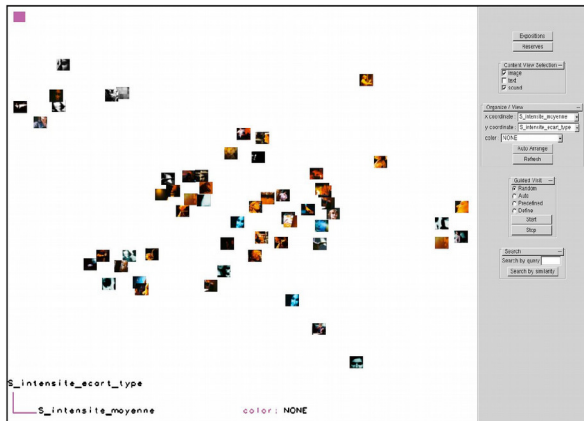
Fig 2. Gallery mode

he can use. The first is a simple 'keyword query', which searches for a keyword within the text or names of the objects. The second is a 'search by similarity'. The user selects an object, or group of objects, and the system searches for items that are similar (according to the descriptors). In both cases, the search is carried out in both the gallery and reserve, and a list of results is displayed in the gallery, ordered by similarity.

B. ReCollection at the time were IDDAs & Big Data technologies are now available

Once all the items of interest have been imported from the reserve, through browsing or searching, and once they have been arranged in the gallery space, the user has a first disposition he can play with. When he will browse the gallery space, his experience will be influenced by the fact that certain objects are close in space, and in time of visitation. Although this is interesting in itself, the system can help the user go further, by defining a set of guided visits, which are simply an order of visitation of selected objects in the gallery.

The type of interface we have chosen to implement these functionalities is a 2D zoomable user interface, inspired by Ken Perlin's Pad [5]. All objects are in the same 2D space, which has no borders. The point of view can be moved vertically and horizontally, and the user can zoom in and out. If he zooms in on an item, until it fills the screen, the sound is played back. This kind of interface has been experimented: Its intuitive approach is seducing to us, particularly in our goal of intuitively collecting digital media. Finally, the spatial metaphor takes advantage of the users' spatial memory and cognitive abilities [7].

ReCollection has typically been designed for supporting figural collections deployment, as the time where IDDAs & Big Data technologies were not yet available. That is why our 'search by similarity' tools were mainly based on PCA.

Nowadays, we are developing some IDDAs to allow the collectors to search by new kind of similarities through remove virtual digital reserves, distributed along the Web.

V. CONCLUSION

Inductive Data-Driven Algorithms really allow us to explore differently digital 'Big Data', if we waive the deductive requirement and permit the inductive heuristics to apply.

But technically, IDDA is not easy to develop, and ethically, it could be dangerous for our decision autonomy because of the lack of traceability that characterizes those IDDA approaches, readily producing 'results' without any linked explanation.

However, we have discovered a 'Big Data' field where IDDAs are fully legitimate and powerful, namely: 'figural' collections constitution, content curation and deployment. This task is much more important for human beings that it looks like, and the lack of efficient tools to support that was clearly identified.

Incidentally, that discovery has disclosed some important Abstract Truth of Big Data: if we accept, against Piaget, that collecting is more native than classifying or categorizing but not less powerful and intelligent, Big Data is the reserve of our future personal digital collections.

REFERENCES

[1] C. Anderson: "Makers: The New Industrial Revolution", New York: Crown Business, 2012.
[2] W. Benjamin: "Paris, capitale du XIXe siècle — le livre des passages", Le Cerf, 1989.
[3] A. Bonardi: "New Approaches of Theatre and Opera Directly Inspired by Interactive Datamining", Proc. Sound & Music Computing Conference, pp. 1-4, Paris, 2004.
[4] P. Brézillon: "Context in Human-Machine Problem Solving: a Survey", Knowledge Engineering Review, Vol. 14, pp. 1-34, 1999.
[5] D. Fox, and K. Perlin: "Pad: An Alternative Approach to the Computer Interface", Proc. ACM SIGRAPH, 1993.
[6] A. Gkoulalas-Divanis, Y. Saygin, and V. Verykios: "Special Issue on Privacy and Security Issues in Data Mining and Machine Learning", Transactions on Data Privacy, Vol. 4, Issue 3, pp. 127-187, December 2011.
[7] L. Hasher, and R. T. Zacks: "Automatic and Effortful Processes in Memory", Journal of Experimental Psychology, 1979.
[8] E. Morozov: "To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems that Don't Exist", FYP Editions, 2014.
[9] F. Pachet, J-J. Aucouturier, A. La Burthe, A. Zils, and A. and Beurive: "Multimedia Tools and Applications," Special Issue on the CBMI 03 Conference, 2006.
[10] F. Pachet: "Content Management for Electronic Music Distribution: The Real Issues", Communications of the ACM, 2003.
[11] J. Piaget, and I. Bärbel: "La genèse des structures logiques élémentaires", Delachaux & Niestlé, 1959.
[12] J. Piaget: "La psychologie de l'enfant", PUF, 2012.
[13] C. S. Pierce: "Pragmatism – The Logic of Abduction", Collected Papers, Vol. 5, pp. 196, 1903.
[14] K. Pomian: "Collectionneurs, amateurs et curieux", Gallimard, 1987.
[15] F. Rousseaux, A. Bonardi, and B. Roadley: "ReCollection: a Disposal/Formal Requirement-Based Tool to Support Sustainable Collection Making", Proc. ICCS Supplement, pp. 131-138, 2008.
[16] A. Rouvroy: "The end(s) of Critique: Data-Behaviourism vs. Due-Process", Privacy, Due Process and the Computational Turn — Philosophers of Law Meet Philosophers of Technology, Chapter 5, Mireille Hildebrandt & Ekatarina De Vries (eds.), Routledge, 2012.
[17] A. Rouvroy: "Le régime de vérité numérique — de la gouvernementalité algorithmique de fait au nouvel état de droit qu'il lui faut", séminaire Digital Studies de l'Association Ars Industrialis, 7 octobre 2014.
[18] D. Schwarz, G. Beller, B. Verbrugghe, and S., Britton: "Real-Time Corpus-Based Concatenative Synthesis with CATART", Proc. DAFx, 2006.
[19] B. Slavkovic, and A. Smith: "Special Issue on Statistical and Learning-Theoretic Challenges in Data Privacy", Journal of Privacy and Confidentiality, Vol. 4, Issue 1, pp. 1-243, 2012.
[20] G. Wajcman, "Collection", Nous, 1999.