1

Ecosystems of Collective Intelligence in the Service of Digital Archives

1.1. Digital archives

The management of digital archives is crucial today and for years to come. It is estimated that every 2 days, humanity produces as much digital information as was produced during the two million years that preceded our existence. In addition to this human production is the information that machines continuously produce. With the cost of digital memory becoming ever cheaper, most of this information is stored in vast databases. In 2025, all of these "big data" will constitute nearly eight zettabytes (trillions of gigabytes) [SAD 15]. In our age, there are very few human activities that do not generate digital archives; each day we feed digital workflows even outside our use of computers, telephones or other digital devices. It is enough for us to turn on a light, run errands, take public transport or watch television to produce digital traces that, for the most part, will never be accessible to us, but which are compiled, indexed and calculated in server farms and management centers.

The status of these digital archives is obviously not the same when dealing with the tweet sent automatically by a cow, the digitization of a course by Gilles Deleuze or the 3D modeling of the Citadelle Laferrière near Cap-Haïtien. Even if these archives are ultimately composed of a set of 0s and 1s and are therefore formally comparable to one another, their

Chapter written by Samuel SZONIECKY.

importance is not equivalent and they particularly vary according to space, time and actor contexts that are faced with this information. The tweet sent by a digital device in relation to a cow's activities[1] is probably not important for most of us, but for the milk producer who wants to follow his herd's movements to correlate the milk composition with the pastures grazed, it is important to know that a certain pasture has an influence on the amount of fat in the milk. Similarly, a certain passage in Gilles Deleuze's courses where he speaks of the importance as a fundamental criterion seems to some people like an almost meaningless phrase while it takes on very great importance for the researcher interested in the relationship between ethics and ontology, but also for the reader of these lines who at this very moment is thinking about this concept just by the fact that they are reading it:

> "What does that mean, this category? The important. No, it is agreed; that is aggravating, but it is not important. What is this calculation? Isn't it that? Isn't it the category of the remarkable or the important that would allow us to establish proportions between the two intransigent meanings of the word proportion? Which depends on and results from the intensive part of myself and which rather refers to the extensive parts that I have[2]."

These proportions between the inner-being and the outer-having are quite easily transposed into the domain of digital archives. Due to their dynamic, upgradeable and interactive characters, digital archives are ecosystems where each element can be analyzed in terms of existence made up of "intensive parts" and "extensive parts". The example of the digitization of the fort at Cap-Haïtien sheds light on the importance of digital archives that illustrate this "intensive/extensive" double dimension that Deleuze emphasizes to show the correlation between an exterior dimension connected to having and the material, and an interior dimension connected to being and the immaterial. In the case of this historic monument classified as a UNESCO World Heritage Site, digital archiving is the chance to develop both a material and immaterial heritage in one of the poorest countries in the world. The creation of an international research program focusing on the issues of augmented realities, the teaching and education of students on these issues, and the mobilization of artists for the innovative use of these technologies are three examples of immaterial heritage development. At the

1 http://criticalmedia.uwaterloo.ca/teattweet/
2 http://www2.univ-paris8.fr/deleuze/article.php3?id_article=24

same time, these activities allow for consideration of material heritage development through the implementation of an economy that uses these digital archives to create new services aimed at tourists on cruises passing by this country. Here, the impact of the digital archive goes beyond the scope of a company or that of knowledge by having repercussions on the whole economy of a country through a joint development of material and immaterial heritage.

Consequently, the fundamental issue of digital archives consists in examining their importance at both the material and the immaterial level in order to estimate their relevance in terms of balance between the finality of the digitization process and the uses made of it. Given the breadth that digital archives take on today and their impact on our lives, we must examine the importance of these archives at both the personal and the collective level. These investigations can only be done through long-term collective work that must take place through a pooling of analyses and the constitution of a collective intelligence capable of lending humanity the means to avoid handing over to machines the full responsibility of semantic choices necessary for the interpretation of archives [CIT 10]. Solutions already exist or are being developed as initiatives taken by the W3C to harmonize information management practices; others remain to be discovered from a technical, epistemological, political or ethical point of view.

1.2. Collective intelligence

It is rather trivial to explain what collective intelligence is through the anthill analogy [FER 97] or all other insect societies [PEN 06]. This conception leads to a very partial vision of the phenomenon of collective intelligence and brings about a questionable ethical position in the case of human organizations. The conception of a collective intelligence modelled on insect societies tends to reduce the human participant in this intelligence to a simple and basic being, whose entire complexity must be removed to make each individual react like the whole. As Bernard Stiegler remarks, therein lie the stakes of a war for control of societies through symbols [STI 04]. Furthermore, it is one of the recurring criticisms vis-à-vis collective intelligence that would only be intelligent in name, and would only serve to centralize memory to better control it without allowing new knowledge to emerge [MAY 06].

What sets humans apart from ants is their ability to reflect on the information flows in their interior and thus express a reflective conscience [LEV 11]. As Yves Prié explains, reflexivity is the ability to get back in touch with oneself in order to construct from memory representations allowing the regulation of one's actions [PRI 11]. This definition, which places reflexivity in an individual context, can nevertheless be understood in a collective framework as well, where individuals share their reflexivity to work collectively in accordance with the consciences of each individual. There we find the basic principles of a science that aims to elaborate a consensus and allows us to define collective intelligence as the sharing of reflexivity in order to complete an action that could not be done by a single person.

But before they can benefit from this collective "ability to act" [RAB 05], the actors must agree to direct their personal interests towards an altruistic sharing of their will. This is possible by formalizing and sharing knowledge while also accepting their validation by collective constraints in order to make the task interoperable and reusable for a community. All of the difficulty of collective intelligence remains in this ability of individuals to agree to restrain their own expressions through formalism, for it quite often challenges habits of reflection. They must not deceive themselves about the primary motivations of humans, which do not necessarily go in the direction of the ethical development of harmonious collaboration. As Yves Citton states, sometimes it is necessary to use tricks to make practices evolve and to anchor them in new social organizations [CIT 08]. It is rather indicative to see that research conducted by Anita Woolley to define a collective intelligence factor confirms that the abandonment of selfish interests in favor of an altruistic approach increases a group's capacity for collective intelligence. In fact, it shows that each individual's intelligence has far less impact than the social sensibility of a group's members, allowing them to listen and not monopolize the discussion in particular [WOO 10].

The issue of restraining individual intelligence in favor of completing a collective action today goes through technical devices and particularly through graphic interfaces that will formalize semiotic systems whose goal is to facilitate individual expression in correlation with the constraints necessary for sharing that expression. The use of a computer language like WIKI is a clear example of going through this constraint to facilitate

the interoperability of an individual expression and completing an encyclopedia's project. These collective intelligence projects do not stop at one computer language; they bring with them an entire knowledge ecosystem at the heart of which these projects will be able to develop through the successive completion of the individual actions.

1.3. Knowledge ecosystems

These are the solutions to these issues that we are going to analyze by taking concrete examples in domains as diverse as corporate innovation or personal archives, but which also have in common the use of collective intelligence to exploit digital archives. To provide a strong coherence to the diverse examples and to handle all of the complexity of the issues they present, we will analyze the solutions following the analogy of ecosystems. In these solutions, which implement collective intelligence approaches in relation to the use of digital archives, we will see how these practices can be analyzed by understanding information not as inert objects but as autonomous beings that develop distinctive ways of life [LAT 12].

The goal of our proposed model consists of developing a generic method for analyzing the ecosystems of knowledge that make up a complex universe of simultaneously complementary and antagonistic relationships between a multitude of human, mechanical, institutional, conceptual, documentary, etc. relationships. With this model, we hope to provide researchers with the means to describe their fields of research and the arguments they defend through the modeling of informational beings. The goal is to be able to render analyses interoperable through the automatic comparison of these beings. To achieve comparative analyses of these ecosystems, we model the informational beings by crossing the Gille Deleuze's Spinozan logical principles [DEL 68] with those of Philippe Descola [DES 05]. Concerning Deleuze, we return to the three dimensions of existence (extensive parts, relationships, essences) correlated with three types of knowledge (shocks, logic, intuition). As for Descola, we use the ontological matrices that characterize the relationships between physicalities and interiorities. More specifically, we focus on the analogy ontology that actually corresponds to the case of digital archives and collective intelligence, given digital

physicalities' unlimited transformational capacity and the multiplicity of interiority relationships proposed by collective intelligence:

> "This continued struggle between a vertiginous ocean and relationship networks, always in the process of multiplying their connections, strictly defines analogism, a word that wonderfully summarizes and paints our objective world, our cognitive tasks, our subjective dreams, and the groups that are born today and will do the politics of the future." [SER 09, p. 85]

With the help of these principles, we form unique representations that we describe as monads. They are made up of four groups: documents, actors, concepts and relationships. Within each group, the elements maintain relations of differential semantics [BAC 07, p. 142] following the relative position of an element in relation to two axes, that of the father and the brother in a tree.

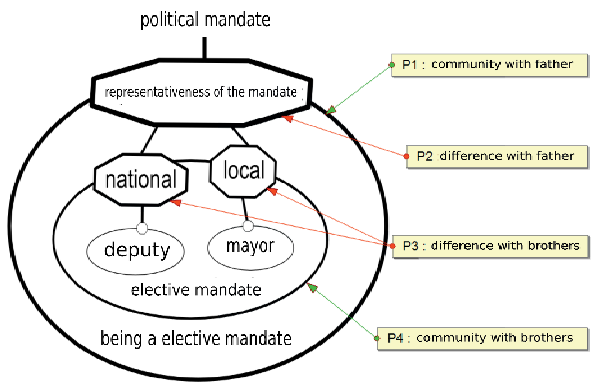

Figure 1.1. *Principles of differential semantics*

The levels defined by the position of the elements in a tree of father–son hierarchies put in contact with the element number in each group gives a precise metric of the monad. This metric allows the level of complexity of a being to be known in order to automatically compare interpretations that cover the same documents, the same actors and the same concepts. We call this metric the Existential Complexity Index (ECI), and we are developing a tool to automatically calculate this index using modeling of a being.

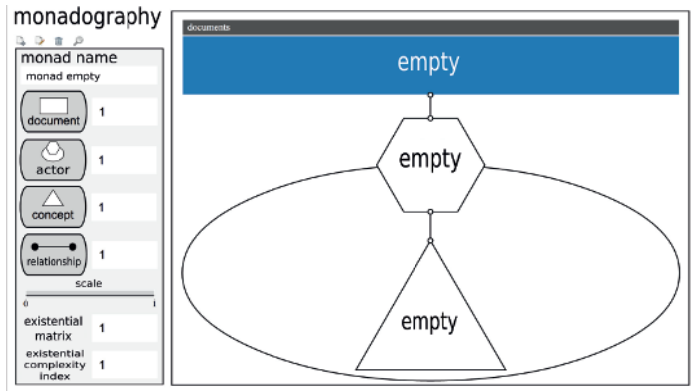

Figure 1.2. *Modeling an empty being*

Each monad and the associated ECI is a unique description of the state of an ecosystem of knowledge at a given moment for a given person. This state gives a particular perspective on the ecosystem; it does not seek exhaustivity, but rather the expression of an interpretation that serves to support arguments and creates the potential for controversies from which the consensus necessary for collective action may emerge.

1.4. Examples of ecosystems of knowledge

1.4.1. *Modeling digital archive interpretation*

The research conducted in the field of digital humanities produces new archive sources that are challenging the division traditionally used by historians and the literati to distinguish between "primary" sources, those produced by the object of study, and "secondary" sources, those produced by research activity. The use of digital technologies leads to the creation of "secondary" archives in the form of databases that, if they are accessible and interoperable, automatically become new "primary" sources for a reflexive analysis of research activities or for other researchers studying the same field. The creation of these digital archives and, more specifically, the durable dimension of their use, conditions the researcher's task by putting an emphasis on the formalization of the task in such a way that it becomes open, interoperable and lasting. This scientific imperative is imposed upon researchers more and more by the simple fact that they work on projects where the digital dimension is central, as it guarantees financing. The

question then arises, how can this data be produced and made visible without being an expert in computer science or knowledge engineering?

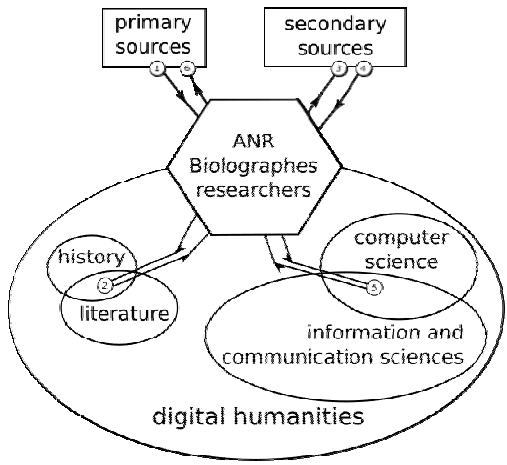

Figure 1.3. *Recursive cycle of sources*

Muriel Louâpre and Samuel Szoniecky aim to tackle this question by analyzing the task performed in the framework of the ANR Biolographes project. This very concrete terrain allows for examination of the nature of digital archives produced by research to extract the special features particular to the field of human science. After a presentation of the digital practices implemented in this type of research, the specific case of visualization methods is dealt with by a review of the primary tools available on the Web in order to critique the epistemological and practical limits. Using the same body of data, the authors show the utility of these tools for quickly testing the coherence of data, for visualizing networks, or for multiplying the approaches and defining new research perspectives. Finally, they reflect on a generic method for modeling influence networks using a prototype developed specifically to help researchers describe their interpretations so that they are interoperable with other perspectives. The goal of this process is to provide cognitive cartographies serving as an aid for the elaboration of a scientific consensus.

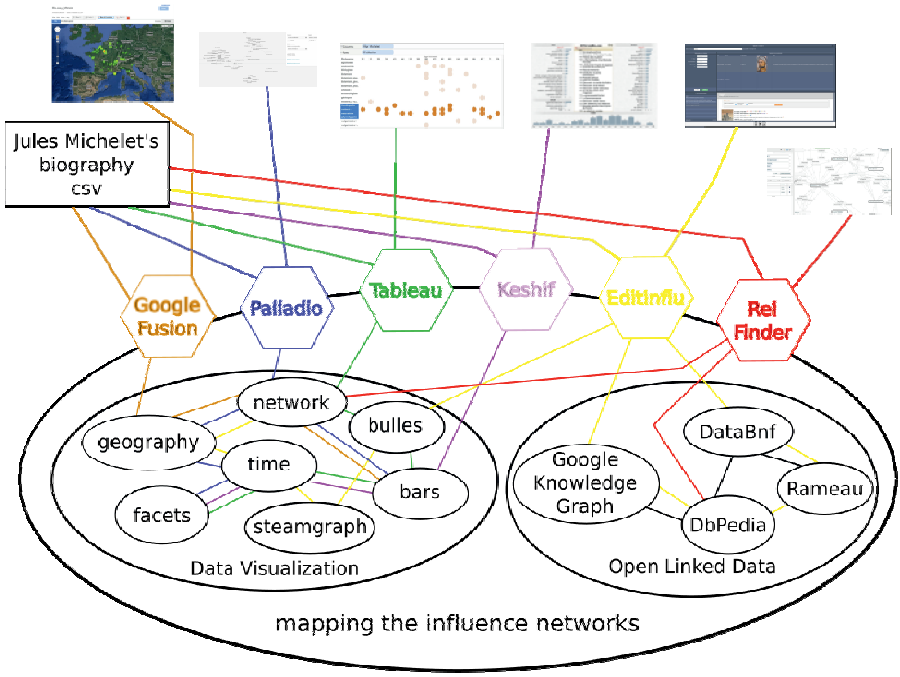

Figure 1.4. *Mapping the influence networks. For a color version of the figure, see www.iste.co.uk/szoniecky/collective.zip*

From these reflections emerges the result of a sometimes-difficult dialogue between researchers coming from different fields of expertise. Faced with the digital "black box", digital models can be imposed upon researchers whose needs in terms of information processing are too often not explained concretely. Even if the lure of a button that can simply be pushed to obtain the relevant information starts to disappear after disappointments and frustrations during the dialogue with the machine, the lack of knowledge engineering training remains flagrant at times. Beyond knowing what the machine can do, it is important for humanities researchers who use digital technology to understand in what way they also bring reorganization to the collective task and research practices.

1.4.2. Editing archives via the semantic web

As we explained above, there are multiple examples of digital archive creation and they not only concern the field of research, but also cultural heritage. Lénaik Leyoudec is interested in the process of editing these digital archives, wondering about the possibilities of preserving the meaning and intelligibility of heritage documents. To explore these issues, he references the differential semantics defended by Rastier [RAS 01] and Bachimont [BAC 07] (Figure 1.1) to deduce an interpretive approach that can be broken down into three consecutive phases: semiotic analysis, document validation method and architext editing. As with the propositions of Muriel Louâpre and Samuel Szoniecky, Lénaik Leyoudec emphasizes the interpretation of the digital archive and the need to equip this process in order to preserve it in the best way possible.

In the framework of an experiment on various audiovisual funds that possess "semiotic objects" belonging to the "private cinema" register, a precise analysis of the cinematographic structure shows how the interpretive approach allows the definition of "memory indicators" at different levels, depending on whether there is interest in a specific plan (micro), a related plan (meso) or all of the segments (macro). This first level of semiotic analysis is enriched by an analysis of the cinematographic indicators specific to family films to bring about the emergence of "perceptive saliences" like so many "memorial diegeses" that will serve as the basis for archive editing. The editing principle proposed goes through the transcription of memory indicators into as many annotations that will define a generic typology: "person", "place", "object", "date" and "activity". What is being played at in this stage of editing is the mobilization of *Linked Open Data* resources like Wikidata.

Fortified by this ambition, a digital device is developed to respond specifically to the needs of family film editing. Devised as an ecosystem of "écrits d'écrans" bringing a semiotic polyphany into play, this tool accompanies the user in the interpretation process by facilitating document annotation. Particularly through a timeline representing the sequences of audiovisual flow, it allows the construction of a graphic in the form of networks for navigation between the categories, a research interface to find the annotations and a device for linking categories with the *Linked Open Data* resources.

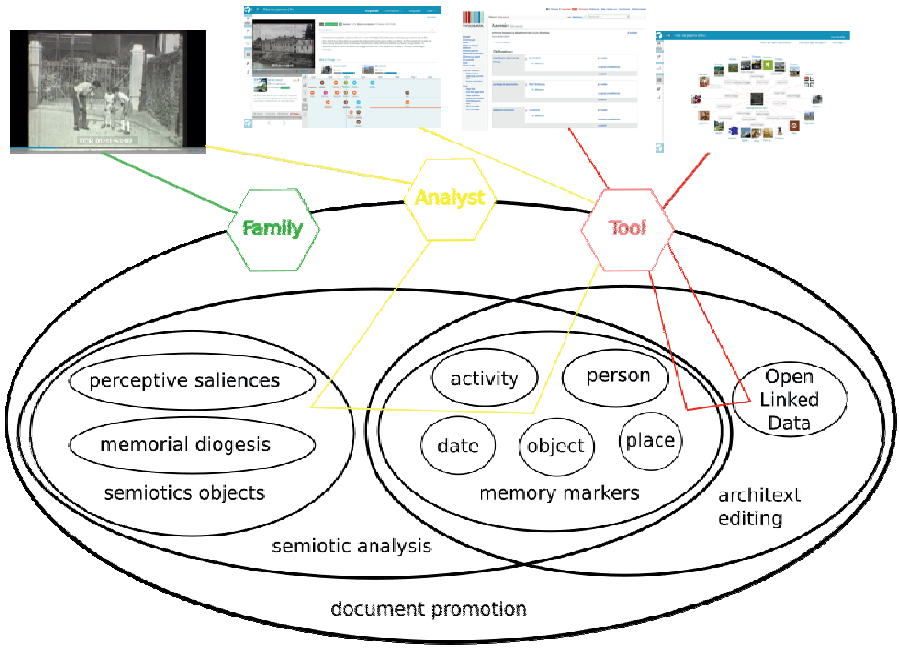

Figure 1.5. *Editing the archives via the semantic web. For a color version of the figure, see www.iste.co.uk/szoniecky/collective.zip*

Numerous questions were raised by this experiment, matching general issues concerning digital archives and the place of collective intelligence in its validation. One of the primary issues concerns the preservation of the document's integrity. Each edition of the document, each interpretation, modifies the primary resources, sometimes by enriching it and sometimes by altering it. Archaeologists know well that as the dig moves forward, they destroy sources of information. Conversely, digital technology allows continuous archiving of resources and their annotations; everything can be preserved. But is this really the most important thing? Is it better to enrich digital memories or to stimulate humans' interpretive experience? If preference is given to the latter approach, it is clearly not necessary to preserve everything for the simple fact that nothing exhausts humans' interpretive capacity, as is shown by the multitude of interpretations for a single book over millennia or a simple sunrise.

1.4.3. *A semantic platform for analyzing audiovisual corpuses*

The previous solution proposes a tool dedicated to the analysis of family films by using *Linked Open Data* to increase the interoperability of interpretations; other researchers are working on similar tools with the aim of facilitating the subjective appropriation of audiovisual data to transform them into a meaningful object. The ANR Studio Campus AAR[3] project has allowed for the development of a tool dedicated primarily to academia and research that increasingly uses audiovisual data as research and educational material. In this context, archives are devised as a *hub* serving as a reference between different communities that form communication ecosystems and lead to a semiotic turning point given the specificity of activities concerning these data.

Structured like foliage of outlines oscillating between the content and the expression, the semiotics of the audiovisual data spreads out according to genres (fiction, documentaries, etc.) and a compositional hierarchy that imposes organization structures and restricts interaction with the data. To describe this system of signs, this tool's creators use the landscape analogy to define a metalanguage and methods of description. In doing so, they make the concrete management of audiovisual data analysis, publication and reediting activities possible.

The Studio Campus AAR sets out to accompany users following two complementary perspectives, the activities of construction and those of audiovisual data appropriation. These activities are made up of steps, themselves structured into procedures that will serve as the basis for orchestrating the data rewriting practices at the thematic, narrative, expressive, discursive and rhetorical levels. These writing/rewriting operations mobilize complex cognitive operations in an intercultural context of re-coverage.

Devised as a software infrastructure based on cognitive and semiotic approaches, this tool aims to provide actors in the audiovisual world with the means to deal with a document in order to transform it into an intellectual resource for cultural education, research and mediation. To achieve this, the solution is organized around an RDF database and a work environment proposing the functionalities necessary for activities of re-coverage: addition

3 For more information about the project, see http://campusaar.hypotheses.org/. To access the experimental Campus AAR portal, see http://preprod.campus-aar.armadillolab.fr/campus/

of an archive, analysis with an ontology, management of individuals, publication/republication, research, modeling the discourse universe.

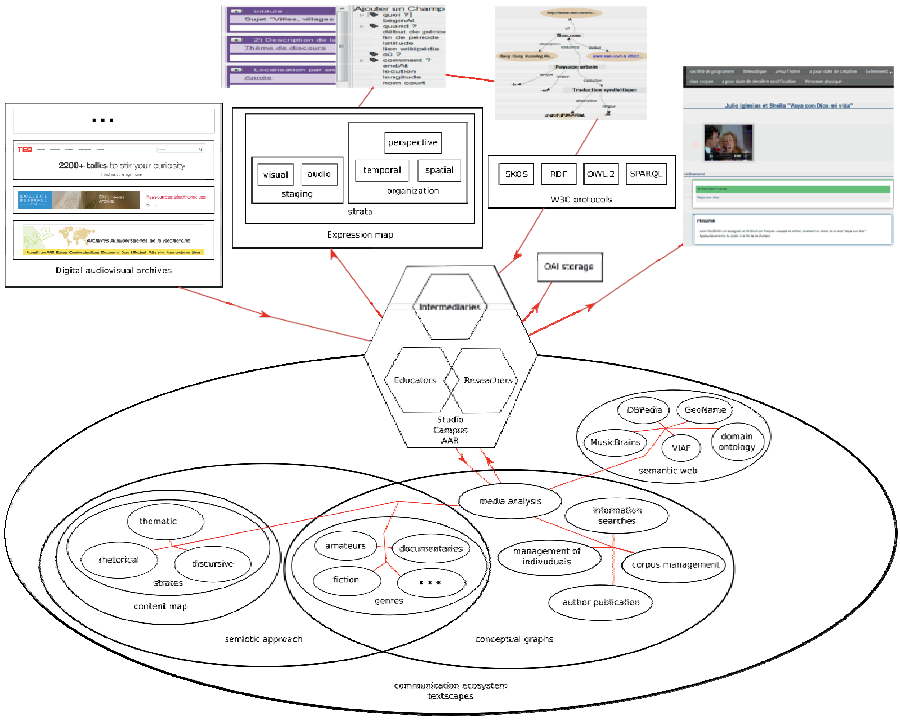

Figure 1.6. *Studio Campus AAR*

At the heart of this platform, knowledge graph editing constitutes a crucial point, particularly for giving those not specialized in knowledge engineering the means to model and analyze the corpuses with languages originating from works of the W3C like RDF, OWL2 and SPARQL. The means of achieving this consist of providing examples of ontology or ontological structures in the form of patterns defining restriction trees. Once the graphs are edited, they can be resolved following different argumentation algorithms that automatically analyze the corpus to deduce content suggestions. The graphical representation of a knowledge graph is another challenge that the Studio Campus AAR is trying to tackle, particularly to reduce the complexity of editing and to respond to the criteria of simplicity, adaptability, dynamism and reusability.

There are various applications of these knowledge graphs that cover all needs through audiovisual analysis. First of all, the media analysis, which consists, for example, of describing the subjects mentioned in the document in the form of strata divided on the audiovisual steam's timeline. This description uses various ontological reference documents and SKOS vocabulary by proposing description patterns via dynamic formulas that suggest ontology entities while the user is typing. These principles are also applied to the management of individuals who will be gathered for faceted questioning, which completes the information search applications via SPARQL requests. Some other applications of this tool to be mentioned are the management of corpuses and author publication.

To finish, Studio Campus AAR offers a complete platform for analyzing audiovisual documents by means of knowledge graphs using formal reference languages (RDF, OWL2, SKOS, etc.) that make the analyses produced durable and interoperable. In this sense, this tool illustrates the work necessary for the formalization of digital archives, so that these will provide knowledge allowing collective intelligence to be developed.

1.4.4. *Digital libraries and crowdsourcing: a state-of-the-art*

Even before being able to promote digital archives, they must first be created by digitizing sources that have not yet been digitized. This task, very simple when the source is recorded directly using digital tools like a word processor or a digital camera, becomes much more difficult when the sources come from a library, or the increase in volume and sometimes their fragility make it difficult to go from an analog to a digital version, and more still the exploitation of digital data that cannot yet be understood by machines. Mathieu Andro and Imad Saleh introduce an original typology of the collective intelligence solutions that can be put into practice to optimize this task through analysis of the notion of "crowdsourcing" and how it is practiced in libraries.

"Crowdsourcing" literally means mobilizing the masses as a resource to carry out a task, but there are different definitions according to whether outsourcing, conscious involvement, volunteering, collaboration, etc. are considered. Whatever the case may be, these practices can be considered to go very far back in time, for example, connecting them to the appeals made in the 18th Century to resolve scientific problems like determining the

longitude of a boat at sea; also, the conceptual origins of this notion find their roots in socialist, Marxist, anarchist, humanist or liberalist ideologies, ideologies that actually place the debate in the political domain, particularly on questions of the "uberization" of libraries.

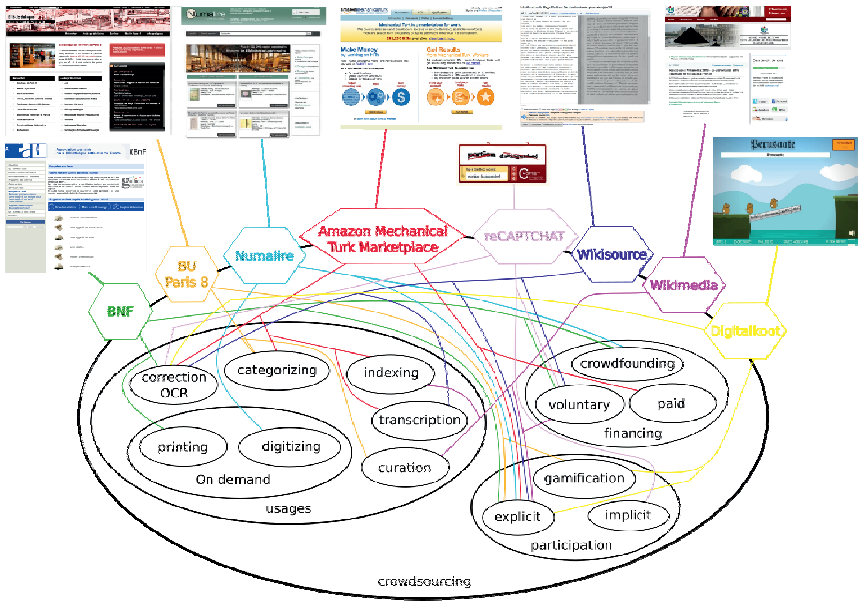

Figure 1.7. *Crowdsourcing in libraries. For a color version of the figure, see www.iste.co.uk/szoniecky/collective.zip*

To analyze these collective collaboration practices, categorizing according to the degree of participant engagement offers a non-negligible quantitative criterion, but one that can be enriched by other, more qualitative criteria. The authors propose, for example, differentiating the implicit practices like gamification or ludification, which consists in appealing to participants' desire to play. "Crowdfunding" constitutes one of the other large categories of "crowdsurfing", where participation is essentially financed like digitization or on-demand printing, for example, which makes it possible to have players pay for a part of the hard work done.

In libraries, there are various challenges of externalizing micro-tasks to Internet users. In addition to reducing costs for correcting errors made by optical character recognition (OCR) tools, these practices would allow the collection to be reedited so as to enrich the existing indexes at the book level

with more precise categorization at the page or even the sentence level. However, the management of libraries is not always open to outside participation, especially the devaluing of employees' jobs, particularly their expertise in categorizing and indexing. Among the other difficulties that halt the development of these collective intelligence projects, we can include the employment of a person dedicated to stirring communities often perceived as useless, the low quality of production and the poor reintegration into information systems, and the difficult evaluation of these projects.

It can be seen here that "crowdsourcing" projects in libraries focus on various issues that allow a better understanding of the relationships between digital archives and collective intelligence. Despite all of these difficulties and the fact that the masses are not always very sensible, "crowdsourcing" is nevertheless a practice that brings about numerous innovations in the fields of technology, economics, politics and even personal development. Let us hope that these experiences will lead to concrete solutions so that we may better coexist in hyper-connected societies.

1.4.5. *Conservation and promotion of cultural heritage*

Human activities leave numerous material and immaterial traces that together make up the cultural heritage whose durable and interoperable promotion is today going through knowledge modeling. To do this, the community of this domain has developed formal languages that take the form of metadata norms like the Dublin Core, LIDO, MODS, EDM, etc. These are completed through the use of controlled vocabularies like KOK, SKOS, RAMEAU, etc. by lexical databases like Wordnet and by ontologies like CIDOC CRM. However, four primary difficulties make knowledge modeling for cultural heritage difficult: the acquisition of data, knowledge modeling, usage and interoperability.

Concerning the acquisition of data, the problem of balance between the complexity of heritage objects, the complexity of implicit expert knowledge and the complexity of formal languages must be resolved. For example, it is often difficult for experts who have their own vocabularies and systems of description to use ontologies whose organization and way of working are different. To facilitate this communication between implicit user knowledge and formal knowledge, it is possible to model ontological paths that will guide the user in the formal description of his or her knowledge. Another

way to perform this task consists of automating data input through automatic language processing technology or through the integration of different data sources. In this case, the problems of contradictory data must nevertheless be managed through the use of a named graph.

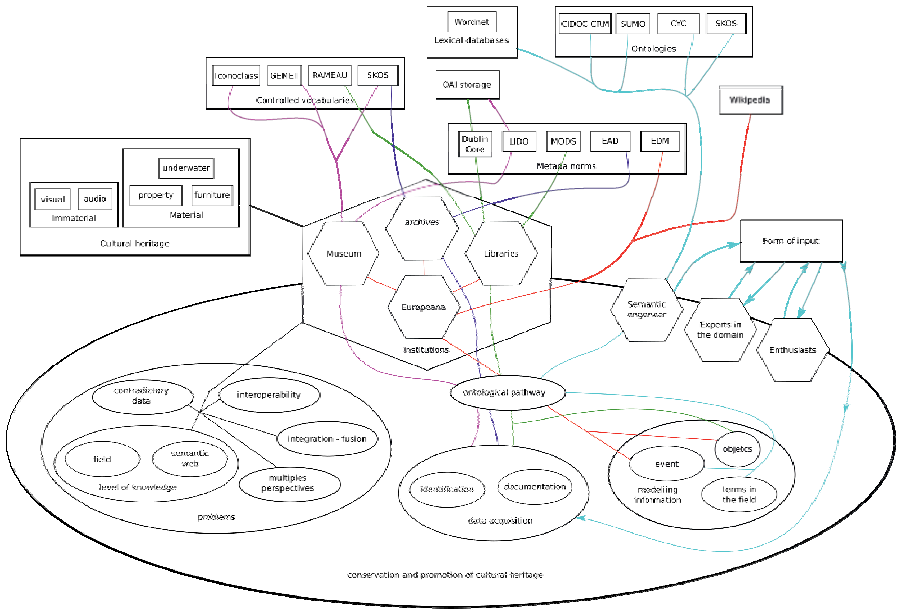

Figure 1.8. *Conservation and promotion of cultural heritage*

The diversity of approaches for modeling information is a central issue. Depending on whether the models come from the field of museums, libraries or archaeology, the approaches are different and the harmonization of these is not always clear. There are methods for automatically calculating the approximation between various formal models that for some use the extension of basic ontological classes and for others appeal to thesauri to enrich the terms in the field.

User profiles condition the uses that will be made of the computer system for the conservation and promotion of cultural heritage. These uses will evolve according to their level of knowledge of the semantic web's technologies, their expertise in the domain and the nature of the terminology. The interfaces of visualization and interaction with information from then on become a fundamental issue so that collective intelligence can be developed

effectively. If they are too complex, the tool will not be used; if they are too simple, they will not serve the users' needs.

As there are multiple ways of describing knowledge, interoperability becomes a challenge, particularly according to the structuring choices that will be made. Even if there are also tools to compare these different structures, the first solution to this type of problem consists of using knowledge models with an elevated level of conceptualization like the OAI-PHM protocol.

Here again, we can see that the use of formal languages undoubtedly contributes to the emergence of a collective intelligence through the qualities of durability and interoperability that semantic technology brings about. Nevertheless, their implementation often remains difficult, is constrained and demands that users adapt their practices. To facilitate this appropriation of semantic technologies, a new actor appears who, by modeling semantic pathways, builds the bridge between ontological complexities and those of experts or enthusiasts in a domain.

1.4.6. *Modeling knowledge for innovation*

The examples that we have just dealt with show how digital technology can help with the implementation of a collective intelligence and facilitate the task of researchers by giving their analyses a durable and interoperable character. The solution that we now present aims to structure the skills offered through a knowledge model extracted from digital archives that researchers create to respond to evaluation demands like those of AERES. The goal here is not limited to accompanying a research task, but rather foresees a prospective dimension by using the model to deduce recommendations for decision makers who, in this case, are not document analysts but rather interpreters of these analyses whose task is to finalize a decision, especially in terms of investment in an innovation. The information system that is developed will allow a real-time evaluation of research activities through a continuous enrichment of experiment returns or the addition of new knowledge.

In this case, digital technologies, notably automatic language processing, are used to create synergies between the world of research and the socio-economic world. The primary goal is to provide decision makers with

information about the skills available in a field so as to be able to respond more efficiently to proposal requests and thus develop a network of innovation involving all the actors in an economic ecosystem. To do this, the researchers form the hypothesis that a field ontology allows the automatic extraction of specific information allowing a laboratory's operating fields by means of a cognitive redaction model. This process follows a three-point methodology.

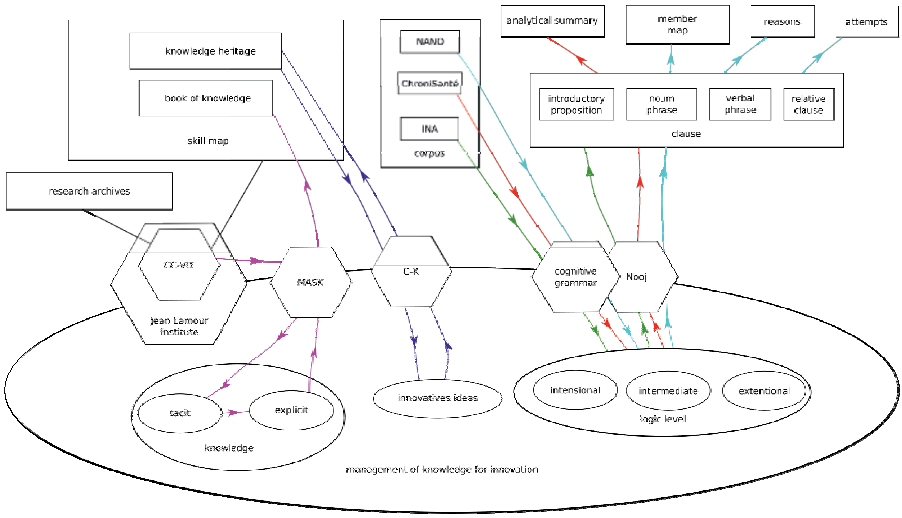

Figure 1.9. *Management of knowledge for innovation*

The first step in this methodology consists of the adaption of Knowledge Management processes to construct the formal descriptions of the organizations that the research will focus on. This stage adopts the outlines of the MASK method, which aims to structure knowledge by adapting it to the C-K theory of innovation to organize the accumulation of knowledge for innovation.

The second step consists of exploiting the digital corpus, thanks to automatic language processing technologies, in order to extract information with a high profit margin. To do this, researchers adopt a morpho-syntactic analysis strategy to extract "organic units" and construct a cognitive grammar from different fields of experimentation.

Finally, the third step aims to organize the information in a class hierarchy in order to structure a skill offer and construct a field ontology. Thanks to the morph-syntactic "patterns" of automatic processing tools, it is possible to construct dictionaries and class hierarchies, and to use them in the analysis of an organism's activities over time or at a given moment in order to identify its skills and expertise in the framework of innovative projects to detect weak signs of a theme, thanks to dictionaries and grammars designed during this research.

1.5. Solutions

These solutions, as numerous as they are diverse, tend to provide an image of incoherence from which no order or expectation can be extracted. However, as chaos, there is definitely order if distance is taken in order to carry out an analysis of the overall arrangement of these solutions. At the systemic level, it can be seen that all of these solutions aim to make the connections that informational existences maintain with one another. In this sense, we can speak of information ecosystems and, using this analogy with a living system, better understand the complexity of the contexts and analyses necessary for understanding them.

From these ecosystemic analyses, it stands out that each of the examples presented seeks to implement formal languages, in most cases issuing from Open Linked Data, which allow an interoperable and durable formalization of the relationships between the documents, actors and concepts. In these modeling tasks that are expressed in the notions of "patterns", "ontological pathways", "digital monads" or "morpho-syntactic patterns", we discern what we call "semantic grains", that is to say, the generic formalization of semantic potential. These grains carry a dynamic algorithm that guides the user in the cognitive maelstrom of digital archives by making connections between the documents, actors and concepts. It is from this continuous growth of semantic potential that the user can construct his or her interpretation by selecting the connections that seem well suited to then share the fruit of these reflections.

Even if the analogy is a bit audacious, it clearly shows the challenges of research on collective intelligence and digital archives and that can be summarized in a few questions. Do semantic grains have a DNA? Who are the creators of semantic grains? Semanticist-engineers? Biologist-cognition

specialists? Can digital archives be promoted, just as a garden is cultivated? Through these questions, what is generally at stake is the ability of human–machine interfaces to make the complexity of formalisms accessible to those who do not specialize in computer engineering without losing the details of expertise in a field.

1.6. Bibliography

[BAC 07] BACHIMONT B., *Ingénierie des connaissances et des contenus: Le numérique entre ontologies et documents*, Hermes Science-Lavoisier, Paris, 2007.

[CIT 08] CITTON Y., "Entre l'économie psychique de Spinoza et l'inter-psychologie économique de Tarde", in *Spinoza et Les Sciences Sociales: De La Puissance de La Multitude À L'économie Des Affects*, Editions Amsterdam, Paris, 2008.

[CIT 10] CITTON Y., *L'avenir des humanités: Economie de la connaissance ou cultures de l'interpétation*, La Découverte, 2010.

[DEL 68] DELEUZE G., *Spinoza et le problème de l'expression*, Éditions de Minuit, Paris, 1968.

[DES 05] DESCOLA P., *Par-delà nature et culture*, Gallimard, Paris, 2005.

[FER 97] FERBER J., *Les systèmes multi-agents: vers une intelligence collective, Informatique, intelligence artificielle*, InterEditions, Paris, 1997.

[LAT 12] LATOUR B., *Enquêtes sur les modes d'existence: Une anthropologie des Modernes*, La Découverte, 2012.

[LEV 11] LÉVY P., *La sphère sémantique: Tome 1, Computation, cognition, économie de l'information*, Hermes Science-Lavoisier, 2011.

[MAY 06] MAYERE A., "L'intelligence collective: une notion en chantier", *Intelligence collective Rencontre 2006*, Presses de MINES, 2006.

[PEN 06] PENALVA J.-M., *Intelligence collective*, Presses de MINES, 2006.

[PRI 11] PRIÉ Y., Vers une phénoménologie des inscriptions numériques. Dynamique de l'activité et des structures informationnelles dans les systèmes d'interprétation (HDR), University Claude Bernard – Lyon I, 2011.

[RAB 05] RABARDEL P., "Instrument subjectif et développement du pouvoir d'agir", *Modèles Du Sujet Pour La Conception: Dialectiques, Activités, Développement*, Octarès Éditions, Toulouse, 2005.

[RAS 01] RASTIER F., CAVAZZA M., ABEILLÉ A. (eds), *Sémantique pour l'analyse. De la linguistique à l'informatique*, Masson, Paris, 2001.

[SER 09] SERRES M., *Ecrivains, savants et philosophes font le tour du monde*, Editions Le Pommier, Paris, 2009.

[SAD 15] SADIN E., *La vie algorithmique: Critique de la raison numérique*. Editions L'échappée, Paris, 2015.

[STI 04] STIEGLER B., *De la misère symbolique: Tome 1. L'époque hyperindustrielle*, Editions Galilée, 2004.

[WOO 10] WOOLLEY A.W., CHABRIS C.F., PENTLAND A. *et al*., "Evidence for a collective intelligence factor in the performance of human groups", *Science*, vol. 330, pp. 686–688, 2010.