

Call for Linked Research for W3C

Workshop: Researcher Centric Scholarly Communication (April 24th, 2018)

22nd February 2018 – IRI

(Giacomo Gilmozzi, Vincent Puig, Yves-Marie Haussonne)

Abstract: The web was originally conceived and developed to meet the demand for automatic information-sharing between scientists in universities and institutes around the world. At IRI, we want to go back to this vision and develop digital tools that enable the contributory construction of meaning, understanding and knowledge. In this article (and in this workshop) we propose to analyze the interest of the categorized contributive annotation model currently tested on three ongoing research projects. This tool has been developed on the basis of a modification of the [hypothes.is](https://www.hypothes.is) platform inspired by the work of the W3C Open Annotation group. In particular we analyze three functions: the group dynamics and related problems inspired by the philosophy of G. Simondon (Hui and Halpin, 2012)¹, the use of meta-categories for transindividuation, and the contributive publication of a glossary.

I. The web we want

We think that the web we are dealing with is not the web we want. This current version of the web is the digital translation of the mainstream economic way of thinking: neoliberalism. But we also think that the web is not (or will not be) necessarily the place of continuous and uninterrupted production of functional stupidity². Using Bernard Stiegler's vocabulary, the web we want is hermeneutical, negentropic³, deliberative and contributive: it is the milieu of production of knowledge and know-hows.

This is why at IRI we chose to develop some digital tools that can rehabilitate web's hermeneutical and contributory functions that were at the core idea of the Internet. The one we want to present you at the Researcher Centric Scholarly Communication Workshop is an annotation platform based on the modification we have implemented on the Hypothes.is project and inspired by the W3C Open Annotation group.

1 See <https://www.iri.centrepompidou.fr/projets/socialweb/>

2 Functional stupidity, as Alvesson and Spicer wrote in their book called *The Stupidity Paradox*, is the inability and/or unwillingness to use cognitive and reflective capacities in anything other than narrow and circumspect ways. It involves a lack of reflexivity, a disinclination to require or provide justification, and avoidance of substantive reasoning.

3 Cfr. Stiegler B. (dir.) (2017). *La toile que nous voulons*. Paris, FYP éditions. With hermeneutical, Stiegler means a web based on interpretation functions (reading, annotating, sharing, categorizing, editorializing and publishing) of texts, videos and images and in line with the tools developed at the Institut de Recherche et Innovation. With negentropic web, the french philosopher means a web which is also the place for the production of knowledge and know-hows and so to diversity-in-order.

In the past years we have developed the idea of a contributive research that inherits two main features. The first one are the principles of action-research, where researchers are not seen as mere observers of the facts of societies but also as contributors to their evolution. The second is the conception of knowledge production that depends on shared digital instruments and organs. This epistemological framework is what Bernard Stiegler calls the general organology⁴, a framework shared by the international network Digital Studies (<http://digital-studies.org>), the journal Etudes Digitales (<http://etudes-digitales.fr/>) and which is being tested by IRI and its researchers in three experiments⁵.

The contributive research has taken on a new dimension with the development of participatory sciences⁶ It can go, in digital context, from a simple production of navigation traces – main object of the data economy and social networks – up to forms of controlled tagging, contributive editing/aggregation (e.g. Wikipedia), editorial and commentary (micro-film reviews) or original publication (blogs).

We propose to analyze the interest of our categorized contributive annotation model questioning the socio-technical functions of the contribution – annotation, indexing, categorization, visualization, group formation, group recommendation and editorialization – by showing how these functions articulate more or less with the algorithmic processing of data.

II. The Hypothes.is Platform Implementations: Meta-categories and Editorialization

Hypothes.is is an open source software project which aim is to give to users a conversation layer over the entire web without needing implementation of any underlying site. In the academic world, it allows researchers to share knowledge on articles, discuss concepts, and then see if there is consensus, but also to analyze the evolution of these controversies in one or more workgroups. Thus, it can speed up while ameliorating the academic research process thanks to the contribution of the various members of a given group. An annotation can thus be private, shared within a group or

4 The *organology* in music classifies and describes the instruments, by extension and in reference to the Organon of Aristotle, we recover this term to designate a general organology, which covers the technical, social and biological conditions of the production of knowledge.

5 Cfr. Contributing Learning Territory project (TAC) for the development of a contributive economy in the north of Paris; in the framework of the ANR Épistème project for the production of knowledge in astrophysics (<https://projet-episteme.org/>). And finally for the discussion on Net-Rights in the NextLeap European project (<https://nextleap.eu/projects/rights.html>).

6 Which includes contributors in the observation and analysis of biodiversity (see in France the success of the site *vigie-nature* of the Museum National d'Histoire Naturelle), of the natural environment or astronomical data.

public. For our purposes, we focus only on annotations that are shared within one (or multiple) group. The group is in fact the condition of possibility for the individuation of the individuals⁷.

Meta-categories. On top of the basic features already available in Hypothes.is⁸, we decided to implement another function. We created four meta-categories (important, keyword, comment and trouble) which are easily recognized by the users thanks to the different colors with which annotations are highlighted – see *Fig. 1* below and *Fig. 2* (p. 6).

Fig. 1 The Annotation Protocol.

Annotation Protocol

important

Use this metacategory when you think the highlighted words are important for understanding of the subject.

mot-clef

Use this one when you think the highlighted expression is a keyword for the subject.

commentaire

Use this one when the highlight is followed by a comment or a question.

trouble

Use this one when you think the highlight is problematic.

Once we highlight a word, a sentence or a paragraph, we can choose between the four meta-categories. As we have seen, the annotations will be highlighted with the color of the chosen category (see *Fig. 1* and *Fig. 2*). This is important to give a sudden idea of what the annotation is referring to. In our experience, annotated texts are easier to understand and quicker to read. This means that researcher can work better and faster.

Once we have selected our meta-category, we can post the annotation to the group. Like in the Hypothes.is project, there is also the possibility to leave comments or short analysis in a box (*Fig. 3*) that helps the members of the group to explain themselves and call for expertise, discussion, to raise a point or a question.

Fig. 3 Comment box.

A dashboard (*Fig. 4*, p. 7) has been conceived to visualize the use of metacategories. It is the starting place for trans-individuation, that is to say the formation of a subject which is partially determined by its group and cultural environment. Users can look through group-annotations, filtering them by metacategory, document or contributor, prioritize them and helping to resolve the controversies.

⁷ See G. Simondon, *L'individuation psychique et collective* (1989). See also the Ars Industrialis Vocabulary: <http://arsindustrialis.org/vocabulaire-ars-industrialis/transindividuation#sdfootnote2anc>.

⁸ See the following link for a graphic explanation of Hypothes.is' functions : <https://web.hypothes.is/#features>

Editorialization. IRI has also developed various functions concerning the *automatic editorialization* of the data collected through annotations. One example is the NetRights-Glossary that we have developed in the context of NextLeap project, an EU funded project that brings together cryptographers, philosophers and sociologists. This glossary is constituted by all the annotations categorized as key-words. Theirs definition(s) are given either by contributors themselves or taken from an external website. It is accessible to all the group members and it permits to create a shared vocabulary for the group (**Fig. 5**). The glossary support multiple keyword definitions, in line with the interdisciplinary and *organological* approach supported by the Digital Studies Network and IRI. Furthermore, contributors' name, the web-reference and the context from which citations are taken are always visible, in line with the academic process of knowledge building. Statistic charts are built automatically while annotating, displaying which kind of category and how many times they have recurred in a given document.

Fig. 5 NetRights Glossary.

Net Rights + Cryptography + Decentralization

The screenshot displays the NetRights Glossary interface. At the top, there are tabs for 'Documents', 'Glossary', and 'Charts'. On the left side, a hierarchical list of categories is shown, including 'blocking', 'censorship', 'common good', 'consent', 'data breach', 'data controller', 'data portability', 'data protection', 'de-anonymization', 'digital rights', 'encryption', 'end-to-end encryption', 'freedom of expression', 'gdpr', 'governance', 'governance of the internet', 'infrastructure', 'internet', 'internet of things', and 'machine'. The main content area is titled 'censorship #' and contains sections for 'Definitions' and 'References'. Under 'Definitions', there is a card titled 'Censorship' with a definition: 'Censorship, the changing or the suppression or prohibition of speech or writing that is deemed subversive of the common good. It occurs in all manifestations'. Below this, it shows a citation from 'mld@h.projet-episteme.org - yesterday #' with tags 'act:definition' and '[censorship]'. Under 'References', there are three cards: 'Internet censorship - Wikipedia', 'Pied Piper's New Internet Isn't Just Possible—It's Almost Here', and 'IPC_engli_4thedition-1.pdf'. Each reference card contains a snippet of text from the source, with some words highlighted in yellow.

Future developments for data editorialization. We are also developing a discussion forum for the keywords presenting problematic or antithetic definitions. In this forum, contributors will have the possibility to discuss about the relevance of the different definitions in order to engage into controversies according to the aim of their research. Another function which is being developed is the automatic editorialization and publication of abstract in a predetermined format with data collected via annotations, the creation of conceptual map summing up the *consensus* or the *dissensus* and also automatic *to-do-lists* based on the action category (act:expertise, act:meeting, etc.).

III. Research problems: The Simondonian model

a) How to go from individual-centered to group-centered social network?

In 2012, in the context of the Arab spring revealing serious threats against individual communication, Harry Halpin and Yuk Hui conducted a research project at IRI⁹ in order to design social networks alternative to the FB principle of maximizing the individual bubble (EdgeRank). Following philosopher Gilbert Simondon individuation theory, they proposed social networks where the group (first entry point) may produce knowledge enriching individuals. Furthermore, Yuk Hui recently argued that the group is in itself a protection of personal data and may provide efficient mechanism for inter-group recommendation ([Beyond personalization and anonymity: Towards a group-based recommender system](#), S Shang, Y Hui, P Hui, P Cuff, S Kulkarni - Proceedings of the 29th Annual ACM Symposium on ..., 2014). This research topic is debated in the Nextleap project for which we developed an Hypothes.is platform for NetRights contribution.

b) How to develop native group-based annotation?

Within the current Hypothes.is data model, contributors may log in a group but only for reading purposes. After writing an annotation they need to publish selecting every time the desired groups. This is not favoring native group-based individuation where reading and writing should not be separated and where every action should be seamlessly visible to all the group members.

The Hypothes.is platform implementations are intended to move towards the Simondonian social-network model by setting the contributors' transindividuation at the center of its purposes.

As Hui and Halpin write:

« *the 'group' [is] based around a common project or calling [which is][...] a projection,*

9 Within the SocialWeb project. See <https://www.iri.centrepompidou.fr/projets/socialweb/>

that is, the anticipation of a common future of the collective individuation of groups. By tying groups to projects, we hold to the fact that individuation is also always a temporal and existential process, rather than merely social and psychological. By projecting a common will to a project, it is the project itself that produces a co-individuation of groups and individuals. Furthermore, by creating a new technical substrate influenced by open standards that are based on this conception of groups, different alternatives can exchange and make elements of their social networks communicable in terms of protocols, data portability, and especially conceptualizations. So while we criticize the limits of social networks and researchers who embrace sociometry as some royal road to understanding social computing, we also want to outline that a new method for understanding – and even programming! – the social and digital is possible, and urgent. »

These are the guidelines that are also drawing our path to what we want to develop within the PIA project in Plaine Commune (Territoire Apprenant Contributif) with a deliberative, decentralized and contributive platform for involving citizens at the local level over data deliberation collected by private and public actors.

Fig. 2 Example of annotated page.

The image shows a screenshot of a document editor interface. The main document area on the left contains text about GDPR pseudonymisation, data breaches, and sanctions. The text is annotated with various tags and highlights. On the right, there is a sidebar with a list of comments or annotations, each associated with a user profile (e.g., VincentPuig, NextLeap) and a date (e.g., 8 feb).

Pseudonymisation [edit]

The GDPR refers to pseudonymisation as a process that transforms personal data in such a way that the resulting data cannot be attributed to a specific data subject without the use of additional information. An example of pseudonymisation is encryption, which renders the original data unintelligible and the process cannot be reversed. The GDPR requires that this additional information (such as the decryption key) be kept separately from the pseudonymised data.

Another example of pseudonymisation is tokenization, which is a non-mathematical approach to protecting data at rest that replaces data with tokens. These tokens have no extrinsic or exploitable meaning or value. Tokenization does not alter the type or length of data, which makes it suitable for databases that may be sensitive to data length and type.

Tokens require significantly less computational resources to process and less storage space in databases than traditionally encrypted data, as they are visible for processing and analytics while sensitive information is kept hidden.

Pseudonymisation is recommended to reduce the risks to the concerned data subjects and also help controllers and processors to manage data breaches. Although the GDPR encourages the use of pseudonymisation to "reduce risks to the data subjects," (Recital 28) pseudonymisation remains covered by the GDPR.

Data breaches [edit]

Under the GDPR, the Data Controller will be under a legal obligation to notify the Supervisory Authority without undue delay. The breach must be reported to the Supervisory Authority within 72 hours after having become aware of the data breach (Article 33). In addition, the data processor will have to notify the controller without undue delay after becoming aware of the breach. However, the notice to data subjects is not required if the data controller has implemented appropriate technical and organizational measures to protect the data, such as encryption (Article 34).

Sanctions [edit]

The following sanctions can be imposed:

- a warning in writing in cases of first and non-intentional non-compliance,
- regular periodic data protection audits,
- a fine up to 10 000 000 EUR or up to 2% of the annual worldwide turnover of the preceding financial year in case of an enterprise (Article 83, Paragraph 4^[19]):
 - the obligations of the controller and the processor pursuant to Articles 8, 11, 25 to 39 and 42 and 43,
 - the obligations of the certification body pursuant to Articles 42 and 43,
 - the obligations of the monitoring body pursuant to Article 41(4).
- a fine up to 20 000 000 EUR or up to 4% of the annual worldwide turnover of the preceding financial year in case of an enterprise (Article 83, Paragraph 5 & 6^[19]).

Comments:

- VincentPuig** (NextLeap) - 8 feb: The GDPR refers to pseudonymisation as a process that transforms personal data in such a way that the resulting data cannot be attributed to a specific data subject. This definition is problematic since using a pseudo has nothing in common with tokenization or producing ID aggregation. In French we can use 2 different words : pseudonymisation et anonymisation.
- VincentPuig** (NextLeap) - 8 feb: encryption
- VincentPuig** (NextLeap) - 7 feb: tokenization

Fig. 4 *Hypothes.is' Dashboard.*

h.

Q

group: NextLeap X Search...

giacomo.gilmozzi Groups ⌵ ⚙ ⌵

244 Matching Annotations

Last 7 days

www.eff.org	A Declaration of the Independence of Cyberspace	2
en.wikipedia.org	General Data Protection Regulation - Wikipedia	7
internetrighsandprinciples.org	IRPC_english_4thedition-1.pdf	10
searchnetworking.techtarget.com	What is Net Neutrality? - Definition from WhatIs.com	1
www.osce.org	empty sheet	4
freedomhouse.org	Freedom of Expression	1
www.britannica.com	Censorship	1
en.wikipedia.org	Internet censorship - Wikipedia	1
www.internetsociety.org	Internet Governance Internet Society	1

NextLeap

NextLeap Group

Created: February, 2018

[Leave this group](#)

Top tags 50

cat:Mot-clef 28

act:definition 27

cat:Important 13

cat:mot-clef 11

decentralization 11

cat:important 10

need:discussion 10

cat:Commentaire 9

cat:commentaire 9

seminar 9

surveillance 9

lang:fr 7

platform 7

political threat 7

issue 6

need:expertise 6

privacy 5

encryption 4

infrastructure 4

internet 4

organization 4

blockchain 3

cat:Trouble 3