SUMMARY

DATA, MUSES AND BORDERS OF CREATIVE ARTS

READING, LISTENING, WATCHING AND GAMING IN THE AGE OF PERSONALISATION

The creative industries, digital content and data

Cultural content as seen through the prism of data

The Holy Grail of recommendation and personalisation

What *data-driven* content and uses for tomorrow?

CNIL
Commission Nationale de l'Informatique et des Libertés

EDITORIAL

O pinion polls regularly remind us that cultural practices take up a large portion of French people's daily life. And yet these surveys are missing something essential. They fail to show us the vital role played by culture in our individual and collective lives.

Reading a book, listening to music, watching films and TV shows - as well as playing video games, no doubt - over and above their entertainment value, are probably the most suitable spheres for the development and permanent remodelling of our identity. It seems to me that the imaginary space they open up to us is a laboratory, a workshop in which we can compare what we think we are with other possible lives; where we can imagine ourselves with other ways of thinking, feeling, loving or, in short, living out our lives. In other words, it is through our contact with artistic content that we constantly invent and understand the parents, friends, citizens, heroes and lovers that we are or aspire to be.

Replicated and distributed on a massive scale for industrial and economic reasons, artistic works are at the same time the canvasses on which shared feelings and beliefs can be played out: they lie at the crossroads of the most collective and the most intimate, at the centre of our public and personal destinies. Being at the forefront of the digital revolution that is radically changing every facet of our societies one by one, CNIL could hardly avoid questioning the impacts of data on the creation and distribution of cultural content - focal points of our identity construction.

The changes on this front are indeed spectacular and as elsewhere, our new practices and consumption patterns are starting to produce personal data. Data feeds personalisation and recommendation algorithms. These algorithms are increasingly indispensable companions to find our way in huge content catalogues. They can foster discovery and diversity, or just as easily lock people into stereotyped tastes or very limited horizons and "filter bubbles". The risk is that the creation and consumption of artistic work will be watered down to such an extent that it no longer gives us an opportunity to encounter "otherness" and different imaginary universes, becoming instead the stage for a pale, endlessly repeated monologue.

Any debate about the value people derive from the use of their personal data should not be confined to the mere fact that the service is free of charge for the user. On the contrary, it should consider what gives cultural consumption its irreplaceable value: namely, that it broadens the field of what is possible and what is conceivable or thinkable.

In the face of these trends, citizens seem to have lost their voice. There is little or no debate: the collection of personal data seems to have become the "necessary evil" for accessing certain innovative digital services. But can we resign ourselves to such a prospect?

Shouldn't we be developing a culture of data that puts us all back at the centre of the digital landscape, and gives data an identity that is more than its mere economic value?

And yet in this sector, by enabling personalisation, data also makes services more useful - maybe even more empathetic. The cultural industries were among the first to be dematerialised; they are something of a testing ground because they have built hybrid business models. It gives us an opportunity to move beyond the over-simplistic stalemate that opposes protection of rights and innovation. Instead of burying the protection of personal data in general terms and conditions that are all too often imprecise and difficult to understand, we should be making it the cornerstone of a relationship of trust with users. Do advanced features require access to the person's location? To the person's movements? To the person's speed of travel? To the various sensors in the person's telephone or connected devices? And tomorrow, to the person's emotions and moods? Why not... but to what end? How? What framework would this require? What limitations should we impose? And most importantly, how can we guarantee that the user will not be passively analysed, supposedly "for his own good"? How can we make sure the user does in fact consent to his data being used for these optional services? How can we demonstrate that these optional features are not just excuses for collecting data that in fact will only ever be needed for sales or advertising purposes?

This IP Report No. 3 aims to fuel debate on the intensive use of personal data by recommendation and personalisation tools, which play a key role today in the consumer market for digital cultural content. It is also a call for economic stakeholders in this sector - and all other sectors - to innovate and show that businesses, which claim to be focusing on the user experience, are also ready to take up the challenge of ethics and trust. Most importantly, it wants the cultural vanguard (both commercial and non-commercial) to take up the subject and invent the digital world of tomorrow.

Isabelle FALQUE-PIERROTIN,
CNIL Chair

CONTENTS

Part 0.1
THE CREATIVE INDUSTRIES, DIGITAL CONTENT AND DATA

Part 0.2
CULTURAL CONTENT AS SEEN THROUGH THE PRISM OF DATA

Part 0.3
THE HOLY GRAIL OF RECOMMENDATION AND PERSONALISATION

Part 0.4
WHAT *DATA-DRIVEN* CONTENT AND USES FOR TOMORROW?

Part 0.1

THE CREATIVE INDUSTRIES, DIGITAL CONTENT AND DATA
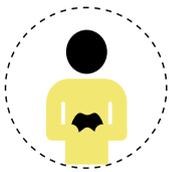
4 SECTORS IN THE SPOTLIGHT

Ebooks

Video and "animated images" content, more specifically video on demand and catch-up TV

Music (creation, production, distribution, listening to recorded music and the emerging market for streaming music)

Video games (PC games, console games, mobile games, and social network games)

TIME SPENT PER WEEK ON CULTURAL AND CREATIVE ACTIVITIES BY FRENCH PEOPLE AGED 10 AND OVER
(2012 by comparison with 2011)

Activity	
Watch TV	**22 hrs 21 min/week** — 3 hrs 11 min/day
Play video games	**12 hrs 17 min/week** — 1 hr 45 min/day (1 hr 45 min more than in 2011)
Listen to radio	**10 hrs 59 min/week** — 1 hr 34 min/day (25 min less than in 2011)
Listen to music other than on radio	**6 hrs 10 min/week** — 53 min/day
Read books	**5 hrs 27 min/week** — 47 min/day (21 min less than in 2011)
Watch videos	**4 hrs 12 min/week** — 36 min/day

Source: GfK - 2012 survey of 2,305 respondents/ EY, Panorama des Industries Créatives et Culturelles en France, November 2013.

DIGITAL CONTENT AND NEW CULTURAL PRACTICES

Digital content is everywhere in our everyday lives. Today it accounts for around 40% of the cultural industry's revenue, as against only 17% in 2010. It is expected to reach 63% in 2018.

Roughly one-third of the French people who use a streaming music service have a paid subscription (source: Médiamétrie CNIL, 2015).

In France, streaming is gaining ground and already generating revenue on a par with music downloads (source: SNEP). The same trend can be seen worldwide, where streaming represents over one-quarter of digital music revenues.

End-2014, roughly one-third of French online users had already watched a video on demand (source: Médiamétrie)

28 million French people play video games (source: EY/CNC)

11% of French people over the age of 15 read ebooks (source: Hadopi)

AS CONTENT GOES DIGITAL, NEW BUSINESS MODELS ARE EMERGING IN THE CULTURAL INDUSTRIES

The growing adoption of digital content is hastening change in the cultural and creative markets. The trend is reflected in their business models, which are changeable, diversified and hybrid, ranging from "free" content financed by advertising to "unlimited" subscriptions and pay-per-use models.

DIGITAL CONTENT AND NEW CULTURAL PRACTICES

Paid subscription: monthly fee for online gaming

AYCE: "All You Can Eat" - unlimited subscription to a catalogue of online games

Freemium: basic functionalities are free, options are extra

F2P: "Free 2 Play" - free game with micropayments charged during the game

P&PM: "Pay and Play More" - unit online purchases made with micropayments

[1] 1 IDATE (IDATE. Content Economics, September 2014).

These markets are a formidable testing ground and challenge the statement that data is "the oil of the digital economy". Admittedly, users' personal data is used as a bargaining counter, but it is also used to personalise the user experience.

France Télévision's Future Media team asks the question in its Metamedia trend report no. 9: "What if the price of culture became variable tomorrow, like the price of plane tickets or hotel rooms? With yield management and real-time price management, prices could vary with the user, the user's tastes, the time of day, or the screen used. " (Read more, page 15)

PERSONAL DATA AT THE CENTER OF THE CREATIVE AND CULTURAL INDUSTRIES

The Wall Street Journal described this landscape as early as 2012: "It takes the average reader just seven hours to read the final book in Suzanne Collins's "Hunger Games" trilogy on the Kobo e-reader - about 57 pages an hour. Nearly 18,000 Kindle readers have highlighted the same line from the second book in the series (...) And on Barnes & Noble's Nook, the first thing that most readers do upon finishing the first "Hunger Games" book is to download the next one." (Read more, page 16-17)

France remains in a relatively enviable position in these sectors, with a few international heavyweights in content catalogues (Universal Music Group, Hachette, Ubisoft, Canal Plus, etc.) and a few gems in the new digital markets (Deezer and DailyMotion, for example).

On these markets, internationally speaking, GAFAM (Google Apple Facebook Amazon Microsoft) are symbolically giving way to ASNS: Amazon for books, Spotify for music, Netflix for video on demand and Steam for video games.

The growing use of data is a trend that cuts across the cultural and creative industries. While data can be used to deliver value-added service, it can also be used to serve very different objectives. Some data is necessary for service delivery, while other data is used to build insights that can benefit the service provider, third parties or the end user

"Platform users can't tell the difference between when they are simply enjoying extremely effective (and free) services and when they are themselves adding data and calibrations to the services in question, to help optimise them and help the business derive added value." (Read the debate between Antonio Casilli and Dominique Cardon on *Digital Labor* on page 58)

These forms of user participation in creating wealth for economic stakeholders can be likened to what some researchers have dubbed *digital labor*. To the saying "If it's free, you're the product", we can now add Antonio Casilli's suggestion: "If you don't pay, then you are a worker".

CULTURAL CONTENT AS SEEN THROUGH THE PRISM OF DATA

THE DEFINING TRAITS OF PERSONAL DATA IN CULTURAL APPLICATIONS: A PROPOSED TYPOLOGY

The personal data collected and processed in connection with cultural and creative content has distinctive features that stem from the very intimate nature of the relationship between a person and cultural content: cultural practices help construct one's personality and identity. The cultural content we use is intimately bound up in our identity: it helps make us what we are.

Data scientists in this field can use very trivial data that is in fact quite distinctive, such as profiling, content descriptions, popularity, enriched data, tastes and context.

Value sources for *data sciences* vary from one sector to another: the type of information that can be deduced from the music a person listens to, from what they read, from the films or series they watch or from their use of video games, is very different.

Without trying to be exhaustive, what are the main types of data associated with the way we use (or create) creative and cultural content? Profile data, data describing the catalogue's content, data enhancing this content, information about popularity, information about each user's tastes, contextual data (see below and opposite and pages 23 to 26 of the Report).

In the musical field, for example, the context will be very important (are we listening to music at work? during sport? at a party?). With books, on the other hand, it will be crucial to analyse how the person interacts with the content (how quickly or slowly they read, what they highlight, etc.) in order to understand what catches their interest. For video on demand, it will be important to understand the structure of the household and try to guess who is watching at different times of day.

TYPOLOGY OF DATA

The main types of data associated with the way we use (or create) creative and cultural content.

Personal data in the most conventional sense: identity, contact details and socio-demographic **PROFILE**. This is the data usually contained in a customer file and traditionally used for customer relationship management.
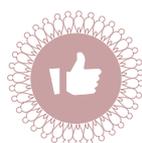
ENRICHMENT data: artist photos or biographies, reviews, scores and rankings, download links, prices, song lyrics. This data might be supplied by professionals (e.g. critics) or generated by users.

CONTENT DESCRIPTION data: not only cataloguing data (artist, author, performer) and characterisation data (duration, genre, subgenre) but also technical data (format, compression, sampling) and legal data (on copyrights and licence, for example).

Data about each user's use, behaviour and **TASTES**: this data can be very general (type, quantity, duration, frequency of purchases/visits, playlists created, etc.) or very specific (passages highlighted in a book, speed of reading, etc.).
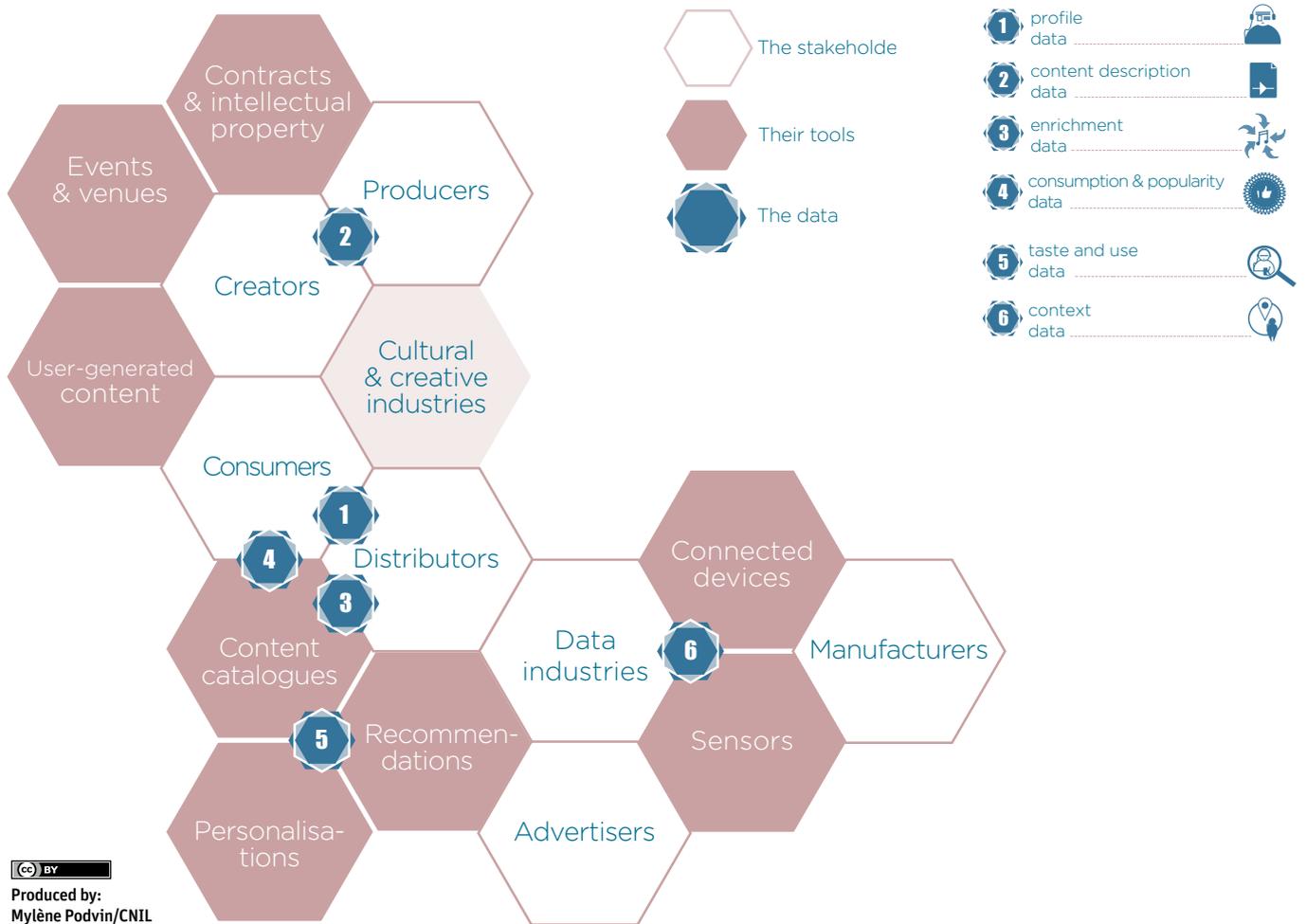
General consumption information, i.e. about collective tastes, such as the **POPULARITY** of a song, the total number of listens, the comments and mentions on social networking services, etc.

CONTEXT data, such as time stamping, location or all of the data yielded by sensors (movements, emotions, physiological state, mood, etc.).

MAP OF THE CULTURAL & CREATIVE INDUSTRIES

Quelles sont les données et où créent-elles de la valeur dans l'écosystème des industries culturelles et créatives ?

The stakeholde

Their tools

The data

1 profile data

2 content description data

3 enrichment data

4 consumption & popularity data

5 taste and use data

6 context data

Contracts & intellectual property

Events & venues

Producers

Creators

User-generated content

Cultural & creative industries

Consumers

Distributors

Connected devices

Content catalogues

Data industries

Manufacturers

Sensors

Recommen-dations

Personalisa-tions

Advertisers

**Produced by:
Mylène Podvin/CNIL**

This map is an attempt to visually represent our analysis of the importance of data in the cultural and creative industries' ecosystem.

It shows how this sector sees the emergence of stakeholders with different and complementary roles, who are being incorporated into a set of more traditional stakeholders. These new stakeholders are seeing their role evolve, since a lot of data can be created, remixed and inferred in their immediate surroundings, close to their tools, their sensors and their devices. These sensors and data science tools (for personalisation and recommendation) are gradually becoming what will make or break the sector.

See the map on .

FROM EBOOKS TO DIGITAL WRITING AND NEW READING EXPERIENCES

"Tell me how you read and I'll shape content to your reading style."

TRENDS AND HYPOTHESES?

• Ebooks move towards digital writing experiences.

• End of linear reading: different types of reading on different scales (extracts chosen to suit the user's tastes, narrative frameworks built for the reader, etc.).

• Dual industrial stakes: keep the reader inside the book and open the book to the reader's community.

• Reading media become more hybrid.

• Use user data to shape the production of certain content.

• Capture the emotions in reading.

KEY DATA

TODAY

■ Technical data on reading (favourite passages, reading speed, depth, etc.) and semantic analysis of the contents.

TOMORROW?

■ Emotions and metadata for algorithmic writing, combined with the captured context.

Examples of emerging applications: Nosy Crow, Wattpad, TimeReader, Glose, IDEO "The Future of the Book"

Read more on pages 28 to 31

FROM PLAYLISTS TO CONTEXTUALISING AND SOUNDTRACKING DAILY LIFE

"Tell me where and when you listen and I'll give you contextualised content"

TRENDS AND HYPOTHESES?

• Very strong identity function, which will be seen in the demand for certain content publicity (sharing, search for social visibility) and, on the contrary, by more discreet consumption.

• Musical consumption involves a lot of zapping, "binge" listening, style switches, saturation mechanisms, etc. It is often linked to the context (I'm alone, I'm at work, I'm at a party, I'm working out, I'm in my car, etc.)

• The "user journey" changes throughout the day, using a wide variety of methods to access music on different devices, media, etc. Music can be omnipresent, for instance, when we are in a shop, at the hairdressers... or in a lift.

• Automated content contextualisation: to suit the individual, the group, something shared, a compilation, etc.

KEY DATA

TODAY

■ Tastes, styles, genres and subgenres, the intensity of collective consumption at different times (popularity by country, by city, etc.)

TOMORROW?

■ The "context" is detected via sensors/connected devices (this is already happening with smartphones): location, emotion, lifestyles and events (party, work, family, etc.), places, types of transport, etc.

Examples of emerging applications: Zero Button Music Player, Prizm, Lucie Hive

Read more on pages 32 to 35

FROM VOD TO ADAPTED AND ADDICTIVE CONTENT

"Tell me what you watch and I'll give you new immersive, addictive content."

TRENDS AND HYPOTHESES?

• Constant effort by video on demand service providers to strike a balance between the experience of discovery and the depth of the catalogue (access to a lot of content, even old content).

• To survive and prosper, subscription VoD needs to retain its users and make it easy and comfortable for them through curation.

• Data is rapidly moving up the value chain and today has become a significant factor in decision making about content production: films and series, which are very expensive to produce, are increasingly calibrated and fine-tuned by algorithms to be successful.

• Lowering of entrance barriers to production and distribution ("from GoPro to Youtube").

• Local versions of global content.

KEY DATA

TODAY

■ Tastes and preferences, consumption habits, composition of the household

TOMORROW?

■ Personalisation "on the fly", based on emotions and behaviours, state of mind and physiological data

Examples of emerging applications: Spideo Mood-Based Discovery, Studio 4.0, AwesomenessTV, Twitch, Oculus Story Studio, Sensory Stories

Read more on pages 36 to 41

FROM ONLINE GAMES TO AUGMENTED HUMAN-MACHINE INTERACTION IN IMMERSIVE ENVIRONMENTS

"Let me analyse how you behave and I'll give you hyper-personalised, addictive content"

TRENDS AND HYPOTHESES?

• Video game business models, which had been relatively simple up until now, are rapidly becoming more diversified and hybrid, with on one side the spectre of "free" games and on the other gaming "hits" whose production costs outstrip those of Hollywood films.

• The growing personalisation of "add-on" offers (bonuses, specific content, special offers, perks etc.).

• Players spend dozens of hours actively "in" the game (today figuratively, tomorrow increasingly literally, thanks to the new virtual or augmented interactions): video games are tremendous tools for analysing reactions and behaviours.

• The arrival of new human-machine interactions (especially virtual and augmented realities).

KEY DATA

TODAY

■ Gaming statistics, the user's choices (propensity to pay for options, bonuses, add-on content) in a context that is as dynamic as possible.

TOMORROW?

■ The user's behaviours, "preferences", choices and emotions, in real time. Information coming directly from body sensors (stress, heartbeats, brain waves).

Examples of emerging applications: emotion detection, new narrative forms, Muse and Emotiv, Oculus Rift, Projet Soli, HoloLens, Magic Leap...

Read more on pages 42 to 47

Part 0.3

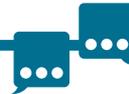# THE HOLY GRAIL OF RECOMMENDATION AND PERSONALISATION

Recommendations are used to attract new customers, build customer loyalty and optimise revenues.

On the technical side, recommendation can combine very different tools and approaches, making it all the less transparent and comprehensible for the user.

Paradoxically, it can harm "user experience", so sometimes a less seamless interaction can be a good thing and give the system a more human feel: frictions can be desirable.

From this angle, being explicit about how personal data is used might appear to be at odds with the goal of making technology invisible in the experience. In fact, though, these "flaws" may foster trust and are opportunities to innovate and stand out from the competition.

To reconcile innovation and the protection of rights, we have to make ethics and trust an integral part of the user experience!

© André Gunthert

EXTRACT FROM THE INTERVIEW…

OLIVIER ERTZSCHEID,

SENIOR LECTURER IN INFORMATION SCIENCE AT THE UNIVERSITY OF NANTES

"The key question when it comes to recommendation, whether we are talking about cultural content or political or religious opinions expressed by our friends on social networks, is the question of Eli Pariser's filter bubble: do algorithms foster diversity or, on the contrary, homophily (the reinforcement of our own tastes, beliefs, opinions and affinities)? All of the scientific studies on the subject, (…) show that two levels of interpretation are superposed: (…) at the micro level, there is an increase in diversity, and a decrease in diversity at the macro level.

(…) The problem is that it is virtually impossible to reconstruct this process of algorithmic recommendation afterwards: because the recommendation has no tangible form, the consumer has no way of recontextualising it and, all too often, simply puts up with it. Traceability works in one direction only."

EXTRACT FROM THE DEBATE BETWEEN

DOMINIQUE CARDON AND ANTONIO CASILLI,

SOCIOLOGISTS AND AUTHORS OF QU'EST-CE QUE LE DIGITAL LABOR ? (INA ÉDITIONS, 2015)

ANTONIO CASILLI: "What counts is data, or even metadata, clicks and connection measurements. With digital labor, the accent is on simply extracting data from the mass of users, not just from the most specialised or most qualified one. This type of production does not presuppose any particular skill or prior training. Being logged-on is enough."

DOMINIQUE CARDON: "Individual productions are only valuable through the database that links them to a huge mass of other individual productions. Individual work is linked to the emergence of a common good, a utility produced by the aggregation of individual activities. (…) From now on, the parties can feel authorised to demand individual accounts, because the whole is only the expression of the economic interests of the firms that develop misleading algorithms."

"THE HUMAN ASPECT WILL INEXORABLY DISAPPEAR FROM RECOMMENDATION AND BE REPLACED BY ALL-POWERFUL ALGORITHMS"

Probably not. Recommendation is produced by humans more than anything else. What's more, social recommendations ("your friend X liked this content") are considered quite effective by some services. In the field of cultural content especially, experts continue to play a crucial role. For video, the *Canal+* Group's new *"Suggest"* engine is banking heavily on the group's editorial credibility to differentiate itself from *Netflix*, adding the work of its editorial experts to its powerful algorithms. In the music field, many services, including the brand new *Apple Music*, offer content selected by musicians or journalists. According to *TechRadar*, the *Slacker Radio* streaming application popular in the United States said that 84% of its premium customers spent more time listening to human-curated playlists than those generated by algorithms on the basis of a track or an artist.[1]

"ALGORITHMS KEEP CUSTOMERS HAPPY AND ARE EFFECTIVE RECOMMENDATION METHODS"

It's hard to answer this question. Does recommendation actually work? Are users really happy with it? Users are often disappointed by the suggestions made by recommendation algorithms, even if the platforms' marketing departments lavish praise on them. In an interview, novelist Aurélien Bellanger said he was always a bit disappointed to see that Amazon persistently recommended he buy the book he wrote himself[2]...

Fact Checking

The sometimes very disappointing magic of algorithms

"ALGORITHMS RECOMMEND... WHAT SUITS THE SELLER"

It's possible... and consistent with the economic requirements. The recommendation market can be analysed as a typical example of a two-sided market. The influencer addresses two segments - users and content producers - offering content for the first segment and consumers for the second. The interests of the two groups are not naturally aligned: the producers' prime concern might be to recoup the costs on the most expensive content. So what happens to the neutrality of the recommendation in this case? Is the recommendation made in the consumer's interest, to offer him the content most in phase with his tastes and expectations, or in the interest of the content producers, to boost a blockbuster or prolong the life cycle of a cultural good? Won't the service give preference to its own products or the content on which it makes the biggest profit?
And in a subtler manner, do the stakeholders recommend content simply to improve the algorithm? A research team at Boston's MIT found that certain products from the *Amazon* or *Netflix* catalogues, for example, come with a disproportionate amount of information about users' tastes. If the service provider recommends these content items to all of its users then, based on the resulting ratings, it will be in a far stronger position to classify users into profile groups[3]. Users are therefore an essential cog in any recommendation mechanism. Without their unwitting digital labour, the system becomes less effective (see insert on *digital labor* below).

"THE RISK IS THAT ALGORITHMS WILL IMPOVERISH THE CREATIVE FIELD"

Maybe... but you can't be sure of it! There is a real risk of impoverishing cultural discovery by sacrificing serendipity to algorithmic efficiency: products that do not, at first sight, seem to match a user's profile or supposed tastes are less likely to be suggested to him. It is a worrying trend for content creators and the authorities charged with promoting cultural diversity (see insert Chapter I-2). *Spideo,* for instance, denounces the spiral that can lead a recommendation system to constantly suggest the same content[4] (see insert "Three questions for Olivier Ertzscheid"). An algorithm is not an immanent entity either: it is the product of a computation based on criteria, much like any other editorial choice. Different criteria or analytical sources will create different paths for the same user. What threatens cultural diversity is less the dominance of algorithms than the dominance of a monopoly *stakeholder.* Lastly, some entrepreneurs, artists and researchers believe that algorithms are anything but the enemies of creation! They see a new type of creative process emerging, one that uses algorithms as a tool.

WHAT RECOMMENDATION IS AND ISN'T

"ALGORITHMS RECOMMEND WHAT EVERYONE LIKES. THEY'RE NOT GOOD AT RECOMMENDING THINGS THAT ARE COMPLETELY NEW, OR HELPING PEOPLE DISCOVER THE UNEXPECTED."

Sometimes. Some recommendation methods (such as collaborative filtering, for example) have powerful retroactive effects and will therefore boost the reputation of content that was popular to start with. In this model, popularity breeds popularity, and the algorithm ends up evaluating how long a hit has been a hit, more than its actual quality. The most familiar unintended consequence of automated recommendation tools is what is known as the "cold start": new content is not rated, is not popular, and the system has little information about it, so it will never get recommended. In this case, the algorithm produces only self-fulfilling prophecies.

"ALGORITHMS MAKE RECOMMENDATION LESS TRANSPARENT"

Often. There have always been questions about just how independent critics, prize juries and, more recently, "influential bloggers"[5] are. But the question is resurfacing from another angle as recommendation becomes increasingly automated.
Recommendation may still be merely a suggestion that users can either accept or decline, challenge or agree with, recommendation stakeholders wield considerable power, which is compounded by the opacity of the algorithms used. Users will never really know how the recommendation is calculated, nor that they are perhaps locked into content bubbles... If, in the future, recommendations are based on data captured by connected devices (see Part 4), it will no doubt be even more difficult to obtain any real transparency as to the reasons for the recommendation. This is why some stakeholders have considered legislating on the transparency of algorithms.[6]

[1] SMITH, Chris. "How Spotify, Netflix and Amazon control your online habits", *TechRadar*, January 2014.
[2] Interview with the editorial team, November 2014.
[3] See the press release on the MIT websit, November 2014.
[4] SPIDEO, Beat Netflix at its own game, May 2015.
[5] Given that the role of these interfaces is to influence opinion and reduce the asymmetry of information
[6] See, for example, the study by the Conseil d'Etat, Le numérique et les droits fondamentaux, september 2014.

TO RECONCILE INNOVATION AND THE PROTECTION OF RIGHTS, MAKE ETHICS AND TRUST AN INTEGRAL PART OF THE USER EXPERIENCE

Once the principle of data's utility value has been factored in, it is finally possible to move beyond the sterile opposition between innovation and the protection of rights. The general principles of data privacy are no longer seen as restrictions, but as the basis for building, or restoring, a relationship of trust between users and these services. This is not only the CNIL's position, it also features prominently in the 2014 Conseil d'Etat study, *"Le numérique et les droits fondamentaux"*[25] and in the Conseil national du numérique report *"Ambition numérique"*[26]: concrete steps must be taken to uphold people's right to make their own choice (right to data self-determination), and ensure that platforms and their tools act fairly. The controversy that broke out in certain media and on the social networks in August 2015 over Spotify's new confidentiality policy shows that any misunderstanding of these questions can spark a crisis between the services and certain users. Spotify was obliged to respond within days and clarify its intentions, whereas they should have been clear.[27]

For services providing personalised access to creative and cultural content, four points should be addressed as a matter of priority :

▶ USER INFORMATION AND CONSENTS

The services need to move beyond the traditional vision of terms and conditions, where ill-informed users have no real choice but to give their consent by signing a contract that no-one really reads. For example, service providers should be encouraged to use installation screens, access authorisation screens, context-sensitive notifications, etc. that are not designed to mislead users or simply apply the law in a formal, minimal way, but to genuinely explain what will be done and used. Another useful possibility would be to explore the idea of providing dashboards that show exactly what data is actually being used.

▶ RIGHT TO DATA PORTABILITY

As a consequence of this need to explain and intelligently inform users, data - which, as we have seen, can play an important economic role - must not become a sort of war chest that businesses seize and use to prevent customers from changing service provider. Data must be recoverable and reusable by the individual, in standardised formats. This right to portability, which is written into the draft European general data protection regulation , also guarantees that personal data will not be used by the stakeholders that monopolise them to impede competition, to the detriment of individuals

▶ TRANSPARENT ALGORITHMS

Today, a company has to be transparent about its prices, its industrial processes, its human resources practices, its purposes, the data it collects and its environmental impact. Tomorrow, it may also have to be transparent about its algorithms, their rules and their settings. It would be good, therefore, to have a certain transparency in recommendation algorithms. In the cultural field, the main goal is to guarantee platforms' neutrality with regard to cultural diversity and ensure that users are not confronted with incomprehensible and obscure recommendations. In this respect, the sector may be more advanced than others: in the future, they, too, would do well to design their recommendation tools so that they are not seen as unchallengeable authorities that must be obeyed...

▶ FAIRNESS

Lastly, fairness must prevail when it comes to data processing purposes.

Data about a person's consumption of cultural and entertainment content can prove to be very sensitive because of what it reveals about the person, his tastes and his interests, and because of what can be inferred about his lifestyle. For example, analysing the history of a person's use of *Spotify* or *Netflix* can make it easy to see when the composition of the household changed with the arrival of a child (i.e. when the type of content used changes to include more children's nursery rhymes and cartoons). What makes this data so personal is also the fact that people generally listen to, read, watch and play things that they like and that reflect who they are. This may seem self-evident, but it underscores the fact that few people today employ strategies to conceal or obfuscate the content they use, unlike other domains where behaviour can be more strategic[28]. Social networks are a prime example: people highlight what makes them look good and hide what bothers them. Consumption data about cultural content also reflects people's identity and personality (see Part 2.1, which talks about the Video Privacy Protection Act (VPPA) in the United States). For these reasons, and to maintain the trust between users and platforms, it is essential that this data not be used for marketing or concealed advertising purposes, but to personalise the service provided, in the interest of all of the parties involved and in compliance with the French data privacy law (*Loi Informatique et Libertés*).

[25] CONSEIL D'ÉTAT. *Le numérique et les droits fondamentaux*, septembre 2014
[26] CONSEIL NATIONAL DU NUMÉRIQUE, *Ambition numérique*, juin 2015.
[27] Sur cette polémique, voir par exemple : BOLESSE, Cécile. "*Après la polémique, Spotify éclaircit ses conditions d'utilisation*" 01net, septembre 2015.
[28] Cf. le chapitre « La révolution du web social : demain tous des people ? » dans CNIL. *Vie privée à l'horizon 2020*, *Cahier IP n°1*, novembre 2012.

[25] CONSEIL D'ÉTAT. *Le numérique et les droits fondamentaux*, September 2014
[26] CONSEIL NATIONAL DU NUMÉRIQUE, *Ambition numérique*, June 2015.
[27] For more about this controversy, see, for example: BOLESSE, Cécile. "*Après la polémique, Spotify éclaircit ses conditions d'utilisation*" 01net, SEPTEMBER 2015.
[28] Cf. the chapter entitled "The social web revolution: will we all be "celebrities" in the future?" in CNIL. *Privacy towards 2020*, IP Report No. 01, November 2012.

WHAT DATA-DRIVEN CONTENT AND USES FOR TOMORROW?

In the cultural and creative content markets, changes in the way data is used and mined is shaping a new landscape. The user experience is becoming more intimately personalised; the work and the user experience are becoming more intricately connected; the border between creation by the author and content generated by and for the user is becoming blurred. Certain major trends sweeping across the various sectors seem particularly robust and provide a rough idea of what might happen in the next few years.

Predictive technology is making it possible for the distributors of cultural content to balance out supply and demand, the first step towards a knowledge of each individual user.

EMPATHY AND CONTEXT AWARENESS + ALGORITHMS = MORE PERSONALISED PREDICTION

FEELINGS, MOODS AND EMOTION ANALYSIS + CONNECTED DEVICES AND SENSORS = AMBIENT SERVICES

MORE IMMERSIVE AND INTERACTIVE FICTIONS + CROSS-MEDIA DELIVERY = STRONGER ENGAGEMENT BETWEEN INDIVIDUALS AND CONTENT

The combination of data and hardware is radically changing the value chain in traditional sectors and bringing new business models to the fore.
Multi-source data and hyper-contextualisation collection raises the question of the neutrality and relevance of algorithms.

FOUR SCENARIOS EXPLORING POSSIBLE FUTURES FOR THE "DATA + CULTURE" COMBO:

> **HYPERMOVIE GENERATOR**
  (see extract on next page)
> **LIFE TRACKS** (see extract on page 15)
> **TRAUMDATA**
> **THE COLLECTIVE MOOD MASTER**

A COLLABORATIVE CREATIVE APPROACH, COMBINING DESIGN THINKING AND FORESIGHT, WITH THE HELP OF WEDESIGNSERVICES AND FIVEBYFIVE.

Two workshops were organized at CNIL premises, with the participation of a small group of people. We searched a balanced distribution of participants to ensure a diversity of skills and visions. These participants belonged to several categories of actors, experts and stakeholders:
▸ pure players offering new modes of digital cultural consumption,
▸ more established industries of content, creators and suppliers of new technologies (algorithms, connected objects, APIs, often embodied by startups),
▸ regulators,
▸ researchers and academics.

If theses collaborative workshops were at the core of our scenario building process, other tools (documentary research or qualitative interviews) were combined to generate the resources and templates necessary for a shared creative work.

These workshops had a very exploratory objective: to imagine some creative and cultural industry uses and services in the next five years and create the first brick of potential data-driven business models.

PHASE #1 ▸ **Decrypt the current economic models and their breaking pointse**
PHASE #2 ▸ **Map this current state of the art to better imagine future scenarios**
PHASE #3 ▸ **Prototype some economic models with a 5-year horizon**

EXAMPLE: CREATIVITY TEMPLATES COMPLETED DURING THE SECOND WORKSHOP

Template 1 - Persona Board (available in appendix)
For the exercise to be as close as possible to the future expectations of cultural content users, participants had to step into the shoes of a typical user 5-year from now. They had to define and analyzed its consumption of cultural content, its needs, its motivations and reluctances to share personal data.

Template 2 - Project Board (available in appendix)
These canvas enabled participants to formalize the service they imagined, while drafting some market points (investment, data needed to operate the service and revenue model).

SECOND WORKSHOP PARTICIPANTS (FULL LIST IN APPENDIX)

SCÉNARIO 1 /

HyperMovie Generator

« Everyone can be a great movie creator »

HyperMovie Generator promises a world where "everyone can be a great movie creator".
The service makes it very easy for people to create original video content, using a content library and their own personal creations. The system helps people explore different creative avenues, based on a genre, a character, a subject and/or an audience. From a functional point of view, the service is supplied with data on content and analyses of audience preferences, which *HyperMovie Generator* obtains through its agreements with film and tv shows brokerage platforms, which hold large aggregate databases on preferences.

The business model is based on a free application combined with a premium subscription package that provides access to several additional service levels.
First, the paid subscription provides unlimited use. Then, it enables users to share content more widely, thanks to agreements with multi-channel networks (MCNs) on video platforms (*YouTube, DailyMotion*) that let hundreds of thousands of creators offer their content to hundreds of thousands of viewers. *HMG* negotiated deals with several of these MCNs to act as a "talent scout" for them. An MCN will then sign up the selected creators so that they can each manage their own distribution channel while using the MCN's promotion, marketing and especially sales tools. *HyperMovie Generator* is in fact the "first level" of the rocket. If the power users at HMG manage to produce engaging content that attracts positive responses, they are offered a chance on a MCN suited to their style (intellectual, youth, reporting, music clips, etc.). Even if very few manage to make a living out of their productions, the handful of success stories is more than enough to attract budding creators to *HyperMovie Generator. HMG* even manages to attract creators who are by no means amateur but who appreciate free access to resources (both production resources and sales resources) that would normally only be available through major producers and distributors.

The targets are therefore established and emerging authors and creators. Studios constantly on the lookout for blockbusters are also a prime target audience, insofar as *HyperMovie Generator* makes it possible to test concepts and narrative formulas. *HyperMovie Generator* owes its success to the fact that it seems to have bridged the gap between user-generated content and the biggest production studios.
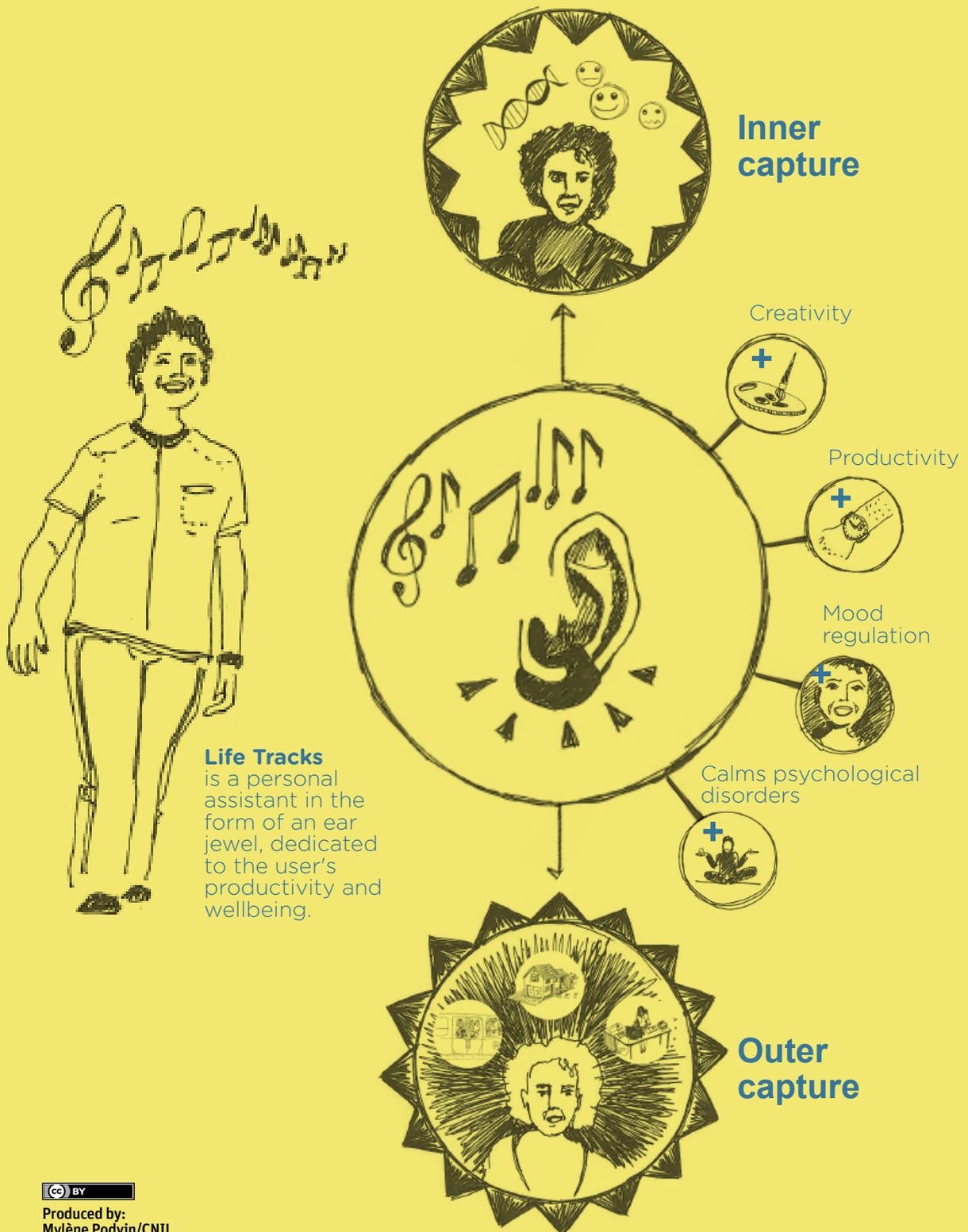
HMG has also become a specialist in dynamic content adaptation and now advises the big studios on how to incorporate discrete narrative adjustments in their blockbusters, or contextual references to specific audiences. For example, digital technology can be used to change the location of certain film scenes so that each audience can relate to them more easily: American movie-goers will see a scene against a New York skyline, while Chinese viewers will recognise the city of Shanghai.

---

⑤ REVENUES HyperMovie Generator derives the bulk of its revenue from charging for the application's audience and from a share in the subsequent revenues of content negotiated by contract between the author, HMG and the distributor. More and more brands are turning directly to HMG to create "brand content", i.e. content linked to the brand and which is subsequently sold as a "turnkey" toolbox for reuse by the authors.

🏃 CHALLENGES The difficulty consists in capturing and massively aggregating extremely granular data on the audiences and their preferences. The aim is therefore to define the incentive mechanism that will prompt viewers to provide information about themselves and help content creators get to know them better.

💡 INSPIRATIONS

▸ *Personalised books and automatic creation:* LostMyName
▸ *L'Expansion* "YouTube : les MCN ou réseaux de chaînes, accélérateurs d'audience", 2014
▸ *Lagniappe – livraison dix-huit, newsletter by Nicolas Nova dated 28 June 2015*
▸ *Popular highlights* - a practice that consists in the reader highlighting his favourite passages and the producer collecting and storing these short extracts. It can be a source of inspiration for imagining a possible future of literature. The designer Nicolas Nova talks about the *Networked Optimization* project led by two artists, Silvio Lorusso and Sebastian Schmieg, with help from "Amazon Kindle users". They produce paper books with a novel feature: they only display the phrases highlighted by their readers. The artists hope it will make people think about ways to optimise writing, and point out that, through highlighting, "the act of reading becomes a data-mining process".

Inner capture

Creativity

Productivity

Mood regulation

Calms psychological disorders

Outer capture

Life Tracks is a personal assistant in the form of an ear jewel, dedicated to the user's productivity and wellbeing.

**Produced by:
Mylène Podvin/CNIL**

Life Tracks

The Life Tracks application would combine the user's genetic profile and emotional state to deliver personalised musical content to "soundtrack" the user's activities.

Read IP Report No. 03 in
its entirety: visit **www.cnil.fr**

IP REPORTS

DATA, MUSES AND BORDERS
OF CREATIVE ARTS
READING, LISTENING, WATCHING AND GAMING
IN THE AGE OF PERSONALISATION

CNIL.

CNiL.
Commission Nationale de l'Informatique et des Libertés