



**HAL**  
open science

# Arts de la scène et Big Data. Retracer et analyser le processus de création d'un spectacle grâce à la visualisation de données

Clarisse Bardiot

## ► To cite this version:

Clarisse Bardiot. Arts de la scène et Big Data. Retracer et analyser le processus de création d'un spectacle grâce à la visualisation de données. H2PTM, Oct 2015, Paris, France. pp.231-242. hal-02337596

**HAL Id: hal-02337596**

**<https://hal.science/hal-02337596>**

Submitted on 29 Oct 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

---

# Arts de la scène et Big Data

Retracer et analyser le processus de création d'un spectacle grâce à la visualisation de données

Clarisse Bardiot

Laboratoire DeVisu, Université de Valenciennes et du Hainaut-Cambrésis  
Campus Mont Houy, 59313 Valenciennes Cedex 9  
clarisse.bardiot@univ-valenciennes.fr

---

RÉSUMÉ. Au théâtre, le Big Data en est officiellement à ses débuts. La première conférence explicite sur le sujet, en 2013, concerne les archivistes et les chercheurs, lesquels sont confrontés au Big Data des collections et des fonds d'archives numérisés. Il est néanmoins important de considérer les documents numériques produits et conservés par les artistes et leurs équipes. Ils recèlent aujourd'hui les traces principales des spectacles et de leurs processus de création. Comment retracer un processus de création à partir de centaine, voire de milliers de documents ? En quoi la visualisation de données peut-elle contribuer à l'analyse du dossier génétique d'une œuvre ? La première partie de l'article revient sur des expériences de distant reading dans les recherches en arts de la scène. La seconde partie examine plus particulièrement la création de ReCall, un environnement open-source pour documenter, analyser les processus de création et simplifier la reprise des œuvres.

ABSTRACT. The application of Big Data to performing arts is officially in its early stages. The first conference explicitly dedicated to the issue, given in 2013, focused on archivists and researchers faced with Big Data in digitized collections and archival funds. It remains nonetheless as important to consider digital material produced and kept by artists and their support teams. Today, they contain the main footprints of performances and creative processes. How can a creative process be reconstructed from thousands of documents? How data-visualization can contribute to the analysis of a creation process ? The first part of this paper analyses distant reading experiments in performing arts research. The second part describes the creation of ReCall, an open-source environment to document, analyse the creative process and simplify the re-enactment of performances.

MOTS-CLÉS : Big Data, arts de la scène, visualisation de données, distant reading, processus de création.

KEYWORDS: Big Data, performing arts, data-visualization, distant reading, creation process.

---

### 1. Arts de la scène et Big Data : des préliminaires incertains

Depuis 2005, Lev Manovich consacre ses recherches aux « cultural analytics »<sup>1</sup>, soit « *the use of computational and visualization methods for the analysis of massive cultural data sets and flows* » (Manovich, 2014). D'après lui, « *Digitization of large sets of cultural artifacts from the past and the rise of social media in 2000s open new possibilities for the study of cultural processes* »<sup>2</sup>. Les études réalisées concernent tant l'ensemble des œuvres de Mondrian ou de Rothko que les jeux vidéo ou encore l'évolution graphique des couvertures de Time Magazine depuis sa création. L'enjeu est ni plus ni moins de reconsidérer ce que nous entendons par « culture » ainsi que les méthodologies qui sont appliquées à ce champ (Manovich, 2013). Le Big Data est à l'origine de ce bouleversement dans l'étude des phénomènes culturels, lequel s'inscrit dans le champ plus large des Digital Humanities.

Dans le domaine des arts de la scène, la rencontre avec le Big Data relève davantage des préliminaires incertains que de la relation régulière et établie. La création d'un groupe « Digital Humanities in Theatre Research » à la FIRT (Fédération Internationale pour la recherche Théâtrale) en 2014<sup>3</sup> montre que l'utilisation de technologies numériques pour la recherche théâtrale, afin de donner notamment de nouvelles perspectives historiographiques, est une prise de conscience certes récente mais croissante. La première communication explicite sur le sujet qui nous occupe ici, intitulée « Big Data and the Performing Arts », a eu lieu le 9 novembre 2013 dans le cadre du colloque annuel de l'American Society of Theatre Research<sup>4</sup>. Elle est le fait de Doug Reside, conservateur numérique au département des arts de la scène à la New York Public Library de 2011 à juillet 2014. Lui-même avait lancé un appel à communication ainsi formulé :

*The emergence of large digitized collections of humanities resources has made it possible to meaningfully address research questions that previously would have taken many lifetimes to answer. However, theater historians have undertaken relatively little of this kind of work. Despite*

---

<sup>1</sup> . Le terme a été créé par Manovich en 2007.

<sup>2</sup> . <http://nycdh.org/members/manovich/> (page consultée le 16 septembre 2014).

<sup>3</sup> . L'initiative de ce groupe a été prise lors de la conférence annuelle de la FIRT à Barcelone, en juillet 2013. Les travaux de Derek Miller sur Broadway ont notamment été présentés dans ce contexte, tout comme d'autres recherches en cours.

<sup>4</sup> . La conférence de l'ASTR (Dallas, 7-10 novembre 2013) a été accompagnée par l'organisation d'un THATCamp les 20 et 21 juin 2013 à New York, intitulé « Performing Arts: Presenting and Archiving Performance in the Digital Age », et par une Unconference intitulée « DH & Theatre Research » à Dallas, le 8 novembre 2013. Cf. <http://performingarts2013.thatcamp.org/> et <http://dhandtheatresearch.wordpress.com/>.

*large datasets of digitized theater reviews, industry news, and production information [cast lists in Playbill Vault or Internet Broadway Database], theater scholars have by and large continued to do close readings of texts and events – and have not yet attempted what Franco Moretti has called distant reading: analyzing not one small set of texts, but an entire corpus of digitized data. (Reside, 2013)*

Doug Reside appelle de ses vœux une révolution similaire à celle opérée dans les études littéraires, sous l'impulsion de différents chercheurs, dont Franco Moretti. Soit sortir de l'étude d'un texte, d'un roman, voire d'un auteur, pour embrasser des corpus composés de centaines d'écrits ; s'intéresser à « *la configuration dans son ensemble, et non une de ses phases* » (Moretti, 2008, p. 46). Pour Moretti, la « lecture de loin » (*distant reading*) est « *une forme spécifique de connaissance* » qui permet de passer « des textes aux modèles » (Moretti, 2008, p. 33). Il s'agit de voir, d'observer, de regarder les relations entre des milliers de documents et non plus de faire une lecture attentive du contenu de chacun d'entre eux. La visualisation remplace alors la lecture. Comme le constate Bruno Bachimont :

*Le contenu documentaire intégré dans le corpus n'est plus considéré comme un contenu, mais comme un fait, une donnée syntaxique et symbolique, indépendamment de toute interprétation qu'on pourrait en faire, que l'on peut alors comparer et mesurer aux autres composantes du corpus. Son intelligibilité n'est plus acquise par la lecture et l'interprétation qu'on peut en faire mais découle désormais de la contribution qu'il apporte comme fait à l'analyse globale du corpus auquel il appartient. (Bachimont, 2014, pp. 70-71)*

## 1.1. Arts de la scène et distant reading

### 1.1.1. *Le close reading privilégié par les chercheurs, malgré la numérisation des fonds d'archive*

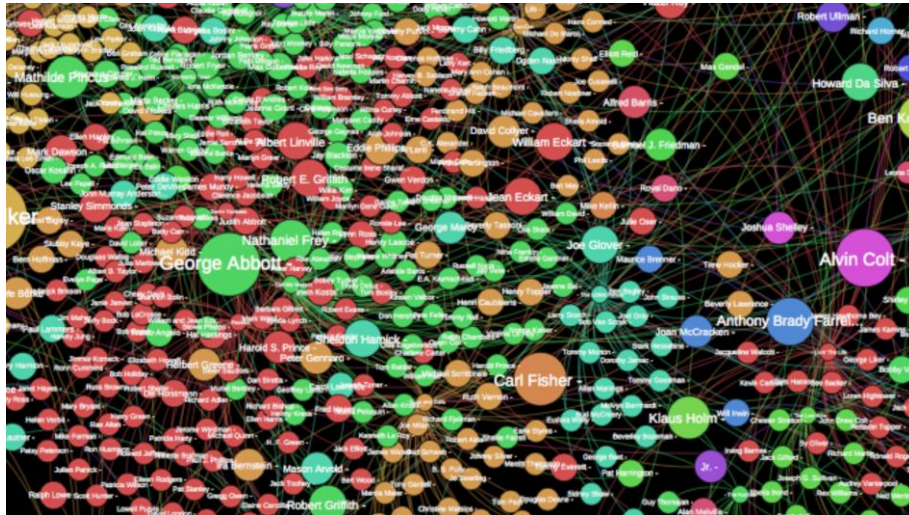
Doug Reside souligne avec justesse que dans le domaine des arts de la scène, les corpus sont déjà existants, en particulier grâce aux différentes vagues de numérisation des fonds d'archives. En Europe, citons les campagnes de numérisation liées au programme ECLAP (European Collected Library of Artistic Performance) qui a permis de numériser plus d'un million de documents issus de 32 institutions. Autre exemple : la numérisation (en cours) de l'ensemble des archives du théâtre national irlandais, l'Abbey Theatre, soit là aussi la mise à disposition sous format numérique de plus d'un million de pages couvrant 110 ans d'histoire du théâtre en Irlande. Ces campagnes de numérisation ont largement été conçues dans l'optique de conserver des documents fragiles (par exemple les films sur supports magnétiques) et de les rendre accessibles au plus grand nombre via des plateformes en ligne, dans un souci de diffusion de la culture. Pourtant, la numérisation des données ne modifie

#### 4 Nom de l'ouvrage

pas les pratiques courantes des chercheurs en arts de la scène, adeptes du *close reading*. Il s'agit pour eux d'une commodité et non pas d'un changement radical de méthodologie. On constate le même phénomène dans des projets de documentation numérique des arts de la scène, comme par exemple les travaux entrepris par le chorégraphe William Forsythe dès les années 1990 avec *Improvisation Technologies*, puis *Synchronous objects* et *Motion Bank* (Allsopp *et al.*, 2007 ; deLahunta *et al.*, 2013). S'ils ont ouvert des perspectives passionnantes sur la manière dont les technologies numériques peuvent aider à l'analyse des spectacles, la plupart d'entre eux sont dans la lignée du *close reading*.

##### 1.1.2. Travaux pionniers en *distant reading* dans la recherche en arts de la scène

Et pourtant, les fonds numérisés portent aussi en eux la promesse du *distant reading*. Ils permettent de déplacer le regard, de l'éloigner de la page, de la représentation, du metteur en scène, de l'acteur, pour élargir le champ des arts de la scène et faire apparaître des relations, des faits, qui ne peuvent surgir qu'avec ce changement d'échelle. Ainsi pour Derek Miller (Harvard), qui conduit le projet *Visualizing Broadway*, « *Digital computing techniques that process large amounts of data, then display results in multiple visualizations, can help scholars see the field of theatrical production in its rich complexity, without relying on anecdote* » (Miller, 2014). Influencé par les travaux de Franco Moretti, Derek Miller analyse l'histoire de Broadway, soit plus d'un siècle de productions théâtrales à New York, à partir de l'ensemble des données disponibles (durée de l'exploitation d'une pièce, nombre de reprises, genres, casting, thèmes...) en prenant en compte l'ensemble des productions (donc aussi et surtout les échecs) et non pas seulement les succès (esthétiques, commerciaux) à partir desquels l'histoire de Broadway a jusqu'ici été écrite. Cela lui permet par exemple de montrer que les personnes centrales du point de vue de la production théâtrale ne sont pas les grands noms que l'histoire a retenus (metteurs en scène, acteurs, producteurs) mais les scénographes, les costumiers, les régisseurs (fig. 1). Surtout, car cette information était déjà connue, cette analyse permet de montrer les réseaux et les circulations entre les différentes équipes de manière très précise et de répondre à des questions comme par exemple : à combien de productions un scénographe contribue-t-il en moyenne par saison ; travaille-t-il avec quelques personnes attirées ou au contraire de manière très éclectique ; est-il lié à un lieu spécifique ou bien au contraire passe-t-il d'un théâtre à un autre, etc. D'où un appel de la part de Derek Miller à reconsidérer les métiers et les parcours de ces créateurs de l'ombre. De la même manière, et dans la lignée des travaux actuels sur les *complex networks*, en se basant sur les critiques publiées dans les journaux, il fait apparaître les différents thèmes traités dans les productions de Broadway, leurs connexions et leur évolution au cours d'une période donnée.



**Figure 1.** Derek Miller, *Visualizing Broadway*.

Extrait de la visualisation du réseau des professionnels dans les années 1950  
(Miller, 2014).

Un autre projet en cours, *Les registres de la Comédie Française*, initié dès 2007 par Jeffrey Ravel (MIT), consiste à numériser les registres journaliers de la Comédie-Française de 1680 à 1793 pour en extraire des connaissances scientifiques grâce à la mise en place d'outils d'analyse statistique et de visualisation des données (Sthul, 2014). Plusieurs universitaires sont invités à utiliser ces données pour en tirer différentes analyses, différentes interprétations. Car l'un des intérêts de ces campagnes de numérisation est de permettre à plusieurs chercheurs de travailler sur un même corpus : indépendantes du chercheur<sup>5</sup>, les données sont combinables et recombinaisons à l'infini. Toute la question est alors de définir l'angle de vue, la direction du regard, la méthode de la fouille.

<sup>5</sup> . Pour Franco Moretti, cette indépendance des données vis-à-vis du chercheur est un gage d'objectivité (Moretti, 2008, p. 51). Rappelons en effet qu'une grande part de la recherche en études théâtrales est consacrée au recueil des archives, voire à leur constitution. On comprend combien les démarches évoquées ici déplacent alors les enjeux de la recherche.

## 1.2. Born Cultural Heritage

### 1.2.1. *Les documents nativement numériques des arts de la scène*

Comme on vient de le voir dans les exemples précédents, dans le domaine des arts de la scène, les recherches liées au Big Data concernent essentiellement des corpus de documents numérisés. Peu s'intéressent encore à l'exploitation des documents numériques produits par les artistes et leurs équipes, ce que les anglosaxons nomment le *Born Cultural Heritage*. Leurs disques durs contiennent des milliers de documents et de données, souvent très difficiles à identifier, à classer et encore plus à conserver. Cependant, ils recèlent aujourd'hui les traces principales des spectacles (captations vidéo, photographies, textes, revues de presse d'articles publiés sur internet, etc.) et de leurs processus de création (carnets de notes numériques, photographies et vidéo numériques des répétitions, documents numériques divers témoignant de recherches préparatoires, programmes informatiques élaborés pour le spectacle, etc.). De plus, aujourd'hui, au-delà du phénomène particulier des Digital Performances – soit des œuvres scéniques à composantes technologiques – toutes les régies techniques sont numériques, et une partie du processus de création a largement lieu via les ordinateurs et les réseaux : échanges de mails, traitements de texte, rendez-vous à distance via des dispositifs de téléprésence (voix sur IP), images et vidéos numériques pour rassembler des idées, des pistes de travail, usage des réseaux de partages d'images pour mettre à disposition des documents visuels pour l'ensemble de la compagnie, croquis effectués sur tablette numérique, etc. Ces documents numériques ne sont pas exclusifs et complètent d'autres types de documents (à numériser) : maquettes, éléments de hardware, éléments du spectacle (costumes, décors...), carnets de notes manuscrits, etc.

### 1.2.2. *Medium Data et étude génétique des œuvres*

Les artistes sont les premiers conservateurs de ces documents, à la fois numériques et matériels. Cela pose de nombreux problèmes, notamment pour ce qui concerne la pérennité des œuvres elle-même, remise en cause par l'obsolescence programmée des technologies. Nous n'aborderons pas ce point ici et nous nous intéresserons uniquement à la question de l'étude génétique des œuvres dans le contexte du Big Data, ou plutôt du Medium Data. Le Medium Data est un intermédiaire entre les « quatre V » du Big Data et le « Small Data », dont les données sont analysables individuellement, en close reading. Il désigne des corpus importants, qu'il devient quasiment impossible de traiter humainement individuellement, mais qui cependant ne rencontrent pas exactement les mêmes problématiques que les milliards de documents du Big Data. Quel type d'analyse peut-on effectuer ? Comment cheminer dans ces centaines, parfois ces milliers de documents pour reconstituer le processus de création ? En quoi la visualisation de données peut-elle contribuer à l'analyse du dossier génétique d'une œuvre ? En comparant différents processus, est-ce possible de repérer des formes, des modèles –

pour reprendre la terminologie de Franco Moretti ? Autrement dit, l'étude de la génétique des œuvres avec des outils numériques permet-elle de faire apparaître des récurrences en fonction des genres artistiques (théâtre, danse...), des méthodes (improvisation/texte pour prendre un exemple simple), des époques, des conditions de production, etc. ?

Pour mener à bien une telle analyse, l'élaboration d'un outil spécifique s'impose. En effet, la question du numérique dans les arts de la scène recouvre différents aspects qui dépassent la question de l'analyse quantitative stricte et donc de l'emploi de logiciels déjà existants dans le domaine du Big Data. Ces autres aspects sont les suivants :

- l'obsolescence rapide des technologies numériques a pour conséquence la nécessité de la documentation des œuvres afin de les mettre à jour tout en respectant les intentions artistiques initiales ;
- les documents doivent pouvoir s'articuler de manière chronologique pour commenter une œuvre en fonction du déroulement du spectacle et ainsi réaliser des « documentaires augmentés » autour d'une œuvre (pour donner un exemple concret, agréger les documents autour d'une captation vidéo d'une mise en scène) ;
- les artistes et leurs équipes (premiers conservateurs des œuvres et des documents) doivent pouvoir utiliser cet outil, de manière collaborative, pendant le processus de création lui-même avec la possibilité d'ajouter des notes de différentes natures (sur les documents ou sur une *timeline*) ;
- cet outil doit être en mesure d'agréger les documents nativement numériques et les documents numérisés ;
- la représentation du processus de création ne doit pas effacer sa complexité.

Autrement dit, l'outil en question doit non seulement permettre de documenter les processus de création mais également les œuvres elles-mêmes et aider à leur conservation, à leur entrée au répertoire, voir à leur reprise. Il doit offrir un va-et-vient entre *close reading* et *distant reading*, entre articulations diachronique et synchronique.

## 2. Rekall, un environnement pour visualiser les processus de création

C'est dans ce contexte que s'inscrit le développement du logiciel Rekall, actuellement disponible en version bêta et en *open source*. J'ai initié ce projet dès 2007 dans le cadre du programme DOCAM sur la « Documentation et la Conservation des Arts Médiatiques » mené par la Fondation Daniel Langlois à Montréal <sup>6</sup>, puis ai travaillé en collaboration avec Buzzing Light et Thierry Coduys

---

<sup>6</sup> . Initialement intitulé Éclats, le projet a été rebaptisé Rekall en 2013.



pour le développement et la conception actuelle à partir de 2012. Dans un premier temps, ReCALL a été conçu pour les arts de la scène, en prenant en compte le processus de création, la réception et les différentes formes d'une œuvre. En effet, il n'est pas rare qu'un spectacle connaisse différentes versions (cas du *Work in progress* par exemple) et soit, à l'heure du transmédia, décliné sous forme d'installation, d'application ou encore de film. Deux objectifs principaux ont guidé la conception de ReCALL : aider les artistes à documenter leurs œuvres pour pouvoir en assurer la reprise et palier à l'obsolescence des technologies numériques ; aider les chercheurs à étudier la génétique des œuvres dans un contexte de documents nativement numériques et de Medium Data.

Ces deux objectifs, tout comme les différents aspects évoqués ci-dessus peuvent sembler *a priori* très éloignés les uns de autres. Cela est vrai en termes d'objectifs, même s'ils peuvent s'avérer complémentaires, mais faux du point de vue de la « matière première » : quel que soit l'objectif poursuivi, ce sont les mêmes matériaux (les documents issus du processus de création et les traces des œuvres) qui sont collectés, analysés, visualisés, selon des modalités et des temporalités différentes. Ainsi, la collecte s'opère pour les artistes pendant le processus de création lui-même, au fur et à mesure des étapes de travail ; il s'agit d'une archive vivante, en mouvement perpétuel. Le chercheur intervient une fois le processus achevé, une fois que les documents se sont en quelque sorte stabilisés, figés, et qu'ils deviennent les traces de ce qui a eu lieu.

## 2.1. Traitement du Medium Data dans ReCALL

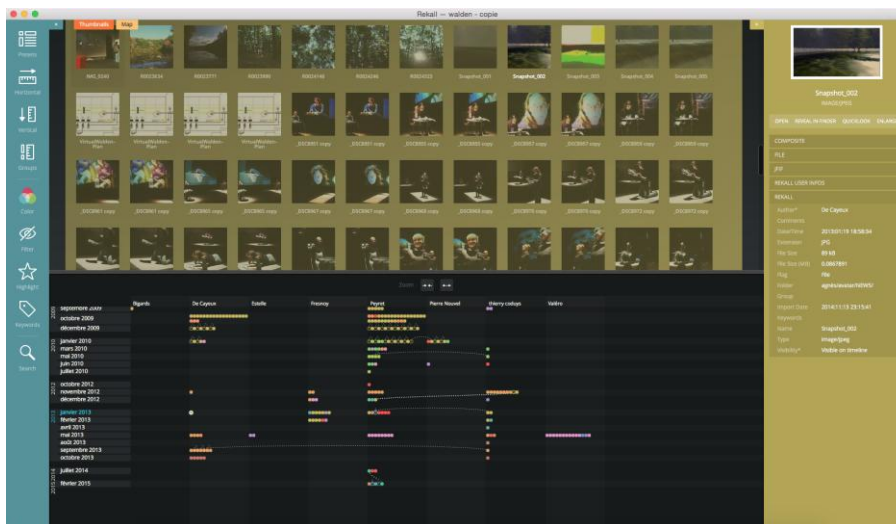
Afin de répondre aux différents objectifs cités plus haut, ReCALL présente deux environnements de travail : visualisation de données et timeline. Le mode timeline ne fait appel qu'à une petite sélection des documents. Il est conçu essentiellement dans une perspective d'annotation des documents, afin de compléter les informations qui ne sont pas perceptibles et intelligibles via la visualisation de données. Nous présenterons donc ici plus spécifiquement le mode visualisation de données, spécifiquement dédié à la partie Medium Data.

Dans les paragraphes qui suivent, ainsi que dans les figures, nous nous appuyons sur l'analyse en cours de *Re :Walden* de Jean-François Peyret, œuvre créée de 2009 à 2013 et qui a connu différentes versions (spectacle, concert, installation). Le corpus comprend plus de 2500 documents. Une petite portion a été utilisée pour réaliser une première phase de tests avec le logiciel et en éprouver les fonctionnalités. Cette œuvre a permis la réalisation d'une première étude de cas et a accompagné les différentes étapes de développement du logiciel dès ses prémices.

### 2.1.1. Un environnement multimodal

ReCALL permet d'agrèger dans une même interface un nombre infini de documents de différentes natures ayant trait à une œuvre. ReCALL rassemble dans un même

espace multimodal des textes, des images, des sons, des vidéos, des programmes, etc. Chaque point visible dans l'interface représente un document. Qu'ils soient indispensables à la représentation de l'œuvre (partition, patch, bande son,...) ou utiles à la compréhension du processus de création (photo de repérage, commentaire, enregistrement de répétition, plan du lieu de représentation...), ReKall offre une vision synthétique de la quantité, de la qualité et de l'organisation des documents entre eux. À gauche, un menu permet de choisir les différentes options de visualisation et de filtres. Au centre, la partie supérieure offre une vue d'ensemble des documents sous forme de vignette. Lorsqu'un document est sélectionné, il s'affiche dans cet espace. La partie inférieure est l'espace réservé à la visualisation de données, avec un affichage des documents selon les axes x et y. À droite, la colonne affiche les métadonnées du document sélectionné (fig. 2).



**Figure 2.** Interface principale de ReKall en mode visualisation de données.

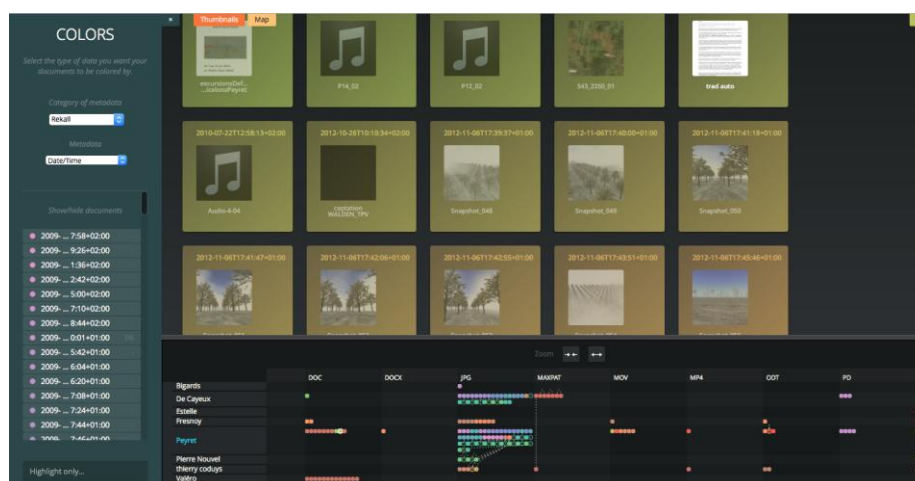
### 2.1.2. Une visualisation multidimensionnelle

L'interface de ReKall propose différents modes d'affichage des données, selon trois axes (horizontal, vertical et couleur), ce qui permet de trier et d'analyser les documents à travers une visualisation multidimensionnelle. En plus des axes x et y, la couleur est la troisième dimension de visualisation des documents. Elle permet par exemple de visualiser en un clin d'œil la diversité des documents présents dans le projet (les vidéos en rouge, les images en vert, etc.). Différents modes de visualisation sont préenregistrés et proposés à l'utilisateur mais celui-ci est libre d'en inventer d'autres.

## 10 Nom de l'ouvrage

Il est par exemple possible de voir à quelle date les différentes personnes impliquées dans le projet ont créé ou modifié des documents (fig. 2). Sur l'axe horizontal figurent les noms des créateurs des documents. Sur l'axe vertical la date, par ordre chronologique. La couleur représente le type de document. Ce mode de visualisation permet de comprendre à quel moment du processus de création interviennent l'éclairagiste, le scénographe, les comédiens, etc. Étaient-ils tous présents lors des différentes phases de résidence ? Ou bien au contraire ces dernières se sont-elle davantage élaborées en fonction de contraintes techniques ou de problématiques artistiques spécifiques ?

Changeons la manière d'afficher ces mêmes documents (fig.3). En x figurent les extensions de fichier, ce qui permet de repérer très facilement les logiciels utilisés dans une production ; en y les auteurs des documents ; la couleur rend compte de la date de création, dans un nuancier où les couleurs les plus froides sont les dates les plus anciennes et les couleurs les plus chaudes les couleurs les plus récentes. Cette vue permet de comprendre plus précisément quelles sont les attributions et les compétences de chaque personne.



**Figure 3.** Autre visualisation des mêmes documents, avec menu d'affichage des couleurs.

D'autres fonctionnalités permettent de filtrer les documents, et de ne travailler par exemple que sur le corpus produit par un auteur spécifique, ou encore de les grouper ou de mettre une sélection en surbrillance. Cette dernière fonctionnalité est par exemple très utile pour faire apparaître l'évolution d'un concept au cours d'une création.

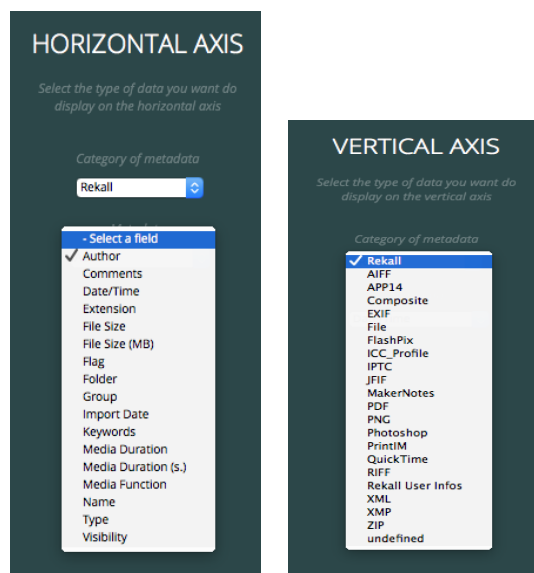
## 2.2. Extraction des métadonnées

Rekall se base sur les métadonnées présentes dans chacun des documents pour en extraire des informations cruciales (auteur, date de création, lieu de création, mot-clé, etc.). Ces informations sont ensuite utilisées par les outils d'analyse et de visualisation de l'information, afin de révéler des comportements, des usages. En jouant avec les axes et différents filtres qui proposent autant d'angles de vues sur une création donnée, les outils de Rekall permettent d'analyser graphiquement l'œuvre au travers de ses documents.

Toutes les opérations de tri reposent sur les métadonnées. Les métadonnées de chaque document sont affichées par catégorie (fig. 4). La catégorie Rekall permet de modifier les métadonnées sans altérer le document d'origine. En effet, l'un des principaux problèmes des métadonnées générées de manière automatique par les logiciels courants est qu'elles sont parfois erronées : en fonction par exemple de l'ordinateur sur lequel une personne a travaillé, ou bien lors de transferts de documents via des sites de partage, par mail ou via clé usb, le nom de l'auteur ou la date de création peuvent être altérés.

C'est pourquoi Rekall propose dans son interface la possibilité de changer les métadonnées des documents, soit un par un, soit en groupe, sans toutefois intervenir sur les métadonnées du document d'origine. Des systèmes de filtre et de visualisation permettent de repérer très vite les principales erreurs d'attribution et de date. La catégorie Rekall permet également d'ajouter des métadonnées, ce qui est ici une fonction d'annotation des documents. Nous avons ainsi par exemple ajouté un champ « commentaire » qui permet d'écrire du texte libre et un champ « mots-clés ». Ce dernier est très utile pour indexer les documents. Il permet par exemple d'attribuer les documents à des concepts propres à l'œuvre étudiée. De cette manière, il est alors possible de faire apparaître l'évolution d'une idée, le surgissement d'un thème, ou l'abandon d'une piste de travail au cours du processus de création.

À terme, Rekall prévoit un certain nombre d'alertes pour que l'utilisateur complète les métadonnées manquantes d'un document qui semble fondamental pour l'œuvre (absence de date, d'auteur, de version du logiciel utilisé, etc).



**Figure 4.** Fenêtre de sélection des catégories de métadonnées et de champ de métadonnées sur les axes horizontaux et verticaux.

### 3. Conclusion

Avec la multiplication des données nativement numériques engendrées lors des processus de création, le Medium Data est devenu une problématique pour les chercheurs en études théâtrales travaillant sur la génétique des œuvres contemporaines. Ceci entraîne une réflexion sur la création d'un outil *ad hoc*, qui puisse à la fois être utile aux artistes (premiers conservateurs de leurs œuvres) et aux chercheurs, ces deux groupes travaillant à partir des mêmes matériaux, mais dans des optiques différentes. Nous avons montré ici les principes qui ont été à la base de la conception du logiciel Rekall, pour la partie qui concerne la visualisation de données. Ce logiciel est en cours de création et il est encore trop tôt pour présenter des études de processus de création finalisées. Cependant, il permet déjà d'envisager un certain nombre de fonctionnalités utiles à l'étude génétique des œuvres. Pour le moment, Rekall permet d'étudier une seule œuvre à la fois, ce qui offre la possibilité de mener des études comparatives, soit chez un même artiste sur plusieurs œuvres, soit chez différents artistes. Dans un second temps de développement, il sera possible de regrouper au sein d'une même analyse plusieurs œuvres. Au vu des premiers tests effectués, Rekall, de par la richesse des modes de visualisation proposés, est en mesure de s'adapter à différentes esthétiques, à différentes méthodes de travail et d'en montrer les modalités précises.

#### 4. Bibliographie

- Allsopp R., deLahunta S. (sous la direction de), *Performance Research « Digital Resources »*, vol. 11, n° 4, 2007.
- Bachimont B., « Le nominalisme et la culture : questions posées par les enjeux du numérique », *Digital studies Organologie des savoirs et technologies de la connaissance*, Limoges, Fyp éditions, IRI, Institut de recherche et d'innovation, 2014, p. 63-78.
- deLahunta S., Whatley S. (sous la direction de), *International Journal of Performance Arts and Digital Media*, vol. 9, n° 1, 2013.
- Manovich L., *Software Takes Command*, New York, London, Bloomsbury Academic, 2013.
- Manovich L., *Software Studies: Cultural Analytics*, <http://lab.softwarestudies.com/p/overview-slides-and-video-articles-why.html> (page consultée le 16 septembre 2014).
- Miller D., *Visualizing Broadway: A Research Project in the Digital Humanities, IFTR/FIRT*, Warwick, <http://www.visualizingbroadway.com/Visualizations/Broadway/>, 2014.
- Moretti F., *Graphes, cartes et arbres: modèles abstraits pour une autre histoire de la littérature*, Paris, Les Prairies ordinaires, 2008.
- Reside D., *The Big D: Big Data and the Performing Arts*, *Theatre Library Association*, appel à contribution, <http://www.tla-online.org/> (page consultée le 20 février 2013), 2013.
- Stuhl A., 2014. Browsing Theater History with CFRP, *HyperStudio – Digital Humanities at MIT*, article de blog, <http://hyperstudio.mit.edu/blog/browsing-theatre-history/> (page consultée le 8 mai 2015), 19 décembre 2014.