

Titre: Outil d'aide à l'inspiration basé sur l'exploration d'une archive
Title: d'exemples de récits centrés sur les données

Auteur: Louri Glen Kae Compain
Author:

Date: 2024

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Compain, L. G. K. (2024). Outil d'aide à l'inspiration basé sur l'exploration d'une
Citation: archive d'exemples de récits centrés sur les données [Master's thesis,
Polytechnique Montréal]. PolyPublie. <https://publications.polymtl.ca/58333/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/58333/>
PolyPublie URL:

**Directeurs de
recherche:** Thomas Hurtut
Advisors:

Programme: Génie informatique
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Outil d'aide à l'inspiration basé sur l'exploration d'une archive d'exemples de
récits centrés sur les données**

LOURI GLEN KAE COMPAIN

Département de génie informatique et génie logiciel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
Génie informatique

Avril 2024

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Outil d'aide à l'inspiration basé sur l'exploration d'une archive d'exemples de
récits centrés sur les données**

présenté par **Louri Glen Kae COMPAIN**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Jinghui CHENG, président

Thomas HURTUT, membre et directeur de recherche

Philippe DOYON-POULIN, membre

REMERCIEMENTS

Ce projet a pu être mené à bien grâce au soutien de plusieurs personnes, que j'aimerais remercier.

Tout d'abord, je tiens à exprimer ma profonde gratitude envers mon superviseur de recherche, le professeur Thomas Hurtut, pour son soutien constant tout au long du projet. Je le remercie sincèrement de m'avoir guidé et conseillé durant mes travaux, et surtout pour sa grande patience tout au long de ces deux années.

Je souhaite également remercier les professeurs Jinghui Cheng et Philippe Doyon-Poulin, qui ont accepté d'agir à titre de membres du jury d'examen de ce mémoire.

Je tiens à remercier les équipes du Devoir et de Radio Canada qui ont généreusement accepté de contribuer à l'expérimentation. Ce sont leurs retours qui ont permis à ce projet de trouver sa vraie valeur.

Mes sincères remerciements vont au CRSNG (Conseil de recherches en sciences naturelles et en génie du Canada) pour la bourse d'études qui m'a permis de compléter ce programme.

Je suis reconnaissant envers ma famille et en particulier mes parents, Thierry et Brigitte Compain, pour leur encouragements et leur soutien du début à la fin de mes études. Enfin je veux remercier les amis des deux côtés de l'Atlantique qui ont été là pour moi quand je doutais de mon succès.

RÉSUMÉ

Les Récits Centrés sur les Données (RCD) sont un format narratif qui s'appuie sur la visualisation de données. Les articles de RCD nécessitent une démarche de conception poussée, qui fait émerger des défis supplémentaires par rapport aux articles plus traditionnels basés principalement sur du texte et des photos. Cette démarche de conception comprend une composante d'idéation pour générer des idées novatrices de représentations visuelles.

L'objectif de ce travail de recherche est de proposer une méthode visant à faciliter la conception de RCD en assistant les concepteurs dans la phase d'idéation. Pour cela, nous explorons l'utilité des exemples dans le processus d'idéation d'un concepteur. Notre intention est de créer un outil qui améliore l'accès des concepteurs à des exemples de RCD.

Pour cela, nous avons créé trois composantes qui seront ensuite implémentées sous la forme d'un outil, que nous appellerons dans ce mémoire CurratedDDS. La première est une taxonomie que nous avons proposée en adaptant les éléments de plusieurs taxonomies préexistantes, dans l'objectif de caractériser les RCD afin de permettre aux utilisateurs d'explorer les exemples. La seconde est une archive d'exemples de RCD que nous avons rassemblés à partir de compilations des meilleurs RCD produits par des journaux en ligne, puis annotés à l'aide de la taxonomie. La dernière composante est une interface qui permet à un concepteur de RCD d'explorer des exemples de l'archive à l'aide de la taxonomie.

Pour évaluer l'efficacité et l'impact de CurratedDDS pour le public cible, nous avons conduit une évaluation sous la forme d'une étude de cas sur un groupe de 6 professionnels des médias. Les participants ont testé l'outil dans le contexte de leur pratique professionnelle sur une période de deux semaines, puis nous avons collecté leurs retours sous forme d'entretiens semi-directifs.

Leurs retours nous permettent de conclure que notre taxonomie décrit les RCD de façon adéquate pour les besoins de nos utilisateurs, mais que certains des éléments de vocabulaire que nous avons employés ne correspondent pas à leur expérience. Ces retours établissent aussi que notre archive contient des exemples pertinents, mais comporte plusieurs limitations. Enfin, ces retours montrent que CurratedDDS a permis aux participants d'accéder à des exemples pertinents et a un effet positif sur leur processus d'idéation.

ABSTRACT

Data Driven Story (DDS) are a narrative format that relies on data visualization. DDS articles require an thorough design approach, which brings up additional challenges compared to more traditional articles, based primarily on text and photos. This design effort includes a significant need for ideation to generate innovative ideas for visual representations.

The goal of this research project is to create a method to help the design of DDS by assisting designers in the ideation phase. To achieve this goal, we explore the usefulness of examples in a designer's ideation process. Our intention is to create a tool that improves designers' access to DDS examples.

To accomplish this, we have created three components that will then be implemented in the form of a tool, which we will call CurratedDDS in this thesis. The first is a taxonomy that we proposed by adapting elements from several pre-existing taxonomies, with the aim of characterizing DDS in order to allow users to explore the examples. The second is an archive of DDS examples that we have gathered from DDS best-of produced by online newspapers, and then annotated using the taxonomy. The final component is an interface that allows a DDS designer to explore examples from the archive using the taxonomy.

To evaluate the efficiency and impact of CurratedDDS for the target audience, we conducted an evaluation in the form of a case study with a group of 6 media professionals. Participants tested the tool in the context of their professional practice over a two-week period, and then we collected their feedback through semi-structured interviews.

The participants' feedback allows us to conclude that our taxonomy adequately describes DDS for the needs of our users, but that some of the terms we used do not correspond to their experience. This feedback also establishes that our archive contains relevant examples, but has several limitations. Finally, this feedback shows that CurratedDDS provided participants with access to relevant examples and had a positive effect on their ideation process.

TABLE DES MATIÈRES

REMERCIEMENTS	iii
RÉSUMÉ	iv
ABSTRACT	v
TABLE DES MATIÈRES	vi
LISTE DES TABLEAUX	ix
LISTE DES FIGURES	x
LISTE DES SIGLES ET ABRÉVIATIONS	xi
LISTE DES ANNEXES	xii
CHAPITRE 1 INTRODUCTION	1
1.1 Problématique	2
1.2 Objectifs de recherche	3
1.2.1 Objectif général	3
1.2.2 Objectifs spécifiques	4
1.2.3 Hypothèses du projet	4
1.3 Plan du mémoire	4
CHAPITRE 2 REVUE DE LITTÉRATURE	6
2.1 Les Récits Centrés sur les Données dans le journalisme de données	6
2.1.1 L'origine et le rôle des Récits Centrés sur les Données	6
2.1.2 Décrire et représenter les Récits Centrés sur les Données	8
2.2 Le rôle des exemples dans l'idéation pour le journalisme de données et les Récits Centrés sur les Données	11
2.2.1 La créativité dans la visualisation de données	12
2.2.2 Les exemples contribuent à la créativité	12
2.3 Taxonomies des Récits Centrés sur les Données	14
2.3.1 <i>Narrative Visualization : Telling Stories with Data</i>	14
2.3.2 <i>Emerging and Recurring Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Recent Stories</i>	15

2.3.3	<i>Classifying Data Journalism</i>	16
2.3.4	<i>Narrative Design Patterns for Data-Driven Storytelling</i>	19
2.3.5	<i>Exploration and Explanation in Data-Driven Storytelling</i>	19
2.3.6	<i>Patterns in Award-Winning Data Storytelling</i>	21
2.3.7	<i>A Conceptual Framework For Interactive Cartographic Storytelling</i>	23
2.3.8	<i>Cartographic Design as Visual Storytelling : Synthesis and Review of Map-Based Narratives</i>	23
2.3.9	<i>Visualization design patterns</i>	25
2.3.10	<i>The Stories We Tell About Data : Media Types for Data-Driven Storytelling</i>	27
2.4	Synthèse	29
CHAPITRE 3	MÉTHODE PROPOSÉE	30
3.1	Taxonomie des Récits Centrés sur les Données	30
3.1.1	Taxonomies préexistantes	30
3.1.2	Comparaison des taxonomies existantes	41
3.1.3	Ajustements des dimensions et valeurs	49
3.1.4	Regroupements des dimensions	52
3.1.5	Version finale de la taxonomie	54
3.2	Archive	56
3.2.1	Collecte d'éléments pour l'archive	56
3.2.2	Annotation de l'archive	59
3.3	CurratedDDS	59
3.3.1	Principe de fonctionnement	60
3.3.2	Interface graphique	61
3.3.3	Côté serveur	65
CHAPITRE 4	ÉVALUATION	66
4.1	Protocole expérimental	66
4.1.1	Méthode d'évaluation	66
4.1.2	Sélection des participants	66
4.1.3	Présentation du projet et utilisation en contexte	67
4.1.4	Entretiens	68
4.2	Résultats et analyse	69
4.2.1	Retours sur la taxonomie	69
4.2.2	Retours sur l'archive	75
4.2.3	Retours sur CurratedDDS	78

4.2.4	Travaux futurs	81
4.2.5	Limitations de l'évaluation	86
CHAPITRE 5 CONCLUSION		89
5.1	Synthèse des travaux	89
5.2	Améliorations futures	90
RÉFÉRENCES		94
ANNEXES		97

LISTE DES TABLEAUX

Tableau 2.1	Récapitulatif de Classifying Data Journalism	17
Tableau 2.2	Correspondances entre dimensions et groupes dans la taxonomie de <i>Narrative Design Patterns for Data-Driven Storytelling</i>	19
Tableau 2.3	Dimensions de flexibilité dans la taxonomie de <i>Exploration and Explanation in Data-Driven Storytelling</i>	20
Tableau 2.4	Taxonomie de <i>Patterns in Award-Winning Data Storytelling</i>	22
Tableau 2.5	Taxonomie de <i>Cartographic Design as Visual Storytelling : Synthesis and Review of Map-Based Narratives</i>	25
Tableau 2.6	Taxonomie de <i>Visualization design patterns</i>	26
Tableau 2.7	Taxonomie de <i>The Stories We Tell About Data : Media Types for Data-Driven Storytelling</i>	28
Tableau 3.1	Dimensions de <i>Narrative Design Patterns for Data-Driven Storytelling</i> après le premier filtrage	34
Tableau 3.2	Tableau modifié à partir du tableau 3.1 des dimensions de <i>Narrative Design Patterns for Data-Driven Storytelling</i>	35
Tableau 3.3	Récapitulatif des dimensions extraite des taxonomies	40
Tableau 3.4	Comparaison des options de médias	41
Tableau 3.5	Comparaison des options de séquence	42
Tableau 3.6	Comparaison des options de visualisations de données	44
Tableau 3.7	Comparaison des options de séquence de visualisation	46
Tableau 3.8	Récapitulatif des fusions de catégories	48
Tableau 3.9	Récapitulatif de la taxonomie	55
Tableau 3.10	Sources identifiées	58
Tableau 4.1	Récapitulatif de la nouvelle taxonomie	84
Tableau A.1	Récapitulatif des définitions de groupes de la taxonomie	98
Tableau A.2	Récapitulatif des définitions de dimensions de la taxonomie	99
Tableau A.3	Récapitulatif des définitions de valeurs de la taxonomie (Partie 1) . .	100
Tableau A.4	Récapitulatif des définitions de valeurs de la taxonomie (Partie 2) . .	101
Tableau A.5	Récapitulatif des définitions de valeurs de la taxonomie (Partie 3) . .	102

LISTE DES FIGURES

Figure 3.1	Interface de l'outil CurratedDDS	62
Figure 3.2	Barre supérieure de l'interface	62
Figure 3.3	Barre supérieure de l'interface, avec toutes les options affichées	62
Figure 3.4	Compteurs pour les valeurs de la dimension "Type"	64
Figure 3.5	Présentation des exemples	65

LISTE DES SIGLES ET ABRÉVIATIONS

DDS	Data Driven Story
RCD	Récit Centré sur les Données
IPTC	International Press and Telecommunications Council

LISTE DES ANNEXES

Annexe A	Définitions des éléments de taxonomie	97
Annexe B	Guide d'entretien	103

CHAPITRE 1 INTRODUCTION

Le journalisme est souvent considéré comme l'un des piliers du bon fonctionnement d'une démocratie. À l'ère numérique, le format traditionnel (un article composé de texte et d'illustrations photographiques ou dessinées, imprimés sur du papier) n'est plus le médium unique du journalisme. L'avènement d'internet et des formats en ligne, puis la concurrence des médias sociaux, ont créé le besoin d'adapter les formats journalistiques à cet environnement. L'introduction de *The Data Journalism Handbook : How Journalists Can Use Data to Improve the News* annonce même la formule choc "Journalism is under siege" [1]. Le RCD est l'un de ces nouveaux formats, porteur d'une nouvelle approche, mais aussi de nouveaux défis.

Le journalisme de données peut être défini comme "Journalisme qui exploite, à des fins statistiques dans le but de mettre des faits en évidence et d'apporter des preuves de ces faits, les informations puisées dans les données ouvertes ou directement sur le Web." [2]. Un RCD peut alors être perçu comme une extension de ce principe, qui propose une importante composante de visualisation de données et d'interaction pour faire ressortir les données au premier plan de l'article, et les rendre plus accessibles et plus vivantes pour les lecteur. L'article *78 Long Minutes* [3], de Keith Collins, Albert Sun, Eleanor Lutz et Larry Buchanan, publié le 28 Mai 2022 en est un exemple, repéré pour sa qualité par l'un des journalistes participant à l'évaluation présentée dans ce mémoire. Après la tuerie d'Uvalde, aux États Unis, des accusations ont été portées contre la police pour son délai d'intervention démesuré. Cet article exploite les horodatages des communications de la police pour placer le texte de l'article autour d'une grande visualisation centrale : une ligne du temps. Cette méthode laisse de grandes zones vides là sur les périodes de temps qui ont suscité la controverse. L'exploitation de données temporelles permet ici de passer d'une déclaration écrite (l'article s'ouvre par "Police on the scene did not stop the gunman until well over an hour after he walked into the school, raising questions about whether lives could have been saved if they had acted sooner."), à une représentation visuelle concrète de ces 78 minutes.

Néanmoins, produire un RCD implique une démarche de design et de création différente d'un article plus classique contenant uniquement du texte et de la photographie, avec des défis qui lui sont propres. La forme de l'article doit être radicalement repensée pour s'adapter à son fond pour chaque nouveau projet. Pour reprendre l'exemple de *78 Long Minutes*, les auteurs ont eu l'idée de créer un format minute par minute, et de représenter chaque minute sans action avec un vide, puis ils ont dû construire une page web spéciale avec un format différent des articles habituels du New York Times, afin d'accommoder la visualisation centrale. L'une

des clés de ce processus est l'idéation, définie comme “le processus créatif de production, développement, et communication de nouvelles idées ” [4]. Dans notre contexte, il s'agit de la production d'idées nouvelles pour traiter et représenter les données afin de créer des RCD novateurs et pertinents. Faciliter cette démarche d'idéation permettrait donc d'aider les concepteurs de RCD dans la recherche d'idées innovantes, et de proposer à leur audience des projets de RCD plus vite et de meilleure qualité.

1.1 Problématique

Créer un RCD présente des défis uniques. Les concepteurs ont besoin de générer des idées innovantes pour mettre en valeur leurs données. L'une des ressources à leur disposition pour cette idéation est la veille professionnelle à travers la recherche et la consommation d'exemples. Quand une équipe se lance dans un projet de RCD, s'inspirer d'un projet pré-existant conduit par une autre équipe ou un autre média permet d'avoir une réflexion par rapport au traitement des données, à leur méthode de représentation [5], ou encore à la clarté du produit fini et à son accessibilité pour le public [6].

La recherche, l'observation et le partage d'exemples est donc une pratique importante dans le processus d'idéation et conception d'un RCD. Mais pour cela, il faut d'abord se procurer des exemples utilisables. Les concepteurs recherchent leurs exemples dans les médias qui publient des RCD, trient celles qui sont intéressantes pour leurs projets, et en retirent les idées qui leur permettent d'avancer. Cela représente du temps et des efforts.

Pour trouver les exemples, les concepteurs de RCD ont différentes méthodes. Des premières conversations avec les partenaires du projet de ce mémoire ont montré que leurs stratégies sont diverses, et sont basées sur des outils qui ne sont pas toujours conçus pour cette mission. Certains utilisent des archives personnelles, de longues listes d'articles enregistrés sur des sites tels que Pocket ou Pinterest, parmi d'autres contenus, avec un système de classification limité ou inadapté. D'autres reposent sur leur mémoire et se rappellent avoir vu un article intéressant quelques mois plus tôt, qu'ils vont ensuite chercher à partir de leur mémoire. Souvent ils emploient le sujet, des éléments du titre, et la source dont ils se souviennent ou croient se souvenir. Parmi les ressources plus formelles, les infolettres des journaux sont un moyen pour les concepteurs de RCD d'accéder à des exemples de projets. Les résultats de concours des meilleures visualisations de données, permettent de trouver des projets de grande qualité. Une recherche internet peut parfois faire ressortir des projets. Les départements spécialisés de certains journaux ont même des comptes sur les réseaux sociaux qui partagent leurs projets phares. Mais ces ressources sont très décentralisées, doivent être parcourues une à une et sont parfois redondantes. Un projet de journal peut paraître dans une infolettre, être promu sur

les réseaux sociaux et puis remporter un prix. La collecte d'exemples est donc une tâche complexe, reposant sur des ressources qui sont peu uniformes.

De plus, une autre problématique émerge, concernant la description des exemples de RCD. Un concepteur en train de rechercher un exemple pour un projet de RCD peut taper des mots clés dans un moteur de recherche, ou lire des titres dans une liste d'articles. Mais ces informations ne contiennent pas de vocabulaire concernant la démarche adoptée par les concepteurs originaux de l'exemple. À titre d'illustration, la structure narrative de type “Verre à Martini” est une méthode proposée par Segel et Heer en 2010 [7]. Les RCD construits selon cette structure ne contiennent pas la mention de cette méthodologie, et il est donc difficile de rechercher des exemples montrant comment plusieurs organes de presse ont construit leurs propres projets en utilisant la méthode du “Verre à Martini”. Il existe donc un besoin pour une méthode permettant de rechercher des exemples de RCD en utilisant un vocabulaire descriptif portant sur la forme de l'article et pas seulement le fond.

1.2 Objectifs de recherche

1.2.1 Objectif général

Le but de ce projet de recherche est de proposer une collection de ressources permettant d'améliorer l'accès des concepteurs de RCD à des exemples.

Une première étape pour adresser ce besoin serait de proposer une archive d'exemples disponible en ligne, une archive d'exemples de RCD, rassemblés depuis de multiples sources, organisés et annotés avec des informations supplémentaires. Le besoin suivant serait de rendre cet archive accessible, en proposant un outil conçu spécifiquement pour remplir cet objectif. Sous la forme d'une interface web, il permettrait d'explorer l'archive à l'aide d'un système de filtrage à facettes afin de faire ressortir les exemples les plus pertinents.

De plus, un tel outil permettrait d'introduire des méthodes de classification qui ne pourraient pas être employées dans un contexte plus général. En effet, dans un outil dédié aux RCD, il devient possible d'employer un vocabulaire spécifique à ce médium. Par exemple, proposer d'accéder aux projets utilisant la méthode du “Verre à Martini”. Ceci permettrait d'apporter un nouvel angle pour sélectionner des exemples pertinents en utilisant un vocabulaire qui décrit aussi les méthodes utilisées pour donner sa forme à l'article, plutôt qu'uniquement les éléments de fond.

Notre objectif serait donc de proposer trois produits interconnectés : Une taxonomie des RCD, une archive d'exemples de RCD annotée à l'aide de la taxonomie et un outil permettant d'explorer l'archive en filtrant à l'aide de la taxonomie pour y trouver les exemples les plus

pertinents. Notre hypothèse générale est donc que ces trois éléments pourront être utiles à des concepteurs de RCD lors de leur démarche d'idéation, afin de rassembler les exemples qui pourront leur servir d'outils de réflexion afin de produire de nouvelles idées de RCD. Ainsi, la création de RCD se verrait facilitée pour les concepteurs, permettant de mieux faire profiter leur public de RCD plus nombreux et de meilleure qualité.

1.2.2 Objectifs spécifiques

Les objectifs spécifiques de ce projet sont les suivants :

- O1 : Proposer une nouvelle taxonomie des RCD, qui permette de classer les RCD avec une large variété d'options de description, afin d'offrir des possibilités de tri variées.
- O2 : Rassembler une archive d'exemples de RCD pertinents, issus de sources de qualité, et annotés selon la taxonomie proposée.
- O3 : Concevoir un outil de recherche d'exemples de RCD qui permette à nos utilisateurs d'explorer l'archive pour y trouver des exemples, en triant son contenu à l'aide de la taxonomie. Nous appellerons cet outil CurratedDDS.

1.2.3 Hypothèses du projet

La réalisation de ces objectifs repose sur les hypothèses suivantes, qui seront évaluées lors de notre recherche :

- H1 : La taxonomie proposée caractérise adéquatement les RCD pour leur usage par des concepteurs de RCD.
- H2 : La taxonomie utilise un vocabulaire qui correspond à la réalité de la pratique des concepteurs de RCD.
- H3 : L'archive proposée contient une sélection d'exemple pertinente pour les concepteurs de RCD comparativement aux exemples auxquels ils sont exposés hors de l'archive.
- H4 : L'outil CurratedDDS améliore l'accès à des exemples pour des concepteurs de RCD, comparativement à leurs pratiques habituelles sans cet outil.
- H5 : L'utilisation de CurratedDDS facilite le travail d'idéation des concepteurs de RCD.

1.3 Plan du mémoire

Ce mémoire est divisé de la façon suivante : Le second chapitre propose une revue de la littérature portant sur les RCD et sur le rôle des exemples dans le processus d'idéation. Le

chapitre 3 détaille le processus de création et la structure de la taxonomie. Le chapitre 4 explique la méthodologie proposée pour concevoir l'outil CurratedDDS. Le chapitre 5 porte sur le protocole de l'expérience qui a été menée pour évaluer nos hypothèses en coopération avec une équipe de concepteurs de RCD. Le chapitre 6 présente et analyse les résultats de l'expérience. Pour finir, le chapitre 7 résume les grandes lignes de ce projet, présente ses limitations et propose des recommandations pour de futures études.

CHAPITRE 2 REVUE DE LITTÉRATURE

Ce chapitre porte sur la littérature préexistante qui a guidé la réalisation de ce projet. La première section présente les RCD, leur place dans le journalisme de données et leur intérêt pour les médias, ainsi que la façon dont ils sont représentés et modélisés. La deuxième section résume le rôle des exemples pour l'idéation lors d'un projet créatif, et en particulier dans le cas d'un RCD.

2.1 Les Récits Centrés sur les Données dans le journalisme de données

Un RCD tel que nous l'avons défini est un format narratif utilisé dans le journalisme des données. Cette section propose une revue de la littérature sur les RCD et la place qu'ils occupent dans le journalisme de données, mais aussi sur les méthodes qui ont été proposées pour les décrire et les classer.

2.1.1 L'origine et le rôle des Récits Centrés sur les Données

Les RCD s'inscrivent dans une convergence de l'évolution de la démarche journalistique et de l'émergence de l'aspect narratif dans les visualisations de données. Leurs spécificités ont mené les RCD à prendre une place unique comme format de journalisme de données et à devenir un outil de communication.

La place des Récits Centrés sur les Données dans le journalisme de données

L'introduction du livre *Data Driven Storytelling* [8] remarque une convergence entre la communauté de chercheurs en visualisation de données et la communauté des journalistes, qui se rejoignent autour de l'idée de Data Driven Story¹, une tendance qui s'inscrit dans le phénomène de démocratisation de l'accès aux données. Fu et Staskov [9] parlent d'ailleurs d'intersection Journalisme-Visualisation dont le sujet central serait le "Data storytelling".

The Data Journalism Handbook : How Journalists Can Use Data to Improve the News [1] explique que l'intérêt du journalisme de données réside dans une nouvelle profondeur d'analyse de l'information. L'analyse de données permet aux journalistes de quitter le rôle de premiers propagateurs de l'information (rôle pour lequel les réseaux sociaux ont mis fin à leur monopole) et de prendre le rôle d'interprétateurs de l'information.

1. Nous avons traduit Data Driven Story par Récit Centré sur les Données

Kosara et Mackinlay, auteurs de *Storytelling : The Next Step for Visualization* [10], remarquent que le “journalisme visuel” se distingue des autres modes d’utilisation du récit avec une visualisation de données. Le journalisme de données à une capacité à incorporer efficacement des graphiques et à impliquer le lecteur sans détourner son attention de la donnée.

Dans *Data stories : Rethinking journalistic storytelling in the context of data journalism* [11], Weber, Engebretsen et Kennedy expliquent que la visualisation permet aux journalistes de données d’implémenter des éléments et structures d’explication et d’argumentation dans une démarche de “montrer et ne pas dire”. Ils remarquent que le design des projets et le mode de présentation des données prennent une nouvelle place et introduisent l’idée qu’une “data story” avec un design visuel convaincant aura plus de succès qu’un article centré sur le texte pour attirer l’attention des lecteurs.

Dans la recherche académique en visualisation d’information, l’intérêt des chercheurs se porte aussi sur l’idée de raconter une histoire dans la visualisation pour augmenter son impact. En 2010, Segel et Heer [7] décrivent une classe de visualisation de données émergentes, combinant récit et visualisations interactives, qui apparaît dans les travaux des journalistes en ligne. Ces visualisations, qu’ils nomment Data Stories, sont des “visualisations avec des composantes narratives” principalement issues de la presse en ligne. Elles se distinguent des RCD en couvrant un territoire plus vaste qui intègre aussi des présentations publiques illustrées par des données ou des vidéos. Six ans plus tard, Stolper, Lee, Riche et Stasko [12] remarquent que l’un des grands thèmes du développement de la visualisation de données depuis l’article de Segel et Heer est l’intégration toujours plus avancées de la visualisation et du récit. Là où leurs prédécesseurs avaient des liens vers un autre article ou une visualisation, en 2016 les visualisations sont implantés directement dans le texte. Dans DataStory [13], Duarte propose quatre raisons pour lesquelles raconter une histoire est un moyen de communication préférable à une simple statistique :

- Une histoire engage un plus grand nombre de régions du cerveau, et donc une attention plus poussée
- Une histoire crée une connexion entre l’auteur et son audience.
- Une histoire crée un engagement personnel de l’audience par rapport au sujet.
- Une histoire pousse à agir plus efficacement que les chiffres.

Ces quatre effets sont des résultats souhaitables dans le contexte du journalisme.

Les intérêts d'un Récit Centré sur les Données pour le journalisme de données

L'un des intérêts d'un RCD est de gagner la confiance du public. *More Than Data Stories : Broadening the Role of Visualization in Contemporary Journalism* [9] établit que l'introduction de données pour étayer le récit contribue à établir l'autorité de ce récit. *The Data Journalism Handbook : How Journalists Can Use Data to Improve the News* [1] explique qu'il y a peu de marge de manœuvre pour contredire une visualisation de données.

Un autre intérêt des RCD est leur efficacité de communication. Weber et Rall, dans *Data Visualization in Online Journalism and Its Implications for the Production Process* [14] avancent que les visuels rendent une information complexe plus facile à comprendre pour le public. Dans *More Than Data Stories : Broadening the Role of Visualization in Contemporary Journalism* [9], Fu et Staskov citent *Readings in information visualization : using vision to think* [15] pour avancer que la visualisation de données permet de stimuler la mémoire de travail, d'améliorer la reconnaissance de schémas, et de rendre l'information plus facile à manipuler. Autant de bénéfices qui contribuent à communiquer plus efficacement avec un contenu dense en visualisations plutôt qu'un format de texte.

Les auteurs de *Data stories of water : Studying the communicative role of data visualizations within long-form journalism* [16] remarquent aussi que les visuels déclenchent des émotions plus fortes et fournissent des indices plus immédiats que le texte. Ces deux phénomènes augmentent l'impact du RCD.

2.1.2 Décrire et représenter les Récits Centrés sur les Données

Les RCD se distinguent des articles classiques uniquement ou principalement centrés sur le texte. Plusieurs démarches séparées ont été engagées par des chercheurs pour décrire les caractéristiques des RCD et des formats qui y sont apparentés, comme la visualisation narrative, qui englobe des vidéos explicatives par exemple. La recherche académique est aussi guidée par deux types d'approches : une qui consiste à s'adresser directement aux praticiens pour leur proposer des outils, et une qui consiste à travailler sur le plan théorique pour permettre à de futurs outils d'aller plus loin.

Différents modèles de représentation pour différentes interprétations des Récits Centrés sur les Données

Dans *Narrative Visualization : Telling Stories with Data* [7], Segel et Heer proposent un système pour décrire les visualisations narratives (qu'ils définissent comme des visualisations avec une composante narrative). Ce système est dominé par le genre narratif (qui dépend du

nombre et de l’ordonnancement des éléments visuels) et l’équilibre du contrôle entre auteur et lecteur. Leur démarche apporte une vision centrée sur la structure de la publication, et sur la façon dont la forme du RCD va influencer la réception du message par l’audience.

Dans *Patterns in award winning data storytelling* [17], Ojo et Heravi proposent une extension des travaux de Segel et Heer, en ajoutant des dimensions qui portent sur la démarche des concepteurs, leurs outils et leurs travail de collecte et de traitement des données. Une partie du modèle continue de considérer la forme comme vecteur de sens, les dimensions telles que l’interactivité et la représentation des données s’inscrivent dans cette démarche, tandis que d’autres s’en séparent pleinement, par exemple les outils technologiques.

Exploration and Explanation in Data-Driven Storytelling [18] décrit les RCD selon deux axes. Celui de la flexibilité et de l’interprétation forme le premier axe, qui étend l’idée d’équilibre du contrôle proposée par Segel et Heer. Le second axe se découpe en trois catégories : Vue, Focus et Séquence. Ce qui est montré à voir au public, ce sur quoi se concentre l’attention du public, et comment le public passe-t-il d’une vision à l’autre. La démarche est encore une fois l’utilisation de supports visuels (la forme) pour soutenir le message (le fond).

Solper, Lee, Riche et Stasko [12] proposent 20 types de visualisations narratives rassemblées en 4 grandes catégories. Leur intention est de séparer les visualisations narratives des outils d’exploration de visualisation de données en proposant des techniques qui caractérisent les RCD. Ces techniques, catégorisées entre “Communiquer le récit et expliquer les données”, “Lier des éléments d’histoire séparés”, “Améliorer la structure et la navigation” et “Fournir une exploration contrôlée”, sont axées sur des éléments structurels entourant la visualisation pour transformer les données en un récit.

Narrative Design Patterns for Data-Driven Storytelling [19], propose des “Narrative Design Patterns”, répartis en cinq groupes. La démarche ici est encore une fois de proposer des méthodes pour travailler avec la forme du RCD afin de communiquer à l’aide de cette forme. Les schémas proposés permettent d’engager le public, de susciter des émotions, et d’argumenter pour soutenir les données avec le récit.

À l’opposé la démarche de *Narrative Design Patterns for Data-Driven Storytelling* [19], *Classifying Data Journalism : A Content Analysis of Daily Data-driven Stories* [20] propose une classification plus mathématique, en proposant de compter les auteurs et d’établir un ratio entre le nombre de mots et le nombre de visualisations, et des interactions définies par un niveau de 1 à 6 plutôt que des méthodes spécifiques. Les auteurs de cet article utilisent leur taxonomie pour décrire l’existant. Ils proposent des valeurs moyennes de certaines dimensions sur leur panel d’exemples, ou des répartitions par catégories. Ils couvrent donc non seulement des éléments narratifs, liés à la forme, mais aussi des éléments sur le contexte dans lequel un

projet a été réalisé, ainsi que sur des travaux liés au fond, comme la collecte des données.

Certains auteurs, comme Landaverde et Abraham, dans *A Conceptual Framework For Interactive Cartographic Storytelling* [21], ou Roth [22] proposent aussi des classifications spécialisées pour certains sous-groupes spécialisés de RCD, comme les cartes narratives. Landaverde et Abraham proposent une taxonomie à 70 dimensions, allant de caractéristiques générales, telles que l’année de publication, à des éléments très spécifiques aux cartes, comme la méthode employée pour indiquer le nord.

Visualization design patterns [23] propose une approche très technique, avec des blocs de construction qui s’appliquent aux interactions, aux rendus visuels, à la programmation de pages web, à la gestion des données, et au cycle de vie d’un projet de RCD. La démarche vise spécifiquement des producteurs de RCD, et adresse aussi bien la forme que le fond.

The Stories We Tell About Data : Media Types for Data-Driven Storytelling [24] se place dans la définition la plus large du Data-Driven Storytelling, en proposant une taxonomie qui prends en compte des scénarios tels que que l’usage de la danse contemporaine pour représenter des données. Les dimensions proposées reflètent cette vision : elle prend en compte la composition du public, ainsi que le lieux dans lequel il se trouve et différencie les présentations asynchrones du temps réel. Certaines dimensions portent aussi sur les composantes du RCD, les visualisations, les photographies et le texte. L’idée de catégorisation par la forme reste donc présente, mais dans un sens élargi pour y intégrer des formats plus divers.

En considérant ces différentes taxonomie, nous pouvons constater que la vision dominante consiste à documenter comment la forme d’un RCD est utilisée pour enrichir et compléter le fond. La majorité des taxonomies sont pensées en termes d’interactions entre le lecteur et l’article, en termes de structures visuelles, et autour de l’idée que la forme est un moyen de raconter l’histoire qui est contenue dans les données. C’est de façon occasionnelle que les taxonomies abordent directement le fond, souvent en termes d’analyse de données. Beaucoup assument que les concepteurs de RCD ont déjà une histoire à raconter, et proposent des moyens de la raconter à l’audience.

Deux visions pour deux publics : théoriciens et praticiens

Parmi les précédents articles qui proposent des modèles de classification pour les RCD, deux tendances apparaissent dans la façon dont leurs conclusions sont supposées se propager de la théorie à la pratique. Certains envisagent la création d’outils à destination des concepteurs de RCD, tandis que d’autres envisagent la consultation directe d’articles de recherche par ces mêmes concepteurs.

Cinq articles s'inscrivent dans une démarche purement théorique. L'idée étant que leur contenu soit utilisé dans de futurs projets de recherche, ou pour développer un outil. Ils ne s'adressent pas directement aux concepteurs de RCD. *Narrative Visualization : Telling Stories with Data* [7], *Emerging and Recurring Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Recent Stories* [12] et *Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Visual Stories* [25] ont tous trois pour objectifs d'informer le design de futurs outils pour améliorer à terme le travail des designers. Parallèlement, *Cartographic Design as Visual Storytelling : Synthesis and Review of Map-Based Narratives* [22] s'adresse à des chercheurs, et cherche à ouvrir un champ de recherche transdisciplinaire entre la cartographie et les Data Stories. *Patterns in Award Winning Data Storytelling* [17] cherche à améliorer l'évaluation de la qualité des data stories.

Il existe aussi une catégorie d'articles conçus dans l'idée qu'un praticien, un concepteur de RCD, viendra directement lire leurs conclusions et s'en servir dans sa pratique professionnelle. *Visualization design patterns* [23] a été pensé pour être lu par des concepteurs afin d'uniformiser leur vocabulaire. *The Stories We Tell About Data : Media Types for Data-Driven Storytelling* [24], *Narrative Design Patterns for Data-Driven Storytelling* [19], *Exploration and Explanation in Data-Driven Storytelling* et *A Conceptual Framework For Interactive Cartographic Storytelling* sont tout les quatre adressés à des personnes qui participent à la production de RCD, dans l'optique que leurs travaux soient directement utilisés pour mener des projets de RCD.

Quelques travaux, restent ouverts aux deux possibilités. *Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Visual Stories* [25], en particulier projetée à la fois de faire avancer les outils de design et de fournir du matériau directement adressé aux designers.

Cette situation peut être mise en contexte par les travaux de Parsons dans *Understanding Data Visualization Design Practice* [26], qui identifie une séparation entre le monde académique et le monde professionnel. Là où les théoriciens issus du monde académique recherchent la généralisation, qui se retrouve dans les règles et modèles des taxonomies que nous avons décrit comme théoriques, les praticiens issus du monde professionnel visent la production d'artefacts particuliers, via des outils.

2.2 Le rôle des exemples dans l'idéation pour le journalisme de données et les Récits Centrés sur les Données

La conception d'un RCD est une tâche de design visuel, avec les défis que cela implique. L'un de ces défis est l'exercice de la créativité dans l'idéation pour la définition et la réalisation

du projet. L’une des méthodes qui contribuent à cette créativité est l’utilisation d’exemples de projets préexistants pour inspirer le travail de création.

2.2.1 La créativité dans la visualisation de données

Les auteurs de *Digital creativity support for original journalism* [27], avancent que chez les journalistes, la pensée créative est nécessaire pour découvrir et générer de nouvelles associations durant la recherche et l’analyse d’informations et contribue à la génération de nouveaux articles. Ils remarquent que quelques outils existent déjà pour soutenir la créativité, mais que c’est un support très incomplet, qui mène les utilisateurs à utiliser des outils génériques faute d’outils spécialisés.

Dans une optique de design de visualisations, Parson, auteur de *Design Cognition in Data Visualization* [28], avance l’idée que la communication à base de visualisations de données repose sur un travail de définition de problématique antérieur à la création du visuel. Un designer doit alors faire preuve de créativité pour proposer une analyse innovante du problème avant même de proposer la visualisation en elle-même. Les auteurs de *On the role of design in information visualization* [29] considèrent que la création d’une visualisation de données est une démarche de design créatif, ce qui implique une double démarche de définition du problème et de recherche de solutions. Mais ils proposent aussi l’idée d’un besoin de créativité dans l’esthétique du produit fini. Un résultat visuel plaisant fait partie intégrante de la communication.

La créativité est donc un atout majeur pour la conception de visualisations de données, en particulier pour le journalisme de données, et par extension les Récits Centrés sur les Données.

2.2.2 Les exemples contribuent à la créativité

Les exemples permettent de construire une culture professionnelle

Dans *Le Code de la Créativité* [30], Marcus Du Sautoy sépare la créativité en trois formes : La créativité exploratoire, la créativité combinatoire, et la créativité transformationnelle. La créativité exploratoire est la prise de conscience des limites de l’existant pour mieux repousser ces limites, mais sans changer les codes de l’environnement de travail. La créativité combinatoire consiste à prendre plusieurs créations préexistantes distinctes, et à les mélanger pour obtenir un résultat nouveau. Cela peut impliquer en particulier d’appliquer les règles de fonctionnement d’un domaine à un autre. La créativité transformationnelle consiste à changer les règles utilisées dans un domaine ou à remettre en question un postulat communément accepté. Ces trois formes de créativité reposent sur la connaissance de l’existant et de ce

qui a déjà été réalisé. En particulier les créativités exploratoire et combinatoire reposent sur l'utilisation de créations existantes.

Interview Study of Professional Designers' Ideation Approaches [31] montre que dans les métiers du design, les exemples sont utilisés d'une part pour la construction d'une culture du domaine de travail, qui mène à une créativité accrue, et d'autre part comme déclencheur pour la formation directe d'une idée ou la comparaison avec un projet en cours.

Les auteurs de *The Data Journalism Handbook : Towards a Critical Data Practice* [32] proposent une démarche d'enseignement centrée sur les exemples. Pour eux, le meilleur moyen de comprendre la pratique du journalisme de données est d'utiliser des exemples commentés et annotés.

Les exemples jouent le rôle de déclencheurs de l'idéation

Les auteurs de *Understanding How Designers Find and Use Data Visualization Examples* [6], ont interviewé des designers de visualisation de données pour mieux comprendre leurs pratiques. Ils ont découvert que ces designers utilisent des exemples pour stimuler leur créativité de plusieurs façons. Ils s'en servent comme base à partir de laquelle itérer, comme source de nouvelles méthodes de représentation des données, et pour découvrir de nouvelles esthétiques pour améliorer leurs projets. Ceci se retrouve dans les créativités exploratoires et combinatoires proposées par Du Sautoy. Dans ce même article, Bako, Liu, Battle et Liu ont aussi remarqué que les exemples permettent de comprendre à quelles visualisations leur audience est habituée. Cela permet aux concepteurs de RCD d'évaluer le risque que leur projet paraisse ordinaire, ou au contraire trop éloigné des habitudes du public.

Exploring Designers' Practice of Online Example Management for Supporting Mobile UI Design [33] indique que les exemples facilitent la création d'une structure de pensée visuelle, ainsi que la réinterprétation et l'évaluation d'idées de design, avec pour effet une augmentation du nombre et de la créativité des idées chez les designers.

Knaflitz, dans *Storytelling with data* [5], dédie un chapitre à disséquer des illustrations modèles, activité fondamentale de l'apprentissage. Elle avance qu'il y a quelque chose à apprendre de tous les exemples, bon ou mauvais. Un exemple est un outil de réflexion sur l'efficacité de différentes méthodes, ainsi qu'une invitation pour un concepteur de RCD à réfléchir à comment il aurait mené le projet.

L'outil INJECT, proposé dans *Digital creativity support for original journalism* [27], permet l'accès à des exemples généralistes d'articles de presse en ligne. Cet outil a montré des effets positifs sur la créativité des utilisateurs lors de tests dans des salles de presse.

Un accès augmenté à des exemples permet donc aux concepteurs de RCD de faire preuve de plus de créativité et d’aller plus loin dans leurs travaux.

2.3 Taxonomies des Récits Centrés sur les Données

Différentes taxonomies décrivent les RCD avec différentes dimensions. D’autres taxonomies décrivent des formats proches des RCD pour que leurs dimensions soient utilisées afin de décrire un RCD. Cette section présente les taxonomies qui ont été utilisées dans la suite de ce mémoire, notamment pour la construction de la taxonomie du projet.

2.3.1 *Narrative Visualization : Telling Stories with Data*

Segel et Heer [7] ont proposé la plus ancienne des taxonomies utilisées pour le projet de ce mémoire. Elle se concentre sur deux axes : le genre narratif et l’équilibre du contrôle partagé entre l’auteur et le lecteur.

Le genre

Le genre, tel qu’il est utilisé dans cette taxonomie, désigne 7 catégories non exclusives de récits, définies par le nombre et l’ordonnancement de leurs éléments visuels. L’exemple explicatif qu’ils proposent est la différence entre le style magazine, défini par une vue unique, et les comics, un enchaînement de vues ordonnées. Ces catégories sont : magazine, graphique / carte annoté, poster partitionné, organigramme, comic, diaporama, film / vidéo / animation.

Le Genre englobe plusieurs modèles qui ne sont, à première lecture, pas pertinents pour les RCD. Notre cas de figure pour un RCD est un article de presse en ligne. Ce qui exclut une interprétation littérale du poster, de “film / vidéo / animation”, et du magazine. Néanmoins, Segel et Heer décrivent le format magazine comme une visualisation implantée dans une page de texte, ce qui reste tout à fait possible sur une page web. De même, il est possible de rencontrer une page web qui contient une section animée. Suivant la même logique, le format poster peut être séparé de son support. Il est possible qu’un RCD soit conçu pour être embrassé du regard avant que le lecteur ne s’intéresse à des sections organisées de façon non linéaire.

Équilibre entre l’auteur et le lecteur

Le second aspect proposé par Segel et Heer est l’équilibre entre le contrôle que conserve l’auteur sur le récit et celui qui est donné au lecteur. Dans un scénario où le récit est linéaire,

avec des visualisations statiques ou une animation automatique, c’est l’auteur qui décide du message et de comment il est transmis. Au contraire, avec une interactivité très présente, c’est le lecteur qui peut construire son propre récit en s’appropriant le matériau fourni par l’auteur. Segel et Heer avancent que tous les projets de visualisations narratives se placent sur un axe entre ces deux extrêmes. Ils proposent aussi trois cas de figure intermédiaires, des configurations types placées sur cet axe.

- Diaporama interactif : une structure construite sur un enchaînement linéaire de cellules, contenant des interactions. L’auteur a conçu l’enchaînement principal, et impose l’enchaînement de ces cellules, mais le lecteur peut utiliser les interactions dans la cellule pour se l’approprier avant de passer à la suivante.
- Verre à martini : un récit linéaire contrôlé par l’auteur qui propose une ouverture une fois le récit principal passé. Le lecteur peut compléter son expérience à travers des interactions, mais après avoir reçu le message imposé par l’auteur.
- Histoire à forage : L’histoire à forage permet au lecteur de choisir un axe parmi plusieurs, l’explorer, puis revenir en arrière pour choisir un autre axe.

2.3.2 *Emerging and Recurring Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Recent Stories*

Solper, Lee, Riche et Stasko [12] proposent des techniques rassemblées en 4 grandes catégories : “Communiquer le Récit et Expliquer les Données”, “Lier des éléments d’histoire séparés”, “Améliorer la structure et la navigation” et “Fournir une Exploration Contrôlée”, chacun avec ses sous dimensions. Cette taxonomie comprend 20 techniques découpées en 4 catégories.

Communiquer le Récit et Expliquer les Données

La première catégorie, “Communiquer le Récit et Expliquer les Données”, englobe des moyens utilisés pour transmettre de l’information à l’audience en complément des visualisations de données elles-mêmes. Les techniques qui y appartiennent sont : récit textuel, narration audio, flèches d’organigramme, étiquetage, annotations de texte sur des visualisations, infobulles, surlignage d’éléments.

Cette catégorie introduit l’idée de catégoriser les éléments de communications qui entourent la visualisation et servent à compléter son message. La plus simple de ces techniques est le récit textuel, le corps de texte explicatif. Les autres options sont moins conventionnelles, mais remplissent le même rôle de transmettre de l’information explicite pour compléter les visualisations de données.

Lier des éléments d’histoire séparés

La deuxième catégorie, “Lier des éléments d’histoire séparés”, regroupe des méthodes permettant de communiquer une connexions entre deux éléments du récit. Ces techniques permettent de structurer un récit en établissant des connexions entre idées. La principale application proposée dans l’article consiste à relier une visualisation de données à ses explications. Pour ceci, trois méthodes sont proposées : liaison par l’interaction, liaison par la couleur, liaison par l’animation.

Améliorer la structure et la navigation

La troisième catégorie est Améliorer la structure et la navigation, qui propose différents moyens pour l’audience de naviguer entre différents éléments du récit. Elle offre une liste d’outils utilisables pour donner du contrôle au lecteur sur le RCD et sur le récit. Ses éléments sont : boutons suivant/précédent, défilement, fil d’Ariane, boutons d’en-tête de section, sélection du menu, ligne du temps, carte géographique.

Fournir une exploration contrôlée

La quatrième catégorie, “Fournir une Exploration Contrôlée”, comprend différentes méthodes pour guider le lecteur dans son exploration afin de l’empêcher de se perdre ou de dévier trop loin du récit principal. Les auteurs proposent les trois méthodes d’interaction suivantes pour guider le lecteur : requêtes dynamiques, visualisations exploratoires intégrées, visualisations exploratoires séparées.

Cette catégorie mélange les méthodes d’interaction avec le guidage du lecteur par l’auteur. Il y a une progression dans l’interaction : les requêtes dynamiques fonctionnent avec un petit nombre de commandes prédéfinies, alors que les deux types de visualisation exploratoire permettent une interactivité plus autonome pour le lecteur. La séparation des visualisations exploratoires se fait alors sur le critère de leur niveau d’intégration avec le texte.

2.3.3 *Classifying Data Journalism*

Classifying Data Journalism [20] propose 18 variables regroupées en 4 dimensions. Comparativement à l’usage que nous avons fait de ce vocabulaire pour d’autres articles, cela correspond à 18 dimensions regroupées en 4 catégories. Elle est récapitulée dans le tableau 2.1.

TABLEAU 2.1 Récapitulatif de Classifying Data Journalism

Caractéristiques formelles	Date de publication
	Médium
	Longueur en mots
	Nombre d’auteurs
	Sujet
Visualisations de données	Ratio texte/visualisation
	Nombre de visualisations
	Types de visualisations
	Niveau d’interactivité
Sources de données	Obtention des données
	Nombre de sources pour chaque visualisation
	Fournisseur de données
	Pays d’origine des données
Forme et contenu	Format de l’histoire
	Sujet
	Référence à la presse étrangère
	Juxtaposition

Caractéristiques formelles

La première de ces catégories, Caractéristiques formelles, est une description d’un RCD qui couvre des éléments hautement factuels, comme le nombre de mots ainsi que des éléments contextuels généraux.

La date de publication et le sujet sont deux informations critiques sur le fond derrière la forme du RCD. La date de publication informe du contexte dans lequel à été produit un RCD, et le sujet permet d’avoir une notion du contenu de l’article. Stalph propose les catégories de sujet suivantes : Politique, Société, Business, Culture, Sports, Local, Autre. Concernant le médium, différents écrans vont nécessiter différentes constructions d’un RCD, mais beaucoup de projets doivent être portés sur plusieurs formats. La longueur en mots et le nombre d’auteurs sont utilisés par Stalph pour repérer les tendances dans la construction des exemples qu’il a considéré en rédigeant l’article. Cette démarche apporte de la documentation, mais est moins pertinente pour décrire un exemple.

Visualisations de données

La seconde catégorie, Visualisation de données, regroupe les éléments portant directement sur les visualisations de données contenues dans le RCD. Cette catégorie répond à la question “Quelles visualisations de données ont été utilisées ?” mais donne aussi des informations sur

comment elles ont été utilisées. Elle compte quatre dimensions : Le type de visualisation, le ratio texte/visualisation, le nombre de visualisations et le niveau d'interactivité.

Le type de visualisation propose quinze types de visualisations possibles qui peuvent être présentes dans le RCD. Cette liste couvre les visualisations rencontrées par l'auteur, mais ne représente pas toutes les visualisations possibles. Les éléments de cette liste sont les suivants : bar chart, carte, line chart, pie chart, table, infographie, pictogramme, area chart, bubble chart, scatter plot, radar chart, timeline, flow chart, dashboard, autre.

Le ratio texte/visualisation et le nombre de visualisations apportent tous deux des informations de description sur un document et permettent à Stalph de faire des observations sur les tendances dans le domaine des RCD, mais leur utilisation sort de notre contexte. Un concepteur de RCD peut chercher un exemple sur la façon dont des visualisations ont été utilisées ou intégrées au récit, mais rechercher "190 mots par visualisation" est une requête très spécifique. Pour le niveau d'interactivité, l'auteur propose des niveaux, numérotés de 1 à 6 et indiqués dans une note comme "Niveau 1 = pas interactif; Niveau 2 = interactif; Niveaux 3, 4, 5 et 6 = hautement interactifs."

Sources de données

La section "sources de données" propose des informations concernant l'obtention des données qui ont été utilisées dans le RCD. L'obtention des données, le nombre de sources, le fournisseur et le pays d'origine des données sont tous les quatre des informations portant sur l'analyse préliminaire plutôt qu'à la conception du RCD.

Forme et contenu

La quatrième section, Forme et contenu, décrit la composante de récit du RCD. Elle comprend des informations portant plus sur le fond que sur la forme. Il n'y a pas de définition formelle d'un format de l'histoire dans l'article, mais les catégories proposées sont le reportage, l'article de fond, le commentaire, le portrait, le quiz et la catégorie "autre". Le sujet réapparaît une deuxième fois ici à cause d'une redondance dans la traduction : Stalph utilise les mots "topic" et "subject matter" respectivement. Là où "topic" couvre une liste définie de valeurs (politique, société, business, culture, sports, local, autre), "subject matter" est moins spécifique, mais couvre l'éducation, les élections, la migration, les politiciens et partis politiques, le divertissement ou la mode.

2.3.4 *Narrative Design Patterns for Data-Driven Storytelling*

Les dix-huit dimensions de cette taxonomie [19] ont la particularité d’être affectées à entre 1 et 4 groupes chacune. La table 2.2 associe chacune de ces dimensions aux différents groupes.

TABLEAU 2.2 Correspondances entre dimensions et groupes dans la taxonomie de *Narrative Design Patterns for Data-Driven Storytelling*

Dimension	Argumentation	Flux	Framing	Emotion	Engagement
Adresse à l’audience			X	X	X
Rupture du 4ème mur			X	X	
Appel à l’action					X
Comparaison	X				
Concrétisation	X			X	
Rupture de conventions			X		
Défamiliarisation			X		
Exploration		X			X
Familiarisation			X	X	
Révélation graduelle		X		X	X
Les humains derrière les points				X	
Devinette			X		X
Métaphore physique			X		
Répétition	X	X			
Question rhétorique	X		X	X	X
Données silencieuses				X	
Accélérations / décélérations		X		X	X
Les utilisateurs se trouvent				X	X

Ces schémas narratifs offrent une vision qui s’inscrit d’abord dans une démarche narrative, avec peu de spécificité en rapport aux données. Cependant, ces schémas restent conçus pour être utilisés dans un contexte de visualisation de données. Par exemple, le schéma “les humains derrière les points” peut consister à représenter des pictogrammes en forme de soldat, de blessés avec des béquilles ou de tombe pour représenter les victimes d’une guerre. Le schéma s’applique directement à la visualisation. Les auteurs de l’article favorisent une explication de leur dimensions au fil du texte, et reposant sur des exemples, ce qui limite la possibilité de définir strictement certaines dimensions.

2.3.5 *Exploration and Explanation in Data-Driven Storytelling*

La taxonomie proposée par Thudt, Walny Gschwandtner, Dykes et Stasko [18], repose sur la combinaison de deux ensembles de concepts.

Le premier regroupe la vue, le focus, et la séquence. La définition donnée pour la vue est “La façon dont les données sont présentées au lecteur, y compris les codages disponibles et les transformations éventuelles apportées aux données”. Celle du focus est “Le sujet de l’histoire, l’ensemble ou le sous-ensemble de données et les aspects des données qui sont affichés”. La séquence est décrite comme “L’ordre dans lequel les informations contenues dans l’histoire peuvent être vues.”

Le second groupe repose sur la séparation entre Flexibilité et Interprétation. La flexibilité est la liberté laissée au lecteur pour explorer l’histoire et se l’approprier. L’interprétation est décrite comme “tous types de sens ajoutés par l’auteur”. Il s’agit d’une démarche d’explication ou le concepteur d’un RCD utilise son autorité d’auteur pour décider d’une interprétation de l’information. Ce groupe s’inscrit dans la succession de Segel et Heer, où l’exploration est la flexibilité laissée au lecteur, et l’explication le contrôle que se réserve l’auteur pour imposer sa propre interprétation.

La combinaison permet ensuite aux auteurs de la taxonomie de créer des catégories “flexibilité pour la vue”, “flexibilité pour le focus”, “flexibilité pour la séquence”, “interprétation pour la vue”, “interprétation pour le focus”, et “interprétation pour la séquence”. Les dimensions produites à partir de la flexibilité ont ensuite été subdivisées en dimensions, tandis que celles issues de l’interprétation sont des invitations à la réflexion pour des concepteurs de RCD sans lister d’options spécifiques.

TABLEAU 2.3 Dimensions de flexibilité dans la taxonomie de *Exploration and Explanation in Data-Driven Storytelling*

Flexibilité	Vue	Vues séparées multiples
		Vues coordonnées multiples
		Transition animée entre les vues
		Encodage visuel mené par le lecteur
	Focus	Filtrage
		Sélection
		Zoom / Panoramique
		Forage
	Séquence	(1D) Défilement vertical
		(1D) Boutons avant/ arrière
		(2D) Panoramique et zoom sur la carte

La flexibilité de la vue propose des méthodes qui servent à naviguer entre plusieurs vues alternatives. Les vues multiples permettent au lecteur de voir plusieurs versions différentes de la même visualisation de données. La transition animée contrôle les passages entre ces vues. L’encodage visuel mené par le lecteur permet au lecteur de construire sa propre vue

avec des options à sa disposition.

La flexibilité du focus propose des méthodes pour laisser le lecteur choisir où se porte son attention. Il s'agit d'options d'interaction avec la visualisation pour changer les données et leur présentation.

La flexibilité de la séquence propose trois méthodes de navigation dans le RCD. Le défilement vertical et les boutons avant-arrière sont décrites comme 1D car elles proposent d'avancer d'avant en arrière, respectivement de façon continue et discrète. La troisième option, panoramique et zoom sur la carte est plus spécifique aux RCD utilisant une carte, et permet de naviguer dans plusieurs directions.

2.3.6 *Patterns in Award-Winning Data Storytelling*

Dans *Patterns in award winning data storytelling* [17], Ojo et Heravi proposent 10 dimensions issues d'autres articles de recherche. Ils reprennent en particulier le genre narratif de Segel et Heer, en le renommant Représentation, et renomment aussi l'équilibre du contrôle auteur/lecteur en style narratif. Ils proposent aussi la classification des RCDen fonction de plusieurs autres paramètres.

Cette classification propose 10 grandes catégories, dont une est découpée en 4 sous-catégories, pour un total de 59 dimensions. Ces dimensions couvrent une grande variété d'informations sur le RCD. Certaines portent sur la démarche générale des concepteurs (but, audience), sur la narration (éléments du récit), sur les données et leur traitement (données, technique d'analyse), les outils employés (médium, outils technologiques) et sur la forme du RCD (style narratif, interactivité, Représentation). La table 2.4 récapitule cette taxonomie.

TABLEAU 2.4 Taxonomie de *Patterns in Award-Winning Data Storytelling*

But	Informer	
	Persuader	
	Divertir	
	Réconforter	
	Expliquer	
	Terroriser	
Audience	Grand public	
	Groupe spécifique	
Éléments du récit	Personnage	
	Contexte	
	Intrigue	
Médium	Navigateur web	
	Mobile	
	Imprimé	
Données	Source	Capturées : Acquisition directe des données
		Générées : Obtenues comme sous-produit d'un autre processus
		Dérivées : Issues d'une prédiction
	Forme	Qualitative
		Quantitative
	Structure	Structurées
		Non structurées
	Type	Numériques
		Texte
		Spatiales
		Réseau
		Temporelles
Style Narratif	Axé sur l'auteur	
	Axé sur le lecteur	
	Hybride : Verre à martini, diaporama interactif et histoire à forage	
Interactivité	Statique	
	Interactif	
	Interactivité limitée	
	Recherche/Filtrage/Sélection	
Représentation	Magazine	
	Graphique / Carte Annoté	
	Poster partitionné	
	Organigramme	
	Comic	
	Diaporama	
	Film / Vidéo / Animation	
	Autre	
Techniques d'analyse	Analyse statistique	
	Fouille de données	
	Traitement automatique de la langue naturelle	
	Apprentissage automatique	
	Analyse spatiale	
Outils technologiques	Excel	
	Tableau Public	
	OpenRefine	
	Google Fusion	
	R	
	Python	
	SPSS	
	Rubby	
	PHP	
	JavaScript	
	HTML	
	ESRI	
	Logiciel de cartographie	
	Many Eyes	

La façon dont les données ont été obtenues est séparée entre les données capturées (acquises directement), générées (obtenues comme sous produit d'un autre processus) ou dérivées (issues d'une prédiction).

Remarquons que la technique d'analyse est basée sur *Big data : The next frontier for inno-*

vation, competition, and productivity [34], qui propose une liste de techniques plus longue, avec 26 options.

2.3.7 *A Conceptual Framework For Interactive Cartographic Storytelling*

Landaverde et Abraham [21] proposent plusieurs modèles destinés à décrire des récits cartographiques, que nous pouvons concevoir comme un sous-ensemble des RCD. Leur taxonomie regroupe au total 39 catégories, réparties en sous catégories, soit binaires, soit à options multiples, découpées entre deux tableaux et une annexe.

Les sections les plus pertinentes que nous avons considérées sont les suivantes :

- Les indicateurs de navigation. Ces indicateurs proposent une liste de ressources pour permettre aux lecteurs de naviguer dans le RCD. Ils recoupent à la fois le suivi de la progression du récit et le contrôle de cette progression par le lecteur.
- Indicateurs interactifs (Interactive Cues). Ces indicateurs proposent différentes méthodes pour montrer à l'utilisateur comment utiliser les différentes interactions.

2.3.8 *Cartographic Design as Visual Storytelling : Synthesis and Review of Map-Based Narratives*

La taxonomie de Roth [22] se divise en deux sections. Il propose une nouvelle version du genre de Segel et Heer, proposé dans *Narrative Visualization : Telling Stories with Data* [7], ainsi qu'une liste de tropes du récit visuel. Ces deux concepts sont résumés dans le tableau 2.5

Genre narratif

Cette version revue du genre narratif pour refléter les transformations qui ont émergé quand les médias en ligne ont repensé la composition de leurs travaux autour du web, renonçant au portage des formats papier. Il définit les différents genres selon la façon dont ils imposent la linéarité. L'axe de travail principal de Roth est la cartographie, mais son application du genre est transférable aux RCD. Les nouveaux genres proposés par Roth sont définis comme suit :

- Histoire visuelle statique : force la linéarité par le partitionnement de la structure et la clarification de la lecture avec des annotations
- Infographie longue : force la linéarité par la lecture verticale et le défilement du navigateur

- Diaporama dynamique : force la linéarité par l'avancement par le biais d'une série de diapositives
- Animation narrée : force la linéarité par la progression dans le temps de l'affichage numérique
- Expérience visuelle multimédia : force la linéarité par des balises d'ancrage et des liens hypertexte
- Story Map personnalisée : force la linéarité via l'ordre dans lequel un individu contribue au contenu de la carte.
- Compilation d'histoire visuelles : force la linéarité en suivant le déroulement des événements en temps quasi réel ou en mettant à jour le RCD

Tropes

Le second groupe, les tropes sont des outils narratifs utilisés pour faire avancer l'intrigue. Ces outils sont très ancrés dans le récit, et beaucoup ne sont pas divisés en catégories strictes.

Certains éléments de taxonomie, comme le point d'entrée clair, sont des invitations à réfléchir à leur sens et à faire attention à en produire un dans un projet, mais il ne s'agit pas d'une méthode qui peut être appliquée. Certains tropes sont donc hors du champ d'application du projet de ce mémoire

TABLEAU 2.5 Taxonomie de *Cartographic Design as Visual Storytelling : Synthesis and Review of Map-Based Narratives*

Genre	Histoire visuelle statique	
	Infographie longue	
	Diaporama dynamique	
	Animation narrée	
	Expérience visuelle multimédia	
	Story Map personnalisée	
	Compilation d'histoires visuelles	
Tropes	Continuité	Récit linéaire en trois actes
		Rythme de genre
		Annotation
		Interaction
		Transitions
	Ambiance	Conception d'éléments cartographiques
		Marginalia et autres ornements
		Style visuel
		Style persuasif
		Styles de pastiche
	Dosage	Séquençage linéaire
		Conception modulaire
		Scénarisation
		Point d'entrée clair
		Partitionnement
		Pause sélective ou ralenti dans une animation.
		Rythme d'interaction spécifique au genre
		Pop-ups
	Attention	Hierarchie visuelle / figure-fond
		Accentuation visuelle
	Redondance	Motifs visuels et symbolisme
		Figures et Oeufs de pâque
		Hierarchie et accentuation visuelles
	Métaphore	Symboles ponctuels associatifs
		Motifs visuels et symbolisme
		Juxtaposition visuelle
		Humour, ironie et satire
	Voix	Voix du designer
		Voix de personnage
		Voix de l'audience
		Typographie
		Point de vue et projection
		Narration : audio et audiovisuelle

2.3.9 Visualization design patterns

Visualization design patterns [23] est une taxonomie publiée sur le site web InfoVis Wiki². Elle propose des solutions à des problèmes de visualisation d'informations, qui sont par extension applicables à des RCD.

2. https://infovis-wiki.net/wiki/Visualization_Design_Patterns

Elle se divise entre schémas fondamentaux, schémas de programmation et schémas de développement. Le groupe le plus pertinent ici est celui des schémas fondamentaux. La liste d'interactions en particulier propose une gamme de méthodologies précises. L'exception réside dans les schémas unipolaires et bipolaires, qui sont spécifiques à la réalité virtuelle.

La taxonomie de *Visualization design patterns* [23] est récapitulée dans le tableau 2.6.

TABLEAU 2.6 Taxonomie de *Visualization design patterns*

Schémas fondamentaux	Architecture de visualisation	Visualisation
	Interactions	Reconfigurer
		Encodage
		Filtrer
		Abstraction/Élaboration
		Connections
		Exploration
		Sélection
		Autre
		Schémas Unipolaires (réalité virtuelle)
		Schémas Bipolaires (réalité virtuelle)
	Rendu de l'affichage	Appropriation des objets visuels
		Niveau de détails
		Représentation 2D
		Représentation 3D
		Contexte de référence
		Encodage redondant
		Small Multiples
		Datatips
		Légende
		Séparation Visuelle
		Dispositif d'organisation familial
		Dispositif d'organisation non-familier
Schémas de programmation	Modélisation et gestion des données	Modèle de référence
		Colonne de données
		Table en cascade
		Graphe relationnel
		Proxy Tuple
		Expression
	Vue	Caméra
		Liaison de requête dynamique
	Contrôleur	Programmateur (scheduler)
		Opérateur
		Représentateur (renderer)
Schémas de développment	Règle de production	
	Diagnostic	
	Évaluation et Test	

Cette taxonomie provient d'un wiki et les sources sont variables d'une section à l'autre. La section des schémas d'interaction se distingue car elle cite des articles de recherche. Nous donnerons la priorité aux éléments de cette section par la suite.

2.3.10 *The Stories We Tell About Data : Media Types for Data-Driven Storytelling*

Cette [24] taxonomie propose 5 groupes pour caractériser les RCD. Les auteurs travaillent avec une définition plus large de RCD que les autres, et cette taxonomie le reflète, en prenant en compte des scénarios plus divers que notre contexte de journalisme en ligne. Elle est récapitulée dans le tableau 2.7.

Cardinalité de l’audience et espace-temps

La cardinalité de l’audience divise les scénarios en fonction du nombre de narrateurs et du nombre de personnes dans l’audience. 1 pour un individu unique, et N pour plusieurs personnes. Le groupe “espace et temps” propose de séparer des RCD selon si le narrateur et le public sont ou non dans le même espace, et si le récit se fait en temps réel où est conçu pour être consulté après sa création. Dans notre cas d’usage, les RCD sont tous de cardinalité 1 :N, distribués dans l’espace et asynchrones, car la narration est centralisée, et le produit fini est distribué via une page web.

Composantes médiatique et de données

Deux groupes s’organisent autour de l’idée de composante. Le groupe des composantes de données liste les moyens de communiquer des données. La majorité des visualisations de données sont regroupées sous le nom de “Graphiques statistiques”. Le terme “graphe” a été utilisé pour les visualisations de réseaux, graphes de connexions et matrices d’adjacence. Les composantes médiatiques qui sont les composantes d’un RCD autres que les visualisations de données, et qui servent de complément. Les auteurs séparent “animations” et “vidéos”, en considérant qu’une animation est un graphique animé. De même un audio regroupe les extraits de voix et effets sonores autre que la musique.

Séquence de visualisation

Le dernier groupe, la séquence de visualisation, propose les outils qui peuvent permettre au lecteur de contrôler la séquence de présentation des visualisations. Le contrôle de playback permet au lecteur d’arrêter le récit ou de le ramener en arrière. La navigation permet au lecteur d’utiliser le zoom et un glissement pour se déplacer dans la représentation. Le contrôle, de base ou complet, réfère à des contrôles pour transformer les visualisations, allant de la mise en valeur d’un point (contrôle simple) à des options de filtrage, transformation et connexion libres.

TABLEAU 2.7 Taxonomie de *The Stories We Tell About Data : Media Types for Data-Driven Storytelling*

Cardinalité de l'audience	1 :1 : Un narrateur et un public d'une personne
	1 :N : Un narrateur et un public de plusieurs personnes
	N :1 : Plusieurs narrateurs et un public d'une personne
	N :N : Plusieurs narrateurs et un public de plusieurs personnes
Espace et temps	Co-localisé : Le narrateur et le public sont dans le même espace physique.
	Distribué : Le narrateur et le public se sont pas dans le même espace physique.
	Synchrone : Le narrateur délivre son récit au public en temps réel.
	Asynchrone : Le narrateur délivre son récit sous une forme qui peut être consommée par la suite.
Composantes médiatiques	Audio
	Musique
	Photographies
	Vidéo en direct
	Graphique Statique
	Animation
	Texte
Composantes de données	Table de données
	Carte géographique
	Graphique statistique : Visualisations de données
	Ligne du temps des événements
	Séries chronologiques de données
	Visualisation de graphe
	Visualisation de texte : Nuages de mots et concordances visuelles
Séquence de visualisation	Pas d'interaction
	Contrôle de playback
	Navigation
	Contrôle de base
	Contrôle complet

2.4 Synthèse

Il existe des manques aux connaissances dans la littérature. Concernant les approches de l'idéation pour le journalisme de données, il n'existe des travaux portant sur les exemples dans la presse classique d'une part, des travaux sur les RCD d'autre part, mais au meilleur de notre connaissance il n'en existe pas qui porte sur l'effet des exemples lors de la conception d'un RCD. Certains travaux sur les RCD, repérés dans la sous-section 2.1.2 sont conçus dans l'optique que de futurs travaux vont utiliser leurs conclusions pour les adapter en pratique et les rendre disponible à un public de praticiens. Il existe donc une avenue d'exploration pour mettre en place cette adaptation et réduire la distance entre théoriciens et praticiens.

CHAPITRE 3 MÉTHODE PROPOSÉE

3.1 Taxonomie des Récits Centrés sur les Données

La première composante de ce projet est la création d'une taxonomie destinée à décrire les RCD. Cette démarche répond à l'objectif O1 de ce projet : proposer une nouvelle taxonomie des RCD, qui permette de catégoriser les RCD avec une large variété d'options de description, afin d'offrir des possibilités de tri variées. Des taxonomies existent déjà. Notre objectif n'est donc pas simplement de proposer une nouvelle taxonomie, mais aussi de proposer une taxonomie spécifiquement adaptée à nos objectifs plus larges. D'une part, notre intention est d'employer cette taxonomie comme filtre dans un outil, ce qui implique de besoins spécifiques de format pour les dimensions, auxquels les taxonomies préexistantes ne répondent pas forcément. D'autre part, ce projet s'inscrit dans une démarche visant à rendre des dimensions issues du monde académique accessible à un public de praticiens via l'outil, nous cherchons donc à rassembler ces éléments de taxonomie dans la nôtre. Enfin, les concepteurs des taxonomies précédentes ont approché la description des RCD selon des angles différents, et nous chercherons à proposer une taxonomie qui couvre ces différentes approches. Ces principes vont donc guider nos décisions. Nous avons donc collecté des taxonomies préexistantes, destinées à décrire des RCD ou des formats similaires aux RCD, tels que la visualisation narrative, ou les cartes narratives. Nous avons sélectionné les dimensions les plus pertinentes parmi ces taxonomies préexistantes, puis les avons regroupées et catégorisées, avant de raffiner cette taxonomie pour la rendre le plus compréhensible possible.

3.1.1 Taxonomies préexistantes

La première étape de la conception de la taxonomie a été de rassembler les taxonomies préexistantes et en extraire les dimensions pertinentes pour le projet de ce mémoire. Nous avons employé les taxonomies issues des articles suivants :

1. *Narrative Visualization : Telling Stories with Data* [7]
2. *Emerging and Recurring Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Recent Stories* [12]
3. *Classifying Data Journalism* [20]
4. *Narrative Design Patterns for Data-Driven Storytelling* [19]
5. *Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Visual Stories* [25]

6. *Exploration and Explanation in Data-Driven Storytelling* [18]
7. *A Conceptual Framework For Interactive Cartographic Storytelling* [21]
8. *Patterns in Award-Winning Data Storytelling* [17]
9. *Cartographic Design as Visual Storytelling : Synthesis and Review of Map-Based Narratives* [22]
10. *Visualization design patterns* [23]
11. *The Stories We Tell About Data : Media Types for Data-Driven Storytelling* [35]

Notre objectif étant d'utiliser la taxonomie proposée dans ce chapitre comme moyen de trier et de filtrer une future archive, l'une des conditions à respecter pour la dimensions est d'avoir différentes valeurs pouvant être sélectionnées individuellement, sous forme de catégories ordinales ou nominales. En particulier, comme nous l'avons vu dans la revue de littérature, certaines taxonomies ont été pensées pour être lues par un concepteur de RCD, qui va ensuite réfléchir à comment les différentes dimensions affectent ses travaux. Par conséquent, certaines de ces dimensions ne cherchent pas à lister toutes les possibilités, ni même à décrire un concept avec un nombre fini de possibilités. Ces dimensions sont donc des invitations à la réflexion, et ne sont donc pas pertinentes pour notre démarche de catégorisation.

Une autre condition est que chaque dimension doit décrire une information présente dans un exemple et qui soit réutilisable pour la conception d'un RCD. Une information qui porte sur une méthode utilisée lors de travaux préliminaires, par exemple pour l'analyse de données, serait difficile à retrouver dans un exemple pour annoter ce dernier. Nous nous attendons aussi à ce que les informations sur les travaux préliminaires soient peu utiles pour nos utilisateurs, car ils sont plus spécifiques d'un projet à l'autre et risquent d'être déjà terminés au moment de la conception du RCD. Pour reprendre le cas de figure qui nous a amené à poser cette condition, beaucoup d'exemples ne citent pas le langage de programmation utilisé pour traiter leurs données.

Enfin, nous cherchons en priorité à utiliser les dimensions qui représentent un choix de conception de la part des créateurs d'une RCD, dans l'idée que nos utilisateurs pourront examiner ce choix et décider de le reproduire ou non. Inversement, des éléments situationnels qui ne peuvent pas être activement reproduits seront conservés seulement si ils informent de manière significative le contexte dans lequel le projet à été mené.

Narrative Visualization : Telling Stories with Data [7]

Le modèle proposé par Segel et Heer concentre l'attention sur deux axes : le genre narratif et le l'équilibre du contrôle partagé entre l'auteur et le lecteur

Nous avons initialement conservé le genre. Il s’agit de catégories qui décrivent la structure du RCD, et nous pouvons l’utiliser pour décrire les exemples. *Cartographic Design as Visual Storytelling : Synthesis and Review of Map-Based Narratives* [22] en propose une autre interprétation que nous avons considérée par la suite.

Concernant l’équilibre entre auteur et lecteur, nous avons choisi de conserver cette dimension pour caractériser les interactions offertes par le RCD. Étant donné qu’elle prend la forme d’un axe continu entre deux extrêmes, nous avons choisi de la discrétiser avec les cas de figure ponctuels qui ont été décrits : diaporama interactif, verre à martini, histoire à forage. Nous avons aussi ajouté deux cas extrêmes où tout le contrôle est totalement entre les mains de l’auteur ou du lecteur. Cela résulte dans le jeu de valeurs suivant : axé sur l’Auteur, diaporama interactif, verre à martini, histoire à forage, axé sur le lecteur.

***Emerging and Recurring Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Recent Stories* [12]**

Stolper, Lee, Riche, et Stasko ont proposé une taxonomie pour le Data Driven Storytelling, en s’axant sur les méthodes de récit.

La section “Communiquer le Récit et Expliquer les Données” introduit l’idée de catégoriser les éléments de communications qui entourent la visualisation elle-même. Nous pouvons conserver cette section. “Fournir une Exploration Contrôlée” à aussi été initialement conservé comme ressource pour décrire les interactions. Concernant “Améliorer la structure et la navigation“, elle offre une liste très complète d’outils utilisables pour donner du contrôle au lecteur sur le RCD et sur le récit. “Lier des éléments d’histoire séparés” introduit l’idée de connecter des éléments de récit et de documenter les méthodes utilisées, mais ses valeurs entrent en conflit avec notre condition concernant la classification. La façon dont la couleur a été utilisée pour relier des concepts, ou dont l’animation a été utilisée pour relier des concepts, sont des sujets de réflexion pour un concepteur de RCD, mais sans spécifier quelle est cette façon nous ne pouvons pas offrir de classification pertinente. Nous ne retenons donc pas la dimension “Lier des éléments d’histoire séparés”.

***Classifying Data Journalism* [20]**

Classifying Data Journalism [20] propose 18 variables regroupées en 4 dimensions. Comparativement à l’usage que nous avons fait de ce vocabulaire pour d’autres articles, cela correspond à 18 dimensions regroupées en 4 groupes.

Parmi les caractéristiques formelles, nous avons conservé la date de publication et le sujet. La

date de publication informe du contexte dans lequel à été produit l'exemple, et le sujet permet de trier des exemples par leur contenu. Stalph propose les catégories de sujet suivantes : Politique, Société, Business, Culture, Sports, Local, Autre. La longueur en mots n'a pas de valeurs ordinales ou nominales, et ne correspond pas au critère que nous avons établi, qui demande des valeurs catégoriques. Le nombre d'auteurs n'a pas de rapport avec le fond ou la forme, nous n'avons donc pas retenu cette dimension. Concernant le médium, nous considérons que la plupart des projets de RCD sont déployés avec un format réactif qui couvre à la fois les navigateurs web et les mobiles, et le format papier n'est pas pertinent pour notre définition des RCD. La dimension médium ne sera donc pas conservée.

Pour le groupe "visualisation de données", le ratio texte/visualisation pose le même problème que le nombre de mots. Les valeurs ne sont pas catégoriques. Nous ne retenons donc pas le ratio texte/visualisation. L'objectif d'un exemple pour les concepteurs est de fournir une inspiration méthodologique, en mettant l'accent sur l'utilisation des visualisations. Nous considérons donc que le nombre de visualisations n'est pas intéressant en lui-même. En revanche, le type de visualisation est pertinent pour notre projet. Il s'agit d'un choix de conception crucial pour la communication de données. L'idée de niveau d'interactivité est aussi intéressante, mais des niveaux numérotés sont une source d'information limitée, en particulier quand quatre niveaux partagent la même définition. Cette dimension sera conservée, dans l'intention d'être complétée par les dimensions d'interaction issues d'autres taxonomies.

Pour les sources de données, nous considérerons que nos utilisateurs disposent déjà de leurs propres données, et que la façon dont un exemple a obtenu son jeu de données appartient aux travaux préliminaires. Ce groupe ne sera donc pas conservé.

Parmi les éléments de forme et contenu, deux sur quatre n'ont pas de listes d'options que nous pouvons utiliser comme valeurs pour ces dimensions. Nous ne les retenons donc pas. La problématique du sujet soulevé dans la revue de littérature (différence entre "Matter subject" et "Topic") sera résolue en proposant une dimension sujet et en proposant une liste de sujets dérivés d'une autre source. La dimension "Référence à la presse étrangère" n'a pas été conservée car notre archive, détaillée dans la section suivante de ce chapitre, contient uniquement de la presse anglophone des États Unis.

Concernant les visualisations, un élément à été retenu : le type de visualisations, avec 15 valeurs possibles proposées : bar chart, carte, line chart, pie chart, table, infographie, pictogramme, area chart, bubble chart, scatter plot, radar chart, timeline, flow chart, dashboard, autre.

Cette liste ne couvre pas toutes les visualisations possibles, mais fournit un point de départ que nous avons étendu par la suite pour correspondre à l'archive. La dimension Interaction

propose des niveaux de 1 à 6. Cette catégorisation n’a pas été retenue, en faveur de dimensions décrivant l’interaction issues d’autres articles.

Narrative Design Patterns for Data-Driven Storytelling [19]

La première problématique concernant les schémas proposés par *Narrative Design Patterns for Data-Driven Storytelling* [19] est la répartition entre les catégories. Étant donné que la plupart des schémas appartiennent à deux catégories ou plus, il n’est pas possible de reproduire directement ces catégories pour notre modèle sans créer des redondances. Pour résoudre ce problème nous avons suivi utilisé les règles suivantes :

- Associer chaque schéma a une catégorie maximum
- Réduire le nombre de catégories en fusionnant les plus similaires
- Fusionner les schémas les plus similaires les uns aux autres

Premièrement, nous avons filtré les schémas pour ne les laisser que dans les catégories ayant une définition explicite dans l’article d’origine. Cela nous laisse la liste présentée dans le tableau 3.1

TABLEAU 3.1 Dimensions de *Narrative Design Patterns for Data-Driven Storytelling* après le premier filtrage

Dimension	Argumentation	Flux	Framing	Emotion	Engagement
Adresse à l’audience			X	X	
Rupture du 4ème mur			X	X	
Appel à l’action					X
Comparaison	X				
Concrétisation	X			X	
Rupture de conventions			X		
Défamiliarisation			X		
Exploration		X			X
Familiarisation			X	X	
Révélation graduelle		X		X	
Les humains derrière les points				X	
Devinette			X		X
Métaphore physique			X		
Répétition	X	X			
Question rhétorique			X		X
Données silencieuses					
Accélérations / décélérations		X		X	
Les utilisateurs se trouvent					

Nous faisons ensuite les constats suivants :

- La catégorie “Flux” est entièrement composée de dimensions qui apparaissent dans les autres catégories. Nous pouvons donc la retirer de notre version.
- Les dimensions “Données silencieuses” et “Les utilisateurs se trouvent” ont été éliminées de toutes les catégories.

Nous pouvons donc les retirer de notre version. Pour réduire les redondance, nous proposons

les choix suivants :

- La devinette et la question rhétorique sont des choix de cadrage pour susciter de l’engagement. L’engagement étant le but final, elles rejoignent cette catégorie.
- La définition de “concrétisation” dans le champ de l’émotion se rapproche de “les humains derrière les points”, alors qu’elle n’a pas de dimension proche dans le champ de l’argumentation. Nous laissons donc la concrétisation dans argumentation, et “les humains derrière les points” dans émotion.
- La familiarisation devrait rester dans la même catégorie que la défamiliarisation. Nous conservons donc la familiarisation dans le cadrage, et la retirons de l’émotion.

Restent l’adresse à l’audience et la rupture du 4ème mur. En l’absence de raisons plus substantielles, nous choisissons de les laisser dans le groupe le plus petit des deux afin de garder des groupes de taille plus équilibrée afin de rendre la lecture de la taxonomie plus aisée. Ce qui laisse la version présentée dans le tableau 3.2.

TABLEAU 3.2 Tableau modifié à partir du tableau 3.1 des dimensions de *Narrative Design Patterns for Data-Driven Storytelling*

Dimension	Argumentation	Framing	Emotion	Engagement
Comparaison	X			
Concrétisation	X			
Répétition	X			
Rupture de conventions		X		
Défamiliarisation		X		
Familiarisation		X		
Métaphore physique		X		
Révélation graduelle			X	
Les humains derrière les points			X	
Accélérations / décélérations			X	
Adresse à l’audience			X	
Rupture du 4ème mur			X	
Appel à l’action				X
Exploration				X
Devinette				X
Question rhétorique				X

Exploration and Explanation in Data-Driven Storytelling [18]

Cette taxonomie contient six dimensions potentielles. Nous en avons déjà repéré trois qui répondent au besoin de catégories utilisables selon nos critères : flexibilité de la vue, flexibilité du focus et flexibilité de la séquence. Les trois autres catégories, celles de l’interprétation, ne répondent pas à nos critères de valeurs ordinales ou nominales. En l’absence de valeurs à donner à ces dimensions, nous ne les incluons pas dans notre taxonomie. Nous conservons les dimensions listées dans le tableau 2.3.

A Conceptual Framework For Interactive Cartographic Storytelling [21]

Parmi les options proposées dans cette taxonomie, deux groupes ont retenu notre attention : les indicateurs de navigation et les indicateurs interactifs. Les indicateurs de navigation sont les suivants : barre de progrès, boutons de section, timeline, fil d'Arianne, sélection dans un menu, visualisation, carte géographique, liste de pointage.

Il s'agit d'une liste de moyens pour donner au public des moyens de contrôle sur la progression du récit qui va de la classique barre de progrès à des méthodes plus rares, comme une carte avec un chemin illustré qui représente l'avancée du récit. Les indicateurs interactifs sont les suivants : instructions explicites, entrée de requêtes, tutoriel tacite.

Patterns in Award-Winning Data Storytelling [17]

Le but fournit une dimension qui informe sur la démarche des créateurs originaux d'un exemple. Bien qu'il ne décrive pas un élément du RCD, le but informe sur le contexte qui a mené à la création du RCD. Nous avons donc conservé cette dimension.

Concernant l'audience, nous avons choisi de ne pas conserver cette dimension. En effet, les exemples de RCD que nous utilisons sont tous publics et pourraient être classifiés comme destinés au grand public. Nous pourrions proposer l'idée qu'un RCD s'adresse à un groupe spécifique à l'intérieur du grand public, mais en l'absence d'une liste finie nous pouvons difficilement proposer une classification efficace à nos utilisateurs. Spécifier "cet article s'adresse à un groupe spécifique" ne serait pas informatif et identifier manuellement un groupe spécifique pour chaque article serait très complexe en plus d'ouvrir le risque de définir des dizaines de groupes différents.

Concernant les éléments du récit, nous rencontrons un problème similaire. Chaque récit a son propre personnage, sa propre intrigue et son propre contexte. En l'absence de listes de personnages, d'intrigues ou de contexte, nous pouvons soit affirmer "Ce RCD à un contexte", soit expliquer le contexte spécifique du RCD et obtenir une information unique qui ne permet pas le filtrage. Aucune de ces deux options n'est pertinente, donc nous ne retenons pas les éléments du récit comme dimension.

Le médium rencontre le même problème que celui proposé par *Classifying Data Journalism* [20]. Nous considérons que les RCD coexistent sur les formats "mobile" et "navigateur web" et ne sont pas portés au format imprimé. Nous ne conservons donc pas cette dimension.

Pour la section données, nous évaluons plusieurs sous-sections séparément. La source de données appartient aux activités préparatoires du RCD. Nous supposons que nos utilisateurs ont déjà leurs propres données et ne vont pas trouver d'inspiration dans la collecte de données

d'un exemple. Cette dimension n'est pas retenue. La forme des données décrit des informations contenues dans les données. Séparer les données qualitatives et quantitatives permet à nos utilisateurs de choisir un exemple qui correspond à leurs données et à leur démarche. Similairement, le type permet de décrire les données et leur usage dans l'exemple. Des données spatiales ou numériques vont créer un exemple différent. Nous conservons donc les deux dimensions forme et type. Le caractère structuré ou déstructuré des données porte sur leur traitement préliminaire et leur format de stockage, ce qui n'est pas pertinent pour notre démarche.

Le style narratif est une reprise du genre de *Narrative Visualization : Telling Stories with Data* [7], que nous avons déjà retenu. Il en va de même avec la représentation, qui correspond à l'équilibre auteur/lecteur.

L'interactivité propose des catégories progressives qui appartiennent au domaine du design. Nous conservons cette dimension.

La technique d'analyse, comme nous l'avons remarqué dans la revue de littérature, correspond à un extrait de *Big data : The next frontier for innovation, competition, and productivity* [34]. Il existe une longue liste de techniques qui pourraient être utilisées pour l'analyse de données, et beaucoup d'entre elles appartiennent à l'analyse préliminaire. Une petite partie de ces techniques reste pertinente pour certains RCD moins conventionnels, comme certains exemples qui proposent une prédiction au lecteur à partir d'informations qu'il ou elle a donné. Plutôt que de lister toutes les méthodes possibles, nous proposons d'utiliser cette dimension comme inspiration pour une nouvelle dimension binaire "Oui/Non", qui mettra ces exemples en valeur sans chercher à tous les expliquer individuellement.

La dernière dimension de cet article est celle des outils technologiques. L'utilisation d'Excel, Tableau, Python ou R pour analyser ne change pas le RCD qui sera conçu à partir de ces données. Nous pouvons aussi présumer que nos utilisateurs disposent de leurs propres outils, et ne vont chercher à changer d'outils à cause d'un exemple que de façon très exceptionnelle. Enfin, il est aussi probable que les exemples de notre archive ne contiennent pas tous de détails sur les outils utilisés pour les produire. Nous avons donc renoncé à utiliser cette dimension dans notre taxonomie.

***Cartographic Design as Visual Storytelling : Synthesis and Review of Map-Based Narratives* [22]**

Le genre proposé par Roth se trouve dans la succession du genre proposé par Segel et Heer dans *Narrative Visualization : Telling Stories with Data* [7]. La liste des options proposées

dans cette dimension est différente : elle a été transformée pour s’adapter aux formats en ligne. Nous conservons cette dimension.

Concernant les tropes, nous avons repéré deux sous-tropes spécifiquement intéressantes :

- Dans “Continuité”, la liste des transitions propose une collection d’options intéressantes : “fading, panning, swiping, tweening”, que nous avons traduit en français par “atténuation”, “panoramique”, “glissement” et “interpolation” (la fusion ou transformation de plusieurs éléments visuels).
- Dans Ambiance, les styles visuels ont été considérés, mais pas retenus, car ses valeurs ont été conçues pour être l’objet d’une réflexion pour un concepteur de RCD mais pas pour la catégorisation.

Visualization design patterns [23]

Les schémas qui nous intéressent dans cette taxonomie sont les schémas fondamentaux. Les autres schémas, programmation et développement, sont hors champ pour le projet de ce mémoire, car ils portent sur le traitement des données préliminaire à la conception d’un RCD.

Parmi les schémas fondamentaux, les schémas d’interaction sont les seuls à être basés sur des articles de recherche. Ce sont donc ceux que nous considérerons. Parmi ces schémas, nous retirons les schémas portant sur la réalité virtuelle. Ceci nous laisse les catégories suivantes : reconfigurer, encodage, filtrer, abstraction/élaboration, connections, exploration, sélection, autres. Cette liste d’options forme une dimension de schémas d’interaction qui peut entrer dans notre taxonomie.

The Stories We Tell About Data : Media Types for Data-Driven Storytelling [35]

Nous ne conservons pas la cardinalité de l’audience, dimension qui indique si le public se compose d’une ou plusieurs personnes et si plusieurs sources participent à la conception. Tous les RCD qui entrent dans notre définition ont la cardinalité 1 :N (source unique, audience multiple). De même, tous ont la même valeur dans la dimension espace-temps : distribué et asynchrone.

Pour la composante médiatique, les catégories correspondant à la conception du RCD sont utilisables pour le filtrage. Nous conservons donc cette dimension.

Concernant la composante de données, cette dimension est une liste de visualisations de données. Nous avons déjà rencontré des dimensions similaires, avec des options différentes.

Nous allons donc conserver cette dimension pour une future fusion.

La séquence de visualisation propose des options qui sont aussi utilisables pour décrire la forme du RCD et pour le filtrage. Cette dimension sera donc conservée.

Récapitulatif des dimensions conservées

Après ce tri, nous obtenons une liste de dimensions qui serviront de matériau pour la construction de notre propre taxonomie. Ces dimensions sont listées dans la tables 3.3.

TABLEAU 3.3 Récapitulatif des dimensions extraite des taxonomies

Source	Dimension	Valeur
Narrative Visualization : Telling Stories with Data	Genre	Magazine, graphique / carte annoté, poster partitionné, organigramme, comic, diaporama, film / vidéo / animation.
	Contrôle auteur/lecteur	Axé sur l'auteur, diaporama interactif, verre à martini, histoire à forage, axé sur le lecteur
Emerging and Recurring Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Recent Stories	Communiquer le Récit et Expliquer les Données	Récit textuel, narration audio, flèches d'organigramme, étiquetage, annotations de texte sur des visualisations, infobulles, surlignage d'éléments
	Améliorer la structure et la navigation	Boutons suivant/précédent, défilement, fil d'Ariane, boutons d'en-tête de section, Sélection dans un menu, ligne du temps, carte géographique
	Fournir une Exploration Contrôlée	Requêtes dynamiques, visualisations exploratoires intégrées, visualisation exploratoires séparées
Classifying Data Journalism	Caractéristiques formelles	Date de publication, sujet
	Visualisation de données	Visualisations
Exploration and Explanation in Data-Driven Storytelling	Flexibilité de la vue	Vues séparées multiples, vues coordonnées multiples, transition animée entre les vues, encodage visuel mené par le lecteur
	Flexibilité du focus	Filtrage, sélection, zoom/pan, drill down
Patterns in Award-Winning Data Storytelling	Forme des données	Qualitative, quantitative
	Type de données	Numérique, textuel, spatial, réseau, temporel
	Interactivité	Statique, interactif, interactivité limité, recherche/filtrage/sélection
	technique d'analyse	Oui/non
A Conceptual Framework For Interactive Cartographic Storytelling	Indicateurs de navigation	Barre de progrès, boutons de section, ligne du temps, fil d'Arianne, sélection dans un menu, visualisation contrôlée, carte géographique, liste de pointage
	Indicateurs interactifs	Instructions explicites, entrée de requêtes, tutoriel tacite
Cartographic Design as Visual Storytelling : Synthesis and Review of Map-Based Narratives	Genre	Histoire visuelle statique, infographie longue, diaporama dynamique, animation narrée, expérience visuelle multimédia, storymap personnalisée, compilation d'histoire visuelles
	Transitions	Atténuation, panoramique, glissement, interpolation
Narrative Design Patterns for Data-Driven Storytelling	Argumentation	Comparaison, concrétisation, répétition
	Cadrage	Rupture de conventions, défamiliarisation, familiarisation, métaphore physique
	Émotion	Révélation graduelle, les humains derrière les points, accélérations / décélérations, adresse à l'audience, rupture du 4ème mur
	Engagement	Appel à l'action , exploration, devinette, question rhétorique
Visualization design patterns	Interactions	Reconfigurer, encodage, filtrer, abstraction/élaboration, connections, exploration, sélection, autres
The Stories We Tell About Data : Media Types for Data-Driven Storytelling	Composante médiatique	Audio, musique, photographies, vidéo en direct, graphique statique, animation, texte
	Composante de données	Table de données, carte géographique, graphique statistique, ligne du temps des événements, séries chronologiques de données, visualisation de graphe, visualisation de texte
	Séquence de visualisation	Pas d'interaction, contrôle de playback, navigation, contrôle de base, contrôle complet

3.1.2 Comparaison des taxonomies existantes

Nous disposons d’une liste de 26 dimensions issues des taxonomies préexistantes. Ces dimensions doivent être organisées, certaines sont similaires et peuvent être regroupées, d’autres peuvent être raffinées.

Genre

Le genre proposé par *Narrative Visualization : Telling Stories with Data* [7] à été réutilisé dans plusieurs taxonomies, et a inspiré une nouvelle version chez *Cartographic Design as Visual Storytelling : Synthesis and Review of Map-Based Narratives* [22].

La version de Roth étant plus récente et plus adaptée au web, nous retenons ses éléments.

Médias Complémentaires

La dimension “Communiquer le Récit et Expliquer les Donnée”, proposée dans *Emerging and Recurring Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Recent Stories* [12] et la dimension “Composante Médiatique” proposée dans *The Stories We Tell About Data : Media Types for Data-Driven Storytelling* [35] répondent au même objectif. Il s’agit de documenter les composantes qui ne sont pas des visualisations de données, mais qui forment la communication autour de ces visualisations. L’exemple type est un extrait de texte, ou une photographie d’illustration. Nous comparons les options de ces deux dimensions dans le tableau 3.4.

TABLEAU 3.4 Comparaison des options de médias

Emerging and Recurring Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Recent Stories	The Stories We Tell About Data : Media Types for Data-Driven Storytelling
Récit Textuel	Audio
Narration Audio	Musique
Flèches d’organigramme	Photographies
Étiquetage	Vidéo en direct
Annotations de Texte sur des Visualisations	Graphique Statique
Infobulles	Animation
Surlignage d’élément	Texte
	Vidéos en direct

Le “Récit Textuel” et le “Texte” sont similaires. Dans la même idée, les infobulles, l’étiquetage et les annotations de texte sur des visualisations sont trois méthodes qui se résument à des

variations de la valeur “texte”. Nous simplifions donc la liste en regroupant toutes ces idées sous le nom de “Texte”.

Les valeurs “narration audio”, “audio” et “musique” sont aussi trois idées proches. Nous les regroupons sous un terme commun : “audio”, qui couvrira toute forme de contenu audio présent dans le RCD.

Concernant les illustrations visuelles, nous avons “photographie”, “graphique statique”, “animation” et “vidéos en direct”. Deux paires se distinguent : “photographie” et “graphique statique” d’une part, et “animation” et “vidéos en direct” d’autre part. La première paire couvre des illustrations statiques, photographique ou dessinées. La seconde paire couvre les illustrations visuelles mouvantes, qu’il s’agisse de vidéos et d’animations. Nous proposons une composition plus réduite pour regrouper ces deux idées : “image” et “vidéo”.

Il nous reste les flèches d’organigramme et le surlignage d’éléments. Ces deux options ne transmettent pas d’informations en elles-mêmes, mais en mettent d’autres en valeur. Elles ne correspondent pas à la définition de la nouvelle catégorie que nous proposons. Nous retirons donc ces deux valeurs. Il en résulte une catégorie “Médias complémentaires” avec quatre valeurs : Audio, Image, Texte, Vidéo.

Indicateurs de progression

Plusieurs des dimensions que nous avons listées tournent autour de l’idée de contrôler la séquence de visualisation. Le contrôle le plus simple d’un RCD est celui qui permet d’avancer dans le récit, et de revenir en arrière. Les dimensions “Améliorer la structure et la navigation”, “indicateurs de navigation” et “séquence de visualisation” offrent toutes des options pour résoudre cette problématique.

TABLEAU 3.5 Comparaison des options de séquence

Améliorer la structure et la navigation	Indicateurs de navigation	Séquence de visualisation
Boutons Suivant/Précédent	Barre de progrès	Pas d’interaction
Défilement	Boutons de section	Contrôle de playback
Fil d’Ariane	Ligne du temps	Navigation
Boutons d’en-tête de section	Fil d’Ariane	Contrôle de base
Sélection dans un menu	Sélection dans un menu	Contrôle complet
Ligne du temps	Visualisation contrôlée	
Carte géographique	Carte géographique	
	Liste de pointage	

Plusieurs de ces valeurs sont similaires et nous pouvons les fusionner et les utiliser telles

quelles :

- “Fil d’Ariane” est présent deux fois
- “Boutons d’en-tête de section” et “boutons de section” sont similaire
- “Sélection dans un menu” est présent deux fois
- “Ligne du temps” est présent deux fois
- “Carte géographique” est présent deux fois

Concernant la séquence de visualisation, le contrôle de playback regroupe plusieurs des idées avancées par les autres articles : la majorité permet une forme ou une autre d’aller-retour. La navigation, le contrôle de base et le contrôle complet appartiennent à notre dimension “interactions”. Nous retenons la valeur “pas d’interaction” pour les situations où il n’y a pas de moyen de contrôler la séquence.

Le défilement et la barre de progrès sont deux valeurs qui portent sur l’avancée linéaire. Nous considérerons le défilement comme l’état par défaut, et donc un scénario couvert par “pas d’interaction”. Une barre de progrès se distingue par l’existence d’un élément visuel pour le suivi. Elle formera donc une valeur dans notre taxonomie.

Il reste les deux valeurs “boutons suivant/précédent” et “liste de pointage”. Ces valeurs portent sur deux outils valides qui n’ont pas de redondance avec les autres. Nous conservons donc ces valeurs.

Concernant le nom de la catégorie qui en résulte, le point commun est le contrôle sur la progression du récit. Nous avons principalement retenu des indicateurs visuels de cette progression, avec l’option de les utiliser pour contrôler cette progression. Nous nommerons donc cette dimension “Indicateurs de progression”. Les valeurs restantes sont : sans indicateur, barre de progrès, boutons de section, ligne temporelle, carte géographique, fil d’Arianne, liste de pointage.

Visualisations de données

Deux des dimensions que nous avons rencontrées proposent de décrire les visualisations de données présentes dans un RCD : La dimension “visualisation de données” de *Classifying Data Journalism* [20], et la “composante de données” de *The Stories We Tell About Data : Media Types for Data-Driven Storytelling* [35]. Elles sont toutes deux présentées dans le tableau 3.6.

TABLEAU 3.6 Comparaison des options de visualisations de données

Visualisation de données	Composante de données
Bar chart	Table de données
Carte	Carte géographique
Line chart	Graphique statistique : Visualisations de données
Pie chart	Ligne du temps des événements
Table	Séries chronologiques de données
Infographie	Visualisation de graphe
Pictogramme	Visualisation de texte
Area chart	
Bubble chart	
Scatter plot	
Radar chart	
Timeline	
Flow chart	
Dashboard	
Autre	

En comparant ces deux possibilités, nous constatons que leurs logiques sont différentes : l’une des deux propose de lister les visualisations type par type, alors que l’autre propose de grandes catégories. La première démarche a l’avantage d’être précise, mais fait face au risque de ne pas proposer une possibilité valide. La seconde démarche est garantie de couvrir toutes les possibilités, mais propose des catégories très irrégulières dans leur taille : la quasi-totalité des visualisations de données existantes sont couvertes par “graphiques statistiques” alors que les lignes du temps ont deux catégories distinctes, et que la théorie des graphes a sa propre catégorie.

Pour résoudre ce problème, nous proposons de choisir une méthode listant des types de visualisations précis, mais en utilisant notre propre liste, qui sera conçue à partir de notre archive pour mieux refléter son contenu. Pour cela, nous utiliserons une liste initiale issue de la littérature, avant de procéder à des ajustements basés sur les articles qui seront inclus dans l’archive.

Nous utilisons la liste de *Classifying Data Journalism* [20] comme base et ajoutons des éléments de la liste de *The Stories We Tell About Data : Media Types for Data-Driven Storytelling* [35] qui ne sont pas présentes : la table de données et la visualisation de graphe. Nous complétons la liste avec des éléments de *Data Visualisation : A Handbook for Data Driven Design* [36]

La liste résultante est la suivante : carte de chaleur, carte choroplèthe, carte de navigation, diagramme à barres, diagramme à bulles, diagramme à colonnes, diagramme circulaire, graphique en aires, graphique de flux, graphique radar, infographie, ligne, ligne du temps, nuage

de points, organigramme, pictogramme, visualisation de graphe, table de données, tableau de bord, tracé en mosaïque, autres.

À partir de cette liste, nous avons parcouru l’archive (détaillée dans la section 3.2) et ajouté à la liste les diagrammes qui n’étaient pas listés. Pour réduire le nombre total d’options, nous avons ensuite supprimé celles qui n’avaient pas d’occurrence dans notre archive. Nous avons aussi choisi de regrouper les diagrammes qui peuvent être identifiés comme des variantes du même concept, par exemple le diagramme à barres et le diagramme à barres empilées.

Il en résulte la liste de visualisations suivante : autre, carte à flux, carte à points, carte à symbole proportionnels, carte choroplèthe, carte de chaleur, carte de navigation, cartogramme, diagramme à barres, diagramme à bulles, diagramme à formes proportionnelles, diagramme circulaire, diagramme d’instance, diagramme de coordonnées parallèles, diagramme de sankey, graphique de flux, graphique en aires, graphique radar, histogramme, infographie, ligne, modèle 3D, nuage de points, organigramme, pictogramme, range chart, réseau, table, tracé en mosaïque.

Degré de liberté de l’utilisateur

L’équilibre entre auteur et lecteur apparaît plusieurs fois sous différents noms depuis sa version originale dans *Narrative Visualization : Telling Stories with Data* [7], jusqu’à la plus récente de nos taxonomies, *The Stories We Tell About Data : Media Types for Data-Driven Storytelling* [35]. Cette idée est aussi présente dans *Patterns in Award-Winning Data Storytelling* [17], et *Cartographic Design as Visual Storytelling : Synthesis and Review of Map-Based Narratives* [22].

Nous avons choisi de conserver la version simple présente dans notre récapitulatif 3.3 en tant que “Contrôle auteur/lecteur”.

Séquence de visualisation

La dimension “Fournir une Exploration Contrôlée” de *Emerging and Recurring Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Recent Stories* [12], la dimension “Interactivité” de *Patterns in Award-Winning Data Storytelling* [17], et la section “Séquence de visualisation” de *The Stories We Tell About Data : Media Types for Data-Driven Storytelling* [35] sont toutes les trois des sections portant sur l’interaction et qui proposent des moyens de progresser dans le récit. Nous avons déjà exploré cette idée dans la section “Indicateurs de progression”, ce qui nous a mené à une catégorie d’indicateurs visuels. Ici, nous proposons une autre dimension qui est le complément de cette dernière, centrée sur l’interaction et sur

le contrôle.

Comparons les trois dimensions qui nous intéressent dans le tableau 3.7.

TABLEAU 3.7 Comparaison des options de séquence de visualisation

Fournir une Exploration Contrôlée	Interactivité	Séquence de visualisation
Requêtes dynamiques	Statique	Pas d'interaction
Visualisations exploratoires intégrées	Interactif	Contrôle de playback
Visualisation exploratoire séparé	Interactivité limitée	Navigation
	Recherche/Filtrage/Sélection	Contrôle de base
		Contrôle complet

Nous avons initialement envisagé de croiser le “Fournir une Exploration Contrôlée” et “Séquence de visualisation” : Nous séparons les valeurs des contrôles “de base” et “complet”, ainsi que les visualisations “intégrées” et “séparées”. Ceci produirait la liste d'options suivantes : sans interaction, contrôle de playback, navigation, contrôle basique intégré, contrôle basique séparé, contrôle total intégré, contrôle total séparé, contrôle par requêtes (requêtes dynamiques). Après avoir produit cette liste, nous avons constaté qu'elle est très longue, et que beaucoup des options les plus complexes ne se reflètent pas dans notre archive.

Nous avons considéré que la navigation et les entrées suivantes sont toutes rares, et correspondent à une forme ou une autre de navigation non linéaire, là où le contrôle de playback est une forme de navigation linéaire. Nous avons donc choisi de conserver cette division entre linéaire et non linéaire.

Après avoir proposé ces options, nous avons choisi de conserver le nom de “séquence de visualisation”, et renommé “sans interaction” en “sans séquençage”. Nous obtenons un dimension intitulée “séquence de visualisation” avec les valeurs suivantes : sans séquençage, navigation Linéaire, navigation non linéaire.

Transitions

Une autre information présente dans plusieurs des dimensions que nous avons repéré est l'idée de transition visuelle. La dimension “flexibilité de la vue” (issue de *Emerging and Recurring Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Recent Stories* [12]) avance une option “Transition animée entre les vues”, et la dimension “Transition” de *Cartographic Design as Visual Storytelling : Synthesis and Review of Map-Based Narratives* [22] propose une variété d'options de transitions.

Nous utiliserons les options les plus détaillées disponibles dans ces deux listes. Remarquons aussi que ces options se recoupent avec une partie des options de “Flexibilité du focus”, là

aussi *Emerging and Recurring Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Recent Stories* [12]. Nous obtenons donc une dimension Transition, avec les options “atténuation”, “panoramique”, “glissement” et “interpolation”.

Interactions

L’une des problématiques restantes, liée aux options des dimensions “flexibilité de la vue” et “flexibilité du focus” que nous n’avons pas utilisées dans les dimensions précédentes, est l’interaction avec le RCD autre que la navigation. Ces deux catégories proposent des options telles que “encodage visuel mené par le lecteur”, “filtrage”, “sélection”, et “forage”.

Une autre dimension qui porte sur ce point est la section “Interactions” de *Visualization design patterns* [23], qui propose les options suivantes : reconfigurer, encodage, filtrer, abstraction/élaboration, connections, exploration, sélection, autres.

Nous remarquons que les options “filtrage” et “sélection” sont déjà présentes dans la liste. De même “encodage visuel mené par le lecteur” et “encodage” sont similaires. Le concept du forage, obtenir plus d’information par l’interaction, correspond à la partie “élaboration” de l’option “abstraction/élaboration”. Nous considérerons donc que toutes les options se recoupent. Nous utiliserons la liste des interactions de *Visualization design patterns* [23].

Récapitulatif des fusions

Suite aux fusions, nous obtenons une liste de 21 dimensions, présentée dans le tableau. 3.8.

TABLEAU 3.8 Récapitulatif des fusions de catégories

Dimensions	Valeurs
Titre	Pas de sous-dimensions
Date	Pas de sous-dimensions
But	Informar, persuader, divertir, réconforter, expliquer, terroriser
Sujet	Pas de sous-dimensions
Forme	Qualitative, quantitative
Type	Numériques, textes, spatiales, réseau, temporelles
Analyse informatisée	Oui, non
Genre	Histoire visuelle statique, infographie longue, diaporama dynamique, animation narrée, expérience visuelle multimédia, story map personnalisée, compilation d'histoires personnalisées, compilation d'histoires visuelles
Médias complémentaires	Audio, image, texte, vidéo
Indicateurs de progression	Sans indicateur, barre de progrès, boutons de section, ligne temporelle, carte géographique, fil d'Arianne, liste de pointage
Visualisations de données	Autre, carte à flux, carte à points, carte à symbole proportionnels, carte choroplèthe, carte de chaleur, carte de navigation, cartogramme, diagramme à barres, diagramme à bulles, diagramme à formes proportionnelles, diagramme circulaire, diagramme d'instance, diagramme de coordonnées parallèles, diagramme de sankey, graphique de flux, graphique en aires, graphique radar, histogramme, infographie, ligne, modèle 3D, nuage de points, organigramme, pictogramme, range chart, réseau, table, tracé en mosaïque
Degré de liberté de l'utilisateur	Axé sur l'auteur, diaporama interactif, verre à martini histoire à forage, axé sur le lecteur
Séquence de visualisation	Sans séquençage, navigation linéaire, navigation non-linéaire
Transitions	Atténuation, panoramique, glissement, interpolation
Interactions	Reconfigurer, encodage, filtrer, abstraction/élaboration, connection, exploration, sélection, autres
Indicateurs interactifs	Instructions explicites, entrée de requêtes, tutoriel tacite
Argumentation	Comparaison, concrétisation, répétition
Cadrage	Rupture de conventions, défamiliarisation, familiarisation, métaphore physique
Émotion	Révélation graduelle, les humains derrière les points, accélération/décélération, adresse à l'audience, rupture du 4ème mur
Engagement	Appel à l'action, exploration, devinette, question rhétorique

3.1.3 Ajustements des dimensions et valeurs

Après avoir fusionné les dimensions, nous proposons une variété d’ajustements afin de les clarifier, de mieux refléter le sens des dimensions résultant d’une fusion, et de combler des manques qui ont émergé.

Création de la dimension tutoriel

La dimension “indicateurs interactifs” propose des méthodes pour expliquer le fonctionnement du RCD et de ses interactions au public. Nous avons renommé la dimension en “tutoriel” et reformulé les valeurs proposées. Les instructions explicites et le tutoriel tacite sont deux concepts qui peuvent être envisagés comme complémentaires, et nous les avons renommé “tutoriel explicite” et “tutoriel tacite”. La troisième option, l’entrée de requêtes, est définie dans *A Conceptual Framework For Interactive Cartographic Storytelling* [21] comme l’utilisation d’infobulles pour inciter l’utilisateur à tester une fonctionnalité. Nous considérons qu’il s’agit d’une variante de tutoriel tacite, déjà couverte par la valeur du même nom.

Ajout des dimensions source et lien

Le principal cas d’usage du projet de ce mémoire est d’utiliser la taxonomie pour classer une archive d’exemples de RCD. Nous observons que nos dimensions laissent un vide critique pour cet objectif. En effet, l’un des principaux besoins de cette archive est de pouvoir ouvrir l’exemple pour le consulter. Or notre taxonomie ne fournit actuellement que le titre pour identifier et retrouver un exemple. Il faut donc des informations complémentaires pour cette tâche. Nous proposons d’y ajouter le journal de publication et un lien vers la page d’origine. Le lien et le titre deviennent les deux seules dimensions où nous conservons une valeur unique pour chaque exemple, afin de permettre de l’identifier.

Valeurs de la dimension “sujet”

Il reste une dimension incomplète : nous avons coupé la liste des valeurs de “Sujet” en proposant de les remplacer par d’autres valeurs, dans l’intention de produire une liste qui couvre mieux la diversité de sujets que nous nous attendons à rencontrer. Pour cela, nous avons choisi de nous référer aux travaux de l’International Press and Telecommunications Council (IPTC), qui propose une liste de sujets [37]. Cette liste propose plusieurs niveaux de profondeurs, avec des nombres croissants de valeurs possibles et toujours plus de détails.

Le premier niveau propose 17 valeurs différentes et le second niveau en propose 122. 17

valeurs sont déjà un nombre considérable, et 122 valeurs différentes seraient excessives. Nous retenons donc la liste de 17 valeurs.

Nous avons donc considéré les valeurs officielles en français pour la dimension “sujet”, et avons décidé d’en modifier deux en nous référant à la version anglophone. La valeur “Gens animaux insolite”, qui porte sur les récits émotionnels sur des histoires de personnes ou d’animaux, a été remplacée par “Intérêt humain”, basé sur l’anglais “Human Interest”. Nous avons aussi remarqué la présence des catégories “Social” et “Société”, difficiles à différencier. En nous référant à la liste anglophone, il apparaît que “Social” a été traduit depuis “Labour” et porte sur les sujets ayant trait au milieu de l’emploi. Nous avons donc choisi de la renommer “Emploi”.

Il en résulte la liste suivante : arts, culture, divertissement et médias ; conflit, guerre et paix ; criminalité, droit et justice ; désastres et accidents ; économie et finances ; éducation ; emploi ; environnement ; intérêt humain ; météo ; politique ; religion et croyance ; santé ; science et technologie ; société ; sport ; vie quotidienne et loisirs.

But : valeur “terroriser”

Nous avons remarqué que parmi les valeurs de la dimension “but” se trouve une option “terroriser”. Nous avons constaté que ce terme est connoté négativement, et avons considéré que nos utilisateurs ne vont pas volontairement chercher un exemple de méthodes pour terroriser leur audience. L’option a donc été retirée.

Division de la séquence linéaire

En annotant l’archive, nous avons remarqué que beaucoup d’exemples avaient une séquence linéaire, mais qu’une partie se distingue par leur structure discontinue. Plusieurs des dimensions que nous avons considérées proposent une idée qui transparaît dans la dimension “indicateurs de progression” mais qui a été partiellement perdue : la séparation entre les séquences continues, par exemple un défilement vertical de page web, ou les séquences discontinues, par exemple un diaporama. Nous séparons donc la valeur “Séquence linéaire” de la dimension “Séquence de visualisation” entre “Séquence linéaire continue” et “Séquence linéaire discontinue”

Fusion des schémas d’engagement et d’empathie

Parmi les schémas narratifs, les schémas d’engagement et d’empathie sont deux dimensions qui sont assez proches. Il s’agit de choisir une démarche pour rapprocher le lecteur du récit qui

lui est proposé et l’impliquer plus avant. Nous considérons l’empathie comme une méthode pour engager le public. Nous allons donc fusionner les deux dimensions sous le nom “schémas d’engagement et d’empathie” et lister toutes les valeurs des deux.

Ajout de valeurs “sans”

Plusieurs des dimensions proposent une liste de méthodes de conception pour un RCD, mais sans avoir de garantie que tous les exemples de notre archive, ou futurs projets de nos utilisateurs, vont utiliser une de ces techniques. Nous ajoutons donc la valeur “Sans” aux dimensions qui ouvrent la possibilité de n’utiliser aucune des valeurs proposées. Ces dimensions sont : tutoriel, séquence de visualisation, modes de transitions, indicateurs de progression, interactions, schémas argumentatifs, schémas de cadrage, schémas d’engagement de d’empathie.

Choix de vocabulaire

La première problématique que nous avons traitée est d’ajuster notre vocabulaire, pour simplifier et clarifier la taxonomie, et la rendre la plus facile possible à utiliser. Pour cela, nous avons travaillé sur la langue, les anglicismes et leurs équivalents français, puis nous avons retravaillé les formules employées pour les simplifier.

Langue Premièrement, nous avons revu la langue. La plupart de nos sources étant anglophones, il y a des termes et formules qui ont besoin d’être réévaluées, et inversement des termes que nous avons traduit en français par défaut et qui ont besoin d’être clarifiés car leur forme anglaise est plus commune.

Une des problématiques que nous avons rencontré est la transition. Les valeurs en français sont peu communes, alors que leurs formes anglophones sont plus communes. Nous avons choisi de spécifier la forme anglaise entre parenthèses après la forme française. Les nouvelles valeurs sont donc atténuation (fading), glissement (swiping), interpolation (tweening), panoramique (panning).

Nous avons aussi constaté que parmi les valeurs d’“Indicateurs de progression”, deux sont peu compréhensibles. Le “fil d’Ariane” et la “liste de pointage” sont des formules francophones qui ne sont pas immédiates. Leurs formes anglophones, sont respectivement “breadcrumbs” et “checklist”. Nous avons utilisé la même méthode que pour les transitions.

Raccourcissement des termes Pour clarifier notre taxonomie, nous avons retravaillé plusieurs termes. Le genre est un concept qui a été désigné sous plusieurs noms dans les taxonomies préexistantes. Nous avons conservé par défaut le nom “genre” car il s’agit du plus ancien terme. Étant donné que ce terme existe aussi pour d’autres concepts, comme la littérature ou le cinéma, nous avons considéré que le terme était un risque de confusion. Nous avons donc choisi de le renommer “structure de l’article”, qui est plus proche des valeurs possibles pour cette dimension. Nous avons changé “analyse informatisée” en “analyse”. Spécifier l’aspect informatique nous a semblé redondant. La séquence linéaire discontinue a été nommée de façon complémentaire à la séquence linéaire continue, mais nous l’avons renommée pour la clarifier. La clé de sa compréhension étant la séparation du récit en étapes discrètes, nous la renommons en “séquence linéaire par étapes”.

3.1.4 Regroupements des dimensions

Nous disposons désormais de 22 dimensions distinctes. Pour rendre notre taxonomie plus claire et plus simple à utiliser, nous avons regroupé ces taxonomies en groupes de plus haut niveau pour les organiser.

Schémas narratifs

Le premier groupe que nous avons considéré est celui des schémas narratifs. Quatre dimensions sont issues de l’article *Narrative Design Patterns for Data-Driven Storytelling* [19] sans avoir été modifiées avec des éléments externes à l’article. Elles obéissent donc à la logique initiale qui gouverne les schémas narratifs : “Un schéma narratif est un dispositif narratif de bas niveau qui sert une intention spécifique. Un schéma peut être utilisé individuellement ou en combinaison avec d’autres pour donner forme à un récit”.

Métadonnées

Ensuite, nous avons considéré deux dimensions qui ont reçu jusque-là le moins d’attentions : “titre” et “date”. Il s’agit d’informations très générales sur le RCD, qui ne correspondent pas vraiment à son contenu mais informent sur des informations extérieures qui sont nécessaires. Elles rejoignent nos deux nouvelles dimensions, “journal de publication ” et “lien”, pour former un groupe de métadonnées sur le RCD.

Contexte

Dans la continuité des métadonnées, nous proposons un groupe intitulé contexte. Ce groupe rassemble certaines des informations les plus générales sur la démarche des créateurs d'un RCD et le cadre dans lequel il a été créé.

Nous y intégrons le but et le sujet, qui informent sur l'objectif que poursuivent les créateurs, ainsi que le genre, qui informe sur le format utilisé. Représenter un RCD avec son contexte pourrait par exemple prendre la forme d'une storymap (genre) visant à expliquer (but) la politique d'une région (sujet).

Contrôle

Nous proposons un groupe intitulé contrôle pour regrouper tous les outils destinés à construire le contrôle du public sur le RCD que nous avons recensés.

Cette définition permet de regrouper les dimensions suivantes ;

- Degré de liberté de l'utilisateur, qui mesure l'ampleur du contrôle.
- Séquence de visualisation, qui couvre les options qui permettent de progresser dans le récit.
- Mode de transition, qui couvre les effets visuels liés à cette séquence.
- Indicateurs de progression, qui regroupe les indicateurs visuels de l'avancement de la lecture.
- Interactions, qui couvre les options d'interaction avec les visualisations.

Bien qu'ils ne répondent que tangentiellement à l'idée de contrôle, nous avons intégré "modes de transition" et "indicateurs de progression" car nous avons considéré qu'il s'agissait d'une extension de la séquence de visualisation.

Contenus

Notre groupe suivant est le groupe des contenus. Il est fondé sur l'idée initiale de placer ensemble les visualisations de données et les médias complémentaires, deux dimensions qui répondent toutes deux au besoin de décrire comment l'information a été communiquée. Nous avons ajouté une troisième dimension à ce groupe, le tutoriel. L'idée est que le tutoriel est un élément d'information qui doit être communiqué au public.

Données

Reste le groupe “Données”. Nous avons deux dimensions qui décrivent directement les données, Type et Forme, et la dimension Analyse traite des procédés qui ont pu y être appliqués pour les transformer. Nous disposons donc de trois dimensions qui permettent de former un groupe portant sur les données et leur traitement.

3.1.5 Version finale de la taxonomie

La version finale de notre taxonomie compte 20 dimensions, réparties entre les 6 groupes “métadonnées”, “contexte”, “Données”, “contenus”, “contrôle” et “schémas narratifs”. Elle est présentée dans le tableau 3.9. Nous avons aussi proposé des définitions des groupes et dimensions, qui sont détaillés en annexe A.

TABLEAU 3.9 Récapitulatif de la taxonomie

Groupe	Dimension	Valeurs
Metadonnées	Titre	Valeur unique
	Date	Valeur unique
	Source	Valeur unique
Contexte	But	Informar, persuader, divertir, réconforter, expliquer
	Sujet	Arts, culture, divertissement et médias ; conflit, guerre et paix ; criminalité, droit et justice ; désastres et accidents ; économie et finances ; éducation ; emploi ; environnement ; intérêt humain ; météo ; politique ; religion et croyance ; santé ; science et technologie ; société ; sport ; vie quotidienne et loisirs
	Structure	Animation, compilation, diaporama, image statique, infographie déroulante, multimédia, storymap, autre
Données	Forme	Qualitative, quantitative
	Type	Numériques, textes, spatiales, réseau, temporelles
	Analyse	Oui, non
Contenus	Visualisations de données	Autre, carte à flux, carte à points, carte à symbole proportionnels, carte choroplèthe, carte de chaleur, carte de navigation, cartogramme, diagramme à barres, diagramme à bulles, diagramme à formes proportionnelles, diagramme circulaire, diagramme d'instance, diagramme de coordonnées parallèles, diagramme de sankey, graphique de flux, graphique en aires, graphique radar, histogramme, infographie, ligne, modèle 3D, nuage de points, organigramme, pictogramme, range chart, réseau, table, tracé en mosaïque
	Médias	Audio, image, texte, vidéo
	Tutoriel	Sans, explicite, implicite
Contrôle	Degré de liberté de l'utilisateur	Axé sur l'auteur, diaporama interactif, verre à martini histoire à forage, axé sur le lecteur
	Séquence	Sans séquençage, navigation linéaire, navigation non-linéaire
	Transitions	Atténuation, panoramique, glissement, interpolation
	Indicateur de progression	Sans indicateur, barre de progrès, boutons de section, ligne temporelle, carte géographique, fil d'Ariane, liste de pointage
	Interactions	Reconfigurer, encodage, filtrer, abstraction/élaboration, connection, exploration, sélection, autres
Schémas Narratifs	Schémas Argumentatifs	Comparaison, concrétisation, répétition
	Schémas de Cadrage	Rupture de conventions, défamiliarisation, familiarisation, métaphore physique
	Schémas d'Engagement de d'Empathie	Sans, accélération/décélération, rupture du 4ème mur, devinette, exploration, les humains derrière les points, question rhétorique, révélation graduelle

3.2 Archive

L'étape suivante du projet de ce mémoire consiste à rassembler des exemples de RCD pour alimenter notre archive. Cette étape correspond à l'objectif O2 : Rassembler une archive d'exemples de RCD pertinents, issus de sources de qualité, et annotés selon la taxonomie proposée.

3.2.1 Collecte d'éléments pour l'archive

Pour concevoir l'archive, l'étape la plus importante est de collecter des entrées afin de fournir un contenu pour l'archive. Elle se divise entre le choix des sources et l'annotation de l'archive.

Choix des sources

La première étape de notre démarche a été la collecte de sources où nous pourrions trouver des RCD de bonne qualité pour construire notre archive. Pour cela, nous avons exploré deux axes : les travaux académiques préexistants et la presse en ligne.

Travaux académiques préexistants Notre idée initiale était d'utiliser comme source les articles que nous avons employés pour créer la taxonomie. Leurs auteurs ayant basé leurs travaux sur l'examen de RCD préexistant, leurs articles auraient formé une base solide pour notre archive. En explorant cette possibilité, nous avons constaté que ces listes sont très irrégulières. En considérant nos sources :

1. *Narrative Visualization : Telling Stories with Data* [7] : Une liste de 58 visualisations est proposée avec l'article.
2. *Classifying Data Journalism* [20] :
3. *Narrative Design Patterns for Data-Driven Storytelling* [19] : Une petite quantité d'exemples sont présentés sous forme de captures d'écran illustratives.
4. *Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Visual Stories* [25] : Une liste de 45 exemples est fournie avec l'article.
5. *Exploration and Explanation in Data-Driven Storytelling* [18] : Une petite quantité d'exemples sont présentés sous forme de captures d'écran illustratives.
6. *A Conceptual Framework For Interactive Cartographic Storytelling* [21] : La liste d'exemples couvre uniquement des cartes géographiques.
7. *Patterns in Award-Winning Data Storytelling* [17] : Les exemples utilisés sont représentés avec une capture d'écran partielle du tableur ou ils sont listés.

8. *Cartographic Design as Visual Storytelling : Synthesis and Review of Map-Based Narratives* [22] : Les exemples sont peu nombreux et présentés sous forme de captures d'écran limitées.
9. *Visualization design patterns* [23] : Aucun exemple n'est listé.
10. *The Stories We Tell About Data : Media Types for Data-Driven Storytelling* [35] : Plusieurs exemples sont cités, mais ils sont peu nombreux, dispersés dans l'article et sont dominés par des exemples de RCD exceptionnels qui sortent de notre cadre, comme la représentation d'un algorithme par de la danse.

Deux de nos sources ont des listes d'exemples significatives, qui pourraient fournir une base. D'autres pourraient nous permettre d'en trouver d'autres, pour un total de l'ordre de 150. Une grande partie de ces références provient de *Narrative Visualization : Telling Stories with Data* [7], et sont donc vieilles de près de 14 ans. Nous avons donc choisi de rechercher une méthode alternative.

Presse en ligne Notre idée de substitution consiste à utiliser une source différente : les journaux en ligne. Nous avons cherché à identifier des publications qui produisent régulièrement des RCD, pour collecter leurs projets récents afin d'alimenter l'archive.

Afin de sélectionner les publications les plus pertinentes, nous avons recherché une méta-source, une référence qui puisse éclairer notre choix. Le blog EagerEye <https://eagereyes.org/>, tenu par Robert Kosara, apparaît dans les références de *Data-Driven Storytelling Techniques : Analysis of a Curated Collection of Visual Stories* [25]. Durant plusieurs années, Kosara a fourni annuellement des listes de “best-of de l'année” de grands journaux, mettant en valeur des dizaines de RCD de bonne qualité à chaque itération. À partir de ces listes, nous avons pu identifier plusieurs publications qui compilent annuellement leurs meilleurs travaux.

Une fois ces publications identifiées, nous avons ensuite rassemblé leurs compilations les plus récentes¹. Quand cela était possible, nous avons extrapolé à partir de certaines de ces listes pour retrouver leurs autres instances. Par exemple, Kosara a mis en valeurs une liste des meilleurs RCD de l'année produite annuellement par Propublica, et nous avons récupéré les éditions de cette liste sur toutes les années où elle a été publiée. Il en résulte une liste de 54 listes de RCD issus des publications suivantes : Al-Jazeera, BBC, Bloomberg, FiveThirtyEight, Los Angeles Times, New York Times, Propublica, Wall Street Journal.

Ces listes comptent parfois plusieurs dizaines de références, sans redondance puisqu'elles correspondent chacune à une combinaison différente de journal et d'année. Nous disposons

1. La collecte a été effectuée au début de l'année 2023. Les compilations remontent donc jusqu'à 2022

donc d'une liste d'exemples volumineuse (Près de trois mille recensés avant que nous ne commençons à les trier), et qui répondent à un certain standard de qualité.

La problématique suivante reste celle du volume, mais prise dans le sens inverse : combien d'exemples pouvons nous annoter ? Une innovation apparue dans un RCD récent est plus pertinente qu'une innovation apparue dans un RCD publié en 2014 (la plus ancienne date de notre collecte). Nous avons donc pris la décision de donner la priorité aux exemples les plus récents. A partir de ce point, nous avons choisi d'annoter année par année, en partant des exemples les plus récents, deux de 2022, et après avoir annoté chaque année d'évaluer la taille de l'archive et le temps qui a été nécessaire pour atteindre ce point.

La liste finale couvre les années 2020 à 2022, avec un total de 780 exemples. Les compilations d'articles qui nous ont fourni ces exemples sont celles listées dans le tableau 3.10.

TABLEAU 3.10 Sources identifiées

Titre de la liste	Page d'origine
<i>The New York Times, 2022 : The Year in Visual Stories and Graphics</i>	https://www.nytimes.com/interactive/2022/12/28/us/2022-year-in-graphics.html
<i>The New York Times, 2021 : The Year in Visual Stories and Graphics</i>	https://www.nytimes.com/interactive/2021/12/29/us/2021-year-in-graphics.html
<i>The New York Times, 2020 : The Year in Visual Stories and Graphics</i>	https://www.nytimes.com/interactive/2020/12/30/us/2020-year-in-graphics.html
<i>The Washington post ; 2022 : The year in graphics</i>	https://www.washingtonpost.com/media/interactive/2022/2022-year-graphics/
<i>Bloomberg, 2020 in Graphics</i>	https://www.bloomberg.com/graphics/2020-in-graphics/#xj4y7vzkg
<i>Bloomberg, 2021 : The Year in Graphics</i>	https://www.bloomberg.com/graphics/2021-in-graphics/#xj4y7vzkg
<i>Bloomberg, 2022 : The Year in Graphics</i>	https://www.bloomberg.com/graphics/2022-in-graphics/#xj4y7vzkg
<i>The Wall Street Journal, 2020 The Year in Graphics</i>	https://www.wsj.com/graphics/year-in-graphics-2020/
<i>The Wall Street Journal, 2021 The Year in Graphics</i>	https://www.wsj.com/articles/2021-the-year-in-graphics-11640817091
<i>The Wall Street Journal, 2022 The Year in Graphics</i>	https://www.wsj.com/articles/2022-the-year-in-graphics-11671737421
<i>ProPublica, 2022 : Our Year in Visual Journalism</i>	https://www.propublica.org/article/year-in-visual-journalism-2022
<i>ProPublica (2021) ProPublica's Year in Visual Journalism</i>	https://www.propublica.org/article/2021-year-in-visual-journalism
<i>LA Times, 2020 : The Year in Data and Graphics</i>	https://www.latimes.com/projects/data-and-graphics-2020/

Nous remarquons que Al-Jazeera, la BBC et FiveThirtyEight ne sont pas représentés, car leurs compilations les plus récentes que nous avons pu retracer à partir des listes de Kosara sont antérieures à 2020. Autre observation, toutes ces sources sont anglophones, et avec les

coupures, nous avons perdu les deux références qui ne sont pas étasuniennes.

3.2.2 Annotation de l'archive

Une fois les exemples identifiés, nous les avons annotés à l'aide de la taxonomie. Pour cela, nous avons créé un tableur avec une colonne pour chaque dimension. Initialement, nous n'avons rempli que les deux premières colonnes : Titre et Lien. Nous avons ensuite parcouru cette liste, en ouvrant chaque lien pour consulter l'exemple, et rempli la colonne correspondants à chaque dimension avec les valeurs de cette dimension que nous avons repéré dans l'article. Cette partie de l'annotation a été réalisée par deux personnes, sur une période totale de quatre mois.

L'archive a été, comme annoncé dans la section 3.2.1, étendue avec des années de plus en plus ancienne au fur et à mesure que nous l'avons complétée. Nous avons aussi modifié la liste de valeurs de la dimension “visualisations”, en ajoutant les types de visualisations rencontrés au cours de l'annotation.

Nous avons aussi constaté que certains exemples de notre liste sont éloignés de notre définition de RCD. Il s'agit par exemple d'entrées provenant de *Propublica, 2022 : Our Year in Visual Journalism*, car le journalisme visuel couvre aussi des photo-reportages. Notre objectif est de fournir de l'inspiration pour la conception de RCD, et non pas de lister des RCD. Nous avons donc choisi d'intégrer des exemples d'articles qui n'étaient pas des RCD mais qui ont du potentiel en tant qu'inspiration. Pour identifier ce qui fait qu'un exemple a du potentiel en tant qu'inspiration, nous avons considéré qu'il doit présenter au moins un élément qui s'écarte du format classique d'un texte illustré par des photographies. Nous avons donc considéré que le minimum nécessaire pour qu'un exemple soit présent dans l'archive est la présence d'au moins une des caractéristiques suivantes :

- La présence de visualisations de données
- L'utilisation d'une méthode particulière appartenant à la dimension “Analyse”
- Une valeur autre que “Sans” dans une des dimensions du groupe “Contrôle”, qui indique la présence d'une interaction.

3.3 CurratedDDS

La dernière partie de ce chapitre est la conception de l'outil CurratedDD², qui répond à l'objectif O3 : Concevoir un outil de recherche d'exemples de RCD qui permette à nos uti-

2. L'outil est disponible en suivant ce lien : <https://script.google.com/macros/s/AKfycbydgUIUH8NoH8cJpVmKf2WVyDwmC3T8k0QEUUaB9Bw/dev>

lisateurs d'explorer l'archive pour y trouver des exemples, en triant son contenu à l'aide des informations de la taxonomie.

3.3.1 Principe de fonctionnement

Objectif de conception et démarche

Notre objectif en concevant CurratedDDS est de permettre à nos utilisateurs d'accéder à l'archive, et de rechercher des exemples pertinents pour leurs travaux. La principale méthode pour obtenir des exemples pertinents reposera sur le filtrage des entrées de l'archive, en utilisant les annotations des entrées.

Nous proposons le flux de travail suivant :

- L'utilisateur choisit les dimensions et valeurs de la taxonomie qui lui semblent pertinentes pour son projet.
- Le côté serveur de l'outil filtre les exemples qui correspondent à cette liste.
- L'interface de l'outil présente les exemples à l'utilisateur.
- L'utilisateur choisit les exemples qui l'intéressent le plus, avant de les consulter sur leurs pages originales.

Nous avons donc besoin de présenter la taxonomie à notre utilisateur, lui permettre de formuler une requête à partir de la taxonomie, effectuer le filtrage à partir de cette requête, puis présenter ces résultats à l'utilisateur.

Outils utilisés

CurratedDDS a été développé pour être accessible sur les ordinateurs de travail des participants. Nous avons donc développé l'outil sous la forme d'une page web hébergée sur la plateforme Google App Script. Le prototype a été développé en utilisant les langages de programmation favorisés par cette plateforme, HTML, CSS, Javascript et Google Script.

Nous avons structuré le code entre trois fichiers : Code.gs, Index.html et le fichier archive. Le fichier Index accueille l'interface visuelle, les interactions avec la page, ainsi que la formulation et la soumission de la requête. Le fichier Code accueille la gestion de la comparaison de la requête avec l'archive. Le fichier archive est un tableur google sheet qui contient les entrées collectées et annotées.

3.3.2 Interface graphique

L'interface graphique de CurratedDDSse divise entre une barre de formulation des requêtes et un tableau des résultats. La figure 3.1 présente l'interface.

Barre de formulation des requêtes

Pour permettre à l'utilisateur de formuler sa requête, nous avons utilisé un formulaire, présenté sous la forme d'une barre d'outils en haut de page. Étant donné que nous avons un grand nombre de dimensions, nous avons choisi d'utiliser les groupes de la taxonomie comme menus déroulants. La présentation initiale, la figure 3.2 liste uniquement les groupes, et permet de choisir les groupes à ouvrir.

Chaque menu contient ensuite une liste des dimensions qui font partie du groupe, avec l'option de dérouler les valeurs de la dimension. Les menus, tous déployés, sont présentés sur la figure 3.3. Chaque dimension est gérée comme une entrée du formulaire. Selon la dimension, nous avons utilisé plusieurs formes pour ces entrées. Nous avons utilisé des cases à cocher, boutons radio et listes déroulantes selon les dimensions.

Les dimensions “structure de l'article”, “forme des données”, “type”, “médias”, “séquence”, “transition”, “interactions”, “schémas argumentatifs”, “schémas de cadrage”, et “schémas d'engagement et d'empathie” sont toutes des dimensions dont les valeurs peuvent être complémentaires, avec de petits nombres de valeurs possibles. Par exemple, un RCD peut utiliser à la fois des données de type numérique et temporel. Nous avons donc utilisé des cases à cocher.

Les dimensions “journal de publication”, “année de publication”, “but”, “sujet”, “analyse”, “tutoriel”, “degré de liberté de l'utilisateur” et “indicateurs de progression” sont des dimensions où une seule valeur peut s'appliquer en même temps au même exemple. Nous avons utilisé des listes déroulantes pour les dimensions avec le plus grand nombre de valeurs, afin de les rendre plus compactes. Il s'agit des dimensions “journal de publication”, “but”, “sujet”, et “indicateurs de progression”. Nous avons aussi utilisé une liste pour l'année de publication. Les autres dimensions, “analyse”, “tutoriel” et “degré de liberté”, sont représentées avec des boutons radio, plus volumineux mais qui permettent de voir toutes les options. La dernière dimension, “visualisations”, est très longue et permet des combinaisons de valeurs. Nous avons utilisé une liste déroulante avec cases à cocher.

Explication de la taxonomie Nous avons estimé que la taxonomie a besoin d'explications. Nous avons donc ajouté des infobulles contenant les explications des groupes, dimen-

▼
▼ | Métadonnées ⓘ
▼ | Contexte ⓘ
▼ | Données ⓘ
▼ | Contenus ⓘ
▼ | Contrôle ⓘ
▼ | Schémas Narratifs ⓘ
Effacer

Métadonnées

Contexte

Données

Contenus

Contrôle

Schémas Narratifs

Effacer

Titre

But

Forme des données
☐ Qualitative (439)
☐ Quantitative (615)

Journal de publication

Sujet

Type

Structure de l'article

☐ Animation (63)
☐ Compilation (17)
☐ Diaporama (15)
☐ Image Statique (3)
☐ Infographie déroulante (699)
☐ Multimédia (1)
☐ Storymap (31)
☐ Autre (17)

☐ Numérique (614)
☐ Réseau (6)
☐ Spatial (496)
☐ Temporel (481)
☐ Texte (114)
☐ Autre (88)

Analyse ☒ Individuel (39) ☐ Non (741)

Visualisation

Médias

Tutoriel

Degré de liberté de l'utilisateur

Séquence

Transitions

☐ Axé sur l'auteur (698)
☐ Verre à Martini (31)
☐ Diaporama interactif (10)
☐ Forage (14)
☐ Axé sur le lecteur (25)

☐ Sans séquençage (9)
☐ Linéaire continue (scroll) (731)
☐ Linéaire par étapes (boutons, slides) (17)
☐ Navigation non-linéaire (14)

☐ Sans transition (539)
☐ Atténuation (fading) (97)
☐ Glissement (swiping) (164)
☐ Interpolation (tweening) (77)
☐ Panoramique (panning) (76)

Schémas argumentatifs

Schémas de cadrage

Schémas d'engagement et d'empathie

☐ Sans (175)
☐ Comparaison (451)
☐ Concrétisation (305)
☐ Répétition (338)

☐ Sans (479)
☐ Défamilliarisation (37)
☐ Familiarisation (193)
☐ Métaphore physique (71)
☐ Rupture de conventions (71)

☐ Les humains derrière les points (163)
☐ Question rhétorique (30)
☐ Révélation graduelle (154)

Indicateurs de progression

Interactions

☐ Sans interaction (/548)

sions et valeurs. Nous avons conservé les explications écrites pour la taxonomie. Les groupes sont l'unité de plus haut niveau, et la première chose que voit l'utilisateur en accédant à l'outil. Nous avons donc rendu leurs infobulles immédiatement visibles en les associant à des boutons, avec le logo d'information "i". Les dimensions ont reçu des infobulles explicatives qui s'affichent quand le curseur de l'utilisateur passe sur le nom de la dimension. Certaines valeurs ont reçu le même traitement quand nous avons estimé que la dimension avait spécifiquement besoin d'explications.

Fonctionnalités complémentaires En plus de la sélection d'options, nous avons trois autres interactions à considérer avec la requête. La première de ces interactions est la soumission du formulaire, pour finaliser la requête. Afin de réduire le nombre de clics nécessaires pour l'action principale de l'outil, et simplifier son utilisation, nous avons choisi de créer une interface responsive, et de soumettre automatiquement la requête à chaque modification. Quand un utilisateur sélectionne ou désélectionne une option, la requête est immédiatement formulée et les exemples correspondants affichés. La deuxième de ces interactions est la réinitialisation du formulaire. Puisque nous soumettons une requête à chaque changement, un utilisateur qui décide de reprendre sa recherche du début va causer une longue liste de requêtes non désirées, et peuvent devenir chronophages. Pour résoudre ce problème, nous proposons un bouton "réinitialiser" pour désélectionner immédiatement toutes les options. Étant donné que les options sont initialement cachées derrière 6 menus déroulants, nos utilisateurs doivent ouvrir les menus qui les intéressent. Nous avons envisagé un cas d'usage où les utilisateurs chercheraient à ouvrir tous les menus déroulants pour faire leur choix, ou à fermer tous les menus après une requête pour passer de la formulation de la requête à la lecture des exemples proposés. Afin de faciliter cette démarche, nous avons ajouté une flèche d'ouverture/fermeture totale, placée à gauche des menus déroulants, afin d'en faire le premier bouton disponible.

Nous avons aussi observé des scénarios extrêmes tels qu'une combinaison très spécifique qui ne correspond à aucun exemple, ou bien une requête simple qui laisse des centaines de possibilités. Pour améliorer cette situation, nous avons choisi d'ajouter un compteur de valeurs. Quand la requête est récupérée, nous comptons le nombre d'exemples correspondant à chaque valeur, pour afficher ce nombre à côté de la valeur. L'objectif est de permettre à nos utilisateurs de savoir combien de résultats ils obtiendront en filtrant avec la valeur en question, et donc d'évaluer si un nouveau filtre va trop réduire leur liste d'exemples, ou au contraire ne pas suffisamment la réduire. Ce compteur est illustré par la figure 3.4.

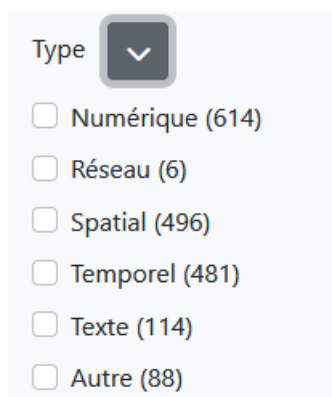


FIGURE 3.4 Compteurs pour les valeurs de la dimension “Type”

Présentation des résultats

La deuxième partie de l’interface est la présentation des résultats. Nos utilisateurs peuvent envoyer une requête à notre archive, mais il reste à leur fournir des résultats lisibles. Nous affichons les exemples sous la forme d’un tableau, visible sur la figure 3.5, avec une ligne pour chaque exemple et une colonne pour chaque dimension.

En raison du nombre de colonnes (21), afficher toutes les colonnes compresse fortement le texte, et rends l’information difficilement lisible. Nous avons donc choisi d’utiliser un tableau large, qui repose sur un défilement horizontal, plutôt que d’afficher tout son contenu simultanément.

Afin de rendre cet affichage plus clair, nous avons mis en place trois mesures. Premièrement, nous avons utilisé un double en-tête pour le tableau, avec les groupes et dimensions, afin d’établir une correspondance claire avec la taxonomie. Deuxièmement, nous avons utilisé une alternance de couleurs entre le gris et le blanc pour les lignes, afin de faciliter leur suivi. Ceci était particulièrement nécessaire car le grand nombre de colonnes rend le tableau très large. Troisièmement, la principale interaction que nous attendons de l’utilisateur est la consultation d’exemples. Nous avons donc mis en valeur le lien pour ouvrir l’article. Il prend la forme d’un bouton, “Ouvrir” dans le tableau, avec une couleur plus vive que le reste de l’interface pour le rendre plus visible.

Nous avons aussi ajouté des interactions avec le tableau pour faciliter l’exploration des résultats d’une requête. Cliquer sur une dimension permet de trier alphabétiquement les exemples, en fonction des valeurs de la dimension. Une barre de recherche a aussi été ajoutée, qui permet de chercher des mots clés dans toute l’archive.

Search:										
Métadonnées				Contexte			Données		Contenus	
Titre	Lien	Journal	Année	But	Sujet	Structure	Forme	Type	Analyse	Visualisation de données
\$600 Unemployment: What Happens When a Stimulus Lifeline Ends	Ouvrir	The New York Times	2020	Expliquer	Économie et finances	Infographie déroulante	Quantitative	Numérique, Spatial	Non	Nuage de points
101 Best L.A. Restaurants of 2020	Ouvrir	Los Angeles Times	2020	Inform, Divertir	Vie quotidienne et loisirs	Infographie déroulante	Qualitative	Texte	Non	-
13,000 Missing Flights: The Global Consequences of the Coronavirus	Ouvrir	The New York Times	2020	Expliquer	Économie et finances, Désastres et accidents, Santé	Infographie déroulante, Animation	Qualitative, Quantitative	Numérique, Spatial, Temporel	Non	Graphique en aires, Modèle 3D, Pictogramme
150 Years Ago Brooklyn Renumbered All Its Streets. It Was a Disaster.	Ouvrir	The New York Times	2021	Expliquer	Autres	Infographie déroulante	Qualitative	Texte	Non	-
179 Reasons You Probably Don't Need to Panic About Inflation	Ouvrir	The New York Times	2021	Inform	Économie et finances	Infographie déroulante	Quantitative	Numérique, Temporel	Non	Diagramme d'instance, Ligne
2,000 Headlines. Here Is a First Draft of Trump's Legacy.	Ouvrir	The New York Times	2021	Inform	Politique	Infographie déroulante, Animation	Qualitative	Temporel, Texte	Non	-
2020 election: How women of color impact Congress diversity	Ouvrir	Los Angeles Times	2020	Inform	Politique	Image statique	Qualitative, Quantitative	Numérique	Non	Graphique en gauffre
2020 in photos and videos	Ouvrir	Los Angeles Times	2020	Inform, Expliquer	Désastres et accidents, Intérêt humain, Politique, Santé, Société	Compilation	Qualitative, Quantitative	Texte, Numérique, Temporel	Non	Autre
2020 Iowa caucus results	Ouvrir	Los Angeles Times	2020	Inform	Politique	Infographie déroulante	Qualitative, Quantitative	Numérique, Spatial	Non	Carte Choropleth, Graphique à tarte

FIGURE 3.5 Présentation des exemples

3.3.3 Côté serveur

Le traitement de la requête est fait par le fichier Code.gs. La requête retourne un objet `formObject`. Ce `formObject` est transformé en liste de critères dimension par dimension. Un critère est une chaîne de texte reproduisant le contenu attendu dans une case du tableur. Vérifier la présence de cette chaîne de texte nous permet de vérifier si un exemple répond bien à la demande. Dans le cas des dimensions qui peuvent avoir plusieurs valeurs simultanées, nous divisons ces critères en liste de sous-critères, une liste de chaînes de texte qui doivent toutes être présentes. Ceci permet de tester séparément deux valeurs qui appartiennent à une même dimension mais ne sont pas successives. Par exemple, si un utilisateur fait une requête pour “Comparaison” et “Répétition” dans les schémas argumentatifs, il recevra des entrées d’archive annotées “Comparaison, Répétition” et “Comparaison, Concrétisation, Répétition”. L’algorithme importe tous les exemples de l’archive, puis les compare aux critères. Seuls les exemples qui correspondent à tous les critères sont conservés, et transmet au front-end pour être affichés.

CHAPITRE 4 ÉVALUATION

4.1 Protocole expérimental

Cette section présente le protocole d'évaluation de l'outil CurratedDDS développé dans le cadre de cette recherche. Elle porte principalement sur la méthode d'évaluation, la sélection des participants, et la conduite des entretiens. Comme ce projet nécessitait une évaluation par des personnes, le protocole d'évaluation a été préalablement approuvé par le comité d'éthique de la recherche (CÉR) de Polytechnique Montréal¹.

4.1.1 Méthode d'évaluation

Notre outil a été conçu pour s'adresser à des professionnels qui conçoivent des visualisations de données. Il s'agit d'un public restreint et spécialisé. Nous cherchons à collecter des résultats qualitatifs concernant l'impact de CurratedDDS sur leur pratique professionnelle. En effet, l'adéquation de la taxonomie, la pertinence des exemples, et l'effet de CurratedDDS sur le travail d'idéation sont tous des données qui n'ont pas de mesure quantitative.

Nous avons donc choisi de conduire une étude de cas sur un groupe de professionnels des médias. Ceci nous permet de collecter des données qualitatives en profondeur sur le champ restreint que représente notre prototype. Le principe de cette étude de cas consisterait à introduire notre outil à des utilisateurs potentiels, à les inviter à utiliser l'outil pour une période de temps, puis de collecter leurs retours sous la forme d'entretiens semi-directifs. Nous cherchons des retours qualitatifs d'utilisateurs sur l'outil, et à travers ce dernier sur la taxonomie et l'archive qui ont été utilisées pour sa conception.

4.1.2 Sélection des participants

Dans le cadre d'un projet Alliance CRSNG, le journal Le Devoir a participé à cette évaluation. Les concepteurs de RCD du Devoir ont été invités à participer à notre étude de cas. Les participants sont des designers, journalistes de données et ingénieurs travaillant pour le partenaire industriel. Nous avons aussi pris contact avec plusieurs journalistes de données de Radio Canada pour augmenter la taille de notre échantillon. Notre échantillon compte 7 concepteurs de RCD, 5 de Le Devoir et 2 de Radio Canada.

L'un des participants a mis fin à sa participation à l'expérience suite à la période d'utilisation

1. Numéro de certificat CER-2324-30-D

libre, car il a estimé que ses tâches professionnelles ne lui permettaient pas d'utiliser CurratedDDS dans la période de temps impartie. Nous identifierons les participants des entretiens comme P1, P2, P3, P4, P5, P6.

Ces participants étant des professionnels dans le domaine des médias et de la visualisation de données, nous assumons qu'ils sont majeurs, à l'aise avec les ordinateurs, et n'ont pas de défauts visuels les empêchant de travailler avec des visualisations de données. Un filtrage n'est donc pas nécessaire pour nous assurer qu'ils sont en condition de participer à notre expérience.

4.1.3 Présentation du projet et utilisation en contexte

La première étape de l'expérience est la mise en place. Elle consiste à présenter le projet aux participants et à leur donner une période de temps pour tester l'outil dans leur contexte professionnel.

Présentation du projet

Une fois nos participants recrutés, nous avons organisé une présentation du projet. Étant donné que nous nous adressons à un groupe limité de professionnels, nous considérons que nous nous adressons à un public qui peut être formé à l'utilisation de l'outil. Nous avons donc invité les participants à une rencontre en ligne afin de leur présenter notre démarche ainsi que les tâches attendues et l'outil employé.

Cette réunion comprend : Une explication brève du contexte du projet, une explication de la tâche attendue durant la phase d'utilisation, une phase de questions-réponse portant sur le projet et la tâche attendue, une démonstration de l'outil employé, et une phase de questions-réponse sur l'utilisation de l'outil. La rencontre a été répétée 3 fois : une fois en rencontre groupée avec les participants du Devoir, qui ont rejoint le projet en équipe, et deux fois en rencontres individuelles pour les participants de Radio Canada, qui ont rejoint le projet individuellement.

Période d'utilisation en contexte

La phase suivante du déroulement de l'expérience consiste à suivre les participants lors d'un projet de conception de RCD mené dans le cadre de leur activité professionnelle. Les participants ont été invités à utiliser CurratedDDS durant la phase d'idéation de leur prochain projet, avec une durée attendue de 2 semaines. Les participants sont autonomes dans leur

utilisation de l'outil et libre de leur utilisation. Les expérimentateurs étaient disponibles par courriel afin de fournir une assistance technique uniquement.

Durant la phase de présentation, les participants ont soulevé la question de l'utilisation de CurratedDDS a des fins d'explorations en dehors d'un projet en cours. Un scénario qui leur semblait prometteur était une vigilance professionnelle sur les exemples de RCD sans chercher d'exemples pour un projet particulier. Nous avons pris cette remarque en compte, et communiqué aux participants qu'ils pouvaient utiliser CurratedDDS librement pendant la période d'utilisation.

4.1.4 Entretiens

La dernière étape du projet prend la forme d'une série d'entretiens individuels afin de collecter les retours des utilisateurs sur l'outil CurratedDDS. Ces entretiens sont semi-directifs, afin de collecter les retours des participants en donnant la priorité à la flexibilité et à la profondeur de leur réponse.

Organisation des entretiens

Les entretiens ont été organisés individuellement. La date initialement prévue était deux semaines après le lancement de la période d'utilisation libre. L'organisation des entretiens a mené à une dispersion des entretiens dans les jours qui ont suivi cette date. Deux entretiens ont été reportés au mois suivant en raison des disponibilités des participants.

Les entretiens ont été conduits en visioconférence. Ces entretiens ont été enregistrés pour garantir une prise de notes précise et une analyse plus exacte de l'échange. La durée estimée des entretiens était de une heure. Les entretiens en pratique ont duré entre 40 et 80 minutes. Leur consentement pour cet enregistrement a été demandé aux participants dans le formulaire de consentement avant l'expérience, et vérifié verbalement avant le début de l'entretien.

Guide d'entretien

Pour mener les entretiens, nous avons rédigé un guide structuré par les sections suivantes :

- Une introduction, qui couvre la prise de contact, présentation, et explications de l'entretien.
- Une série de questions sur l'outil visant à évaluer nos hypothèses H4 et H5. Nous y avons aussi intégré des questions portant sur les exemples pour évaluer notre hypothèse H3. Les hypothèses sont les suivantes :

- H3 : L’archive proposée contient une sélection d’exemples pertinente pour les concepteurs de RCD comparativement aux exemples auxquels ils sont exposés hors de l’archive
- H4 : L’outil CurratedDDS améliore l’accès à des exemples pour des concepteurs de RCD, comparativement à leurs pratiques habituelles sans cet outil.
- H5 : L’utilisation de CurratedDDS facilite le travail d’idéation des concepteurs de RCD.
- Une série de questions sur l’utilisabilité de l’outil, pour séparer les problèmes liés à l’ergonomie de l’intérêt des fonctions principales de CurratedDDS.
- Une série de questions sur la taxonomie visant à évaluer nos hypothèses H1 et H2 :
 - H1 : La taxonomie proposée caractérise adéquatement les RCD pour leur usage par des concepteurs de RCD.
 - H2 : La taxonomie utilise un vocabulaire qui correspond à la réalité de la pratique des concepteurs de RCD.
- Une conclusion, pour permettre à la personne interviewée d’ajouter ses commentaires qui n’étaient pas couverts par les questions principales et mettre fin à l’interview.

La liste de questions complète se trouve dans l’annexe B.

4.2 Résultats et analyse

Suite à l’expérimentation, nous disposons des transcriptions de nos échanges avec six participants. Nous avons utilisé ces échanges pour évaluer nos hypothèses et tirer des conclusions. Pour cela nous avons considéré les retours faits sur la taxonomie, ceux faits sur l’archive, les retours sur l’influence de CurratedDDS sur le travail des concepteurs de RCD et les idées avancées pour améliorer CurratedDDS.

4.2.1 Retours sur la taxonomie

Nos hypothèses H1 et H2 portent sur la taxonomie.

- H1 : La taxonomie proposée caractérise adéquatement les RCD pour leur usage par des concepteurs de RCD.
- H2 : La taxonomie utilise un vocabulaire qui correspond à la réalité de la pratique des concepteurs de RCD.

Longueur de la taxonomie

Une des observations récurrentes des participants est la longueur de la taxonomie. Cette dernière comporte 6 groupes, 20 dimensions et 119 valeurs. Tous les participants de l'expérience ont remarqué que ce nombre est excessif, et déborde l'utilisateur en le noyant sous un nombre excessif de possibilités. Une minorité d'options est au final très souvent utilisée, tandis que d'autres ne semblent pas pertinentes pour notre objectif, qui est d'employer la taxonomie comme outil de filtrage. Les sections "contrôles" et "schémas narratifs" ont été très peu utilisées afin de filtrer les exemples.

Une remarque faite par un des participants de l'expérience est la différence entre les éléments de taxonomie pertinents pour le filtrage, avec des dimensions qu'un utilisateur pourrait rechercher activement, et les éléments de taxonomie pertinents pour la description de l'exemple. Ainsi, les dimensions pourraient être divisées en deux grands groupes : Un pour les dimensions pertinentes à la fois pour le filtrage et la description, et un pour les dimensions utiles uniquement pour décrire le RCD. Ceci permettrait de simplifier la recherche sans perdre d'information dans la taxonomie. Une autre idée avancée par un participant (P1) est que le problème réside principalement dans l'interface utilisateur de CurratedDDS, et qu'un remaniement de l'interface permettrait de rendre la taxonomie plus claire sans la modifier.

Qualité des dimensions

Concernant la qualité des dimensions de la taxonomie, les retours sont variables selon les dimensions.

Données Les données ont été mentionnées plusieurs fois, avec des éléments positifs et des manques. Deux participants (P1 et P2) annoncent avoir utilisé fréquemment le groupe "Données" sans précisions supplémentaires. Le type de données a été mentionné comme une des dimensions fréquemment utilisées, mais le participant P2 a aussi remarqué que la dimension présente des problèmes de clarté : la valeur "numérique" n'est pas claire, et cause de la confusion. Pour le citer "tout est numérique", et le terme ne fait sens que comparativement aux autres options, telles que "spatiales" et "temporelles". Un autre remarque que les données spatiales et temporelles sont des données numériques. Deux participants (P1 et P3) mentionnent avoir utilisé la forme des données comme moyen de recherche dans l'archive, mais un autre remarque que la séparation qualitative/quantitative est peu pertinente pour lui, car les visualisation de données sont le plus souvent basées sur des données quantitatives. Un participant (P3) remarque aussi que le type et la forme des données sont intéressants mais que son utilisation habituelle de ces concepts est inconsciente en traitant son jeu de données.

La troisième dimension du groupe données est l'analyse. Trois participants (P1, P2 et P6) portent des retours négatifs sur cette dimension, et aucun n'a mentionné y avoir trouvé un intérêt. Le terme manque de clarté, et porte à confusion. Passer d'"analyse informatisée" à "analyse" a été une simplification excessive, car tous les RCD contiennent une forme d'analyse faite par ses créateurs.

Une remarque faite par deux des participants (P2 et P3) porte sur une information manquante dans les données : la source. Nous avons estimé en rédigeant la taxonomie que la source des données n'était pas pertinente pour notre cas d'usage, car nous avons considéré que la collecte des données est une étape antérieure à la conception du RCD. Nous nous attendions donc à ce que nos utilisateurs aient déjà leurs propres données, et ne s'intéressent pas à l'origine des données des exemples. Il apparaît que certains participants préfèrent savoir comment les auteurs d'un exemple ont obtenu leurs données. Ils considèrent que connaître l'origine des données d'un RCD leur permet d'avoir confiance dans l'exemple, et leur permet d'identifier des sources potentielles pour leurs futurs projets. Retirer la source de notre taxonomie réduit donc sa pertinence.

Métadonnées Les métadonnées ont été mentionnées brièvement à plusieurs reprises, mais sont peu commentées. Trois participants (P1, P3 et P4) ont mentionné utiliser régulièrement les métadonnées. Concernant les dimensions de ce groupe, tous les utilisateurs ont mentionné avoir régulièrement utilisé l'année et s'être référé au journal de publication. Ils ont aussi remarqué que le titre était souvent une source d'information précieuse pour comprendre un intérêt.

L'un des participants, P6, remarque qu'un ajout pourrait être fait pour étendre les métadonnées. En effet, plusieurs participants (P1, P3, P4 et P6) ont regretté l'absence d'une illustration ou d'un résumé des exemples dans l'archive, avant de les ouvrir. Le participant P6 remarqué que beaucoup de journaux en ligne intègrent à leur page web une étiquette qui contient un aperçu de l'article destiné aux réseaux sociaux. Cet aperçu est spécifiquement prévu pour communiquer un résumé et un visuel en dehors du site du journal d'origine.

Contexte Trois des utilisateurs (P1, P3 et P4) , disent utiliser régulièrement le groupe "contexte". Les dimensions "but" et "sujet" sont régulièrement mentionnées comme utiles, par tous les participants. Bien que le sujet et le but ne soient pas des dimensions exclusives aux RCD les participants ont particulièrement utilisé ces deux informations pour rechercher des exemples qui correspondent à leur propre projet. La dimension structure, quant-à-elle, a été peu mentionnée, et certains participants (P4 et P6) ont mentionné que certaines de ses

options étaient difficiles à interpréter. L'un des principaux intérêts de cette dimension semble être d'identifier les animations. Un participant, P4, remarque aussi que la structure est une des dimensions la plus négativement affectée par la manque d'hétérogénéité technique de l'archive. Effectivement, sur 871 entrées, 699 appartiennent à la même catégorie, l'infographie déroulante.

Contenus Dans le contenu, nous remarquons que la dimension “visualisation” a été mentionnée comme une des dimensions les plus utiles par 5 des participants, et comme la plus utile d'entre toutes par un participant, P5. Les utilisateurs vont aussi bien chercher des visualisations spécifiques qu'évaluer un exemple en fonction des visualisations listées. Son intérêt est néanmoins restreint par le nombre d'options. Il s'agit de la dimension qui possède le plus grand nombre de valeurs possibles dans notre taxonomie (31 valeurs en comptant “Autres”). Deux participants (P1 et P5) ont remarqué que ce nombre est excessif et cause une confusion. Un autre, P5, remarque que la position de la dimension “visualisation” n'est pas assez visible par rapport à son importance, et est noyée parmi les autres groupes et dimensions.

Trois utilisateurs (P1, P3 et P5) ont montré de l'intérêt pour la dimension médias, en particulier pour repérer des exemples contenant de l'audio et des vidéos. Une remarque faite par un utilisateur (P1) est le manque d'une valeur “graphique”, qui indiquerait la présence de graphiques statistiques et data visualisations. Bien que la grande majorité de nos exemples répondrait positivement à cette valeur, l'utilisateur a remarqué que cela permettrait aux utilisateurs d'éviter la confusion avec la valeur “image” de la dimension.

Le tutoriel est une des dimensions qui ont été le moins bien reçues. Trois participants (P1, P4 et P6) ont indiqué qu'ils n'en avaient pas l'utilité, et aucun n'a remarqué d'intérêt particulier à cette dimension. Il s'agit, pour citer le participant P6, d'un “petit filtre ultra-précis”.

Contrôles Concernant le groupe intitulé “contrôles”, une remarque récurrente est la contribution du groupe à l'excès d'options de la taxonomie. Il comporte cinq dimensions, dont plusieurs ont reçu des retours mitigés.

Le degré liberté de l'utilisateur a suscité de l'intérêt chez un des participants, P4, qui l'a mentionné comme une des dimensions qu'il utilisait régulièrement, mais d'autres ont remarqué que le vocabulaire n'était pas forcément familier, et que sans infobulles explicatives la dimension aurait été inutilisable.

Les interactions sont une des dimensions qui divisent le plus : trois participants de l'expérience (P3, P4 et P5) ont annoncé beaucoup s'en servir, tandis que d'autres considèrent l'option comme superflue (P1 et P6). Une des critiques récurrentes était l'excès d'options dif-

férentes, qui entrave son utilisation. L'un des participants a suggéré qu'une simplification de la dimension, peut être jusqu'à une binarité "Oui/Non", pourrait la rendre plus pertinente.

La dimension "séquence" a été considérée comme "trop précise" par plusieurs participants (P2, P5 et P6). En particulier en considérant que la problématique dominante de la taxonomie est la surcharge d'options différentes. Le participant P4 remarque que cette dimension est intéressante mais nécessite une explication, en particulier le "verre à martini".

Les transitions ont aussi été considérées comme superflues. Les utilisateurs (P1, P2, P3, P4, P6) les considèrent soit comme trop précises soit comme appartenant à une autre étape de la conception. L'un des participants (P3) remarque qu'il ne se voit pas chercher un format de transition particulier, mais peut en remarquer un dans un exemple et décider de le reproduire. Un autre (P6) indique que les transitions sont un embellissement final, qu'il ne prend pas en compte dans la phase de conception initiale.

Les indicateurs de progression ont reçu très peu d'attention de la part des participants. Aucun d'entre eux ne l'a mentionné positivement ou négativement lorsqu'ils ont été interrogés sur les dimensions. Considérant que les contrôles sont collectivement l'objet de critiques sur le nombre d'options, il semble que cette dimension ait été noyée dans la masse et ignorée. Nous ne pouvons pas conclure que cette dimension est hors de propos, mais elle n'est visiblement pas vitale à la pertinence de la taxonomie.

Schémas narratifs Le dernier groupe, schémas narratifs, est celui qui a été le plus négativement reçu. Aucun participant n'a mentionné y avoir trouvé un intérêt particulier, mais quatre d'entre eux (P1, P2, P5, P6) ont remarqué qu'ils ne savaient pas ce que pourrait apporter ce groupe, ou bien ont considéré qu'il s'agit d'informations très spécifiques et ayant peu de chances d'être utilisées. La plupart des participants n'ont pas commenté les dimensions de ce groupe, mais l'un d'entre eux, P3, nous offre une piste pour expliquer le peu d'intérêt pour les schémas narratifs : en commentant certains schémas, il remarque que certains concepts tels que "les humains derrière les points" sont des outils qu'il utilise implicitement mais ne correspondent pas à une démarche de conception intentionnelle. Il est donc possible que les utilisateurs utilisent des schémas narratifs, mais en utilisant une démarche ou un vocabulaire différents, qui rendent le groupe peu utilisable afin de filtrer notre archive.

Observations sur la construction de la taxonomie

Une observation qui a été faite est le fait que certaines dimensions sont pertinentes pour décrire des RCD, et correspondent à une réalité de terrain dans la conception des RCD, mais ne sont pas pertinentes pour la recherche d'exemple. En effet, un participant, P5, indique que

certaines décisions ne sont pas prises au moment de la conception du RCD et du brainstorming initial. Par conséquent, elles ne sont pas présentes à l'esprit des utilisateurs quand ces derniers sont à la recherche d'inspiration.

Autre remarque, la pertinence perçue des éléments de taxonomie semble corrélée à leur positionnement dans l'interface de CurratedDDS. Les groupes les plus à gauche ont été les plus utilisés et positivement reçus, les groupes les plus à droite ont été plus négativement reçus. En considérant que l'une des grandes failles de notre taxonomie est le grand nombre d'options, il est possible que les dernières options lues sont celles qui ont été le plus affectées par ce problème. Alternativement, une autre possibilité réside dans l'ordre des groupes : Nous avons ordonné les groupes de celui qui nous semblait le plus général au plus spécifique.

Correspondance du vocabulaire avec la réalité

L'une de nos interrogations en rédigeant la taxonomie était de savoir à quel point ce vocabulaire correspond à la réalité du terrain vécue par les participants. La réponse est positive sur plusieurs dimensions, mais semble négative sur plusieurs d'autres. Les utilisateurs ont utilisé le vocabulaire employé par certaines dimensions sans avoir à s'arrêter pour interroger sur sa signification, et ont réutilisé ces termes pendant l'entretien car il s'agit d'un vocabulaire familier. Il s'agit souvent des mêmes dimensions qui ont été le plus utilisées : “sujet”, “visualisation”, “but”. Une partie des termes sont aussi moins familiers mais compréhensibles. P3 indique par exemple avoir immédiatement identifié le schéma “les humains derrière les points” comme quelque chose qu'il aime intégrer à ses projets, mais sur lequel il n'avait jamais mis de nom. De la même manière, la dimension “degré de liberté de l'utilisateur” soulève un problème. Trois participants (P1, P2, P5) remarquent que le “verre à martini” n'est pas une formule qui leur est familière. Après une explication, ils ont été capable de faire correspondre ces termes à une méthode qu'ils utilisent, mais qu'ils ne désignent pas ainsi. Certains pointent aussi que certains termes, tels que les noms de certains schémas narratifs, correspondent à des idées qui leurs sont familières mais qu'ils ne sont pas habitués à nommer verbalement dans le cadre de leur travail.

Un participant (P5) décrit le vocabulaire de la taxonomie comme académique, et remarque que beaucoup de concepteurs de RCD sont à sa connaissance des personnes formées comme journalistes classiques ou comme développeurs informatiques, et que seule une poignée d'entre eux ont spécifiquement suivi un cursus universitaire dans l'optique de concevoir ce type de projets. Il émet l'hypothèse que ce phénomène cause une différence entre le vocabulaire académique et le vocabulaire utilisé par les praticiens. Un autre participant, P6, avance une idée similaire. Il indique qu'il s'identifie plus comme un praticien qu'un théoricien, et que cela

peut expliquer pourquoi certains termes ne lui sont pas familiers. Un autre participant (P4), décrit le vocabulaire comme structurant, et indique que l'utilisation prolongée d'un outil tel que CurratedDDS pourrait le mener à changer son vocabulaire de travail, en particulier si ses collègues utilisent les mêmes outils.

Évaluation des hypothèses H1 et H2

L'hypothèse H1 : La taxonomie proposée caractérise adéquatement les RCD pour leur usage par des concepteurs de RCD est partiellement validée. Nos utilisateurs ont pu utiliser les dimensions de la taxonomie pour explorer des RCD et formuler leurs requêtes sur CurratedDDS. Néanmoins, ils ont accompli ce but à l'aide de parties de la taxonomie, et sans en utiliser certaines autres.

L'hypothèse H2 : La taxonomie utilise un vocabulaire qui correspond à la réalité de la pratique des concepteurs de RCD, n'est que partiellement validée. Certaines sections du vocabulaire de la taxonomie étaient familières aux utilisateurs, mais plusieurs leur étaient inconnues et ont dû leur être expliquées via des infobulles. Les termes qui n'ont pas été reconnus ont pu être associés à des pratiques réelles, mais cette façon de les décrire n'était pas familière aux participants.

4.2.2 Retours sur l'archive

Notre hypothèse H3 porte sur l'archive.

- H3 : L'archive proposée contient une sélection d'exemple pertinente pour les concepteurs de RCD comparativement aux exemples auxquels ils sont exposés hors de l'archive

Pertinence des contenus

Le retour général sur les exemples est positif. Interrogés sur l'intérêt et la qualité des exemples de l'archive, 5 participants ont eu un retour positif, et le sixième (P4) a constaté que l'archive est intéressante mais que les limitations, détaillées plus bas, réduisent son intérêt.

La qualité des exemples en eux même est sujette à un retour positif. Quatre participants (P1, P2, P3 et P5) pointent l'intérêt des contenus. Deux (P1 et P6) remarquent que l'archive contient des sources familières auxquelles ils font confiance. Un autre (P2) compare l'archive à sa méthode habituelle, une recherche sur google, et remarque que l'archive offre une garantie de pertinence des exemples comparativement au moteur de recherche. Une autre remarque est

le fait que l'un des utilisateurs (P1) est que l'archive donne une bonne idée des avancements dans le domaine de la data visualisation pour le journalisme.

Au-delà de la qualité des exemples individuels, certaines caractéristiques de l'archive globale sont intéressantes à évaluer. L'un des points d'intérêt est la variété d'exemples de l'archive. Deux participants (P1 et P5) ont remarqué directement que la diversité d'articles et de journaux était un apport de l'archive. De plus, plusieurs participants ont remarqué avoir fait des découvertes inattendues dans l'archive. En particulier, l'un des participants (P1) remarque avoir découvert de nouvelles publications dans des médias qui font pourtant partie de sa consommation habituelle. Nous constatons donc que notre archive permet une bonne sérendipité qui permet aux utilisateurs de faire des découvertes intéressantes.

Limitations de l'archive

Néanmoins, plusieurs limitations de l'archive émergent et limitent sa pertinence pour les concepteurs de RCD.

Diversité linguistique et géographique Notre archive a été composée à partir d'une source étasunienne, et par conséquent les exemples qu'elle contient sont tous issus de la presse anglophone des États-Unis. Cinq des six participants (P1, P2, P4, P5 et P6) de l'expérience ont remarqué que cette tendance nuit à la qualité de l'archive.

La pertinence d'un exemple dépend du contexte culturel et linguistique dans lequel il s'inscrit. Différents pays, ou différents groupes linguistiques, ont des pratiques journalistiques distinctes, ou des publics aux attentes différentes. Chaque exemple s'inscrit donc dans un contexte spécifique, et son utilisation comme inspiration va être impactée par la différence entre ce contexte et celui de l'utilisateur. Nos participants s'adressent habituellement à un public québécois, ou canadien francophone, différent du public auquel s'adressent les exemples de l'archive, étatsuniens anglophones. Bien qu'ils puissent trouver une inspiration technique dans un article anglophone, ces derniers ne représentent qu'une partie du paysage médiatique, et ne sont pas forcément les plus adaptés aux besoins des utilisateurs.

Le participant P1 remarque qu'un article dans une langue que l'utilisateur ne maîtrise pas, comme l'allemand ou l'espagnol reste une source d'inspiration potentielle en termes d'éléments visuels et d'interaction. Il peut même s'agir d'une meilleure source d'inspiration puisque ces exemples sortent plus des habitudes de consommation médiatiques de l'utilisateur. Plusieurs extensions possibles de l'archive ont été proposées par des participants (tous sauf P2) : contenus francophones, québécois ou français, contenus hispanophones, contenus anglophones non-américains.

Articles récents La troisième limitation de l’archive qui a été constatée est l’absence d’une alimentation continue en exemples. L’archive contient des RCD datant de 2020 à 2022, rassemblés et annotés au début de l’année 2023. Notre méthode de collecte d’exemples repose sur des listes d’articles publiées en fin d’année, et celles de 2023 n’étaient pas encore parues au moment de l’expérimentation. En décembre 2023, les exemples dataient donc d’au moins douze mois. Nous avons supposé que l’intérêt technique et méthodologique des exemples était, à défaut d’être intemporel, persistant dans le temps. Un exemple de RCD pourrait rester viable plusieurs années, tant qu’il n’est pas rendu obsolète par un changement technologique considérable ou un changement significatif des tendances de conception visuelle. En pratique, nous avons surestimé cette persistance, car les utilisateurs favorisent les exemples portant sur l’actualité récente. Ils cherchent de l’inspiration pour leur projet avec un sujet d’actualité à l’esprit, et cherchent à savoir comment d’autres sources traitent de ce sujet. Les exemples portant sur des sortis de l’actualité perdent donc de leur pertinence.

Le manque de RCD récents limite donc la qualité de nos entrées en tant qu’exemples, et par extension la pertinence de l’archive.

Distribution des valeurs de dimensions Une autre faille, remarquée par l’un des utilisateurs (P4), est la rareté de certaines valeurs et combinaisons de valeurs de la taxonomie dans l’archive. Toutes les valeurs individuelles des dimensions de la taxonomie sont représentées, mais certaines sont plus rares que d’autres, et cela peut devenir problématique. Par exemple, le participant souhaite rechercher un exemple de diaporama interactif (une valeur rare de la dimension “Degré de liberté de l’utilisateur”). Seuls dix de nos exemples correspondent à cette requête, et en ajoutant d’autres filtres le nombre risque de tomber à 0. Certaines des valeurs les plus rares sont donc en sous-nombre, et donnent l’impression d’une archive “sous-alimentée” comparativement à sa taille. Ce manque d’hétérogénéité limite la pertinence de l’archive.

Évaluation de l’hypothèse H3

Nous constatons que l’archive propose un contenu pertinent, qui donne accès à des exemples intéressants et variés, qui permettent de faire des découvertes. Néanmoins, il existe plusieurs limitations qui remettent en question cette pertinence. En particulier, la diversité, bien qu’ayant été décrite comme satisfaisante par certains participants, a aussi été sujette à critique. Il semble que la diversité interne à l’espace que nous couvrons (médias étasuniens anglophones entre 2020 et 2023) est satisfaisante, mais que cet espace est trop restreint.

4.2.3 Retours sur CurratedDDS

Nos hypothèses H4 et H5 portent sur l'influence de CurratedDDS sur les utilisateurs et leurs travaux, en particulier via l'accès aux exemples.

- H4 : L'outil CurratedDDS améliore l'accès à des exemples pour des concepteurs de RCD, comparativement à leurs pratiques habituelles sans cet outil.
- H5 : L'utilisation de CurratedDDS facilite le travail d'idéation des concepteurs de RCD.

Impact des exemples

Certains participants (P1, P2, P4) utilisent les exemples pour explorer les pratiques des autres médias en comparant les exemples à leurs propres travaux, et en examinant comment les thèmes, données et représentations graphiques ont été traités. Le participant P5 indique qu'un exemple permet aussi d'identifier un projet déjà fait par une autre organisation afin de ne pas copier leurs travaux.

Les exemples sont aussi utilisés comme outils de communication pour illustrer des idées similaires et échanger des idées de conception, ainsi que comme outils de brainstorming en équipe. Un participant, P5, ajoute qu'un exemple peut être utilisé par des développeurs discuter de la faisabilité technique d'une idée avec les journalistes.

Un autre aspect est l'intérêt des exemples pour structurer des projets. Au-delà d'une caractéristique individuelle de RCD à reproduire, plusieurs participants (P3) ont indiqué trouver dans des exemples de nouvelles idées pour construire leur projet à plus grande échelle. D'autres, comme P5 et P6, utilisent les exemples pour une accumulation de petites inspirations qui vont transformer leurs projets.

Une remarque qui a été faite par P6 est la différence d'impact des exemples au cours du temps. En début de projet, quand les concepteurs sont encore en train d'explorer des possibilités, un exemple peut radicalement transformer le projet. Plus tard, quand les grandes lignes sont déjà établies, les exemples permettent de repérer des éléments de design plus restreints qui vont mener à des modifications moins significatives.

Deux participants (P2 et P5) ont indiqué que les exemples leur permettent d'évaluer à quel point une idée sera compréhensible pour le public. Ils ont pointé que c'était l'une des plus grandes problématiques qu'ils cherchent à adresser, car une idée qui semble innovante peut s'avérer être au-delà de la compréhension du public une fois implémentée. Une remarque de P6 rejoint cette idée : Il pointe que l'alternative à un exemple serait de créer un prototype, au prix de temps et d'efforts de travail bien plus élevés. Les exemples permettent donc d'écarter

des idées à faible valeur.

Un participant (P4) pointe une autre utilisation qu'il fait des exemples. En effet, il les utilise comme outil d'apprentissage. Pour cela, il reproduit des exemples afin de développer les compétences qui auraient été nécessaires pour le produire. Ceci correspond à l'idée avancée par un autre participant, qui commente que plus un concepteur de RCD consulte d'exemples, plus ses compétences professionnelles se développent.

Concernant la créativité spécifiquement, les participants font des remarques sur l'apport des exemples en termes de diversification de leurs travaux. Ils considèrent que des exemples de bonne qualité permettent de trouver des idées intéressantes, de varier les représentations de données, et de penser à des concepts moins conventionnels.

Impact de CurratedDDS

Ces constats nous informent sur les exemples de façon générale, mais la question de l'intérêt de CurratedDDS reste non résolue.

Impact de CurratedDDS sur l'accès aux exemples Concernant l'impact de CurratedDDS sur l'accès des concepteurs de RCD à des exemples, nous avons collecté plusieurs retours. Quatre utilisateurs, (P1, P3, P4 et P5) pointent que CurratedDDS les pousse à rechercher des exemples différemment. L'un d'entre eux (P1) remarque qu'utiliser CurratedDDS le fait sortir de ses habitudes de consommation médiatique, et un autre (P4) remarque qu'avoir un outil dédié à la recherche d'exemples l'a poussé à prendre le temps de chercher activement des exemples. Trois autres (P2, P5 et P6) ont remarqué que CurratedDDS leur a permis de trouver des exemples qu'ils n'auraient pas trouvés dans sans cet outil. Deux participants (P2 et P3) remarquent aussi CurratedDDS sauve du temps pour la recherche d'exemples.

Quatre participants (P1, P2, P3 et P6), remarquent aussi que CurratedDDS leur permet spécifiquement d'accéder à des exemples inspirants ou qui leur permettent de découvrir de nouvelles façons de présenter l'information. CurratedDDS contribue donc bien à la créativité via les exemples.

Un autre retour est la facilité d'accès aux exemples. Cinq participants sur six (P2, P3, P4, P5 et P6) ont fait un retour indiquant que CurratedDDS a facilité leur recherche d'exemples. La tendance dominante de ces retours porte en réalité sur l'archive. La concentration d'exemples dans un même environnement permet d'accéder à ces derniers plus aisément.

Une autre remarque a été faite par un participant, P6. Il dispose déjà d'une archive personnelle, constituée au fil du temps, mais que cette dernière est moins structurée et navigable

que CurratedDDS.

Néanmoins, l'un des participants (P2) a remarqué qu'il n'avait pas pu trouver d'exemples pertinents pour les comparer avec un projet spécifique qu'il conduisait au moment de l'expérimentation. Il apparaît que pour des recherches très précises, CurratedDDS ne permet pas de spécifier les besoins de façon adéquate.

Impact de CurratedDDS sur le travail des utilisateurs Nous avons aussi des retours généraux sur CurratedDDS, qui correspondent aux retours portant sur les exemples.

Deux participants (P2 et P3) indiquent que CurratedDDS leur permet de sauver du temps dans leur travail en facilitant la recherche d'exemples. L'outil permet de trouver des exemples de RCD plus facilement et plus rapidement que les méthodes habituelles des participants.

Trois participants (P1, P2, P3) ont aussi confirmé que le bénéfice de recherche d'exemples en termes de diversification des exemples se traduit en termes de bénéfices de la qualité des exemples. L'outil permet donc de faire des découvertes d'exemples pertinents qui contribuent à l'idéation pour les projets de nos utilisateurs.

Nous constatons néanmoins que plusieurs limitations apparaissent, qui réduisent l'intérêt de CurratedDDS. L'un de ces points est la limite de l'interface graphique. La conception de l'interface de l'outil contient plusieurs failles qui le rendent difficile à utiliser à son plein potentiel. La remarque la plus fréquente est le fait que le nombre d'options et la diversité des outils de sélection débordent l'utilisateur. Une majorité d'utilisateurs (P2, P3, P4, P4, P5) remarquent sous une forme ou une autre que le nombre d'options de filtrage est excessif, et que cela cause une perte de valeur de l'outil. P4 indique qu'il y a une courbe d'apprentissage avant de pouvoir se repérer efficacement. Un autre exemple de ce problème est le menu déroulant des visualisations, qui utilise un format différent pour intégrer une barre de défilement et parcourir les nombreuses options de cette dimension. La différence visuelle ne rend pas l'outil inutilisable, mais cause un inconfort qui complique l'utilisation de CurratedDDS.

Les problématiques liées à l'archive et à la taxonomie, détaillées plus haut, sont aussi une cause de perte d'efficacité et d'intérêt. Le manque de diversité linguistique et géographique, la date des articles, la distribution des valeurs de dimensions, la taille excessive de la taxonomie et les problèmes de vocabulaire sont tous des problèmes qui, sans le rendre inutilisable, entravent le bon fonctionnement de CurratedDDS.

Une autre limitation de l'influence de CurratedDDS est le cycle de travail des utilisateurs. Plusieurs participants (P1, P2, P3 et P4) ont remarqué que leur travaux suivent un cycle de développement découpé en phases implicites. Certaines de ces phases sont plus adaptées

à l'utilisation de CurratedDDS car elles nécessitent des exemples et donnent le temps aux utilisateurs de les rechercher activement. D'autres phases sont moins adaptées à la recherche d'exemple car elles se focalisent sur la réalisation d'un projet déjà conceptualisé, qui n'a pas besoin d'une nouvelle source de créativité.

Évaluation des hypothèses H4 et H5

Les retours portant sur CurratedDDS, son rôle dans l'accès aux exemples et ses effets sur la pratique professionnelle des concepteurs de RCD sont mitigés. Bien que CurratedDDS augmente l'accès aux exemples et que l'effet positif de ces derniers soit confirmé, les limitations de l'archive, de la taxonomie et de l'interface de l'outil l'empêchent de réaliser ce potentiel. Les retours positifs laissent penser qu'aucune de ces limitations n'est rédhibitoire, mais elles réduisent la pertinence de l'outil.

L'hypothèse H4, "L'outil CurratedDDS améliore l'accès à des exemples pour des concepteurs de RCD, comparativement à leurs pratiques habituelles sans cet outil", est validée. Les retours montrent que les utilisateurs ont bénéficié d'un accès amélioré à des exemples en utilisant CurratedDDS.

L'hypothèse H5, "L'utilisation de CurratedDDS facilite le travail d'idéation des concepteurs de RCD", est partiellement validée. Les retours des participants indiquent un effet positif de CurratedDDS sur le travail des concepteurs de RCD, mais les limitations de la taxonomie, de l'archive et de l'interface de CurratedDDS réduisent l'impact positif de CurratedDDS sur le travail d'idéation des utilisateurs.

4.2.4 Travaux futurs

Les résultats de cette expérimentation nous ont permis d'identifier des idées de travaux futurs qui permettraient de donner une continuité à ce projet.

Améliorations de la taxonomie

Concernant la taxonomie, plusieurs idées d'améliorations ont émergé durant les entretiens. Nous pouvons proposer des altérations à la taxonomie afin d'améliorer CurratedDDS. Plusieurs demandes d'ajout ont été faites par les participants. L'une des recommandations était d'ajouter une colonne de métadonnées contenant le méta-tag, c'est-à-dire le résumé et l'image destinés au partage sur les réseaux sociaux. Ceci permettrait de donner une identité visuelle aux exemples dans le tableau de présentation, et de donner des informations sur le contenu de l'article. Ce résumé permettrait aussi d'utiliser la barre de recherche du tableau pour

des mots clés qui ne sont pas présents dans le titre. Il s’agit cependant d’un ajout sujet à caution, car cette entrée sera très volumineuse dans tout affichage des exemples, comme CurratedDDS, et ne peut pas être utilisée comme filtre par catégories. Une autre porte sur l’information concernant les verrous d’accès payants, afin d’identifier les exemples payants et permettre de filtrer l’archive selon cette caractéristique. Les remarques sur la diversité culturelle et linguistique de l’archive ont été accompagnées d’une proposition d’identifier dans la taxonomie la langue et le pays d’origine d’un RCD. A l’heure actuelle, toutes les entrées de l’archive sont des exemples anglophones publiés aux États Unis, mais dans un futur ou l’archive serait étendue, cela permettrait à un utilisateur d’explorer des exemples selon leur pays d’origine. L’un des participants (P1) propose aussi d’ajouter la valeur “graphiques” à la dimension “média”. Bien qu’il s’agisse d’une redondance avec la dimension “visualisations”. Bien que cela représente une redondance avec la dimension “visualisations”, le participant estime que ce changement ajoutera de la clarté aux valeurs de “média” en distinguant la valeur “image” comme une illustration, par opposition aux graphiques.

Plusieurs éléments de taxonomie ont été remis en question lors de l’expérimentation. Des coupures, transformations et réorganisations peuvent donc être proposées. Deux dimensions, “tutoriel” et “analyse”, ont été perçues par les participants comme superflues et comme des contributeurs au surplus d’information qui limite l’intérêt de la taxonomie. Ces deux dimensions n’ont pas reçu de retours positifs de la part des participants. Nous remarquons aussi que ces deux catégories distinguent des caractéristiques qui sont rares dans notre archive : “analyse” a 39 valeurs “oui” contre 741 “non”, et 700 des RCD de notre archive n’ont pas de tutoriel, ce qui laisse 80 exemples pertinents avec des tutoriels. Il s’agit donc de dimensions qui ont un apport très limité. Nous pouvons retirer ces deux dimensions afin de réduire la taille de notre taxonomie.

Nous avons aussi observé que la dimension “indicateurs de progression” a été laissée de côté par les participants. Elle est aussi dominée par la valeur “sans indicateurs”, avec 734 occurrences, contre 46 occurrences réparties entre les 6 autres valeurs possibles. La dimension met donc en valeur 6% des entrées de l’archive en mettant en avant une information qui ne semble pas apporter de valeur ajoutée. Dans un contexte où le nombre trop grand de dimensions dans la taxonomie est une source de problèmes, retirer “indicateurs de progression” de la taxonomie nous permet de simplifier avec une perte réduite d’information.

Trois participants (P1, P2 et P3), avancent l’idée de placer une partie groupes de la taxonomie dans une recherche avancée, qui permettrait de séparer la taxonomie entre les dimensions les plus utilisées et les plus reconnues, et celles qui s’adressent à des cas plus rares ou à des requêtes plus spécifiques. Cette séparation permettrait de réduire la surcharge d’information

en permettant à un utilisateur de se concentrer sur un groupe ou l'autre selon ses besoins. Les groupes "contrôle" et "schémas narratifs" sont deux groupes de dimensions qui ont été considérés comme "trop précis" par les utilisateurs. Étant donné que ces éléments de taxonomie ont reçu moins d'intérêt de la part des participants, et que l'une des limitations de la taxonomie est le nombre excessif de dimensions, nous avons envisagé de retirer tout ou partie de ces groupes. Certaines utilisations restant possibles, utiliser ces deux groupes pour les intégrer à une recherche avancée dans un éventuel outil permettrait de les préserver tout en dégageant les autres éléments de taxonomie. Nous avons donc regroupé les métadonnées, le contexte, les contenus et les données dans une grande catégorie "recherche simple", et regroupé les contrôles et les schémas narratifs comme "recherche avancée".

Enfin, les interactions et les visualisations sont deux dimensions qui ont été appréciées par les participants mais qui proposent trop d'options. Des participants ont avancé que simplifier les valeurs de ces dimensions permettrait de les rendre plus intéressantes à utiliser. L'un d'entre eux a proposé de réduire "interactions" à une option binaire oui/non. Nous proposons donc de revoir ces dimensions, et de fusionner certaines valeurs ou de regrouper les valeurs les plus rares sous des catégories "autres". Concernant les interactions, 548 exemples sur 780 sont marqués "sans interactions", ce qui appuie la proposition de réduire à une dimension binaire. Cependant, deux participants (P4 et P6) ont mentionné rechercher des interactions spécifiques. Réduire à une option binaire cause donc une perte d'information. Trois des valeurs de la dimension (comparaison, connection et encodage) ont un nombre d'occurrences considérablement plus réduit, et les regrouper sous l'étiquette "autres" permettrait de former un groupe de taille équilibrée par rapport aux autres (34 occurrences). Concernant les visualisations, la présence de 30 valeurs est une limitation, car les utilisateurs doivent fournir un effort accru pour se repérer parmi les valeurs et sélectionner celle qu'ils recherchent. Couper des catégories causerait une perte significative d'information. Une solution envisageable serait de grouper les visualisations sur deux niveaux. Ceci permettrait de décrire un RCD selon la présence d'une carte, ou de spécifier quelque chose de plus précis, comme une carte choroplèthe.

Le tableau 4.1 récapitule les modifications apportées à la taxonomie. Il indique si les anciennes dimensions ont été conservées ou rejetées, et ajoute les nouvelles dimensions et les nouveaux regroupements proposés.

TABLEAU 4.1 Récapitulatif de la nouvelle taxonomie

Niveau	Groupe	Dimension	Valeurs	Status
Recherche de base	Métadonnées	Titre	Valeur unique	Conservé
		Date	Valeur unique	Conservé
		Source	Valeur unique	Conservé
		Résumé	Valeur unique	Ajouté
		Verrou d'accès payant	Oui/non	Ajouté
	Contexte	But	Informar, persuader, divertir, reconforter, explique	Conservé
		Sujet	Arts, culture, divertissement et médias ; conflit, guerre et paix ; criminalité, droit et justice ; désastres et accidents ; économie et finances ; éducation ; emploi ; environnement ; intérêt humain ; météo ; politique ; religion et croyance ; santé ; science et technologie ; société ; sport ; vie quotidienne et loisirs	Conservé
		Langue et pays	Anglais(USA), (futurs valeurs à ajouter avec l'expansion de l'archive)	Ajouté
		Structure	Animation, compilation, diaporama, image statique, infographie déroulante, multimédia, storymap, autre	Conservé
	Données	Origine	Valeur unique	Ajouté
		Forme	Qualitative, quantitative	Conservé
		Type	Numériques, textes, spatiales, réseau, temporelles	Conservé
		Analyse	Oui, non	Retiré
	Contenus	Visualisation	Autre, carte à flux, carte à points, carte à symboles proportionnels, carte choroplèthe, carte de chaleur, carte de navigation, cartogramme, diagramme à barres, diagramme à bulles, diagramme à formes proportionnelles, diagramme circulaire, diagramme d'instance, diagramme de coordonnées parallèles, diagramme de sankey, graphique de flux, graphique en aires, graphique radar, histogramme, infographie, ligne, modèle 3D, nuage de points, organigramme, pictogramme, range chart, réseau, table, tracé en mosaïque	Modifié
		Médias	Audio, graphique, image, texte, vidéo	Modifié
		Tutoriels	Sans, explicite, implicite	Retiré
Recherche avancée	Contrôle	Degré de liberté de l'utilisateur	Axé sur l'auteur, diaporama interactif, verre à martini histoire à forage, axé sur le lecteur	Conservé
		Séquence	Sans séquençage, navigation linéaire, navigation non-linéaire	Conservé
		Transitions	Atténuation, panoramique, glissement, interpolation	Conservé
		Indicateur de progression	Sans indicateur, barre de progrès, boutons de section, ligne temporelle, carte géographique, fil d'Arianne, liste de pointage	Retiré
		Interactions	Sans, autres, abstraction/élaboration, exploration, filtre, reconfiguration, sélection	Modifié
	Schémas Narratifs	Schémas Argumentatifs	Comparaison, concrétisation, répétition	Conservé
		Schémas de Cadrage	Rupture de conventions, défamiliarisation, familiarisation, métaphore physique	Conservé
		Schémas d'Engagement de d'Empathie	Sans, accélération/décélération, rupture du 4ème mur, devinette, exploration, les humains derrière les points, question rhétorique, révélation graduelle	Conservé

Amélioration de l'archive

En considérant les limitations de l'archive qui ont été relevées lors de l'expérimentation et les propositions qui ont été avancées par les participants, nous pouvons envisager de futures modifications destinées à améliorer l'archive.

Pour la problématique de la diversité culturelle et linguistique, nous proposons d'ajouter les sources proposées par les participants afin d'élargir notre champ de sélection. Cinq sources ont été mentionnées directement par les participants : South China Morning Post (Hong-Kong, anglophone), Le Monde (France, francophone), Le Figaro (France, francophone), Radio Canada (Canada, francophone et anglophone), La Nacion, (Argentine, hispanophone), La BBC (Royaume Uni, anglophone). Ces sources nous permettent un élargissement à des aires culturelles plus variées, ainsi qu'à trois langues.

Une autre possibilité à aussi été identifiée : les concours de journalisme de données. Ces sources permettraient d'ajouter des contenus de qualité à l'archive, et d'augmenter sa diversité, puisque les lauréats peuvent provenir de publications autres que nos sources habituelles. Le Concours Canadien de Journalisme² dispose d'une catégorie "Présentation/Conception graphique". Les Sigma Awards³ sont un concours international de journalisme de données. Des exemples de RCD peuvent être identifiés dans leur shortlist. Les Digital Publishing Awards⁴ sont un concours plus large. La catégorie "Journalist of the Year" permet d'identifier des lauréats dont les travaux sont à leur place dans l'archive.

Amélioration de CurratedDDS

Modifier la taxonomie et l'archive permettrait d'améliorer CurratedDDS, mais une de nos limitations reste la lisibilité de l'interface graphique. Pour faciliter de futurs usages de CurratedDDS, nous cherchons à simplifier l'utilisation du filtrage, et à rendre l'affichage des résultats plus lisible.

La première série de modifications de l'interface de l'outil porte sur le filtrage. Les retours des participants indiquent qu'il est nécessaire de clarifier cette interface pour la rendre plus aisée à utiliser. Le changement le plus conséquent sera la division du panneau déroulant des filtres en deux parties. En utilisant la nouvelle division de la taxonomie entre "recherche de base" et "recherche avancées", nous proposons de mettre en place un panneau déroulant qui permet d'afficher les filtres de la "recherche de base", ainsi qu'un bouton "recherche avancée" pour ouvrir une page de recherche qui permette d'accéder à tous les filtres. Nous proposons

2. <https://nna-ccj.ca/fr/>

3. <https://www.sigmaawards.org/>

4. <https://www.ukaop.org/digital-publishing-awards/>

aussi d’uniformiser les commandes de filtrage. L’un des participants, P1, a remarqué que le mélange entre formats cause de la confusion. Initialement, les différents formats d’entrée pour les commandes avaient été choisis pour minimiser le volume des dimensions avec le plus d’options. Réduire le volume des filtres est moins intéressant que rendre ces derniers plus lisibles. Tous les menus déroulants sont remplacés par des cases à cocher ou des boutons radio, afin de rendre les formats plus cohérents. La réduction du nombre de formats permet de rendre l’interface plus cohérente. Enfin, la recherche réactive, qui se lance automatiquement quand l’utilisateur change ses filtres, est retirée et remplacée par un bouton “Rechercher”. Certains participants ont remarqué que le lancement automatique les empêchait de formuler immédiatement une requête selon plusieurs dimensions. Ce phénomène cause un gêne qui dégrade l’utilisabilité de CurratedDDS.

Deux propositions ont aussi été faites par les participants pour améliorer le tableau qui affiche les résultats. La première de ces propositions est la sélection des colonnes affichées. Une problématique qui est apparue lors de l’expérience est la taille du tableau d’affichage. Étant donné que nous exposons toutes les informations dont nous disposons sur chaque entrée de l’archive, le tableau présente 22 colonnes, ce qui impose un défilement horizontal pour les consulter. En permettant à l’utilisateur de choisir quelles colonnes afficher ou ne pas afficher, nous permettrons à ce dernier de sélectionner les éléments les plus pertinents, et de ne pas avoir à naviguer pour passer d’une dimension à l’autre. L’autre proposition est l’ajout d’un code couleur pour mettre en valeur des groupes et dimensions, pour mieux les distinguer dans le tableau, et pour connecter visuellement l’interface de filtrage aux colonnes du tableau. Étant donné que nous avons 23 colonnes (22 dans le tableau d’origine, quatre qui ont été ajoutées et 3 retirées), nous pourrions proposer 23 couleurs. Ce nombre risque néanmoins d’être trop grand et de porter à plus de confusion. Alternativement, nous proposons d’utiliser une couleur par groupe, ce qui permettrait de nous limiter à 6 couleurs.

Une autre idée qui a été soulevée par un participant (P6) est l’inclusion d’un agent conversationnel intégré à l’outil pour assister l’utilisateur dans la formulation de requêtes. L’agent serait utilisé pour permettre à un utilisateur de décrire ses intentions en langage naturel, puis effectuer la conversion vers les filtres les plus appropriés. Cette idée implique un défi technique considérable, mais permettrait aussi une approche différente, et potentiellement plus fluide, des options de filtrage de CurratedDDS.

4.2.5 Limitations de l’évaluation

La mise en pratique de notre protocole expérimental a mis en lumière certaines de ses limitations, qui réduisent la portée de nos conclusions.

Verrous d'accès payant

Une des observations qu'on fait les participants est l'impact des verrous d'accès payant sur leur utilisation de CurratedDDS. Une supposition que nous avons faite en préparant l'archive est que nos utilisateurs, journalistes de métier, disposent d'un moyen d'accéder aux travaux d'autres journaux si ces derniers se trouvent derrière un verrou d'accès payant, afin de consulter des exemples.

Durant les entretiens il est apparu que cet accès prend la forme d'abonnements personnels à différentes sources, qui diffèrent donc d'un utilisateur à l'autre. Ainsi, nous avons constaté que certains participants de l'expérience disposent d'une large gamme d'abonnements et peuvent naviguer dans tous les exemples de l'archive, tandis que d'autres sont restreints à une portion seulement des contenus. L'un des utilisateurs (P4) remarque avoir initialement attendu une grande diversité, puis avoir constaté que cette diversité s'est limitée aux articles et journaux gratuits et à ceux du New York Times, seul journal de l'archive auquel il est abonné. Nous estimons que ce participant a eu accès à 36% de l'archive. Il a pu utiliser CurratedDDS pour trouver des exemples pertinents, mais avec une efficacité réduite. Les participants P1 et P2 indiquent aussi une perte, plus réduite, d'accès à l'archive en raison des verrous d'accès payants. Nos conclusions concernant l'intérêt de l'archive sont donc limitées aux personnes qui peuvent en premier lieu accéder au contenu de cette archive.

Deux des utilisateurs (P4 et P6) remarquent qu'il n'est sans doute pas possible de contourner complètement cette limitation puisque c'est une caractéristique du journalisme en ligne. Pour citer l'un d'eux, "l'information a un coût". Néanmoins, de futurs projets pourraient atténuer cette problématique en intégrant l'accès à des abonnements dans leurs critères de recrutement, ou en proposant une archive construite autour de l'accessibilité, par exemple en compilant des exemples gratuits, ou autour des abonnements de leurs participants.

Impact de la période d'expérimentation

Une autre limitation de notre protocole expérimental réside dans son organisation et les dates sur lesquelles l'expérimentation a été menée. Nous avons lancé la période d'utilisation au mois de décembre, et les entretiens se sont poursuivis jusqu'au début du mois de janvier. Cette période a été décrite par certains participants comme la clôture des projets des fêtes de fin d'année. Ce contexte a marqué leur utilisation de CurratedDDS. Certains ont utilisé l'outil pour des comparaisons avec des projets déjà en cours, d'autres pour leur veille professionnelle, et d'autres encore de manière préparatoire pour préparer leurs projets du début de l'année suivante. Les participants ont signalé que cette période n'est pas forcément

représentative de toute leur pratique. Certaines démarches, comme la phase de lancement de projet, n'étaient pas d'actualité au moment de l'évaluation. Par conséquent, nos résultats ne sont donc pas représentatifs de certaines pratiques professionnelles des concepteurs de RCD, qui sont associées à d'autres périodes de l'année.

Un futur projet pourrait activement cibler des périodes de temps spécifiques pour évaluer les effets de CurratedDDS en couvrant une plus large variété de situations vécues par les utilisateurs.

Taille de l'échantillon

Notre échantillon, limité en nombre, nous permet de mener étude de cas, mais pas d'arriver à des conclusions à grande échelle. Notre population se compose de 6 journalistes québécois francophones, et n'est donc pas représentative de la totalité des concepteurs de RCD dans le monde.

De futures recherches pourraient considérer un échantillonnage plus large, qui permettrait de mieux représenter la population des professionnels qui participent à la conception de RCD. Un tel échantillon pourrait permettre d'ajouter des relevés démographiques afin de prendre en compte les années d'expérience, ou la formation d'un participant.

Approche descendante

L'un des éléments de la démarche de ce projet était de proposer un outil qui rendrait accessible aux praticiens des éléments de taxonomie précédemment confinés aux milieux académiques. Pour cette raison, l'approche de création de la taxonomie est descendante, de la théorie vers la pratique. Une approche ascendante, centrée sur l'utilisateur aurait permis de répondre plus efficacement aux réalités de l'expérience des praticiens, ce qui aurait permis une taxonomie plus adaptée aux besoins des utilisateurs et de créer une interface plus ergonomique. Une façon dont cette approche aurait pu être mise en place aurait pu intégrer une première série d'interview formelles pour consulter les participants quant à leurs besoins. Suite à ces entretiens, nous aurions été capables de créer la taxonomie en fonction de ces besoins, puis d'annoter l'archive avec une taxonomie mieux conçue, puis de mettre en place l'outil. Enfin, nous aurions mené une deuxième série d'entretiens, similaire à celle qui a vraiment été mise en place afin de collecter des retours.

CHAPITRE 5 CONCLUSION

Les RCD sont un format de journalisme de données en ligne qui utilise une importante composante de visualisation de données et d'interaction pour proposer une communication visuelle de l'information. Faciliter le travail d'idéation des concepteurs de RCD permet de les assister dans leur pratique professionnelle et de favoriser le développement de nouveaux projets innovants. L'un des moyens de faciliter ce travail d'idéation est de faciliter l'accès à des exemples de travaux préexistants.

5.1 Synthèse des travaux

Dans l'optique d'améliorer l'accès des concepteurs de RCD aux exemples, nous avons proposé une taxonomie, utilisée dans un outil que nous avons également conçu, permettant l'exploration d'une archive que nous avons collecté et annoté.

Nos contributions sont : (1) la proposition d'une nouvelle taxonomie des RCD, qui permet de décrire les RCD avec une large variété d'options de description, afin d'offrir des possibilités de tri variées, (2) une archive d'exemples de RCD pertinents, issus de sources de qualité, et annotés selon la taxonomie proposée, et (3) un outil de recherche d'exemples de RCD qui permet à ses utilisateurs d'explorer l'archive pour y trouver des exemples, en triant son contenu à l'aide de la taxonomie.

La taxonomie nous permet de caractériser les RCD. Elle se compose de 6 grands groupes, les métadonnées, le contexte, les données, le contenu, le contrôle et les schémas narratifs, pour un total de 20 dimensions. Cette taxonomie a été construite en adaptant les taxonomies proposées par 11 articles de recherche préexistants, listés dans le chapitre 3 (sous-section 3.1.1). Nous avons utilisé cette taxonomie d'une part pour décrire les RCD de l'archive, et d'autre part comme outil de filtrage pour CurratedDDS.

L'archive se compose de 780 exemples de RCD issus des compilations annuelles des meilleurs projets de 6 publications en ligne, publiées entre 2020 et 2022. Cette archive a été annotée à l'aide de la taxonomie proposée, puis utilisée comme base de données afin d'alimenter CurratedDDS.

Pour valider la pertinence de cette méthode, une évaluation a été réalisée auprès de concepteurs de RCD du journal Le Devoir, partenaire de ce projet, rejoint par des concepteurs de RCD de Radio Canada, pour un total de 6 participants. Cette évaluation a pris la forme d'une période d'essai de l'outil par les participants, suivie de retours sous la forme d'entretiens.

Cette expérience avait pour but d'évaluer les hypothèses suivantes :

- H1 : La taxonomie proposée caractérise adéquatement les RCD pour leur usage par des concepteurs de RCD.
- H2 : La taxonomie utilise un vocabulaire qui correspond à la réalité de la pratique des concepteurs de RCD.
- H3 : L'archive proposée contient une sélection d'exemple pertinente pour les concepteurs de RCD comparativement aux exemples auxquels ils sont exposés hors de l'archive
- H4 : L'outil CurratedDDS améliore l'accès à des exemples pour des concepteurs de RCD, comparativement à leurs pratiques habituelles sans cet outil.
- H5 : L'utilisation de CurratedDDS facilite le travail d'idéation des concepteurs de RCD.

Les résultats montrent que les utilisateurs ont pu utiliser la taxonomie pour caractériser des exemples de RCD, mais aussi que le vocabulaire de la taxonomie, bien que pertinent, avait besoin de plus d'explications pour être compréhensible. Les hypothèses H1 et H2 sont donc partiellement validées.

Les résultats ont aussi montré que l'archive permet aux concepteurs de RCD de découvrir des contenus pertinents et qui diffèrent de leurs habitudes de consommation, mais présente plusieurs contraintes, notamment en termes de diversité linguistique et culturelle. L'hypothèse H3 est donc partiellement validée.

Enfin, ces mêmes résultats établissent que CurratedDDS améliore l'accès des utilisateurs à des exemples de RCD relativement à leurs méthodes habituelles et que ces exemples ont un effet positif sur le travail d'idéation, mais ce dernier est impacté par certaines limitations. L'hypothèse H4 est donc validée et l'hypothèse H5 partiellement validée.

5.2 Améliorations futures

L'analyse des résultats nous a permis de proposer plusieurs améliorations de la taxonomie, de l'archive et de CurratedDDS afin d'améliorer la méthode proposée. La taxonomie a reçu plusieurs adaptations sous la forme de dimensions ajoutées, retirées et modifiées, ainsi que sa division entre deux niveaux de détail. Des propositions d'extension de l'archive ont été faites afin d'améliorer sa couverture. Les modifications proposées pour CurratedDDS proposent une interface plus lisible et plus facile d'utilisation, pour améliorer l'expérience utilisateur. En particulier, l'ajout d'illustrations visuelles de certains concepts est une piste à explorer pour améliorer cette lisibilité et rendre plus concrets certains concepts introduits par la taxonomie. Nous avons aussi soulevé plusieurs limitations de notre protocole, qui permettent

de mettre en perspective nos résultats. Nous avons constaté que la gestion des verrous d'accès payant, la période sur laquelle a été conduite l'expérimentation et notre échantillon ont tous un impact sur les résultats de l'expérimentation.

Notre objectif initial était d'assister les concepteurs de RCD dans leur démarche d'idéation. Nous avons choisi d'explorer l'axe de la recherche d'exemples, ce qui nous a amené à proposer une méthode destinée à fournir des exemples aux concepteurs de RCD. D'autres angles pourraient être explorés pour décrire le rôle des exemples dans l'idéation. Une question qui pourrait être soulevée est comment un concepteur va-t-il identifier un exemple qui l'intéresse au moment de la lecture, après un éventuel filtrage. En effet, nous avons proposé une méthode pour explorer une archive d'exemples de RCD, mais une fois ces filtres appliqués, les utilisateurs doivent évaluer si les exemples qui leurs sont proposés sont effectivement pertinents et répondent à leur besoin. Examiner cette prise de décision permettrait de mieux comprendre ce qui amène un concepteur à favoriser un exemple de RCD et à en tirer plus ou moins de bénéfices. Un aspect de cette problématique, soulevé par les participants, est le besoin d'information visuelle, par exemple sous la forme d'extraits d'illustration de l'exemple. Ajouter de tels éléments visuels pourrait faciliter la prise de décision lors de la sélection d'exemples.

Une alternative serait de proposer d'autres formes d'assistance adressées aux concepteurs de RCD. Pour la démarche d'idéation nous pourrions imaginer un outil qui prenne un autre angle que les exemples, comme la conception de prototypes, ou la définition de projet. Un tel outil pourrait alors être employé en plus de CurratedDDS afin de cumuler leurs effets. Pour cela, il serait nécessaire d'identifier les défis liés à la conception de prototypes de RCD, par exemple comment représenter les éléments de taxonomie, ou bien les problématiques à soulever lors de la définition d'un nouveau projet de RCD. Une fois ces éléments identifiés, il faudrait alors proposer un moyen de les rendre accessible à un concepteur de RCD de façon efficace, mais qui ne vienne pas les limiter, par exemple en contraignant la définition de projet. En nous écartant de l'idéation elle-même, si nous voulons assister les concepteurs de RCD, il serait possible de créer des outils destinés à la construction de RCD basés sur la taxonomie proposée dans ce mémoire. Un participant (P5) a remarqué que la conception d'un RCD obligeait à sortir des patrons de code préexistants afin de créer une nouvelle page web. Proposer une bibliothèque de code facilitant la mise en place de certaines configurations décrites par notre taxonomie pourrait fournir un outil de productivité à des concepteurs de RCD. Pour cela, il faudrait identifier les éléments de RCD qui sont pertinents à intégrer à cette bibliothèque, puis proposer des implémentations efficaces et simples d'utilisation. Un défi serait d'établir la frontière entre ce qui doit être imposé par la bibliothèque afin de simplifier le contenu, et la liberté qui doit être laissée à l'utilisateur pour adapter le projet à ses travaux.

L'un des obstacles qui empêchent CurratedDDS de devenir un outil accessible au public est le caractère statique de l'archive. Nous avons collecté des exemples puis les avons annoté manuellement, et pour maintenir l'archive à jour il sera nécessaire de répéter ce processus. En l'absence d'une maintenance active, les contenus vont donc progressivement vieillir et peuvent perdre de leur intérêt. Il pourrait être intéressant d'assurer la continuité de ce projet en perpétuant l'archive avec des RCD récents. L'une des méthodes envisageables serait la mise en place d'un système communautaire pour alimenter l'archive. Construire un système permettant à des volontaires de soumettre un article annoté permettrait de mettre en valeur des articles divers, appréciés de la communauté, sans avoir à maintenir une présence permanente pour annoter des exemples en continu. Cette méthode présenterait deux défis. Le premier serait la qualité des exemples. Si ils sont ajoutés à l'archive par des volontaires, les articles seront moins contrôlés, ouvrant la possibilité de malentendus ou de conflits entre deux annotations différentes issues de deux interprétations d'un exemple. Le second défi serait la constitution d'une communauté autour de CurratedDDS. L'alimentation d'une archive d'exemples par des volontaires n'est possible que si des personnes choisissent de soumettre des contenus. Pour cela il faudrait donc non seulement mettre en place l'infrastructure pour rendre CurratedDDS disponible au grand public et aux utilisateurs potentiels, mais aussi s'assurer de le publiciser et de le rendre suffisamment performant et accessible pour susciter l'intérêt, afin de motiver la soumission spontanée de nouvelles entrées.

Une autre méthode qui pourrait être envisagée serait la mise en place d'une intelligence artificielle d'annotation automatique. Un tel système prendrait en entrée la page web pour identifier les différentes valeurs que prennent les dimensions, et annoter immédiatement un exemple. Ceci permettrait d'automatiser la mise à jour de l'archive, et donc d'assurer la présence d'entrées récentes pour de futurs utilisateurs. Cette proposition comporte ses propres défis. Elle repose sur la construction d'un modèle capable d'identifier les valeurs de chacune des dimensions. Ceci impliquerait une analyse du texte, pour identifier des informations telles que le sujet, mais aussi une analyse d'image pour identifier une visualisation de données présentée sous forme d'image, ou encore une analyse du code de la page, pour identifier des fonctionnalités de contrôle. Enfin, il faudrait identifier des éléments d'organisation, par exemple identifier une exploration en fin de récit qui caractérise une structure en verre à martini. Au-delà de ce modèle, il serait aussi intéressant de mettre en place un système de collecte automatique peut être mis en place en amont de l'annotation automatique.

Entre le début de ce projet et sa conclusion, les modèles d'intelligence artificielle générative ont reçu une attention publique grandissante, en partie grâce à des Large Language Models très médiatisés tels que ChatGPT. L'une des visions de ces outils était la génération d'exemples ou d'idées, y compris pour des concepts de visualisation de données. La création

de CurratedDDS s'inscrit donc dans un contexte où d'autres outils proposent une méthode très différente pour assister le processus d'idéation. Ce contexte ouvre plusieurs possibilités de futures recherches. D'un côté, il serait possible de comparer des exemples archivés, comme ceux de CurratedDDS, avec des exemples générés, textes ou images produits pour les besoins des utilisateurs par un modèle génératif. Ceci permettrait de comparer leurs utilités pour l'idéation, ou de développer des méthodes pour tirer le meilleur de l'un ou l'autre groupe en fonction de la situation. D'autre part, il serait envisageable d'utiliser des exemples de RCD pour créer une intelligence artificielle générative dédiée, qui proposerait des exemples de RCD, utilisant des caractéristiques propres aux RCD. Il serait possible de créer un modèle génératif basé sur notre méthode, qui utiliserait la taxonomie pour définir un exemple à générer, et entraîné sur une archive similaire à la nôtre. Un tel projet permettrait de proposer des exemples de RCD en grand nombre et correspondant à des requêtes spécifiques. Il rencontrerait néanmoins plusieurs obstacles. Le plus immédiat serait la création d'un modèle générant des RCD, c'est-à-dire un ensemble de visualisations de données, d'illustrations, de texte et d'interactions, structurés de façon à construire le récit. Un moyen de résoudre cette problématique pourrait être de s'inspirer de modèles générant du code, tels que GitHub Copilot, afin de produire une page web. Une autre problématique serait l'entraînement. Nous disposons actuellement de 780 exemples dans notre archive, d'autres entités peuvent disposer de leurs propres archives, mais entraîner un modèle génératif pourrait nécessiter des archives d'exemples de l'ordre du millier, voire plus, pour être correctement entraîné. Ceci est d'autant plus vrai que certaines dimensions ont des valeurs rares. Acquérir ces données serait un défi logistique, mais soulève aussi des questions de propriété intellectuelle. Nous pouvons proposer à un utilisateur de consulter un article du New York Times, mais utiliser les images, le texte et le code source de cet article pour entraîner une intelligence artificielle nécessitera au minimum une autorisation.

RÉFÉRENCES

- [1] J. Gray, L. Chambers et L. Bounegru, *The data journalism handbook : How journalists can use data to improve the news.* " O'Reilly Media, Inc.", 2012.
- [2] O. québécois de la langue française, "journalisme de données — vitrinelinguistique.oqlf.gouv.qc.ca," <https://vitrinelinguistique.oqlf.gouv.qc.ca/fiche-gdt/fiche/26522905/journalisme-de-donnees>, 2013, [Accessed 08-02-2024].
- [3] K. Collins, A. Sun, E. Lutz et L. Buchanan, "78 Long Minutes (Published 2022) — nytimes.com," <https://www.nytimes.com/interactive/2022/05/28/us/school-shooting-timeline-uvalde-texas.html?bgrp=a&smid=url-share>, 2022, [Accessed 05-02-2024].
- [4] Wikipédia, "Idéation — wikipédia, l'encyclopédie libre," 2023, [En ligne; Page disponible le 10-janvier-2023]. [En ligne]. Disponible : <http://fr.wikipedia.org/w/index.php?title=Id%C3%A9ation&oldid=200333606>
- [5] C. N. Knaflic, *Storytelling with data : A data visualization guide for business professionals.* John Wiley & Sons, 2015.
- [6] H. K. Bako, X. Liu, L. Battle et Z. Liu, "Understanding how designers find and use data visualization examples," *IEEE Transactions on Visualization and Computer Graphics*, vol. 29, p. 1–11, 2022.
- [7] E. Segel et J. Heer, "Narrative visualization : Telling stories with data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 16, p. 1139–1148, 11 2010.
- [8] N. H. Riche, C. Hurter, N. Diakopoulos et S. Carpendale, *Data-driven storytelling.* CRC Press, 2018.
- [9] Y. Fu et J. Stasko, "More than data stories : Broadening the role of visualization in contemporary journalism," *IEEE Transactions on Visualization and Computer Graphics*, p. 1–20, 01 2023.
- [10] R. Kosara et J. Mackinlay, "Storytelling : The next step for visualization," *Computer*, vol. 46, n°. 5, p. 44–50, 2013.
- [11] W. Weber, M. Engebretsen et H. Kennedy, "Data stories : Rethinking journalistic storytelling in the context of data journalism," *Studies in communication sciences*, vol. 2018, n°. 1, p. 191–206, 2018.
- [12] C. D. Stolper, B. Lee, H. Riche et J. Stasko, "Emerging and recurring data-driven storytelling techniques : Analysis of a curated collection of recent stories,"

2016. [En ligne]. Disponible : <https://www.microsoft.com/en-us/research/publication/emerging-and-recurring-data-driven-storytelling-techniques-analysis-of-a-curated-collection-of-recent>.
- [13] N. Duarte, *Data story : explain data and inspire action through story*. Ideapress Publishing, 2019.
- [14] W. Weber et H. Rall, “Data visualization in online journalism and its implications for the production process,” *2012 16th International Conference on Information Visualisation*, 07 2012.
- [15] S. K. Card, J. Mackinlay et B. Shneiderman, *Readings in information visualization : using vision to think*. Morgan Kaufmann, 1999.
- [16] *Data stories of water : Studying the communicative role of data visualizations within long-form journalism*, vol. 42, n°. 3. COMPUTER GRAPHICS forum, 2023.
- [17] A. Ojo et B. Heravi, “Patterns in award winning data storytelling,” *Digital Journalism*, vol. 6, n°. 6, p. 693–718, 2018. [En ligne]. Disponible : <https://doi.org/10.1080/21670811.2017.1403291>
- [18] A. Thudt, J. Walny, T. Gschwandtner, J. Dykes et J. Stasko, *Exploration and explanation in data-driven storytelling*, ser. Data-Driven Storytelling. AK Peters/CRC Press, 04 2018, p. 59–83.
- [19] B. Bach, M. Stefaner, J. Boy, S. Drucker, L. Bartram, J. Wood, P. Ciuccarelli, Y. Engelhardt, U. Koeppen et B. Tversky, *Narrative design patterns for data-driven storytelling*, ser. Data-Driven Storytelling. AK Peters/CRC Press, 04 2018, p. 107–133.
- [20] F. Stalph, “Classifying data journalism a content analysis of daily data-driven stories,” *Journalism Practice*, vol. 12, p. 1332–1350, 10 2017.
- [21] N. Abraham et C. Landaverde, “A conceptual framework for interactive cartographic storytelling,” Thèse de doctorat, 2018. [En ligne]. Disponible : <http://essay.utwente.nl/85868/>
- [22] R. E. Roth, “Cartographic design as visual storytelling : Synthesis and review of map-based narratives, genres, and tropes,” *The Cartographic Journal*, vol. 58, p. 1–32, 09 2020.
- [23] P. Parsons, “Visualization design patterns - infovis :wiki,” infovis-wiki.net, 03 2022. [En ligne]. Disponible : https://infovis-wiki.net/wiki/Visualization_Design_Patterns
- [24] Z. Zhao et N. Elmqvist, “The stories we tell about data : Surveying data-driven storytelling using visualization,” *IEEE Computer Graphics and Applications*, vol. 43, n°. 04, p. 97–110, jul 2023.

- [25] C. D. Stolper, B. Lee, N. H. Riche et J. Stasko, *Data-driven storytelling techniques : Analysis of a curated collection of visual stories*, ser. Data-Driven Storytelling. AK Peters/CRC Press, 04 2018, p. 85–105.
- [26] P. Parsons, “Understanding data visualization design practice,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, n^o. 1, p. 665–675, 2021.
- [27] N. Maiden, K. Zachos, A. Brown, D. Apostolou, B. Holm, L. Nyre, A. Tonheim et A. van den Beld, “Digital creativity support for original journalism,” *Commun. ACM*, vol. 63, n^o. 8, p. 46–53, jul 2020. [En ligne]. Disponible : <https://doi.org/10.1145/3386526>
- [28] P. C. Parsons, “Design cognition in data visualization,” dans *Visualization Psychology*. Springer, 2023, p. 197–215.
- [29] A. V. Moere et H. Purchase, “On the role of design in information visualization,” *Information Visualization*, vol. 10, n^o. 4, p. 356–371, 2011.
- [30] M. Du Sautoy, “The creativity code,” dans *The Creativity Code*. Harvard University Press, 2019.
- [31] T.-K. Laamanen et P. Seitamaa-Hakkarainen, “Interview study of professional designers’ ideation approaches,” *The Design Journal*, vol. 17, n^o. 2, p. 194–217, 2014.
- [32] L. Bounegru et J. Gray, *The Data Journalism Handbook : Towards a Critical Data Practice*. Amsterdam University Press, 2021.
- [33] Z. Wu, Q. Xu, Y. Liu, Z. Peng, Y. Xu et X. Ma, “Exploring designers’ practice of online example management for supporting mobile ui design,” *Proceedings of the 23rd International Conference on Mobile Human-Computer Interaction*, 09 2021.
- [34] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh et A. Hung Byers, “Big data : The next frontier for innovation, competition, and productivity,” 2011.
- [35] Z. Zhao et N. Elmqvist, “The stories we tell about data : Surveying data-driven storytelling using visualization,” *IEEE Computer Graphics and Applications*, vol. 43, n^o. 04, p. 97–110, jul 2023.
- [36] A. Kirk, *Data Visualisation : A Handbook for Data Driven Design*. SAGE, 2016.
- [37] J. Christopher et K. Richardson, “Media topics,” IPTC, 06 2023. [En ligne]. Disponible : <http://iptc.org/standards/media-topics/>

ANNEXE A DÉFINITIONS DES ÉLÉMENTS DE TAXONOMIE

TABLEAU A.1 Récapitulatif des définitions de groupes de la taxonomie

Élément	Définition
Metadonnées	Informations générales sur l'article
Contexte	Cadre du projet et démarche de l'article
Données	Jeu de données et traitement qui lui est appliqué
Contenus	Visualisations et autres contenus de l'article
Contrôle	Contrôle donné à l'audience sur l'article
Schémas Narratifs	Options proposées pour travailler avec des schémas narratifs (définition de Narrative Design Patterns for Data-Driven Storytelling :A narrative pattern is a low-level narrative device that serves a specific intent. A pattern can be used individually or in combination with others to give form to a story.)

TABLEAU A.2 Récapitulatif des définitions de dimensions de la taxonomie

Dimension	Définition
Titre	Titre de l'article
Date	Année de publication
Source	Lien vers la page de l'article
But	Quel est l'objectif de communication de l'auteur ?
Sujet	Quel est le sujet de l'article selon le standard de l'IPTC ??
Structure	Assortiment du nombre de d'unités visuelles, assemblées dans le temps et l'espace, et de l'ordonnement de leurs éléments visuels.
Forme	Les données employées sont-elles qualitatives, quantitatives ou les deux ?
Type	Quels types de données sont contenus dans le dataset ?
Analyse	L'article utilise-t-il des techniques d'analyse informatisée complexes telles que le Machine Learning, le Traitement de la Langue Naturelle, ou le Data Mining pour extraire des informations supplémentaires du dataset ?
Visualisations	Quels types de visualisation font partie de l'article ?
Médias	Quels autres médias sont présents autour des visualisations ?
Tutoriel	Quelle méthode est employée pour montrer au public comment tirer le maximum de l'article ?
Degré de liberté de l'utilisateur	Comment le projet équilibre-t-il l'exploration autonome de l'utilisateur et le contrôle de l'auteur ?
Séquence	De quelle façon avancent les étapes de la visualisation ?
Transition	Un effet visuel est-il associé l'avancement entre étapes ?
Indicateurs de progression	Indicateur visuel de l'état d'avancement du lecteur
Interactions	Options proposées au lecteur pour interagir avec les visualisation / se les approprier
Schémas argumentatifs	Schémas utilisés pour persuader et convaincre l'audience.
Schémas de cadrage	Schémas utilisés pour intégrer l'audience dans l'histoire ou la garder à distance
Schémas d'engagement de d'empathie	Schémas utilisés pour engager plus profondément l'audience

TABLEAU A.3 Récapitulatif des définitions de valeurs de la taxonomie (Partie 1)

Dimension associée	Valeur	Définition
Structure	Animation	Progression linéaire avec l'avancée du temps
	Compilation	De multiples histoires rassemblées ensemble, ordonnées de façon à créer une progression
	Diaporama	Avancée à travers une série de diapositives successives
	Image Statique	Image unique qui forme une histoire en partitionnant la vue. L'avancée de l'histoire vient de l'ordre de lecture
	Infographie déroulante	Histoire qui progresse en faisant défiler la page
	Multimédia	Progression via l'ouverture d'hyperliens successifs
	Storymap	Progression avec l'évolution d'une carte
	Autre	Autre forme de progression
Type	Numérique	Données numériques chiffrées
	Réseau	Données représentant des connexions entre entités
	Spatial	Coordonnées géographiques, trajets
	Temporel	Informations datées
	Texte	Données issues de l'analyse d'un corpus de texte
	Autres	Données ne répondant pas à l'un des types précédents
Tutoriel	Sans	Pas de tutoriel ou d'explications
	Explicite	Une boîte de texte ou un paragraphe explique les fonctions interactives.
	Implicite	Une mise en surbrillance ou animation guide l'utilisateur dans l'utilisation des fonctions interactives.
Degré de liberté de l'utilisateur	Axé sur l'auteur	Une approche purement axée sur l'auteur a un chemin linéaire strict à travers la visualisation, s'appuie fortement sur la messagerie et n'inclut aucune interactivité.
	Verre à martini	La structure en verre à martini commence par une approche axée sur l'auteur, utilisant initialement des questions, des observations ou des articles écrits pour introduire la visualisation. Une fois que le récit prévu par l'auteur est terminé, la visualisation s'ouvre sur une étape axée sur le lecteur où l'utilisateur est libre d'explorer les données de manière interactive.
	Diaporama interactif	La structure du diaporama interactif suit un format de diaporama typique, mais intègre une interaction au milieu de la narration dans les limites de chaque diapositive. Cette structure permet à l'utilisateur d'explorer plus avant des points particuliers de la présentation avant de passer à l'étape suivante de l'histoire.
	Forage	La structure de forage présente un thème général et permet ensuite à l'utilisateur de choisir parmi des instances particulières de ce thème pour révéler des détails et des histoires supplémentaires.
	Axé sur le lecteur	Une approche purement axée sur le lecteur n'a pas d'ordre prescrit d'images, pas de messagerie et un degré élevé d'interactivité.

TABLEAU A.4 Récapitulatif des définitions de valeurs de la taxonomie (Partie 2)

Dimension associée	Valeur	Définition
Séquence	Sans	La structure de l'article fait qu'il n'y a pas de commande pour progresser.
	Linéaire continue	La progression se fait en défilant la page.
	Linéaire par étapes	La progression se fait en avançant dans des étapes ponctuelles.
	Navigation non linéaire	La progression se fait en navigant dans un espace non linéaire.
Transition	Sans	Le contenu de l'écran contient la totalité de l'histoire.
	Atténuation (Fading)	Le contenu de l'écran se diffuse et est remplacé par le nouveau contenu.
	Glissement (Swiping)	Le contenu de l'écran glisse vers l'extérieur
	Interpolation (Tweening)	Le contenu se transforme pour faire place au nouveau contenu.
	Panoramique (Panning)	Le contenu de l'écran glisse de façon à donner l'illusion que le nouveau contenu se trouvait hors écran.
Indicateurs de progression	Sans	La progression n'est pas indiquée par un indicateur visible
	Barre de progrès	La progression est indiquée par une barre qui se remplit progressivement
	Boutons de section	La progression est indiquée par la mise en valeur progressive d'indicateurs représentant les sections
	Ligne temporelle	La progression est indiquée par une ligne temporelle et un marqueur qui indique l'état de la lecture
	Minicarte avec chemin	La progression est indiquée par un chemin sur une carte miniature
	Navigation par fil d'Arianne (breadcrumbs)	La progression est indiquée par un guide de sous sections indiquées qui mènent jusqu'à celle en cours
	Liste de pointage (checklist)	La progression est indiquée en cochant des cases dans une liste de sections
Interactions	Abstraction/élaboration	Montre plus ou moins de détails
	Comparaison	Montre deux items pour les comparer
	Connection	Montre des items associés
	Encodage	Montre une représentation différente
	Exploration (navigation)	Montre autre chose
	Filtre	Montre quelque chose conditionnellement
	Reconfiguration	Montre un arrangement différent
	Sélection	Marque quelque chose comme intéressant

TABLEAU A.5 Récapitulatif des définitions de valeurs de la taxonomie (Partie 3)

Dimension associée	Valeur	Définition
Schémas argumentatifs	Pas de définition	
	Comparaison	Le narrateur présente deux ensembles de données ou plus et en tire des conclusions respectives.
	Concrétisation	Illustre des concepts abstraits avec des objets concrets. Chaque point de données est représenté par un objet visuel individuel, ce qui les rend moins abstraits que les statistiques agrégées.
	Répétition	La répétition peut également aider à l'argumentation, car elle peut augmenter l'importance et la mémorisation d'un message, et peut aider à lier différents arguments sur un ensemble de données donné. La répétition peut être utilisée comme moyen de rechercher une réponse dans les données.
Schémas de cadrage	Pas de définition	
	Défamiliarisation	La défamiliarisation défie les attentes du public, en présentant quelque chose de familier d'une manière inattendue.
	Familiarisation	La familiarisation aide à créer un cadre familier pour le public afin qu'il puisse relier des éléments de données à ce qu'il sait déjà.
	Métaphore physique	Représenter un élément du récit avec un symbole plus tangible
	Rupture de conventions	La rupture de convention peut également être utilisée en établissant d'abord une convention graphique à travers, par exemple, la couleur, les échelles ou le rythme visuel dans le récit, puis en la perturbant à un point spécifique.
Schémas d'engagement et d'empathie	Pas de définition	
	Accélérations / décélérations	La narration est ralentie et concentrée sur un seul aspect des données, tout en laissant l'utilisateur seul avec les données. Le contraire des ralentissements et des pauses est l'accélération de la narration ou de la présentation des données.
	Rupture du 4ème mur	En s'adressant directement au public, le narrateur brise le quatrième mur.
	Devinette	Inviter l'audience à deviner les données ou le résultat d'une série d'événements présentés dans le récit peut leur permettre de contrôler le déroulement de l'histoire.
	Exploration	Le public est confronté à une interface complexe où les paramètres des données peuvent être définis librement.
	Les humains derrière les points	Le schéma des "humains derrière les points" consiste à présenter des histoires individuelles à travers des détails à la demande dans une visualisation
	Question rhétorique	En s'adressant directement au public, une question rhétorique peut connecter le narrateur au public et déclencher sa volonté de répondre à cette question, même si ce n'est que pour un instant.
	Révélation graduelle	Un schéma courant est la révélation, dans laquelle des éléments des données sont progressivement révélés, conduisant finalement à l'image d'ensemble et à l'argument final.

ANNEXE B GUIDE D'ENTRETIEN

Introduction

Lancement de l'appel, accueil de la personne interviewée, présentation, mise en place.

Présentation de l'objectif d'entretien

Tout d'abord merci d'avoir accepté de participer à cet entretien. L'objectif de cet échange est de collecter des retours du test sur l'outil qui vous a été proposé dans le cadre d'une expérience sur l'accès à des exemples de Data Driven Stories.

Consentement à l'enregistrement

Avant de commencer, je souhaite vous demander votre autorisation pour enregistrer cette entrevue. L'enregistrement a pour but de garantir une prise de notes précise et une analyse de notre conversation. Êtes-vous d'accord pour que j'enregistre notre entretien ?

Droits de la personne interviewée et sécurité

Vous pouvez à tout moment demander à mettre fin à l'entretien, sans condition ou justification. Toutes les informations seront anonymisées dans l'analyse finale des données. Le stockage des données d'entretien sur une plateforme cloud fermée, exclusivement accessible aux membres du laboratoire jData participant à l'expérience, c'est-à-dire moi même et Thomas Hurtut. Vos réponses sont sur une base volontaire, et vous pouvez à tout moment refuser de répondre ou mettre fin à l'entretien. Avant de débiter, avez-vous des questions ?

Retours d'utilité de l'outil

Questions de base

- Commençons par ma première question : est-ce que vous avez eu des occasions d'utiliser CurratedDDS, l'outil du projet ?
- Dans quelles situations avez-vous utilisé l'outil ?
- Comment avez-vous utilisé CurratedDDS ?

Influence

- Pensez vous que CurratedDDS a eu un impact sur votre temps et vos efforts de travail ?
- Lequel ?

- Pouvez vous me dire comment l'outil a pu vous aider dans votre recherche d'idées ?
- Pouvez vous décrire des moments où l'outil a pu vous aider dans votre recherche d'idée ?
- Pouvez vous me dire comment l'outil a pu vous limiter ou vous contraindre dans votre recherche d'idées ?
- Pouvez vous décrire des moments où l'outil a pu vous limiter dans votre recherche d'idée ?
- Estimez vous que CurratedDDS a pu influencer votre travail d'une autre façon ?
- Pouvez vous me décrire des moments où ça a été le cas ?

Exemples

- Qu'avez-vous pensé de l'intérêt et de la qualité des exemples ?
- Quels impacts ces exemples ont-ils pu avoir sur votre travail ?
- Pouvez-vous me décrire les situations où cet impact a été visible ?
- Pensez-vous que les exemples ont pu avoir sur votre créativité ?
- Pouvez-vous me décrire les situations où cet impact a été visible ?
- Selon-vous, quels changements seraient nécessaires pour améliorer la pertinence de ces exemples ?

Retour d'utilisabilité de l'outil

Questions de base

- Passons maintenant à l'utilisabilité de l'outil. Pouvez vous me donner une impression générale ?

Ergonomie

- Qu'avez-vous pensé de l'ergonomie de l'outil ?
- Y-a-t-il des choses qui vous ont posé des difficultés pour utiliser l'outil ? Lesquelles ?
- Est-ce que vous pouvez me décrire ces situations plus en détail ?
- Comment changeriez-vous l'outil pour le rendre plus facile à utiliser ?

Interactions

- Qu'avez-vous pensé des interactions proposées ?
- Pensez-vous que les interactions ont influencé la façon dont vous avez abordé les exemples ?
- Est-ce qu'il y a des exemples concrets ?
- Comment changeriez-vous les options d'interaction pour les rendre plus intéressantes ?

Changements

- Quels changements apporteriez-vous à l'outil pour le rendre plus utile ?

Retours sur la Taxonomie

Questions de base

- Le point suivant que je souhaite aborder est la taxonomie utilisée pour le système de filtrage. Nous avons proposé un certain nombre de catégories et d'options de description pour ces filtres. Quelle est votre impression de ce vocabulaire que nous avons utilisé ?

Clarté

- Que pensez-vous de la clarté de cette taxonomie ?
- Quelles modifications feriez-vous pour la rendre plus claire ?

Pertinence

- Comment estimez-vous la pertinence des options proposées ?
- Quelles sont les options proposées qui vous semblent le plus ou le moins pertinentes ?

Influence

- Pouvez vous me dire comment les concepts a pu vous inspirer dans votre recherche d'idées ?
- Est-ce que vous pouvez de décrire des moments ou cette influence s'est faite sentir ?
- Pouvez vous me dire comment la taxonomie a pu vous limiter ou vous contraindre dans votre recherche d'idées ?
- Est-ce que vous avez des exemples de situations précises où ça a été le cas ?
- Y a-t-il des aspects spécifiques de ces descriptions qui vous ont particulièrement influencé ?

Améliorations

- Comment amélioreriez vous cette taxonomie ?
- Y-a-t-il des choses à ajouter ? A supprimer ?

Utilisation

- Pouvez-vous commenter comment vous avez utilisé la taxonomie pendant le projet ?
- Avez vous remarqué des tendances dans votre façon d'utiliser les filtres ?

Fin

Invitation à aborder d'autres sujets

- Y-a-t-il des sujets qui n'ont pas été couverts par cet entretien et qui devraient être discutés ?
- Souhaitez-vous ajouter quelque chose avant de terminer ?