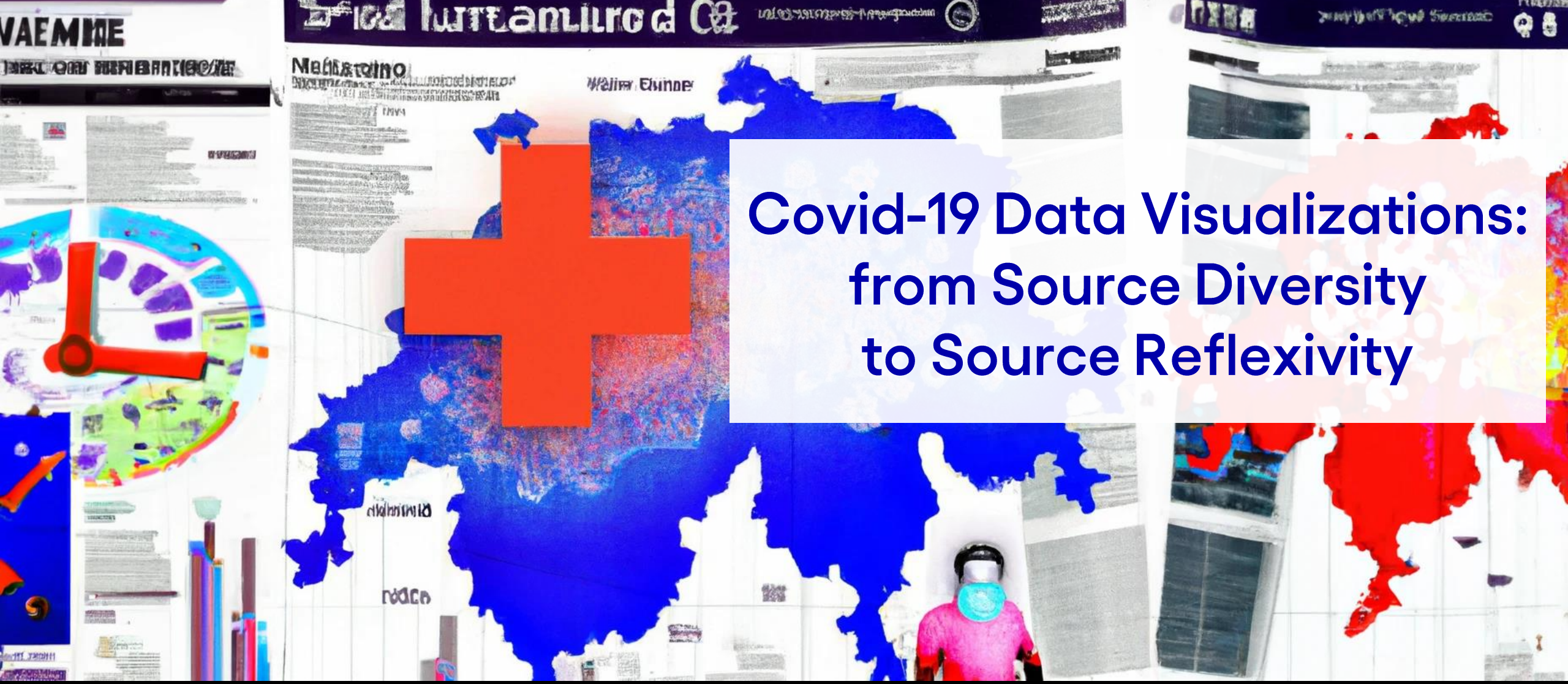


Covid-19 Data Visualizations: from Source Diversity to Source Reflexivity

Andrew Robotham, Céline Dupuis and Nathalie Pignard-Cheynel

Academy of Journalism and Media, University of Neuchâtel, Switzerland

unine
Université de Neuchâtel

ECREA 25.09.2025

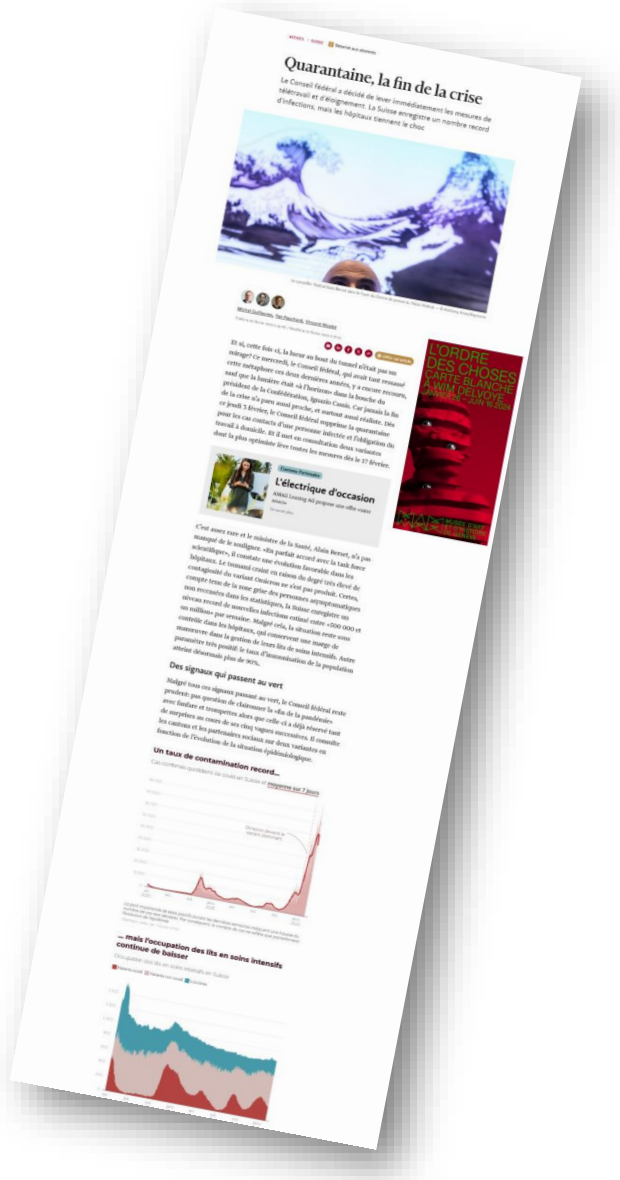
What are we trying to understand in this study?

Research questions

RQ 1: How (often) did different Swiss news media use data visualizations in their online news articles, and what types or data visualizations were used?

RQ2: What kind of sources were used/cited for these data visualizations and did their use vary according to media and over time?

RQ3: Did the data visualizations include text that presented the source critically or reflexively?



Covid-19 data journalism
Data visualizations:
from Source Diversity
to Source Reflexivity

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

The broad(er) context

The wider project Swiss National Science Foundation

1. Corpus of approx. 180,000 online news articles from 5 news media
2. Research interviews with 20 professionals (journalists & some public administration officials)
3. Audience reception experiments
4. Prototyping of new and innovative story formats

Covid-19 data journalism
Data visualizations:
from Source Diversity
to Source Reflexivity

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

The broad(er) context

The wider project Swiss National Science Foundation

1. Corpus of approx. 180,000 online news articles from 5 news media
2. Research interviews with 20 professionals (journalists & some public administration officials)
3. Audience reception experiments
4. Prototyping of new and innovative story formats

The study I am presenting at ECREA

In this dataviz study we actually "only" considered 89,000 online news articles and their 608 dataviz

Covid-19 data journalism
Data visualizations:
from Source Diversity
to Source Reflexivity

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

Why are we interested in data visualizations, sources and metadiscourse during Covid-19?

Data visualizations in general in news articles

- There is research to suggest that (given certain conditions) data visualizations serve as incentives, attention grabbers and focal points of stories (de Haan et al. 2018)
- Within their "host article", they provide a full set rather than an individually selected indicator (which may be more easily taken out of context)

> "visual displays present the highest bandwidth channel from the computer to the human" (Ware 2004,2)

Neuchâtel journalisme
ew Robotham
eline Dupuis
Pignard-Cheynel

Why are we interested in data visualizations, sources and meta-discourse during Covid-19?

Covid-19 data sources matter because:

- Data journalism has been shown to be highly reliant on government data (Borges-Rey 2020, Stalph 2018)
- Research has suggested that not mentioning the source of data in infographics compromises trust (de Haan et al. 2018)

→ Issues related to trust and the use of government data were likely exacerbated during the pandemic

> "Only the two data visualisations based on survey data with unknown data sources were rated below three by approximately one third of the participants."
> Tong 2023, 7

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

Why are we interested in data visualizations, sources and metadiscourse during Covid-19?

Metadiscourse matters because

- News media found themselves walking a tightrope between a watchdog function and a public service function
- Being transparent about the data is generally seen as good practice because:
    - It may reflect data journalists' critical engagement with the underlying data and its source (Lowrey et al. 2019)
    - (For this reason,) metadiscourse about source and data is generally believed to improve audience trust

> "Findings indicated that data journalists may peer into the top of the data 'black box', but they are unlikely to look deeply" Lowrey et al. 2019, 79

Covid-19 data journalism
Data visualizations:
from Source Diversity
to Source Reflexivity

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

A quick word on method

Four French-language Swiss online news media reflecting diverse audiences and editorial strategies

Public broadcaster	Legacy	Health and science	Regional
RTS INFO	LE TEMPS	HEIDI.NEWS IL EST TEMPS DE RACONTER LE MONDE	24heures

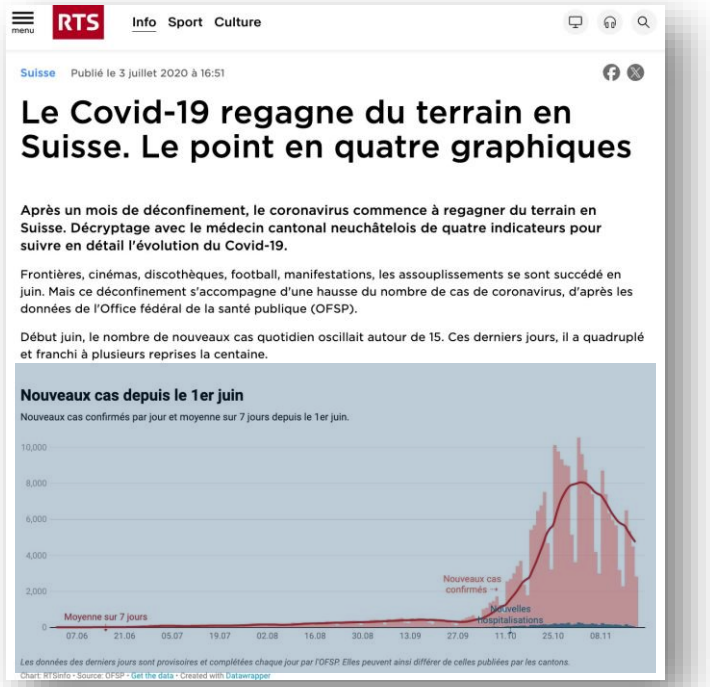
Covid-19 data journalism
Data visualizations:
from Source Diversity
to Source Reflexivity

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

A quick word on method

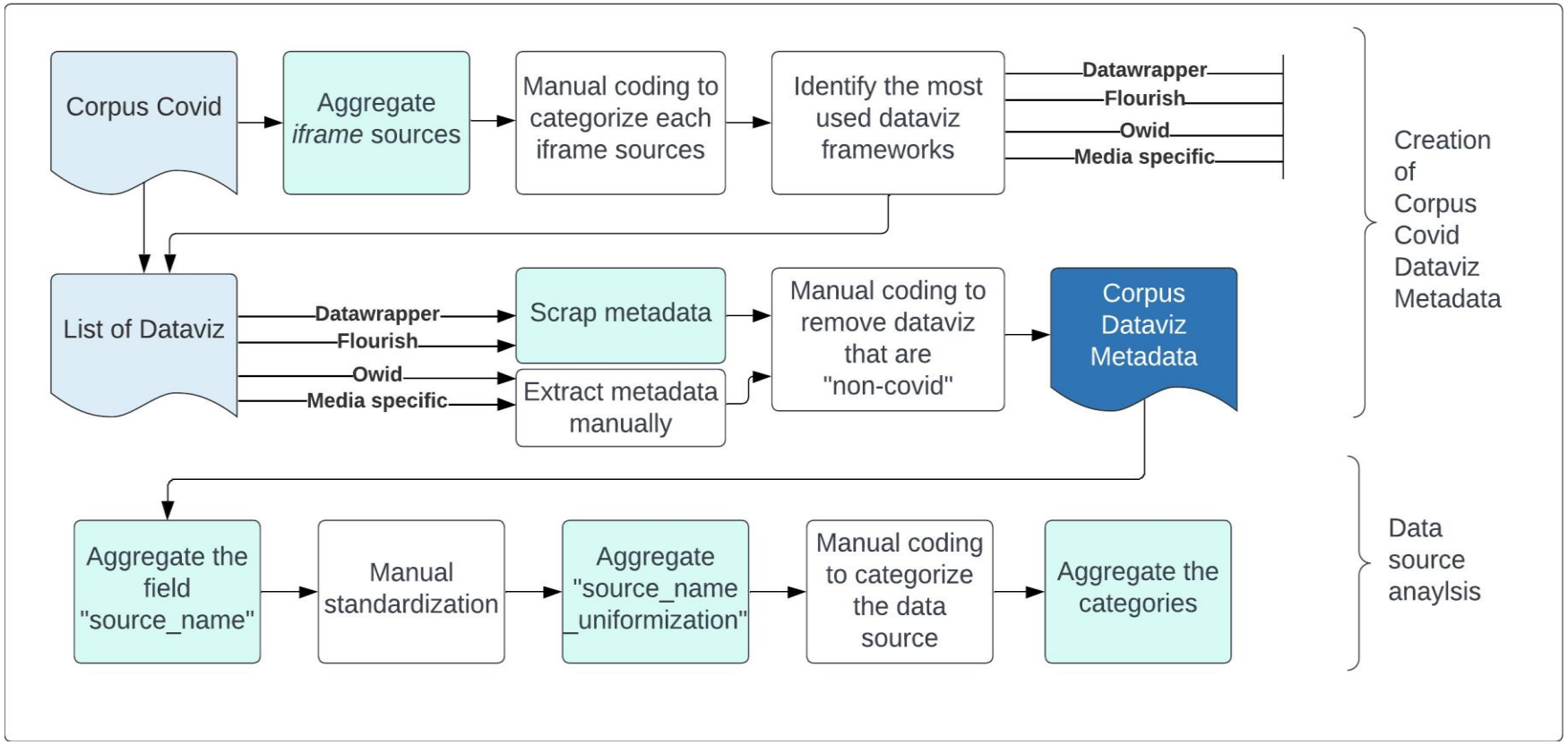
For practical purposes, a data visualization in our study is defined as a graphic element that is integrated into the news article via an embedded `<iframe>`



unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

A quick word on method



Covid-19 data journalism
Data visualizations:
from Source Diversity
to Source Reflexivity

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

A quick word on method

Simply put:

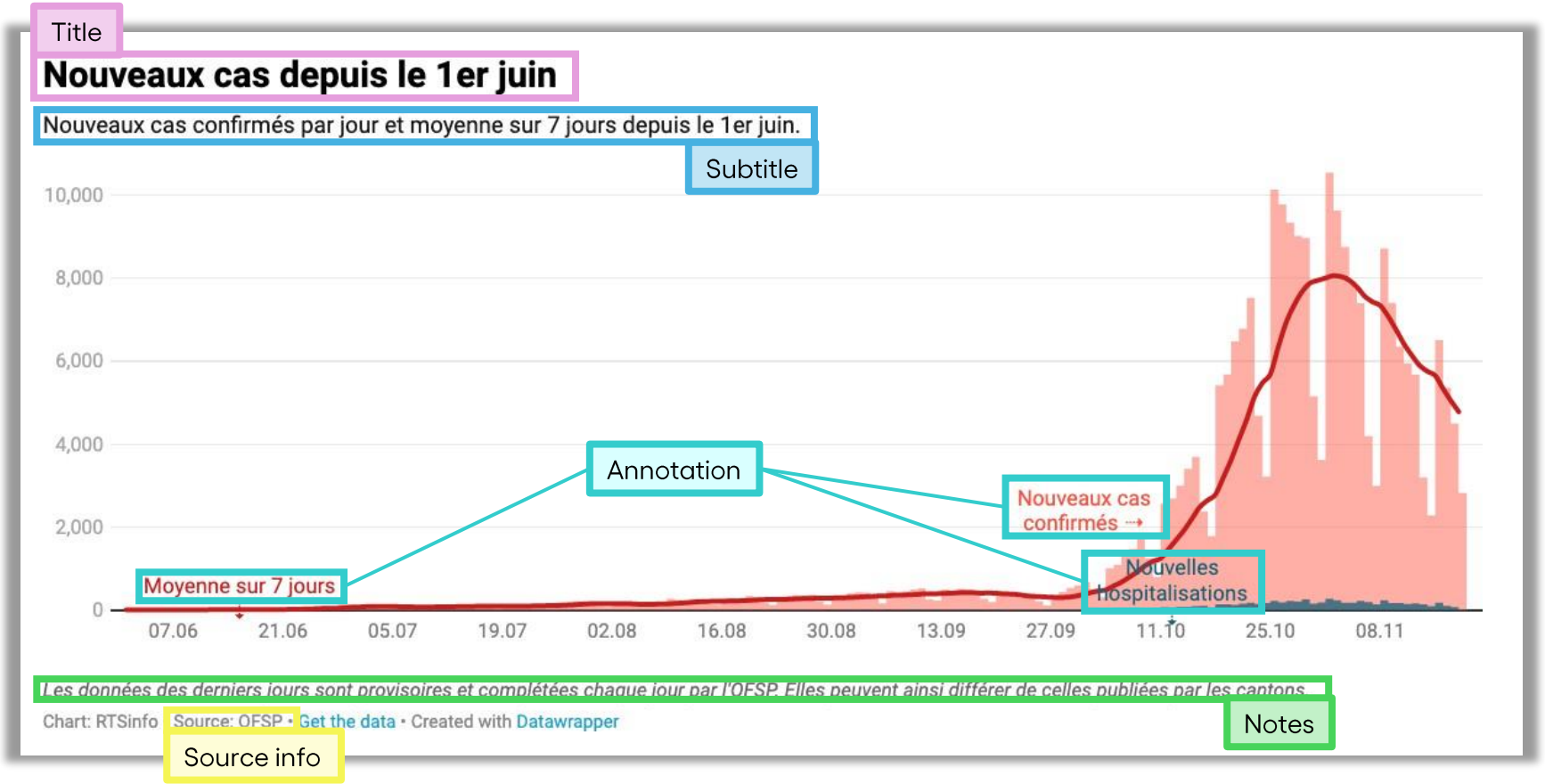
1. Create of a giant corpus consisting of all online articles from 4 Swiss media (URLs)
2. Identify those that are «about Covid»
3. Find all `<iframe>`
4. Identify `<iframe>` that are data visualizations (and not for example Tweets)
5. Exclude `<iframe>` data visualizations that are not related to the pandemic
6. Identify duplicates (and calculate range)
7. Extract all relevant metadata
    - Title, subtible
    - Source of the data (and standardize)
    - Other text (legend, notes, etc.)
8. Standardize and categorize (e.g type of source, etc.)
9. Analze, analyze, analyze…

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

A quick word on method

A data visualization and its textual elements



unine
Université de Neuchâtel
Académie du journalisme
et des médias

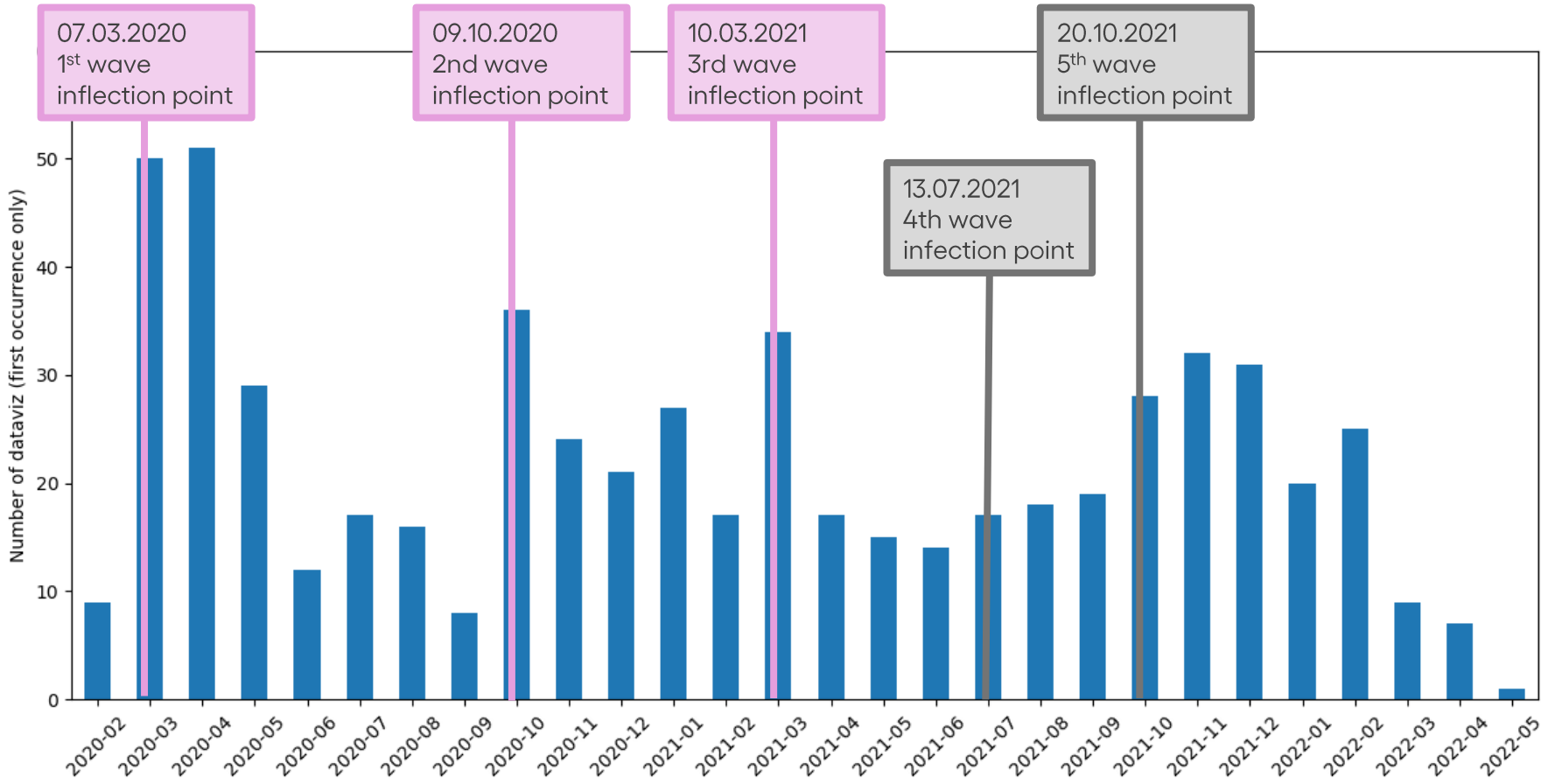
Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

RQ 1

How (often) did different Swiss news media use data visualizations in their online news articles, and what types of data visualizations were used?

Results: dataviz waves

Number of new data visualizations published during each month of the pandemic (first occurrence only)



Covid-19 data journalism
Data visualizations:
from Source Diversity
to Source Reflexivity

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

Results : an overview

	RTS (broadcast) N=33,916	Le Temps (legacy) N=32,969	Heidi.news (sci. & health) N=7548	24 Heures (local) N=14,978	Total N=89,411
Nb of Covid-19 articles	6701 articles	8844 articles	2751 articles	2101 articles	20397 articles
Nb of articles with at least 1 dataviz	920 articles	495 articles	88 articles	177 articles	1677 articles
Nb of occurrences of dataviz	2176 occurrences	779 occurrences	161 occurrences	246 occurrences	3362 occurrences
Nb of unique dataviz (counts as 1 when in multiple articles)	348 unique dataviz	102 unique dataviz	124 unique dataviz	34 unique dataviz	608 unique dataviz
Nb of single use dataviz	250 single use dataviz	81 single use dataviz	114 single use dataviz	29 single use dataviz	474 single use dataviz

Covid-19 data journalism
Data visualizations:
from Source Diversity
to Source Reflexivity

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

Results : some media's articles we much more likely to contain dataviz

	RTS (broadcast) N=33,916	Le Temps (legacy) N=32,969	Heidi.news (sci. & health) N=7548	24 Heures (local) N=14,978	Total N=89,411
% Covid-19 articles with at least 1 dataviz	13.7%	5.6%	3.2%	8.4%	8.2%

Daily live blogs with many multiple use dataviz (cases, deaths, etc.)

Often made to measure

Some .jpg .png graphs not included

Covid-19 data journalism
Data visualizations:
from Source Diversity
to Source Reflexivity

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

Results : some media reused certain dataviz often, others did not

	RTS (broadcast) N=33,916	Le Temps (legacy) N=32,969	Heidi.news (sci. & health) N=7548	24 Heures (local) N=14,978	Total N=89,411
Average nb of uses of unique dataviz	6.25	7.64	1.3	7.2	5.5
% single use dataviz	71.8%	79.4%	91.9%	85.3%	80.0%

Often made to measure

Some .jpg .png graphs not included

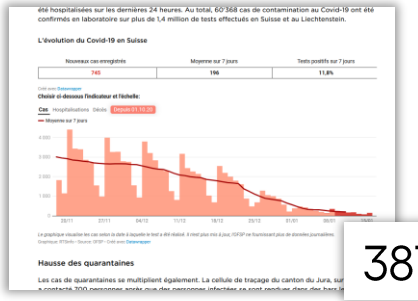
Covid-19 data journalism
Data visualizations:
from Source Diversity
to Source Reflexivity

unine
Université de Neuchâtel
Académie du journalisme
et des médias

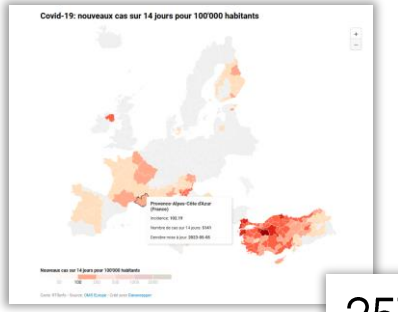
Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

Results: maps and timelines were reused

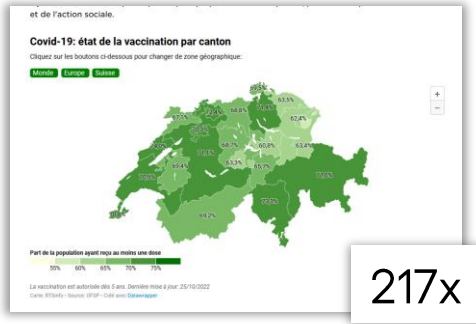
RTS INFO (broadcast)



387x



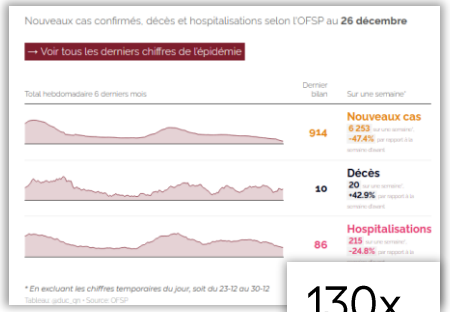
257x



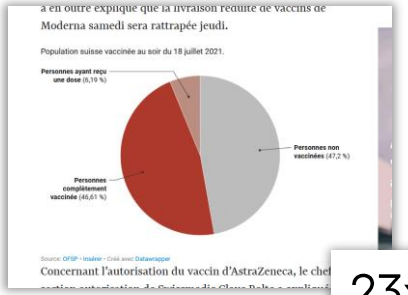
217x

LE TEMPS (legacy)

401x

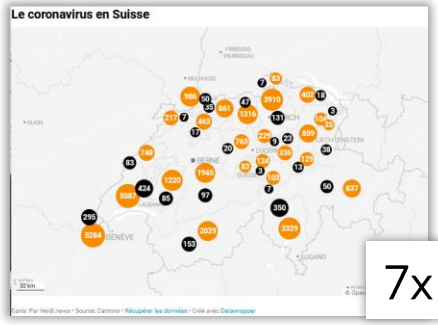


130x

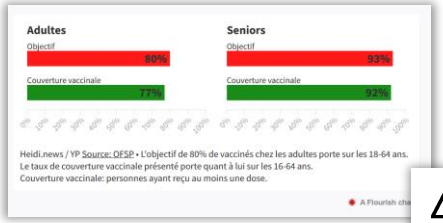


23x

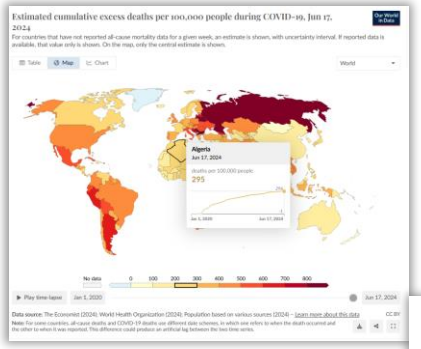
HEIDI.NEWS (specialist)



7x

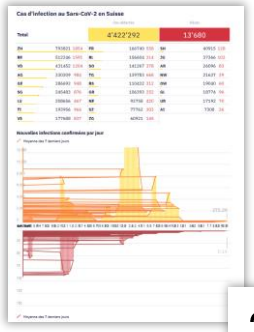


4x

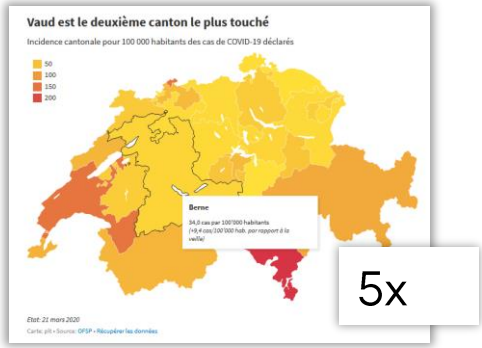


3x

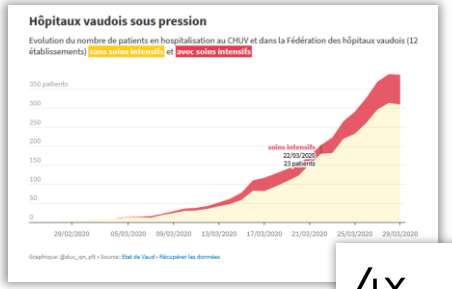
24 heures (regional)



29x



5x



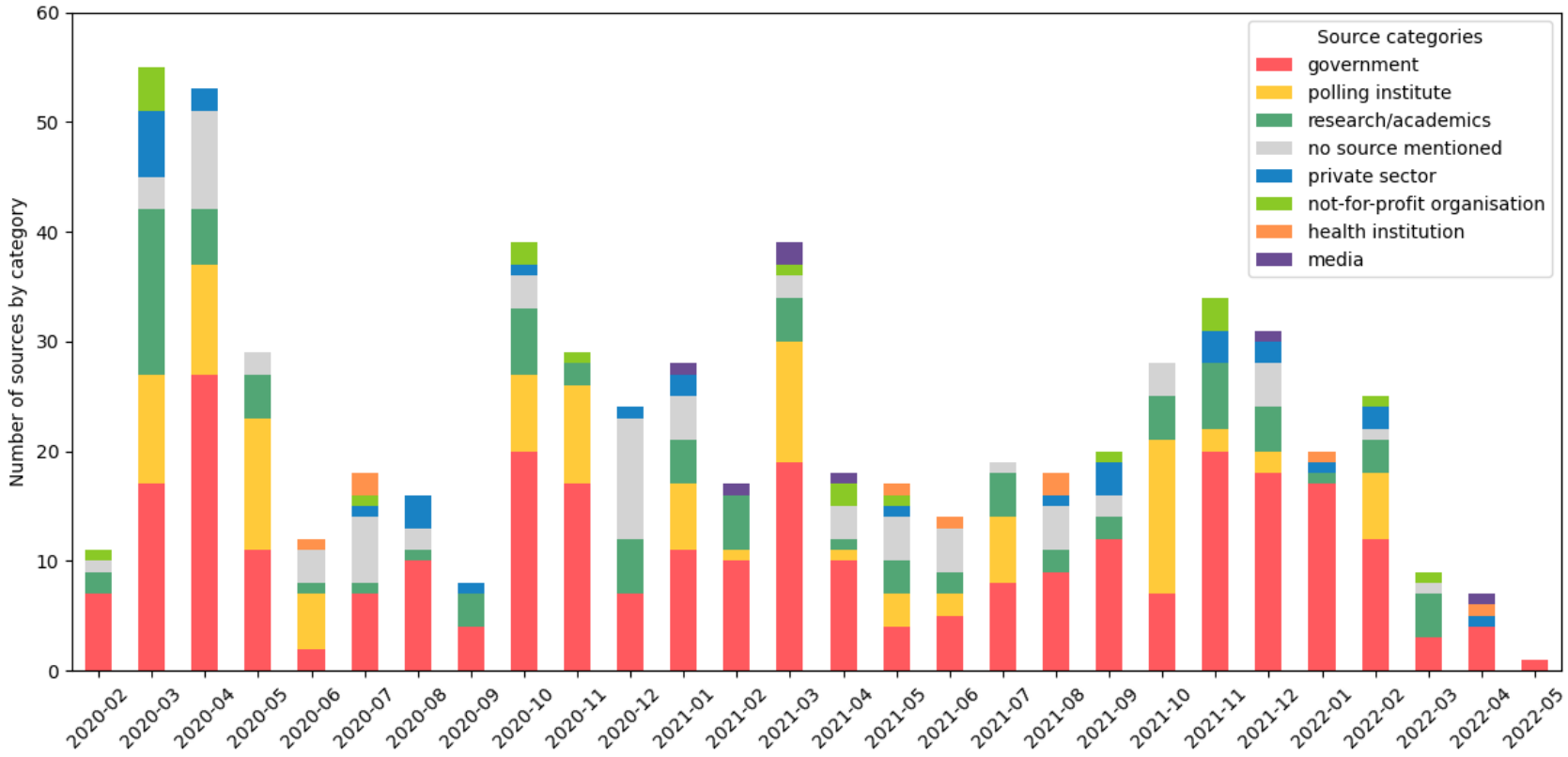
4x

RQ 2

What kind of sources were used/cited
for these data visualizations and did their use vary
according to media or over time?

Results: government data dominates

Breakdown of categories of sources
(as stated by text within the embedded code)



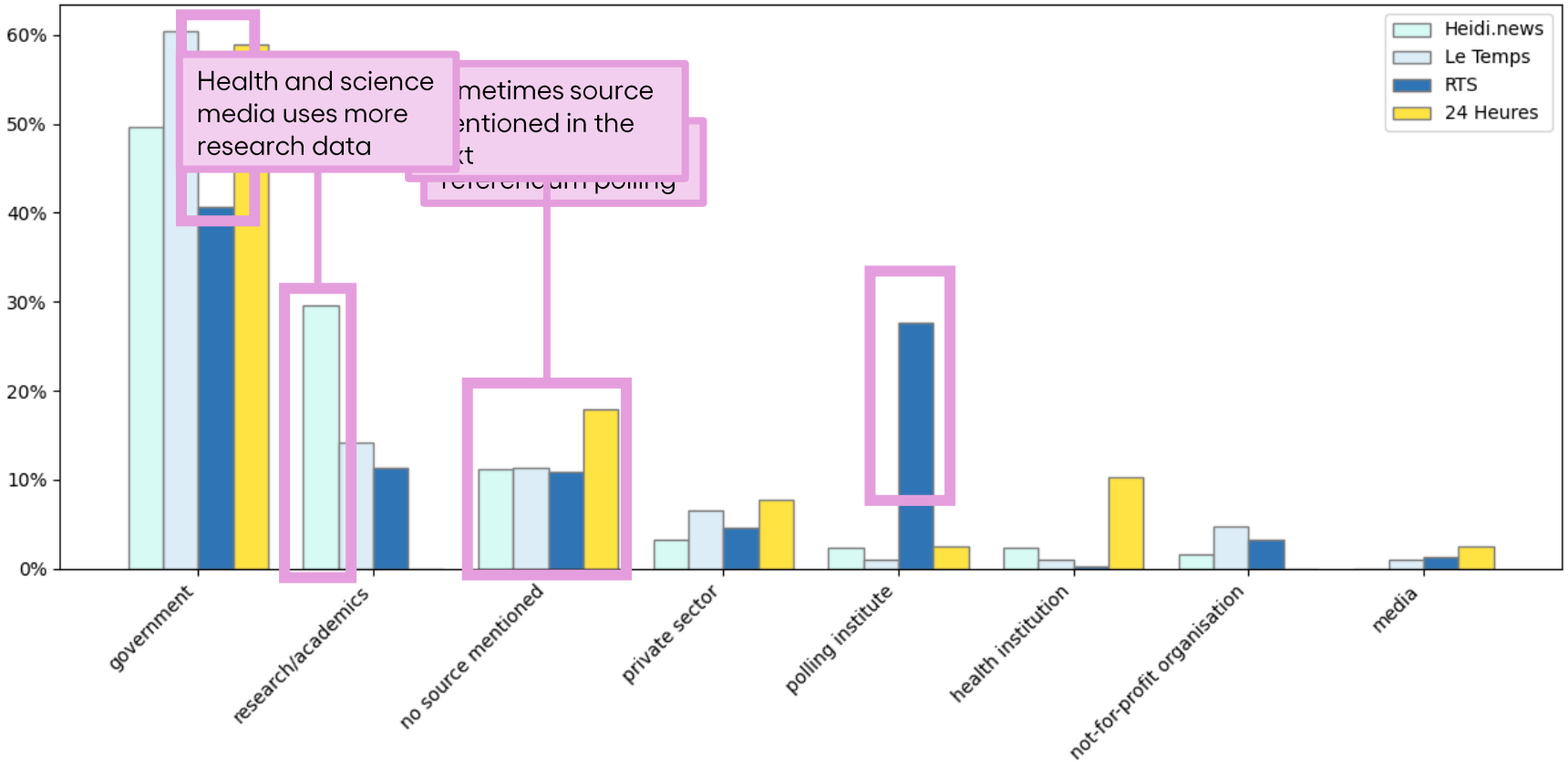
Covid-19 data journalism
Data visualizations:
from Source Diversity
to Source Reflexivity

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

Results: media reflect their editorial DNA

Breakdown of categories of sources according to different media
(sources as stated by text within the embedded code)



Covid-19 data journalism
Data visualizations:
from Source Diversity
to Source Reflexivity

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

Results: sources

	RTS 348 unique dataviz	Le Temps 102 unique dataviz	Heidi News 124 unique dataviz	24 Heures 34 unique dataviz
Number of different sources	58 sources	25 sources	33 sources	16 sources
Top 5 (unique)	1. Sotomo 2. OFSP 3. OFS 4. OWID 5. gfs.bern 67% of the total	1. OFSP 2. OWID 3. OFS 4. OMS 5. UE 74% of the total	1. OFSP 2. OWID 3. Johns Hopkins 4. OFS 5. Seco 67% of the total	1. OFSP 2. Etat de Vaud 3. CHUV 4. Cantons 5. Canton de Vaud 62% of the total
Top 5 (range)	1. OFSP 2. OWID 3. Sotomo 4. OFS 5. gfs.bern 62% of the total	1. OFSP 2. OFS 3. OWID 4. OMS 5. UE 96% of the total	1. OFSP 2. OWID 3. Johns Hopkins 4. OFS 5. Seco 67% of the total	1. OFSP 2. Cantons 3. Etat de Vaud 4. CHUV 5. Canton de Vaud 86% of the total

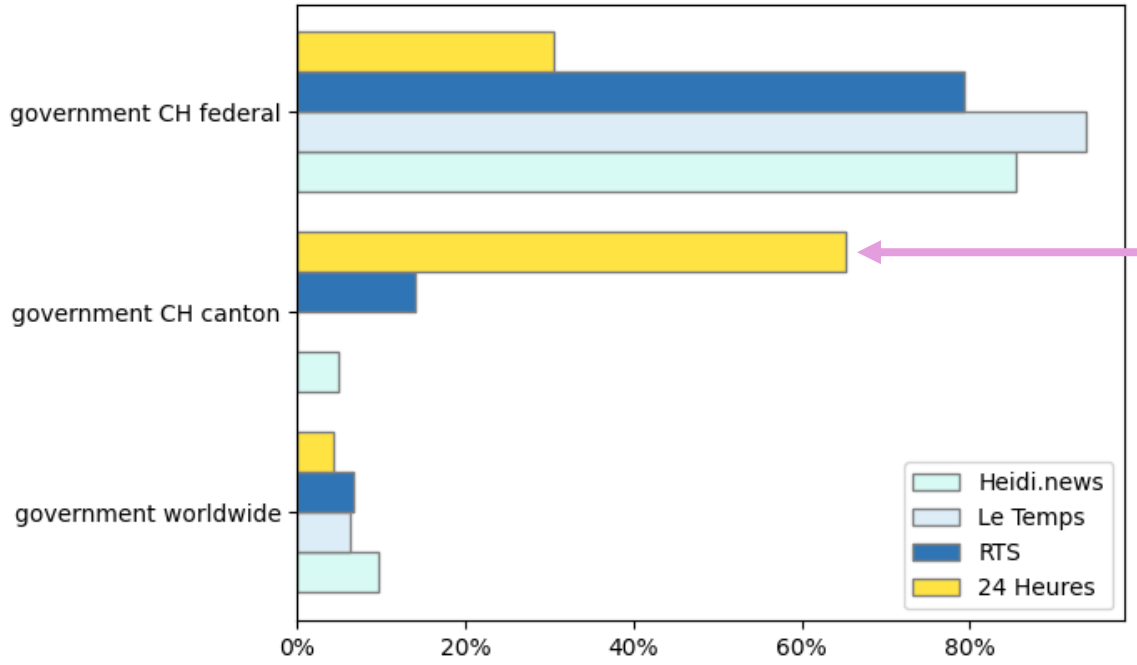
Covid-19 data journalism
Data visualizations:
from Source Diversity
to Source Reflexivity

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

Results: media reflect their editorial DNA

Breakdown of categories of sources sources according to different media
(as stated by text within the embedded code)



Regional media used data from regional government....

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

RQ 3

Did the data visualizations include text that presented the source critically or reflexively?

Classification system for title, subtitle and notes

Category	Definition	Example
Descriptive	Description of what is represented in the data visualization	Number of cases per 10'000 persons in the last 7 days.
Date/update	Information related to when the data was collected, analyzed or updated	*Last updated: 17/05/2022*
Interaction	Information about interactive feature and instructions about how to use them	*Click or move your finger on the screen to see the number of tests per thousand people during a given week*
Question	Questions asked about the data or questions asked to respondents the survey for which the data is provided in the data visualization (similar to description)	*How do you evaluate the level of risk you are exposed to?*
Interpretation of the data	Interpretation of what the data reveals (= framing or angle)	*The number of deaths recorder between 18 and 15 October increased to reach a level that had not been reached since late April*
Details about legends, units or method	Information about visual choices, explanations about mathematical method	*The red area represents the number of deaths attributed to Covid-19. The grey zone represents the unexplained excess deaths*
Details about the data and its status	Simple factual information about the source data	*Data for the week between 29 March and 4 April are currently incomplete*
Details about the data that includes metadiscourse	Detailed data about the source data that includes a warning and/or a more in-depth explanation or contextualization of the data Note: this is different from the interpretation category above since it explains how method impacts or distorts reality rather than how one reality or phenomena is related to another	*The spike in new detected cases is probably in part imputable to a change in counting method and/or the less restrictive screening policy than previously*

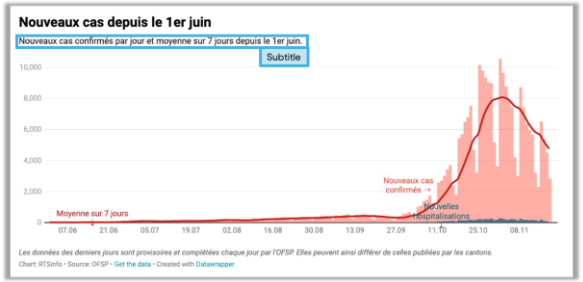
Classification system for title, subtitle and notes

Category	Definition	Example
Descriptive	Description of what is represented in the data visualization	Number of cases per 10'000 persons in the last 7 days.
Date/update	Information related to when the data was collected, analyzed or updated	*Last updated: 17/05/2022*
Interaction	Information about interactive feature and instructions about how to use them	*Click or move your finger on the screen to see the number of tests per thousand people during a given week*
Question	Questions asked about the data or questions asked to respondents the survey for which the data is provided in e data visualization (similar to description)	*How do you evaluate the level of risk you are exposed to?*
	terpretation of what the data reveals framing or angle)	*The number of deaths recorder between 18 and 15 October increased to reach a level that had not been reached since late April*
Details about legends, units or method	Information about visual choices, explanations about mathematical method	*The red area represents the number of deaths attributed to Covid-19. The grey zone represents the unexplained excess deaths*
Details about the data and its status	Simple factual information about the source data	*Data for the week between 29 March and 4 April are currently incomplete*
Details about the data that includes metadiscourse	Detailed data about the source data that includes a warning and/or a more in-depth explanation or contextualization of the data Note: this is different from the interpretation category above since it explains how method impacts or distorts reality rather than how one reality or phenomena is related to another	*The spike in new detected cases is probably in part imputable to a change in counting method and/or the less restrictive screening policy than previously*

Categories related to source and data metadiscourse

Results: subtitle mostly describes the dataviz & variables

Category	Definition
Descriptive	Description of what is represented in the data visualization
Date/update	Information related to when the data was collected, analyzed or updated
Interaction	Information about interactive feature and instructions about how to use them
Question	Questions asked about the data or questions asked to respondents the survey for which the data is provided in the data visualization (similar to description)
Interpretation of the data	Interpretation of what the data reveals (= framing or angle)
Details about legends, units or method	Information about visual choices, explanations about mathematical method
Details about the data and its status	Simple factual information about the source data
Details about the data that includes metadiscourse	Detailed data about the source data that includes a warning and/or a more in-depth explanation or contextualization of the data Note: this is different from the interpretation category above since it explains how method impacts or distorts reality rather than how one reality or phenomena is related to another



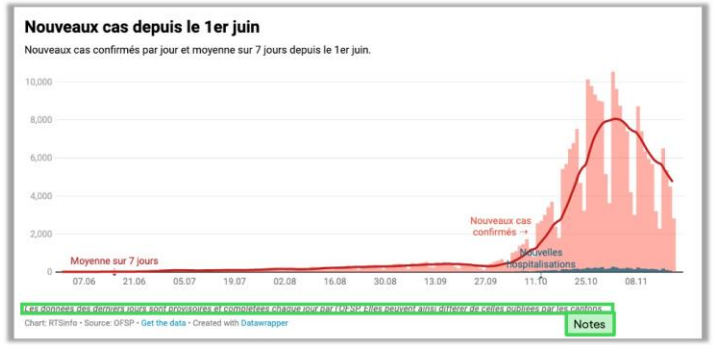
When a subtitle is included (63% of dataviz), it is mostly **descriptive**.

The second most common type of information in subtitles is information related to the date or last update of the data.

This was common to all media and may be explained by the fact that much Covid-19 coverage involved updating or monitoring evolving data.

Results: notes mostly provide basic details about the data

Category	Definition
Descriptive	Description of what is represented in the data visualization
Date/update	Information related to when the data was collected, analyzed or updated
Interaction	Information about interactive feature and instructions about how to use them
Question	Questions asked about the data or questions asked to respondents the survey for which the data is provided in the data visualization (similar to description)
Interpretation of the data	Interpretation of what the data reveals (= framing or angle)
Details about legends, units or method	Information about visual choices, explanations about mathematical method
Details about the data and its status	Simple factual information about the source data
Details about the data that includes metadiscourse	Detailed data about the source data that includes a warning and/or a more in-depth explanation or contextualization of the data Note: this is different from the interpretation category above since it explains how method impacts or distorts reality rather than how one reality or phenomena is related to another



When a note is included (28% of dataviz), it mostly provides further specific **details about the data**.

Metadiscourse explaining the limits of the dataset or possible impact on interpretation is very rare.

Results

RQ 1: *How (often) did different Swiss news media use data visualizations in their online news articles, and what types or data visualizations were used?*

- Data visualizations appeared in 8.2% of news articles dedicated to Covid-19
- Many new data visualizations were created at the beginning of the first waves
- The largest media published the most dataviz and rused them most often
- Overall, beyond a small number of multiple use data visualizations, approx 80% were single use
- Maps and timelines of cases, hospitalizations and deaths (and later vaccinations) were most common, and often rused

Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

Results

RQ2: *What kind of sources were used for these data visualizations and did their use vary according to media and over time?*

- The source of the data was mentioned in most cases (85-90%)
- Government data was by far the most commonly used source for Covid-19 related data visualizations (40-60%)
- Beyond a slightly chaotic situation in the first weeks, the breakdown of sources used did not seem to evolve all that much over time
- Links to the full original dataset were rare
- Data sources did reflect the editorial DNA of the different media, albeit mildly
    - Health and science media used more academic data sources
    - The national boradcaster conducted a lot of polling, often related to upcoming referenda
    - Local media used more local government data

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

Results

RQ3 *Did the data visualizations include text that presented the source critically or reflexively?*

- The text accompanying data visualizations (title, subtitle, notes) contained almost no information qualifying the data, presenting the source critically or demonstrating reflexivity (there were some dedicated articles)
- Most information was descriptive, presenting the variables and the data
- When the text did qualify the data (28%), it mostly related to its timeliness (i.e. how up to date the data was

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

Conclusion

RQ1 The first waves of the pandemic gave rise to the creation of numerous data visualizations

RQ1 The key health related data (cases, hospitalizations, deaths) was widely used and reused in data visualizations especially in the form of maps and timelines.

RQ1 We recorded some differences between the media studied, but also many similarities. Overall, these differences reflected the DNA of these media.

RQ2 The reliance on government data reflects previous research (it is in fact even more pronounced for the COVID-19 pandemic than for data journalism in general as identified in previous studies)

RQ3 There was little reflexive or critical metadiscourse qualifying the source and data

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

Conclusion

While the use of government data is not in itself problematic (when used, it is often the best available data), in view of existing literature we believe that:

1. Journalists have a duty to critically engage with this data (as with all other data)
2. metadiscourse related to this reflexive positioning likely has a rhetorical function, especially when it comes to building trust
3. There remain many opportunities for data stories that do not rely on government data, that might shine a different light on aspects of the pandemic that government data cannot account for (although these are often time consuming and costly given the current state of the media industry)

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

Limitations and outlook

No in-depth qualitative analysis of the dataviz

→ Each dataviz considered «equal»

Study of how (data)journalists present source reflexively to their audiences

→ Says little about underlying logics and reflexivity (or lack of)

Disconnected from the article's body text

→ Possible critical source metadiscourse within the text instead of the dataviz (likely very rare)

Effects not measured

→ Says nothing about reception (audience side)

A second more qualitative study?

Follow-up interview research with the journalists who created the dataviz

Ongoing experimental audience research

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel

References

Cited

Borges-Rey, Eddy. “Unravelling Data Journalism: A Study of Data Journalism Practice in British Newsrooms.” In *The Future of Journalism: Risks, Threats and Opportunities*, 170–80. Routledge, 2020.

Lowrey, Wilson, Ryan Broussard, and Lindsey A. Sherrill. “Data Journalism and Black-Boxed Data Sets.” *Newspaper Research Journal* 40, no. 1 (March 2019): 69–82. https://doi.org/10.1177/0739532918814451.

De Haan, Yael, Sanne Kruikemeier, Sophie Lecheler, Gerard Smit, and Renee Van Der Nat. “When Does an Infographic Say More Than a Thousand Words?: Audience Evaluations of News Visualizations.” *Journalism Studies* 19, no. 9 (July 4, 2018): 1293–1312. https://doi.org/10.1080/1461670X.2016.1267592.

Focacci, C.N., P.H. Lam, and Y. Bai. “Choosing the Right COVID-19 Indicator: Crude Mortality, Case Fatality, and Infection Fatality Rates Influence Policy Preferences, Behaviour, and Understanding.” *Humanities and Social Sciences Communications* 9, no. 1 (2022). https://doi.org/10.1057/s41599-021-01032-0.

Mellado, Claudia, Daniel Hallin, Luis Cárcamo, Rodrigo Alfaro, Daniel Jackson, María Luisa Humanes, Mireya Márquez-Ramírez, et al. 2021. “Sourcing Pandemic News: A Cross-National Computational Analysis of Mainstream Media Coverage of COVID-19 on Facebook, Twitter, and Instagram.” *Digital Journalism* 9 (9): 1261–85. https://doi.org/10.1080/21670811.2021.1942114.

Stalph, Florian. “Classifying Data Journalism.” *Journalism Practice* 12, no. 10 (November 26, 2018): 1332–50. https://doi.org/10.1080/17512786.2017.1386583.

Tong, Jingrong, ed. *Data Journalism and the COVID-19 Disruption*. Abingdon, Oxon: Routledge, 2024.

Tong, Jingrong. “From Content to Context: A Qualitative Case Study of Factors Influencing Audience Perception of the Trustworthiness of COVID-19 Data Visualisations in UK Newspaper Coverage.” *Journalism*, July 27, 2023, 14648849231190725. https://doi.org/10.1177/14648849231190725.

Ware, Colin. *Information Visualization: Perception for Design*. 2. ed., [Nachdr.]. The Morgan Kaufmann Series in Interactive Technologies. Amsterdam Heidelberg: Elsevier/Morgan Kaufmann, 2004.

Other key references

Hallin, Daniel C., Claudia Mellado, Akiba Cohen, Nicolas Hubé, David Nolan, Gabriella Szabó, Yasser Abuali, et al. 2023. “Journalistic Role Performance in Times of COVID.” Journalism Studies 24 (16): 1977–98. https://doi.org/10.1080/1461670X.2023.2274584.

Parasie, Sylvain. *Computing the News Data: Journalism and the Search for Objectivity*. New York: Columbia University Press, 2022.

Ramsälv, Amanda, Mats Ekström, and Oscar Westlund. 2023. “The Epistemologies of Data Journalism.” New Media & Society, January, 146144482211504. https://doi.org/10.1177/14614448221150439.

Tandoc, Edson C., and Soo-Kwang Oh. “Small Departures, Big Continuities?: Norms, Values, and Routines in *The Guardian*’s Big Data Journalism.” *Journalism Studies* 18, no. 8 (August 3, 2017): 997–1015. https://doi.org/10.1080/1461670X.2015.1104260.

Zamith, Rodrigo. “Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism.” *Digital Journalism* 7, no. 4 (2019): 470–89. https://doi.org/10.1080/21670811.2018.1554409.

Covid-19 data journalism
Data visualizations:
from Source Diversity
to Source Reflexivity

unine
Université de Neuchâtel
Académie du journalisme
et des médias

Andrew Robotham
Céline Dupuis
Nathalie Pignard-Cheynel