Summary sheet: Research categories

Research category 1: Funding model

Classification level 1	Classification level 2
Crossborder journalist networks	Global network
	International/national/local collaborations
	Financial sources: Foundation grants, private donations, public funds
	Revenues: non-profit organization
Number of informative outputs	Topics visibility (free circulation of contents)
Competitiveness	Monetization of the story and sustainability
Supply and demand	Not paid for as a service for readers
Metrics	User traffic
Impact	Level: micro (changes in awareness, attitudes, and behaviours) / meso (changes at group level: collective opinions) / macro (changes at the level of institutions or power holders)
	Type: Civic / Symbolic / Deliberative
Reputation	Awards, prizes

Research category 2: Work practices

Classification level 1	Classification level 2
Organizational form	Intra/inter-newsrooms
	Intertwined newsrooms
	Hubs
	Clusters
Professional conditions	Quality of the journalistic coverage / High quality journalism
	Global mindset
	International perspective
	Multicultural setting
	Cross-national solidarity
	Common ethical standard
	Cooperation (individual ends) versus collaboration (common ends)
	Win-win attitude
	Trust in each other interactions
Skills and working methods	Neutral intermediaries / host / editorial coordinator
	Projectification (project-based journalism)
	Team building process
	Inspiration
	Exclusivity
	Specialisation
	Mutual assistance
	Expansion of research capacity
	Transfer of knowledge and knowhow
	Pooling of resources
	Data skills
	Access to local expert
	Access to information (FOI)
	Collaboration on similar topics (join content and storytelling)
	Synchronization (issue salience)
	Capacity of replication

Research category 3: Thematic diversity

Classification level 1	Classification level 2
Topics	Overarching long-term themes
	Policy Agenda Area

	Corruption, conflict of interest, fraud, mislead, nepotism, influence peddling, favouritism, abuse, harassment, misconduct, discrimination, misuse
	Politicians, corporations, organizational criminals
Reasons to cover a topic	Accountability
	Expose wrongdoing (stories of malfeasance and mismanagement)

Research category 4: Connectors

Classification level 1	Classification level 2
Constellation of actors	Editors, journalists / freelancers, sources, data scientists, coders, designers, photographers, fact-checkers, competitors, publics, funders, public institutions, NGO's, advocacy groups, activists, intermediary organizations, facilitators, interlopers, technical platforms, etc.
Digital resources	Collaboration software (managing), digital security software, data extraction software, data analysis software, algorithms, code, tools, encryption software (Virtual Private Networks, Tor browser, PGP-encrypted emails), defensive and offensive technology
Shared resources	Datasets, leaks, researching raw material, crowdsources databases, open-source tools
Infrastructures	Material platforms and locations
	Online meeting arenas
	Offline meeting arenas
Networks	Non-hierarchal decentralized research network
	Alfter (2018): Loose network; one-off assistance; loose collaboration; close collaboration
	Houston (2021): Temporary and separate; temporary and cocreating; temporary and integrated; ongoing and separate; ongoing and cocreating; ongoing and integrated
	Heft et al. (2019): High profile collaboration; low-intensity cooperation
	Jenkins and Graves (2022): Co-op model; contractor model; NGO model
	Krüger et al. (2019): Organizations that organized or co-organized (at least once) one of the two most important conferences in the field; International organizations or projects producing investigative journalism; International organizations or projects supporting investigative journalism by providing financial or other resources; National organizations or projects producing investigative journalism; National organizations or projects supporting investigative journalism by providing financial or other resources
	Flow of work: bottom-up / top-down initiatives (Heft et al., 2019)

Words with a repetition rate ≥ 20

RC / Words	Suma Freq
RC1	**251**
funding	42
funds	50
grant	20
grants	20
impact	35
model	28
models	20
nonprofit	36
RC2	**11967**
able	24
access	97
address	29
algorithms	38
analyse	40
analysis	83
approach	29
available	69
basic	56
basics	33
build	54
challenge	20
challenges	43
clean	29
common	36
communication	28
course	22
cover	48
covering	25
create	32
data	1079
database	93
databases	43
data-driven	22
dataset	40
datasets	66
develop	31
developed	24
different	119
discuss	72
document	31
documents	111
easy	30
editorial	22
evidence	28
examples	68

excel	59
experience	51
experienced	28
experiences	45
experts	31
explain	50
explore	45
files	48
find	145
finding	35
findings	35
FOI	62
found	40
free	30
freedom	53
get	225
graph	21
hack	23
hard	25
help	98
identify	27
independent	28
information	222
interactive	27
interest	58
interested	29
investigate	86
investigated	25
investigating	51
investigation	171
investigations	103
investigative	271
journalism	408
journalistic	71
know	80
knowledge	37
language	38
law	26
leaks	36
learn	192
learned	20
learning	28
legal	44
legislation	39
life	27
line	23
look	82
looking	32

used	135
useful	33
using	142
way	79
ways	34
websites	23
wobbing	29
work	242
working	105
works	34
write	21
years	83
RC3	**4168**
activities	20
authorities	35
billion	20
building	24
business	69
cities	23
climate	73
commission	26
companies	89
company	27
complex	38
concepts	21
corporate	41
corruption	78
council	21
court	37
covid	21
crime	57
crisis	26
development	21
economic	20
emissions	24
energy	21
environment	26
environmental	26
EU	289
Europe	228
European	291
Europe's	24
euros	28
EU's	31
field	43
finance	26
financial	36
focus	47

food	44
fraud	33
government	33
governments	25
health	43
housing	60
human	31
ideas	40
illegal	20
important	59
industry	61
influence	20
institutions	38
interesting	21
investment	20
involved	27
issue	22
issues	33
labour	36
lobby	24
lobbying	34
market	34
medical	25
medicines	21
million	20
money	115
new	154
offshore	26
Panama	20
papers	35
parties	20
people	116
policy	21
political	40
politicians	28
power	52
powerful	41
prices	22
public	187
relevant	28
right	76
rights	22
science	22
secret	31
state	36
states	33
subsidies	22
surveillance	21

tax	66
tender	24
tenders	29
topic	28
topics	34
trade	23
war	23
water	22
RC4	**4827**
across	109
around	59
audience	26
borders	39
citizens	37
code	26
coding	25
collaborate	30
collaboration	77
collaborations	49
collaborative	49
colleagues	36
community	27
consortium	36
cooperation	21
Correctiv	29
countries	128
country	70
crossborder	195
Dataharvest	36
digital	94
editor	27
editors	21
file	21
global	38
Google	33
group	26
groups	40
including	50
international	116
internet	28
join	65
journalist	81
journalists	519
level	50
local	105
media	176
meet	97
member	35

members	32
national	97
network	110
networks	63
newsrooms	50
online	88
open	99
organisations	52
organized	35
others	39
participants	38
partners	32
place	21
platform	38
platforms	20
readers	24
regional	23
reporters	54
session	231
share	100
shared	24
sharing	22
social	55
software	38
source	34
sources	92
structures	24
system	45
systems	34
tech	25
technology	21
together	87
us	94
users	23
web	65
whistleblowers	24
workshop	71
world	77
Total	**21213**

Centroid documents

Group 0

"Europe's big uncovered follow-the-money story: Public spending amounts for vast parts of our economy, large parts of the contracts can be traced through the EU's and national tender portals, however the field is hardly covered journalistically yet. Adriana Homolova and her colleagues from The Investigative Desk (De Onderzoeksredactie) are focusing on the IT spending of Dutch authorities. She will show how to make and read spending networks with Open TED data for your own country. Another option is to follow the EU-tender legislations' product codes, so-called CPVs, to look into EU structures applicable in all EU countries. Learn about the data structures and join the network of interested journalists".

Group 1

"Get started with R: Intro & importing data: R is a statistical coding language which has become one of the key tools for journalists. It's great at handling large datasets, and many news organisations are using R now, especially for visualisation. We will talk about why R is a good language to learn, how to get started, how to import the data you need and start working on it".

Group 2

"Get started with data journalism: Data journalism has been the posterchild of progressive investigations for the past decade in Europe and Dataharvest has been its herald. In 2019, we offer a one-day data journalism masterclass for seasoned reporters and editors, who have yet to integrate data discovery, negotiation, analysis and transparency in their storytelling and reporting toolkit. In a step-by-step workshop, we'll piece together how data can take your reporting far and above anecdotal evidence; how to source information and data from various jurisdictions; and how to analyse and summarise it for original and versatile reporting".

Group 3

"Personal data wobbing on the web: Through a process similar to FOIA, any European citizen can ask for their personal data from any public or private institution. This is very powerful, and could be very helpful for investigative journalism efforts around algorithmic bias/surveillance/privacy. Paul-Olivier Dehaye will give some examples of use in different fields: programmatic advertising, keystroke fingerprinting, Facebook newsfeed curation, etc.".

Full results analysis available at
https://drive.google.com/drive/folders/1e3gvrZkPjGr2bG1p1Vc4sDVDTI4PSwo3?usp=drive_link