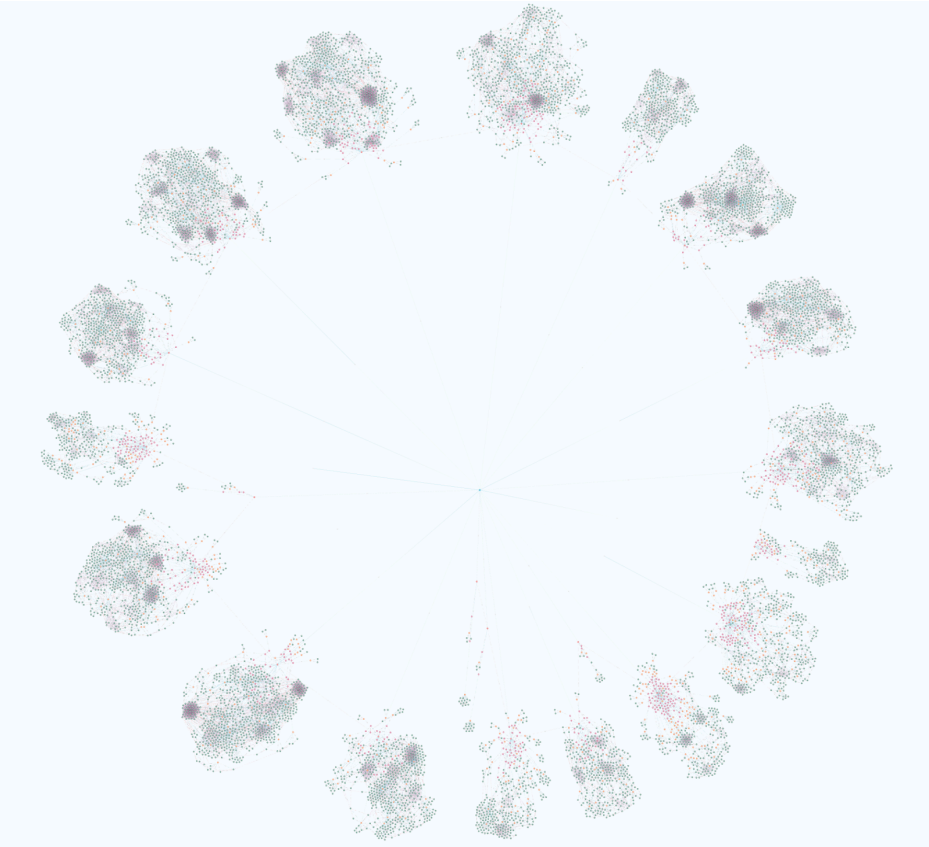
Testrapport CirculairNLP



by
Axel Frederiks (2202630)
Jonah Siemers (2202907)
Menno Rompelberg (2204333)
Oscar Theelen (2204154)

19/04/2024

Supervisors

Pascal Slaats
Jeroen Baijens
Shannen Dolls

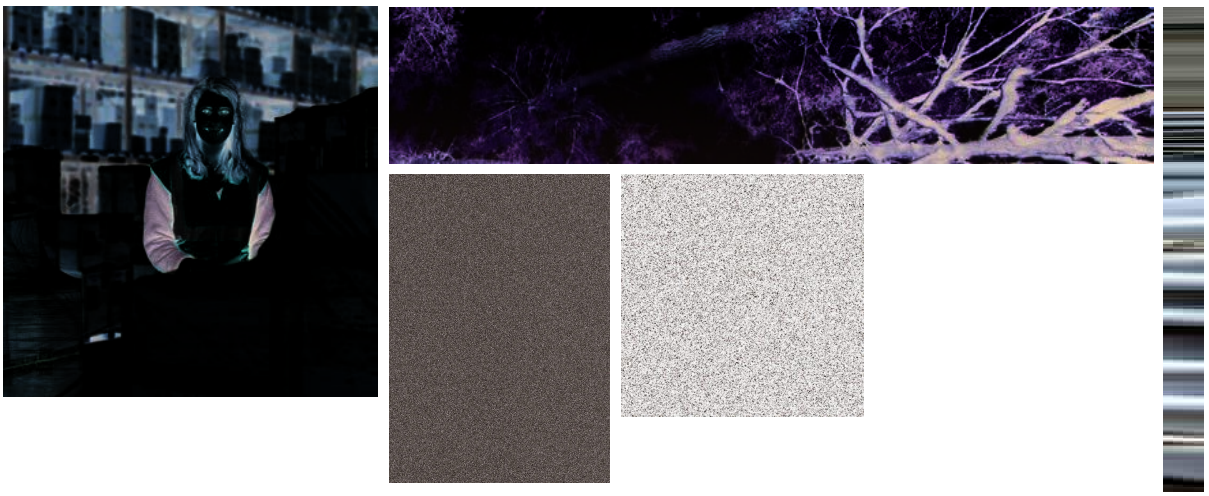
Voor het testen hebben we 6 pdf documenten gebruikt voor de volledige tests.Van deze documenten zijn er 4 aangeleverd door de opdrachtgever.

1 File Data Extraction

1.1 Images

Files	# images extracted	# images garbage	# correct
Allianz	0	0	-
CMA	170	57	33.5%
EFRAG	49	49[1]	0%
DPEF	75	33	44%
VDL	38	4	10.5%
NASDAQ	167	146	87.4%

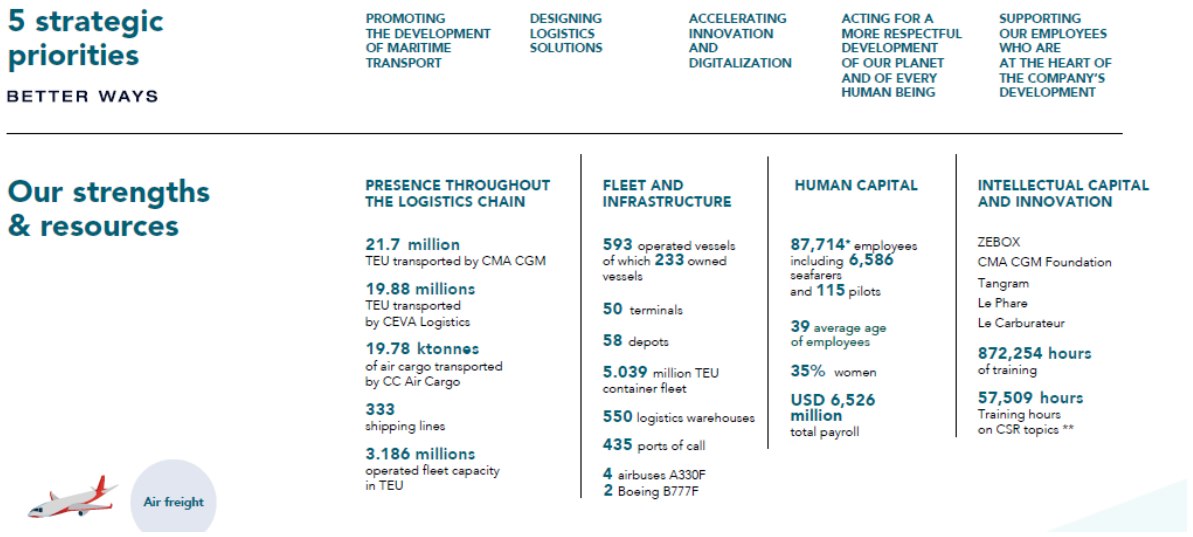
Garbage examples:



[1] 44 van de 49 corrupted images zijn onleesbaar, omdat het .jpx-files zijn.

1.2 Tabellen

Tabellen werken met OCR en hebben dus een grote error marge.
Tabellen hebben geen eenduidige definitie. Om die reden worden alleen tabellen die duidelijk tabellen horen te zijn geteld onder wanted tables. Het volgende wordt bijvoorbeeld niet geteld als tabel:



Voor deze test is alleen CMA handmatig gecontroleerd aangezien dit bestand al 2 uur duurde.

Files	gemiste tabellen	slechte tabellen	correcte tabellen	deels correcte tabellen
CMA	3	6	7	8

CMA:

<table>
<tr><td>Garbage</td><td>

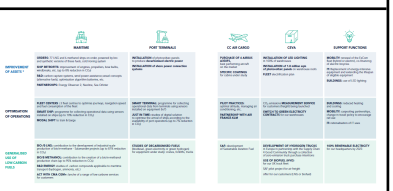
	Col0	Frei forwa
0	R	
1		oad
2		

</td><td>
</td></tr>
<tr><td>Garbage</td><td>

	Col0	Col1	Protecting employees Energy consumption and the emission of greenhouse gases Marine and on-shore pollution Air quality Competitive environment and antitrust laws
0		Diversity and inclusion Human rights throughout the value chain Marine and inland biodiversity Personal data Products and services with sustainable added value Sustainable supply chain Economic embargoes and sanctions Transport of sensitive and/or dangerous goods Business ethics and integrity	
1	Local communities Circular economy with our stakeholders End of life of equipment Entrepreneurial culture and innovation		

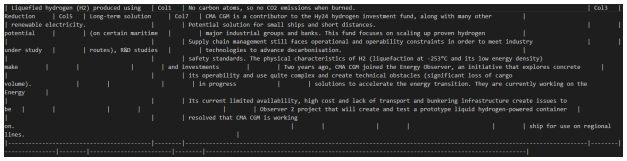
</td><td>
</td></tr>
<tr><td>Garbage</td><td>

2019	2022	2023	2024	Col4	2025	2030	2050

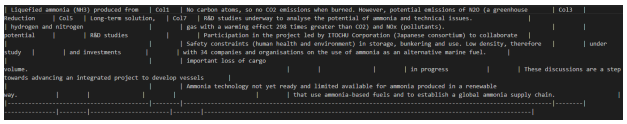
</td><td> </td></tr>
<tr><td>Missed</td><td>-</td><td>  </td></tr>
</table>

Garbage		**02. ASSESSMENT PROCEDURES** In 2022, for the third consecutive year, the Group was awarded the ECOVADIS "Platinum" medal, rewarding its CSR commitment and duty of vigilance actions. To maintain this status, the Group has obtained various international certifications (ISO 9001, ISO 14001 and ISO 45001), accelerated the evaluation of its suppliers and strengthened supplier risk management through the Group's Accelerate programme.
Correct		
Correct		
Correct		
Correct		
Correct		

Correct		
Correct		
Garbage		
Garbage		-
Partly		



	Col0	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
0	Liquefied hydrogen (H2) produced using renewable electricity.	Col1	No carbon atoms, so no CO2 emissions when burned.	Col2	CMA CGM is a contributor to the H2V hydrogen investment fund, along with many other (on certain maritime routes), R&D studies under study	Col3	Long-term solution, Potential solution for small ships and short distances.	Col4	Observer 2 project that will create and test a prototype liquid hydrogen-powered container ship for use on regional lines.
1					major industrial groups and banks. This fund focuses on scaling up proven hydrogen				
2					Supply chain management still faces operational and operability constraints in order to meet industry				
3					technologies to advance decarbonisation.				
4					safety standards. The physical characteristics of H2 (liquefaction at -253°C and its low energy density) make and investments		Two years ago, CMA CGM joined the Energy Observer, an initiative that explores concrete		
5					its operability and use quite complex and create technical obstacles (significant loss of cargo volume).		In progress		solutions to accelerate the energy transition. They are currently working on the Energy
6					Its current limited availability, high cost and lack of transport and bunkering infrastructure create issues to be				
7					resolved that CMA CGM is working on.				



	Col0	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8
0	Liquefied ammonia (NH3) produced from hydrogen and nitrogen	Col1	No carbon atoms, so no CO2 emissions when burned. However, potential emissions of N2O (a greenhouse gas with a warming effect 298 times greater than CO2) and NOx (pollutants).	Col2	R&D studies underway to analyse the potential of ammonia and technical issues.				
1					Participation in the project led by ITOCHU Corporation (Japanese consortium) to collaborate with 34 companies and organisations on the use of ammonia as an alternative marine fuel.				
2					Safety constraints (human health and environment) in storage, bunkering and use. Low density, therefore important loss of cargo volume.				
3									
4									
5									

Partly

	INDICATOR	UNIT	SCOPE	2022	2021	OBJECTIVE
0						
1	Maritime activities 100 100					
	Ports & Terminals					
	35 23					
	ISO 9001 Quality Certification % activities					
	CEVA inland and					
	77 77					
	logistics activities					
2	Maritime activities 100 100					
	Ports & Terminals					
	35 29					
	ISO 14001 Environmental Certification % activities					
	CEVA inland and					
	51 51					
	logistics activities					
3	Ports & Terminals					
	30 16					
	activities					
	ISO 45001 Health and Safety Certification %					
	CEVA inland and					
	51 46					
	logistics activities					
4	Maritime activities 5 4					
	Major incidents Ports & Terminals					
	Number 0 3					
	activities					
	Depots 0 3					
5	Rescue operations Number CMA CGM 12 7					
6	Promoting education and training					

INDICATOR	UNIT	SCOPE	2022	2021	OBJECTIVE
		Maritime activities	100	100	
ISO 9001 Quality Certification	%	Ports & Terminals activities	35	23	
		CEVA inland and logistics activities	77	77	
		Maritime activities	100	100	
ISO 14001 Environmental Certification	%	Ports & Terminals activities	35	29	
		CEVA inland and logistics activities	51	51	
ISO 45001 Health and Safety Certification	%	Ports & Terminals activities	30	16	
		CEVA inland and logistics activities	51	46	
		Maritime activities	5	4	
Major incidents	Number	Ports & Terminals activities	0	3	
		Depots	0	3	
Rescue operations	Number	CMA CGM	12	7	
Promoting education and training					
Number of training days per employee	Day		1.46	1.55	3 days by 2023
Number of hours of training (total)			872,259	787,988	
Number of hours of training - e-learning	Hours	CMA CGM CEVA Logistics	295,027	328,169	
Number of hours of training - in person			577,232	459,819	
Number of hours of CSR training*			57,509	First year reported	1 day by 2023
Average length of service	Number	CMA CGM	6.1	6	
Number of International Corporate Volunteer young talents	Number	CMA CGM	63	50	
Number of apprentice young talents	Number	CMA CGM	13	23	
Promoting diversity and inclusion					
Non-discrimination policy	Document	CMA CGM CEVA Logistics	Yes	Yes	
Ratio of women to men	%	CMA CGM CEVA Logistic	35/65	37/63	
Gender equality index	-	CMA CGM	92/100	84/100	
		CEVA Logistics	84/100	79/100	

Partly

Promoting diversity and inclusion	Col1	Col2	Col3	Col4	Col5

Partly

	INDICATOR	UNIT	SCOPE	2022	2021	OBJECTIVE
0						
1	CMA CGM					
	Share of women in management % 32 32					
	CEVA Logistics					
2	Share of women in the Top 200 % CMA CGM 11.5 First year 30% by 2025					
	by remuneration CEVA Logistics reported					
3	CMA CGM					
	Share of women on the Board of Directors % 42 42					
	CEVA Logistics					
4	Share of non-French nationals % CMA CGM 58 First year 61% by 2023					
	in the Top 200 by remuneration CEVA Logistics reported					
5	CMA CGM					
	Average age of employees Age 39 39					
	CEVA Logistics					
6	Intergenerational ties - Distribution					
	of staff by age group					
	Under 20 0.3 0.2					
	20 to 29 24.5 21.0					
	CMA CGM					
	30 to 39 % CEVA Logistics 31.5 25.0					
	40 to 49 24.3 28.0					
	50 to 59 14.6 20.8					
	60 and over 4.8 5.0					
7	Cultural diversity - Distribution of staff					
	by geographical area					
	Europe / France 31.1 33.0					
	CMA CGM					
	%					
	Africa/Middle East CEVA Logistics 11.8 13.0					
	Americas 24.5 27.0					
	Asia/Oceania 32.5 27.0					
8	Supporting local communities					
9	100% local					
	Percentage of countries that have launched					
	% CMA CGM 94 74 initiatives					
	at least one local initiative					
	by 2023					
10	«Containers of Hope» Number					
	CMA CGM 926 201					
	humanitarian aid delivered					
11	Ensuring respect for human rights					

INDICATOR	UNIT	SCOPE	2022	2021	OBJECTIVE
Share of women in management	%	CMA CGM CEVA Logistics	32	32	
Share of women in the Top 200 by remuneration	%	CMA CGM CEVA Logistics	11.5	First year reported	30% by 2025
Share of women on the Board of Directors	%	CMA CGM CEVA Logistics	42	42	
Share of non-French nationals in the Top 200 by remuneration	%	CMA CGM CEVA Logistics	58	First year reported	61% by 2023
Average age of employees	Age	CMA CGM CEVA Logistics	39	39	
Intergenerational ties - Distribution of staff by age group					
Under 20			0.3	0.2	
20 to 29			24.5	21.0	
30 to 39	%	CMA CGM CEVA Logistics	31.5	25.0	
40 to 49			24.3	28.0	
50 to 59			14.6	20.8	
60 and over			4.8	5.0	
Cultural diversity - Distribution of staff by geographical area					
Europe / France	%	CMA CGM CEVA Logistics	31.1	33.0	
Africa/Middle East			11.8	13.0	
Americas			24.5	27.0	
Asia/Oceania			32.5	27.0	
Supporting local communities					
Percentage of countries that have launched at least one local initiative	%	CMA CGM	94	74	100% local initiatives by 2023
«Containers of Hope» humanitarian aid delivered	Number	CMA CGM	926	201	
Ensuring respect for human rights					
Human rights policy	Document	CMA CGM CEVA Logistics	Yes	Yes	
Vigilance Plan	Document	CMA CGM CEVA Logistics	Yes	Yes	

Partly

	INDICATOR	UNIT	SCOPE	2022	2021	OBJECTIVE
0						
1	Acting ethically, with integrity, and ensuring compliance in business practices					

Acting for fair trade

INDICATOR	UNIT	SCOPE	2022	2021	OBJECTIVE
Acting ethically, with integrity, and ensuring compliance in business practices					
Code of Ethics	Document	CMA CGM CEVA Logistics	Yes	Yes	
Share of employees who have received training on ethics and compliance issues	%	Group	85	90	
Share of all whistleblower reports made that were processed and closed during the year	%	Group	77	76	
Total number of alerts received	Number	Group	320	321	
Number of alerts processed and closed under the Group's alert system	Number	Group	245	244	
Offering innovative and sustainable services to our customers					
Share of inland transport by rail or barge	%	CMA CGM	64	71	70% by 2025
Entities financed or supported by the Group	Number	CMA CGM	130	First year reported	
Entities financed or supported by the Group whose business is in line with the Group's CSR challenges and objectives	%	CMA CGM	58	First year reported	
Number of containers transported via the humanitarian services office	TEU	CMA CGM	29,974	26,692	
CO_2 emissions reduced (ACT with CMA CGM+)	Tonnes	CMA CGM	76,297	74,174	
CO_2 emissions - indirectly reduced* in tonnes (ACT with CMA CGM+)	Tonnes	CMA CGM	142,973	64,889	
CO_2 emissions offset (ACT with CMA CGM+)	Tonnes	CMA CGM	1,146,570	602,203	
Total number of TEUs ACT with CMA CGM+	TEU	CMA CGM	1,463,791	784,377	
EcoVadis	Rating	CMA CGM	Platinum	Platinum	
Carbon Disclosure Project	Rating	CMA CGM	B	C	
Ensuring a sustainable supply chain					
Third Party Code of Conduct	Document	CMA CGM	Yes	Yes	
Suppliers assessed	Nombre	CMA CGM	14,884	14,594	100% by 2025

Partly			
Partly			
Partly			
Missed	-		
Missed	-		

	INDICATOR	UNIT	SCOPE	2022	2021	OBJECTIVE
0						
1	CMA CGM Environmental policy Document Yes CEVA Logistics					
2	fighting climate change					

Acting for the planet

INDICATOR	UNIT	SCOPE	2022	2021	OBJECTIVE
Environmental policy	Document	CMA CGM CEVA Logistics		Yes	
fighting climate change					
Number of dual-fuel vessels	Number	Maritime activities	32	23	95 by 2026
Nominal capacity of dual-fuel vessels	TEU		480,000	335,656	
Energy consumption					
Fuel consumption of the maritime fleet √	Millions of tonnes	Maritime activities	8.104	8.09	
of which LNG	Thousands of tonnes		409.29*	204.7	
of which biofuels	Thousands of tonnes		99.771	22.7	
Share of alternative fuels in the energy mix of maritime activities √	%		6.28	2.81	
Energy consumption of warehouses	MWh	CEVA Logistics	270,147	274,096	
Share of low-carbon electricity in warehouses	%		22.0	17.7	100% by 2025
Percentage of warehouse space covered by LEDs	%		75.7	63.2	100% by 2023
Surface area of photovoltaic panels on warehouse roofs	m²		666,042	506,561	1.8 million m² in 2025
Truck fleet fuel consumption	Millions of litres		73.60	68.53	
CO_2 emissions					
TTW CO_2e emissions from maritime fleet fuel consumption √	Millions de tonnes	Maritime activities	25.39	25.80	NetZero by 2050
WTW CO_2e emissions from maritime fleet fuel consumption √	Millions de tonnes		30.11	30.57	
Indirect reduction of CO_2e emissions from the purchase of biomethane under guarantees of origin	Tonnes		118,535	59,070	
Reduction of maritime CO_2e emissions relative to 2008 benchmark (gCO2/TEU-km)	%		-50	-48	
CO_2e performance of maritime activity CO_2/TEU.km TTW √	gCO2e/ TEU-km		51.8	53.9	
CO_2e performance of maritime activity CO_2/TEU.km WTW √	gCO2e/ TEU-km		61.4	63.9	
Scope 3 CO_2 emissions	Millions of tonnes		9.9	10.5	
CC Air Cargo Scope 1 CO_2 emissions	Thousands of tonnes	Air Cargo	97	First year reported	

	INDICATOR	UNIT	SCOPE	2022	2021	OBJECTIVE
0						
1	CO2e emissions from warehouse energy Tonnes 92,898 94,328 consumption (Scopes 1 & 2) CO2e emissions from truck fleet fuel Tonnes CEVA Logistics 188,586 178,849 consumption (Scopes 1 & 2) Scope 3 CO2e emissions M ofi l tl oio nn ns 3,3 F reir ps ot ry te ea dr e s					
2	Protecting air quality					

INDICATOR	UNIT	SCOPE	2022	2021	OBJECTIVE
CO_2e emissions from warehouse energy consumption (Scopes 1 & 2)	Tonnes	CEVA Logistics	92,898	94,328	
CO_2e emissions from truck fleet fuel consumption (Scopes 1 & 2)	Tonnes		188,586	178,849	
Scope 3 CO_2e emissions	Millions of tonnes		3,3	First year reported	
Protecting air quality					
NOx emissions from maritime fleet fuel consumption √	Thousands of tonnes	Maritime activities	575	590	
SOx emissions from maritime fleet fuel consumption √	Thousands of tonnes		71	74	
Fine particle emissions from maritime fleet fuel consumption √	Thousands of tonnes		54	56	
Number of owned vessels retrofitted for shore power connection	Number		66	61	
Number of available mobile shore power AMP connectors	Number		66	First year reported	
Percentage of owned fleet retrofitted with shore power connection systems	%		28	32	
Owned vessels equipped with scrubbers	Number		60	First year reported	
Percentage of owned vessels equipped with scrubbers	%		26	First year reported	
Protecting biodiversity					
Number of major marine oil pollution incidents	Number	Maritime activities	0	0	0 accidental pollution per year
Number of containers lost at sea	Number		69	0	0 containers lost at sea per year
Vessels equipped with ballast water treatment technology	Number		177	126	
Number of vessels equipped with FORS	Number		62	60	
Promoting a circular economy					
Ships dismantled	Number	Maritime activities	0	0	
Solid waste	m³ per vessel		144	135	
Oil sludge generated	Thousands of tonnes		179	173	

	Percentage of owned vessels equipped with scrubbers	%	Col2	26	First year reported	Col5
0	Protecting biodiversity					
1	0 accidental Number of major marine Number 0 0 pollution oil pollution incidents per year 0 containers Number of containers lost at sea Number 69 0 lost at sea per Maritime year activities Vessels equipped with Number 177 126 ballast water treatment technology Number of vessels Number 62 60 equipped with FORS					
2	Promoting a circular economy					

HFO	0.26 tCO_2e/t	Source: EN 16258
DO	0.68 tCO_2e/t	Source: EN 16258
VLSFO	0.64 tCO_2e/t	Source: GLEC
LNG	0.89 tCO_2e/t	Source: GLEC
BIOFUELS	0.554 tCO_2e/t	Source: RED II EU Directive 2018/2011

Key performance indicators and other quantitative results considered most important		Level of assurance
People	Total number of employees	Limited
	Share of women in management	
	Turnover rate	
	Number of training days per employee	
	Lost Time Injury Frequency Rate	
	Severity rate for employees of CEVA, terminals, seafarers	
	Number of fatal accidents	
Planet – Maritime Activity	Fuel consumption of the maritime fleet	Reasonable
	Share of alternative fuels in the energy mix of maritime activities	
	CO_2 emissions from maritime fleet fuel consumption (Well to Wake and Tank to Wake)	
	CO_2 performance of maritime activity CO_2 / TEU.km	
	SOx emissions from maritime fleet fuel consumption	
	NOx emissions from maritime fleet fuel consumption	
	PM related to the maritime fleet fuel consumption	
	Scope 3 CO_2 emissions of the maritime activity	Limited
	Number of owned vessels adapted for shore power connection	
	Production of oil sludge	
	Number of major marine oil pollution incidents	
	Number of containers lost at sea	
Planet – Inland Activity	Share of inland transport by rail or barge	Limited
	Truck fleet fuel consumption	
	CO_2 emissions from truck fleet fuel consumption	
	Energy consumption of warehouses	
	CO_2 emissions from warehouse energy consumption	
	Share of low-carbon electricity in warehouses	
	Scope 3 CO_2 emissions from subcontracted sea and air transport	
Responsible Trade	Share of alerts processed and closed among all reported alerts	Limited
	Suppliers assessed	
	CO_2 emissions reduced and offset with « ACT with CMA CGM+ »	

1.3 Hierarchical structure

File	Correct	Relatief Correct	Fout
Allianz (sort=True)	63	18	27
Allianz (sort=False)	103	0	1

Legenda:
= Goede plek/volgorde
= Foute plek/volgorde
= Foute plek/volgorde maar relatief aan het vorige element wel goed.

Allianz (sort=True):

title_1 = Competence Change Future
subtitle1_1 = ALLIANZ GROUP… …REPORT 2017
subtitle3_1 = About the report
paragraph_1 = This report… …Sector Supple-
subtitle4_1 = ASSET MANAGEMENT
paragraph_2 = As of… …Allianz Group
subtitle3_2 = Corporate Responsibility Governance and Strategy
paragraph_3 = and also … respective document).
subtitle3_3 = Company description
paragraph_4 = The Allianz… …70 countries.
subtitle4_2 = CORPORATE RESPONSIBILITY GOVERNANCE
paragraph_5 = The highest… …the underwriting
subtitle4_3 = INSURANCE BUSINESS
paragraph_6 = Our retail… …and strategies.
subtext4_1 = 1_ESG… …Allianz SE
paragraph_7 = associated with… …Group's cen-
subtitle4_4 = RISK MANAGEMENT
paragraph_8 = In the… …Responsibility function.
subtitle4_5 = ESG APPROACH
paragraph_9 = The types… …investment, and
subtitle4_6 = STAKEHOLDER ENGAGEMENT & MATERIALITY
paragraph_10 = As we… … November 2015.
subtitle4_7 = CORPORATE RESPONSIBILITY STRATEGY
paragraph_11 = We have… …and management.
subtext4_2 = 1_Cluster… …their
paragraph_12 = Strategy… …opportunities
subtext4_3 = revenue… …Allianz SE
paragraph_13 = An in-depth… …countries.
subtext3_1 = www.allianz.com/esg-framework.
paragraph_14 = We enable… …of compa-

subtitle5_1 = KPIS: TARGETS AND ACHIEVEMENTS
paragraph_15 = 583 ESG… …until 2019.
subtitle4_8 = ENVIRONMENTAL CONCEPT
paragraph_16 = Allianz is… …seek to:
subtitle3_4 = Environmental Matters
paragraph_17 = reduce the… …Management System
subtitle4_9 = CONCEPTS
paragraph_18 = Within our… …of Management
subtitle4_10 = CLIMATE CHANGE STRATEGY
paragraph_19 = At Allianz,... …integral part
subtitle5_2 = KPIS: TARGETS AND ACHIEVEMENTS
paragraph_20 = To ensure… …who have
subtext5_1 = Non-Financial… …Allianz SE
subtext2_1 = GHG emissions… …per employee
subtitle4_11 = ENCOURAGING… …(CONCEPT)
paragraph_21 = In 2016… …our operations.
subtitle3_5 = Social Matters
paragraph_22 = This chapter… …these fields.
subtitle4_12 = EMERGING CONSUMERS CONCEPT
paragraph_23 = Our first… …in Indo-
subtitle4_13 = SOCIAL INCLUSION CONCEPTS
paragraph_24 = As a… …Responsibility
subtext4_4 = 1_The majority… … Allianz Re.
paragraph_25 = nesia. Our… …full default
subtext5_2 = Non-Financial… …Non-Financial
paragraph_26 = they seek… …companies
subtitle5_3 = KPIS… …ACHIEVEMENTS
subtext2_2 = Encouraging… …during 2018
subtitle5_4 = KPIS… …ACHIEVEMENTS
subtext2_3 = Global… …defined
subtitle4_14 = RESPONSIBLE… …CONCEPTS
paragraph_27 = Our strong… …is our
subtitle4_15 = DATA PRIVACY_CONCEPTS
paragraph_28 = Digitalization… …organization
subtext5_3 = Non-Financial… …SE
paragraph_29 = Equally… …and
subtitle5_5 = KPIS:... …ACHIEVEMENTS
paragraph_30 = Status… …understand-
subtitle3_6 = Employee Matter
paragraph_31 = ing… …KPIs/targets
subtitle5_6 = KPIS… …ACHIEVEMENTS
paragraph_32 = Status… …May 2018
subtitle3_6 = Human-Rights Matters
paragraph_33 = This… …KPIs/targets
subtitle4_16 = CONCEPTS
paragraph_34 = To successfully… …the CEO
subtitle4_17 = CONCEPTS
paragraph_35 = We actively… …Examples include

subtext5_4 = Non-FInancial… …Allianz SE
paragraph_36 = Simplithon… …open-door
subtext3_2 = governance
paragraph_37 = formats… …and HR
subtitle3_8 = Compliance… …Bribery Matters
paragraph_38 = Our Board… …family commitments
subtitle4_18 = CONCEPTS AND PROGRAMS
paragraph_39 = The Allianz… …Conduct was
subtitle5_7 = KPIS… …ACHIEVEMENTS
subtext2_4 = AES… …and corruption
subtext4_5 = 1_Figure… …SE
paragraph_40 = In view… …interna-
subtitle5_8 = KPIS… …ACHIEVEMENTS
paragraph_41 = Status… …languages
subtext5_5 = Non-Financial… …SE
paragraph_42 = our attention… …HGB
subtitle3_9 = Independent… …Non-Financial
paragraph_43 = In a limited… …of the Non-
subtitle4_19 = RESPONSIBILITIES… …DIRECTORS
paragraph_44 = The executive… …Germany
subtitle4_20 = INDEPENDENCE… …FIRM
paragraph_45 = We have… …requirements
subtitle4_21 = ASSURANCE CONCLUSION
paragraph_46 = Based on… …with
subtitle4_22 = PRACTITIONER… …RESPONSIBILITY
paragraph_47 = Our responsibility… …come to
subtext5_6 = Non_Financial… …SE
subtitle4_23 = INTENDED… …REPORT
paragraph_48 = We issue… …auditor]
subtext5_7 = Non_financial… …SE

Allianz (sort=False):

subtitle1_1 = ALLIANZ GROUP… …REPORT 2017
title_1 = Competence Change Future
subtext4_1 = Non-Financial… …Allianz SE
subtitle3_1 = About the report
paragraph_1 = This report… …respective document).
subtitle3_2 = Company description
paragraph_2 = The Allianz… …70 countries.
subtitle4_1 = INSURANCE BUSINESS
paragraph_3 = Our retail… …and reinsurance.
subtitle4_2 = ASSET MANAGEMENT
paragraph_4 = As of… …website at
subtitle3_3 = Corporate Responsibility Governance and Strategy
paragraph_5 = At Allianz… …that arise.
subtitle4_3 = CORPORATE RESPONSIBILITY GOVERNANCE
paragraph_6 = The highest… …and strategies.
subtext5_1 = subtext5_1… …Allianz SE
paragraph_7 = Our group-level… …Responsibility function.
subtitle4_4 = STAKEHOLDER ENGAGEMENT & MATERIALITY
paragraph_8 = As we… …this report.
subtitle4_5 = CORPORATE RESPONSIBILITY STRATEGY
paragraph_9 = We have… …energy investments.
subtitle4_6 = RISK MANAGEMENT
paragraph_10 = In the… …these concepts.
subtitle4_7 = ESG APPROACH
paragraph_11 = The types… …and management.
subtext5_2 = 1_Cluster… …Allianz SE
paragraph_12 = An in-depth… …website at
subtext3_1 = www.allianz.com/esg-framework.
paragraph_13 = In the… …re- and
subtitle5_1 = KPIS: TARGETS AND ACHIEVEMENTS
paragraph_14 = 583 ESG… …until 2019.
subtitle3_4 = Environmental Matters
paragraph_15 = This chapter… …and targets.
subtitle4_8 = CONCEPTS
paragraph_16 = Within our… …carbon-neutral company.
subtitle4_9 = CLIMATE CHANGE STRATEGY
paragraph_17 = At Allianz… …website at
subtitle4_10 = ENVIRONMENTAL CONCEPT
paragraph_18 = Allianz is… …of Management
subtitle5_2 = KPIS: TARGETS AND ACHIEVEMENTS
paragraph_19 = To ensure… …mid-term
subtext4_2 = Non-Financial.. … Allianz SE
subtext2_1 = GHG emissions… …operations
subtitle3_5 = Social Matters
paragraph_20 = This chapter… …fields
subtitle4_11 = SOCIAL INCLUSION CONCEPTS

paragraph_21 = As a… …Responsibility
subtext5_3 = 1_The majority
paragraph_22 = Strategy… …money
subtitle4_12 = ENCOURAGING FUTURE… …(CONCEPT)
paragraph_23 = In 2016… …pupils
subtitle4_13 = EMERGING CONSUMERS CONCEPT
paragraph_24 = Our first… …default
subtext4_3 = Non-Financial… …SE
paragraph_25 = and failure… …companies
subtitle5_3 = KPIS… …ACHIEVEMENTS
subtext2_2 = Encouraging… …target defined
subtitle4_14 = RESPONSIBLE… …CONCEPTS
paragraph_26 = Our strong… …treated fairly
subtitle5_4 = KPIS… …ACHIEVEMENTS
subtext2_3 = Global… …2018
subtitle4_15 = DATA PRIVACY CONCEPTS
paragraph_27 = Digitalization… …organization
subtext4_4 = Non-Financial.. …SE
paragraph_28 = Equally… …disclosure).
subtitle5_5 = KPIS… …ACHIEVEMENTS
paragraph_29 = Status/progress… …2018
subtitle3_6 = Human-Rights Matters
paragraph_30 = This chapter… …targets
subtitle4_16 = CONCEPTS
paragraph_31 = We actively… …concepts
subtitle5_6 = KPIS… …ACHIEVEMENTS
paragraph_32 = Status/progress… …decisions
subtitle3_7 = Employee Matters
paragraph_33 = This chapter… …targets
subtitle4_17 = CONCEPTs
paragraph_34 = To successfully… …include
subtext4_5 = Non-Financial… …SE
paragraph_35 = Simplithon… …commitments
subtitle5_7 = KPIS… …ACHIEVEMENTS
subtext2_4 = AES for… …website at
subtext3_2 = governance
subtitle3_8 = Compliance… …Matters
paragraph_36 = In this… …concepts and
subtitle4_18 = CONCEPTS… …PROGRAMS
paragraph_37 = The Allianz… …corruption
subtext5_4 = 1_Figure… …SE
paragraph_38 = In view… …records
subtitle5_8 = KPIS… …ACHIEVEMENT
paragraph_39 = Status/progress… …testing
subtext4_6 = Non-Financial… …SE
subtitle3_9 = Independent Practitioner… ...Non-Financial
paragraph_40 = We have… …Non-financial Report
subtitle4_19 = RESPONSIBILITES… …DIRECTORS

paragraph_41 = The executive… …or error
subtitle4_20 = INDEPENDENCE… …FIRM
paragraph_42 = We have… …requirements.
subtitle4_21 = PRACTITIONER… …RESPONSIBILITY
paragraph_43 = Our responsibility… …information
subtitle4_22 = ASSURANCE CONCLUSION
paragraph_44 = Based on… …conjunction with
subtext4_7 = Non-Financial… …Allianz SE
subtitle4_23 = INTENDED USE… …REPORT
paragraph_45 = We issue… …public auditor]

1.4 Garbage filter

Bestanden	Spaties	Nummers	Geen letters	Undecided
Allianz	222	20	121	61
CMA	214	353	263	201
EFRAG	30	398	483	304
DPEF	204	79	209	136
VDL	72	99	197	236
NASDAQ	179	205	321	213

Bestanden	Sub 3 woorden of garbage	Andere talen of gemiste woorden	Weblinks
Allianz	23	33	8
CMA	90	112	-
EFRAG	160	138	6
DPEF	104	31	-
NASDAQ	106	107	-

De undecided waarden voor Allianz

Weblinks	Sub 3 woorden of garbage	Andere talen of gemiste woorden	
www.allianz.com/annualreport.	(AS RRIM),	Allianz	Reporting
www.allianz.com/annualreport.	(EMS)	Allianz	spective websites at
www.allianz.com/corporate-	% in-	Allianz Elementar Versicherungs-Aktiengesellsc...	Status/progress:
www.allianz.com/sustainability.	% of	Allianz ESG	Status/progress:
www.allianz.com/sustainability.	€ 5.6 bn.	Allianz S.p.A. in Milan, Italy	Status/progress:
www.allianz.com/sustainability.	289c to 289e HGB.	Allianz Suisse Versicherungs-Gesellschaft AG i...	Status/progress: 60
www.allianzgi.com	At	Allianz’s	Status/progress: 68
www.pimco.com.	In 2017 20.6	Austria	Status/progress: 72
	In 2017 20.6	defines	Status/progress: In 2017, we achieved a 17
	Repor	implemented in 2000.	Status/progress: In 2017, we achieved a 29
	Repor	kids (relaunched in 2016).	Switzerland
	Repor	KPIs/targets.	To Allianz SE, Munich
	Repor	Munich, 26 February 2018	Undecided
	Repor	of Allianz SE.	using
	sessments	ppa. Annette Daschner	was 32.2 GJ.
	t	PricewaterhouseCoopers GmbH	Wirtschaftsprüfer
	t		Wirtschaftsprüfungsgesellschaft
	t		
	t		
	t		
	trophes;		
	ment.		
	In 2017 20.6		

De undecided waarden voor CMA

Sub 3 words and or garbage		Different language or missed words	
CO	- 1/3 by 2025	compared to 2008.	- Responsability:
CO	- etc.	compared to 2008.	– Decarbonisation
CO	- In 2022,	compared to 2008.	"Excelerate" programme
CO	- IT	conditioning	(compared to 2021)
CO	-10% in 2023	conditioning, etc.	& SUSTAINABILITY
CO	-40% in 2025	d’alerte	15. IN UKRAINE: CEVA
CO	(gCO	d’alerte	219,270 tonnes of CO
CO	(gCO	d'évaluation	3,296 ktonnes CO
e	**CO	d'évaluation	8,000 tonnes
e	/TEU-km)**	DATABASE	Europe/France:
e	2 0 2 2 C S R R E P O R T	des mesures	FRANCE-KLM
e	2 0 2 2 C S R R E P O R T	des mesures	FRANCK MAGARIAN,
e	20 to 29:	des risques	GENERALISED
e	200 by 2023	des risques	Groupe CMA CGM
e/TEU-	200 by 2025	du Ministère des Affaires Etrangères	HappyIndex
e/TEU-km	30 to 39:	november 2022	CMA CGM Asia
E/TEU-KM)*	30.1 MtCO	december 2022	CMA CGM France
e/TEU.km)*	30.11 (25.39 on a TTW	december 2022	CMA CGM France
ECOVADIS	30.57 (25.80 on a TTW	december 2022	CMA CGM GBS Chongqing
eq	335,656 TEU		CMA CGM: DECARBONISING
eq.	39.3 MtCO	ACCELERATING	CABAU WOEHREL CHRISTINE,
etc.).	40 to 49:	accelerating	California
ETS).	468,000 TEU	adaptées	carpooling
CMA CGM’S CO	50 to 59:	adaptées	CEVA
by 2023	61.4 (51.8 on a TTW	airbuses A330F	CEVA
CMA CGM	63.9 (53.9 on a TTW	Americas:	CEVA

CMA CGM	A350FS,	ANL Australia’s VolunTEU programme provides	IN CÔTE D’IVOIRE:
CMA CGM	in TEU	Asia/Oceania:	IN FRANCE
IFREMER	nd	BANQUES ALIMENTAIRES	in gCO
IMO	ND	BEIT EL BARAKA	IN LEBANON
IMO	o	BIO-/E-LNG:	IN SLOVAKIA, GEFCO AUTOMATES
IMO	of	BIO/E-METHANOL:	including
IMO	R&D:	biodiversity.	including
in	RD	Boeing B777F	IT:
in 2015.	rd	La cartographie	IT:
in 2022	RD	La cartographie	ktonnes
in 2022	s	Le Carburateur	October 2021
in 2022.	s	Le mécanisme	of Marseille.
in 2023	SAF:	Le mécanisme	of tonnes
in 2023 (compared to 2022)	st	Le suivi	OPTIMISATION
in CO	TEU	Le suivi	optimisation, etc.
in CO	TEU	Les procédures	organisations.
th	USD 6,526	Les procédures	owned
th	WWF	MONITORING	PAUL HAERI,
TH		Monitoring	programme in France.
th		Monitoring	PROMOTING
		rationalisation of IT uses	representing 206kt
		Savannah, Georgia	RESTOS DU CŒUR
		SECOURS POPULAIRE	roadmap
		shared by CMA CGM	RODOLPHE SAADÉ,
		SmartShips programme,	Rodolphe Saadé,
		Syroco, incubated by ZEBOX.	VIGEO EIRIS
		TANYA SAADÉ ZEENNY,	WHISTLEBLOWING
		TANYA SAADÉ ZEENNY,	TEUs representing
		TELEWORKING	to Ukraine.
		ZEBOX	UNICEF,

De undecided waarden voor EFRAG

Sub 3 words and or garbage		Different language or missed words		Weblinks
– EUR 20	, EBA (2018)	, compared	Normalised	https://eur-lex.europa.eu/legal-content/EN/TXT...
– EUR 3	, ECB (2020)	Adapted	Normalised	https://quantis-suite.com/Scope-3-Evaluator/
– KPI	, ECIIA (2014)	as follows:	Normalised	https://sseinitiative.org/
, AECA (2020)	, ecoact	Austria	Norway	https://sseinitiative.org/
, CDP	, ecoact (2016)	Austria	Norway	www.efrag.org
, CDP (2017)	, ecoact (2019)	Austria, Belgium, Bulgaria, Croatia,	Omitted	www.efrag.org
, CDSB (2014)	, EFAA (2017)	B-1000 Brussels	Organisation	
, CDSB (2018)	, ESMA (2020)	Belgium	Others	
, CDSB (2018)	, ESMA (2020)	Belgium	platform/	
, CDSB (2020)	, ESMA (2020)	Biodiversity	platform/	
, CDSB (2020)	, ESMA (2020)	BrownFlynn	platform/	
, CDSB (2020)	, EY (2018)	Bulgaria	Poland	
, CDSB (2020)	, EY (2019)	Bulgaria	Poland	
, CDSB (2020)	, EY (2020)	Bulgaria, Denmark, Italy, Poland,	Poland, Portugal, Romania, Slovakia,	
(IRM-FESG)	, EY (2020)	december 2008	Portugal	
(iv)	, FRC (2019)	Classification/taxonomy	Portugal	
(OECD)	, FRC (2020)	Croatia	referenced	
(SDGs)	, IMA (2020)	Croatia	reported;	
(TCFD)	, IMA (2020)	CSR-Rich.-Umsetzug. of 11/04/2017	REPORTING	
(v)	, in	Cyprus	Reporting	
(XBRL,	, Kathryn Thomas (2016)	Cyprus	reporting	
(XBRL,	, Mazars & ecoDa (2020)	Denmark	reporting	

(XBRL,	, OECD (2014)	Denmark	Reporting	
< EUR 300	, or	disaggregated by	reporting	
> EUR 20	, PwC (2016)	Donato Calace	Reporting	
1000 (CZK)	, PwC (2019)	Estonia	reporting	
12000 (HUF)	, SASB (2020)	Estonia	Reporting	
15 (HRK)	, TCFD (2017)	Estonia, France, Greece, Hungary,	Reporting	
156 (DKK)	, TCFD (2020)	Europe (2017 & 2019);	reporting	
170 (PLN)	, WBCSD (2019)	European	Reporting	
175 (SEK)	, WBCSD (2019)	European	Reporting	
2016/17:CU2 of 25/10/2016	, WBCSD (2020)	EuropeanLab@efrag.org	reporting	
30 (HRK)	, WBCSD (2020)	February 2021	Reporting	
3000 (ISK)	, XBRL (2020)	Frameworks/standards/guidance	Reporting	
313 (DKK)	, XBRL (2020)	FRAMEWORKS/STANDARDS/GUIDANCE	Reporting	
350 (SEK)	. IMA (2020)	France	reporting	
38 (BGN)	(2017/C 215/01)	France	Reporting	
6000 (HUF)	(AECA)	generated	reporting	
6000 (ISK)	(EFFAS)	Germany	reporting (2020)	
76 (BGN)	(EMAS)	Germany	Reporting ESG	
85 (PLN)	(i)	Globescan	reporting.	
a)	(i)	goes mainstream	required	
a)	(i)	goes mainstream	Romania	
a)	(i)	goes mainstream	Romania	
a)	(ii)	goes mainstream	Slovakia	
A6.	(ii)	goes mainstream	Slovakia	
aisbl - ivzw	(ii)	Greece	Slovakia, Slovenia, Spain, Sweden;	
Amd. 237 of 06/06/2017	(ii)	Greece	Slovakia, Spain	

Amd. 2564 of 03/09/2017	(iii)	Greece, Hungary, Ireland, Italy, Latvia, Lithu...	Slovenia	
Amd. 3 of 19/05/2017	(iii)	Hungary	Slovenia	
Amd. 462 of 14/12/2016	(iii)	Hungary	Spain	
Amd. 51 of 12/12/2016	(iii)	info@efrag.org	Spain	
Amd. XIII-94 of 15/12/2016	(ILO)	Integrated	Spanish	
Amd. ZGD-1J of 29/03/2017	Bua 58/2016 of 3/05/2016	Integrated	Sustainability	
b)	c)	Integrated	Sustainability	
b)	c)	Investing	Sweden	
b)	c)	Ireland	Sweden	
b)	CDP	Ireland	XBRL in Europe	
Bedenik, Petra Barišić	CDSB	Italy		
D.Lgs. 254/2016 of 30/12/2016	CDSB	Italy		
d)	EU Eco-	Japanese		
e)	EU27	KPI disaggregated		
EBRA	EUR 300	KPIs		
EFAA	EY	Latvia		
EFAA	EY	Latvia		
EFFAS-ESG	EY	Liechtenstein		
EFRAG	EY	Liechtenstein		
EIOPA (2020)	EY	Lithuania		
EMAS	EY	Lithuania		
ESMA	f)	Lithuania, Luxemburg, Malta, Norway,		
i)	g)	Ludwig-Fröhler-Institut		
IDC	j)	Luxembourg		
IMA	KPMG	Luxembourg		
NFRD.	L33014 / 03/2017	Malta		
Nidžara Osmanagić	OECD	Malta		

Ord. 1.938 of 17/08/2016	OMFIF-Refinitiv	manufactured,		
PRI	S.I, 360 of 30/07/2017	McKinsey		
PwC	SASB	mendations		
SMEs.	SASB	Netherlands		
TCFD recom-	SASB, IR, GRI, CDSB,	Netherlands		

De undecided waarden voor DPEF

Sub 3 words and or garbage		Different language or missed words
• I&D:	6.A	december 2021
1.A	6.A	(Cuernavaca).
1.A	6.B	6 - ÉTHICS
1.B	6.B	Animale SA (%)
1.B	6.C	at December 31, 2021
1.C	6.C	Autovaccines
1.C	771 M	AVIS DE L’OTI
161mg/PCU 2021 to 91 6mg/PCU)	79.7% (H)	by evapoconcentration.
2.A	at	Carbone
2.A	DPEF / 2020	Ceva Santé Animale
2.A	DPEF / 2021	Ceva Santé Animale
2.B	DPEF / 2021	Ceva Santé Animale
2.B	DPEF / 2021	Ceva Santé Animale
2.B	DPEF / 2021	Ceva Santé Animale
2.C	DPEF / 2021	Ceva Santé Animale
2.C	DPEF / 2021	Ceva Santé Animale
2.C	DPEF / 2021	Ceva Santé Animale
20.3% (F)	DPEF / 2021	Cevolution
3.A	DPEF / 2021	ciples
3.A	DPEF / 2021	EY & Associés
3.A	DPEF / 2021	EY & Associés
3.B	DPEF / 2021	France (Brest - Ecat-ID - Desvac).
3.B	DPEF / 2021	IMPLEMENTING
3.B	DPEF / 2021	Injectables
3.C	DPEF / 2021	Intranet
3.C	DPEF / 2021	invested in I&D in 2021
3.C	DPEF / 2021	pation
4.A	DPEF / 2021	Philippe Aubain
4.A	DPEF / 2021	purchasing
4.B	DPEF / 2021	Sharing of expertise
4.B	DPEF / 2021	Thème
4.C	DPEF / 2021	
4.C	DPEF / 2021	

40.1% (F)	DPEF / 2021	
5.A	DPEF / 2021	
5.A	DPEF / 2021	
5.A	DPEF / 2021	
5.B	DPEF / 2021	
5.B	DPEF / 2021	
5.B	DPEF / 2021	
54kT CO2 eq.	DPEF / 2021	
59.9% (H)	DPEF / 2021	
in 2021	DPEF / 2021	
jects	DPEF / 2021	
NA	DPEF / 2021	
NA	DPEF / 2021	
NA	DPEF / 2021	
o	DPEF / 2021	
o	DPEF / 2021	
o	DPEF / 2021	
o	st	
th	st	

De undecided waarden voor VDL

Deze zijn niet uitgewerkt omdat het bestand nederlands talig is.

De undecided waarden voor NASDAQ

Sub 3 words and or garbage		Different language or missed words	
-40 % en 2025	(CCI)	november 2021	(France)
-40% by 2025	(g CO	november 2021	27% America
28% 40 to 49	(SDGs)	december 2021	27% Asia/Oceania
2O5O	/ TEU-km)	december 2021	33% Europe/France
363 K	/TEU	– Decarbonation	70% in 2050, compared to 2008.
45 K	/TEU trans-	– Worldwide specialised	a programme to optimise its owned
474 K	/TEU-km in	(1 4000 teus) in Rotterdam	A rewarded
50 m3	/TEU-km in 2021.	aimed at acce-	ACCELERATING
506 561 m	1 800 000 m	Airbus operated	ACCELERATING
AI R C A RG O	10% by 2023	Announced by Rodolphe Saadé CEO	achieved.
APPRO-	141 063t of CO	biodiversity	by 2030 compared to 1990.
AXES	15,9 M	biodiversity	by 50% in 2050, compared to 2008.
BIO LNG	182 K	biodiversity.	Capitale de la Mer.
by 2025	20,8% 50 to 59	BIOFUEL	Centre
BY 2050	2020 to 53 g CO	BIOFUELS TO	Chapître
by 2050.	2020, to 1,15 tonnes of CO	BIOMETHANE/	CHRISTINE CABAU WOEHREL,
C M A CG M	21,4 M	CMA CGM Estonia has implemented	committed to
CEVA	21% 20 to 29	CMA CGM FRANCE	COMMITTING TO
CEVA	25% 30 to 39	CMA CGM FRANCE	compared to 2008
CEVA	262 K	CMA CGM FRANCE	compared to 2008
CEVA	CMA CGM	CMA CGM Scandinavia has developed	compared to 2008
CFA	CMA CGM	CMA CGM UK	compared to 2019
CO	CMA CGM	Europe.	compensated.
CO	CMA CGM –	excluded	CONTRIBUTING TO A
CO	CSR com-	explained by:	COOPERATION
CO	digitized.	FRANCK MAGARIAN,	In February 2021, CMA CGM
CO	duce)	implemented.	In Latin America:
CO	ECOVADIS	INTERMODAL	In Marseille:
CO	ehouses, etc.	investing in:	including

CO	eq.	INVOLVING	India
CO	EVALUA-	its CO	Indonesia
CO	in 2021.	JACQUES SAADE in Rotterdam in	initiated by Rodolphe
IMO	in 2022.	January 1st, 2023.	LAURENCE DAMBRINE,
IMO	in 2022.	mentoring programme.	Lebanon
IMO	in 2050	MOBILISATION	LEBANON
kok	in 2050	Monitoring of	Lebanon:
ktons	LTIFR	November 2021 -	OPTIMISED,
Lagrange.	ment)	OBJECTIFS	organisation
lators.	N-4.	October 2021	performing
mics.	NA	October 2021	POWERD
mittee	NC	October 2021	PRIATE
MONITO-	NC	October 2021	Programme
of	NGOs.	October 2021 in Dublin.	programme (Lebanon)
of 2022.	rintendents (MSI).	roadmap	PROGRAMMES AT
of 335,656 TEU	S TR ATEGY	ROADMAP 2050	PROMOTING
of bio LNG	S TR ATEGY	Rodolphe Saadé,	PROMOTING
of CO	ST	Rodolphe Saadé,	PROMOTING
tals.	TION	Rodolphe Saadé, CEO of CMA CGM	PURSUING
TANYA SAADÉ ZEENNY,	tion,	Saadé, CEO of CMA	QUALIOPI
TANYA SAADÉ-ZEENNY,	to 10%.	seafearers	RESPONDING TO
TEU	VIGEO EIRIS -	ticos programme in Côte d'Ivoire.	STRUCTURING R&D
TEU	vs. 2008	TO BIOMETHANE	Zebox
TEU-km	XXX	tonnes of CO	Zhengzhou
		up a programme entitled Roadmap	

2 File Data Storage

2.1 Tijd prestaties

Files	verstreken t in (s)	# pagina's	Gemiddelde t per pagina (s)
Allianz (2018)	51.28	11	4.66
CMA	1290.33	83	15.55
EFRAG	1811.18	78	23,22
DPEF	1306.75	41	31,85

Er is te zien dat de runtime van de neo4j upload per pagina varieert van bestand tot bestand, dit is zeer afhankelijk van hoeveel inhoud er staat op de desbetreffende pagina. Alle bestanddelen van een pagina worden namelijk ook nog met elkaar vergeleken om connecties te leggen.

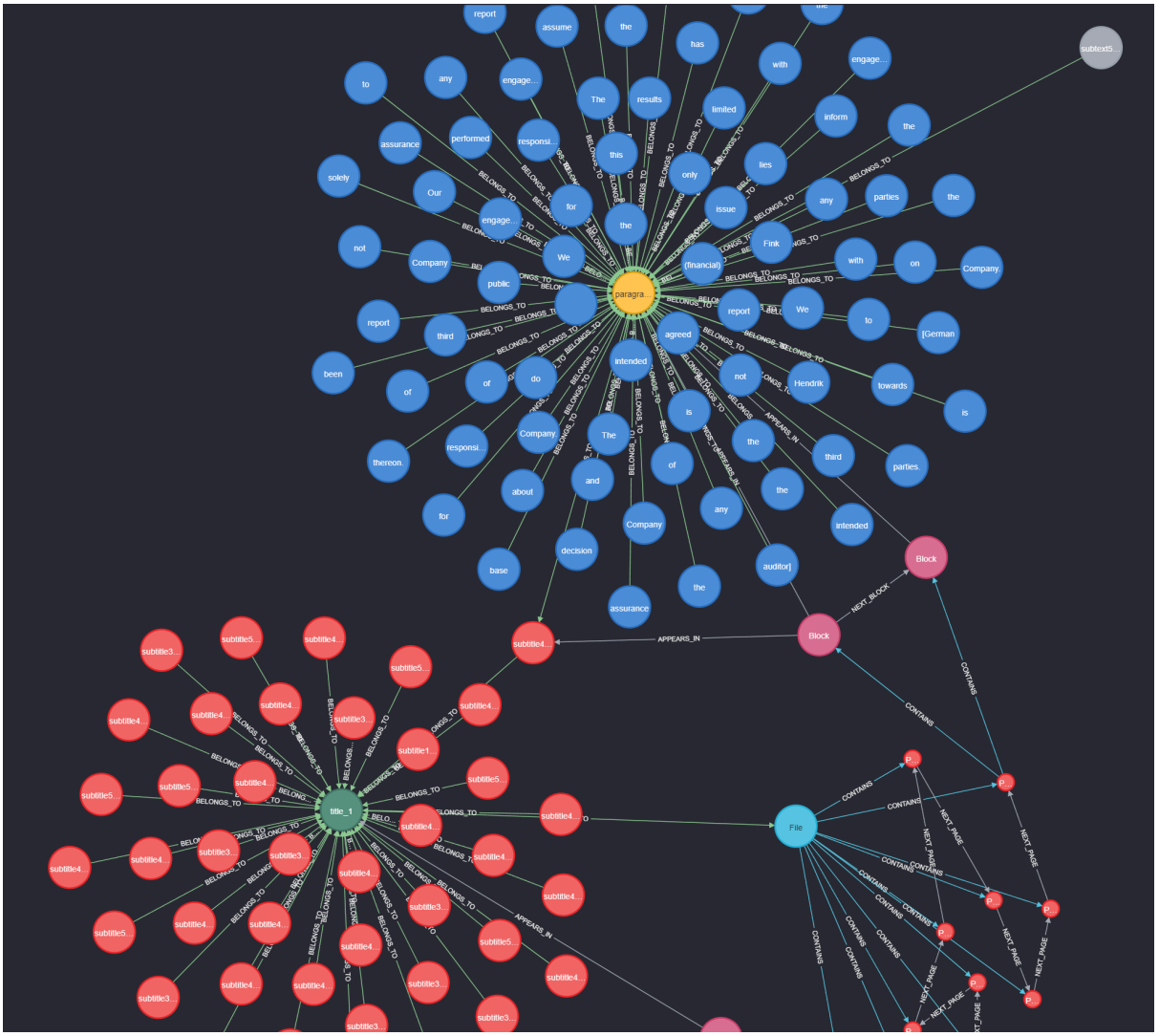
Gemiddelde tijd per pagina: 18,82 seconden, dit getal zegt een stuk minder dan de stijgende waarden in de tabel waaruit is af te leiden dat de prestaties van de neo4j database achteruitgaan als het aantal nodes toeneemt. Het toenemen van het aantal nodes in de database heeft de grootste impact op hoe lang het gemiddeld duurt om een pagina in de database te zetten.

Het structurize gedeelte gaat aanzienlijk sneller dan het aanvankelijk uploaden van de data, maar ook deze hangt zwaar af van de hoeveelheid data.

2.2 Structuur

De absolute structuur loopt als volgt:
Absolute structuur: File > nPage > nBlock > Lines > nWord

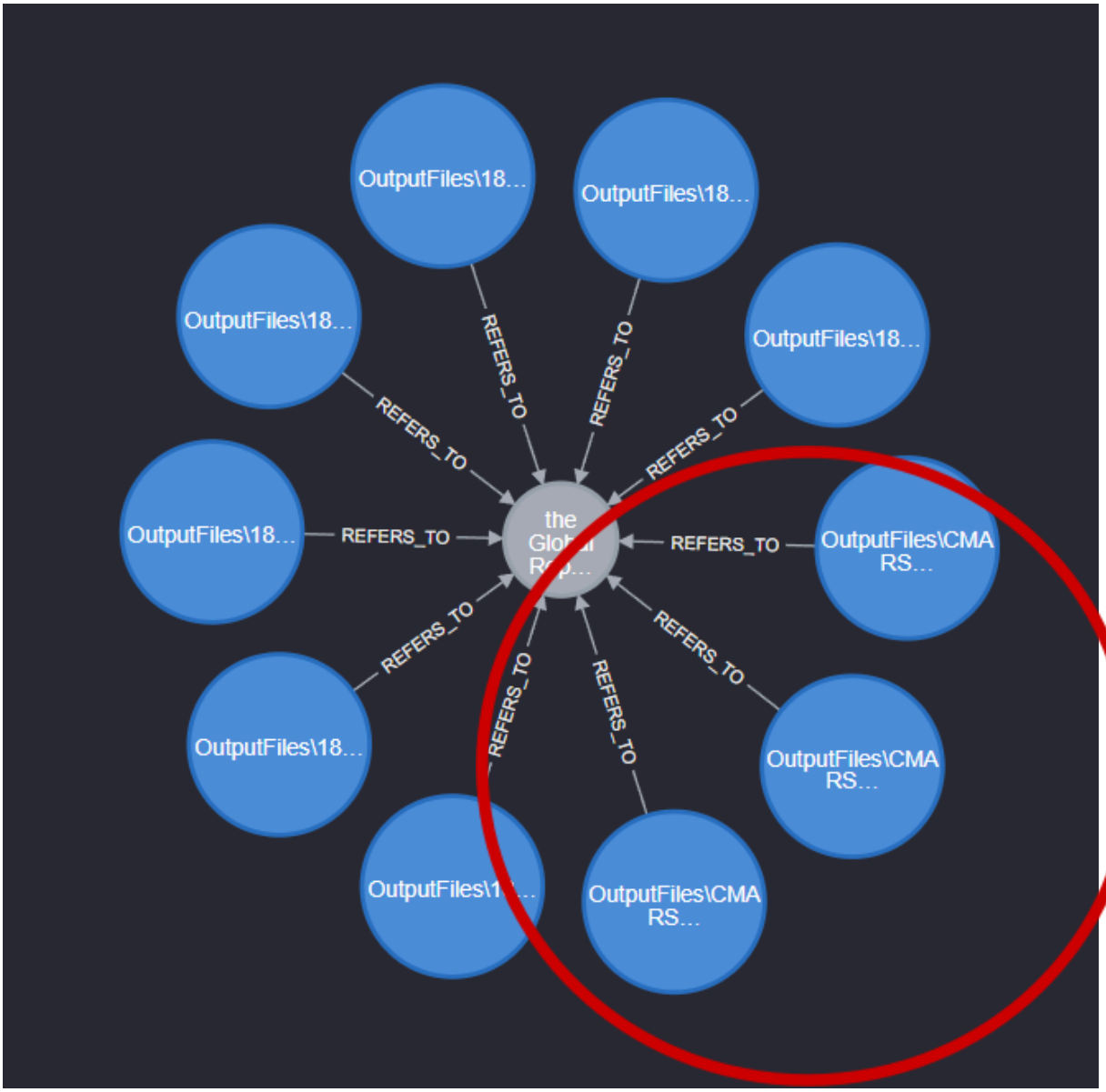
Ook worden images en tables aan pages gelinkt met de -CONTAINS-> relatie.



Geïnterpreteerde structuur: File > nTitle > nSubtitle > nParagraph > nPargraph_Word

In deze structuur is ook al meteen de link te zien tussen de absolute en de geïnterpreteerde structuur doordat de absolute “blocks” met de geïnterpreteerde elementen worden verbonden.

Deze structuur zal zoals de naam al suggereert een geïnterpreteerde indeling van de PDF volgen. Aan de hand van -BELONGS_TO-> relaties is te zien hoe de structuur in elkaar zit. De blocks die voorkomen in een geïnterpreteerde node worden ermee verbonden door middel van een -APPEARS_IN-> relatie zodat de originele structuur achterhaald kan worden.



In deze afbeelding is te zien dat de woorden(de blauwe nodes) verbonden zijn met een -REFERS_TO-> relatie. Sommige nodes komen uit hetzelfde bestand, maar er is ook te zien dat er enkele verschillen van de rest (zie omcirkelde nodes). Dit betekent dat verwijzingen worden gekoppeld over de hele database. Hierdoor zijn verbanden te leggen tussen bestanden op basis van de entiteiten die naar voren komen uit de tekst

3 Coreferencing

Door bij het draaien de coreferencing functie te kiezen. Kan de gebruiker paragraph_”paragraaf nummer” invullen om zo deze paragraaf contextueel kloppend terug te krijgen. Met contextueel kloppen wordt bedoeld: Verwijzingen van entiteiten zijn vervangen met de entiteiten zelf zodat de lezer het stuk per zin kan gebruiken zonder de gehele context te kennen.

Input tekst:

As of 31 December 2017, with approximately € 1,960 bn of assets under management, we are one of the largest asset managers in the world that manage assets by means of active investment strategies. We run our Asset Management business through two distinct invest-ment management units, Allianz Global Investors (AllianzGI) and PIMCO, both of which operate under the umbrella of Allianz Asset Management GmbH (AAM). Particular strongholds include the United States of America, Germany, France, Italy, the United Kingdom and the Asia-Pacific region. As of 31 December 2017, Allianz managed € 1,448 bn in third-party assets. For additional information on our business model, please refer to the Group Management Report of the Allianz Group Annual Report 2017 on our website at The 2017 Non-Financial Report covers the entire Allianz Group

Nieuwe text:

As of 31 December 2017, with approximately €1,960 in of asset under management, Allianz are one of the largest asset managers in the world that manage asset by means of active investment strategies. Our run Allianz Asset Management business through two distinct investment management units, we Global Investors (Allianz GI) and PICO, both of which operate under the umbrella of we Asset Management GmbH (AAM). Particular strongholds include the United States of America, Germany, France, Italy, the United Kingdom and the Asia-Pacific region. As of 31 December 2017, we managed €1,448 in in third-party assets. For additional information on Allianz business model, please refer to the Group's Management Report of the the Group's Annual Report 2017 on Allianz website at The 2017 Non-Financial Report covers the entire we the Group's

Deze functionaliteit werkt zeker niet perfect, maar werkt voor nu acceptabel.

Performance van het model:

Benchmark	**F-score**
OntoNotes	80.9
LitBank	80.2
PreCo	88.3

Opmerkingen:

- De F-score is een maat voor de prestaties van een model op het gebied van coreferentie-identificatie. Een hogere F-score geeft een betere prestatie aan.
- De prestaties van het model op OntoNotes, in een diverse reeks tekstuele contexten.
- De prestaties van het model op LitBank, het aanpakken van coreferentie-resolutie in literaire teksten, waar nuances en complexiteit vaak een uitdaging vormen.

- De prestaties van het model op PreCo, waarbij sprake is van informeel taalgebruik en variabiliteit in de zinsstructuur.

Referenties

Allianz Group and Allianz SE [Allianz]. (2018). Competence change future: Allianz Group and Allianz SE combined separate Non-financial Report 2017. In *allianz.com*. https://www.allianz.com/en/investor_relations/results-reports/annual-reports/nf-report.html

albertochigua & Github. (2020, 19 november). *In the extract images from a PDF, some are shown with inverted colors · Issue #725 · pymupdf/PyMuPDF*. GitHub. https://github.com/pymupdf/PyMuPDF/issues/725