

'Combinations of Concern - CoC'

*Python to automate
an Analytical Risk Based Monitoring (AR)
project identifying
discrepancies across fields
from a gov. open-source dataset
of covid-19 vaccine registrations.*

PHUSE EU CONNECT 2022

AR02

Aldir MEDEIROS FILHO

Some of covid vaccines containers used in Brazil



24/01/2021 Chagada das vacinas de Oxford/Astrazenica / Fiocruz, vin...
by Governo Santa Catarina



Data dictionary -

columns (out of 32) allowing to identify a directly or indirectly a vaccine provider

Colum nr - c_	Campo	Description in English
13	Paciente_endereço_cep	Subject's declared residency Postal Code 5 dígitos
18	Estabelecimento_município_codigo	Gov. IBGE code nr for where vaccine was applied
23	Vacina_categoria_code	Population category code nr
24	Vacina_categoria_nome	Population category label
25	Vacina_lote	Vaccine batch nr
26	Vacina_fabricante_nome	Vaccine manufacturer name
27	Vacina_fabricante_referencia	Vaccine manufacturer official registration nr with BR gov.
28	Vacina_data_Aplicacao	Date vaccine dose was applied
29	Vacina_descricao_dose	Dose Description
30	Vacina_codigo	Specific vaccine provider code
31	Vacina_nome	Vaccine's name
32	Sistema origem	Original system used enter registration data

Combination of Concern - CoC

when the cross check of different fields
from an individual register (one register per dose)
recorded in the federal OpenDataSUS dataset
shows conflicting information
impeding to confirm which vaccine an individual received.

CoC version 01 Up to CoC_07 (Nov 21)	CoC version 02 From CoC-08 (Jan 22)
c_26_manufacturer name	c_25_vaccine_batch_number
c_30_vaccine_code	c_30_vaccine_code
c_31_vaccine_name	C_31_vaccine_name

CoC version 01 up to Nov 2021

The most clear example of CoC was for a register showing

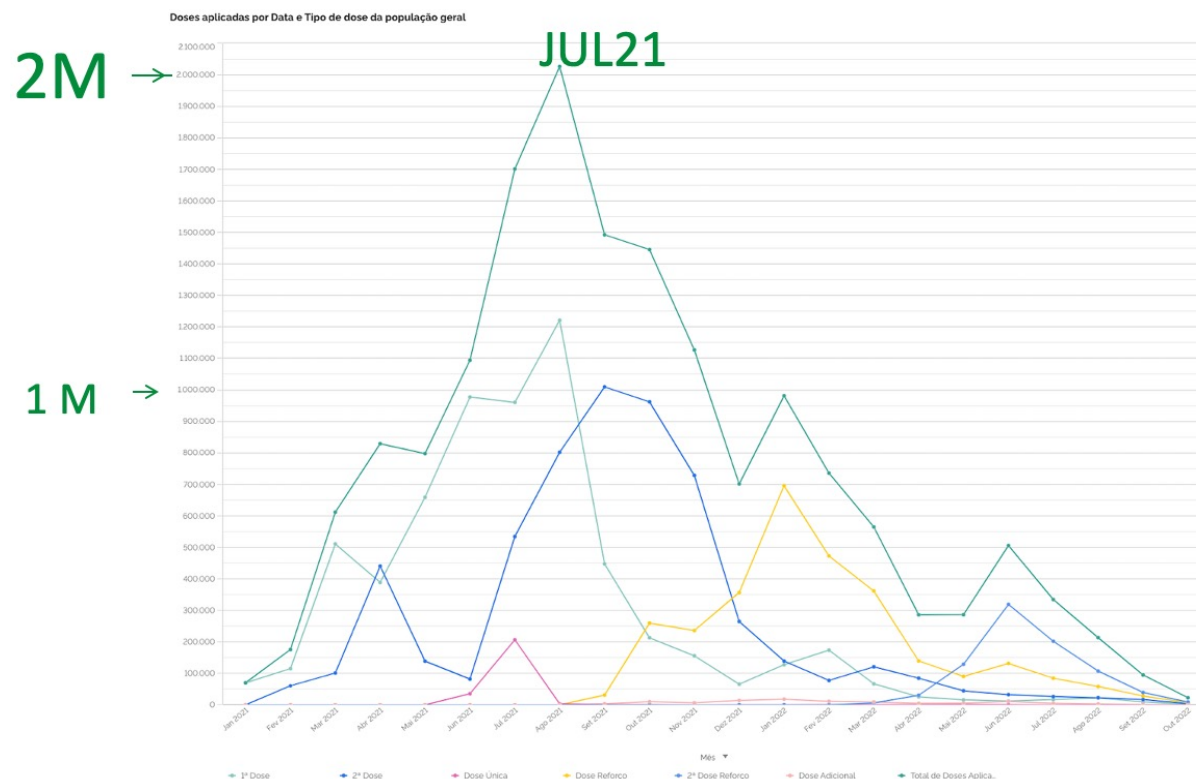
C26_manufacturer's name	C30_code and/or c31_name
AstraZeneca covid-19 vaccine (Serum from India, manufacturer of the generic)	Sinovac/Coronavac (from China).

Up to CoC_07 (early Nov 2021)
the overall proportion of CoCs per month
was ~ 12% of all monthly registers

CoC version 02 from mid January 2022

- CoC version 02 started with CoC_08, after the Brazilian Ministry of Health websites were restored from 'the hack' that lasted 'from early Dec. 2021 to mid Jan. 2022'.
- 'following' the 'hack', the government's teams 'simplified' column 26. The main problem is that we no longer could identify 'the true' source vaccine manufacturer from the fields c26.
- That obliged me to create a CoC version 2, focusing on c25_batch number and looking for information from sources outside OPDS dataset to be able to possibly identify the real manufacturer for each vaccine batch in column 25 (c25).
- From there I created new column to be populated with coded flag alerts about the Combinations of Concern focused on the c25.

LEFT (LOCALIZAUS 23 Oct 22) – Total nr of applied doses per month per LOCALIZASUS VACINOMETRO
RIGHT (Historic of Google Collab IDE & Pandas & Python CoC monthly iterations) per OpenDataSUS



Historic of Google Collab IDE & Pandas & Python CoC monthly iterations								
Original iteration nr	simplified CoC monthly iteration nr	Dataset Freeze	OPDS original raw file size (GB) on hard drive	TOTAL rows (registrations)	added registrations since last CoC	Catch up' from previous months before last CoC.	% 'catch up' per added registrations	Nr of vaccinations effectively applied since previous CoC iteration
1 & 2	1 & 2	01-Jun-21	1.1	2.059.426				
3	3	01-Jul-21	1.7	3.133.344	1.116.119			
4	4	01-Aug-21	2.6	4.854.329	1.720.985	282.647	16%	1.438.338
5	5	01-Sep-21	3.6	6.822.321	1.967.992	241.618	12%	1.726.374
6	6	01-Oct-21	4.7	8.839.765	2.017.444	341.322	19%	1.676.122
7	7	01-Nov-21	5.4	10.148.626	1.308.861	146.408	10%	1.162.453
8	8	17-Jan-22	5.9	11.921.415	1.772.789	123.801	8%	1.648.988
9	9	01-Feb-22	6.1	12.481.710	560.295	112.719	14%	447.576
10	10	01-Mar-22	6.5	13.317.696	835.985	310.159	36%	525.826
14	11	01-Apr-22	6.9	14.092.523	774.827	375.503	48%	399.324
15	12	01-May-22	7.0	14.387.231	294.708	56.036	19%	238.672
16	13	01-Jun-22	7.2	14.858.441	471.21	231.373	49%	239.837
17	14	01-Jul-22	7.5	15.339.609	481.168	25.72	5%	438.664
18	15	01-Aug-22	7.7	15.756.690	417.081	115.105	28%	301.976
19	16	01-Sep-22	8.3	17.321.648	however, ~ 1.275.000 registers not covid related			
19_BIS	16_BIS	02-Sep-22		16.045.255	288.565	80.239	28%	208.326
20	17	01-Oct-22	7.9	16.163.979	118.724	26.343	22%	92.381
21	18	31-Oct-22	7.9	16.220.784	56.805	9.567	17%	47.238



Case Study Region

- Population:
 - [Brazil](#) : 217,240,060 ([7th](#))
 - [Santa Catarina](#)(SC): 7,252,502

SC Area:

Total	95,730 km ² (36,961 sq mi)
• Rank	20th

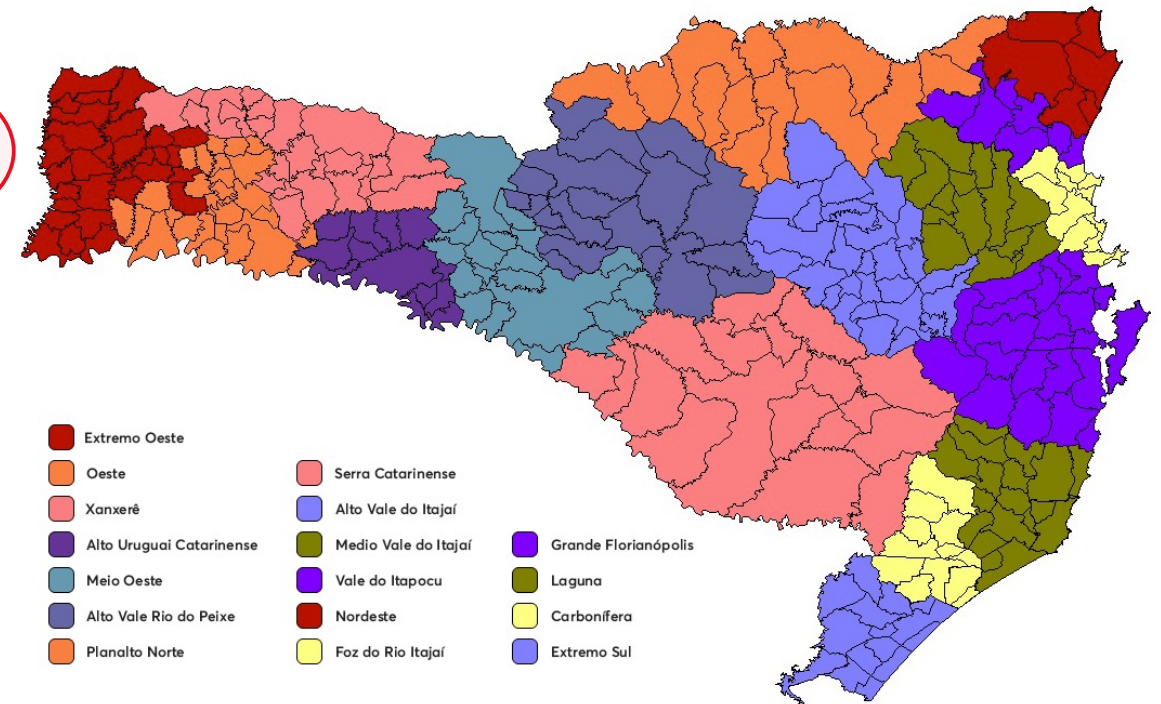
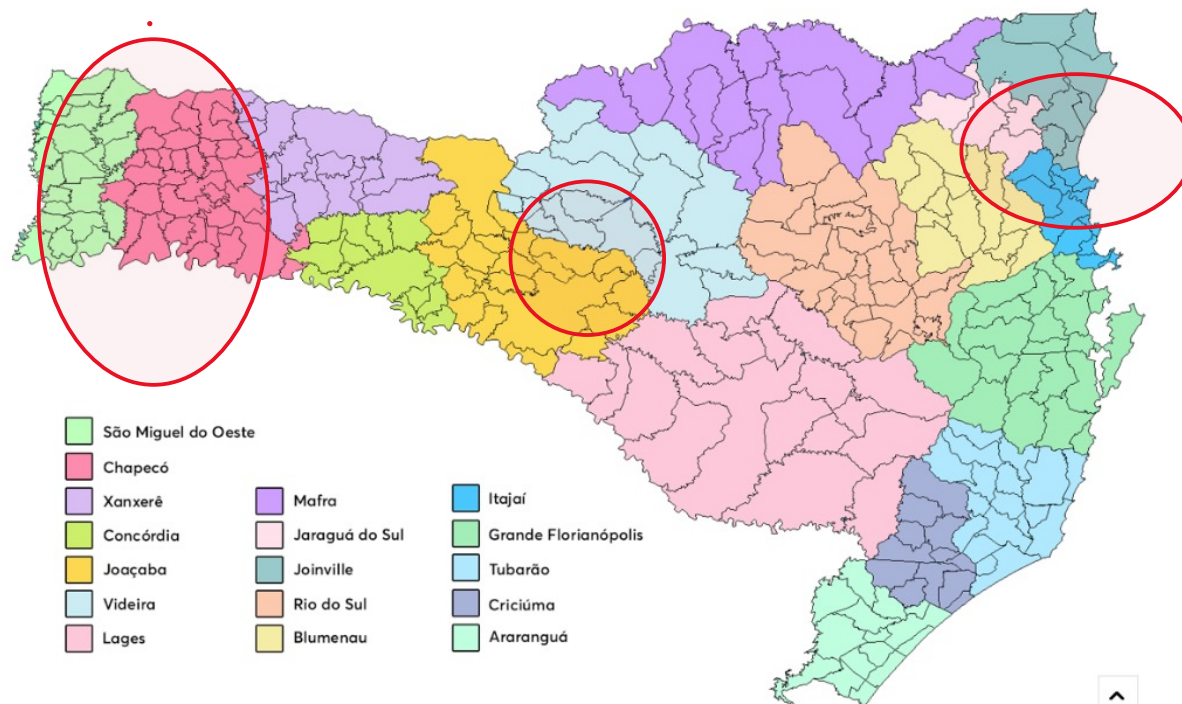


Santa Catarina

295 counties

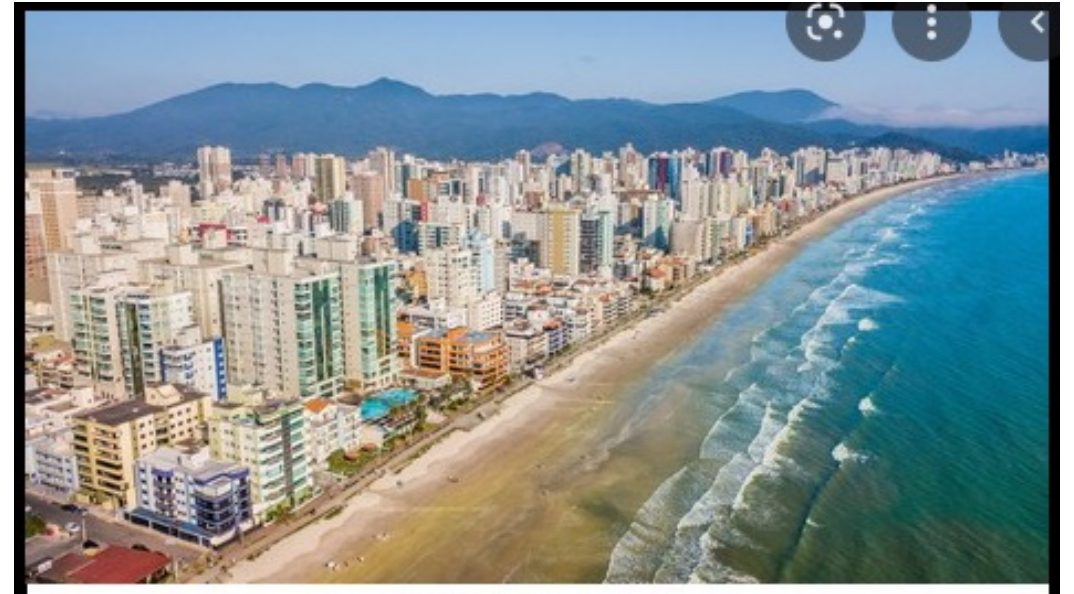
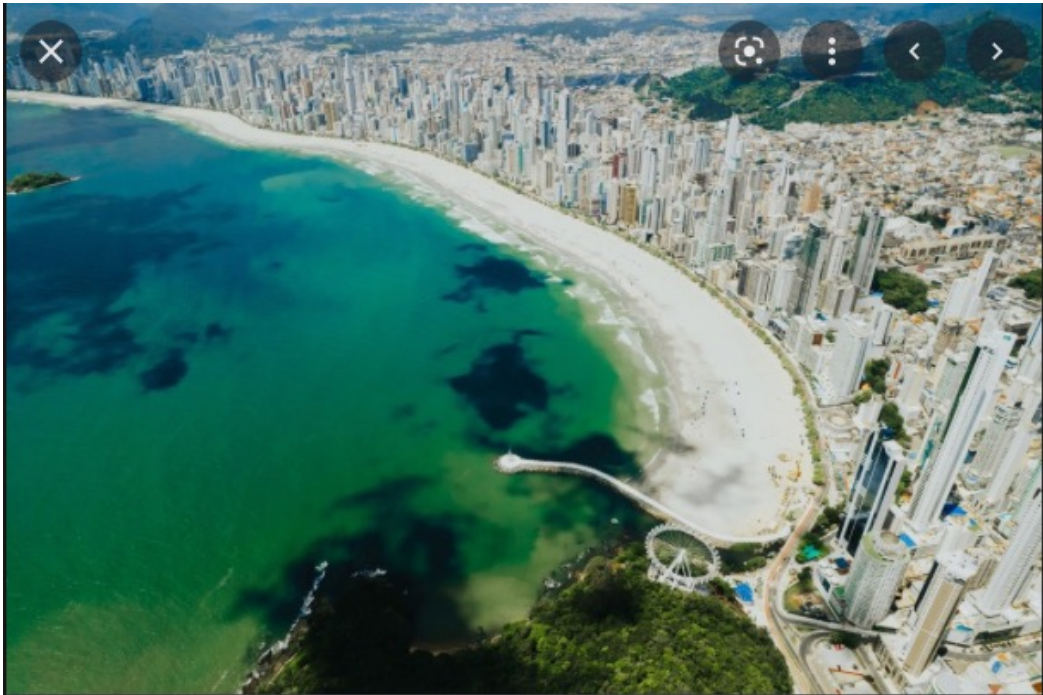
FIRST CHALLENGE

- IN – vaccine distribution by UDVE regionals
- OUT – vaccine application registration by PDR regionals



Camboriu (left)
Itapema (right)

Some SC cities are very vertically clustered!!!



Challenge 2:

Identifying IN and OUT different terminology and nr of counties

IN				OUT																	
DIVE SC UDVES				PLANO DIRETOR DE REGIONALIZACAO - PDR - 2018						ESTIMATED POPULATION BY AGE RANGE											
Programmer code	UDVE REGION NAME	Estimated Population pelo IBGE	Nr of counties	MACRO REGIAO DE SAUDE PER PDR 2018	REGIAO DE SAUDE	POP 2018	MUNICIPIOS			05 a 11	12 a 17	18 a 19	20 a 29	30 a 39	40 a 49	50 a 59	60 a 69	70 a 79	80 +'		DIF IN MINUS OUT
1	UDVE SMO	181,755	22	GRANDE OESTE	Extremo Oeste	231,848	30		Extremo Oeste	18,555	14,959	5,025	35,800	36,990	30,725	32,025	24,477	13,539	6,800	218,895	-12,953
2	UDVE XANXERE	164,649	17		Xanxere	200,114	21		Xanxere	18,693	15,353	5,255	32,438	31,321	26,756	25,416	17,974	9,764	4,486	187,456	-12,658
3	UDVE CHAPECO	457,339	39		Oeste	360,933	27		Oeste	31,911	27,636	9,535	60,634	62,854	49,999	46,611	31,480	16,410	7,506	344,576	-16,357
4	UDVE CONCORDIA	144,090	13	MEIO OESTE E SERRA CATARINENSE	Alto Uruguai Catarinense	143,337	13		Alto Uruguai Catarinense	11,646	9,761	3,297	20,316	22,207	20,065	19,796	15,328	8,751	4,189	135,356	-7,981
5	UDVE JOACABA	202,037	21		Meio Oeste	191,303	20		Meio Oeste	16,898	14,298	4,956	29,876	29,922	26,665	25,157	17,713	10,043	4,858	180,386	-10,917
6	UDVE VIDEIRA	287,962	19		Alto Vale do Rio do Peixe	293,133	20		Alto Vale do Rio do Peixe	29,131	24,212	8,353	48,747	44,683	40,225	36,161	23,483	12,852	5,824	273,671	-19,462
14	UDVE LAGES	287,853	18		Serra Catarinense	288,479	18		Serra Catarinense	26,710	21,824	7,386	44,414	43,840	38,790	36,759	26,015	14,621	6,965	267,324	-21,155
7	UDVE MAFRA	381,125	13	PLANALTO NORTE E NORDESTE	Planalto Norte	376,993	13		Planalto Norte	36,543	30,096	10,346	60,111	59,273	54,752	46,589	31,335	15,656	7,632	352,333	-24,660
8	UDVE JARAGUA DO SUL	282,193	5		Nordeste	1,023,135	13	7 counties	Vale do Itapocu	30,067	26,516	9,091	52,687	55,607	45,712	35,876	22,447	10,445	4,894	293,342	-40,465
9	UDVE JOINVILLE	775,243	8					6 counties	Nordeste	68,393	57,622	19,569	119,987	123,867	105,953	89,890	62,105	29,212	12,730	689,328	
10	UDVE ITAJAI	731,766	11	FOZ DO RIO ITAJAI	Foz do Rio Itajai	698,912	11		Foz do Rio Itajai	67,704	57,831	19,891	121,057	125,284	103,184	86,126	58,456	27,675	12,106	679,314	-19,598
11	UDVE RIO DO SUL	300,397	28	VALE DO ITAJAI	Alto Vale do Itajai	295,201	28		Alto Vale do Itajai	27,486	22,747	7,687	47,365	49,313	40,919	37,787	25,322	14,025	6,801	279,452	-15,749
12	UDVE BLUMENAU	808,052	14		Medio Vale do Itajai	782,458	14		Medio Vale do Itajai	70,195	63,156	21,729	130,919	142,256	115,337	99,499	66,914	32,160	15,289	757,454	-25,004
13	UDVE GRANDE FPOIS	1,229,335	22	GRANDE FLORIANOPOLIS	Grande Florianopolis	1,189,947	22		Grande Florianopolis	100,796	89,996	31,315	191,665	217,253	181,603	151,015	109,470	55,291	25,608	1,154,012	-35,935
15	UDVE TUBARAO	371,748	18	SUL	Laguna	365,687	18		Laguna	31,349	26,007	8,661	55,333	60,422	50,410	51,076	36,725	19,657	8,891	348,531	-17,156
16	UDVE CRICIUMA	442,578	12		Carbonifera	433,675	12		Carbonifera	38,126	32,619	11,073	69,832	77,767	59,007	58,501	40,298	18,603	8,271	414,097	-19,578
17	UDVE ARARANGUA	204,380	15		Extremo Sul Catarinense	200,339	15		Extremo Sul Catarinense	18,595	15,263	5,086	31,469	32,107	26,891	27,452	18,939	10,015	4,570	190,387	-9,952
Totais		7,252,502	295			7,075,494			TOTAL	642,798	549,896	188,255	1,152,650	1,214,966	1,016,993	905,736	628,481	318,719	147,420	6,765,914	
									AGE RANGE	05 a 11	12 a 17	18 a 19	20 a 29	30 a 39	40 a 49	50 a 59	60 a 69	70 a 79	80 +'		-309,580
																					-4%

Challenge 3

DOSE INTERVALS PERMUTATIONS PER VACCINE MANUFACTURER, AGE & GROUP PRIORITY AND/OR VACCINE AVAILABILITY

ESQUEMAS VACINAS COVID-19

PESSOAS ACIMA DE 18 ANOS - IMUNOCOMPETENTE - NÃO GESTANTE

	Dose 2 (D2)	1ª Dose Reforço (DR1)		2ª Dose de Reforço (DR2)*	
		Intervalo	Imunizante	Intervalo	Imunizante
CoronaVac	4 semanas após D1	4 meses após D2	AstraZeneca, Janssen ou Pfizer	4 meses após DR1	AstraZeneca, Janssen ou Pfizer
AstraZeneca	8 semanas após D1	4 meses após D2	AstraZeneca, Janssen ou Pfizer	4 meses após DR1	AstraZeneca, Janssen ou Pfizer
Pfizer	8 semanas após D1	4 meses após D2	AstraZeneca, Janssen ou Pfizer	4 meses após DR1	AstraZeneca, Janssen ou Pfizer
Janssen	NÃO Indicada	2 meses após dose D1	AstraZeneca, Janssen ou Pfizer	4 meses após DR1	AstraZeneca, Janssen ou Pfizer

*Apenas para pessoas com 80 anos ou mais.

PESSOAS ACIMA DE 18 ANOS - IMUNOSSUPRIMIDO - NÃO GESTANTE

	Dose 2 (D2)	Dose adicional		Reforço	
		Intervalo	Imunizante	Intervalo	Imunizante
CoronaVac	4 semanas após D1	8 semanas	Mesmo imunizante do esquema primário	4 meses após dose adicional	AstraZeneca, Janssen ou Pfizer
AstraZeneca	10 semanas após D1	8 semanas	Mesmo imunizante do esquema primário	4 meses após dose adicional	AstraZeneca, Janssen ou Pfizer
Pfizer	8 semanas após D1	8 semanas	Mesmo imunizante do esquema primário	4 meses após dose adicional	AstraZeneca, Janssen ou Pfizer
Janssen*	NÃO recomendada	8 semanas	Janssen	4 meses após dose adicional	AstraZeneca, Janssen ou Pfizer

*PACIENTE COM SEGUINTE ESQUEMA:

D1: Janssen
DA: Pfizer

Recebe DR com intervalo de 4 meses da DA.

Neste caso a DR deverá ser preferencialmente, da plataforma de RNA mensageiro (Comirnaty/Pfizer) ou, de maneira alternativa, vacina de vetor viral (Janssen ou AstraZeneca)

PESSOAS ACIMA DE 18 ANOS - GESTANTE E PUÉRPERAS IMUNOCOMPROMETIDAS

	Dose 2 (D2)	Dose Adicional (DA)		Dose Reforço (DR)	
		Intervalo	Imunizante	Intervalo	Imunizante
CoronaVac	4 semanas após D1	8 semanas após D2	Mesmo imunizante do esquema primário	4 meses após a dose adicional (DA)	Preferencialmente Pfizer, se contraindicação (CRIE) utilizar CoronaVac
Pfizer	8 semanas após D1	8 semanas após D2	Mesmo imunizante do esquema primário	4 meses após a dose adicional (DA)	Preferencialmente Pfizer, se contraindicação (CRIE) utilizar CoronaVac

PESSOAS ACIMA DE 18 ANOS - GESTANTE

	Dose 2 (D2)	Dose Reforço (DR)	
		Intervalo	Imunizante
CoronaVac	4 semanas após D1	4 meses após D2	Pfizer + Se contraindicação (CRIE) utilizar CoronaVac
Pfizer	8 semanas após D1	4 meses após D2	Pfizer + Se contraindicação (CRIE) utilizar CoronaVac

ADOLESCENTES - 12 A 17 ANOS

	Dose 2 (D2)	Dose Adicional (DA)	Dose Reforço (DR)
Pfizer	8 semanas após D1	-	Não indicado
Pfizer *imunossuprimido	8 semanas após D1	8 semanas após D2	A partir de 4 meses após a dose adicional
CoronaVac	4 semanas após D1	-	Não indicado

CRIANÇAS - 5 A 11 ANOS

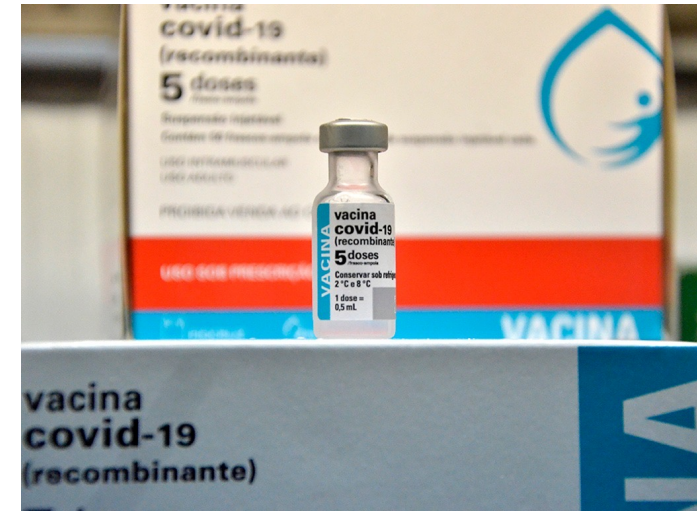
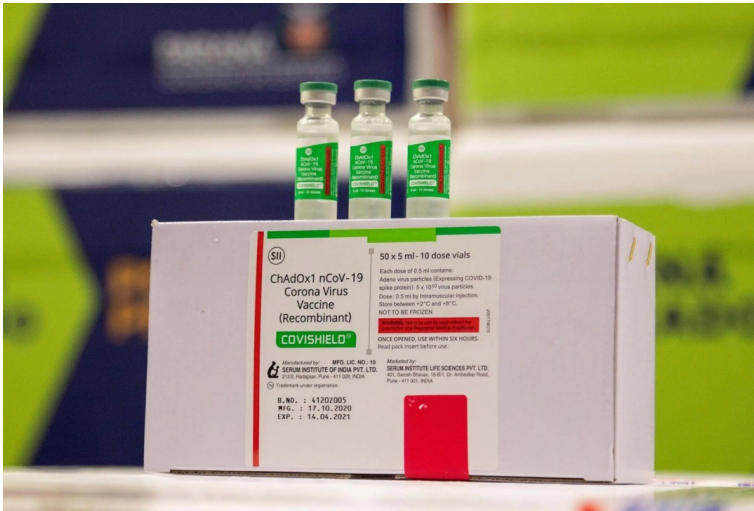
	Dose 2 (D2)
Pfizer (5 a 11 anos)	8 semanas após D1
CoronaVac (6 a 11 anos)	4 semanas após D1

*Exceção pacientes imunossuprimidos

#EUESCOLHISERVACINADO



CHALLENGE 4 – ‘ONE BRAND’ FOR ‘MANY MANUFACTURER SOURCES’ i.e. AstraZeneca



atualização 6 Jan 2022

atualização em 22 Feb 22

CoC_v2_project Notebooks on 31 Oct 22

Part I: Data Wrangling

- Ntbk_01 - Downloading, data wrangling, adding UDVE column and parqueting;
- Ntbk_02 - Defining & Reviewing 'Combinations of Concern (CoC)' versions 1 and 2, first EDA cross_checking of UDVEs & CoCs;

Part II: Checking some specific set of variables (columns)

- Ntbk_03: Review columns c21_to_c24
- Ntbk_04: Doses_Types_and_V_Manufacturers
- Ntbk_05: Subject_ID_AGE_crosschecks
- Ntbk_06: MULTIDOSE_UNIDOSE_DUPLICATES

Part III: Focusing on data from 'FLORIANOPOLIS' (SC's capital)

- Ntbk_07: FLORIPA_c_18_Multidose_Unidose_Duplicated
- Ntbk_08: FLORIPA_c_13_Multidose_Unidose_Duplicated
- Ntbk_09: FLORIPA_c_13_within_FLORIPA_c18_Multidose_Unidose_Duplicated

Part IV: Reviewing Pediatric data (age range 5 to 11 yo) for SC and 'Florianopolis'

- Ntbk_10: SC_PEDIATRICS

CoC_v2 - flags

The flags in this column follow my Risk Based coding criteria about consistency between columns c25, c30 and c31.

The cross checking of CoC_v2 with counties and UDVEs also serves as indicators to identify, and monitor RBM's QTLs (Quality Tolerance Limits)

Table XX – CoC_v2 flag_codes																					
CoC_v2 flag code	“Color” code	Rationale																			
G	Green	'c25_v_lote' identification is correctly spelt AND 'consistent' with available data in columns 'c30' and 'c31'																			
LG	Light Green	'c25_v_lote' identification is 'lightly misspelled' (a zero instead O or vice-versa'. However cross checking with other sources suggests it matches a documented vaccine shipment AND/OR is 'consistent' with 'c30' or 'c31'																			
Y	Yellow	Per original new dataset from 17 Jan 2022, the AztraZeneca brand is the only vaccine provider (out of the 4) that has now two different set of identification for the adult population over 18 yo. <table><tr><th colspan="2">c30_v_codigo</th><th colspan="2">c31_v_nome</th></tr><tr><td>85</td><td></td><td colspan="2">ASTRAZENECA/FIOCRUZ - COVISHIELD</td></tr><tr><td>89</td><td></td><td colspan="2">ASTRAZENECA - ChAdOx1-S</td></tr></table> Despite conflicting information across some official government documents a more reliable one suggests that 'ChAdOx1' would refer only to 'ready-to-use' AZ vaccines manufactured in Belgium and provided by the COVAX agreement 'COVISHIELD' would refer to any of the other four sources (i.e.; 'Ready to use Monodose seringe' or 10 dose' vials manufactured by Serum India, 10 doses vial produced from IFA delivered by Serum India, 5 or 10 dose vials manufactured in ??, etc,' Code yellow is when review of fields c25, c30 and c31 suggest that there was a mix up between 'ChAdOx1" and 'COVISHIELD' data entered in the register.				c30_v_codigo		c31_v_nome		85		ASTRAZENECA/FIOCRUZ - COVISHIELD		89		ASTRAZENECA - ChAdOx1-S					
c30_v_codigo		c31_v_nome																			
85		ASTRAZENECA/FIOCRUZ - COVISHIELD																			
89		ASTRAZENECA - ChAdOx1-S																			
R	Red	When we have a correct spelled or misspelt batch number that is a mismatch with the data in columns c30 and/or c31. In other words, batch number from one manufacturer coded with the vaccine code from another.																			
U	Unknown	When the batch number is so badly misspelt to a point that we cannot match with any possible batch number AND & OR I could not identify a reliable source that could confirm the potential lot number AND & OR the batch numbers were too small to suggest that it was related to a real batch number																			
UR	U RED	When we have names, phone or ID numbers that are clearly not batch numbers																			
P_ALERTE	Pediatric Alert	Since Jan 22 and up to Oct 22 we have only two providers for pediatric vaccines against covid-19. <table><tr><th colspan="2">c30_v_codigo</th><th colspan="2">c31_v_nome</th></tr><tr><td>Over 18 yo</td><td>5 to 11 yo</td><td>Over 18 yo</td><td>5 to 11 yo</td></tr><tr><td>86</td><td>98</td><td>SINOVAC/BUTANTAN - CORONAVAC</td><td>SINOVAC - CORONAVAC</td></tr><tr><td>87</td><td>99</td><td>PFIZER - COMIRNATY</td><td>PEDIÁTRICA - PFIZER COMIRNATY</td></tr></table> Back on Jan 2022 Pfizer started to also provide specific pediatric vaccines batches. For the Pediatric SINOVAC we still have a confusing picture. P_ALERTE is when my review shows that the batch number in c25 does not matches the c30 or c31 or c3 patient age.				c30_v_codigo		c31_v_nome		Over 18 yo	5 to 11 yo	Over 18 yo	5 to 11 yo	86	98	SINOVAC/BUTANTAN - CORONAVAC	SINOVAC - CORONAVAC	87	99	PFIZER - COMIRNATY	PEDIÁTRICA - PFIZER COMIRNATY
c30_v_codigo		c31_v_nome																			
Over 18 yo	5 to 11 yo	Over 18 yo	5 to 11 yo																		
86	98	SINOVAC/BUTANTAN - CORONAVAC	SINOVAC - CORONAVAC																		
87	99	PFIZER - COMIRNATY	PEDIÁTRICA - PFIZER COMIRNATY																		

UDVEs and CoC_v2_flags

CoC_v2	G	LG	Y	R	U	UR	P_ALERTE	Total	G_%	TOT_HIGH_FLAG_ALERTS	HIGH_FLAG_ALERTS_%
c35_SC_UDVE_Nr											
udve01	464,520	1,602	97	17	46	0	343	466,625	99.55%	503	0.11%
udve02	350,978	3,601	2,287	260	493	61	1,940	359,620	97.6%	5,041	1.4%
udve03	1,110,175	11,417	2,868	82	1,369	236	1,966	1,128,113	98.41%	6,521	0.58%
udve04	383,835	2,132	16	7	25	17	468	386,500	99.31%	533	0.14%
udve05	476,473	1,332	82	1	118	5	685	478,696	99.54%	891	0.19%
udve06	591,190	30,104	6,345	133	2,314	136	1,494	631,716	93.58%	10,422	1.65%
udve07	753,690	39,296	27,370	636	2,904	272	9,207	833,375	90.44%	40,389	4.85%
udve08	497,440	19,548	5,805	16	333	72	22	523,236	95.07%	6,248	1.19%
udve09	1,616,989	21,864	7,721	24	629	289	5,236	1,652,752	97.84%	13,899	0.84%
udve10	1,544,868	15,578	5,679	5	1,086	93	3,758	1,571,067	98.33%	10,621	0.68%
udve11	612,258	2,645	1,031	0	354	0	325	616,613	99.29%	1,710	0.28%
udve12	1,586,532	18,000	8,617	62	853	200	2,999	1,617,263	98.1%	12,731	0.79%
udve13	2,829,100	81,834	25,634	354	1,820	292	52,504	2,991,538	94.57%	80,604	2.69%
udve14	667,838	1,405	313	7	75	138	1,279	671,055	99.52%	1,812	0.27%
udve15	870,755	2,970	611	2	479	8	1,572	876,397	99.36%	2,672	0.3%
udve16	961,935	2,577	514	16	471	326	1,251	967,090	99.47%	2,578	0.27%
udve17	444,124	1,619	0	5	353	0	247	446,348	99.5%	605	0.14%
Total	15,762,700	257,524	94,990	1,627	13,722	2,145	85,296	16,218,004	97.19%	197,780	1.22%

UDVE 13 (22 counties) and CoC_v2_flags



	CoC_v2	G	LG	Y	R	U	UR	P_ALERTE
c18_e_municipio_codigo								
420060	16,041	0	0	0	0	0	0	23
420070	21,809	49	0	0	0	0	0	54
420090	12,627	168	383	0	17	0	0	394
420110	8,565	246	1	2	10	0	0	0
420120	26,313	26	0	0	0	0	0	49
420230	161,311	292	0	0	2	0	0	363
420370	21,740	206	0	0	3	0	0	79
420540	1,225,706	53,015	11,573	281	487	75	0	27,893
420570	63,132	337	1	0	0	0	0	147
420600	37,928	14	0	2	7	4	0	35
420980	8,334	213	0	0	95	0	0	32
421020	7,441	0	0	0	0	0	0	51
421150	27,757	40	1	2	8	0	0	55
421190	419,152	8,884	7,598	0	121	12	0	6,346
421230	19,371	0	0	0	1	0	0	38
421430	8,112	20	11	0	0	0	0	0
421570	59,824	347	1	8	22	1	0	94
421590	7,969	97	0	0	0	0	0	9
421630	54,900	76	0	0	1	0	0	153
421660	540,295	10,503	4,852	54	854	12	0	14,925
421725	13,325	98	2	5	67	188	0	2
421800	67,448	7,203	1,211	0	125	0	0	1,762

Part II – Key learning

UNIDOSE	Individuals with only one register so far - 'D1' or 'SINGLE_DOSE' or 'D2' or 'D3' or 'D4' or 'D5' or 'UNK1' or UNK2 only
MULTIDOSE	Individuals with more than one dose register ('D1' or 'SINGLE_DOSE' or 'D2' or 'D3' or 'D4' or 'D5' or 'UNK1' or UNK2 only)
DUPLICATE	Individuals from the group multidose who have a registration with a duplicated dose ('D1' or 'SINGLE_DOSE' or 'D2' or 'D3' or 'D4')

**CoC_18_Ntbk_06 Santa Catarina SC_311022 _c2_expected_same_individuals_DATASET	Total registrations	Total c2_p_id (ou nr de pessoas)	original labels c29_v_descricao_dose	clustered by me as	Totals per DOSE	REGISTERS & INDIVIDUALS UNIDOSE_only.shape	REGISTERS c2_p_id_short MULTIDOSE registers.shape	INDIVIDUALS c2_p_id_short MULTIDOSE unique	REGISTERS Multidose_DX_only_ duplicate_row_only .shape	INDIVIDUALS Multidose_DX_only_ duplicate_row_only .unique
status per OpenDataSUS on 31 Oct 2022**	16.220.310	6.741.027				949.475	15.270.835	5.791.552		
			'1ª Dose'	D1	6.257.644	632.129	5.625.515	5.563.387	123.579	61.451
			'Única'	UNICA	2.503	54	2449	2.449	8	4
			'Dose'	'UNK1'	251.405	94.778	156.627	154.095	5.026	2.494
			'2ª Dose' 'Revacinação'	D2	5.701.310	108.780	5.592.530	5.554.620	75.390	37.480
			'3ª Dose' 'Reforço' '1º Reforço' '1ª Dose Revacinação'	"D3"	3.014.770	90.693	2.924.077	2.876.093	94.617	46.633
			'2º Reforço' '2ª Dose Revacinação'	"D4"	886.510	20.251	866.259	863.272	5.686	2.699
			'3º Reforço' '3ª Dose Revacinação'	"D5"	7.560	161	7.399	7.390	15	6
			'Dose Adicional'	'UNK2'	97.931	2.611	95.320	93.358	3.912	1.950
			'4ª Dose Revacinação'	'UNK3'	677	18	659	658	2	1

Part III key learnings – Big Cities are NOT ‘islands’.

Nr of vaccine doses taken in Florianópolis (c18 = 420540) from residents, officially, from other cities & regions (c13)

CoC_18	Ntbk_07	Ntbk_07	Ntbk_08	Ntbk_08	Ntbk_9
'FLORIANOPOLIS' summary 31 Oct 22	c18 IBGE code = 420540 FLORIANOPOLIS	C18 = 420540 with c13 CEP NOT IN RANGE 88000 to 88099	c13 Registers with CEP range 88000 to 88099 ALL c18 codes	c13 CEP range 88000 to 88099 NOT IN c18 = 420540	c13 CEP range 88000 to 880099 WITHIN c18 = 420540
Registers	1.319.335	413.063	998.452	320.883	906.272
Individuals	581.915	205.551	418.072	163.843	382.669



c29_v_descricao_dose / Brasil_Regiao	N	CO	NE	SE	S	UNK	Total
D1	5,767	4,432	6,738	17,034	90,038	32,763	156,772
D2	5,185	4,213	6,312	16,290	83,369	29,929	145,298
D3	2,757	2,370	2,781	9,652	44,427	15,667	302,070
D4	638	747	685	3,262	15,624	6,345	27,301
D5	1	2	1	11	50	8	73
UNK1	125	156	128	763	2,820	1,003	27,374
UNK2	28	28	15	99	585	215	970
Total	14,501	11,948	16,660	47,111	236,913	85,930	413,063

South Region (minus Floripa_c13)	RS	PR	SC	Total
D1	26,009	9,331	54,698	90,038
D2	23,684	8,523	51,162	83,369
D3	13,057	4,673	26,697	173,407
D4	4,366	1,293	9,965	15,624
D5	13	2	35	50
UNK1	875	348	1,597	15,674
UNK2	147	47	391	585
Total	68,151	24,217	144,545	236,913

Part IV – Key
learning –
Pediatric
Population
SOMETHING IS
NOT RIGHT IN
FLORIANOPOLIS

		PEDIATRICS D1 - IN and OUT						
	c35_SC_UDVE_Nr	Estimated Pop 5 to 11 yo	TOTAL VACCINE D1_DOSES SUPPLIED BY_DIVE IN	TOTAL VACCINE D1_DOSES APPLIED PER_OPDS OUT	OPDS_OUT_D1 TRUE_PEDIATRIC (c30 = 98 or 99)	OPDS_OUT_D1 NOT_TRUE_PEDIATRIC (c30 = 86 or 87, or other)	Pct TOTAL_D1 OPDS_OUT per DIVE_IN	Pct OPDS_D1_OUT TRUE_PEDIATRIC per DIVE_IN
	udve01	14,416	15,480	11,094	6,879	4,215	72%	44%
	udve02	15,427	14,570	9,167	4,595	4,572	63%	32%
	udve03	39,316	46,860	25,617	15,081	10,536	55%	32%
	udve04	11,646	13,340	8,998	5,284	3,714	67%	40%
	udve05	17,638	19,740	10,665	6,465	4,200	54%	33%
	udve06	28,391	27,580	18,706	11,416	7,290	68%	41%
	udve07	36,543	37,870	21,530	9,041	12,489	57%	24%
	udve08	26,748	20,670	6,437	5,160	1,277	31%	25%
	udve09	71,712	67,710	30,904	23,040	7,864	46%	34%
1st SC most ppulated city	Joinville	54,059	51,070	22,080	19,915	2,165	43%	39%
	udve10	67,704	61,250	29,345	17,485	11,860	48%	29%
	udve11	27,486	27,240	11,677	7,824	3,853	43%	29%
	udve12	70,195	63,500	27,145	17,758	9,387	43%	28%
	udve13	100,796	106,980	60,534	14,305	46,229	57%	13%
2nd SC most populated city	Florianopolis	38,297	42,690	22,320	349	21,971	52%	1%
	udve14	26,710	26,300	15,323	8,330	6,993	58%	32%
	udve15	31,349	33,750	18,161	10,267	7,894	54%	30%
	udve16	38,126	32,750	16,961	10,830	6,131	52%	33%
	udve17	18,595	18,010	7,547	4,800	2,747	42%	27%
	Total	642,798	633,600	329,811	178,560	151,251	52%	28%

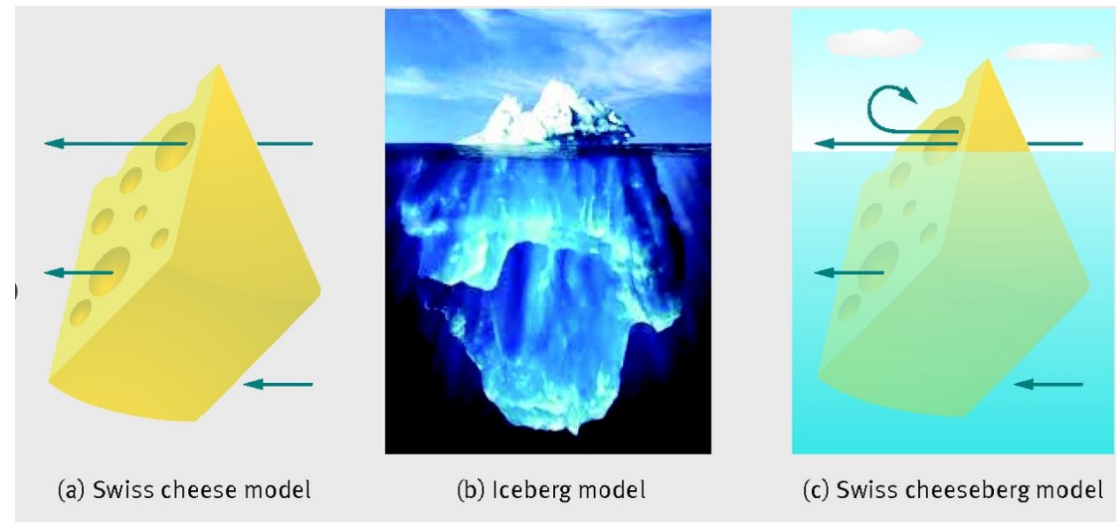
HEALTHCARE DATASETS

COVID-19 PANDEMIC ERA AND NEAR FUTURE

The holes are
getting much
bigger.

Three potential causes increasing the 'SWISS CHESSBERG' effect

- Healthcare staff reduction:
- Global disruptions of Medical Supply and 'Gold Standards'
- Short to medium term effect of Cyber-attacks:



Please check AR02 paper's Part IV for more details

Conclusions

Proof of Concept

- The simple triad, Google Collab IDE & Pandas and Python passed my 'proof of concept' as potential tools to be used for ClinOps managerial purposes for clinical trials.
- With discipline and collaboration from Regulatory and QA departments we can make them better suited for ICH & GCP & ALCOA+ requirements and therefore acceptable by regulatory authorities.

Did you said automate?

- *'Python's ecosystem to automate a Clinical Risk Based Monitoring..'* Did that happen?
- The overall project code strategy must spot and warn about any dataset changes since last run.
- Triad Google Collab IDE & Pandas & Python allowed a 'semi-automation and makes the code maintenance and replication from notebook or project iteration to the other much easier.

- THANK YOU FOR YOUR ATTENTION

Dr Siddhartha Mukherjee
in 'Eric J. Topol.
Emperor's Mukherjee on 'Three Laws of Medicine'.
Medscape, Oct 12, 2015'

*'The Science of Medicine
tells us that
uncertainty, imperfections, priors, outliers, and biases,
are endemic.*

*They are not going to go away today,
tomorrow or 50 or 100 years from now.*

*The sooner we instil these ideas
in ourselves, and our patients,
the more likely it is that we will have a much more
collaborative effort to solve the major problems of
human disease.*

We need to do that.'

'Interpretable Machine Learning –
A Brief History, State-of-the-Art and Challenges'
Christoph Molnar, Giuseppe and Bernd Bischl

*'..., ML models are usually
not used in a static and isolated way,
but
are embedded in some process or product,
and interact with people.*

*A more dynamic and holistic view of the entire process,
from data collection
to the final consumption of the explained prediction
is needed.*

*This includes thinking
how to explain predictions to individuals
with diverse knowledge and backgrounds and
about the need of interpretability
on the level of an institution or society in general.'*