

Data sharing is the future.

With a FAIR database of individual patient experiences

Nikola Biller Andorno
Giovanni Spitale
Institute of Biomedical Ethics and History of Medicine



AIMS





Institute of Biomedical Ethics and History of Medicine

- 1. Do we need a shared data structure?
- 2. Understand the structure and properties of our data
- 3. Detail our data infrastructure
- 4. Present our in/out data flows
- 5. Discuss about future possibilities

Do we need a shared data structure?

Proposal (from June 2021):

Open a discourse on the data access model, keeping in mind cost of the data collection, possible exploits for commercial purposes, and infrastructure;

Start to reason on the feasibility of an international "meta-database", i.e. a harmonized catalogue of existing datasets and a set of rules for developing new datasets;

Reflect on the criteria for retrospective harmonization;

Leverage on existing experience - from the COVID study but also from other fields, e.g the Maelstrom catalogue (epidemiology)

https://www.maelstrom-research.org/page/publications

1. A shared data structure

It makes sense if:

- We want to increase interoperability, data exchange, data pooling and collaborations;
- We want to foster secondary research, teaching, and education;
- We want to probe new dissemination strategies;
- We want to set common quality standards;
- We want to reduce development and maintenance costs.



It does not make sense if:

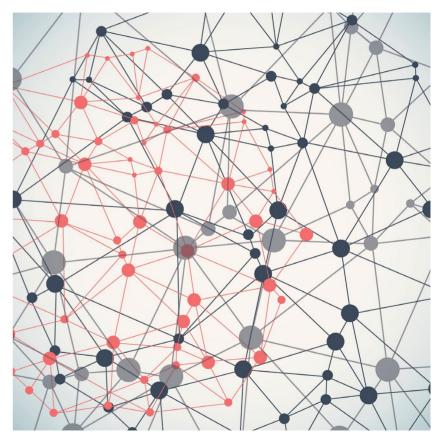
- We are jealous;
- We are shy;
- Other reasons I cannot think of.



1. A shared data structure

It requires harmony:

- Data are generated in the same way;
- Data have the same shape;
- Data have the same classification properties;
- Data are pulled for similar purposes.



That's what I see in my wildest dreams



OPEN question 1:

Do we want to increase interoperability, data exchange, data pooling and collaborations; foster secondary research, teaching, and education; probe new dissemination strategies; set common quality standards; and reduce development and maintenance costs?



OPEN question 2:

Are our data harmonic? Are they generated in the same way, with the same shape, with the same classification properties?

Do we use them for similar purposes?

Our data



Individual Patient Experiences
Subjective narratives in which
participants tell their own
experience from their own point of
view

Database

Data are indexed and structured,
accessible to multiple actors for
different purposes

Local branch
One among many!

DIPEX.Ch
Gesundheitserfahrungen
Expériences de Santé
Esperienze di Salute
Health Experiences

DIPEx International is a network comprising research groups from 14 countries.

We all follow the same methodology and generate similar data.



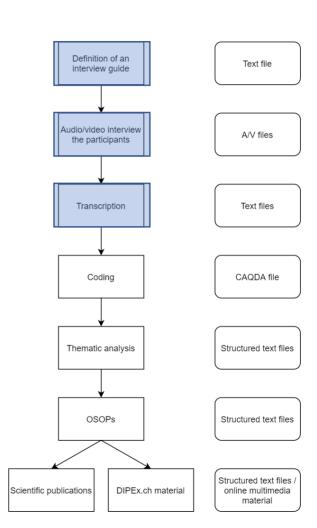


Interview guide

Semi-structured document listing the questions to ask and the prompts to give to the interviewee. Starts with an open section then follows specific topics of interest.

Interview files

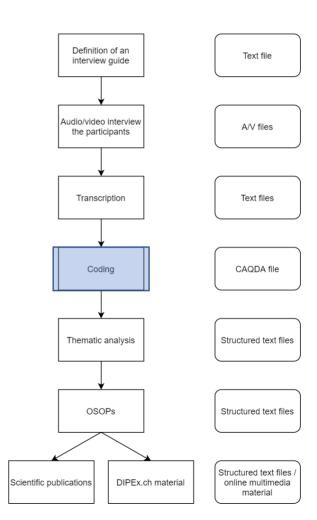
The interview is audio and/or video recorded – according to the preferences of the itwee – and then transcribed as text.





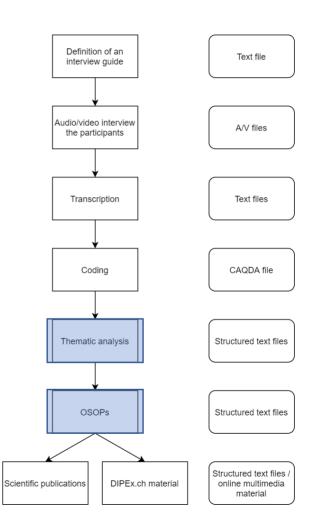
Coding

The interviews are loaded in a software for computer assisted qualitative data analysis. We define a specific coding tree and manually code the text (= assing one or more labels to a meaningful passage)



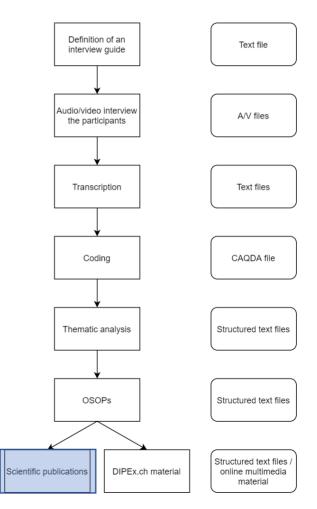


Thematic analysis and OSOPs
We select specific topics that allow us to tell the 'collective history' of a given experience weaving individual voices together. We attribute codes to topics, retrieve the quotes, and put the story together.





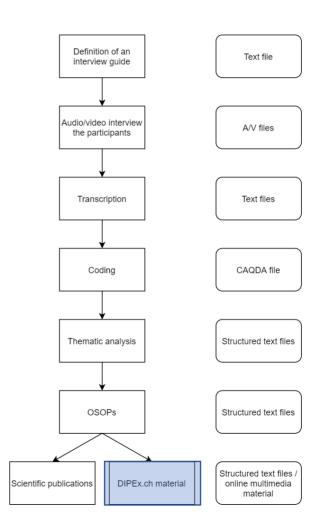
Year/Publication	+ Language	\$
2020		
Duff, C., Kokanović, R., Flore, J., Thomas, SDM., Callard, F. & Blackman, L. (2020). Perspectives on person-centred care for borderline personality disorder: A critical research agenda. Health Sociology Review, 29(1), 1-15. https://doi.org/10.1080/14461242.2020.1715815 ①	English	
Davies, L., LeClair, K. L., Bagley, P., Blunt, H., Hinton, L., Ryan, S., & Ziebland, S. (2020). Face-to-Face Compared With Online Collected Accounts of Health and Illness Experiences: A Scoping Review. Qualitative Health Research, 30(13), 2092–2102. https://doi.org/10.1177/1049732320935835	English	
Flore, J., Kokanović, R., Duff, C. & Callard, F. (2020). The antidepressant in women's lifeworlds: Feminist materialist encounters. BioSocieties. https://doi.org/10.1057/s41292-020-00189-2	English	
Ghio, D., Muller, I., Greenwell, K., Roberts, A., McNiven, A., Langan, S.M. & Santer, M., 2020. 'It's like the bad guy in a movie who just doesn't die': A qualitative exploration of young people's adaptation to eczema and implications for self-care. British Journal of Dermatology, 182(1), 112-118. https://doi.org/10.1111/bjd.18046		
Holmberg, C. & Breuning, M. (in print). Personal Experiences of Illness. In S. Scrimshaw, S. Lane, R, Rubenstein & J. Fisher (Eds.). The SAGE Handbook of Social Studies in Health and Medicine (2nd ed.). Sage.	, English	
Johnston-Ataata, K., Flore, J., Kokanović, R., Hickey, M., Teede, H., Boyle, J. & Vincent, A. (2020). 'My relationships have changed because I've changed': Biographical disruption, personal relationships, and the formation of an early menopausal subjectivity. Sociology of Health and Illness, 42(7), 1516-1531. https://doi.org/10.1111/1467-9566.13143	l English	
Locock, L., Graham, C., King, J., Parkin, S., Chisholm, A., Montgomery, C., Gibbons, E., Ainley, E., Bostock, J., Gager, M., Churchill, Dopson, S., Greenhalgh, T., Martin, A., Powell, J., Sizmur, S. & Ziebland, S. (2020) Understanding how front-line staff use patient experience data for service improvement: An exploratory case study evaluation. Health Services and Delivery Research, 8(13). https://doi.org/10.3310/hsdr08130	_	
Locock, L., Montgomery, C., Parkin, S., Chisholm, A., Bostock, J., Dopson, S., Gager, M., Gibbons, E., Graham, C., King, J., Martin, Powell, J., & Ziebland, S. (2020). How do frontline staff use patient experience data for service improvement? Findings from an ethnographic case study evaluation. Journal of Health Services Research & Policy, 25(3), 151–161. https://doi.org/10.1177/1355819619888675	_	
O'Brien, N., Law, S., Proulx-Boucher, K., Menard, B., Skerritt, L., Boucoiran, I., Cox, J., Andersson, N. & de Pokomandy, A. (2020). Co-designing care improvements for women living with HIV: A patient-oriented deliberative dialogue workshop in Montréal, Québec. CMAJ Open, 8(2) E264-E272. https://doi.org/10.9778/cmajo.20190158	English	





Website material

Based on the OSOPs, we prepare selected material to be put online and serve as an important resource for patients, relatives, caregivers, healthcare professionals, and students.



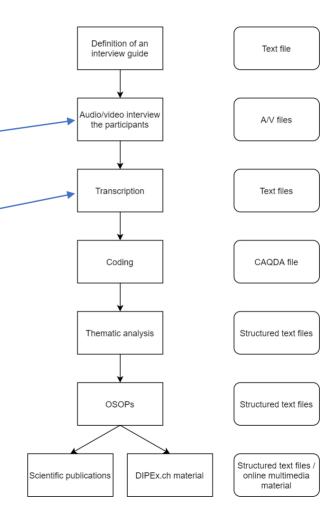


1st informed consent

The interviewee allows the research team to record the interview (A or V). The material is fully available for research and teaching.

2nd informed consent

The interviewee revises the transcript and specifies their preference about the online use (yes / only specific parts / no) + (video / audio /text)





Key message 1

We collect semi-structured interviews of people telling their own experience with selected health conditions. We transcribe and code the interviews, conduct thematic analysis and use the resulting material for publications and divulgation.

Our data are generated in the same way, hence have the same ontological properties.

Our data are analyzed in the same way, hence have the same classification (=relational) properties. Each experience belongs to a topic, which belongs to a category, which belongs to a module. Each experience belongs to one interviewee.

This is how we (CH) do it

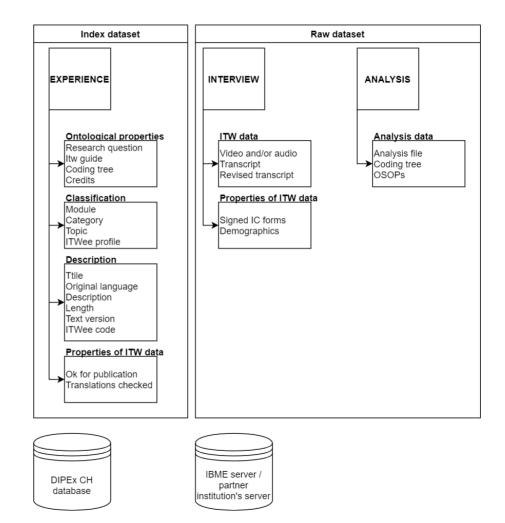
There might be other ways for sure. This is ours, and my guess is that it's gonna be similar to yours.

The point is: if the structure is similar, we might meet the 'harmony' criteria detailed above.



Raw dataset The DNA of our research data; located on IBME's servers

Index dataset
The index and mRNA of our research data (extended metadata); located on UZH's MariaDB instance



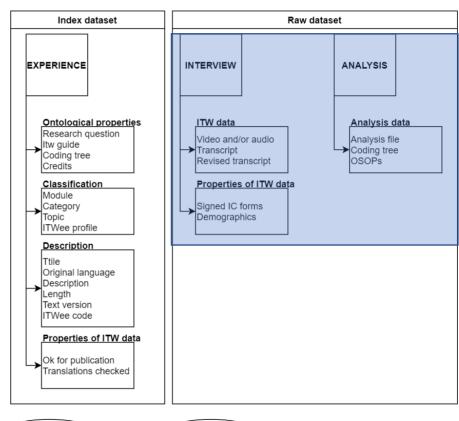


Interviews

Material resulting from the interview: original audio / video; transcript; revised transcript; informed consents; demographic variables of the interviewee

Analysis

Analysis material regarding all the interviews: CAQDA master file, coding tree, OSOPs





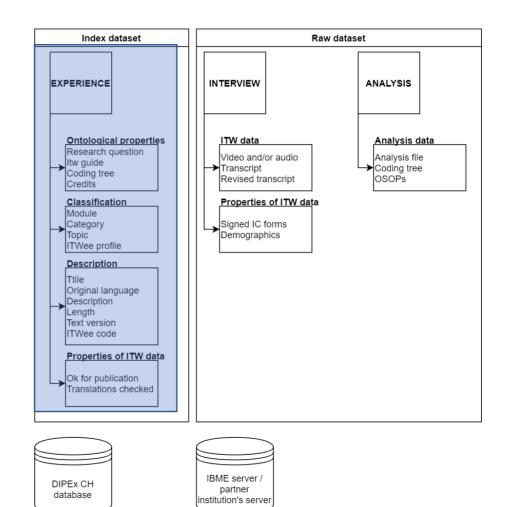




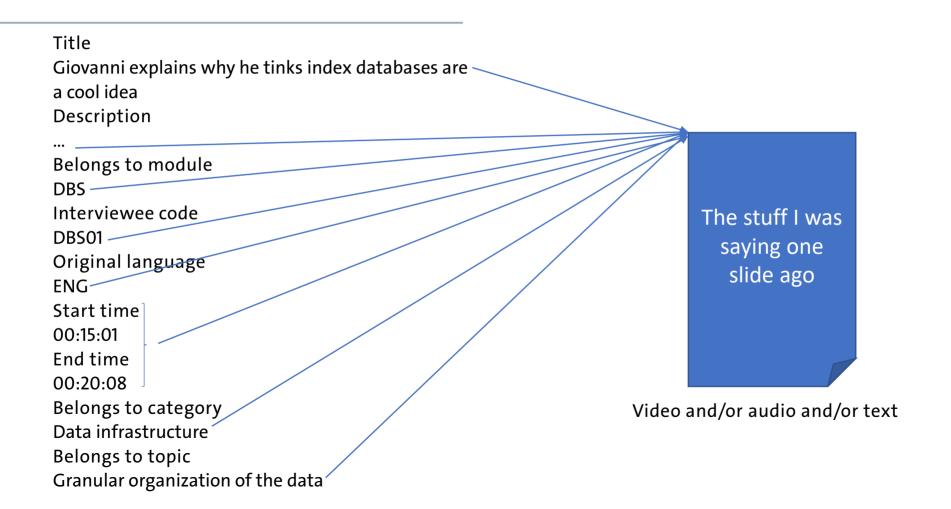
Experience

Minimal viable entity of the dataframe, defined as one narrative block with thematic coherence.

Each experience has ontological properties (= how it was generated/extracted); classification properties (=triple set module/category/topic); descriptive properties; data management properties.



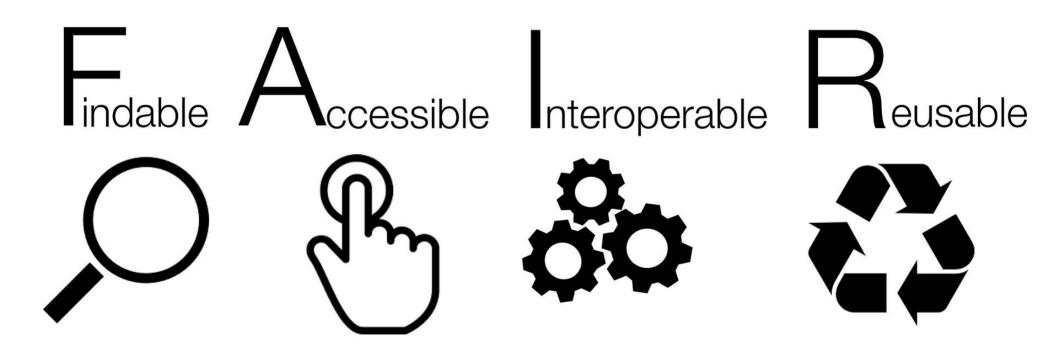






	#	Nome	Tipo di dati	Lunghezza/set	Senza s	Permetti	Riem	Predefinito	Commento	Confronto	Espressione	Virtualità
9	1	ID_Experience	INT	11				AUTO_INCREME				
	2	Belongs_to_module	VARCHAR	3		~		Nessun valore pr				
	3	ITW_code	VARCHAR	100		~		Nessun valore pr				
	4	ITWee_code	VARCHAR	5		~		Nessun valore pr				
	5	ITW_original_langu	VARCHAR	3		~		Nessun valore pr				
	6	EXP_start_time	TIME			~		Nessun valore pr				
	7	EXP_end_time	TIME			~		Nessun valore pr				
	8	EXP_duration	TIME			~		Nessun valore pr			timediff(`EXP	STORED
	9	EXP_video_link	VARCHAR	255		~		Nessun valore pr				
	10	EXP_thumbnail	VARCHAR	255		~		NULL				
	11	EXP_category	VARCHAR	500		~		Nessun valore pr				
	12	EXP_topic	VARCHAR	1000		~		Nessun valore pr				
	13	EXP_name_ENG	VARCHAR	500		~		NULL				
	14	EXP_name_DEU	VARCHAR	500		~		NULL				
	15	EXP_name_FRE	VARCHAR	500		~		NULL				
	16	EXP_name_ITA	VARCHAR	500		~		NULL				
	17	EXP_description_ENG	TEXT			~		NULL				
	18	EXP_description_DEU	TEXT			~		NULL				
	19	EXP_description_FRE	TEXT			~		NULL				
	20	EXP_description_ITA	TEXT			~		NULL				
	21	EXP_textversion_ENG	TEXT			~		NULL				
	22	EXP_textversion_DEU	TEXT			~		NULL				
	23	EXP_textversion_FRE	TEXT			~		NULL				
	24	EXP_textversion_ITA	TEXT			~		NULL				
	25	Translation_OK	TINYINT	1		~		NULL				
	26	Publication_OK	TINYINT	1		~		Nessun valore pr				

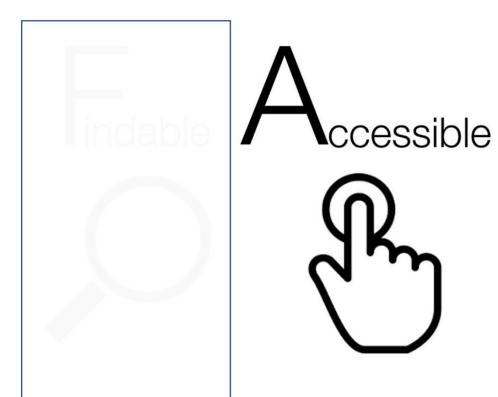






Findable data

Everyone (people and machines) with access to the database can navigate the metadata; can know what data we have; can search them using any of the fields building complex queries.



Accessible data

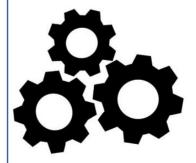
The communication protocol is free and open; it allows (different levels of) authentication and authorization; the metadata are always available.



Interoperable data

Data use open formats (as much as possible) = shared language for knowledge representation; data include references and pointers = from where the data come from, where the data are used.

nteroperable



eusable



Reusable data

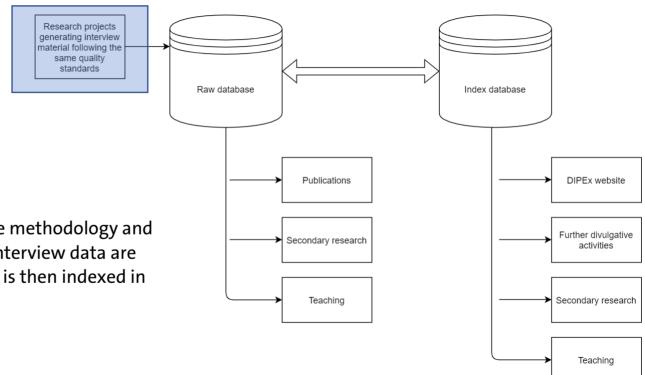
The ultimate goal of FAIR is to optimise the reuse of data. Metadata and data are well-described so that they can be replicated and/or combined in different settings (by people and machines).





Key message 2

We store data and metadata under FAIR principles, building a future-proof data infrastructure which allows people and machines to access the data.



Data in:

Interview data generated with the same methodology and with the same quality standards. Raw interview data are stored in the raw dataset. Their content is then indexed in the index database.





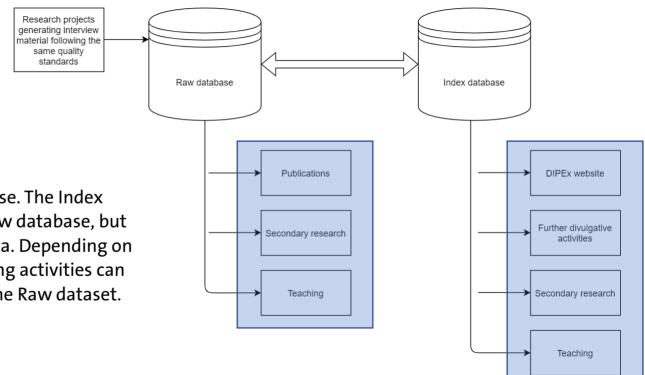
This table lists all the available combinations of categories and topics for the selected module.

It does not allow editing. If you want to modify a category or a topic, please use the appropriate form.

Please consult it before filling in the related fields in the drop-down menus on the left $\!\!\leftarrow$

Module	Category	Topic
CMI	Communication	ways to communicate
CMI	Communication in the ICU	trust and safety
CMI	Communication in the ICU	Team and family support
CMI	Communication in the ICU	Willingness and time to communic
CMI	Communication in the ICU	possibilities of misunderstanding
CMI	Communication in the ICU	DEEPL: The general perception that
CMI	General perception / overview of the i	DEEPL Environmental perception.
CMI	General perception / overview of the i	DEEPL General perception of relativ
CMI	State of consciousness and memory is	DEEPL Forgetting / remembering v
CMI	State of consciousness and memory is	Delusion/hallucination/ dreams/ Fe
CMI	State of consciousness and memory is	DEEPL Waking up in ICU
CMI	State of consciousness and memory is	Coma/Sedation





Data out:

Data can be explored via the Index database. The Index database contains the metadata of the Raw database, but also organized sections from the same data. Depending on the purpose, secondary research or teaching activities can use data from the Index dataset or from the Raw dataset.



Create minified dataframe for video editor

cols = ['ID_Experience', 'ITW_code', 'EXP_name_ENG', 'EXP_start_time', 'EXP_end_time', 'Consent']
df = experience_df_subset[cols]
df

Video
Audio
Video
Video
Audio
Audio
Audio
Audio
Audio
Text

Text

Video

Audio Audio Audio

Text

Audio

Video

NaT

]:		ID_Experience	ITW_code	EXP_name_ENG	EXP_start_time	EXP_end_time
	3	13	CMI03_05122019_EN	Charles C talks about the efficiency of the ICU	0 days 00:00:20	0 days 00:01:31
	4	14	CMI17_29052020_FR	For Mr. Quentin Q. everything was quick and co	0 days 00:02:49	0 days 00:04:15
	5	15	CMI01_22102019_FR	Mrs. Anna A. remembers the noise, the lights.	0 days 00:00:07	0 days 00:02:18
	6	16	CMI07_11122019_FR	Mr. Georges G. praises the professionalism of	0 days 00:01:54	0 days 00:03:06
	7	17	CMI13_10032020_DE	Mr Markus M always felt secure.	0 days 00:16:04	0 days 00:17:20
	10	20	CMI04_05122019_FR	Mrs. Deborah D. experienced a positive atmosph	0 days 00:09:48	0 days 00:10:56
	11	21	CMI05_05122019_FR	Mrs. Eveline E sometimes felt unnoticed	0 days 00:08:12	0 days 00:08:59
	12	22	CMI03_05122019_EN	Mr. Charles C. used headphones against noise.	0 days 00:13:35	0 days 00:14:52
	13	23	CMI06_11122019_FR	Mr. Fabio F. hears and sees, but accepts the u	0 days 00:12:19	0 days 00:13:24
	14	24	CMI09_12012020_DE	Mr. IGOR I. remembers the ceiling of the room \dots	0 days 00:11:23	0 days 00:12:22
	15	25	CMI01_22102019_FR	Mrs. Anna A. mentioned the noise by other pati	0 days 00:04:24	0 days 00:05:15
	16	26	CMI21_10022020_DE	Mrs Yvonne Y. would have liked more support.	0 days 00:33:13	0 days 00:33:48
	17	27	CMI21_10022020_DE	Mrs. Yvonne Y needed support for herself.	0 days 00:14:28	0 days 00:15:32
18	18	28	CMI12_12022020_DE	Ms. Lena L. suffered from the long waiting tim	0 days 00:14:39	0 days 00:15:56

Mr. Urs was distressed by his wife's situation...

Mr. Clau C. was dissatisfied with his wife's c... 0 days 00:05:27 0 days 00:06:49

Mrs. Anna A talks about her hallucinations. 0 days 00:32:02 0 days 00:32:45

Preparing lists of video and audio to edit for selected experiences ...in 2 lines of code

19

20

21

CMI25_30092020_DE

CMI01_22102019_FR

30 CMI19a_06032020_DE



Filter experiences by module

This creates a slice of the experience dataframe based on the selected module.

print(experience_df['Belongs_to_module'].value_counts()) module = input("specify the sortname of the module you want to work with (e.g. AAA):") experience_df_subset = experience_df[experience_df[Belongs_to_module'] == module] display(experience_df_subset) CMI 41 TST 4 DEM 1 MSC PDI 1 COV Name: Belongs_to_module, dtype: int64 specify the sortname of the module you want to work with (e.g: AAA): CMI ID_Experience Belongs_to_module ITW_code ITWee_code ITW_original_language EXP_start_time EXP_end_time EXP_duration 0 days 0 days 0 days 3 13 ENG CMI03 05122019 EN CMI03 00:00:20 00:01:31 00:01:11 0 days 0 days 0 days CMI17 FRE 14 CMI17_29052020_FR 00:04:15 00:02:49 00:01:26 0 days 0 days 0 days 5 15 CMI01_22102019_FR CMI01 FRE 00:00:07 00:02:18 00:02:11 0 days 0 days 0 days 16 CMI07_11122019_FR CMI07 FRE 00:01:54 00:03:06 00:01:12

Getting all the experiences belonging to a module ...in 3 lines of code



Filter experiences by topic

This creates a slice of the dataframe containing only the experiences belonging to a specific topic

[23]:	3]: experience_df_subset[experience_df_subset['EXP_topic'] == 'Delusion/hallucination/ dreams/anxiety.']								
[23]:		ID_Experience Belongs_to_module		ITW_code	ITWee_code	ITW_original_language	EXP_start_time	EXP_end_time	EXP_duration
	21	31	СМІ	CMI01_22102019_FR	CMI01	FRE	0 days 00:32:02	0 days 00:32:45	0 days 00:00:43
	22	32	СМІ	CMI03_05122019_EN	CMI03	ENG	0 days 00:05:35	0 days 00:07:20	0 days 00:01:45
	23	33	CMI	CMI08_16122019_DE	CMI08	DEU	0 days 00:06:06	0 days 00:07:06	0 days 00:01:00
	24	34	СМІ	CMI05_05122019_FR	CMI05	FRE	0 days 00:27:48	0 days 00:29:16	0 days 00:01:28
	25	35	СМІ	CMI10_15012020_DE	CMI10	DEU	0 days 00:06:43	0 days 00:08:20	0 days 00:01:37
	26	36	СМІ	CMI20_29022020_DE	CMI20	DEU	NaT	NaT	NaT
	27	37	СМІ	CMI23_03072020_FR	CMI23	FRE	0 days 00:20:30	0 days 00:21:18	0 days 00:00:48

Getting all the experiences belonging to a topic ...in 1 line of code



Text version of selected experiences

This creates a list containing all the english text of the selected experiences

- [29]: experience_df_subset['EXP_textversion_ENG'].tolist()
- ["Hallucinations, I saw myself in a dark cave, yes, images like that. Sometimes yes, caves with branches, earth, dark and yself no, there's nothing like that, you're just in the hospital in intensive care, you're getting your ceiling back. That's it.", And there, it was a more difficult thing. I was in a coma for four days and I was losing weight and it was difficult to feed sense of smell and of taste, which I reckon was augmented a hundred times. So, one little grain of sugar in water was in und me, not from a directional point of view, but that was very difficult. And, the only problem I had there then, was getti the taste was too strong. And each time I would take a glass of even water, or orange juice, I would be vomiting, because before I really managed to convince someone that I wasn't exaggerating. And we think that the reason was the morphine "In every dream I was lying somewhere, I could not move. \n I was lying in a room, I uh / so, mostly in a room or in a gar. ss. Fear. Not being able to escape from the situation. So, I was just lying there. Whether it was in a restaurant or in a hos 'Well, talking like that, like they were talking to those two girls, I think that - and not talking to me; I thought I was next to ey were doing / talking about their classes, their stuff and I was following. And I thought to myself that it\'s true that ever ut I don't have a bad memory. \nI: So you always think that people, even with their eyes closed, can hear? \nE: Yes. And nderstand. \n\nI: And afterwards, did you understand what was going on with the famous picnic, what it could be, that it at they were having a picnic, then that they were going to eat, then that they were going to leave. But I thought they were "I: Tell me about the dreams. What were they like?\n\nE: They were very likely things that I heard from my family. From v wanted to fly me to the health institution (place_3). Rega (Swiss Air Rescue) was a topic. Then someone told that there v wnstairs in the Migrolino. Down there was the hospital bed. Rega was refueling down there, waiting for me. Then they pu no light. Then they were able to restart it. Then they drove me to the nursing home to (place_2). Like so to place_2. Then ree bed. It was just in the hallway. Then I had such problems with my lungs and I kept wanting to tell the doctor that, but s. For whatever reason a cow. I don't know. But, funnily enough, it was always the same doctor.", " I: And how many of your dreams were reality? \nP: Unfortunately, more than I thought, which scared me the most. I als
 - atients. And so I thought for days that this was a dream and suddenly he walks past my room. All the doors are always c ou in some way? \nE: No, I don't know. I talked to him too, but only a day later or so. And I apologized, although I don't kn for? But then I didn't ask any further. But he was maybe a bit rough or too strict with me. I think he was also there the nig find that bad. I found it more reassuring and she was worried about me and yes, she only meant well.\n",

Getting all the text (in English) of the selected experiences ...in 1 line of code



Applying a simple NLP pipeline to the selected text ...looks like in 2 line of code but I'm actually cheating (they are 43)

Lemma frequency and word cloud ¶

This tokenizes the selected text, calculates lemma frequencies, and creates a word cloud.

#Display lemma frequencies of first 10 lemmas in corpus
display(lemma_df.head(10))
Generate word cloud from lemmas
generate_wordcloud(lemma_df, "lemma", "count", 50, "white")

	lemma	count	frequency
0	like	11	4.7210
1	think	9	3.8627
2	thought	8	3.4335
3	know	8	3.4335
4	coma	5	2.1459
5	yes	5	2.1459
6	talking	5	2.1459
7	room	5	2.1459
8	dreams	4	1.7167
9	lying	4	1.7167



Key message 3

We defined rules for incoming data (quality standards) and procedures for outgoing data. Being the system flexible and open, multiple strategies for outgoing data are possible.

5. Future possibilities



5. Future possibilities

New partnerships to increase incoming data

Qualitative data are a useful lens to study complex phenomena, and proved crucial in improving person centered care, patient empowerment, and HCP training. More data (and better data) are a crucial need to generate evidence based improvement in these areas.

Do we need it? Do we want it? Let's stress test the concept and the structure with limit cases!

A database infrastructure allows to conceive multiple interfaces, specific for their purpose.

Currently we have one data out interface (DIPEx.ch). We could 1. probe the needs (of patients, students, relatives, researchers ...) and 2. build new specific interfaces.

New interfaces for outgoing data

Key message 4

We could use a common structure for our 'index datasets'. This would allow:

- Internal interoperability and data sharing;
- Only the data which are authorized for publication are exposed;
- No risk to compromise or expose the original data ('raw dataset');
- Endless possibilities to develop new distribution tools;
- Enhanced possibilities for secondary research;
- New collaborations (inside and outside DIPEx International).

Plus:

- Cutting to 1/14th the total cost of data infrastructures;
- Cutting to 1/14th the total cost of website development and maintenance;
- Lowering the financial threshold for new members (fair FAIR data ☺).

THANKS FOR YOUR TIME!



