1:53

Alexey

**This week, we'll talk about technical writing and data journalism. We have a special guest today, Angelica. Angelica is a researcher at the Institute of Informatics and Telematics in Italy. Her research interests include data science, machine learning, text analytics, data visualization, data journalism, web applications, and recently she became also interested in data engineering. She is also a professor at the University of Pisa, where she teaches data journalism. Welcome, Angelica.**

2:22

Angelica

Hi, nice to meet you. For me, it's a pleasure to be here with you today.

# Angelica’s background

2:29

Alexey

**Yeah, likewise, Thanks for accepting the invite to be here. I also want to mention that the questions for today's interview are prepared by Johanna Bayer. Thanks, Johanna, for preparing the questions. Angelica, before we go into our main topic of writing and data journalism, I wanted to ask you about your background. Can you tell us about your career journey so far?**

2:52

Angelica

Yes. Indeed, since I was a teenager, my big dream was to become a computer scientist. Because I had a computer (this was the late 90s period) and I decided to understand how a computer works. And so my dream continued – in university, I studied computer science. Then I devoted all my career to the research field. In fact, now I’m a researcher. At the beginning, I worked on network security with the focus on cryptography, reputation, and similar stuff. I got my PhD in this field.

But after my PhD, I realized that web applications were more appealing to me than computer security, so I moved to this field. Next, I landed in data science, and I worked on data science projects for many years. Only recently, I'm moving to data engineering, because I like to explore new things. As you can see, my career is slightly different than that of other people, because I always remained in the research field. I changed only the topic that I studied. For me, I can see that I’m a privileged person because I can decide the topic to study and this is a very interesting aspect of academia and research.

4:41

Alexey

**Just curious, what are you researching now in data engineering? To me, it looks like engineering topics are more practical. It's more about know-how and less about research. But maybe I'm a bit far from that. So I'm really curious – what kind of research topics are there in data engineering?**

5:03

Angelica

Yes, indeed. I'm studying the topic and so I don't know if I will have an idea how to improve something. But given my expertise, a possible idea could be to add some security aspects to the data – to maybe help recognize if data is manipulated or something like this. I don't know. I'm studying the topic at the moment. I'm also writing a book about this topic. So we will see.

5:47

Alexey

**Okay. That's quite on-topic to today's conversation, right? Writing. You are learning data engineering by writing a book about that?**

5:56

Angelica

Yes.

# Angelica’s books

5:57

Alexey

**That's quite a good way to learn something. [chuckles] [Angelica agrees] You already have a few books, right? So you have at least one that I know of, which is Comet for Data Science [Angelica agrees] Have you written any other books?**

6:14

Angelica

In the past I've written some novels, but I was very young – some novels, short tales, but nothing interesting. Now I'm writing a book about Presto, the very famous database also used by Meta. Yes.

6:41

Alexey

**I think they changed their name recently, right? So it's not Presto anymore. Or is it two different things?**

6:44

Angelica

No, no, no. The original name was Presto and then they forked into two different databases, which are Presto and Trino. Now there are two different projects supported by two different foundations.

7:07

Alexey

**Ah, okay. At our company, OLX, we used to use Presto and then one day, I connected to Presto and it said Trino. So I was like “Okay, what happened?” [laughs] Then it turned out that our data engineers updated the version and it became Trino.**

7:25

Angelica

They are quite similar, but they derive from two different organizations.

7:32

Alexey

**I guess it's like LibreOffice and OpenOffice, right? They have the same history, but then at some point, they branched away from one another.**

7:39

Angelica

Yes. I think so.

# Data journalism

7:43

Alexey

**One of the topics that I wanted to talk to you about is data journalism. This is one of your research interests. You're also a professor at the University of Pisa, when you teach this. [Angelica agrees] So what is data journalism?**

8:01

Angelica

Yes, data journalism is data-driven journalism in the sense that, similar to any data science project, it collects, analyzes, filters data, to create interesting stories. Data journalism is slightly different from data storytelling or data narrative, which are other fields because data storytelling, for example, is the art of telling stories from data. They could seem like the same thing, but they are not, because the objective of data journalism is to tell the news and to use data to confirm the story it is telling. For example, you can see long articles that talk about something and then they use data to confirm the story which is narrated.

Additionally, for example, with respect to data science, which is another aspect – in a data journalism story, you usually don't have the analysis part, which is a big part in data science. Usually you don't use machine learning or data analysis models in general – you don't train a model in data journalism. The huge part is to build a storyboard and to collect data. While in data science, you use big data, in data journalism, you also focus on small datasets because the idea is to have accurate information.

You have to tell the truth and you cannot approximate things. For example, if you train a model in data science or in machine learning, this model has an accuracy, for example, of 80%. But in data journalism, you have to say true news – you cannot have an accuracy of 80%. It should be 50%. This is the main difference. There are also other differences. But to give an overview, this is the first aspect.

10:56

Alexey

**Can you give an example? The articles I'm reading in the economist or similar journals or newspapers – are they doing data journalism? Or is it just regular journalism?**

11:11

Angelica

No, they’re not. In general, I tell my students to build data journalism stories but in practice, I don't write my data journalism stories. I help my students to build their stories, to collect the data. There are many websites. For example, one interesting website is the Washington Post website.

If you search on Google for “The Washington Post data journalism,” you will find their group (their laboratory). They write wonderful stories and you can learn how to organize a story. But usually almost all the newspapers have a dedicated data journalism laboratory. Very specific.

12:29

Alexey

**It's not about fact checking, right? It's different. It's more about supporting your agreements with data.**

12:34

Angelica

Yes.

12:35

Alexey

**I'm just trying to think of an example. Can you think of anything? Something you’ve maybe read recently? I know that Musk bought Twitter. Have you seen any stories about that, which are a good example of data journalism?**

12:52

Angelica

Yes. For example, there was an article by the Washington Post, which talks about the barriers (the walls) that were built in 2015. And to avoid that, migrants went from Southeast Asia to Europe. They use some data and graphs (visualizations) to show how this process happened across the years. If you want, after, I can send you the link. But at the moment, I don't have it. It's a very, very impressive story. Other stories analyze, for example, how oppressors in some lands use the lands to oppress the local people – to build coffee plants and similar stuff. They use the data to prove this.

14:34

Alexey

**Yeah, interesting. Yes, please do send the link. It would be quite interesting to take a look. What I understood regarding the main difference between usual journalism and data journalism is that, in usual journalism, you have a story, but it's not necessarily based on data. You just present some facts but you don't show diagrams – you don't show graphs, you don't show anything to support the story. While in data journalism, you have some data source, you have some data visualization and then you base your story on the data. Right?**

15:11

Angelica

Yes, I think that the challenging aspect in data journalism is to search for data. Because data is hidden on the web, and data journalists must discover it. Instead, for example, in data science, so you already have data – your company provider provides you with your data, so you have to analyze them. In data journalism, you have to search for new data. Also, small data can be used. It's not important to have big data. The most important aspect is that the data should be as accurate as possible. Otherwise, you deliver fake news.

# How Angelica got into data journalism

16:13

Alexey

**How did you get into data journalism? Is it something that you started your career with? Or is this something that accidentally happened to you? I think you mentioned that you actually researched cryptography, right? [Angelica agrees] Then from doing cryptography, how did you end up doing data driven? They're quite unrelated, right? Or are they?**

16:37

Angelica

No, they are not related. When I moved to web applications, my boss asked me to start a course about web applications. We started with a course at university but then, he had this smart idea to change the course to data journalism. I don't know why. [cross-talk]

17:06

Alexey

**They’re not related at all, right? [chuckles]**

17:07

Angelica

No, not really. [chuckles]

17:09

Alexey

**He just said, “Okay, let's change to a completely random different topic.”?**

17:14

Angelica

Yes. Because he understood that web applications became old, I don’t know. Then he decided to involve me in this new topic. I accepted it. I was surprised by his choice, but usually, he makes good choices. In the end, I think that he was right because I liked this course. Over the years, the course has improved. Also, thanks to the feedback provided by students.

Here at the University of Pisa, there is a mechanism where every year, students provide feedback about each professor and how the course must be improved. Over the years, we have improved this. We are still improving it, but I think now it's quite mature.

# The field of digital humanities and Angelica’s data journalism course

18:33

Alexey

**Who are the students? What kind of students is this course for? Is it for technical students – for students who study computer science and want to learn more about journalism? Or is it for those who are studying less technical topics and want to get into data?**

18:49

Angelica

They are in a particular area, because they are humanists with a background in computer science. They study digital humanities and have knowledge both of the humanities and computer science. It's a new course, which is not everywhere – at least in Italy, it's not everywhere. The name is humanistic informatics, which is the rough translation.

19:27

Alexey

**That’s pretty interesting. To be honest, this is the first time I hear about such a field.**

19:33

Angelica

Yes. This profile is very required in the digital humanities field. For example, to build or work in archives, in cultural heritage, digital collections, and similar stuff.

19:50

Alexey

**How technical is this? From your description, I think you need to be a good writer to write the story. The story has to be there. But how technical should you be, in addition to being a good writer, in order to be a data journalist?**

20:13

Angelica

Yes. There are two options. You may not know anything about programming languages – you may be able to use tools like Tableau. In this case, if you don't use anything, it's okay. Or, if you have a technical background, usually the most popular programming language that students use is Python. But this is at a high level. You don't focus, for example, on heavy programming. You just need to know how to clean some data, how to manipulate them. It's very easy. Everyone can do this. But the alternative could be Tableau, so you don't need to be technical at all.

21:12

Alexey

**Is the course in Italian, or in English?**

21:17

Angelica

Unfortunately, it's in Italian.

21:23

Alexey

**Do you know any other courses that are in English that are open? Let's say if I want to go and learn more about data journalism, do you know any open courses that I can check out?**

21:37

Angelica

I don't know if it's open, but I think there is one course on Coursera. I can send you the link. Maybe you can share it with the community. But I don't know if it's still available. It describes the aspects more from a journalism point of view. Also the books that I have read on data journalism, focus on the journalism and then they describe a little data. Maybe my next book could be on data journalism. [chuckles] But it's a very niche sector, so maybe there will not be an audience for this book. I don't know.

# Technical articles vs data journalism articles

22:38

Alexey

**But as I understood, the audience that you write for in data journalism is the general public. Let's say you write an article for The Washington Post. The readers for this article are just ordinary people, right? They don't have to have a background in a specific area.**

22:57

Angelica

It depends on the topic of your article. If you write an article for The Washington Post, yes – probably the general user is okay. But if you write an article for a children’s magazine, for example, the audience are children. So you have to think about your audience before writing your article. But this is also true when you write a book – a technical book or an article.

Everything you write, before writing, you have to think about the audience. I think this is the first step. What does this mean? It means that you have to focus on the audience’s needs. So you know who will read your book or your article? What are their interests? Which skills do they have? And are you writing for a technical audience or not? This is a very important aspect before writing everything.

24:20

Alexey

**The reason I ask this question is that I know that you do a lot of technical writing. I don't know how many articles you wrote on Medium – it must be hundreds by now, right? [cross-talk]**

24:32

Angelica

70 maybe. [chuckles]

24:35

Alexey

**You're quite a prolific writer. The articles you wrote that I know about, they are technical articles. It's technical writing, which I guess is quite different from the articles we've been talking about so far – data journalism. Would you say that data journalism is not technical writing?**

24:55

Angelica

Yes, it's not technical writing. It's for general audiences. Also, the language you use is different. But there are a variety of languages that you have to use. For example, if you write a scientific paper, you need to write in a certain way. If you write a technical book, you have to use another type of language. If you write for data journalism, you have to write another way. So you should focus on your audience – this is the problem.

25:33

Alexey

**We kind of have two topics then, right? We have data journalism and we have technical writing. We already talked about data journalism, but what is technical writing? What makes writing technical? If we just add code, does it become technical? Or is there more to that?**

25:51

Angelica

I think that when you’re doing technical writing you should talk about something technical, for example, how to set up something, how to perform a task, how to do something. Technical writing differs from, for example, a research article because a research article discusses some concepts more from a theoretical point of view. You can still have some implementation parts, but they demonstrate your ideas.

In technical writing, instead, I think that you describe how to solve a problem. For example, if you write a technical book, this book can be a reference manual where you teach something. Also, the language you use should be sure, in the sense that if you say something, you cannot say, for example, “You can do this.” You have to say, “Do this,” because you are sure of what you are saying. There are these differences.

27:16

Alexey

**So, something like, “How to prepare the environment for a GPU”. That would be a very technical article, right? Because you have a sequence of steps and you have some screenshots, you have pieces of code, and then you just follow this. [Angelica agrees] I guess this one is easy. But let's say, if we’re talking about how to do data visualization. This is also technical, right? [Angelica agrees] You walk the reader through a guide – a how-to? [Angelica agrees] Is it always like that? You just walk the reader through a sequence of things and show how to get to the result they need?**

28:05

Angelica

No. You can also describe, you can report some things, I think. You can provide some examples about something. In my book, Comet for Data Science, I also describe some general concepts to introduce the reader to some concepts.

28:32

Alexey

**Like “What is machine learning,” for example, right?**

# Transforming reports into data storytelling

28:34

Angelica

Yes, for example – the general overview. But it's different from a research article where you have to explain the details about concepts. Another thing, I think that the main difference between research articles and technical articles, is that in research articles you discover something new – you describe something new. In a technical article, instead, you describe existing things.

29:19

Alexey

**Maybe, you know that there are companies like StackOverflow that run surveys. Every year, they ask people (members of StackOverflow) about their interests, about the tools they use, about the tools they want to use, tools they want to learn, the tools they hate, languages they hate, languages they love, their salaries, etc. It's quite a massive questionnaire. I don't know if you ever took part, but maybe you took part in similar ones. I think O'Reilly does this as well. There are quite a few companies that are doing this.**

**So they collect this data, and then one or two months later, they release a report (a PDF). How would you classify this report? Is it data journalism? Is it technical writing? Is it data storytelling? What is it?**

30:10

Angelica

It depends on how the report is written. If you only report data as they are then it’s a report. If you extract a story from your data it can become data storytelling or also a data journalism story. Recently, I have also read a book about data storytelling, which used this example of a Stack Overflow report, and they described how to transform this raw data into a story. This was very interesting. This was a case of a self-published book because the author decided to self-publish and it was a very, very interesting book. Maybe I can send you a title on that so you can share it with the community.

31:33

Alexey

**So let's say we have this report. If we just do basic data analysis – we aggregate data, we show all these pie charts, line charts, bar charts, and all these things and we put them in a PDF – and we say that, “Okay, this is the most popular language, this is the most hated language, this is how much data scientists earn,” then this is just report. [Angelica agrees]**

**But if we write an article about different programming languages, and we say, “Java and C are the most popular programming languages,” and then we add a chart that shows that, that proves, “This is based on this dataset,” and then we maybe talk about the story behind some of these programming languages – then it would be more like data journalism. Right?**

32:25

Angelica

Yes. In addition to that, to transform a report into a data journalism story, for example, you can add context to your data. For example, in the case of programming languages, you can group data by the year – the age of the people that are answering the questions – and you can see, for example, that the 80s generation likes Java. Now you can start searching for why this category of people likes Java more than the other languages. Maybe they studied Java at university, and so on.

Here, you transform your raw data into a story by adding context. Finally, the main objective when you want to transform your data into a story is to add what we call “wisdom,” which is to attach to your data, an ethic, a message, a call to action to your audience – because the final objective of data storytelling, or a data journalism story in general, is to call the audience to action. Then this is not a report anymore. It's a story.

34:24

Alexey

**In a report, do we have a call to action? We could, right? “Share it with your friends.”**

34:33

Angelica

It's not sufficient. [laughs]

34:37

Alexey

**Okay. [laughs] I guess companies usually do this to get some exposure, right? So the real reason they're doing this is so that people start sharing it – they link to their websites so that there are more users coming in.**

34:50

Angelica

Maybe in this case – the call to action in the previous example of Java people could be to try to get these people who like Java to learn Python. “Invite these people to learn Python because Python is more recent.” I don't know, something like that.

# Are reports to stakeholders considered technical writing?

35:13

Alexey

**I see a comment in the live chat from Adonis. “So reports to your stakeholders can be classified as technical writing too?” Angelica, what do you think about this? Is it an accurate observation?**

35:26

Angelica

I think that it depends on how the report is built. Because I'm currently studying data storytelling – there are many books about this. One of the most popular ones is Storytelling With Data by Cole Knaflic. It's a very famous book which says that when you talk to your stakeholders, you have to tell them stories and not just the report. You have to invite them to make some decisions based on the data, obviously. So I strongly encourage you to transform your sad and boring reports into data stories.

# Data visualization in articles

36:20

Alexey

**We have a question. “How do you make sure that an article that you're writing has the right amount of data visualization? Not too much, not too little – just the right amount.”**

36:33

Angelica

I think that to explain the concept, maybe one graph is enough. So you have to use one concept per graph. Don't show many concepts in the same graph. It's very confusing. Then, if you have to explain many concepts, you need many charts. But I think that the audience will get confused, because the idea is to transmit just a single message. All the graphs that you put in your report must be in accordance with this message.

Don't think about whether there are too few graphs, because the most important aspect of your report is not the amount of graphs, but the message that you want to transmit. I think that in some cases, a table is better than a graph because it's clearer. If you have just two or three data points, it's useless to use a graph. Don’t use a pie chart. [chuckles]

38:19

Alexey

**If you only have three data points, a chart will simply take up too much space.**

38:26

Angelica

I fight pie charts with all my heart, because I think that when it comes to pie charts – if there are many, many slices in the pie chart, you don't understand anything. If there are only two slices in your pie chart, it gives you the Pac Man idea. I don't know if you know Pac Man, which is a character from some video games. It seems that the greater slice wants to eat the small one. [laughs] This has a negative effect on the audience. So don't use pie charts. [chuckles]

# Article length

39:12

Alexey

**You mentioned an interesting thing – a purpose for an article should be to convey a single message and all the visualizations you use (all the charts) should support this single message. Since you wrote so many different articles (170) how long are these articles usually? Are they six minutes long to read?**

39:38

Angelica

Yes – five, six minutes. On Medium, there are also some very long articles – about 15 minutes. To tell the truth, I don't have the time to write these long articles, so I keep them short. I’ve also started a publication on Medium, which is called Syntax Error (at the moment, it's very small) where the main objective is to solve some problems very shortly – like Stack Overflow in Medium. Maybe in the Slack community, if someone wants to write or participate with this publication, they could drop me a message, and I will be happy to add them as a writer.

40:40

Alexey

**Just share the link and we will include this in the description. That's an interesting idea.**

# The process of writing an article

40:47

Alexey

**Can you walk us through the process of creating an article? You said that your articles are usually five, six minutes long – so they're super-focused on conveying a single message. Correct me if I'm mistaken – this is what you start your article with, thinking of what this message could be. Right?**

41:07

Angelica

Yes – an overview of the topic. You start by giving an overview of the topic, then you describe how you solved your problem. Usually, I write technical articles that solve some common programming problems. I describe an example which solves the problem and then I give a summary. The structure is very, very simple. You could write something more complex, but in my case, it's very simple. I also usually provide the code of the article in a separate repository, which is well-appreciated on GitHub. It’s small, but it's appreciated.

42:06

Alexey

**The format, as I understood it, is – first you have a problem, then you show the solution, and then at the end, you talk about the result. Right? In this format, you come up with some outline. You have this outline in this form and then you add text, some code, and illustrations. [Angelica agrees] How many illustrations do you usually add?**

42:32

Angelica

I generally add an illustration at the beginning to capture the attention of the reader, and then maybe technical illustrations related to the problem to be solved. For example, if I need to draw a graph, I show the graph.

42:56

Alexey

**So when do you think about illustrations? When you have an outline, but you haven't started writing the article? Or at the end you see, “Okay, there is just too much text. I need to add a picture.”? How does it usually work for you?**

43:11

Angelica

No, when I finish the topic, I stop the article. But if the article is too short, I try to extend it, for example, by extending the example or extending the overview, adding some general concepts and similar things.

# Finding writing topics

43:40

Alexey

**How do you find topics? You said that you usually focus on a specific problem, and then show how to solve this specific problem. Where do you find these problems? Is it something where, let's say, you are working on a particular thing and then you have this error? [Angelica agrees] Is this your main source of inspiration? Or how do you usually find topics?**

44:03

Angelica

Usually I write on the problems that I personally have because I think that maybe they could be useful for the community. But some other times, I read on social networks like LinkedIn or Twitter for new libraries, for example – I try to test them and then they write an article about the topic. But most of the time, I take note of my problems and how I solved them. For me, this is also a way to keep track of how I solve my problems.

In fact, I would like to suggest to the community to also take note of how they solve their problems. Because we are used to searching on Google, but maybe if a problem comes again, twice or three times, we need to search again, on Google for the same problem. Instead, if you keep track of the problem by writing an article, we have the solution already available. And this is my strategy. Maybe if after 10 months, I have the same problem, I know exactly how to solve it.

# How Angelica got into writing a book (communication with publishers)

45:35

Alexey

**These articles that you write – they're very focused, very short, and to the point. And then you also wrote a book. A book is not a five-minute read – it's something more comprehensive. I have a lot of questions about the book, but maybe you can tell us how you ended up writing a book? You said you were writing some novels, but then at some point you decided to write a technical book. So how did this happen?**

46:06

Angelica

The situation was different because I didn't choose to write a book.

46:22

Alexey

**Did somebody force you? [chuckles]**

46:25

Angelica

No, no, no. [chuckles] My idea, my dream since I was a child was to write a book. But at some point, I received an email from an acquisition editor, who asked me to write a book. I was surprised by this. Maybe he read my articles and contacted me. At the beginning, I didn't know if it was the best thing to do. But then I decided to accept and then I wrote this book. Also, he proposed the topic to me. So instead, the topic was proposed by them (by the publisher) and I accepted it.

47:24

Alexey

**But it didn’t come out of the blue, right? Was it something related to some of the articles you wrote? You probably must have written an article about Comet for Data Science and that's why they thought of you.**

47:37

Angelica

Yes, I wrote an article about Comet – an overview of Comet. I’m also a contributor to the Heartbeat publication, which is a publication by Comet. So maybe they contacted me for these reasons.

48:00

Alexey

**Interesting. Basically, the recipe is you write and publish your articles on Medium and some other platforms, and then publishers might reach out to you. Right? At least this is how it happened with you?**

48:20

Angelica

Yes, I think that it could be a great strategy if you want to write something – to build a portfolio of articles which show your capabilities and your writing skills and then maybe someone from the acquisition team of a publisher can contact you. An alternative could be to contact an acquisition editor directly. For example, you can go on LinkedIn, search for a publisher, you search for the acquisition editor in the company, and you ask them if they are interested in some books.

49:13

Alexey

**So you approach them with a topic. Maybe you go to LinkedIn, and you write “Packt acquisition editors” or “Manning acquisition editors”, and then you see some people and send them a connection request writing, “I'm interested in the writing a book about Comet,” or “I'm interested in writing a book about (some specific topic).” And then they might say, “Okay. You know what? Let's actually write it.” Right?**

49:41

Angelica

Yes, because the alternative (the official way) is to fill a form and send a proposal. But in this case, I think you have less chance to be accepted because you're one of the thousands of people who sent a book proposal. If you are sponsored by an acquisition editor instead, I think that you have a higher probability of being accepted.

# The process for writing a book

50:19

Alexey

**It's not an easy process. I know that because I’ve also written a couple of books. I'm curious, how did it actually happen? What was the process from the moment you got contacted by an acquisition editor, to the point when it was actually out in print?**

50:40

Angelica

It was a *long* journey.

50:42

Alexey

**I know. [laughs] It's very long. Too long.**

50:44

Angelica

It's another job. From this first book, I learned many things and I'd like to share them with you. The first thing is to read the contract carefully before signing it. I think that the first step is to read the particular timelines where the delivery process is defined. The publisher shares with you a possible timetable (timeline) where he says exactly when you have to deliver a chapter. Usually, due to time-to-market resources, the publisher expects a chapter every two weeks. I have done... [cross-talk]

51:41

Alexey

**That’s a very tight schedule.**

51:43

Angelica

Yes. I have done this for the book Comet for Data Science, but I don't ever do this otherwise.

51:53

Alexey

**That’s like, “Okay, but when do I sleep?” [chuckles] Right? “When do I have a life?”**

51:59

Angelica

Yes. And so, if you can, ask the publisher to extend this period to at least four weeks per chapter. And if a chapter requires more time (maybe it requires a lot of coding and so on) also ask them for six weeks. I think that you should also add a holiday period – a vacation period – because the contract has a section where you can insert this aspect.

Then, when everything is okay for you, you can start writing – assuming that you already have a table of contents and so on (a list of chapters). If you have to write a chapter every four weeks, in practice, you need to write a page a day. You have 28 days for each chapter. If you remove weekends, maybe you have 20 days. You can write a chapter of 20 pages only by approaching one page per day, which is very easy.

53:32

Alexey

**Even before that – even before you have a contract (the contract and the timeline that you mentioned) – you need to actually feel the timeline. You need to know what the milestones are. “What are the chapters? What are the sections?” I guess there is also a lot of work before the contract. I don't know how it happened with you, but I guess these acquisition managers come in with a title and say, “Okay, I have these titles. Which of them was interesting for you?” And then you need to actually work out a proposal. Right? How does it happen?**

54:17

Angelica

Firstly, you need to download the publisher’s template and you have to fill it. But the main question you have to ask (and you have to tell the truth when answering this question) “Do I know the topic?” Because if you don't know the topic, at least in general, don't write the book.

54:44

Alexey

**[laughs] That’s a valid point. I was kind of under the assumption that if you want to write a book, then you know it. [chuckles]**

54:52

Angelica

[chuckles ]It could be that you don't know the topic because maybe you’re working on machine learning and you want to write something about TensorFlow, for example. But if you don't work with TensorFlow, don't write a book on TensorFlow. At least, you must know an overall overview of the topic.

Then, I think before writing everything, you need to search for other similar books. In the research field, this is called “the state of the art”. You need to see what is already available in the market. Because if your book doesn’t add anything with respect to the previous books, there is no sense to write it. [cross-talk]

55:53

Alexey

**It’s called market research, right?**

55:55

Angelica

Yes.

55:57

Alexey

**It could be. I think I came across this. I don't know what publishers call it. Your point is that the book should be unique in at least one aspect, right? [Angelica agrees] It shouldn't be just a copy of another book when it comes to the table of contents.**

56:15

Angelica

Then once you have the topic, you have to think about the audience of your book. This is similar to the previous… [coughs]

56:36

Alexey

**Yeah, maybe take a moment.**

# Brainstorming

56:47

Angelica

Okay, sorry. You have to think about the audience and then I think you can start writing on paper everything about this topic. Everything that comes to your mind – it’s the brainstorming phase, where do you decide what to include and what not to include in your book.

57:16

Alexey

**I think what I did was create a mind map. You take a blank page and in the center I put the topic like, for example, “machine learning engineering”. Then there could be branches like “Okay, what can I talk about when it comes to this topic? I can talk about machine learning, I can talk about engineering,” and then I kind of branched out from each of these things. Then you convert that into a proposal eventually.**

57:47

Angelica

Yes. Then I think that you can group the similar topics that you have found into macro areas. For example, if you have found 40 arguments, you can group them into 10-12 macro areas and this will become the chapter of your book. This is a general process. To tell the truth, I didn't follow this step. For the book Comet for Data Science, I had everything clearly in my mind. I said “I start saying this, then this, then this, and finally this.” [chuckles] The editor slightly changed something, but they accepted the proposal as it was.

58:51

Alexey

**So you have this proposal. You work for some time on the proposal and then the result (the proposal) is basically the table of contents, right? You have what the sections are, what the chapters are, and then I think you also describe the target audience.**

59:07

Angelica

Yes. Also for some publishers, you also have to provide the state of the art (similar works) and the potential audience. But you must be very specific in regards to the audience. For example, you must say if the audience's entry-level or intermediate engineering, if there are some requirements for the audience. Because an entry level book is simpler – you talk about the basics and so on. But for an intermediate level, you have to know the topic very well because you assume that the audience already knows some things.

# Reviews and revisions

60:09

Alexey

**So then there is a proposal, which is hopefully accepted by the publisher. Then you sign the contract, where you should watch out for the things you described – make sure it's manageable and that you have time to actually enjoy life, and sleep, and other things. Then, if it's one chapter per month, you need to write around one-two pages per day to finish it, right?**

60:38

Angelica

Yes, but the process is not as simple as it seems. Because when you submit the first chapter, you write the second chapter, but after, for example, two weeks, you get the first chapter back with reviews. Firstly, you get notes back from your editor to change some editing aspects and then from reviewers. If you wrote quite well, this is easy, because the reviewer simply says “Okay.” But if you made some mistakes, then you need to work on chapter one and chapter two. This process overlaps with all the chapters.

61:44

Alexey

**I can see how it can snowball. Like when you write chapter four, you still have some unfinished comments from chapter one, then chapter two, and then you go mad.**

61:01

Angelica

Yes, but the idea is to keep yourself organized. I think that if you leave at least half an hour per day or one hour per day, it could be sufficient to deliver everything in time. But one hour is the rule for the time.

# Conclusion

62:23

Alexey

**I just noticed that actually, we’ve run out of time. This is a very exciting topic and time flies.**

62:31

Alexey

**Maybe a very quick question before we finish. There is a question from Funkan. The question is, “Can a very routine 200-stroke article on, (for example, Italy's inflation rate in November 2022) be considered data journalism? Or is it more like a report?”**

62:52

Angelica

I didn't get the point? Stroke?

62:56

Alexey

**“200 stroke” I think 200 pages maybe? A report from Reuters on Italy's inflation rate in November 2022. Can this be considered data journalism? Are you familiar with Reuters? I think they are some sort of publication.**

63:17

Angelica

Actually, I don't. So I don't know the topic. Maybe I can investigate this and give a better answer.

63:30

Alexey

**Maybe we can find this article, share it with you and then you can answer in the notes. [Angelica agrees] It will be published in a couple of weeks anyways. Okay. It was fun talking to you. Thanks for joining us today. Thanks, everyone, for asking questions. I think we can wrap it up. Everyone, enjoy your weekend.**