**Mark:** [00:00:00] Welcome to the Endless Knot Podcast,

**Aven:** where the more we know,

**Mark:** the more we want to find out.

**Aven:** Tracing serendipitous connections through our lives

**Mark:** and across disciplines.

**Aven:** Hi, I'm Aven.

**Mark:** And I'm Mark.

**Aven:** And today we're talking about ancient papyrus scrolls. But before we get to that, and an interview, I wanted to talk a little bit about something that is specifically relevant to our Canadian listeners and most relevant to our Sudbury listeners, but it may be of interest to others as well.

There is a current campaign going on called Defeat Depression. This is a national mental health fundraising awareness and anti stigma campaign that provides hope for people affected by depression and other mood disorders. There are events happening across Canada, so I encourage our Canadian audience to look into events going on in your community.

The campaign runs until the end of May. [00:01:00] So there may well be things going on when this drops, but afterwards as well. Specifically in Sudbury, the Defeat Depression events are raising money for the Mood Disorder Society of Canada, but also for a local mental health organization known as NISA, the Northern Initiative for Social Action, which is an inclusive, peer run, recovery oriented mental health organization located in downtown Sudbury, which offers peer support along with active living and creative and occupational programming.

So the events going on in Sudbury can be found by going to the website that I'm going to put in the show notes or searching Defeat Depression Sudbury, but in particular, it will be culminating with a walk run to raise money on May 25th. This will start at Bell Park in Sudbury, and you can register by again going to the link that I'm putting in the show notes or looking up Defeat Depression Sudbury.

And you can either register to take part or pledge money to those who are or simply donate to the cause. This [00:02:00] is a particularly relevant cause for us, one that we care about, and so I wanted to draw attention to it. So, as I said, check that information out in the show notes, or look up Defeat Depression in Canada to find community events near you.

Now, let's turn to the subject of today's episode. We're going to be talking about a volcano. I mean, not really, but sort of. So we're going to be interviewing a member of the Vesuvius Challenge. And what is the Vesuvius Challenge, Mark?

**Mark:** Well, in 2015 a team led by Dr. Brent Seales at the University of Kentucky used X ray tomography and computer vision to read a still rolled scroll.

Virtual unwrapping has since emerged as a growing field with multiple successes. Their work went on to show that the elusive carbon ink of the Herculaneum scrolls can also be detected using X ray tomography.

**Aven:** The Vesuvius Challenge is a machine learning and [00:03:00] computer vision competition that was launched in March 2023 to bring the world together to read the Herculaneum Scrolls using the technology that comes out of Brent Seales and his team's work.

Along with smaller progress prizes, A grand prize was issued for the first team to recover four passages of 140 characters from a Herculaneum scroll. And today we're going to talk to Dr. Stephen Parsons, and he is the project lead for the Vesuvius Challenge.

**Mark:** Stephen started with the project as an undergraduate in 2014 and has never fully gotten away.

After graduating and working as a program manager at Microsoft, he joined the DRI team full time to help set technical goals and translate them into functional technology. He is driven by the team's challenge of pushing the boundary of what is deemed possible and the opportunity to recover knowledge that is often thought to have been lost forever.

He began the PhD program in computer science at UK in spring of 2019 and defended July [00:04:00] 2023.

**Aven:** So today we talked to Stephen about his work, about the Vesuvius Challenge, about lost texts from the ancient world, and a number of interconnected, surprising connections along the way. So we'll play that now.

So, hi Steven. Thank you so much for joining us.

**Stephen:** Hi. Thanks for having me. I'm excited to talk to you.

**Mark:** It's really good to have you here. And you know, your project is the sort of thing that really interests us because it pulls in so many different areas and so many different approaches.

And we're really interested in the way things connect. And so the first question I want to ask you is, about that and this project. Can you tell us about an unexpected connection in your life, a link between your work and the rest of your life, or between parts of your work and your personal experiences that affects the way you think about the world, or was helpful to you?

or surprising, or led you in a new direction? Or [00:05:00] interconnection between two parts of your life that had unexpected results?

**Stephen:** Yeah I've had fun thinking about this, and I, didn't want to bias my thoughts by listening to anyone else's answers, but now I'm curious how everyone approaches this.

I'm sure this goes in many directions. And yeah, I was thinking about it, and there is a clear area that I feel has recently been on my mind and this might just be because I'm in writing mode right now. So I'm, preparing the technical paper behind some of the recent results we announced that we'll submit to a journal.

So I'm in writing mode. And one connection that comes up all the time in my life is that, I'm always finding that my writing, which is either technical writing in English or is code, is more alike, or the process of writing that is more alike my wife's writing. And she writes young adult fiction than I would have thought.

And we're always, we've had the conversation a hundred times now, and still every day, one of us will [00:06:00] gripe about some challenge of, writing and the other one to our surprise-- I don't know why we're surprised anymore-- we'll, we'll relate to it. I just find it really interesting. How similar it is, especially the similarities between writing fiction and code, for instance, which I would think are two quite different things, but they really are both this act of trying to clarify and end up with a concrete product.

That's the end goal, but you start with this vague idea in your head. So the process of distilling it into a real thing is very similar. And we face the same trade offs of, you know, you try to outline as best you can for fiction or for code. You'll try to think through or diagram or plan what it is exactly your program is going to do before you write the code to do it.

And in both cases, inevitably, you could do as much of that as you want, and the instant you sit down to actually do it, you find all the things you didn't think through, and the whole thing kind of unravels, but of course, you know, it still was useful to do the exercise. yeah, little [00:07:00] things like that.

They're always, always similar. So that's 1. I have related thoughts, but I don't know if you had any thoughts or questions based on that.

**Mark:** Well, I just want to sort of react to that saying, that's really fascinating, and actually I've thought a lot about, since I'm particularly interested in language and linguistics, I have thought about some of the comparisons that can be made between computer language and a human language.

And in fact, although human language is theoretically capable of producing Endless variety. We tend to speak often in very formulaic ways in the same way that you might in coding, reuse little chunks of code because you know exactly what this is going to do and you just plug it in, in the right spot.

And we do the same

**Stephen:** thing when we talk. Yeah, most code, like most language, I guess is, is repeated many times over.

**Aven:** And because it's been made fit for purpose, like, you've refined it to the point where it does exactly what you want it to do in exactly the right place, and so why wouldn't you reuse it?

And in fact, that's exactly [00:08:00] what we do in, you know, why would I come up with a new way of saying hello when the old way of saying hello is the one that everyone expects and understands because it doesn't, it's not actually better to be innovative in a lot of our transactions, because it confuses people.

Yes, they understand the new way of phrasing it, but Why would that help? No,

**Stephen:** code is the same. Very much you want to save the innovation for where you

**Aven:** need it. Mm hmm. And I imagine that's true for your wife's writing, too, and for fiction.

Obviously, we want to be creative and we want to be innovative, but you don't actually want to be creative and innovative, say, when you're doing the background work of constructing dialogue. because that will just draw attention if you use he expostulated instead of he said, that's not helpful.

That makes people look at the word expostulate. I knocked my microphone. I was so excited. I was expostulating.

**Stephen:** Yeah, you're right. And there's so much I've learned so much through her. You know, I read fiction. Casually my wife reads fiction voraciously and sort of on a different level where she just reads so much of it that she [00:09:00] has become familiar with all of these building blocks and can really with almost a skim, like what for me would be challenging for me to get much out of a book.

I mean, she listens a lot to audiobooks, and will do it while doing other things, and still, she's so familiar with the common tropes and building blocks that she can take a lot from that and see how something's been constructed, and you know, most of it's pretty formulaic, so you can focus on the parts that were interesting to this work and stuff you might want to apply to your own work.

I found that really fascinating.

**Aven:** That's really interesting. We have other things to talk about, so I will not keep going down this rabbit hole, but now I just want to talk about, like, the interesting combinations. Maybe we need your wife on, because I want to talk about, like, formula in fiction versus formula in oral, you know, and the way we think about classical texts and how people read things and tropes and anyway, but I won't.

But I will ask you, what does your wife write and what's her name that she writes under? Just, you know, in case it's of interest.

**Stephen:** Oh, I'd, be thrilled to say. My wife's name is Caitlin [00:10:00] Hill, and she writes young adult fiction.

She writes romantic comedies so far, and you should be able to find her by searching her name, yeah. She has two books out, and there's another book coming out in May.

**Aven:** Awesome. Thank you.

Okay, great. Now I'm going to put a lid on that because that is a really interesting conversation, but we need to talk about what we're here to talk about, so could you, you're talking about writing up results. So can we take a step back then to talk about the project?

So this is the Vesuvius Challenge. Could you give us, your little short rundown on it? And then maybe we can talk more about some of the details.

**Stephen:** Yes, it will depend on where we want to begin the story. I can give sort of the effort to read the Herculaneum Scrolls. Non invasively, we say, meaning with X ray or other images that don't cause any physical change.

Mm hmm. That goes back about 20 years now. Mm hmm. Largely with my PhD advisor. And there's a long history there of, of ideas, and a lot of the technology just wasn't ready until recently. So, some of these ideas were too early The basics were there, but the [00:11:00] imaging wasn't good enough, or the computational power wasn't there yet, or we didn't have the ability to store or process these large images.

It's just sort of all converging. So anyways, that's a long history. It includes some of my own work in the last six years or so doing my PhD. And then now we're presenting results from Vesuvius Challenge, which is the latest chapter in this story. And that was born from a collaboration between my PhD advisor at the University of Kentucky, Brent Seal es and his team, including myself, and Nat Friedman, largely, who instigated and funded the initial effort, as well as his business partner, Daniel Gross, and then a bunch of other sponsors, and now a global community of people. So Vesuvius Challenge, we took a lot of our work and results as of a year and change ago, and we released them to the world and we invited anyone to contribute and extend what we had done. And largely what that means is that we had proven the concept of some of these [00:12:00] methods, but we hadn't gotten them yet to reveal complete texts from inside the rolled scrolls, which is what we really wanted. There are a couple of technical hurdles left.

So we invited anyone to do this and we issued some cash prizes. And launched still less than a year ago, which is remarkable. March 15th of last year. So we're coming up on it. And it's just been a crazy year. We've had so many people be involved. We've had thousands of competitors. We have a community now of all of this technical expertise, but we also have scholarly interest in the community.

Things have changed. really come a long way. There's a lot to talk about. I'll stop there for now.

**Mark:** And I think that's a really important point that you said right at the beginning that, these are non destructive techniques which is really important because, 50 years from now, there may be new technology available, new approaches that can be made, but the scrolls themselves will remain exactly as they are now.

And they can, do all this again with new technology. And I think that's a really important first principle.

**Stephen:** Absolutely. Yeah. One of the fun [00:13:00] things with the work is even on a smaller scale and shorter timeline things only get better so we can improve the method here and there. And what we have now is essentially a software pipeline with many steps.

And we know there's a lot of room for improvement in each one of those steps. And every time you Crank it up a notch on one part of the pipeline, the whole thing improves a bit and it never goes backwards. So we know that the images we see today will only get better going forward and that's always been rather exciting to me.

**Aven:** Now, just before we go any further because we have a broad audience of different backgrounds, some of whom, for instance, are real into etymology, but not necessarily the classical world, why don't we do a very brief explainer on the Herculaneum scrolls, just to start off with, if you don't mind doing that.

**Stephen:** It's a great idea. Should have probably included that in the

**Aven:** Yeah,

**Stephen:** the Vesuvius challenge is specifically concerning the Herculaneum scrolls. And my, my work also on the technical side isn't necessarily confined to [00:14:00] those, although they are a unique enough technical challenge that my work has ended up focusing pretty explicitly on, on the Herculaneum scrolls.

So this is a library of scrolls from Herculaneum, Italy, which was buried and carbonized or That means something close to burnt, but not quite by the eruption of Mount Vesuvius at the same time that it buried Pompei in the year 79 mm-Hmm. . So these scrolls were in a villa, a rather luxurious villa, so yes, this is a, it's a library from a villa. The villa and Herculaneum were buried by this eruption and it buried everything in about 80 feet or what, 20 meters, of ash that all hardened into rock ultimately.

But. At the time of the burial, it was hot enough to bring everything up to combustion temperature, but it also extinguished all the oxygen. So what happened was very much literally the same process as making charcoal. So these papyrus scrolls tried to burn, but there was no air to breathe, so they were turned into these lumps of, [00:15:00] charcoal.

And now they're extremely brittle, and if you attempt to open them physically, which many have done over the last 250 years, it destroys them. It's very destructive. So we use non invasive imaging to get around that.

**Aven:** And the exciting thing and the thing that makes any of this possible is that they were written on with ink and there is a difference between the carbonized ink and the carbonized papyrus, right?

That's the sort of base level of, physical difference that allows any of this to be, you know,

**Stephen:** Yes, there is a very, very subtle difference. One of the, really the focus of my own doctoral work was that difference in the fact that it's very small. Unfortunately, that's one of the major technical challenges here.

So some other materials around the world have been a little easier for these methods to work with. For example, The lab that I studied under worked with the En-Gedi scroll in 2015 and 2016. And that scroll is written with iron gall ink, where the ink itself essentially has iron in it. And so in [00:16:00] X ray, it shows up as these bright spots quite clearly under the right circumstances, it's still kind of tricky, but once you do the virtual unwrapping pipeline, you can see the ink. And the challenge with the Herculaneum scrolls is they are written with lamp black ink that's soot or close to pure carbon and they're written on papyrus, which as a plant material is close to pure carbon.

So chemically there's almost no difference between the ink and the papyrus. So the ink doesn't show up any brighter. So what we found is you have to zoom in really, really, really close and you can see a textural difference where the surface of the ink has a different texture. It might not be a different brightness or a different color in x ray, but if you look close enough, You can see that it's smooth and has cracks, whereas the papyrus is this rough, waffle looking grid.

**Aven:** Right. But it's, not a chemically different thing, right? I kind of vaguely understood that, but that is even more annoying and impressive than I realized. It

**Stephen:** is annoying. Yeah. That was [00:17:00] my PhD right there.

**Mark:** Did you, have input from chemists or other scientists about the sort of the physical side of this early on?

**Stephen:** Yeah, absolutely. There's been a number of works that have studied that, trying to better understand the chemical composition of the ink in the papyrus, and we've learned a lot.

Unfortunately, none of it has led yet to a magical imaging method that finds high contrast and can penetrate to the interior of the rolled scrolls. But we've learned a lot, and we do this using detached scroll fragments. So there have been a couple hundred years worth of physical unrolling efforts before this fancy magical modern imaging came around.

And you can't blame people for trying because they never really could have foreseen that this would be possible. So there's a long history of what now looks quite destructive, but really, to give them credit, was remarkably cautious work to physically unroll these scrolls, and a lot [00:18:00] was lost in the process, but we have now thousands and thousands of scroll pieces where there is visible text on the surface.

So that's led to a couple of things. One is, there is active scholarship with, that collection. So we do know some about the textual contents of the scrolls. And the other thing is we can study those fragments to learn more about the ink and the papyrus. And what we found is that it varies a lot because the collection represents about three centuries worth of text.

of scrolls made by different hands. And of course you didn't go to Walmart to get ink in that time. Everything was rather handmade and probably varied day to day and certainly varied over the centuries and different scribes I'm sure you had different techniques and so on. So there are sometimes these interesting chemical signatures in the ink that make it clearly different from the papyrus, at least with surface imaging, but they don't tend to replicate really well across

the complete collection. So some of them have faint traces of lead, for instance, or calcium or other elements, [00:19:00] but that's been tricky. And even when those do work well on the surface, it's hard to make those work for a penetrative imaging method.

**Aven:** Right. So you mentioned, you just referenced this a little bit in passing, but the collection as a whole, we, think we know some stuff about what the texts are that we're looking for in broad strokes. It's probably a collection of philosophical works, mostly in Greek that we know about from other textual sources, as well as from, I guess, those fragments that have already been Interpreted.

Is that correct?

**Stephen:** Yes. Yeah. And there are, many fragments that have been interpreted. It's, painstaking work. So people specialize and spend entire careers physically sitting in this office in Naples, most of them looking at these fragments under a microscope and doing a really intricate three dimensional jigsaw puzzle, trying to piece it together.

Mm hmm. And there's been a lot of text that has come out of that, so yes, we do know that most of it is in Greek, about 20 percent of it is in Latin, most of the Greek [00:20:00] is philosophy, and specifically most of it's Epicurean philosophy, and most of that is by an author named Philodemus. So we know more or less what to look for from the remaining scrolls, at least from what's been so far excavated from this one room.

And sure enough, the recent results seem to be more Philodemus.

**Mark:** It is a remarkable gift, though, that, you know, the first full scroll you were able to transcribe was a heretofore unknown work. I mean Absolutely. So lucky. That's not guaranteed, is it? Just randomly, you know, whichever one that you picked happened to be.

**Aven:** On the one hand, it's not guaranteed. On the other hand, statistically, not that surprising, given how small an amount of the ancient world's texts survive. We have such a small amount that made it through the manuscript tradition. Yeah. Yeah.

**Stephen:** I should say, I think I failed to bring up when introducing Herculaneum that this is the only library from classical antiquity to have been preserved, period.

**Aven:** in any way at all.

**Stephen:** Yeah. It's the only [00:21:00] one. So as you're well aware, most of the texts that we have, there are very few of them. And what we do have has been copied many times and changed over the centuries. So this is a very rare glimpse into a, it's a direct witness to the time period.

**Aven:** Yeah. The only thing at all close is the other papyrological stuff that we have, but we have very little literary text that's actually directly preserved from the period in Egypt. We have some, but they're all scraps and they were not from a library. They were from rubbish dumps. From Oxyrhynchus. Yeah, exactly. We have our, incredibly precious lines of Gallus and the the one manuscript of Menander that was a mummy wrapping.

And like, that's basically where it runs out. That's what we've got.

**Stephen:** Yeah. It's been so interesting from my perspective on the technological side to learn how specialized, the scholarly side is, so words such as papyrologist would have meant nothing to me, but of course people spend entire careers specializing not just in ancient manuscripts, but [00:22:00] specifically those written on papyrus.

or really more specialized than that. I mean, many of them work a whole career

**Aven:** on Herculaneum. Yeah. And then there's among papyrologists, there's the vast majority of them work on documentary papyrus, meaning on the, the wills and the shopping lists and the contracts and the things like that, which is the majority of what survives.

And then there's a tiny little bit of literary papyrus work and it's, It's yeah. They are among the, I think, hardest working and least glorified classicists out there, frankly. Papyrology is just maddening, from the outside anyway. As someone who never did it but has worked with some of the results, I'm astonished anyone can do it.

It's just so much work. I

**Stephen:** agree. It is a ton of work and it's so, But it's really

**Aven:** painstaking work. Yeah, extremely laborious and careful. And I mean, I know that what we'll, we'll get to the way that what you're doing helps with that, [00:23:00] but it doesn't really, it doesn't really replace it. And also it's, a different corpus that you're working with.

Yeah.

**Mark:** And we'll get into some of the, technical details about, how this is all done. But on that notion of, this first scroll being an unknown text. Do you think it would have made any difference had it been something that we had a later copy of, or whatever?

Would you have rather, happened upon something that we already had a copy of, or is it just as good, or maybe better, to have one that we didn't know anything about before?

**Stephen:** I have two ways of thinking about that. One is just my personal interest, which is, more excited by finding something new.

On the technical side, it does become a little bit relevant and really now it's becoming more relevant. Now that the methods are good enough that we have a starting point, and it becomes relevant because the technical methods rely on us labeling ink and not ink in order to train machine learning models, and that is a difficult thing to do when you [00:24:00] have only fragmentary results from an unknown text.

So already what we're trying to do is to work with our papyrological colleagues, so they can help us annotate these images, we can run a model and it will generate an image that shows where it finds ink, but it will have gaps. And I cannot look at that image and interpret much. And I certainly can't fill in any gaps.

And when I say gaps, I mean, really small stuff like individual strokes of individual characters, but the papyrologists who specialize not only in papyrus, but in, for example, the specific hand we are reading are the people you want for that job and they can help us modify our images in this sort of iterative process.

So they're able to do that even with this unknown text because they can read Greek and they can read fragmentary Greek. I'm sure it would be even a little bit more effective if it was a known text where we could fill in much larger gaps,

**Aven:** but you'd have something to check it against if you had, if you had even a, not perfect.

which you wouldn't, of course, have a perfect, but even if you had a close match on [00:25:00] an existing text, then, I mean, I am also torn on this question. Part of me, of course, like, increasing the amount of texts we have in the ancient world is a pipe dream for everybody. Like, it's something that we've always wanted, and it's amazing. The other part of me that is very aware of how manuscripts are edited, so this is something that Mark has worked more on because he's a medievalist, but that process of taking even a modern manuscript, you know, from the 14th, 15th century, like even something as recent as that, much less something from the 8th or 9th century taking a manuscript, and filling in those gaps and also doing this, there's this whole technical work, which is taking the manuscript as it gives you and then being like, but it's wrong because monks are stupid.

And coming up with the text that actually lay behind that, like correcting the mistakes. Yeah.

**Stephen:** There's a lot I don't know.

**Mark:** That's something that we're often told not to do. In medieval studies. To assume that they know more about [00:26:00] this language than we do. Yeah. So if they're writing something, they're right.

**Aven:** Yeah. Almost certainly. And we're wrong. That's one of the sort of principles of paleography. However,

**Stephen:** Are you talking about corrections at a small level, like a character or word level?

**Aven:** Yeah, at a letter or word level, generally. Yeah. Yeah. So there's this whole process, there's a whole set of rules that you use to, you assume they're right, unless it's absolutely, so massively incorrect that it couldn't possibly be right, or it makes no sense.

But then if it's, you assume it's the harder reading rather than the easier reading, because why would they, you know, make a mistake into making it less comprehensible, than more comprehensible. It's a whole discipline, right?

Paleography. Yeah.

**Stephen:** So I was recently exposed to the, and I may mispronounce this, the Leiden or the Leiden conventions. Hmm.

**Aven:** Lighten, I think, right? Yeah, yeah.

**Stephen:** just the syntax of how these things are transcribed. Mm-Hmm. and I had to learn the basics because we had our papyrological team generate transcriptions of, this work and had to do independently annotation.

Yeah. Mm-Hmm. , yes. [00:27:00] They very, yeah. They used that annotation and we had them do that work independently so that I could compile those results and we could evaluate how consistent they were to measure what we had and whether the transcriptions were unanimous or where they differed and so on. And there were some very interesting places where there's this notation I kept having to refer to the a document describing the Leiden Conventions for what, what all the different brackets and braces mean.

Yeah. And, there were places where you know, we would have five or six scholars unanimously transcribe a sentence as scribe error. Right. That, which I thought was so fascinating that they all could, see that.

**Aven:** Yeah. And partly

**Mark:** it is about, knowing about a particular, like, when you get your eye into one manuscript and you notice, okay, this scribe seems to do this a lot.

Then you can predict that as being

**Aven:** an error. They're likely to fall into.

**Mark:** Yeah, that they're likely this, this particular scribe tends to, make this mistake. Then you can say with a little more certainty, Okay, we can, we can emend that. We can [00:28:00] change it. Because I know that this scribe, keeps making the same mistake,

**Stephen:** which yeah, just, just within the last couple of days, I learned about a, I will get the details wrong on this, but in broad strokes, there's some effect in this period of Greek where it would be preferable in this context to avoid a word ending in a vowel followed by a word beginning in a vowel, that would create a sound that wasn't desirable.

And some authors took more care to avoid it than others. And so the presence or absence of that can inform who may have authored a piece and there's some debate about this with respect to these particular works we're working on. Yeah,

**Aven:** no, exactly. And, the bit that you've, that you've managed, that the team, the whole team and the teams that have won have managed to translate and try to, or transcribe and then it's been translated are very small, right?

We're still talking a very small excerpt of, I mean, it's amazing. It's big, not trying to underplay it at all, but in the grand scheme of [00:29:00] things, so that those, those choices, like what does the scribe tend to do? What are the things that the author tends to do? Those are a bit constrained. They don't have a whole lot of data to work off yet.

But there are these principles that you can work on, you know, the kinds of error a scribe is likely to make. they're more likely to regularize irregularities in the grammar than they are to introduce an irregularity, like something like that. But yeah, one of the reasons that I, would love it if along the way, once this goes, we do find a manuscript that we have, is because it would be so fascinating to find out if these principles of editing, especially in classics, where there does tend to be a lot more some might call it interference, others might call it emendation, where editors tend to be quite willing to say, no, the manuscript that we have from the Middle Ages or the Renaissance is incorrect, and this is what the actual author originally wrote you know, changing on the level of words or letters, or sometimes even more than that, not often, but sometimes more than [00:30:00] that. Yeah. it would be fascinating to find a manuscript from the first century AD and compare it to the edited, you know, compare it to the Middle Ages and then to the manuscript that survives and then the manuscript that has been edited, especially where they made those emendations and to be able to actually go back and be like, were they right?

Have our principles carried through? Which editors, you know, often there's multiple editions of a text and editors make different choices as to what they think the correct reading is. It would be, fascinating and a certain amount of unholy glee if they get it all wrong.

**Mark:** There's going to be a lot of ripple effects from this to so many

**Aven:** different areas.

Yeah, a hundred percent. Yeah,

**Stephen:** you want what in my technical world we call ground truth to measure against.

**Aven:** Which we, we simply don't have. That has, happened occasionally with certain texts. where we find older versions later, it's never gone back that far, but sometimes there'll be something that Renaissance manuscript and maybe [00:31:00] someone will find an 11th century manuscript of a part of it or something.

And that correction has been able to be made occasionally, but it's never back to, and of course, these scrolls are not all going to be written by the hand of the author either. We're talking about stuff that has been copied at least a few times, but still, it's so much closer.

Okay. So let me try once again to constrain myself to actually asking the basic questions, which I'm doing a bad job of. So we have these scrolls. they contain theoretically readable text in actuality, essentially unreadable by normal means. And so the project that has been going on, as you said at the beginning, for at least 20 years, and I remember, when it was first talked about first brought up the possibility that it was going to be possible to read these carbonized scrolls, because it has felt like such a tease for so long.

And this is why people have tried unwrapping it, as you say, the only library Ever from the actual period that we're ever likely to find and there they all sit on their shelves. inaccessible. And so it has been sort of this, [00:32:00] when we, I remember as a grad student hearing about the possibility we might be able to read it and how just like amazing that seemed.

**Stephen:** It's amazing in many ways. I feel like we're really lucky we have any of them at all. Yeah, especially given the time period when they were discovered, there was nothing remotely resembling modern imaging on the horizon. They just like, they could not have even entered the imaginations of the people and they still left hundreds of them untouched, hoping in some vague sense that some, better methods would come along.

It's a miracle they didn't just open them all.

**Aven:** It is. Oh, and I mean, when you think of the things that were done at Pompeii to some of the early discoveries they made there, I mean, I think in some ways we're lucky there was no gold and there was no paint on them and they weren't pretty. And then that was, that's part of the reason that they, you know, were untouched, but you're right.

It, it does feel kind of miraculous that, Yeah,

**Stephen:** they were in this villa. And my understanding is that for a long time, the prize of the villa was not the scrolls, but it was the sculptures [00:33:00] and wall paintings,

**Aven:** I think. Yeah.

**Stephen:** A lot of luxury items that were sort of raided, but the scrolls were second class for quite some time.

**Aven:** So the work that was going on before with your PhD supervisor and then with you and has been, so I'm going to explain this in the way that my very untechnical brain understands it. And you are going to correct me if that works. Sounds great. The, basic approach has been to use x rays to, penetrate to the scrolls without unwrapping them. And then to use technology that I understand was originally developed for medical imaging. That is, is it tomography? Is that where you do x rays at sort of a different depth and that produces a whole bunch of slices. So you sort of have a picture at each depth and you do it by like half a millimeter or whatever, whatever the depth that you want.

And then you stack those together. And that gives you either 1D or 3D, depending what you want, I imagine you want a 3D image of what the scroll looks like. And that I imagine was a whole bunch of work that took years just to do [00:34:00] that part. And if that makes sense.

**Stephen:** Yeah, and then our work begins.

**Aven:** Yeah, and then there's also the the ability to manipulate that image to then, as it were, unroll that image and say, okay, well, at each depth, if that were laid out flat, it would go this way. And to do that, which my brain, I can't even do the basic required spatial reasoning to do, navigate the city.

I can't even, if you'll excuse the pun, wrap my head around how that works. But I take it I do understand that that's what you're doing? So you're unrolling it. So now you've got an image, And as you said, with this very, very fine, slight distinction between where there's ink and where there's not.

But to any normal vision, it would just be black. And so that presumably was another piece of, technical work. But then where this challenge and where your work specifically comes in is doing that reading, like, how do we identify where the ink is and where the ink is not, and how do we turn that into a [00:35:00] transcription of what is on the papyrus?

**Stephen:** Yeah, that is pretty much the pipeline. the challenge and my work have focused really on every one of those steps, not so much on the imaging and on tomography, but every subsequent step, everything that we would call virtual unwrapping, which is taking the image that comes from the CT scan and tracing the layers and doing the flattening or the unrolling of them, and then trying to detect the ink on their surfaces.

Right, okay. Our work here and, through the challenge have touched all of those. Every one of those has needed improvements in order to tackle these particular scrolls. The Herculean scrolls, I have to say, are, such a compelling technical challenge alone. You could remove the,

**Aven:** The fact that there's new text?

Yeah.

**Stephen:** Yeah, and just the significance behind this being the only library Preserved from antiquity and all of that. And even if these were not particularly interesting material to scholars, still technically alone, they are, [00:36:00] to me, I think it's funny. I think they're like comically on the boundary of achievable and it's only now that, that it's doable.

And that's true on many fronts. The imaging is only recently good enough. Computation is only recently good enough in a number of different ways. They are really just barely possible, and they're such a fascinating, I, think that's really interesting.

**Mark:** So, to what extent does sort of feedback play in these different steps like for instance, once the papyrologists look at it, can their insights then get, is that useful to improve earlier steps in that chain?

**Stephen:** Yes. I think it's going to become even more important. And we do use that already. Really what we have is we're just now entering a situation where we have enough for them to be able to do anything with it. And now that we have something we have sort of a foot in the door and they're able now to give feedback on the images that goes back into, certainly goes back into helping [00:37:00] refine the ink detection machine learning models.

But on the technical side, there are some approaches we're exploring that combine these steps in ways also that would incorporate more feedback. So this gets a bit technical, but one example is that the tracing the shapes of these layers is a process we call segmentation. And it's really difficult because the papyrus is a complicated material and the scrolls were crushed.

So it's like a crumpled sheet of paper. So they go in all different directions and there are many, many layers and they're all tightly packed together. So there's not always a gap between one layer and the next layer. So you can't really see what's what. In x ray. So tracing them is hard. And what we're hoping is that the ink detection might actually help inform the tracing.

So right now it's a pretty linear process where we take the CT scan, we segment or trace all these layers, we virtually unfold them, and then we look for the ink on their surfaces. But it may be true [00:38:00] that before you trace or unfold anything, You could just run a detector across the entire image that will tell you where there's ink and that will help inform how to trace the surface, because the ink will only sit on the surface, if that makes any sense.

So these steps, right now, we're sort of anticipating the film.

**Mark:** you see the ink sandwiched between layers of papyrus. And so you, at any one point, you'll see ink, papyrus, ink, papyrus, and that will help you figure out where the, boundaries are, where the, layers are.

**Stephen:** Exactly. Because without seeing the ink layers of that sandwich, all you see is one big blob of papyrus.

Right.

**Aven:** And the fact that papyrus, for those who don't know, the way you construct a writing surface out of papyrus is you take reeds, you flatten them by pounding them, and then to make your life really miserable, you take them and you place them on top of each other crossways and you pound them. So you lay [00:39:00] one set going one way and one set going the other way, and you pound them really hard until the various substances in them kind of form a glue. And so what you end up with is a crosshatched double layer as your first, as one layer of papyrus, right? Like what we call a layer of papyrus is actually two layers of reeds, or probably sometimes more than that, but like an uneven level of layers, and then you sand it down a little bit, and then you write on that.

**Stephen:** Yeah, the whole construction of the scrolls is a really well executed cruel joke on those of us trying to do this in the year 2024. It's just Let's pick the most difficult material to write on. This really complicated structure of even one sheet is actually made of multiple layers of stuff.

It looks much different and much more complicated than, for example, Vellum does in CT, which is much more uniform. So, let's take a really complicated material, hammer it a bunch, roll it up really tightly with a ton of layers, and we'll write on it with ink that doesn't have any [00:40:00] contrast because it's chemically the same as the stuff we're writing on, and then we'll crush the stuff under, yeah, we'll just totally crush it and

**Aven:** burn it.

But

**Stephen:** of course, we have to be grateful, it's funny, I don't know if that's the right word, we're grateful for all that because that's the only reason we have this material at all, right?

Yeah. One thing that is fun about this challenge the, the challenge of, of overcoming this damage is that with these, we can at least have some vague optimism that as our methods improve, we'll be able to get more and more out of them.

**Aven:** Mm hmm. and just for those who don't know I believe, anyway, that this part of the library, you mentioned this, I think, in passing, the part of the library that's been excavated, this isn't all of, like, the number of scrolls we know that exist is not the entirety of the library of this villa.

There's a big chunk of it that, as far as I know, hasn't been excavated yet, even. Correct. And there's a whole, like, sub level, isn't there, that goes below that they expect to also have scrolls in it, though?

**Stephen:** There's at least one, possibly multiple, sub [00:41:00] levels. Mm hmm. Yeah, absolutely. There's a majority of the villa has not been excavated and it is speculated that there could totally be more scrolls.

There are different working theories on what the arrangement was based on what we have so far. So one of them, for example, is that this room where most of the scrolls came from was Philodemus's working library. And if true, perhaps there were other authors who had their own working libraries in other rooms down the hall, for example. Yes. It's not necessarily the most common theory, I don't think, but yeah, we don't, we don't know

**Aven:** the fond hope of people like me who are literature people is that that there's somewhere in the depths of this villa, I'm sure not to be found until years after I'm dead, but that there will be more, you know, Latin poetry, say. Yeah. And down the hall across the way was Virgil's working library. Yeah. Oh yeah. No, I mean, it won't be because he was dead sadly, but boy, oh boy, if it, if [00:42:00] there is, you know, there are so many things we don't have, this is where the, the daydreams start because there's so many things that we don't have, but anyway.

**Stephen:** Yeah. I've enjoyed learning different peoples' daydream stuff. What do you hope is there? Yeah, both from lay people, myself included, because People don't really know, they'll just come up with all kinds of interesting answers. But among the classicists I, I would have thought, as a complete un expert, that Greek philosophy is about as cool as it could get, but a lot of people would really love Latin poetry, it turns out.

Like, eh, Greek philosophy, we have, you know, whatever. Yeah, where's the poems?

**Aven:** Yeah, it is a game the specific way that we played it as a grad student is, what would you give up 10 books of Livy's history for?

What missing piece, because we have a lot of Livy's history, not all of it, but we have a lot, we have 40 books or something and it's, it's, there was 120, we're missing a big chunk of the stuff we wish we had, but. nonetheless. What would you give up 10 books of Livy's [00:43:00] history for? Like what, what work that we don't have?

And it obviously has to do with what people study and what they care about, but everybody has their own answer to that, for sure. if you ask most Classicists, that then we'll have an answer to hand as to what that is. What's yours? I mean, the one that I, I probably wouldn't enjoy the poetry as much, well, I might, but the thing that would solve a lot of questions, or at least probably ask a lot more, is Gallus's poetry. This is Gallus as a writer who was in the first century B. C., and his poetry is quoted by a bunch of people that we have, whose poetry survives, including Virgil. He's referenced, we're told by everyone that he invented the genre of love elegy, that he was one of the best poets ever.

He's in all the lists of the best poets ever. And we only have nine lines of him since some was discovered on a papyrus in Egypt in the 20th century. Before that, we had three words, separate words, that were preserved in, or I'm getting the numbers wrong, but it's like three or four words that were preserved in grammarians writing [00:44:00] about like how he'd used an interesting Ending on a word.

That's all we have. And yet he's, he's hugely influential and everybody talks about him and he was also an interesting political figure, who was in charge of Egypt. He was a governor of Egypt until he seems to have done something wrong that made Augustus mad at him and he killed himself. He was, he was forced to commit suicide because he'd done some traitorous thing.

So like, Is there

**Stephen:** any possibility given times and places that any of that could be in the, in Herculaneum?

**Aven:** Absolutely. he was writing in the, 30s and 20s BC. and he was very famous. So there's absolutely no reason he couldn't be in that library. So, you know, that's the kind of thing that, that we lie awake dreaming about.

**Mark:** I think for me, it would be Ennius for similar reasons. Yeah, no, that makes sense. Because Ennius was the big epic writer before Virgil.

And then once Virgil came on the scene, people stopped

**Aven:** copying [00:45:00] Ennius, I guess. Yeah, they just gave up on him. He had been the school text in the same way Homer was for the Greeks, Ennius, E N N I U S because it's, there are too many words that sound the same. Ennius had written this big long epic about the history of Rome, basically.

It was the history of Rome, but as an epic poem. And it was the school text that everybody learned their Latin by and learned, and memorized in school. And then Virgil came along and wrote the Aeneid and it became the school text within a year or two of his death, basically. And they stopped, I mean, they didn't stop immediately, but basically Ennius disappeared.

And now we only have fragments of him, quotes, quotations in other works. But we know he was the most, like, he was hugely influential. we can tell that Virgil is reacting to him and other people react to him and people talk about how old fashioned he was and all these things, and we have more of him than we have of Gallus, but that's because he wrote a huge amount.

He also wrote a bunch of plays and he wrote, he wrote a whole bunch of stuff. None of it survives.

**Stephen:** That's amazing. My mind just reels, imagining all [00:46:00] these, these things it could be. Of course with Herculaneum specifically, we know that at least a large part of it is likely more Philodemus, but it could be anything.

And that is so exciting.

**Aven:** Yeah, exactly. And I mean, obviously, I joke that I want the Latin poetry, like it doesn't matter. Any text, even from the basic sense of we get more Greek text, we understand more Greek words. it's, like, even as simple as that. Forget about what they say. We have a better, witness to how Greek was being spelled, what its grammar was, what the vocabulary meant, usages, like, we have to rewrite the dictionaries.

Okay, must stop dreaming.

**Aven:** So, okay, let's just talk very quickly about the actual Vesuvius challenge, because there's so much, I mean, this is such a huge project, but I know that the challenge is a part of it. And not the whole project of reading the Herculaneum scrolls is the big project, but this is part of it. it launched in March last year, as you said, and it was a series of challenges, right?

There was a [00:47:00] series of goals to be met and anybody could work on them. And then each goal as they were met, the results of that goal became public, publicly available to build on. Is that essentially the framework?

**Stephen:** Yeah, we launched with a grand prize that at the time a year ago was really highly ambitious.

We had really no idea whether we would meet it. And certainly if we got close, we didn't know how close we would get, kind of made up the bar as we went. We ended up meeting it basically exactly. One team and one team only exactly met the bar we had set in terms of number of passages and number of, of letters.

But yes, we had that grand prize. We knew that was really ambitious. So we. We tried to design a series of progress prizes that would guide the community towards that, and we wanted to encourage collaboration. We didn't want just one gigantic grand prize where everyone worked towards that in secret. So, yeah, for any prize, grand prize, or these smaller progress prizes, we wrote into the rules that to accept the prize you have to make your methods open source.

And so that. it worked really [00:48:00] well throughout the year. So we had a lot of contributions that we would award and then they would immediately become open source and the entire community would sort of level up and proceed from there.

**Mark:** And so since, they were different, groups working towards the same goal, do you think it's going to be the case where, although, one attempt was maybe not overall as successful as another, nevertheless, it might provide some information better than say another, method. And can these be kind of combined into producing an overall result that's even better?

**Stephen:** A hundred percent. We see this both on the technical and on the scholarly sides. So technically there were many submissions for the grand prize and for various intermediate prizes throughout the year that didn't win or were runners up or what have you, but had very interesting ideas in them and those got absorbed into other things.

if you didn't win a prize, you [00:49:00] weren't forced to open source your methods, but a lot of people collaborated just out of sheer enthusiasm and excitement and desire to collaborate. We actually, in hindsight, maybe didn't have to force the collaboration as much as we thought we did.

People just really like to work on this together. But we also see that on the scholarly side, where 18 submissions. And this is one thing we didn't foresee, but Because of the way the year played out on the technical front, these 18 submissions all generated images of text of the same part of the scroll.

So we had 18 images of the same text and we provided those that passed our technical evaluation, we provided I think 12 of them to the papyrologists and they absolutely reported that there was a very clear winner that had the most Legible text, but in many cases they would reference some of the other submissions for particular letters

**Aven:** or tricky areas.

almost like having multiple witnesses to--- so we call manuscripts, like if you have [00:50:00] multiple manuscripts in the Middle Ages, you call them different witnesses, right? So it's like having multiple witnesses for, To make your, your, same bit. And then absolutely.

**Stephen:** And as things move forward, hopefully these methods will converge.

So the ink detection should become more and more consistent to the point where someday that may not exhibit as strongly, but certainly now where we're still in the early days of building these data sets, the models show different performance and focus on different features. So

**Aven:** yeah, that they read different kinds of letters or different parts of strokes better or whatever.

So you're going to get different things more accurate in each of the different versions at the moment. Yeah.

**Mark:** And so since this is stuff is being released the winning method is being released as open source. Does this mean that scholars will then be able to use these methods and technology to apply to other projects that they might have?

**Stephen:** Yes. certainly in theory, I will say a lot of the methods are quite particularly developed [00:51:00] to the challenges of the Herculaneum scrolls, but also specifically the datasets we've acquired of the Herculaneum scrolls. Right. And so, it's no straightforward matter to take the code that was released recently and scan your own different scroll and press a big green button and have it come out the other side.

But, absolutely, that will be true and our work is to make that more and more true as we go forward. These methods should and will extend to more than just these scrolls.

**Aven:** That brings up the question of the future. and for details, I will link in the show notes, of course, the Vesuvius Challenge website. There's lots and lots and lots of details for anybody who wants to really dig into what's available there. You can go in and find more about any of these aspects, but for the future, I know this isn't over, this was the grand prize, but it's been awarded, but there's more going on, right?

So what's the next steps?

**Stephen:** Yes, we didn't know when we launched a year ago what would happen, but we know now that it went really well and we want to keep it going. So we just, [00:52:00] I think eight days ago, announced a new set of prizes for this year. There is another grand prize, details forthcoming, but the bones of the prize are already Available on our website, and more or less that prize is to get us from 5 percent of one scroll, which is where we ended last year, to closer to 90 percent of four scrolls.

So again, we have a bar that feels dizzyingly high, but we will see. We'll see. We do have another series of monthly progress prizes that are rather open ended for all sorts of contributions, and then we also have some milestones that we've sort of first past the post achievements that we've set out that the bounties are ready to be claimed.

So we will see who gets them and when.

**Aven:** And I mean, in theory, the reason that there's such a big leap is because as these processes get refined, they can be run on much larger data sets, et cetera, quite a lot quicker. Like in theory, the, progress could ramp up quickly

**Stephen:** [00:53:00] in theory. Yes. what we usually see and certainly what we hope for is that improvements to the technical methods have a better than linear return on what you can get out.

So if we improved our segmentation methods, for instance, that doesn't get us just an extra 1 percent of the scroll that will get us from 5 to 10 or 20. And it doesn't take so many of those jumps before you get. Pretty big leaps. I'll say too, that's the plan for this year. We want to really spend this year refining those methods to maximize what we can get out of the images we have.

and we want to spend some of this year doing some of the research on the imaging. There are still some open scientific questions on how you scan these best. We now know that the way we did it in 2019, which is the data used in Vesuvius Challenge, is sufficient, at least to recover the amount of text we've seen so far.

It's very possible that we could do better though, or that we could do just as well with less expensive or more accessible imaging. So we have some research work to do this year on that front and with any luck, subsequent years would take [00:54:00] what we learned this year and sort of run the table on the rest of the collection.

So go read all the scrolls.

**Aven:** Mm hmm. Have they all been imaged? I'm imagining the answer to that is no, but how much of them, what percentage of them have been imaged in this, what you said, 2019?

**Stephen:** Yeah, very few have been imaged in this way. Maybe five or six scrolls in total have been imaged by CT, but only two to four of those, depending on how you count, were imaged at it in a way that we now I believe is probably good enough.

Okay. So we have basically four to date out of approximately 300. It's hard to count this inventory because of its fragmentary nature.

**Aven:** Cause it's basically a big pile of carbon. Yeah. Yeah.

**Stephen:** Yeah. And you know, how do you count one scroll that is now broken into four lumps? One of which is in France and three are in different drawers in Italy and who knows.

**Aven:** Right. Yes, actually, that was a question that I didn't know the answer to, is where these scrolls are now [00:55:00] housed and who's, like, institutionally I don't want to go spend too much time on the administrative part of this, but institutionally, who owns these scrolls? Who owns the project? Who owns the results?

I mean, you're talking the code is open source, but you've got a bunch of different people working on, transcribing the text. But, traditionally in classical scholarship, when somebody transcribes and, and publishes a text, be that a small fragment of a papyrus or a whole manuscript or whatever, that is a piece of scholarly work that they put their name to, and they, I mean, they don't own it, they publish it, but, Each piece gets published as a scholarly contribution. So how is that part of this working, or at least what are the broad strokes of it? I don't need you to get into all the details.

**Stephen:** That is a very interesting question and I'll answer what we know today. Some of that we're trying to figure out because we're sort of introducing new questions on how we handle this.

So the majority of the scrolls from Herculaneum, the vast majority, are in Naples in the National Library, in [00:56:00] the Officina della Papiri, or the Office of the Papyri. And they have thousands of trays of these opened fragments. The nature of this collection is so varied it's hard to overstate how different all the different pieces are because you have everything from just what is now a bunch of carbon dust in the bottom of a drawer that was once a scroll to tiny tiny fragments much smaller than your thumbnail to everything from there up to highly successfully physically unrolled scrolls that are in these Long custom built cases multiple meters long where you can see the text.

That's sort of the pinnacle of the physical unrolling. Right. But there's everything in between. So there are thousands of trays of open stuff in all sorts of different conditions. And there are hundreds more. Drawers of still rolled scrolls. They left the most challenging ones for us. So most of that's in Naples.

There were a few of them, quote, gifted, I don't know the [00:57:00] exact circumstances, to Napoleon at one point.

**Aven:** Yeah, let's not ask that question.

**Stephen:** Now reside in the Institut de France, and actually those ones are the, well, one of those in particular, is the scroll from which we have the Vesuvius challenge results. So we collaborate with the library in Naples and we collaborate with the Institut de France.

There are some in the Bodleian library in Oxford. Of course. A small number. I think there's another British institution. I want to say British library. That would be a small collection, but all of, yeah, but all of these, those three others are really small compared to the quantity in Naples. And then regarding the ownership, that is where it gets really interesting.

And we're trying to figure it out. So as it stands to image these requires an agreement between the research group, or in this case Vesuvius Challenge, which is sort of this, it's a nonprofit, but it's this sort of novel structure to do scientific research via competition. Entity. So there's a university, there's a technical university research lab involved.

[00:58:00] There's this non profit entity running this challenge. There are the host institutions that own the material. And there are the scholars who work with the material who often, at a minimum, closely collaborate with the host institutions, may be employed by the host institutions, but not always. And to image the material in the first place requires an agreement between those bodies what happens to the data, who owns the data, under what conditions the data can be released, and so on. And that's a hard thing to negotiate, and a lot of work goes into it. And now we have the wonderful problem of really facing the consequences of how we wrote those agreements for these scrolls. Of course it all becomes more complicated, more interesting once there's actually text involved.

**Aven:** Right. Exactly. Yeah.

**Stephen:** So, going forward. Yeah. We're trying to figure out the best way to frame that in a way that works for everyone. Cause of course we want to, there are these sort of competing legitimate interests. One of which is the careful scholarly work that has to happen.

And the other is this aspect that has done so much for Vesuvius challenge, which is this [00:59:00] sort of accelerated open environment where everything is shared quite openly, and it's a delicate balance. We're trying to figure it out.

**Aven:** Yeah, that's why I asked the question, because, frankly, the mind reels at how that is all being balanced.

And not because I think scholars are going to be protective of it, or, have a problem with that, but in these fields, collaboration isn't the norm, really. that's not true of archaeology. So I don't want to say something that I'm going to get colleagues mad at me about.

But you know, the, the, traditional approach to text in particular has been one person sits down and works on these pieces, whatever pieces they're working on, and publishes it. And then somebody else comes along and also publishes it if they, if feel that they have a different reading or whatever.

And then that becomes this whole process of deciding who's right and all the rest of it. And it's not that those things are owned. The point of it is for them to be out in the world so that people can work with them. But there is, an obvious importance to that intellectual property being ascribed to the right person [01:00:00] and the work because that careers are based on that and remuneration is based on that and, institutional reputations are built on that. So I can see it being extremely, I'm sure everybody involved wants all of this stuff to be out there in the world, but getting it there in a way that, because the other question to ask, which I think you've already sort of covered is the funding for all of this.

And a question that could be asked is, you know, what is the funding? What are the funders getting out of it? And what are each of the people in institutions who are contributing to it getting out of it? And I don't ask that because I think people are terrible and not altruistic, just, you know, what is it that the, that each person is going to be able to gain from the contributions, I guess.

**Stephen:** It's true. Yeah. It takes a lot of care. I mean, you can't have an agreement where you are going to scan them all and release it all and the host institution basically gets nothing. Yeah. It gets, nothing doesn't have to mean necessarily financial or anything specific, but there has to be some interest.

Yeah. Yeah, there are many stakeholders and for just normal reasons, they have [01:01:00] different interests they bring to the table and it has been quite a fascinating challenge to balance that. We've come up, with the particular results. That we have now. We've come up with at least a temporary solution that, that works for the current situation, which is not going to exactly resemble the future situations.

The current situation is we had the results from this scroll that were imaged before. Vesuvius challenge or any of this was conceptualized. So the imaging was performed under an agreement that didn't concern any of this work of the last year. And it does have provisions for the release of the data, but under certain conditions.

And then we also have a team of p did a lot of work doing the evaluation. And so out of respect for the Pretty crazy amount of work they put into the transcription. They need the ability to do some publication of this result. And so there are little ways we're trying to. Of course, meanwhile, we want to share the result as widely as possible and as openly as possible.

**Aven:** Right. Because otherwise it can't feed into the ongoing work because from a data point of view, you [01:02:00] need that data to build on. Of course.

**Stephen:** Yeah. Yeah. So we're careful about how we share the images. There's copyright on the images and there are only certain passages have been publicly released at the highest resolution.

Right. Those that have sort of agreement from our partners and the rest are temporarily. Downsampled a little bit in resolution, whereas you can join the Vesuvius challenge discord where our community resides and within that enclave, this is all open, right? But yeah, there's, it's been an interesting, yeah, yeah.

So that's, we're very much figuring that one out as we go. And I'm eager to see what we do. I agree with you. It's a good point that everyone involved does want the stuff shared with the world. but there are, of course, as you said, people's. livelihoods at stake here among other things. Yeah,

**Aven:** exactly.

And I mean, that brings us brings me to you as well. Do you mentioned that you were, right at the beginning? You said you were working on writing up some of the material for publication. I introduced [01:03:00] who you are, but I've completely neglected to ask you how you got involved in all of this and what your background is.

So in a completely backwards way, why don't we just take a moment to talk about that and how you got involved in all of this and what you're doing with the material yourself, like how this, how this is playing into your career in that way.

**Stephen:** Yeah, I got involved with this in 2013. That was when I first, I was an undergraduate at the university of Kentucky in computer science, and I wanted to do some research.

So I, I found Dr. Seales's research really fascinating. So I, reached out and ended up doing some work in his lab on a related project. So I worked on a manuscript called the St. Chad Gospels, which is not so damaged. You can open that one physically and image it in normal ways, but there's still interesting technical work to be done.

So I did a project with that. And then I did some other things. I graduated and I moved to Seattle and worked in industry for a while, ended up moving back home to Kentucky a little more than a year later. [01:04:00] And got back together with Dr. Seales and he offered me a staff position to continue some of the research while I was sort of in between other things trying to figure out what I wanted to do.

So I came back into the lab, I started working on this ink detection problem for the Herculaneum Scrolls. Got pretty absorbed by it. As I was saying, I just find the whole technical challenge really compelling and ended up deciding If I'm going to do multiple years of graduate research in a lab, I might as well get the credential. So I switched from being a staff member to a PhD student did my PhD on the subject near the end the Vesuvius challenge Happened, which I could not have foreseen.

And now I'm the project lead of Vesuvius Challenge. So I've taken over that role from JP, who did a wonderful job getting it going last year. I sort of advised throughout last year, and now I'm running that and we will see how, how it affects my career. I feel extremely grateful to work on this.

There's so many amazing people involved, the material is so fascinating it's pretty unclear what the specifics look like actually, but it's clear to me [01:05:00] that something will keep happening. Right. And that it will be very interesting. So I'm sort of riding that wave at the moment and we'll see where this goes, takes me.

**Aven:** Oh, yeah. goodness knows in the world of academia, that's way more secure and well thought out future plan than most grad students ever have. So

**Stephen:** that's, that's, that's nice to hear. It hasn't felt that way from my seat, but, but I know that things have, have gone.

**Aven:** Yeah, no, but that's, that's mainly an indictment of the entire academic sector rather than necessarily praise for you.

But yeah, no, it's, fabulous. And yeah, so that's, I know that the Vesuvius Challenge is its own entity, but your supervisor and the work you were doing that's the University of Kentucky. And then the Challenge is its own non profit funded by the two main funders and then with other donor support and some grants.

Is that sort of overall how the funding is working?

**Stephen:** Yeah. Vesuvius Challenge is entirely donor funded and we list most of them on the website and [01:06:00] and yeah, we still collaborate closely with Dr. Seales's lab at the University of Kentucky and with the partners at the institutions. Yeah. So it is interesting.

I mean, this. entity, and the way it operates is new. So we are figuring it out, but so far it's working well.

**Aven:** so you're very much from a computer science background and, been kind of coming to the actual sort of classics piece of this from that side, you know, as you, as you're getting more into, as you're finally getting more into actual text, presumably that piece of it is coming more to the fore than it had for a long time.

Yes, that is true. Yeah.

**Stephen:** Yeah. I always knew I wanted to do something combining computer science with insert really vague humanities related. Humanities or something globally minded. I, I got dual degrees as an undergraduate in computer science and in international development, thinking I might work on technology with that lens for a while and sort of changed course.

But yeah, you know, most of this, of course, hasn't been super deliberate. It's been [01:07:00] pursuing interesting opportunities as they have come up.

**Aven:** Well, of course, that's dear to our heart, the intersection of different fields, different spheres. I mean, I think this is an obvious place where that fruitfulness can be seen, right?

The, the combination of people come from different backgrounds. This is a very, very clear example of how specialists in different fields can collaborate and produce something that nobody could do on their own. Yeah, it's been Papyrus they can't read.

**Stephen:** It's been amazing. And there are always, I think, bound to be some tensions when you combine fields with different perspectives.

And, That's probably always been true with this work. It has been really interesting and certainly at times quite challenging with Vesuvius Challenge because we're taking two quite different worlds, not just academia, but specifically the classics and not just technology, but specifically the sort of Silicon Valley approach to moving fast and doing work in the open.

Those are two very different cultures, so there's bound to be sort of a collision point and sitting in the middle of that is at [01:08:00] times challenging, but we do see now, I think what fruit there is or we see now what we can achieve if you can manage to figure it out at the middle of that.

**Aven:** Well, and that tension, I'm sure to some degree is productive in itself, right?

The, cultures having to clash and work out and re examine the way that they're used to doing things is also in itself, I think sometimes a very valuable thing.

**Stephen:** Absolutely. Yeah. I couldn't agree more. I think, for instance, when we were talking about what the scholarship and data release looks like going forward, we don't know yet, but it's virtually guaranteed to be a more open approach than has really been seen before in that field.

So that will be Virtually guaranteed for sure. And I think is

**Aven:** a Faster.

**Stephen:** Is a major And fascinating contribution.

**Aven:** For sure.

So, we have kept you for ages, and I could ask you about 7, 000 more questions. But, I don't want to keep you too long. And, after all, there'll be next year's results to talk about.

**Stephen:** That's true. That's true. We don't know [01:09:00] exactly what will happen this year, but I can guarantee it will be interesting.

**Aven:** Yeah. So we don't, this conversation doesn't have to stop, but I will draw it to a close now.

Just so that we can give you your life back rather than me asking more questions about individual papyri for the rest of the afternoon. But thank you so much. As I've said, people go to the website. I think that's the obvious thing. you've just launched a new call. So as you said, right, basically to be involved in it, you just have to say you want to be involved in it.

Is that about right?

**Stephen:** That's absolutely correct. Yes, we have. I'm sure you'll put the website scrollprize. org somewhere that that can point you to our discord community and so on and yeah to be involved you just have to show up and express interest. We have people in the community who don't have a technical background who are doing their own papyrology with the results.

There's a channel where that's being discussed and we have collaborations between people of that background and technical people. And some of the prizes even are not, they're, they're [01:10:00] largely technical, but not entirely. Some of them we're looking for contributions, for example, to our documentation on the website or improved tutorials that help anyone from any background get into this.

So even if you don't have a technical background, that might be a good fit.

**Aven:** Great. Well, thank you so much. It's been a fascinating discussion. Yeah, really enjoyed it. And I'm thrilled that it's happening. And I'm really, really glad that we got the chance to talk to you.

**Stephen:** Me too. I've really enjoyed it. Thank you so much for having me. We'll have to update you next year on whatever happens.

Yeah,

**Aven:** we'll be coming back to this. Great.

**Stephen:** I look forward to it.

**Aven:** Thanks so much.

For more information on this podcast, check out our website, www.alliterative.net, where you can find links to the videos, blog posts, sources and credits, and all our contact info.

**Mark:** And please check out our Patreon where you can pledge to support this show and our video project. You can go directly to the videos at youtube.com/alliterative.

**Aven:** Our email is on the website, but the easiest way to get in touch with us is Twitter. I'm at [01:11:00] @AvenSarah, A V E N S A R A H,

**Mark:** and I'm @alliterative. To keep up with the podcast, subscribe on your favorite podcast app or to the feed on the website.

**Aven:** And if you've enjoyed it, consider leaving us a review on Apple Podcasts or wherever you listen.

It helps us a lot.

We'll be back soon with more musings about the connections around us. Thanks for listening.

**Mark:** Bye.