Jeff Deutch: I'm a researcher at Syrian Archive, where I've been working for the past four or five years. I have a background in sociology and am finishing up a PhD at Humboldt University in Berlin. At Syrian Archive I'm working on documenting and developing different workflows and methods for mostly open-source investigations relating specifically to human rights violations committed in Syria. I also do investigations myself.

Syrian Archive was founded in 2014 by Hadi al Khatib. He'd been doing a series of digital security and investigative trainings mostly with Syrian journalists, human rights defenders, and lawyers who were at this time going across the border to get trained and return to doing human rights and documentation work in Syria. The Syrian border was still open, and Hadi was doing trainings in Turkey and some other places. He was meeting with these participants and already, early on in the conflict, participants were saying that a lot of content published online on platforms was being taken down. So Syrian Archive started as an emergency thing to save this documentation from being erased. Now I think there's about 12 people in the team. We've also recently launched a similar project in Yemen, Yemeni Archive, as documentation groups in Yemen are facing similar challenges to those we have seen in Syria. We're using the same kinds of open-source technology tools and workflows that we been developing over the last few years and applying them to different contexts, such as Sudan, Myanmar, and Nicaragua.

Q: TK

JD: We use a workflow that is adapted from the Electronic Discovery Reference Model developed by the Duke University school of law. It's a conceptual framework geared mostly towards those working in legal environments; in their case it's mostly discovery, but we've adopted and applied it also to the archival, verification, and analysis processes. We've also worked to incorporate various harm reduction and risk mitigation strategies. The first step is identification, and within identification the kind of thing we're looking for are things like what kind of content do we want, right? This means videos, documents, Twitter posts, Facebook posts. What kind of sources do we want? This is the phase where we put together our lists of sources for preservation. For Syria we have about 3,000 or so sources that we're drawing content from, and for Yemen we have about 500 or so sources.

The identification phase also includes things like creating a metadata schema. A metadata schema in our context refers to technical metadata terms, such as exif metadata and file-specific metadata. But it also refers to things which add context to the data, such as a video, to make that data useful and allow users and intended audiences to find what they need. These are things like, what kind of ammunition is used or present in the video, location, latitude, longitude, weather, landmarks, if you can see a particular type of violation happening or not. So we have violation types from the UN office or the High Commissioner for Human Rights, which has the Independent International Commission of Inquiry on the Syrian Arab Republic. we have their categories in our schema. Some of the categories in our schema are populated automatically, and some of them are populated manually. This is the first step with identification.

The next phase in our workflow is collection, and this is mostly an automated process. We use some tools that archive and process digital content from a lot of our sources on a daily basis. Once we add a source, it's automatically archived processed according to the metadata schema that we set out. So this is the archival and processing phase.

The next phases of the workflow goes into verification, analysis, and publication. With verification we have three distinct steps. At minimum, we work to verify the source of the content, verify the location where the video was taken, and verify the dates in which the content was filmed and uploaded. If the source of content is in our existing list of sources, verifying the source is quite straightforward. If the source is not, we determine their credibility by analysing whether the source is familiar to those working on the project or our existing professional network of journalists, media activists, human rights groups and documentation workers. We also look for things like whether the source's content and reportage has been reliable in the past. This is determined by evaluating how long the source has been reporting and how active they are.

To determine where the source is based, we evaluate the content sources are publishing to determine whether content is consistent and mostly covering a specific area or whether the location differs widely. If sources use a logo, we check to see if this logo is used consistently across content. This also helps determine if the source is aggregating content from other news organisations and social media channels or whether they are uploading mostly their own content.

In some cases, we will use sources whose credibility we are unable to determine, or where content is credible in some cases but not others. International media outlets, for example, might be highly accurate and credible when reporting on certain types of events, but entirely wrong when reporting on others. Local media outlets might provide highly in-depth reporting on events in the immediate area that international media outlets miss, but may have little ground access to other areas.

For Syrian Archive, each piece of content goes through basic geolocation to verify that it has been captured in Syria. More in-depth analysis is done on some content when that content is deemed priority content. This is done by comparing reference points, such as buildings, mountain ranges, trees, minarets, with satellite imagery from Microsoft Bing and Digitalglobe, as well as Open Street Map and Google Maps.

Q: TK

JD: Since the Tunisian revolution, there's been a huge shift in the way that conflicts are documented with individuals and citizen journalists filling the gap in coverage that NGOs, international media groups, and monitoring agencies have traditionally filled. This is particularly true In places where foreign journalists are banned, or the reporting environment is too dangerous to report on certain incidents. In Syria, we are seeing more hours of video footage online that documents the conflict than there have been hours in the conflict itself. That's over 8 years of video footage online. People are filming a lot, they're uploading a lot, and they're wanting some of these things that are unreported to be seen. In conflict zones like Syria, you don't have a lot of foreign journalists that are going in. And in many cases, you don't have a lot of international organizations or NGOs going in either. Intergovernmental bodies have to seek permission to get inside, so open source content or user generated content, people posting online, is a way of bringing attention to a lot of the crimes being committed.

Q: TK

JD: At the International Criminal Court there was an arrest warrant put out in 2017 and was renewed in 2018 against a Libyan national by the name of Mahmoud Mustafa Busayf Al-Werfalli, that was mostly

based on Facebook content. This was the first time that at the ICC level, we are seeing user-generated content, social media content, being used as a basis of evidence that a crime had been committed. But user-generated content in general has been used a lot. In the U.S. for instance there's a company called X1 that develops what I would call surveillance software. The company put together a database of legal cases in the U.S. where social media content has been used, and I think they have found something like 8,000 or 10,000 or something. This was in 2015 or '16 so... Already in the U.S. we're seeing that being used, but in Europe, not so much yet.

Q: TK

JD: It's important to highlight what we're collecting isn't necessarily evidence. It's documentation that we're trying to securely preserve, that we're trying to verify, analyze and store to the best of our ability. This includes time-stamping it and hashing it and things like that, but from a legal perspective there's a lot of challenges in using this kind of content in a legal environment. For instance, in Germany you might have to convince a judge that this would be able to be accepted as evidence. Whereas in the U.S., you have a more long standing acceptance that this would be valid. But the goal is definitely to use this to corroborate other forms of evidence, such as witness testimony, physical objects, papers.

Q: TK

JD: One of our goals is having some kind of record, a story that's framed. Most of the people on the team are Syrians themselves, so having Syrians being able to tell a Syrian story is quite important. Ultimately, any kind of justice that's going to be happening won't be happening in Germany or the ICC or the U.S. It's up to Syrian people themselves to decide what to do in terms of their own Truth and Reconciliation processes. And maybe that means that we should at some point be taking our site down. This is something I think that we are totally open to, but right now we've decided after consulting with a variety of archiving and documentation initiatives that the benefits of secure long-term preservation and analysis serve an immediate role as a public good, particularly when it's done in a transparent, detailed, and reliable way. We felt that having this information public can positively contribute to post-conflict reconstruction and stability. And that it can humanise those who are killed or injured, help societies understand the true human cost of war, and support truth and reconciliation efforts. We think that it's important to have this kind of content available so that you can create some kind of counter-narrative to all the propaganda and misinformation out there.

Q: TK

JD: The reason for content take-downs has changed over the past couple of years. At the beginning, we were seeing things like targeted flagging of a video or a Facebook post. So flagging being for example, if I were say that a particular post or video violates terms of service, right? So we were seeing things like coordinated flagging, where many people would simultaneously target particular videos or particular channels, regardless of whether the videos actually violated terms of service. That was a reason for content being taken down. We would also see things like the hacking of channels, and the deleting of a lot of content. But a couple years ago we were seeing that YouTube in particular, but also other platforms such as Twitter and Facebook, was using machine learning to detect so-called extremist content. We saw, overnight, 400,000 videos that we had archived on our infrastructure disappear publicly and become unavailable. We speak to channels and account holders, and we find out the reasons content had been taken down. Then we speak to platforms to try to address the reasons why

the content is down, and to see whether it is possible to reinstate channels whose content had been removed. In the case of YouTube, we brought to their attention the fact that certain channels were not propaganda, not extremist, but were in many cases media houses. Journalists and so on were being removed, channels that had in some cases 200,000 videos.. YouTube in particular has been quite cooperative in terms of reinstating those and making those public again.

Q: TK

JD: YouTube and Facebook and Twitter, social media channels, became these accidental archives where people were uploading content because they didn't have anywhere else to put it. But platforms are not a good solution for long-term safe storage. Particularly when you're talking about documentation of human rights violations. When conflicts start, often times people don't have time to organize effectively in terms of developing alternate strategies for content preservation, and platforms have filled this gap. But platforms are also used differently in different places, by different actors. In some cases, this might be for preservation of content documenting rights violations, in others, it might be for the spreading of misinformation of propaganda. So what you see in Syria might be different from what you see in Myanmar, might be different than what you see in Sudan for instance, but also within those countries different actors are publishing different types of materials. In Myanmar in addition to pages or channels documenting rights violations, you have a lot of Facebook pages in particular that are being used to spread or incite genocide So you don't necessarily want these pages to be public because of the ethical and real threats this content poses. You want them to be taken down, but you also want to have a record of that content so that you can potentially use it for some kind of future accountability mechanisms. In Syria, there is a very highly organised civil society, and you're not encountering those same challenges. In Yemen, issues of access mean there is a lot less documentation being uploaded, and civil society is not as organised online as in Syria, although very highly organised offline. Every conflict is context specific, so I don't think there's one solution you can implement at a platform level that's going to take into account all these things. We would also argue that there can be no satisfactory solution that would address these issues, particularly in the case of dealing with human rights, which has a much higher complexity of ‘problem’ than most commercial tools, interfaces, or platforms are able to solve. Silicon Valley style approaches towards creating platforms often assume a simple problem solved by a technically complex solution with the goal of creating user convenience. Think about the example of connecting available drivers and passengers, or wanting to share your resume with potential employers. Convenience drives users to platforms. For human rights documentation, even correctly identifying the ‘problem,’ or ‘problems,’ or potential users is hard enough, and finding a solution at a platform level that’s going to take into account all these issues and considerations is next to impossible.

Q: TK

JD: If we're talking about accountability, things are different in different places, so I don't think there can be some kind of universal accountability solution put in place. It has to happen more on a case-by-case basis. Facebook having content that's promoting or inciting genocide in Myanmar is really not the same issue as taking down content in Syria that civil society groups are using to document state or non-state violence. And within even Syria, there is a difference in the content that media groups are uploading that shows the effects of an attack, and the content that, say, Ministries of Defence are uploading that show an attack being perpetrated. Is some of this content graphic? Of course. But it’s not necessarily promoting violence or encouraging dangerous activities. Content can be violent because, in many cases, it shows the suffering of people as a result of attacks against them. So it's important to draw a

distinction between content that itself is violent, and content that is documenting violence being committed. There's definitely online violence, but there's also offline violence and, you know, behind all of this violence is people, right? I think platforms aren't off the hook entirely in terms of being used a tools to spreading misinformation, propaganda, or violence, but behind that content that's being spread, there's people who are behind them, who are posting these things, who are spreading these things.

Q: TK

JD: I definitely think some kind of process would be helpful in the U.S., but I don't know where to start. And also, I don't think it would necessarily happen right now, in any form. As you were talking about, all these open wounds, genocide of Native Americans, slavery, Jim Crow, the war on drugs, not to mention everything that the U.S. has been involved in globally. So I really have no idea where that process would even start or who would be involved.

Q: TK

JD: There's a real push right now for not having terrorist content online. In Europe, in New Zealand, in the U.S., there's a push to restrict violence or propaganda or extremist content. I can understand where people are coming from in terms of this, and I think they are very well intentioned. But the work that we do in terms of preserving at risk documentation of human rights violations would be extremely impacted by these types of policies. If you have a policy on content moderation of online extremism or terrorism, then you're getting into questions of how do you define what terrorist content is? Is the state considered a terrorist group? Where do you draw these lines? A lot of the stuff we're archiving and using for investigations and legal work could fall under the criteria that is being set for extremist content, meaning if such policies are enacted, we soon might face much higher rates of content takedown than we are seeing even now. I think there's going to be more of a push to take things down quickly, which has benefits but it also has drawbacks. Maybe you don't want to be spreading a video of a shooting at a mosque in New Zealand. But it's important, especially in areas that don't have as much coverage or access to a global audience, to have these incidents known and recorded. So there's a danger in pushing back too much in terms of what's being uploaded. It's going to affect human rights monitors and human rights defenders and groups that are already marginalised or face issues of having injustices being visible much more so than the platforms themselves .

Q: TK

JD: We're actually doing a project now where we're doing long form interviews with some of the videographers themselves who filmed a lot of this content. So that they can talk about why they filmed it, where they filmed it, how they became involved, what their story is, what their motivation is. What's the story behind the camera, which I think will complement the videos and other forms of open source documentation, as you're putting people back at the center of it.