Bridging the gap between anti-corruption advocacy and technology

User research, insights, and learning strategies

With ever-lowering barriers of access to technology, and thanks to consistent wins in advocating for governmental and corporate data disclosure, investigative journalists and transparency activists have at their disposal a wealth of sources and tools to uncover corruption and advocate for accountability. In recent years, there have been efforts to create guides, tutorials, repositories and online learning tools, with the goal to simplify the process of finding the right tool for the job, and the right dataset for analysis. However, the availability of these platforms didn't always guarantee the expected outcome of a more skilled, evidence-driven civil society. While there are shining examples of activists and journalists across the globe effectively analysing the available data (and advocating for more data openness) in order to combat corruption and demand accountability, they still don't represent a default state — many organisations whose work would be greatly simplified with stronger data skills still struggle to integrate data and technology in their day-to-day work.

Introduction

We believe there is a **training and skill-building gap** that is not being adequately addressed by the available resources: for a variety of reasons, journalists and activists are struggling to up-skill themselves and their teams to a level where the wealth of resources and technology becomes a well-understood, core part of their daily work. Our goal is to identify strategies that could help bridge this gap, and implement pilot projects to test our hypotheses in the real world.

In order to better understand how to bridge this gap, **we conducted user research** across groups that would be commonly recognised as "successful" in utilising data for anti-corruption advocacy. We tried to find out both how they work with data and technology as a team, and how they got to the point of being an effective "data for anti-corruption" outfit.

Based on our interviews, we defined the following steps:

- **A conceptual framework:** what does it mean, in practice, to be an effective team using data and technology for anti-corruption? Which skills are needed the most, and which skills are overrated? What combination of skills and attitudes works best in most contexts?
- **Preliminaries:** which considerations do we need to keep in mind when designing training activities and implementation support for a specific team? Which factors do we need to account for, in order for the training to be as effective as possible?
- **Skill-building activities:** how can we support, in practice, the bridging of this gap? Which activities can we design and implement ourselves, and which ones need wider support?

We expand upon each of these steps in dedicated chapters. In the Annexes, we explore practical operational steps for specific activities, and define our user research methodology.

User Research Overview

A brief overview of our findings, to help understand how they shaped our strategy. For more information about the user research, please go to Annex II.

We interviewed representatives of organisations known for successful work in using data and technology for anti-corruption, as well as experts in data trainings, especially for anti-corruption and investigative journalism. *We cast a wide net geographically, to complement the relatively small number of interviewees (fourteen in total).* Most of our interviewees came from small NGOs (4-20 people) and generally worked in teams of 5-6 on their projects — although some interviewees came from much larger outfits (international investigative journalism projects, for example) and were used to working with, and coordinating, dozens of team members.

We were interested in capturing more than just specifics about their organisational setup, or the tools they use, so we designed our interviews to be open-ended and qualitative, in order to inspire free-flowing conversation.

The main conversation areas were:

- **Pressure points:** what are the main roadblocks in working with data and technology for anti-corruption, both internal (team capacity, etc) and external (lack of good data, etc)?
- **Tools and platforms in use:** what sorts of programs, computer languages, etc does the team use each day, who uses which tools, and how comfortable are they with their level of tech and data use?
- **Capacity building:** how do they currently train in-team and in-organisation, both on tools/tech and techniques (for investigative journalism, data analysis, etc)? What types of trainings would they love to have access to, if money and time were not a problem?

Pressure points

> *"I cannot overemphasise the low degree of tech literacy."*
>
> *"It is very frustrating that majority of the media can't do basic analysis - I mean very very basic. Perhaps not that they can't - but that they don't want to."*

Most interviewees agree that the main pressure points are **low tech literacy** of the staff and collaborators (external journalists, for example), followed by **poor quality and low availability of data.** Depending on the context (geographic or sectoral), usually a massive amount of time is spent procuring, cleaning, and preparing the data for analysis — a task rendered even more frustrating when most staff and collaborators don't have the skills for basic cleaning and sorting of data into spreadsheets. **Inflexible donor funding** is a factor when skills training is required, but difficult or impossible to combine with project-driven donor proposals that don't leave enough space for capacity building. **Uncertain impact** of anti-corruption projects makes it hard for organisations to justify long-term anti-corruption projects, if they aren't able to see or prove results within the relatively short time frame their stakeholders require. **Knowledge management** gets an honorable mention: the bigger the team, the more sharing information becomes complex —a problem that is compounded by low tech literacy skills.

Tools and platforms in use

> *"I tend to still use Excel: obviously very limiting, but lots of times that's all I need."*

The most important tool by far is **spreadsheet software**. It is both the most used in day-to-day work, and the most valuable team asset: the more people on the team know how to use spreadsheet software well, the easier it gets for everyone else. This is especially true for the more highly skilled technologists on the team, whose expertise can shine brighter if they are not overloaded by requests that don't require more than spreadsheet knowledge.

The second step is what we defined as "Basic +:" user-friendly tools that complement spreadsheet software skills with more advanced features to clean (OpenRefine), or visualise (Tableau) data. The interviewees who successfully use Basic+ tools are also the ones with a high overall capacity in basic spreadsheet skills.

Regarding more advanced technological tools and platforms, interviewees either outsource, or have one or two deep divers on their team — people dealing with specific advanced problems.

Capacity building

Overall, interviewees expressed a clear distinction between self-taught and guided capacity building. Self-taught is generally unorganised, and perceived as much less useful (if useful at all) than guided capacity building.

> *"In my context, there's no history of self-directed learning. MOOCs don't work; it has to be one on one."*

Most smaller organisations don't have any kind of structured capacity building strategy, mostly due to lack of funding and time. Usually they count on peer support — sometimes by literally grabbing people at conferences in order to ask them questions about their project. The larger organisations we spoke to mostly have well-defined, complex internal training processes.

A number of interviewees were very supportive of **peer sharing**: either by spending ample time with other successful teams and learning from their practical experience, or by embedding their team members in technical projects with expert developers and project managers. They would all like to have more occasions and spaces where this type of practical sharing can happen.

> *"The way that people learn is by working with people, or listening to real people. I don't have enough clear cases of journalists saying 'I did this MOOC and now I understand this'."*

The Conceptual Framework

"Your conceptual framework is awfully ambitious, but I agree with it."

In order to define learning and training activities, we needed to establish our goal: what is the aspirational "perfect team" we are trying to help organisations achieve? This conceptual framework drives our strategy in defining, prioritising, implementing, and monitoring the success of our activities.

During the interviews, we noticed that a particular shared "perfect team setup" started to emerge: all interviewees shared a very similar combination of skills and approaches in working with data and technology for anti-corruption that ensured better results, smoother inter-team working relationships, and reduced the risk of burnout of single individuals overloaded with the bulk of the work. This helped us conceptualise the idea for a framework of what an effective data team should look like — we then sense-checked this idea with all the interviewees and found it to resonate with everyone we talked to.

Our framework

1. **Everyone is a spreadsheet master**
2. **No more than a deep diver or two**
3. **General knowledge about what other technologies can (and can't) achieve**

1 - Everyone is a spreadsheet master

> *"The most reliable software we can use is Excel: it allows you to start understanding how to write queries, how to use logic, how to see results quite easily."*

Everything we learned from our interviews seems to indicate that, with spreadsheet software skills, *more is indeed more*. Interviewees with team-wide solid spreadsheet skills are the ones who report the least frustration with data analysis, cleaning, and management. They are also the ones most ready to embrace "Basic +" tools like Open Refine, or Tableau — because solid skills in spreadsheet software makes the transition much smoother. The ones with low team-wide spreadsheet skills invariably want to be more skilled, because they recognize how much of their work depends on knowing how to do spreadsheets really well. In addition, they reported that their resident technologists (the *deep divers*) often experience burnout because of the added workload of supporting spreadsheet-type tasks across the team.

This focus on spreadsheets is no mystery, as much as it's no surprise that, in the for profit world, "advanced Excel user" is one of the most appreciated skills on a CV. Far more than just being an ordered way to write lists of things, **spreadsheet software captures the fundamental concepts of dealing with data in two dimensions** (rows and columns). Spreadsheets can identify data types, enable data analysis, perform complex logic operations, design visualisations, import and export other common file formats (csv, JSON, xml), and much more. Spreadsheets are easily shareable, don't require internet connectivity, and quality software options, in most cases, come pre-installed on computers (and there is a super-decent, free, open source alternative).

While spreadsheets obviously have their limitations, they can usually get 80% of the work done, and the fact that data can be exported in a variety of formats means that they are perfect for preparing data before it moves along the chain of complexity (to programmers, database specialists, etc).

Most importantly, **becoming good at spreadsheets means becoming deeply comfortable with the fundamentals of data management, analysis, and visualisation.** Knowing how to clean and prepare a spreadsheet for sorting by date, name, pay grade or country requires having understood:

- How different data types and content types work;
- How to assign and manage metadata (headers);
- How to clean and prepare database records;
- How to perform search queries;
- And much more.

We refer to the skillset described above as **data fundamentals.** Arguably, in our context, knowing data fundamentals is what differentiates a spreadsheet user from a spreadsheet master — and all other more advanced tools and technologies are based on the same data fundamentals knowledge.

In our opinion, the first step towards an effective data team is making sure everyone's spreadsheet skills are raised to a high enough level that each person is comfortably capable of importing a relatively dirty dataset, cleaning and preparing it, investigating it through sorting, filtering, pivot tables and other techniques, and telling a visual story of their findings — all within the spreadsheet software.

2 - No more than a deep diver or two

"We find it impossible to equip everyone with the entire knowledge of the data pipeline, it's not feasible."

Deep divers are experts in a specific advanced programming **language** (Python, Ruby, PHP, Javascript, C++), or **skill** (scraping, API manipulation, data storytelling, SQL manipulation).

The importance of deep divers is hard to overstate: if it's true that spreadsheets get you 80% of the way there, the last 20% can be incredibly complex, or even impossible to conquer without advanced skills. For example, if an organisation is planning to provide an online searchable database of their datasets, they will inevitably need at least a database expert to combine their spreadsheets into an online relational database. Programming languages like R and Python are particularly well-suited for number crunching of huge amounts of data in order to plot regressions, conduct complex (multivariate) analyses, run supervised machine learning algorithms etc, where spreadsheets start to show their limits.

Having a deep diver automatically expands not only the team's overall skillset, but also their understanding of what is possible to achieve with data and technology. Working in close contact with a GIS specialist, for example, helps the rest of the team understand how much is (or isn't) possible, given the right data. This in turn expands their conceptual limits, and lets organisations envision more ambitious projects and goals.

At the same time, deep divers are usually no more than 1 or 2 per team: while it is true that dedicated deep divers are hard to come by, most interviewees agree that having more than a very small core group of deep divers wouldn't be as useful, even if there were resources and opportunity to hire more deep divers. Especially if the entire team has solid spreadsheet skills, the deep diver is free to focus on very specific high level problems.

There is no specific programming language, or tool, that is preferred by the majority of interviewees — usually there is an organic adaptation between the team and the deep diver, where the divers' skills are understood and recognized by the rest of the team, that then learns exactly which puzzles and problems can be solved best by the diver.

Deep divers are usually recruited from the outside, but in some cases have grown into their role internally. The interviewees who have up-skilled their internal staff believe it to be very beneficial for both the organisation and the individual. In fact, the hardest part about up-skilling internal team

members is the ability to recognise there is potential, and provide timely and positively challenging professional development opportunities.

In some cases, the role of deep divers is played by external consultants or organisations. For example, [School Of Data fellowships](#) provide long-term tech and data support to organisations. [The Engine Room](#) offers a number of different deep dive support frameworks: the Matchbox project, the Replication Sprints,, and their fee for service offerings all aim to bridge the tech capacity gap. This type of collaboration is usually as useful as the bond between team and consultant is strong: for example, long-standing, mutually beneficial collaborations with "tech for good" organisations that have high tech skills and are always on the search for meaningful projects. When the deep diver is an external "parachuter," the organisations must have solid understanding and knowledge of how and which technology can support them best — or external support in choosing the right consultant.

3 - Ambient knowledge about what other technologies can (and can't) achieve

> *"We are good at knowing which technologies to go and look for, even if we don't have skills to do it."*

While spreadsheet skills and a dedicated deep diver form the backbone of technical capacity for a data team, sometimes a project might need specific solutions that aren't covered internally. For example, a group might want to tackle GIS analysis on a project while not needing a full-time dedicated GIS person. For those types of cases, a team needs to have solid knowledge *about* different technologies: when they might be useful, what kind of answers they might give, and where to look for the best technological partners. In fact, assuming that any team will have the opportunity to bring all the possible technologies in-house is obviously impossible. Developing this solid ambient knowledge makes it possible for organisations to extend their reach even further, and helps them limit the risk of becoming a one-trick data pony: an organisation that is really good at one specific technology runs the risk of biasing their work based on the toolset they have, and might fail to recognise new opportunities.

One of the ways interviewees expand their ambient knowledge is through conferences and peer sharing: learning how other groups tackled similar problems, but with different technologies, helps organisations think how those examples might be implemented in their own projects.

Another important aspect of having a solid ambient knowledge is understanding the limits of technology: in many cases, a superficial understanding of a technological tool or solution might oversell its potential[1] and cause teams to over-invest, with underwhelming returns, and a waste of money and resources.

---

[1] Two current (2018) examples: machine learning and the blockchain, apparently capable of solving humanity's problems problem humanity ever had.

Preliminary considerations when designing implementation support

While we believe we can define templates for activities and trainings, we also understand that there are many factors that will affect a team's ability to learn new skills, and put them in practice. For this reason, after defining our conceptual framework, but before designing training and activities, we focused on establishing a series of preliminary considerations that need to be understood, on a case by case basis, in order to adapt the activities to fit as well as possible its recipients.

For an example of an exploration questionnaire, please go to Annex I - guided team preparation.

It's not (just) about anti-corruption - but sectoral knowledge is crucial

While this user research is focused on understanding the data and technology needs of teams working on anti-corruption, it soon became clear that the tools and approaches themselves are in no way special or in some way endemic to the anti-corruption sphere. This is also obvious from the conceptual framework above: the same framework could be transposed to advocates working in health, environment, human rights, or any other social good sector.

However, this doesn't mean that the issue area doesn't matter; rather the opposite. While tech skills have no particular affinity for a given issue area, both learning them and using them requires massive practical focus and significant sectoral knowledge. In learning activities, especially in adult learning, working on practical examples is the best way to absorb new knowledge. In implementation, most interviewees agree that having sectoral experts will make or break a project: for example, when dealing with oil exploration or extraction contracts, the team needs to be able to count on significant sectoral expertise in the legal framework of how licenses are awarded and by whom; in procurement monitoring, the team needs to know the ins and outs of government procurement mechanisms in order to effectively monitor for irregularities, etc.

In a nutshell, sectoral knowledge enables teams to *ask the right questions* — which in turn enables them to choose the right data and technology strategy to search for answers.

Data quality and availability will always influence team capacity

You can't (easily) change data you have in your country. This means that anti-corruption data teams in different countries will need to have wildly different levels of skills to achieve the same result. In a country like Ukraine, for example, e-procurement data is readily available and comes pre-packaged in comprehensive business intelligence dashboards that are open to all citizens. In other countries, activists working on procurement monitoring might not be as lucky, and most of their time will be spent painstakingly gathering datasets of dubious quality, and squeezing tiny droplets of content out of them through sheer manual labor.

Improving the geographic (or sectoral) context usually takes significant time and effort, and in many cases it won't be possible to improve data quality within a useful timeframe. Advocating for national

implementation of FOI laws, for example, or lobbying for more transparent procurement practices cannot be approached as a subset of a specific anti-corruption project. They are entire programs in their own right, and usually require years of painstaking negotiation efforts. Understanding the importance of the state of the data ecosystem on project design makes preliminary assessments all the more important, when planning support activities, to make sure that it is done on tools and techniques that will be most useful for that particular context.

Institutional readiness

Building new skills takes time and effort: most of it after the skill-building activities have been completed. No matter how well a training activity has been planned, and how effectively it has been implemented, the only way to ensure that new skills become an active part of a team's toolbox is regular and continued use. For this reason, it is crucial to establish the level of institutional readiness to provide newly-skilled staff with time and space to effectively integrate new skills in their daily operations.

Trainings and other skill-building activities also depend on the participants' preparation: there needs to be a shared agreement that team preparation is crucial for an effective training. This institutional readiness could also come from an initial training/ workshop (what School of Data calls awareness raising "trainings"), or could be conducted through mentorships as part of a programme like the Hivos Open Contracting Programme.

Learning activities

After establishing our learning goals through the conceptual framework, and defining preliminary considerations that need to be addressed before activity design, we focused on designing lists of activities that fit with the framework, and respond to interviewees' needs.

The overall list of activities tries to directly answer to needs we heard in the interviews; we try to list a comprehensive set of activities that would address most of the expressed needs. Some activities in this section are practical: they are understood, use established and tested methodologies, and the resources required are known. Others are new: moonshot ideas that need piloting and understanding further, with resource-heavy or sometimes unknown requirements. Ideally, HIVOS and SCODA would start from implementing the more well-understood ideas, while looking for opportunities to test the more out-there ideas.

Team readiness assessment

In order to understand how to help a team or an organisation achieve the goals of our conceptual framework, we first need to pinpoint their starting point — how close are they to achieving the goal of becoming an effective data team? **The readiness assessment would take the shape of a series of multiple-choice open questions**, in which the team representative is interviewed by the training designer. You can find an example template of the readiness assessment in Annex II.

The assessment works best if it is guided: while teams can self-report on some parts of it, the devil is in the details. For example, when trying to assess the team's spreadsheet skills, it is important to be able to ask follow-up questions to understand what exactly a team means by "high" or "low:" an experienced, tech-savvy interviewer who both knows very well activities they're designing, and the learning objectives of the training, will know how to investigate deeper and paint the appropriate baseline picture. In many cases, an external interviewer will be able to surface the "unknown unknowns" — uncover blind spots in capacity or institutional readiness, as well discover hidden, unrecognized talents.

While it is important to have an expert guiding the assessment, the questions themselves can be formatted into an online tool: for a very good example, see [Alidade](#) by The Engine Room. Alidade is an online guided process for selecting the best technologies based on the organisations' needs and resources. Based on real-life experiences, it works best when an organisation is guided through the answers by an expert facilitator. The added value of the online platform is that organisations are able to save their assessment, come back to it and re-evaluate it as they grow in skills.

Learning Data Fundamentals trainings

These trainings are three- to five-day intensive workshops, equally split between learning skills from experts, and skill-building laboratories based on real-life examples of datasets. The components of a data training are considerably well understood and tested: as a reference standard, School Of Data has a number of tested, successful skill-building techniques and models across the data pipeline. For

Data Fundamentals, the learning objectives should focus specifically on fundamentals of data and the use of spreadsheet software as the technical, software component. The activity would be split into two streams:

- Extensive hard skills training in the fundamentals of data and spreadsheet software
- Real-life use cases and examples of how those skills apply within an anti-corruption investigation

The two streams need to be carefully blended for each training, each building on the other, in a loop of laboratory-style exercises (paired work, achieving specific goals) and discussions around real-life applications (how did organisation A implement skill B, how did skill B fit within the larger team structure, what were the points of failure and hurdles, etc).

The training objectives are directly geared towards achieving part one of the conceptual framework (everyone is a spreadsheet master), as well as part three (general knowledge). The former is a direct skill-goal, while the latter will be gained through the use of real-life examples. These examples need to be carefully chosen based on the team readiness assessment, so that participants can get inspiration from, and learn about, how other teams addressed similar challenges.

What if we want to work on actual challenges the teams have? That is definitely possible, and in many cases it might prove beneficial. However, the risk is that the importance of the learning process gets overshadowed by problems and challenges of the real-life projects that are not linked to learning skills. The choice of working on the teams' projects depends heavily on the readiness assessment findings.

Continued support in implementation

Interviewees were well aware that achieving the goals of the conceptual framework requires time, patience, and effort. An activity such as the Data Fundamentals training will never be able to support long-term consolidation of the newly acquired skills, but it can be designed so that it provides useful jump-off points towards more complex and time-consuming skill building. The main factor to consider is enthusiasm: the somewhat dry and complex skills required can quickly dampen spirits, especially within the activist/journalist for anti-corruption ecosystem where the majority of people come from non-STEM backgrounds.

For this reason, we are planning long-term continued support activities in order to:

- Support teams to consolidate their skills
- Provide deep diver support where needed
- Foster potential new deep divers from within teams
- Help teams overcome inevitable slow-downs in enthusiasm
- Troubleshoot unexpected problems as they arise

Data Fellowships

The long-standing and proven methodology of supporting data fellows in organisations and communities is the perfect accompaniment to the Data Fundamentals training. The added value of looking at the data fellowship program through the lens of this document is the preliminary assessment: organisations like School of Data will be much better positioned to help fellows design their fellowships with a structured preliminary assessment. Data fellowships are also an effective way to ensure that there are individuals on the ground who are able to provide this type of technical advice and skills development to civil society actors.

Guest Sprinters

A strong insight from interviewees was the appreciation of learning *how to think about* data and technology, rather than just learning how to implement tools in practice. Interviewees with little technical knowledge who have participated in design or development sprints have said that their appreciation for, and understanding of, how experts use data and technology, greatly helped them in gaining deeper understanding of, and honing their approach to data and tech.

The model we're following here is that of **sprints**: both [Design Sprints](), and (Agile) [Development Sprints](). Ideally, both would be combined into a single sprint experience that combines design and development. The goal of the sprints isn't to solve huge problems, or create complete solutions: their goal should be small, and incremental, an addition to the existing project components, or an improvement of a component. The real goal is *experiencing the process of* design and development sprint with a group of experts from other domains, in order to expand the general knowledge of what data and tech can and cannot do, and gain appreciation for the intricacies of project and product design and development from different perspectives.

The Guest Sprinters are a sort of reverse fellowship: short, but intense deep dives into process. They can be conducted in two ways: team members can actively participate in "project sprints" with experts in design and development of technological tools, or a visiting expert can join the team for a guided, focused, short development sprint on a small part of their own project. The former is most useful if the team members have a vested interest in the final product of the sprints (such as, for example, cleaning governmental datasets that are relevant to the team's work).

Bridging mentors

During any of the above activities, certain team members might shine more brightly: be more interested in learning not just about data and technology, but also about what it means to be a data and technology champion in their own organisations. These individuals can be upskilled into "bridging mentors" — internal team members that also wish to carry the flag of the importance of integrating data and technology into the organisation's work, and wish to help others make the leap. We would focus on providing these bridging mentors with tools and techniques to improve their methodological thinking — their ability to learn about learning — and provide them with ToT (training of trainers).

Bridging mentors are different than deep divers. While they can be the same person (and usually are), any team member can become a bridging mentor: there is no need for advanced technical skills to be a champion of data and technology within an organisation or a team.

This is one of the "moonshot" ideas — because it is not yet understood, so time and resource requirements are not defined, and it is not clear how much of this knowledge can be templated, and how much of it depends on personality of individuals. The idea is to design a methodology to create bridge mentors.

Further user research

This user research gave us enough insight to build the conceptual framework and formulate the activities, but we're sure there is more to be learned, especially if we wish to expand on specific countries of interest.

The user research structure should be twofold: on one hand, the same interview methodology can be followed, to gain new insight or confirm what we learned until now. On the other, we can test the conceptual framework itself through conversations with experts, and see if they have anything to add or remove, or if they are willing to test it out in their environments.

"How It's Made" series: learning through well-told examples

Many interviewees spoke to us about the importance of learning from their peers, through real-life examples. They were mostly interested in learning about the nuts and bolts of a project: why they chose a certain technology, how well (or badly) did it perform, how much manual work went into the whole process, etc. The learning process they were describing was reminiscent of the "How It's Made" documentary series, where viewers are treated to the creation process of common household objects, from the raw materials to the final product. Another good example comes from the musical podcast "Song Exploder," where listeners get the chance to hear a musician deconstruct one of their famous songs lyric by lyric and instrument by instrument, and tells the story of how it all came together. We believe that such detailed **storytelling of process** should have more space and attention in the field of data and technology for anti-corruption: incredibly interesting and useful information is lost either because it is not documented, or because the final documentation only tells the final, polished story, glossing over the ever-so-useful and interesting implementation details.

Storytelling formats can be divided into **live** and **pre-recorded**:
- Live
  - Webinars
  - Side events at conferences
- Pre-recorded
  - Podcasts
  - Video series
  - Articles

For both formats, content is key: the main first step will be sourcing interesting and insightful stories told by engaging people. As well, both formats need to be carefully scripted. Live formats can be developed on a budget, and therefore are the best way to pilot the activity and test audience interest. For pre-recorded formats, the main challenge is **ensuring high production values**: in the social media ecosystem, any scripted product needs to compete for viewers' attention on the same level with high-budget, high-expertise products made by industry-leading companies. Deciding to develop pre-recorded formats without sufficient resources for an industry-level product might ultimately prove to be a waste of time. Tactical Technology Collective [Exposing the Invisible](#) film series is a great example of the type of production values and resources required to match the quality of the content.

Annex I - Team readiness assessment

An example template list of questions to ask to develop a team readiness assessment, with interviewer's notes

---

What to listen for in general, throughout the interview

- Human system dynamics
    - Do particular interpersonal events or interactions cause expressions of delight, or frustration?
    - Are you sensing recurring behavioral patterns between team members?
- Institutional relationship
    - Are you sensing tensions or difficulties between team and institution?
    - Does the interviewee regularly use phrases like "if only it were possible in this org," "there's no space for new/different/etc"
- Relationship with data and technology
    - Does the interviewee approach technology with uncertainty and suspicion? What might be deeper reasons for this? *Follow up "but why" questions*
    - Listen for levels, and combinations, of the following emotions about data and technology:
        - Excitement
        - Frustration
        - Fear/uncertainty
        - Neutral/functional perspective
    - How do they interrelate? Was it possible to sense the team-wide "zeitgeist" emotional composition?
- Sectoral expertise
    - Does it sound like the team has solid, deep substantive understanding of the topics they are working on?

Question	What to listen for

● Please describe, in detail, your team's composition
   ○ How many are you?
   ○ Which roles do your teammates cover?
	● *Which roles aren't being covered?*
	● *Does the interviewer give more or less emphasis or interest to specific roles or expertises?*

● Please tell us a story about an anti-corruption project you and your team worked on
   ○ What was the problem you were trying to tackle?
   ○ Tell us about planning, design, team roles, implementation, in as much detail as possible
   ○ What were the results? How satisfied are you with them?
	● *Listen for process: what frustrations are floating in the air?*
	● *How do team skills relate to sectoral knowledge?*
	● *How do team skills relate to technical literacy?*
	● *Telltales of human system patterns*
	● *Listen for external factors:*
	   ○ *What is the context they are working in? Is data available, is it of good quality? Inversely, is it hard or impossible to access?*

● Which technology and data tools do you use, and why?
   ○ Which team members use which technologies?
   ○ Do you outsource tech work? If yes, to whom, why, and how satisfied are you with the results?
	● *Overall tech literacy of the team*
	● *Approach to data and technology*
	● *(see general above)*

● What are your goals in improving data and technology skills of your team?	● *Are the goals realistic?* ● *Do they align with your perceived data and tech readiness levels of the team?*

Annex II: Survey Methodology

1. Inception and sourcing interviewees

After defining the user research framework and focus, we created a list of participants that would provide the largest possible scope of opinions and ideas within the time and resource constraints.

We were striving to be as globally applicable as possible. While this is a very ambitious goal, considering how much geographic context plays a role in availability of data, resources (both tech and investigation), internet availability, etc, the user research covered parts of the following regions:

- Eastern Europe and the Balkans
- Africa
- South America
- Global players (not linked to specific regions)

The choice not to focus on Asia, Australia, Europe, or North America is due to a couple of factors:

- **Baseline**: it is assumed that there are enough cultural complementarities between regions that the geo scope described above will give enough template-able information (if any such templatization is indeed possible) to be applicable to the missing regions
- **Resource constraints:** we prefer to focus the limited amount of interview time to regions where there's on average less budget transparency

The audiences were separated into three categories based on their understanding and usage of technologies for anti corruption work: **growers, swimmers and flyers**.

Growers are journalists or activists with concrete experience dealing with anti corruption campaigns, working in an environment where technology is understood but not well-integrated yet, or where they have shown or mentioned that their success was achieved not thanks to technology, but more despite the difficulties in using data and technology, due to their lack of expertise, lack of training materials etc.

Swimmers are journalists or activists with concrete experience dealing with anti corruption campaigns, but with a stronger/firmer grasp on data and technology; more evolved and nuanced understanding of processes, formats, data modeling etc; a more integrated presence of technology (and technologists) within the orgs, either through in-house staff or through smart, well-understood and fruitful relationships with service providers. Aware of the learning curve (more knowledgeable of their known unknowns).

Flyers are high-capacity journalists and activists who use highly advanced technologies for their work; effectively create the anti-corruption tech others will be using one day; are requested to support

and train others on a regular basis, either through tech support in establishing platforms, or through capacity building on using those platforms.

We interviewed fourteen individuals, split coherently between the above categories.The interviewees had the following geographic distribution:

- 6 National (Hungary, Nigeria, Peru, South Africa, UK, Croatia)
- 5 Global
- 1 Regional (LatAm)

2. Survey methodology design
The survey methodology was defined a priori, and refined after the initial interview.

The interview structure
- What does a typical anti-corruption investigation looks like?
  - Which types of expertise do team members have
  - How is data being collected and analysed at the moment
  - What types of campaigns are being built around the investigation
  - How do you deal with trainings, if you do?
- Deeper dive into pressure points
  - What are the main frustrations around working with data for anti-corruption?
    - Methodology
    - Lobbying for data
    - Cleaning
    - Analysing
    - Standards
- Knowledge and use of tools and platforms
  - Which tools and platforms do interviewees use, and why?
  - Effective data team:
- Current capacity building practices
  - How do they currently train teams on using data for anti corruption?
- Which learning platforms and tools are they familiar with?
- The perfect world
  - If available, what kind of learning tools and practices would they prefer using?
  - Are there any good examples outside of anti-corruption that you would want to suggest we look at?
- Anything you'd like to add?

3. Interviews
The interviews were conducted during the months of August and September 2017. A typical interview lasted 45 minutes. Interviews were conducted online, over Skype or similar voice-only tools.
Each interview started with a framing paragraph:

Thank you for your time. I am supporting HIVOS in developing a "data for anti-corruption" training package. Since we're aware there's already much out there, and at the same time we don't want to assume what shape this training package should have, we're conducting interviews with people we intend to use this platform, as well as people who have been training and developing tools for anti-corruption because of their unique perspective. The result of these interviews will be a framework for development of the training package based on what we discover through the interviews.

The interview will take 30-45 minutes, I will be taking notes in real time, and it will not be recorded. I am taking notes on a Google Doc so if you prefer that I take notes offline, do let me know.

Any questions before we start?