GlobalGiving Storytelling Project Phrase 3

Fact sheet
Timeframe: 2012
Goal: Get 100 organizations (of our 500+ org network in East Africa) to use the 2010-2012 storytelling data in some operational, evaluative, marketing, strategic thinking, or advocacy context.
Strategy: Regular community meetings in 6 to 9 locations in Kenya and Uganda where we had existing NGO partner networks. The meetings were an opportunity for organizations to learn what community feedback said, and network with each other around common goals (presumably goals that are influenced by community feedback).
Framework: Borrowed from **positive deviance** (showing local people a picture of themselves and highlighting the best "positive deviant behaviors" found within their group and the **Clinton Global Initiative Model**, whereby we facilitate roundtable discussions of local actors with the goal of having them set goals. "What is one thing you can do between now and our next meeting in 3-4 months?" Those organizations returning would be expected to report out to their peers on how well the one thing they tried worked.
Actual locations: Kakamega, Mathare, Kibera, Kisumu, Masaka, Kampala, Jinja, Busia, Vihiga, Bungoma
Actual goal met: Less than a dozen orgs showed any sustained interest in the project.

Lessons Learned from the org-community feedback part

Ineffective

- Of these locations, only three of them met more than once.
- While attendance remained high, there was no actual increasing in the "doing" or "reporting" parts.
- **Magnet for bad-fit orgs**: Organizations that wanted to use GlobalGiving for fundraising had already moved to our website, and those with no interest in being that kind of a partner were increasingly showing up to this meeting without setting goals or making progress towards them.
- Stories and local feedback were the focus of the meeting but did not play a crucial role in **guiding actions**. In a few cases where they did trigger a discussion that led to actions, most of the actions were about **passing the buck** to another player who ought to act on the problem.

Unsustainable

- With more money and staff we could have made something work, but it was becoming clear that the in-person meetings model would never translate to GlobalGiving's online platform, would never be sustainable on both financial and social measures. In the best cases, organizations were networking with each other through the meetings but neither GlobalGiving nor the story data need be present for such an activity to occur.
- We focus on doing things that are effective within a cost-recovery model. The cost of face-to-face meetings and staff to manage them would never pay for itself.

Why did it fail?

- Our assumption that community feedback could be gathered broadly and cheaply was correct. But turning 60,000 individual "community effort" stories into a set of coherent and meaningful statements or needs from a community required tools that were not available. Perhaps, with better tools and more knowledge about how to translate thousands of voices into dozens of focused, shared petitions or common praise to expand what works, this could work. Failure was a **big-data problem** and not a data collection problem.
- **Weak incentive to act**. Each organization could act on community feedback, but it was easier to do nothing as there was no big carrot for addressing the problem. There was no election (political mechanism), no promise of future grant money (funders' mechanism), no media soapbox (TED), social praise from peers (CGI). This is in contrast to the basic GlobalGiving value-proposition: use us to fundraise and you get money for exactly what you want to do, while building up your public reputation and credibility at the same time.
- Our storytelling method relies on large samples of open-ended narratives that are later filtered. Our filters are acceptable but not amazing. We need to invest more in the work needed to enhance how we screen the flood of incoming data and **curate smaller subsets** of "rich narrative" with depth and nuance about the situation. We also **need to demonstrate** to the sector that a "broad collections with filtering" approach works – this is a diversion from the "safe" survey approach, which makes results publishable but not necessarily accurate because of the huge self-bias problem in M&E.

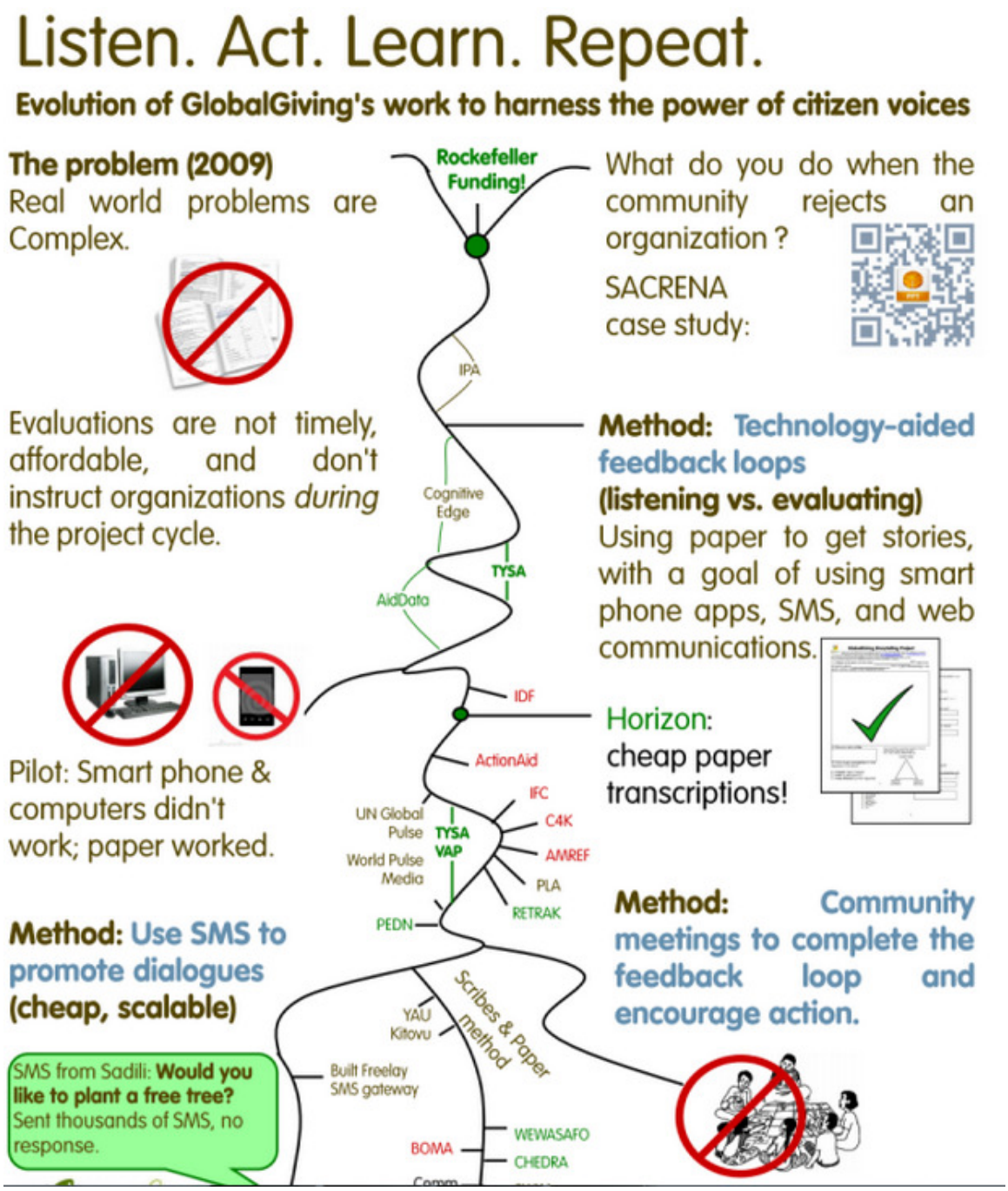
What are we keeping / adding / dropping / changing?

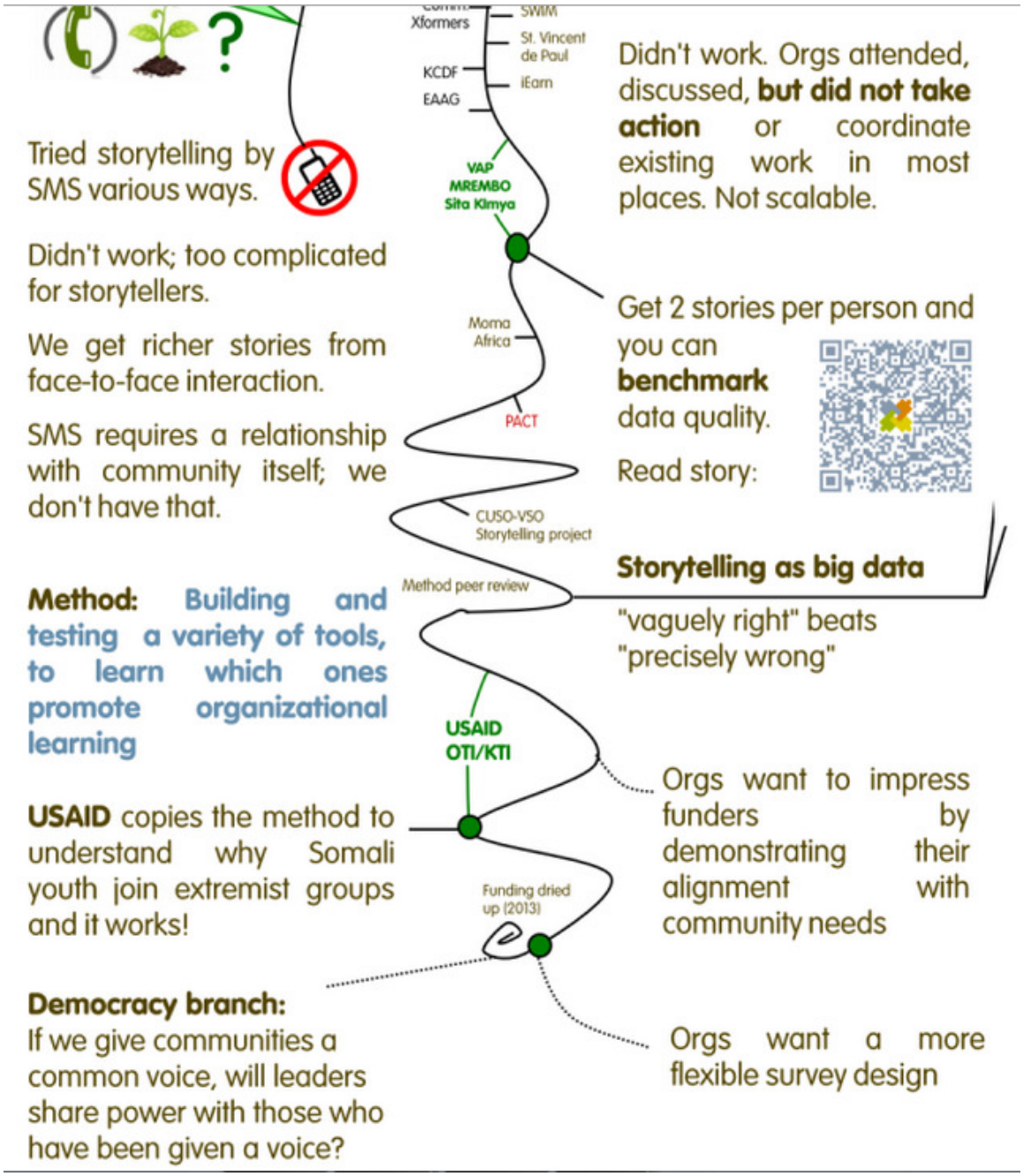
- Keeping: **Open ended feedback** about "community efforts"
- Changing: **Tools and methods for engaging** the organizations.
- Adding: **Marketplace incentives** for organizations that participate (once we're confident from our early-adopter focus groups that the analysis tools and collection method are good).
- Adding: **Mechanism to tie use of stories to stronger external funding proposals** for orgs. The idea is that organizations who participate can include visuals of the data in their next funding proposal, increasing its likelihood of getting funded. Some funders won't care about community feedback either, but at least GG will be able to track which funders DO care and build stronger relationships with these.
- Keeping: **Paper collection**. Many variations of collection that involve direct input of content into the electronic device (computer, phone, tablet) did not work – so we're sticking with paper at one end of the process.
- Adding: **Scan paper stories** from anywhere with an android camera-phone. Instead of picking up papers and mailing them to Nairobi on EasyCoach, we'll let organizations scan the papers.
- Dropping: Sensemaker® software. It wasn't easy to use, and we couldn't allow organizations to use it in the first place for licensing reasons and because it isn't webware.

- Dropping: Local story coordinators. We're relying on our relationship with GG orgs to collect stories. They are relying on GG to make such stories available in real time and with real analysis tools.
- Dropping: Individual financial payments to scribes. I personally think dropping this one is a bad idea, but we lack the money and the infrastructure to make thousands of micro-payments all over the world. We are instead relying on an organization's ability to attract, train, and retain scribes in the future, and helping them get creative with incentives for those who perform well.
- Adding: **Flexible Story Forms** – a frequent request from organizations. Now they will be able to customize which questions appear, and align these questions with their own M&E interests. Benchmarking will still allow GG to give them 'performance feedback' against peers.
- Adding: **Evolving story forms** – The questions that follow the narratives are only somewhat useful. Over years we've changed the focus of the questions from detailing the organization named in a story to mapping the aspects of the story that relate to the community and context (power, conflict, root causes of problem, possible solutions). Our focus ought to be allowed to evolve as users choose which types of questions align with their use of the data – so we are launching a "**flexible storytelling form**" with a pool of 30 optional questions and only 5 required questions. The organization designs the form by selecting which subset of the 30 optional questions appears on the specific paper forms they are helping to collect in their neighborhood.
- Changing: **Within-subjects benchmark:** We also continue to force the "**two story rule**" per storyteller, which ensures that at least 50% of the collected stories are useful. The other 50% are really only useful to one organization, as every organization has played a role in collecting each story, and one out of two could be about themselves (but not both). When organizations collect stories about themselves, these are relatively positive/self biased. This is a fact of life and instead of fighting bias, we filter it. We accept the full set in order to retain the 50% that isn't very biased, and we plan to even use the self-biased part of the org's data to show them the bias so that they can try to collect less biased stories in the future. There are a very small number of cross-narratives (where org A collects stories that mention org B and G works with both A and B). This is not like YELP where everybody in the city has an opinion about the same 20 famous restaurants. Most of the best "restaurants" in this metaphor are street merchants with no advertising budget. But if every customer reviews one street merchant has to also review a second one, we can at least compare merchants' '2nd stories' to each other for a less biased benchmark. And the point here is not to give merchants scores but rather to learn what kinds of food people are craving and what types of food are amazing.
- Changing: Instead of merely training scribes, we did a 2012 pilot on a "**train and retain" model** where many more scribes are recruited, trained, and tested with a small story gathering task of 10 people, but only a few that perform well are retained on a permanent basis. The idea here is to increase the overall story quality without making the training any longer or more complex.

Deeper Dive Report: segments to the funder (Rockefeller) for this project in 2012:

In a pretty picture, this is the whole 3 year process:





2010-2012

Improving the survey design: Through trial and error from 2010-2012, the best questions map out the most complex elements of stories – gray areas – and are themselves difficult to phrase. Many good questions are not precise because language does not have the right words to describe them.

Examples:

Conflict – causes, ideas for solving, effects

Power hierarchy – perception vs reality. Insiders and outsiders view things differently, so questions to map who has power are useful in describing a situation.
Intractible problems – why is this so difficult to solve? Questions that dig into the meaning of complexity in the context of the story. We assume that describing complex social problems will yield surprising and insightful information.
Nature of evil in society – mapping storyteller beliefs as a proxy for community perceptions about justice, equality, norms, aspirations.

Jan-March 2011 and Jan-March 2012

This goal (getting 100 organizations to use stories as data) was a challenge, and we did not make much headway. Most of this report is about what we learned about the problem of getting organizations to adopt the use of a free community feedback tool. It amounts to struggling with two problems: **incentives and analysis tools.**

We believed that the incentive orgs would have to use this data was to build their reputation, as a means to discover better program designs and better target interventions in a community, and to ultimately increase their funding potential.

The tools we designed and tested aimed at providing organizations with an overall perspective on any set of narratives, and summarizing parts of the associated survey questions that best provide context to the why, when, how, who, and what questions.

(1) Search pages with drill down options into subsets of stories

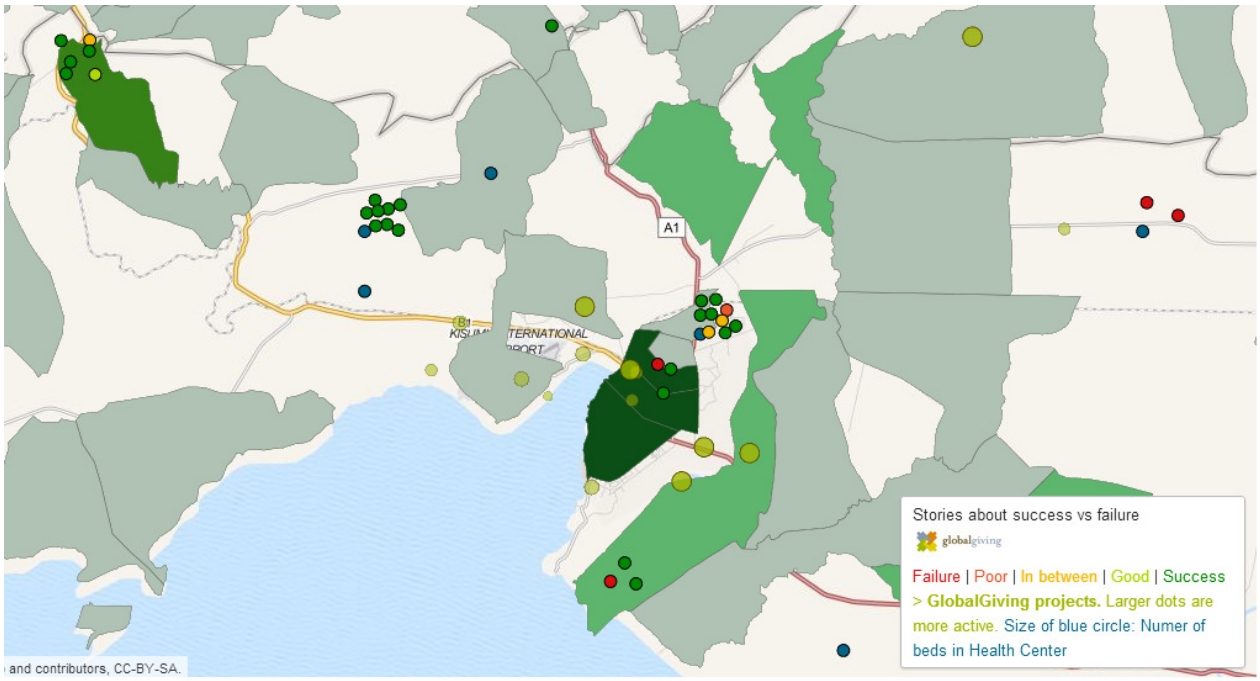
The "search engine" concept remains popular with the partner organizations that we would like to see using these stories. However, a search involving a few words or phrases is not going to be sophisticated enough to build a full analysis from.

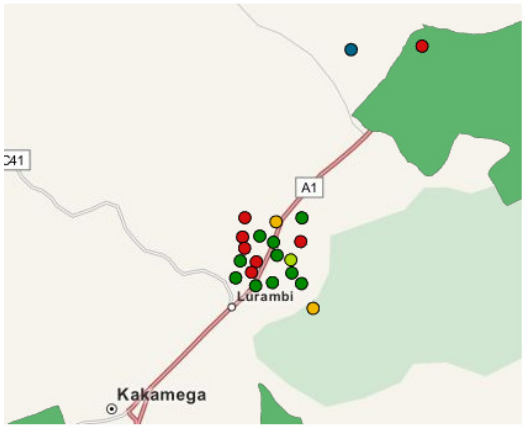
We also tried matching stories a complete set of org data from our globalgiving records and sending them a personalized digest of relevant, localized community feedback stories. This would have worked better if our matching algorithm was better – we're working on it. But an email-based user experience is not ideal. We would like to get to an online interactive version, because the rest of GlobalGiving is an online-interactive-social experience.

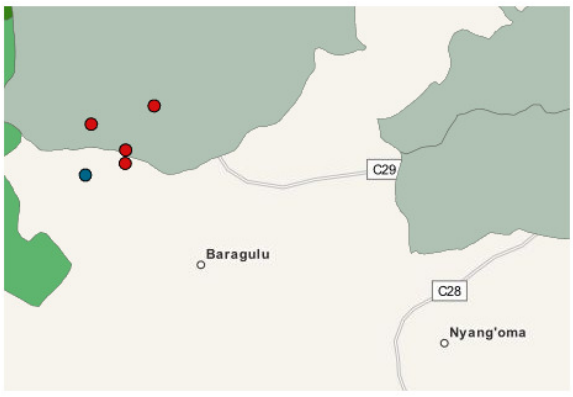
(2) Geomaps

We next tried geomapping tools, arriving at MapBox as the best option. MapBox allows us to display thousands of stories on a map and overlay other information, whereas other toolkits (ESRI online, example: http://www.crimemapping.com/map/la/neworleans) are limited to 800 points on a vector map. MapBox requires us to use Tilemill, which takes hours to generate each map – so we cannot make it fully dynamic. We can however provide frequently updated maps if needed. The problem was not knowing what overlays would tell project leaders the most useful information.

One good Example: HIV stories vs HIV treatment & testing centers (VCT) – Stories about HIV are color coded (red for failure and green for positive). Overlayed are blue dots provided to us by USAID. These are VCT centers. You can see that some VCTs are surrounded by red dots and others are surrounded by green. This illustrated how geo-tagged-stories can be used to make specific funding decisions and target interventions. However, we could find few other use cases where story locations yielded actionable information. There are better tools for service delivery and quality type questions, as our method is not a pur random sample. We have no stories from parts of Nyanza, Kenya (for example) where HIV rates are the highest, but where GlobalGiving has few partner organizations.







More at examples available at:
http://tiles.mapbox.com/stories

This process, besides being slow, was also a little expensive. We could only reasonably host about 5 detailed maps at $50 a month, and without user adoption we scaled it down

at the end of 2012. The larger the map – the larger the cost, and the longer the processing time to generate.

Lesson: geomapping stories is only going to answer the WHERE part of a larger story about an issue, leaving the who, **what**, why, how, and when parts unanswered. Most organizations what to know why and why, which are harder.

(3) Personal consultations with orgs – yeah we did a lot of that. It didn't yield huge returns on time invested. The following map of our 3-year process shows how many organizations we interacted with in more than one meeting in order to arrive at about a dozen good candidate user organizations.

April-May 2012

(4) In-depth training to copy storytelling methods

So the personal consultations continued throughout the year. And in fact, we worked with some larger organizations in workshops that went more than one day. Independent Development Fund (IDF), ActionAid, and PACT worked with us in trying to adopt this scribe storytelling method to their specific program evaluation needs. Unlike the smaller globalgiving partner orgs, these larger grantmaking orgs wanted more flexibility in the storytelling survey, and were willing to try our approach if the benchmarking tools were available. Each was prepared to try SenseMaker® (the paid analysis software we were using), but ultimately none did their own storytelling projects. In parallel, USAID-OTI (Kenya Transition Initiative) and a research group at Limpopo University in South Africa did conduct their own storytelling project based on our scribe method, with satisfactory results. I presented the GG and USAID results at the GLOBAL COUNTERTERRORISM FORUM (GCTF) in Abu Dhabi in Feb 2013. That community found value in the data but were terrified of the process that was necessary to achieve those results. One high-level counter-terrorism government official said, "failure is simply something we cannot accept at any level in our department."

So in-depth training has gone hand in hand with a reframing of the problem of low-risk tolerance. Organizations fear investing in a process that one academic could trash and "ruin their reputation." Others are afraid the results will reveal something unpredictable, or reveal nothing interesting at all. And in practice, both of these are likely to happen.

So why did ActionAid, IDF, PACT not implement it, whereas Limpopo and USAID-OTI did? (And why would a counter-terrorism office never implement it?)

My guesses:

- These other organizations **rewrote the storytelling form** to suit their needs. They gained an element of unbaised feedback from our method, even without the ability to benchmark data. The other three groups did not recognize that their existing M&E frameworks are based mostly on self-reported data with an inherent positive self-bias.

Global counter terrorism efforts around engaging extremist groups are generally not evidence based or empirical, with the exception of Malaysia.

- There was **no internal pressure to adopt a new approach** if the organization doesn't acknowledge any bias problems with their existing evaluation framework. Incidentally, one of the reasons GlobalGiving was willing to experiment with what became the scribe method is that we acknowledged our project-reporting system and not-so-random project visits are not sufficient unbiased to serve us as a quality control system.
- USAID-OTI (the unofficial department for working in failed states) has much more latitude to try new approaches, and bolstered by GlobalGiving's early success, they could justify the action. The team at Limpopo is an "action research team" and the approach aligns with their inductive approach to mapping out complex social problems – as I explain on my blog: http://chewychunks.wordpress.com/2012/02/29/from-survey-questions-to-prescriptive-answers/

Feb-June 2012

(5) Community Feedback Meetings

Lessons from the project as it has evolved (July 2012)

1. Method of collection – We know how to collect tens of thousands of stories from a region. We can do it at low cost, but it remains an *active* process. Local organizations recruit scribes and our coordinators train them, but scribes must have a financial incentive to go out, listen, and write (however small that stipend is). The next challenge will be exploring passive collection methods, and exploring no-cost incentives to foster collection by organizations.
2. Data cleanup – Some of the most important data cannot be collected directly by asking citizens. Two story elements are essential for analysis – latitude/longitude of where the story took place and a unique identifier for the organization that is prominent in the story – but these require significant post-collection analysis. We have improved our computer-matching methods to now assign GPS and NGO information to about 95% and 70% of locations and organizations named in stories, respectively. These data will never be 100% auto-detected (especially NGOs, with hundreds of thousands of them and no global registry), but we have reduced the workload to a manageable, semi-automated process that can scale.
3. Structured data – We began this project with the premise that we should be able to derive a *universal set of questions* that can be used to provide *specific, actionable information* to our hundreds of regional partner organizations. Over time that universal set of questions has reduced to a smaller set of elements that map out basic characteristics of the *story* (who, what, when, where). This greatly **simplifies training** and **broadens the reach** of the survey (few people will devote one hour to tell their story, but most will give 15 minutes). However, the reduced form puts a greater emphasis on generating patterns and meaning from the narratives themselves. If narratives are rich in semantic depth and meaning,

we can then create structured representations of story elements, which can then be used to *quantitatively* detect robust patterns (a robust pattern persists over time and across many locations). This approach to structuring data from narratives depends on actively managing quality story collection and constantly improving our computer-aided analysis and visualization methods.

4. Why and How: the unstructured black hole – We have improved our mapping methods pertaining to the who, what, when (TIME), and where (GPS) of stories, but the Why and How elements resist structuring. When we began, structuring the Who in stories (and the Who of the storytellers) was difficult but we've made progress. In the future we should devote more of our questionnaires to allowing storytellers to explain why they chose this story and how they expect the community effort in the story to affect them, evolve, or respond to the feedback they are providing.
5. Specific, actionable information – remains limited, both for GlobalGiving and its partners. Our own actions are limited to attracting great organizations or exposing fraudulent ones. We have some ability to detect and filter self-report bias and an artificial consensus view using heuristic methods (studying the patterns of stories or parts of stories). But the strength of our analysis tools is around conceptual evaluation – identifying the social drivers or root causes of complex social problems that may improve the interventions our partner organizations implement. Examples include rape[1], street children[2], HIV/AIDS[3], water[4], the Horn of Africa drought and famine[5] and other issues. Getting our partners to use (and value) this type of information will depend on how we package it, or how we align its use with existing incentives (like their ability to attract funding).
6. Hyperlocal and macro level data are interlinked – Once stories have GPS coordinates, we can aggregate trends across many geographic scale and observe how one data set differs. The micro-trends may be more actionable for local organization (for example, noticing that negative stories about HIV cluster around a particular HIV clinic), but the macro trends may inform strategic policy.
7. What makes a good story? – We are currently analyzing a broad swath of characteristics to define what kind of narrative seems to contain the most useful information for organizations. This may help us improve our training and focus our scribes' instructions.
8. Questions about story elements provide a superior measure whether the community effort in the story succeeded or failed. By using these three questions (Was the community effort a good idea? Who benefited? How do you feel about your story?) to measure success-failure, we can generate a much more accurate scale than simply asking the direct question. In our 2010 pilot, nearly all stories were about "success". But in our 2011-2012 data, we find that for every 10 stories we get that are about success, we get 1 negative story and 20 stories with "mixed" outcomes that fall somewhere in between the extremes. This distribution gives us

[1] http://chewychunks.wordpress.com/2011/07/25/comparing-two-rape-prevention-programs/
[2] http://chewychunks.wordpress.com/2012/06/10/analyzing-stories-about-street-children/
[3] http://chewychunks.wordpress.com/2012/02/29/from-survey-questions-to-prescriptive-answers/
[4] http://chewychunks.wordpress.com/2011/08/12/stories-about-water-and-swim/
[5] http://chewychunks.wordpress.com/2011/09/02/storytelling-the-east-african-drought-famine/

a more nuanced view of the world, more power to slice the elements of success and failure, and validates our basic premise that *story elements* provide a superior lens into project outcomes than direct constituency feedback.

9. Broad benchmarks for longitudinal learning – our method gives us a pulse on trends across civil society as a whole in Kenya and Uganda, and can provide benchmarking data for future longitudinal research. Done this way, this data becomes extremely cost-effective, and external stories serve as the reference frame for any location- or topic-specific analysis.
10. Provides a mechanism to transform the focus of the funding environment without coordination – We believe that community voice can used to wisely direct funding to address the problems that matter. This problem has often been conceived as a problem of feedback loops and funder coordination. A viable alternative to test is to provide a community-report generating tool that can bolster grant proposals that align with community priorities. In the process of writing grants, organizations may even adjust their strategies to align with the community perspective. Funders that receive stronger, better organized grant proposals from local organizations that have incorporated community feedback may start to preferentially fund these initiatives. Over time community voice has an influence on what gets funded, without the messy step of coordination. It is a system that works by strengthening the proposals of curious and listening, learning organizations preferentially.
11. True emergent conclusions – Since most of the questions are too general to constrain the possibilities of what people talk about (within the scope of what civil society organizations do), much of what we do learn can be surprising.

Our design principles

1. A simpler survey with broad relevance to all people and all organizations is better.
2. Keeping it simple for the storyteller is okay, since we can usually fill in the gaps in data after the fact.
3. The storyteller's perspective is subjective, and therefore, impossible to derive by any other means than directly asking him or her. Our collection process emphasizes accurately capturing the storyteller's point of view.
4. Language the most powerful and flexible data encoding system. It can encode nearly ever possible concept, and facilitate derivative representations and recodings of the original meaning.
5. Language and narrative is also an error-prone encoding system. However, we compensate for this with huge redundancy. Patterns have meaning when hundreds or thousands of stories show similarities, and these patterns persist over time and across many places (robustness). We rely on large samples of narratives for the emergence of meaning.
6. Our method facilitates pattern detection at both the macro level but the same data can also be contextualized as the hyperlocal level. In principle, we can facilitate actions within a single community[6].

[6] Example – interactive map of HIV stories and HIV testing centers
http://tiles.mapbox.com/stories/map/map-yjb5jlwq

7. Map whole community perspectives instead of only beneficiary feedback.
8. Built-in benchmarks – any analysis tools should emphasize the comparative power of automatic benchmarks. If an organization is interested in the impact of school uniforms, they would automatically get to compare uniforms against a dozen of related interventions mentioned in overlapping stories, such as free lunch, school fees, or orphan scholarships.

More on the project evolution from the simple to the complex, from http://chewychunks.wordpress.com/2013/04/10/pythonic-international-development/. This project is currently stuck in between stages 3 and 4 below. If this was like writing a program…

1. **Amorphic scripting:** At the beginning – you'll write your first program as a **linear script** with a "just give me the answer!" mentality. You won't care about structure, legacy, or efficiency, just what works. Surprisingly, quite a number of non-profits start out with an amorphic scripting mentality to their work. A person sees a crisis and take charge to alleviate suffering. And they can see that they are making progress, up to a point. At some point, usually 2 years into running a new organization (or 6 months for a new programmer) this approach starts to break down. Work is taking more work to make fewer gains. Staff feel like they are tackling the same problem repeatedly.

   This leads to the next level in program design: Modules

2. **Functional programming:** After a few tries at programming with a sequence of instructions, most people realize the power of organizing their bits of code into functions. Modules are collections of functions with a related purpose.

   Nonprofits come to a similar realization and tend to hire full time staff members with specific tasks, such as grant writing, marketing, program managing, advocacy, and youth engagement. They tend to hire the monitoring and evaluation officers last. These appear to offer less bang for the buck until you realize that you are trying to do solve a complex problem. Without monitoring and evaluation tools and methods, you begin to lose track of what worked and what didn't. Your versions of social programs get muddied in your memory.The same happens in programming. The next stage is using a code repository like GitHub for version control and collaboration.

3. **Collaboration and knowledge management:** Repositories (or repos for short) allow programmers share code, or fork the same project into multiple subprojects while tracking changes to the original idea. **GitHub** has ~4.6 million repositories and 2.4 million users; that's more than 1 repo for every community group in the world.The nonprofit world has tried to create "communities of practice" and "best practices" and write training manuals, do capacity building, and establish guidelines – but none of these come close to the simplicity of version tracking and

knowledge dissemination that GitHub has provided the open-source computer programming community.

But even the ability to share code pales in comparison to the power of real-time error checking, which is the next conceptual leap that programmers make.

4. **Continuous redesign and retesting -** eventually a programmer will have created something useful enough that it is time to publish a product. Some tool, software, website, or analysis will require a bit more maintenance. Strangers are using your tools and messing up your code, finding all sorts of bugs as they do things you never anticipated. Your new feature development grinds to a half as you spend all your time maintaining the code wasteland you once so lovingly believed was crisp and elegant.This is where I am. Luckily my friend turned me on to **py.test and TRAVIS.CI**.

   These tools allow you to write a series of debugging scripts that will automatically try to break your code each time you push updates to your GitHub repository. TRAVIS.CI emails me whenever it finds a bug in my code. Py.test allows one to write the "business requirements" of a program before even writing the program itself. It is a different way to manage the problem, but ultimately a better one.Nothing in international development seems to mirror this feature. If it did, it might be some combination of real-time feedback and peer group benchmarking on progress towards stated goals.

5. **Agile, iterative thinking -** and along the way many programmers realize that solving problems is going to be an iterative process, so they adopt agile.

   In international development, I'd summarize this as going from an idea to a working prototype in 14 days. Any idea too big to test in 2 weeks is too big to succeed, period. Either your resources are too scarce to handle the task, or the task is too big and complex to be feasibly managed. Most big ideas can be broken down into 14-day prototype tests. Agile says this is the best approach.

   Test your idea in a small chunk, then refine it and reexamine your assumptions. Let that TRAVIS.CI real-time bug checker crunch on your prototype while you do additional one-time tests, and you'll find yourself becoming very good at "programming" whether you mean coding or running social programs. Both require some flexibility to adapt.

6. **Upgrading the language to unlock new capacities**: A slim few programmers realize that they have been constrained by the language they were using to write code, so they begin working on a better way to write code. Python was one of those examples, but it will not be the last one. In international development, the way people think about problems, share information, allocate funding, and give credit will change in the future, and most likely by actors that have mastered all 5 previous steps first.