OPEN DATA NOW!

Open Data Unconference
30 June - 1 July 2014
Cape Town

CODE *for* SOUTH AFRICA

Kindly Sponsored By

THE *indigo* TRUST

"We have a good foundation for transparency in South Africa"

Adi Eyal
Code for South Africa

Forward

Transparency has become part of the zeitgeist of recent times. With it come a new lexicon of buzzwords: "open source", "open data", open this, and open that. It is all meant to lead to an "active citizenry", "data journalism", and of course, "open government". We have open data initiatives such as the American data.gov and its British twin data.gov.uk, as well as many initiatives around the world, including countries on the African continent such as Kenya and Ghana. In the light of these developments, it is surprising that South Africa has not made any moves in that direction.

We have a good foundation for transparency in South Africa. Section 32 of our constitution states that everyone has 'the right of access to any information held by the state ...'. The Promotion of Access to Information Act of 2000 (PAIA) gives meat to this imperative. However, the waters have become muddied in the actual implementation of the Act, which has often created an adversarial relationship between government and citizens. PAIA requests are often shuttled to government legal departments to evaluate whether the department in question is obliged to make that information public. The so called "secrecy bill" also adds to a culture of suspicion and reluctance to make data available.

Viewed from 10,000 feet up, one might be forgiven for thinking that we have a government that errs on the side of secrecy. It is however the details that matter. We have pockets of excellence within and outside of government. The Independent Electoral Commission releases one of the most detailed datasets on elections results around. It took the bold step in the 2014 elections to provide news organisations with real-time access to election data. Similarly, Statistics South Africa releases hundreds of datasets annually with incredibly detailed results from the national census. Finally, our national treasury was rated first out of 100 countries by the open budget index for the transparency of its budgeting process.

The concept of open data embraces these initiatives and tries to formalise the process through which government willingly and proactively makes data available.

There are already some moves towards a more open government, with South Africa becoming a signatory to the Open Government Partnership, committing to combat corruption and promote an active citizenry (amongst others). The City of Cape Town and the government of the Western Cape are working on their open data policies. The City has also committed to erecting an Open Data portal in 2014.

According to the Open Definition, "A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute

and/or share-alike."

Apart from a liberal licence, some attributes of an open dataset include:

- It is available in bulk, in other words the entire dataset is available rather than individual portions;
- It is machine readable — this means that it is digital (as opposed to paper-based) and that it available in a format that allows for computers to process automatically (think Excel spreadsheets rather PDF documents);
- It is available in open formats (think csv files rather than Excel spreadsheets).

In July 2014, Code for South Africa, supported by Creative Commons South Africa and Silicon Cape, hosted an unconference entitled "Open Data Now". Ninety participants attended, representing a wide range of stakeholder groups including government and quasi-governmental organisations, civil society, academia, programmers, private sector companies, funders and the media. The purpose of the conference was to map out which actors are currently involved in work relating to data availability, and to facilitate the creation and strengthening of relationships between participating organisations.

The unconference format required participants to identify their own open data topics that they felt were important to them. As organisers, we were nervous about whether participants would engage with the process. Their response was overwhelming. Thirty-two one-hour sessions were held, and discussions were robust, energetic, occasionally heated and always thought provoking. During each session, one participant was asked to record the discussion which would then form the body of these conference proceedings. The following pages present the notes collected from participants. The brevity of the recorded notes testifies to the richness and rapidity of the discussions, barely leaving rapporteurs sufficient time to accurately record the discourse.

Despite this, we have attempted to produce a document that gives you, the reader, a taste of the events that took place. A few common themes emerged in many of discussions:

- There is a clear need for an open data portal to collate available datasets at a single location;
- A policy of "open by default" for data is needed in contrast to the current "closed by default" mindset;
- Data needs to be released responsibly to ensure that it does not compromise the privacy of individuals;
- Data providers (such as government) should only be responsible for making data available, and info-mediaries such as the media, civil society and the private sector will package and present data to end-users.

The conference concluded with individuals making commitments to push the boundaries with open data in their specific contexts (listed at the end of this document). More than 50 commitments were made. Given the excitement and energy seen over the two days of the conference, we expect that at Open Data Now 2015, the open data environment in South Africa will be markedly different, and we hope, better.

Adi Eyal
Code for South Africa Founder
adi@code4sa.org

Code for South Africa, A civic technology organisation that uses data to promote informed decision-making.

http://www.code4sa.org

Conference Notes Day 1

How Do We Get Data From Government?

Convenor: Adi Eyal
Participants: Loren Treisman - Indigo
Nomsa Masuku - IEC
Roger Gachago - SAFLII
Rashaad Alli - PMG
Brett Horner - Daily Dispatch
Laban Bagui - UCT
Liesel Pretorius - City Press
Jessica Taylor - International Budget Partnership
Trevor Oosterwyk - StatsSA
Koketso Moeng - StatsSA
Tarryn-Anne Anderson - Paperight
Janet Jobson - The DG Murray Trust
Venue: A
Time Slot: 10:30

Issues

Data is a planning tool. If it is not shared by government then that affects others' ability to plan effectively. For this reason, government should not hold onto data.

Getting data from government departments is difficult. You need to find the appropriate content people to request data from. For example, request procurement data from the head of procurement. The problem is that It is not always clear who one should speak to. Also, even when a request for data is granted, one may only receive data in drips and drabs or in inconvenient formats such as scanned PDF files.

Although government departments all have a mandate to make information available, when requesting data, one is often redirected to the legal department in order to ascertain whether government is obliged to release the data. The relationship with media and civil society is considered to be adversarial. The mindset within departments need to change.

A problem that government experiences is lack of capacity to handle the number of requests for data.

Also, some data cannot be released such as personal/private data although this data can be de-identified.

Some organisations are leading the way. The IEC api developed for the most recent elections was a huge step forward.

Ideas and Recommendations

- Publishing data to an open data portal would improve data accessibility.
- A Database of contacts within departments would be very useful. This would shorten the time to finding the right person who can help with a data request.
- Mindsets need to change. Government should start to understand the value of sharing data.
- We need education/advocacy around open data. For example, SANEF has been putting pressure on government for an open data portal which would include officially sanctioned data. This would include data from treasury and the national planning commission.
- Private sector and civil society need to prove the value of open data by showing practical examples of how it could be useful.
- Identify ways where data could help solve some of governments problems.
- Build non-adversarial relationships within departments and persever.

How Do We Get Data From Government? (A2)

Convenor:	Cor Schutte
Participants:	Lynn Woolfrey - Data First, UCT
	Cor Schutte - City of Cape Town
	Greg Kempe - Code for South Africa
	Kevin Tabisher - City of Cape Town
	Neil Hoorn - City of Cape Town
	Jared Rossouw - Ndifuna Ukwazi
	Luxolo Lengs - StatsSA
Venue:	A
Time Slot:	10:30

Data as a right
Data can be used to promote active citizenship and can lead to the empowerment of citizens. It is important to decide whether data is a "nice to have" vs "a right". Government data should belong to taxpayers.

Internal Capacity
Often, government does not have adequate capacity (skills, budget, experience, etc) to release data. Many technical steps are needed in order to successfully release data. Small departments are often expected to perform the same duties with their data as StatSA with little or no budget. One solution would be to use data brokers. For instance, UCT has signed an MoU with the Department of Basic Education to release data on their behalf.

City of Cape Town as an example
The City of Cape Town has approximately 3,000 datasets. City has data and plan, where do we start? What's the process? How does one start making this available? Which datasets should be released first? One approach would be to provide an overview of what data is available and solicit feedback. Should government be trying to add value to data (through apps, analysis, etc)? It is not clear that this is needed nor does government have the necessary understanding of data users in order to do this. The City wants to release data that is NEEDED, USABLE and UNDERSTANDABLE

Data quality
Data is often released in aggregated format. This may hide inaccuracies.
Service delivery protests may be based on wrong information. The raw data is needed in order to verify.
Budget data should be presented in both summarised and detailed formats.
When making data available, feedback from users eventually results in improved data quality.

Misc
New data sets often drive demand

Challenges
Releasing data comes with responsibilities (e.g protecting personal data)
Municipal data is political, mistakes in raw data can lead to embarrassment. If government is honest and brave, this will eventually lead to better data.
How does one handle pushback from political people?
It is important to ensure that an open data policy is correct because it will be hard to change once it has been adopted.
Education is important, bureaucrats may not understand requests, data policies or the underlying data that is being requested.

Making Open Data More Inclusive and Less Scary

Convenor:	Kelsey Wiens
Participants:	Paul Kennedy - Private
	Philip Kiracofe - Horizen Ventures
	Phillipa Yeats - Praekelt Group
	Belinda Lewis - Praekelt Group
	Hannah Williams - Code for South Africa
	Sarah Goodier - OpenUCT
	Megan Beckett - Siyavula
	Giovanni Gerbi - EWN
	Melanie du Plessis - IEC
Venue:	C
Time Slot:	10:30

There are two main issues regarding the public's relationship with data. The first is that data is considered to be technical and "geeky". Data illiteracy is a major reason why most people don't understand how to use data (and how to interpret statistics used in the media). There also exists a stigma around data, specifically personal data. People don't realise how much data they are releasing and feel as if they their privacy is being violated. This often causes a strained relationship between the public and data.

From IEC
- Users don't know what they can do with data. and can't do
- They are trying to improve accessibility by making data available through Excel, APIs
- Forward-thinking organisation
- Believe that data MUST be publically available
- Distinction between technical and analytical aspects needed

Usability
- Curation is important
- Analysis can be privatised BUT some feel that incorrect interpretation is a problem
- There is a difference between raw and processed data

Distribution
- Data providers are responsible for collecting, cleaning, maintaining and publishing data.
- They are not (solely) responsible for making it accessible to the public.
- There is a need for infomediaries that manipulate and interpret and visualise data for mass consumption.

Recommendations
- Government data custodians need to be educated
- Can tie funding to openness of data
- Language problems – the terminology around data should be softened so as to seem less technical and be more inclusive.
- Visual representation of data is important although a lack of resources is a common limitation for showing data.
- "Best Practice" guidelines needed
- Standardisation is needed to improve accessibility
- Case studies can help change attitudes
- Mapping of open data resources needed

Private Data
- Create a reward system for sharing with tangible outcomes
- Before private data is collected, users need to "opt-in"

Data Use
- Data not just for "watchdog" purposes, also for good
- Data should also be used for positive purposes so as to improve its image.
- Knowledge sharing is important

Where do we get a data coder?

Convenor: Gaile Fullard - PMG
Participants: Douglas Day - dotLogic
Rachel Ward - SAHRC
Alison Tilley - ODAC
Benjamin Ace - Heinrich Boell Stiftung
Gaile Fullard - PMG
Venue: G
Time Slot: 10:30

The general discussion focused on the difficulties of finding good resources to help deliver on projects, rather than anything specific to OpenData. Although the implication was how can we undertake OpenData initiatives if we can't find the right people to do them.

The general discussion focused on some of the major pitfalls of projects to under deliver.

1. Where can I find a freelancer? You can access good freelancers through a variety online sites:
   - Monster.com
   - Linkedin – South Africa Search
2. Warning: using different coders over time, you will end up with spaghetti code.
3. In your organisation you need one person / company that needs to be the architect of the software and own the solutions vision.
4. Engage with an individual or company that will give you support after initial development is complete, someone that will– "own the code"
5. Avoid larger companies and go for smaller niche companies, they value your business more.
6. Ongoing support should be built into the contract
7. Make use of Agile project management (user group SUGSA: Scrum User Group). "Evaluate output at smaller intervals" = collaboration built in.
8. Inspect deliverable each two weeks and feedback to the coders. Iterative development.
9. Money bonus for finishing early "money for nothing" bonus depletes over time so incentive to coder to complete task.
10. Bill Gates advises one to invest the minimum amount of effort and use the technologies available to you know. Once you have a product then you can develop it into future.

Tech Companies and Open Data

Convenor:	Zimkhita Buwa
Participants:	Wellington Radu - Media Monitoring Africa
	Fortune Mgwili-Sibanda - Google
	Charles Lebon Mberi Kimpolo - Thoughtworks
	Ditjhaba Selemela - Thoughtworks
	Zimkhita Buwa - Britehouse
Venue:	H
Time Slot:	10.30 a.m.

General Points
- Thoughtworks has an open data initiative
- Google map maker - useful tool to fix/update your local MP used in a project at schools at a township to fix the data called Map Ups
- Tech companies must always update the open data they use

Issues
- Tech companies in SA need to think more about the poorer consumer
- Open data is an elite issue and not used by people on the ground - challenge to make this more widely applicable
- There needs to be a sense of worth for the users of the data
- Data is not always available. e.g. wards data is but roads data isn't
- Need to identify people in the community who can use open data technology.
- Open data is less important to someone who doesn't have food. The challenge is connecting the two

Recommendations
- Need to connect teachers and young people in the community
- Create a tech company working group
- Should be a portal for all these projects?
- Fund someone at Code4SA to collect a list of these initiatives
- Break portal info into both the data and the tool sets. Portal needs to include:
- data that is open and available
- tools that can use this data
- Option to ask large tech companies for free technology for open data projects e.g. Splunk Open Data Initiative

Political Party Funding and Mapping Financial Interests

Participants: Stefan Gilbert - Institute for Security Studies
                      Dillon Mann - World Wide Web Foundation
                      Hopolang Selebalo - Equal Education
                      Murray Hunter - Right2Know
                      Osman Siddiqi - Code for South Africa
                      Friedrich Lindenberg - ICFJ
                      Karabo Rajuili - My Vote Counts
                      Lonwabo Jwili - IEC
                      Jason Norwood-Young   - Code for South Africa
                      Khadija Sharife   - Investigative Journalist/Author
Venue: J
Time Slot: 10:30

Background/Overview on the topic
There are two main sources of funding for political parties:
1. Public Funds:
   • IEC - granted proportionally to all represented political parties
   • Parliament Constituency allowances
   • Provincial Legislature (serious questions on the constitutionality of these allocations)
2. Unregulated private funding (secret donations).
The threat posed by private funding - corruption (party scandals), lack of accountability and transparency, and political inequality/ maintenance of the status quo.
Other issues:
   • The constitutional argument for reform through regulation: Voting (S19) is meaningless without access to financial info of parties (S32)
   • Political parties are public bodies with private facets. Political parties should legally be public bodies, currently seen as private bodies (see IDASA judgement) which means no legal obligation for financial disclosure.

How do we use open data to move things forward?
Issues to consider in answering this questions:
   • Is it/ how do we distinguish between the political problem, legal problem or tech problems?
   • What are the current Open Data obligations for political parties?
   • What are some of the unintended consequences of advocacy work in this area?
   • What are the available data sources (e.g. CIPRO, Afrileaks etc) which we could use?
   • What are the priority areas for data to be made available? -   2016 Local Elections coming up, how to make use of this political moment.

Mapping of Financial Interests of MP's and MPL's - existing project of ISS
   • Political figures should have to disclose their financial interests
   • The database works as a spider web showing relationships

How can we develop this existing tool to further our goals?
   • Can we get that info? What will we do with that info?
   • Build massive database
   • Top down to municipal level (mapping people)
   • Get information on tenders made in every province

Other Open Data options/suggestions
   • Getting data at the municipal level is particularly important strategically as we prepare for the 2016 Local Government elections.
   • We should use curated information from Jane Duncan's protest data.
   • A simple campaign could request income and expenditure data from political parties.

Attribution for Data

Convenor: James Aphane - IEC
Participants: Neil Hoorn - City of Cape Town
Venue: B
Time Slot: 11:30

Open data is not really free to maintain, there are costs for hosting, maintenance, etc. Therefore the creators of the original data should be acknowlegded where ever and whenever their data us used, either directly or indirectly.

Data should be acknowledged. The author should be credited.

Government data is created using public funds.

Data creators need feedback from users regarding inaccuracies.

Open Data for Social Change - What's Stopping You?

Convenor: Hannah Williams
Participants: Hannah Williams - Code for South Africa
                  Sarah Goodier - OpenUCT
                  Zimkhita Buwa - Britehouse
                  Luxolo Lengs - StatsSA
                  Alfred Mahlangu - AfricanLII
                  Charles Mberi Kimpolo - Thoughtworks
                  Ditjhaba Seemela - Thoughtworks
                  Kelsey Wiens -Creative Commons
                  Kenny Inggs - 22Seven
Venue: C
Time Slot: 11.30

Common barriers:
1. Money. Projects are often pro bono. Easier mechanism to find funding.
2. Access to information
3. Data Quality
   • There is no feedback process for data if it's wrong
   • Stats SA say they will listen to any data corrections
   • Difficult sometimes due to many parties / 3rd parties
   • No-one wants to be wrong
   • Should create encouraging, supportive environment
4. Not always easy to know what data is available
5. Who do I ask? What's the right entry point for data? - e.g Stats SA collects data for Dept of Education - must ask DoE. Do people know StatsSA will sit down and help you find the right data?
6. Inclusiveness of data - challenge in SA is that the people who need help the most are least documented and connected to technology
7. Data collection - how do you get people to participate?
   • feedback / examples
   • rural areas data literacy
   • presenting data on low tech phones
8. Perspective: Data Providers vs Data Consumers
   • **MONEY**
   • opening data takes time and money
   • collecting data takes time and money
   • analysing and presenting data takes time and money
9. Demand - if there is enough demand for data will government release it? How do we create data demand?
10. Who decides what data is released? We need an open data platform for people to share and access data.

Open Data Demand and Collaboration

Convenor: Dillon Mann - World Wide Web Foundation
Participants: Liesl Pretorius - City Press
Justin Arenstein - AMI
Janet Jobson - DG Murray Trust
Greg Kempe - Code for South Africa
Raymond Joseph - Listening Post
Loren Triesman - Indigo Trust
Greg Solik - My Vote Counts
Giovanna Gerbi - Eye Witness News
Osman Siddiqi - Code for South Africa
Venue: D
Time Slot: 11:30

Giovanna:	Challenge selling data journalism to editors
Raymond:	How good is current awareness around tools
	Geeks are building stuff they think reporters want – but not necessarily right
	Journalists often think it's more work
	Not being shown what tools can do but they can help
	Training is not happening
	Wazimap – good example
	Training and advocacy should extend beyond journalists – to activists
	We need a central portal to collect available tools.
Loren:	Consultations with users about the things are built
Greg:	We need to bring users and all other stakeholders together at build stage
	Tech guys need more contact with users. Needs to be demand driven
Loren:	How do we make this happen?
Raymond:	Funders need to specify this at donation stage. More transparency and insistence of deeper consultation at the development stage
Janet:	Know at the beginning how the technology will be used. The tech is an enabler it is not the hero
Dillon:	Central portal with tools and examples of how they can be used for reporters for activists
Raymond:	We need examples of where reporters are using these
	Attractions 101
	Wiki maps tools: no-one uses these yet
Giovanna:	Radio and print had different speeds
Liesl:	Everyone is stretched
	We are not going to consent entire newsroom who are already interested in data – start a slow movement
Raymond:	People don't understand the tool will not do the job. They still have to
Greg:	Mentions: civicpatterns.org
	1st step: convening a dedicated discussion on this
	Then builds up to a website
Janet:	Challenge for civil society
	Many gatekeepers – need to get directly to activists
	Poor level of transmission of new tactics into civil society =
	Must take tools – reality
	Tools need to be intuitive, timing minimal
Loren:	Maybe select a handful of effective organisations. Start with those.
Raymond:	Too much stuff being built, some lies on the shelf
	Media houses aren't funding
Dillon:	Data journalism does not have to result in a numbers story. Rather the story is in the data
	www.pa.org.za
	People's Assembly tool mentioned
	You can know your MP is

Open Data Demand and Collaboration (continued)

	You can get special interests rewards participation in communities
Janet:	Lack of use is a stressful issue. Fundamental relationship solution. South Africans and our parliament
Osman:	"Maybe some lads just don't like going to school"
	Undefined who is the user in the first place
Greg:	You have to reach the middleman before you reach the man on the street
	The tools mean that reporters produce much better stories and save time
Raymond:	Example: Before you do a story in Lavender Hill you need to conduct background research, e.g. demographics
Loren:	Audience at the moment is NOT the people on the street. Most not online. Need intermediaries.
Giovanna:	Challenge for breaking news. No time to research first. Example, shooting just happened need to go and cover it. Radio may not be able to use it.
Greg:	The issue is in problem definition. Radio specific example: Build an app for reporters on the move.
Raymond:	My takeaway: Take a problem to the tech community, let them propose solution don't say "I want a tool that does this…"
Osman:	Solution is removing sunk-cost research
	Shifting people past a technological barrier. Get them to use a new system. All organisations need to work towards effective implementation
Liesl:	What should reporters do differently when working with Code4SA
	Parked - Liesl to ask Adi
Raymond:	Investments will be made more easily scalable if something is reputable open-source eg maybe pa.org.za
Janet:	Suggestion tweak to Wazimap. Give village names rather than ward numbers.
Osman:	Do you crowd source feedback to Wazimap
Raymond:	When Indigo funds something, do you have a monthly requirement and feedback?
Loren:	Yes but we need to be realistic as out grants are small. Only going so far as we are only talking to activists we know
	Did press use tools after press breakfast?
Giovanni:	Yes to a degree but not
Justin:	Only thing we have first works is that we need to embed people in orgs. Scale down not happen immediately.
	Need champions within organisations
	Get samples from Argentina, Kenya etc but none form SA yet
	Tools and tech should be free. Real cost is in using them
	In Kenya tools that have worked but have been heavily marketed
	We shall use industry with media partners
	Try to avoid getting replies to say what we want to build – rather start with a problem
Raymond:	Need to show these tools reduce work for others
Justin:	Mistake to expect IT to do this
	Need to put in extra person- best resources at New York Times, Guardian etc
	Small, quick wins important to demonstrate potential
	Elections are a waste to prioritise investment on
	Elections are the start of a process
	Build with election day

PAIA Process

Convenor: Katherine Eyal - UCT

Participants: Rachel Ward - SAHRC
Shaun Russel - Ndifuna Ukwazi
Brett Horner - Daily Dispatch
Hopolang Selebalo - Thoughtworks
Wellington Radu - Media Monitoring Africa
Jessica Taylor - International Budget Partnership
Greg Solik - My Vote Counts

Venue: E

Time Slot: 11.30 a.m.

Current Activities
- Katherine introduces research project - how do we make PAIA work?
- Ndifuna Ukwazi doing a PU manual - Open data and proactive disclosure vs using a request process
- SAHRC provide a guide on how to use PAIA for investigations and enquiries
- South African History Archive has a PAIA programme
- Askafrica.info is an automated portal - should link PAIA/ civil society groups and help with questions,but doesn't necessarily solve compliance

Problems
- The law is difficult to grapple with, need an advice service
- IR may remove human rights element
- Records management part of reason for non-release of data

General Points
- Link to Information Regulator
- Nigeria has simpler freedom of information system - but requests can be ignored
- You need a formal mechanism for recourse
- Should be voluntary disclosure as well

Benefits of data being released
- Data can be interpreted by experts to inform good policy-making
- Solves the issue of poor inter-departmental information sharing
- Informed citizen decision-making

How Valuable Is Open Data, Really?

Convenor: Roger Gachago
Participants: Adi Eyal - Code for South Africa
Jeff Fletcher - Internet Solutions/Limn
Roger Gachago - SAFLII
Belinda Lewis - Praekelt Group
Phillipa Yeats - Praekelt Group
Douglas Day - dotLogic
Cor Schutte - City of Cape Town
Paul Kennedy - Private
Megan Beckett - Megan Beckett
Philip Kiracofe - Horizon Ventures
Karabo Rajuili - My Vote Counts
Fortune Sibanda - Google
Venue: F
Time Slot: 11:30

- Does data have monetary value? If so, how can it be monetised?
- Value does not necessarily need to be monetary, could have social impact as well
- There is not enough dialogue in ZA and how it was successfully used.
- How to price data? Need to demonstrate ROI
- Don't sell data but rather charge for services
  - Data cleaning
  - Timely access, APIs etc
  - Service level agreements around availability
  - Analytical services on top of data
  - Build products on top of data.
- How can we fund the publication of data?
  - Donor funding - not sustainable. e.g. Wikipedia seeks donations from users.
  - Value added services
  - Public funds - i.e. taxes
  - Commercial data provider - e.g. government can tender out a dataset to a 3rd party who will manage the dataset and charge for access. Similar to toll roads.
- Difficult to estimate the value of data. Value can not always be anticipated. Value is sometimes created after data is made available.

E-Participation and the use of Government Data

Convenor: Laban Bagui
Participants: Paul Murphy - Western Cape Province
Melanie du Plessis - IEC
James Aphane - IEC
Laban Bagui - UCT
Venue: H
Time Slot: 11:30

- Challenge and opportunity
- Policy + Infrastructure + culture
- Too much data
- Need to improve the data
- Need to engage with public
- Need to acknowledge problems
- Need feedback mechanisms to improve datasets
- Some aspects of data can't be made available

IEC - events dictate how to make data available to the public

What To Do When Government (Or Anyone) Says No?

Convenor: Jason Norwood-Young - Code for South Africa
Participants: Tarryn-Anne Anderson - Paperight
                   Keren Ben-Zeev - Heinrich Boll Stiftung
                   Lonwabo Jwili
                   Lynn Woolfrey - DataFirst
                   Gaile Fullard - PMG
                   Murray Hunter - Right2Know
                   Jared Rossouw - Ndifuna Ukwazi
                   Trevor Oosterwyk - StatsSA
Venue: J
Time Slot: 11:30

- PAIA as a tool
  - Applies to public bodies
  - Applies to private bodies in a limited way
- How you ask is important and as well as who you ask
- Best practice: bodies should make sure the IT person is not given the task of being the data person. IT staff provides the infrastructure for data curation but curation itself is a specialist task which often needs subject knowledge and awareness of the elements of disclosure control
  - The more data producers know the better they are at releasing because they do not have unfounded fears around confidentiality breaches
  - but not always. Sometimes PAIA can be used skillfully to ATI
  - product of fear- political embarrassment
- Example: Ndifuna Ukwazi asked the City for sanitation data
  - Failed to get official release
  - Instead crowd-sourced social audit
  - It's possible the data is not being held centrally
- Example: Code for South Africa requested tender contracts from the City. Was told that all contracts were stored as paper documents and would be too costly to scan in.
- How the data is held is important
- And how you ask is important
- Open releasing helps avoid black market release where those from well resourced organisations buy the data - these leads to a proliferation of unofficial versions of the data, with limited usability
- So what do you do when you say no?
  - legal route -> expensive
  - advocacy route -> political pressure
  - Sometimes avoiding PAIA is better
  - Asking is politicised -> develop allies within the organisation
  - Not asking at all through community mobilisation
  - Release alternative info / shadow report
  - Build public support / allies around the info required

Open Data for Social Good / Civic Engagement

Convenor:	Loren Treisman, Jessica Taylor
Participants:	Loren Treisman - Indigo Trust
	Dillon Mann - World Wide Web Foundation
	Neil Hoorn - City of Cape Town
	Kevin Tabisher - City of Cape Town
	Laban Bagui - UCT
	Jessica Taylor - International Budget Partnership
	Katherine Eyal - UCT
	Vuyisa Qabaka - CiTi
	Janet Jobson - DG Murray Trust
	Osman Siddiqi - Code for South Africa
	Liesl Pretorius - City Press
	Phillipa Yeats - Praekelt Group
	Justin Arenstein - AMI
	Alison Tilley - ODAC
	Keren Ben-Zeev - Heinrich Boll Stiftung
	Stefan Gilbert - Institute for Security Studies
	Nomsa Masuku - IEC
Venue:	B
Time Slot:	13:30

Mobile Phones for Delivery of Information
- Consider tools for simple phones e.g. USSD
- There are challenges therefore costs are high for services offline
- SA is still a feature phone / dumb phone market predominantly
- DG Murray Trust is lobbying (with partners) for government legislation to force mobile operators to zero-rate some services. (ISA-SA or other govt levies should be used more effectively)

Miscellaneous Points
- SAPS deals with most of requests (75%) for data from government, often from lawyers
- International Budget Partnership and Social Justice Coalition seeking info from City of Cape Town e.g. service delivery agreements, local budgets. Need info to be available and accessible. Data must be disaggregated enough to be meaningful at local level.
- World Bank and Omidyar Network have an open contract project
- Global standards are being developed with open standards
- Challenge = compliance

Issues
- Don't underestimate language and literacy barriers - voice services e.g. toll free call centres or IVR can be useful
- Access to data and building the tech is less hard than the sustained citizen engagement and ensuring civil society has the capacity to respond to data/reports coming in
- A business argument for giving free (mobile) data for social services (zero rate) should be made. Build social projects and then identify a revenue scheme when it scales, e.g. getting more users, providing locally available content and value added services
- It won't always be possible to generate income especially in transparency space
- Data isn't granular enough - need provincial and lower level data for effective advocacy
- People asking govt / province for data are struggling to get it. Some data is tied up in commercial contracts - should be a public good

Sustainability
- Are not-for-profit solutions sustainable? Perhaps we should charge a little? If data affects a person's life significantly they will pay - e.g. in Kenya people pay premium rate SMS for key info on their doctors
- People will pay for some info but not others - sometimes must be free

Open Data for Social Good / Civic Engagement (continued)

Recommendations

- Target mass based community groups and identify their challenges e.g. church/mosque groups, farmers' groups, unions then link them to civil society organisations to ensure their voices are heard and not just those of key civil society groups
- We need transparency and demand driven data release by default at once
- Build solutions for users
- Partner with mass media to ensure tools are used widely. Target Daily Sun, radio and other widely used papers
- Marketing should have a larger role in civil society projects - ideally use edutainment models for social messaging. People don't want to be lectured to
- The feedback loop between citizens and govt is key. We need to build this into government's way of working
- Mobile data costs are HUGELY prohibitive. Civil society needs to encourage free/ cheaper costs from operators. Instead a lot of money is being spent on hardware solutions. Can civil society set up advocacy around this? Engage corporates like Facebook and Google in the lobbying process - potential revenue streams
- Treasury and others often can't see the point of releasing data as we can't see how it will be used. We need case studies e.g. only 10% World Bank reports read.
- Need to show economic opportunities to opening up data - civil society is driving the concept of open and govt doesn't always listen
- Focusing on a specific need / problem can be helpful to identify what data is important e.g. SJC WASH campaign. People care about their hospital / education
- Often different people are working on different pieces of the puzzle - we need the dots to be linked up more effectively

Questions

- Can tech tools help people to access personal data from big data - this is what people care about. Disaggregation is a huge challenge
- How can govt community centres be utilised more effectively for info sharing and community activism / mobilisation?
- Mbombela govt exploring providing free wifi and giving 1 free GB of data per person costed at R 1.2 million / year. Can this be done elsewhere? Similar is being explored in Knysna
- Western Cape Province s trying to open hotspots with free internet e.g. at hospitals. How can we as civil society tap into this?
- Freemium models, subscriptions, premium rate SMS, corporate partnerships, ads(?) - can you prove to operator that your intervention will lead to more customers?
- What is more effective? Start with demand for data around specific needs or transparency first?

Service Delivery Data

Convenor: Shaun Russell
Participants: Shaun Russell - Ndifuna Ukwazi
                    Adi Eyal - Code for South Africa
                    Tarryn-Anne Anderson - Paperight
                    Petrus Janse van Rensburg - Code for South Africa
                    Lonwabo Jwili - IEC
                    Paul Murphy - Western Cape Government
                    Wellington Radu - Media Monitoring Africa
                    Hopolang Selebalo - Thoughtworks
                    Luxolo Lengs - StatsSA
                    James Aphane - IEC
Venue: B2
Time Slot: 13:30

Discussion Points
- Service Delivery Agreements
- Public services
- School data (Equal Education)
- Disconnect between tech and civil society
- Locate data point on a map (e.g. election data)
- Seeing data around the individuals
- Combining multiple datasets to find trends

Key Notes
Service Delivery Agreements
- Engaging with governments
- Get around negativity from governments
- There is a culture of secrecy.
- Look at legislation for advocating government to release data
- Internal government data issues
    - departments working in silos
    - where does this info actually sit?

Public services
- Engaging with governments
- Internal government data issues
    - departments working in silos
    - where does this info actually sit?
- School data (Equal Education)
- Who is my ward councillor?
- People don't know where their wards are

Disconnect between tech and civil society
- What is the info you want? And what are you going to with it?
- Analysis of service delivery protest datasets
- Combining multiple datasets to find trends

Locate data point on a map (e.g. election data)
- Where are the housing developments- human settlements
    - HDA- Housing Development Agency

Seeing data around the individuals
- Making public data actionable

Opening Up Academic Research

Convenor: Sarah Goodier and Paul Kennedy
Participants: Paul Kennedy - Private
                    Lynne Woolfrey - DataFirst
                    Sarah Goodier - OpenUCT
                    Douglas Day - dotLogic
                    Megan Beckett - Siyavula
                    Greg Solik - My Vote Counts
                    Philip Kiracofe - Horizon Ventures
Venue: C
Time Slot: 13:30

Open research data:

Why?
- Publishing data with publication allows for verification of research
- Can produce new research using previously published datasets

Miscellaneous Points
- Data can have multiple owners, e.g. universities, academics and funders
- Governments and Universities often need to buy publications (and data) back from publishers
- Governments and Universities both have data to share. It is not necessarily their responsibility to manage and clean it up.
- UCT Datafirst uses a share/share-alike licence (with required login for downloads)
- UCT Datafirst has an MoU with StatsSA to add value and share data
- Researchers are often not keen on sharing data as they might lose their competitive edge.

Questions
- Who owns research/data funded using public funds?
- Should researchers share data and what are the copyright/licensing implications of downstream uses?

Recommendations:
- Portal to share data/database
- Need a data sharing policy. UCT has a data management policy. Part of the policy should be to ensure that data be made open as well as specify how this should be done.
- Need stakeholder workshops - e.g. answer questions on the census and to provide suggestions.

Data Visualisation

Convenor: Jeff Fletcher
Participants: Jeff Fletcher - Internet Solutions / Limn
                   Roger Gachago - SAFLII
                   Charles Kimpolo - Thoughtworks
                   Koketso Moeng - StatsSA
                   Fortune Sibanda - Google
                   Melanie du Plessis - IEC
                   Zimkhita Buwa - Britehouse
                   Ditjhaba Selemela - Thoughtworks
                   Hannah WIlliams - Hannah Williams
Venue: D
Time Slot: 13:30

Projects and software
- Thoughtworks have a matric results map available, recently added sanitation data (za-schools-results.heroku.com). It was developed using Leaflet. Will soon open the code.
- There is published election map from IEC
- StatsSA have some Google code, high charts etc. They stopped using D3

Tools
- SAP provides data viz tools (e..g SAP Lumira)
- Tableau Public
- High Charts
- D3
- Leaflet
- Gephi for network diagrams
- Google Charts library (https://developers.google.com/chart/interactive/docs/gallery)

Miscellaneous Points
- IEC provided API to media for elections, although it impose a restrictive data usage policy.
    - https://api.elections.org.za/
- Department of Basic Education releases the Education Management Information System (EMIS)
- Public data can only be charged for if it is not in its raw form
- Two kinds of users
    - Raw data consumers
    - Consumers looking at visual data

Recommendation
- Create a dataviz meet-up group

How Do We Get People To Engage?

Convenor: Benjamin Ace
Participants: Gabriella Razzano - ODAC
                   Benjamin Ace - Heinrich Boll Stiftung
                   Greg Kempe - Code for South Africa
                   Adi Eyal - Code for South Africa
Venue: G
Time Slot: 13:30

Context:
There is an issue of access to technology when promoting open data.

1. Don't try directly as programmers, work through civil society with connections and demonstrate usefulness.
   - Marketing issue - buy-in comes from use
   - People who have organised on a topic - people who don't care as an example
2. Translating data in non-tech language, building familiarity or using tech that 'makes sense'.

Problem:
Different communities dialogue differently - so how do we centralise as we contextualize?
- Discrete solutions -> hyperlocalising to allow organic communication?
- How can the urbanite + rural meet? Identify intermediary - radio? mobile? Other mediums?

Problem:
Breaking down information silos. How many organisations replicate data mining/collection in isolation?
- Perhaps generate a central forum where data requests are put forward to the wider community of NGOs/Civil Society/Researchers/Etc so as to not waste time/money generating information that already exists.

Alternative Ideas:
- Awethu as a blog/call centre?
- Micro-blogging and civil society organiser

Want to move from "issue" to IMPACT
- Coalition-building and want citizens to have access.

Coding for Journalists

Convenor: Giovanna Gerbi
Participants: Jason Norwood-Young - Code for South Africa
Giovanna Gerbi - Eye Witness News
Friedrich Lindenberg - ICFJ
Khadija Sharife - Investigative Dashboard
Venue: H
Time Slot: 13.30

Scraping
- Coding is a craft. Who uses it? Journalists could use it, but it's not an easy skill to acquire. Often, much interesting work can be done with tools that are sprouting up. But for in-depth data work and making news apps, it's still a necessary skill.
- Scraping data: now you don't need to know how to code to use it. New tools have made "scraping" more accessible to journalists, and it will only get easier.
- Problem? It is solvable with coding resources but that would do the job differently

Example - Health Dept has stats on hospitals

- PDF docs can be treasure troves of information
- Document Cloud: Upload PDF and info gets extracted
- Nkandla is a good example of how coding can help an investigative journalist to do an in-depth story

Tools / Websites
- Source.opennews.org
- import.io - website scraping. Click on info that you need and the tool fetches data and updates it
- Churnalism - looks for press releases and sees how much journalist has plagiarised
- Storyful and Listening Post are organisations that source news content from social media and verify the accuracy of that content.
- A tool exists that shows which stories do well on international sites and one can plan a news diary using it.

Data Journalism - how to get information
- Primary source can be DATA and not necessarily from a person
- Great stories buried in data
- Education Department has tables and tables of data for example: 20 kids died at one school.
- Data journalism is just regular old journalism with new tools
- Journalists need to learn new skills such as coding in order to be able to take advantage of all the available data.

Open Data and Financial Health

Convenor:	Kenny Inggs - 22Seven
Participants:	Cor Schutte - City of Cape Town
	Brett Horner - Daily Dispatch
	Belinda Lewis - Praekelt Group
Venue:	J
Time Slot:	13:30

- Western Cape – "Transport index" is a lagging indicator of effectiveness of transport - % salary spent on transport
- Bottom 40% of earners spend 60% of salary on transport – ultimately want to halve this
- Multiple sources for financial health data:
  - Reserve bank
  - Census (issues of self-reporting)
  - Other banks
  - Surveys
- People have privacy concerns even it's anonymised
- 1st step to financial health and responsibility = awareness = show financial data relative to some average
- Could be brought into school/even a financial literacy course
- Especially education around debt – compound interest
- No household level financial data currently exists (census very vague and not detailed)
- Give opt in at a personal level to access everyone's data
- Be able to compare your spend to your community
- Access real financial data to effect public policy and personal behaviour – eg. Spend less on your car by leaving for work later
- Allows you to access the "best practices" of your peers to improve your financial health
- How do we get data that doesn't exist? And how do we share and unlock value of data that we have?
- How do we convince people that "potential risk" is worth greater good?
- How do we educate people around how much data already exists about them? And how this "data" belongs to "us"?
- Personal open data contracts with your financial institutions – eg. They must share with you. Want to expand this with ability of people to open up their data on an opt in basis.
- Banks don't share data – lose potential to design better, more efficient products for their customers
- Credit ratings shared in "people readable" format between banks
- What behaviours/data should be open to up financial responsibility:
  - Revenue and tax data
  - Savings
  - Debt (amount and type)
  - Demographics
  - Do we ask for this data? OR make use of available resources rather to come up with indicators that exist
  - Any and all data can assist with at least business cases and modelling
- Does it help to start with people who need it most or people who we have data for?
  - Eg higher income groups with increased financial health pump this money back into economic growth within the country
- By understanding financial data – private sector can make relatively simple changes to improve lives of employees
  - Eg. Work flexitime, work from home
  - Way to share education through narratives and personal stories
  - Eg. We only have one car and cycle to work etc
- Can help drive a sharing economy?
- Mobiles as a source of data

Transport

Convenor: Cor Schutte - City of Cape Town
Participants: Kelsey Wiens - Creative Commons
                  Philip Kiracofe - Horizen Ventures
                  Osman Siddiqi - Code for South Africa
Venue: A
Time Slot: 14:30

- Mini bus routes - rates: routes are only 20% accurate
- Regulation moving away from province into City of CT
- In discussion with MIT to help organise tech for route Ac
- Open data critical for innovation- need to be made open
  - Business case around why data should be available

What are the top problems to transport?
- Lower cost to user rates, alleviate revenue stream
- Efficiencies - off peak travel
- Accountability
- Culture of public transport

Initiatives and ideas
- Mine Twitter account for common complaints
- Golden Arrow data to plug into a Google map
- Current vision of a single payment across different modes of transport. Centralised routing system and schedule
- The idea behind My CITI to Khayelitsha is to increase the number of trips within Khayelitsha which means barriers to open transport
- PR risk - we will fail, we need to fail well
- Low hanging fruit
  - Bus routes - empty bus
  - Cycling safety
  - WAZE for cyclists

Open Data APIs

Convenor: Jeff Fletcher - Information Systems / Limn
Participants: Adi Eyal - Code for South Africa
                    Lonwabo Jwili - IEC
                    Kenny Inggs - 22Seven
                    Vuyisa Qabaka - CiTi
                    Koteso Moeng - StatsSA
                    Jason Norwood-Young - Code for South Africa
                    Tarryn-Ann Anderson - Paperight
Venue: B
Time Slot: 14:30

Why APIs?
- An API should be published by the data producer
- API provides up-to-date information and ensures that any improvements to the underlying data are immediately available
- The API should be part of the operational database so that data available through the API is always up-to-date.
- APIs are mostly useful for dynamic data.

Implementation
- Can be costly if there are many users (e.g. elections API during the elections)
- Needs to be constantly maintained

Available APIs
- Weather / Wunderground
- IEC
- South African Tourism
- Flight info is copyrighted
- Reserve Bank
- JSE (although delayed)
- Code4SA list
    - wards.code4sa.org - converts addresses to boundaries such as police precincts, wards, census subplaces, etc
    - maps.code4sa.org - provides maps data that can be used to create maps quickly
    - elections.code4sa.org - elections API
    - mpr.code4sa.org - medicine prices
- StatsSA is contemplating developing APIs
    - Electricity
    - CPI
    - Labour force

Data Literacy

Convenor: Charles Kimpolo
Participants: Dillon Mann - WWW Foundation
                   Jessica Taylor - IBP
                   Dijhaba Selemela - Thoughtworks
                   Hannah Williams - Code4SA
                   Wellington Radu - Media Monitoring Africa
                   Paul Kennedy - Science Journalist
                   Loren Treisman - Indigo
                   Liesl Pretorius - City Press
                   Katherine Eyal - UCT
                   Ory Okolloh - Omidyar Network
                   Murray Hunter - Right2Know Campaign
                   Brett Horner - Daily Dispatch
                   Douglas Day - dotLogic Consulting
                   Janet Jobson - DG Murray Trust
                   Izak Minnaar - SABC Digital News
Venue: C
Time Slot: 14:30

- Data literacy
  - Accessibility
  - Audience
- Infomedia – radio
- Target
- Social Enterprise
  - Hybrid
- Data visualisation is a good way to do it
- School advisor (data visualisation tool)
  - School, pass rate, infrastructure, contact details
  - Department Education: data all available online although in PDF format = literacy in delivery
  - Audio visual culture = infographics. This can convert the message more effectively than raw data
  - There are needs to personalise the experience
- Need to make data articulated to call to action and for specific communication
- Lessons can be take from advertising (what works, what doesn't)
- Usability is key for literacy
- Crowd sourcing data
- User generated data
  - Usability
  - Focus groups
- Gameification – these can facilitate data literacy

Different Intermediaries (schools, media, universities, private orgs) ->

Consumable information (change, transform, enrich) ->

Consumer

Right to Privacy

Convenor: MIA
Participants: Benjamin Ace - Heinrich Boll Stiftung
                     Belinda Lewis - Praekelt Group
                     Mervyn Naidoo - IEC
Venue: D
Time Slot: 14:30

- How does SA Electoral committee implement electors security
  - POPI
- What do we view to be our digital privacy rights?
  - Generational – older people are more attached to the notion of privacy
  - Younger people are less concerned about privacy
- Implement digital privacy 101 courses in schools to teach children how to protect themselves online

Open Data in Education

Convenor: Megan Beckett
Participants: Phillipa Yeats - Praekelt Foundation
                   Keren Ben-Zeev - Heinrich Boll Stiftung
                   Hopolang Selebalo - Equal Education
                   Paul Murphy - Western Cape Government
                   Zimikhita Buwa - Britehouse
                   Megan Beckett - Siyavula
                   Petrus Janse van Rensburg - Code4SA
                   Sarah Goodier - OpenUCT
Venue: E
Time Slot: 14:30

- Looking for school infrastructure data at province level (tenders and contracts a struggle, building and maintenance).
- Trying to see what is going on and accountability
- Survey of schools and infrastructure – every 4 years – latest not available to Equal Education yet
- Utilise network to collect data
- Business model around school data – get learners teachers etc to ID problems – selling to government and service providers
- If national government not providing, what about provincial?
- Matric results for public consumption; Thoughtworks
- Integrate data sets (results and infrastructure) to see impact
- Trust data received by government?
- Data on teachers in schools? – EE writing a report on that (info: from parliament etc.)
- All state employees (-aggregated data possible) on state system but personal info very locked down
- SMS line for teacher late-coming…?
- How many computer facilities and how many in use?
- Education movement; quick implementation; lack of skills in this area; hardware can be donated but skills/teachers not necessarily there
- Produce tech to function on devices people have (software productions not hardware investment)
- Siyavula's uses base: >1 million kids use it per month; see what is being followed and when
- What data does Siyavula collect?
  - Based around maths and science (by subject and grade) curriculum
  - Google analytics
  - Mxit data – most hits
- Investment could lead to open data
- Open data: Loads of educational data – possibly build tool to make DBE's life easier – make open
- We have mapping tool with schools broken up into Quintiles, can click through to school data – would be great to make publically available
- Health data on different systems, difficult to aggregate
- Possibly build system to handle basic data to serve provincial and national needs
- Do teachers have time/skill to input data:
  - Possible standardised test results
  - Minimise drain on teachers' time
- StatsSA interested in getting that info too – possibly work together
- Approach collection of data in a positive/economic way
- Could be capacity issues around data collection

Data Integrity

Convenor: MIA
Participants: Karabo Rajuili - My Vote Counts
Neil Hoorn - City of Cape Town
Roger Gachago - Saflii
Kevin Tabisher - City of Cape Town
James Aphane - IEC
Lynn Woolfrey - UCT
Fortune Sibanda - Google, SA
Venue: F
Time Slot: 14:30

Issues/Problems
- Creating data feedback loops: How to create incentives for users to provide feedback? To improve quality
- Bad/weak data quality used as an excuse for not publishing info (eg. Government department)
- Quality of communication system "finding a problem but no one to report to"

Options/problem-solving
- The more data is published the more quality improves. Make it public! Create awareness!
- Quality affected by adherence to standards

Standards
Quality and completeness of data
- 6 standard criteria:
  - Accuracy
  - Relevance
  - Comparability
  - Timelines
  - Security
  - Interpretation
- Source of data will influence perception of quality and accuracy (this can be problematic)
- Becoming a trusted data source take time: example of data first 10 year journey
- Quality is linked to usability

Anonymyzing Data, "Making Data Safe"

Convenor: Greg Kempe
Participants: Lynne Woolfrey - UCT
                  Koketso Moeng - Stats SA
                  Paul Murphy - Western Cape Government
                  Phillip Kiracofe - Horizen Ventures
Venue: A
Time Slot: 15:30

Problem:
Finding middle ground between personal security and losing data relevance by zooming too far outwards. Standards need to be developed that are relevant to government and users for accountability and understanding risk.

Context:
Personal information in data is considered risky to share, however by anonymising data you loose contextual relevance and its raw form. NYC had made data available which ended up compromising personal info.

Open data policy being drawn up.

POPI - Act may dampen push for open data.

Helpful Sources:
StatsSA - raw data can be made available if process is followed.
Anonymization standards - UCT Datafirst flowchart, US Census Bureau, Stats Canada
Considerations:
Risk analyses of datasets - how much zoom balances 'risk' and contextual relevance.

The focus on privacy may be due to a fear for big negative publicity from a small number of mistakes.

Data can be very messy, with big differences between departments, institutions, and datasets

Civil Activism

Convenor: Brett Horner
Participants: Douglas Day - dotLogic Consulting
Murray Hunter - Right2Know Campaign
Venue: C
Time Slot: 15:30

Questions:
How do we create public activism for open data?
How do people engage in the process and publish data?

A lack of consumable technology is prohibiting access to many to publish their data.

Neighbourhood watch is a good source of crime data (Crimestopper).

- Crowdsourcing
  - The good – Crimestopper, Hello Peter
  - The bad – Pigspotter
- Open data activism – need to market this as something else to engage the public, data is not something the public can relate with as important.
- General public doesn't think about 'data', they may think about a service or use'. Will get involved in ways that won't necessarily be thought of as 'data'. For example, reporting broken street lights is not thought of as 'contributing data'

Healthcare Data - Trends and Developments

Convenor: Tarryn-Anne Anderson
Participants: Jeff Fletcher - IS/Limn
                    Kenny Inggs - 22seven
Venue: D
Time Slot: 15:30

Question: How can we reduce costs (time and financial) by mechanisation?

Code for South Africa: Medicine price API and also have hospitals info, but the latter is not yet in API form

Department of Public Health: pricing regulations around medicines and consults.
Birth and death info, age and demographics as well.
Nurse time spent on data capture can prove useful.

Trends in health information might need to come from secondary organisations including insurance companies and suppliers particularly hospital dispensaries track drugs carefully - can be used to extrapolate medicine supply and demand

Various apps / SMS prog to remind people to take medicines - though impacts of SMS programs in compliance is mixed at best depending on the context.

How to Fail Successfully

Convenor: Justin Arenstein
Participants: Roger Gachago - Saflii
Keren Ben-Zeev - Heinrich Boll Stiftung
Neil Hoorn - City of Cape Town
Liesl Pretorius - City Press
Wellington Radu - Media Monitoring Africa
Izak Minnaar - SABC Digital News
Venue: E
Time Slot: 15:30

How to Fail Fast:
1. Clearly define what impacts you want & then design a strategy to achieve it
2. Use free tools / solutions while experimenting
3. Build for re-use so you can make the argument for ROI
4. Reinforce a culture for tweaking / recalibrating strategies if they aren't delivering impacts
5. Once a responsive culture is established, managing failures becomes easier / cheaper
6. Do a lot of small things
7. Only scale the stuff where there is evidence of real traction / demand
8. Create easy entry points for data pioneers with roadmap for up-skilling

The General Public and Open Data - Education, Ethos and Use

Convenor: Belinda Lewis
Participants: Phillipa Yeats - Praekelt Foundation
                   Sarah Goodier - UCT
                   Belinda Lewis - Praekelt Foundation
                   Hannah Williams - Code4SA
                   Jessica Taylor - IBP
                   Cor Schutte - Transport for Cape Town
                   Zimkhita Buwa - Britehouse
                   Londwabo Jwili - IEC
Venue: H
Time Slot: 15:30

Q: How do people extract real value from their own data?
Show immediate demonstrable benefits of providing your personal data - e.g how does your school rate.

Q: How do you build people friendly services on top of open data sets?
Accountability at each stage of the value tree drives uptake
   • Drivers drive better eg. taxi rating system
   • Commuters get cheaper rates or improved routes eg. taxi rating system
   • City gets better civic planning eg. taxi rating system
   • Rating systems provide immediately recognisable value for people based on open data e.g taxi ratings
   • Domestic workers connected better to employment through recommendations

Q: What is the draw card for open data in communities? How can we educate people around this?
Value of data is often based on its use - what are the goals of citizens, what would be valuable to people?
On the ground research is important - don't make assumptions on what people value
Personalisation & feedback helps people connect with data

Q: How do people know that the data exists?
There needs to be some sort of public awareness, but again don't make assumptions about what people understand - research is important here too

Q: Tech not so hard. Policy a bit harder. Processes are really hard
1. Behaviour change UCT case study: solar ovens - no uptake because of cultural gender issues and not actual tech - (Diffusion of Innovation Theory - original paper can be seen here http://solarcooking.org/lesotho1.htm but there's a lot more about it online)
2. People have to feel like they own data. People also have to be accountable for their own data, it helps if people can connect with data on a personal level eg. see how they compare to the community average

Q: How do we ensure all levels of data are useful? How much data do you need for something to be useful?

Outcomes:
1. Combinations of data sets - e.g. data sharing really unlocks value, e.g Department of Women Children and People with Disability lets IEC people know areas with high numbers of poorly sighted people. IEC can provide correct voting forms in these areas
2. Start at grass roots community level to give people info they need to make informed decisions at a community level - don't make assumptions about what you think a community needs
3. Empower and educate people as to the value of their data - even before aggregation.

Corporate Commitment to Open Data

Convenor: Paul Kennedy
Participants: Karabo Rajuili - My Vote Counts
                    Janet Jobson - The DG Murray Trust
                    Benjiman Ace - Heinrich Boll Stiftung
                    Charles Kimpolo - Thoughtworks
                    Ditjaba Selemela - Thoughtworks
                    Hopolang Selebalo - Equal Education
                    Dillon Mann - WWW Foundation
Venue: J
Time Slot: 15:30

Summary
Getting corporates to release data openly is a step too far in SA at the moment and beset with issues over personal privacy and business performance. The prevailing view was that it would be far better to get governments and actors working closely on open data initiatives that businesses derive value from first before moving to more open corporate data. The exception was the idea of open finance statements from publicly listed companies which we all agreed would be a good idea.

Challenges and limitations
IBM, Google, Twitter, mPesa: Big Data Sources.

Problems:
Commercially valuable - Monopoly differentiates the corporation
Privacy

Reasons why it should open:
Argument for "human rights" motivation
Corporate info can impact governance issues
Data on mobile buying habits used for food aid

Political donations:
"name and shame" into sharing data
Could meet resistance through negative impact on commerce

Legal:
Requirements in SA unclear

Action Points:
Civic organisation "observatory" needed – must find a way to be valuable to corporations
Open bidding process for tenders would be a good start
Limited opening of data for projects/hacks as a trial run/case study
Leverage corporate social responsibility programmes

Conference Notes
Day 2

Data Visualisation

Context

The need for visualization of data.
1. To minimise effort for other people in understanding and consuming data.
2. To share technical information with the local tech community.
3. For consumer engagement / education - how do graphs work?
4. Encourage participation in interacting with data and visualizations.

Schedule

1. Develop email group (July).
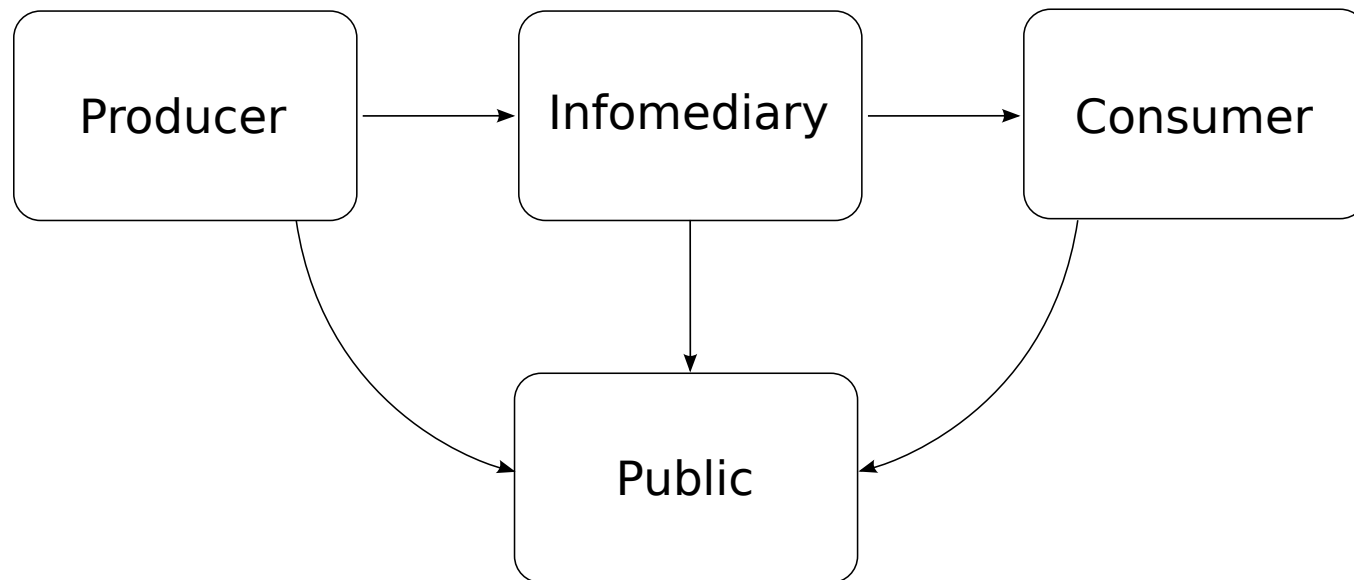2. Initiate community portal (August)
3. Meet ups (September)

Action Plan

1. Initiate a SA data viz community and explore existing players including hackers, researchers, universities, tech companies and free tools.

2. Ideas exchange forum including feedback, portfolios, and links to available data.

3. Tie (2) with an open API Catalogue with guidelines for data visualization.

4. Profile and promote completed work through GCIS newsletter and SVG

5. Email group starting immediately.

Improving Data Literacy

What?

- Literacy needed at all levels
- Different things at each level
- Producer:
  - What datasets
  - What formats (not PDFS), standards
  - What licences
  - What metadata and methods

```
┌──────────┐      ┌──────────────┐      ┌──────────┐
│ Producer │ ───▶ │ Infomediary  │ ───▶ │ Consumer │
└──────────┘      └──────────────┘      └──────────┘
       │                  │                   │
       │                  ▼                   │
       │           ┌──────────┐               │
       └─────────▶ │  Public  │ ◀─────────────┘
                   └──────────┘
```

- Info-mediaries:
  - Tools
  - Skills (technical)
  - What value is being added
  - Interpretation (expert)
- Consumers:
  - Meaning/relevance to them
  - Add value to their lives/work
  - Delivery appropriate and timely

WHO?
- Producers (produce data) : Government, IEC, NGOs, Universities, Corporate, You
- Info-mediary (consume data, produce info): Media, Research community, Developers, Marketing
- Consumers: Capture value from info-mediaries, public, companies, government, EVERYONE

HOW?
- Best practices
- Education
- Case studies

WHERE?
- No "noisy marketplace" for data (currently)
- Hacks/unconference/Indaba

Improving Data Literacy (continued)

- Institution-level education
- Community-based organisations
- Teachers
- Point-of-sale (informed customers)

All of the above are producers, info-mediaries and consumers

WHEN?
NOW!

FEAR/MISTRUST
- Gov:
  - Transparency and feedback
- User:
  - Control (individual)
  - Right to privacy
  - Opt in to specific data and uses

UNDERSTANDING
- Rights
- Do with it? – POPI Act
- Gov – protect and educate
- Translation of all relevant info:
  - Language
  - Easily understandable
  - Visually impaired/deaf etc

VALUE
- Media – good newstories
- Devs/NGOs – more data services for consumers
- Individual empowerment

WHAT DO (NORMAL) PEOPLE NEED
Tools (and education)
- User-centered tools
- More engagement between techies and users/consumers
- Basic data education
- Better dissemination (marketing)
- Single/central hub of tools/datasets (for media and other users and techies)

Building Demand For Open Data

WHO?
- Journalists
- Activists
- Organised citizens
  - NGOs
  - Social movements
  - CBOs
  - "survivalist", church, funeral
- Data analysts
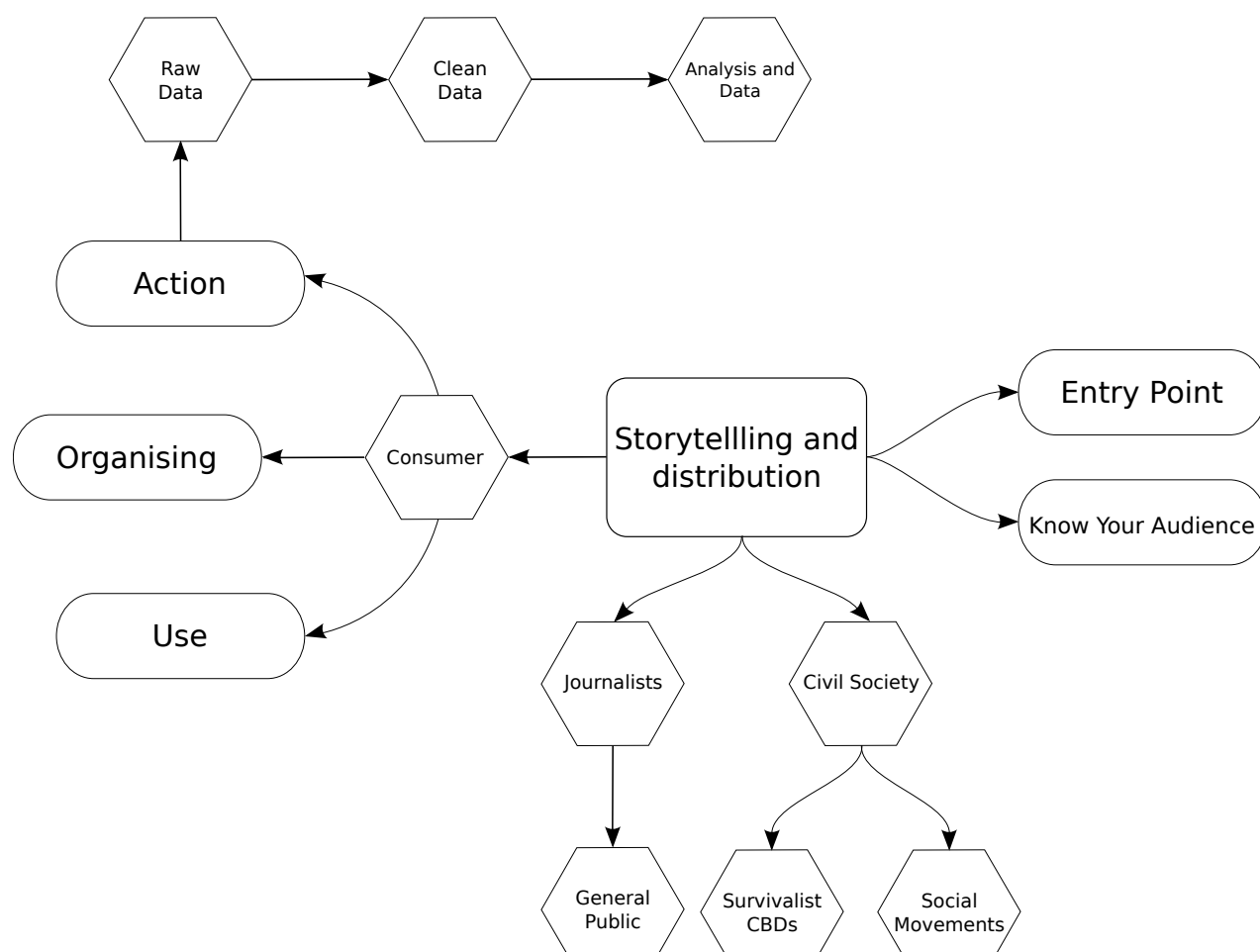
JOURNALISM/NEWSROOMS
- Journalism school
- Investigate reporting
- Graphics

INTERVENTIONS AND IN NEWSROOMS
- "Champion"- ID newsrooms with champions, create champs
  - Highway Africa
  - Power reporting
- Co-ordination and newsrooms (Code4SA):
  - Journalist
  - Designer
  - IT
    (2 day production workshop)
- Embed data analyst in 3 major media houses (according to readership/audience size)

Building Demand For Open Data (continued)

- Tool portal with relevant use case studies
- Training of journalists through existing programmes

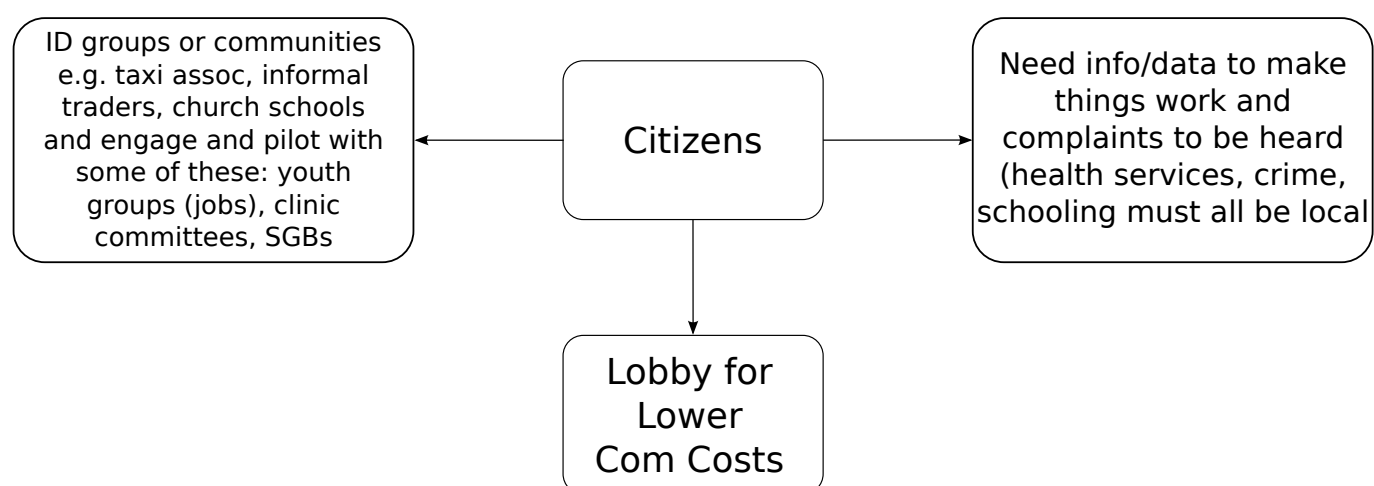**ACTIVISTS/CIVIL SOCIETY**
NGOs/Establishes orgs/mov/CBOs
Barriers:
- Know what is needed
- Know how to get access and resources
- Know how to use

Info/data for advocacy and popular education/infomedia

Info/data to help com members
- Partner tech companies with organise groups. Get more people to come to uptechs. Identify problems and work for solutions.
- Build pool of providers by bringing in put sector (sponsor interns?)
- Data dump: portal not clean data
- Internships: place prog. students in DST progs. or work on portal/work with partner orgs frm
- Case study/ies show impact and training for tools that exist
- Embed people in orgs (data analysts). Techies want stretched.
- Advertise to tech com that social w is cool:
  - Silicon cape
  - Honours students
  - Survivalist org eg. Funeral societies, stockvel
  - University societies
  - Techs and design colleges, final year students

ID groups or communities e.g. taxi assoc, informal traders, church schools and engage and pilot with some of these: youth groups (jobs), clinic committees, SGBs ← Citizens → Need info/data to make things work and complaints to be heard (health services, crime, schooling must all be local

Citizens → Lobby for Lower Com Costs

Catalogue of Open APIs

WHO?
• Better for government to do catalogue (with assistance from private sector)

**HOW?**
• Start with list, grow over time

WHEN AND WHERE?
• Get API catalogue on the GITOC Agenda (Luxolo to do)
• Need presentation
• IEC to provide some info on cost/benefits
• Jason to get info on US government data portal and Kenya portal

EXAMPLES
• Data.gov.uk/data/api (CKAN)
   • Thoughtworks built gov.uk portal
• Epsiplatform.eu
• Data.gov
• Microdata.worldbank.org (NADA – is there an API govt patent?)
• www.quandol.com

Funding of Political Economy – Mapping Financial Interests
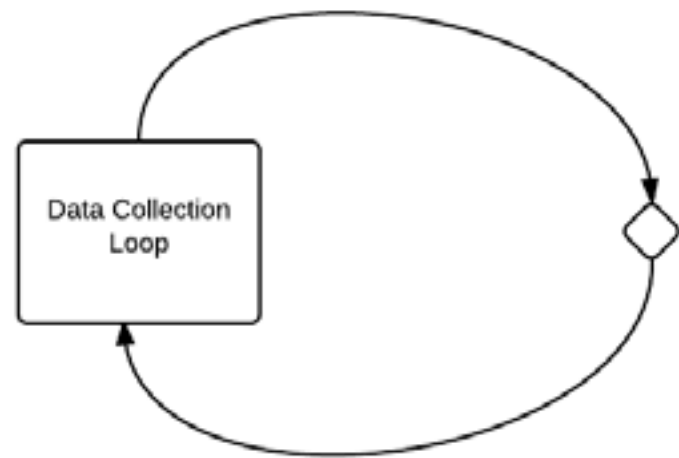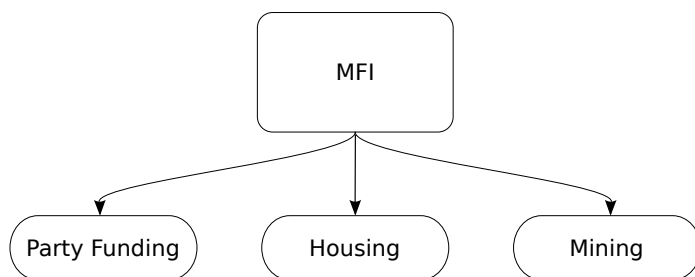
July: MFI Pilot launches

**CREATING CONTENT THAT OTHERS USE**
- Users:
  - Researchers
  - Media
  - Political party member
  - Business/commercial
  - MVC/RZK
  - Coalitions
  - State institutions
- Available resources
  - Data exists – Afrileaks etc
- How is local level info useful to MVC/RZK?
  - Building broad coalition
- Actionable
- Different uses for data (media as diff from campaign)
- Suited of tools for different users – foundational tools – entity analysis
- RZK/CW/MVC both users and sources of data
- Building trustworthy data "rumour to gospel" – needs to be substantiated
- Timeframes:
  - Small pilot, 12 July
  - ISS, organically grow
- Incentives for users to add?
  - Getting local level data
  - Infrastructure to be built
  - Giving editors/journalists tools
  - Facilitating co-op in sharing data/information
  - Users can build "special interest" data set – uploading own information "citizen engagement"
  - Curator – facilitate conversation
- Good places to start?
  - Small grant for information collection
  - Platform can host the data

Data Procurement and its Purposes

Getting data from government:
1. Simplifying PAIA to access records
2. POP 1, Data vs Information, easy to use.
3. Influence what is collected. Information not always new or as in depth.

User-based Value-add Process of Feedback

Discovery
 |
Consult Catalogue
 |
Request Data
 |
Use and Enrich Information
 |
Feedback
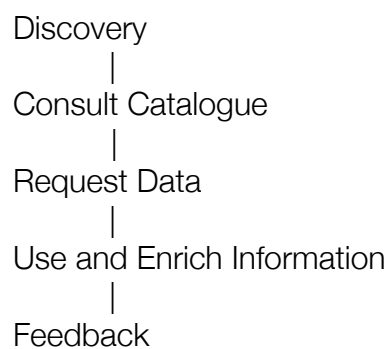
1. DISCOVERY (Not knowing what you don't know)
- Quality meta data.
- Categories in fields (black, white, green)
- Data Capturing including links for fields/international data
- Use open source if available
- Ask for forms to be filled in

2. CONSULT (asking / looking / high level research)
- Knowing about what is available
- Available Online - know where to look and who to speak with
- Legislative availability of data (classification) i.e personal data
- Aggregation / quality checkboxes e.g province vs municipal
- SAP add-on
- Source aggregation

3. REQUEST (Accessibility)
- Access
- PAIA / legal channels
- Online access (direct)
- Informal sharing (black market / non std)
- Diversity (not all eggs in same basket)
- Political (risk appetite) how could be influenced - lobbying / advocacy
- Lead by example

4. ENRICH / USE (Innovations and Transformation)
- Policy research
- Media
- Product Development & Business
- Accountability
- Advocacy through findings
- Collaboration among players
- Regulation of bodies

Data Procurement and its Purposes (continued)

5. FEEDBACK
- Standards
- Collection / integration
- (Get from Neil - City of CT - what is available?)

**Examples of Municipal Level Data Sought**
- Energy generation, loss, and capacity information, including locations and substations
- Utilities - water, solid waste / sanitation, routes for collection, frequency, electricity
- Ward expenditures, projects and SDAs
- Transportation - bus lines / schedule, train lines, taxi, fares, routes, bike lanes
- Environmental maps - wetlands / parks
- Community services - sports fields
- Housing / vagrancy - suburbs / houses
- Maps - property value / tax assessment / valuation, census
- Economic development
- Urban agriculture
- Safety / security, metro crime stats, tickets / violations of bylaws
- Health - clinics (locations), visitations, disease, restaurant inspections
- National education - schools / students, matriculation, scores

Open Data For Positive Social Change

1. Responsibility Mapping

Advocacy & Demand	Inclusivity	EducationPro-vision	Funding	Suppliers	Implementation	Interpretation	Coordination	Impact	Policy
NGOs	Private	Media	Private	Academia	Government	Acedemia	Working Group	Citizens	Government
Government	Government	Government	Government	Media	Private	Media		NGOs	
Private	NGOs	Acedemia	Funders	Private	Acedemia	NGOs		Acedemia	
Acedemia	Researchers	NGOs		Government	NGOs	Private			
Citizens				NGOs		Government			

2. Cycle of Access and Understanding Data
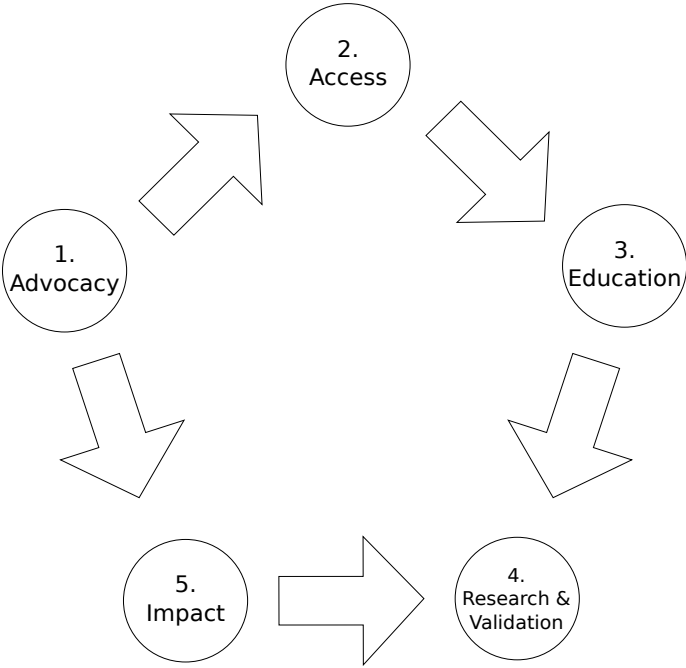
Important to get stable funding coordination for this process.

3. Summarization of Outputs

- Coordinated Working Group including all key players in the Cycle of Access.
- Data Portal which includes a training group and/or materials to educate users on particular data.
- Policy Document highlighting the process by which data should be released.`

4. The Data Portal

The Data Portal needs to be developed with established guidelines and standardization to make it accessible for different kinds of users with different degrees of functionality.

The portal should have room for Advocacy to allow for data requests to be formulated given a context, type of data, and outputs that can be expected with its release.

Training modules should be embedded in the portal to allow users to self-learn how to access it regardless of what their background is. The portal should be searchable, categorized, and easy to use as well.

5. The Working Group

The Working Group is responsible for awareness and marketing, coordination of data sources, and outreach, including documents or

Open Data For Positive Social Change (Continued)

materials on how data can help them. The stakeholders, including academia, citizens, government, NGOs, media, funders, and the private sector collectively form the working group.

To make the working group successful, several key aspects are highlighted below (not in order of importance):
- Marketing to target groups.
- Funding.
- Champions and/or leaders with vested interests
- Full time salaried people dedicated to making it function
- Should be a small and manageable size.
- Identification of groups that could utilize data for outreach to expand presence and demand.
- Formalized minutes and documented conversations through regular meetings.

Outputs from the Working Group should include at least:
- An audit map of existing initiatives (Lynn at UCT is willing to conduct) undertaken by stakeholders and extended groups.
- Explanations on initiative status in terms of stage, success, failure, and reasons.
- Newsletter to expand the network and summarize achievements, challenges, and failures.
- Exploring internal challenges in disseminating information or other functions to fail fast and reinvent as and when necessary.
- Advocacy and education materials on existing data sources.

Questions still remain on how to initiate this process including which players will take the lead at which stage of implementing the Working Group

Commitments

Justin Arenstein - African Media Initiative / International Centre for Journalists:
Commit seed funding and tech support and media coverage/partnership to:
Map influence in money data initiatives
Starting a data viz community
Creation of an API catalogue
Maintain and grow connections and networks established at #OPENDATANOW
Create a R100k fund for open data experiments

Adi Eyal - Code for South Africa
Will disseminate outputs of conference
Will monitor commitments and report back to unconference participants
Convene informal stakeholder discussion group
Will actively seek opportunities to use data and technology for the purpose increasing citizen participation in governance

Hannah Williams - Code for South Africa
Create "marketing pack" for data advocacy
Open data for social change working group
Help with any graphic design
Data viz community
Osman Siddiqi - Code for South Africa
Explore existing data for the purpose of making
Lynn Woolfrey - Data First, University of Cape Town
Audit of SA open data organisations

Nomsa Masuku - Independent Electoral Commission:
Disseminate learnings/proceedings from this interaction amongst institutions for supporting democracy through the forum for the same
Maintain contact with participants and their networks so that my own understanding remains current
Support depth of understanding of the move from information to data in state institutions

Lonwabo Jwili - Independent Electoral Commission
Become an advocate for open data

Loren Treisman - Indigo Trust
Fund programmes using data to stimulate social change
Make relevant connections between civil society, activists, techies, journos
Talk about open data and relevant case studies – social change in international. press/conferences

Jessica Taylor - International Budget Partnership
Work to make disaggregated budget data available to empower people to hold government to account with regards to service delivering

Jeff Fletcher - Internet Solutions / Limn, Paul Murphy - Western Cape Government, Koketso Moeng - StatsSA
Create an SA data viz community

Raymond Joseph - Listening Post
Educate and act as a bridge between techies and journos

Shaun Russell - Ndifuna Ukwazi
Release City Contract and PAIA data received from internal projects
Participate in visualisation forum
Release Khayelitsha sanitation data/mapping information
Greg Kempe - Open Bylaws
Grow tech skills in this space by evangelising to my community and supporting growth of new entrants

Belinda Lewis - Praekelt Foundation
Commit to active participation in the new SA viz community and share all the work we've done in that space.
Philipa Yeats - Praekelt Foundation
Continue to drive open data initiatives in my organisation, especially with new clients and national rollouts of tech services

Izak Minnaar - SABC / South African National Editors Forum:
Champion open data via Sanef
Facilitate data journ training in newsrooms
Driving open data reporting projects at SABC news
Willing to participate in open data working group

SAFLII
Support the "open data now" tech community by working with them as much as possible – SAFLII

Rachel Ward - South African Human Rights Commission:
Champion open data within human rights commission and within the human rights movement more broadly
Link open data with internet rights
Luxolo Lengs - Statistics South Africa:
StatsSA data
GITOC info

Koketso Moeng - Statistics South Africa:
Share datasets and facilitate data viz community (with Jeff)
Develop APIs to our data future project

Cor Schutte - Transport for Cape Town, City of Cape Town
Advocate for "open" within the City

Katherine Eyal - University of Cape Town:
Use and publish more research with open data
Teach the next generation how to use it
Laban Bagui - University of Cape Town:
Advocacy
Be there
Critic

Paul Murphy - Western Cape Government
Advocacy in government space

Paul Kennedy
Work towards an SA/African open data portal
Help demystify opening up data in the academic community

Attendees

Attendees

Benjamin Ace	Heinrich Boll Stiftung	Friedrich Lindenberg	International Center for Journalists
Zaheeda Adams	PMG	Alfred Mahlangu	AfricanLII
Rashaad Alli	PMG	Dillon Mann	World Wide Web Foundation
Tarryn-Anne Anderson	Paperight	Nomsa Masuku	IEC
James Aphane	IEC	Craig McKune	amaBhungane
Justin Arenstein	AMI	Izak Minnaar	SABC Digital News
Mariya Badeva-Bright	AfricanLII	Koketso Moeng	StatsSA
Laban Bagui	University of Cape Town	Paul Murphy	Western Cape Government
Megan Beckett	Siyavula	Mervin Naidoo	IEC
Keren Ben-Zeev	Heinrich Boll Stiftung	Jason Norwood-Young	Code for South Africa
Zimkhita Buwa	Britehouse	Axolile Notywala	Social Justice Coalition
Douglas Day	dotLogic Consulting	Nombulelo Nyathela	Equal Education
Melanie du Plessis	IEC	Carolina Odman	hack4dev
Lianne du Toit	SiliconCape	Ory Okolloh	Omidyar Network
Adi Eyal	Code for South Africa	Trevor Oosterwyk	StatsSA
Katherine Eyal	University of Cape Town	Liesl Pretorius	City Press
Jeff Fletcher	IS / Limn	Vuyisa Qabaka	CiTi
Gail Fullard	PMG	Wellington Radu	Media Monitoring Africa
Abdul Gabier	Western Cape Government	Karabo Rajuili	My Vote Counts
Roger Gachago	Saflii	Gabriella Razzano	Open Democracy Advice Centre
Giovanna Gerbi	Eye Witness News	Jared Rossouw	Ndifuna Ukwazi
Stefan Gilbert	Institute for Security Studies	Shaun Russell	Ndifuna Ukwazi
Sarah Goodier	OpenUCT, University of Cape Town	Tobias Schonwetter	Creative Commons
Tim Harris	City of Cape Town	Cor Schutte	Transport for Cape Town
Neil Hoorn	City of Cape Town	Hopolang Selebalo	Equal Education
Brett Horner	Daily Dispatch	Elliot Selemela	Thoughtworks
Murray Hunter	Right2Know Campaign	Khadija Sharife	Investigative Dashboard
Kenny Inggs	22seven	Fortune Sibanda	Google South Africa
Petrus Janse van Rensburg	Code for South Africa	Osman Siddiqi	Code for South Africa
Janet Jobson	DG Murray Trust	Greg Solik	My Vote Counts
Raymond Joseph	Listening Post	Kevin Tabisher	City of Cape Town
Lonwabo Jwili	IEC	Jessica Taylor	International Budget Partnership
Greg Kempe	Code for South Africa	Alison Tilley	Open Democracy Advice Centre
Paul Kennedy		Loren Treisman	Indigo
Charles Kimpolo	Thoughtworks	Rachel Ward	SA Human Rights Commission
Philip Kiracofe	Horizen Ventures	Kelsey Wiens	Creative Commons South Africa
Ntombovuyo Klaas	Department of the Premier	Hannah Williams	Code for South Africa
Luxolo Lengs	StatsSA	Lynn Woolfrey	DataFirst, UCT
Belinda Lewis	Praekelt Foundation	Pippa Yeats	Praekelt Foundation