

Praja.Data

An Open Data Initiative for state of Karnataka
May 2010



“It's such an untapped resource,
Government data is something we
have already spent the money on... and
when it is sitting there on a disk in
somebody's office it is wasted.”

Sir Tim Berners-Lee, Web Founder



Introduction

- ▶ Information systems inside governments contain a lot of data, and this data may be of interest to citizens. Present day technology can easily make government data available to citizens on regular basis and help increase levels of transparency and trust.
- ▶ This presentation puts forward a proposal for district governments of Karnataka to collaborate on a pioneering project to make government data available to citizens on regular basis.



Opening up Government Data

Objective

Provide citizens access to community data

Unlock pertinent data from government and public systems

Enable innovative applications that make effective usage of this information

Enable quality analysis and debates on social platforms

Benefits

Citizen access to Data cuts the middleman, allows transparency and increases trust

Unlocking data enables administrators and citizens to realize value of capturing data

Encouraging creation of applications that use this data will result in new services and enterprises

Analysis on social platforms allows for Public consultation and involvement in projects

The Proposition

Collect ...

Group data by areas of interest

- Define datasets for well known “domains”

Setup process and governance for preparing datasets

- Understand availability, quality and sensitivity of data

Gather data in a standard formats

- Develop open technical standards for defining government data

And, Publish!

Setup tools for “publishing” datasets

- Datasets go out as “feeds”

Develop technical standards for access

- Application developers will access the “data feeds”

And finally, develop some sample applications

- ... to demonstrate use of open data technology

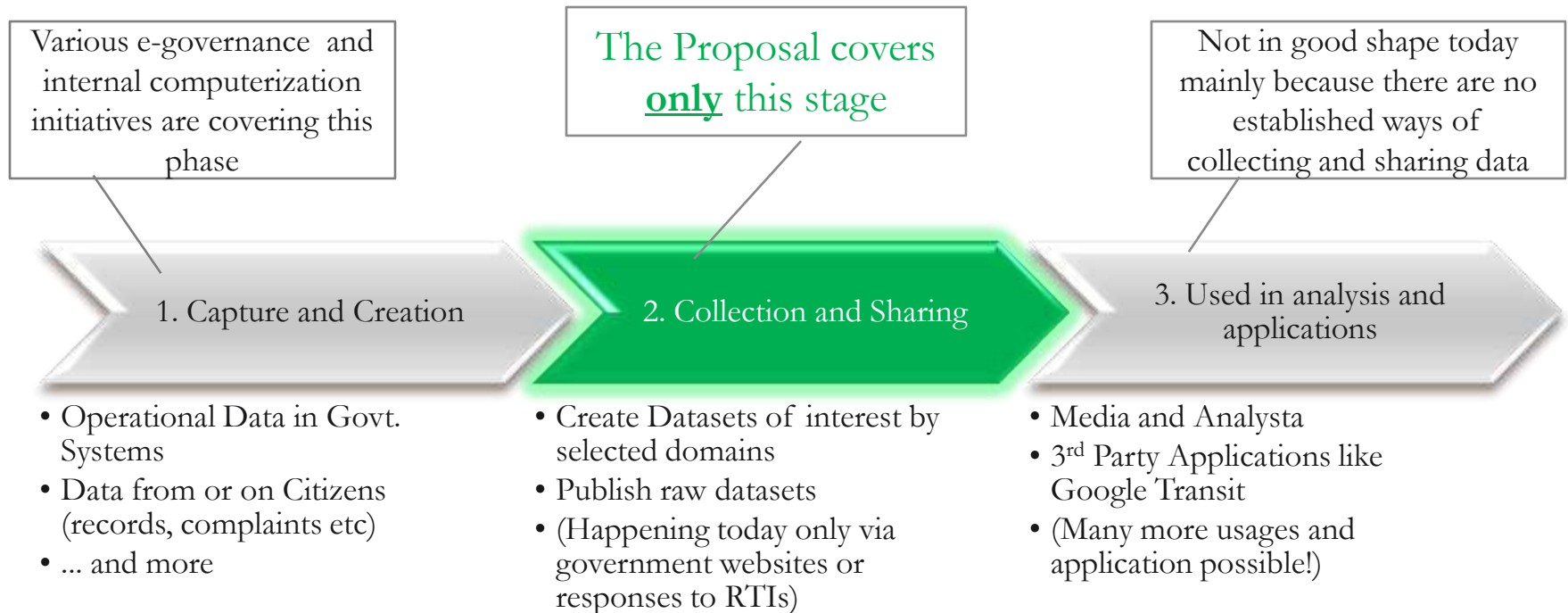


Proposition – Clarity

▶ Life Cycle of Data in Public systems - The 3 Stages

▶ Stage #2 (Collection and Sharing) is critical

- ▶ Drives requirements that provide visibility for your investments in phase #1
- ▶ Enable applications that improve quality of your interaction with citizens and media



Clarification #1 – Focus, Boundaries

- ▶ For the pilot, we will pick 2-3 focus domains, and define a few “data sets” in each of them.
- ▶ Data sets will be “published” per district
 - ▶ So, expect few data sets per domain per district.

- ▶ Focus Domains - samples
 - ▶ Law and Order
 - ▶ Ex: Police station wise break up of crime statistics
 - ▶ Education
 - ▶ Ex: Counts and Attendance records of teachers and students in primary schools
 - ▶ Food Supply - Procurement
 - ▶ Ex: Yard wise data on stocking level, prices of select commodities

Focus is not on publishing everything available, but on publishing coherent data of relevance to citizens.



Clarification #2 – Data Sensitivity

- ▶ Data can be confidential or sensitive
 - ▶ Personal Information on Citizens or government employees
 - ▶ Security Implications
 - ▶ Commercially or Culturally sensitive data
 - ▶ Incomplete data or information that may be materially misleading.

- ▶ Data sets will be defined with sensitivity in mind

Proposal will address sensitivity concerns



The Project – Phases, and partners

Partners

Government
Departments

Educational
Institutions, eGov
standards committee

IT Companies, or
Technology Industry
associations

Phase 1)
Domain
Inventory

Phase 2)
Develop formats
for data sets

Phase 3) Setup
publishing
process

Phase 4) Pilot

- Check what data is available in the given domain
- List data that is available but not digitized

- Develop Domain-neutral vocabulary and technical standards for use by all domains
- Develop vocabulary for each domain

- Tools and procedures for publishing data sets
- Technical standards for access by application developers

- Publish first data set for each domain in scope
- Demonstrate 1 sample application for each domain



Phase #1 : Domain Inventory

Steps involved (for each domain)

- Interview departments that record or archive data
- Create a list of data points available, categorize them
 - Digitized
 - Cost of regular capture
 - Possible to digitize
 - Cost of regular digitization



Output of this phase

- List of data points available in this domain, digitized or not digitized
- Cost estimate for recording these data points every month
 - Capturing data, including cleansing or re-formatting
 - Digitizing & storing data
 - Filtering for sensitivity

Estimated time

2-3 months per domain

People

praja.in/partner
– 1 lead per domain

Government support

Interview time,
inventory of
existing IT
based process



“Information is the currency of
democracy”

Thomas Jefferson





Thank you

Look forward to your responses.

Email admin@praja.in

