CHOne

CANADIAN HEALTHY OCEANS NETWORK

Data Sharing

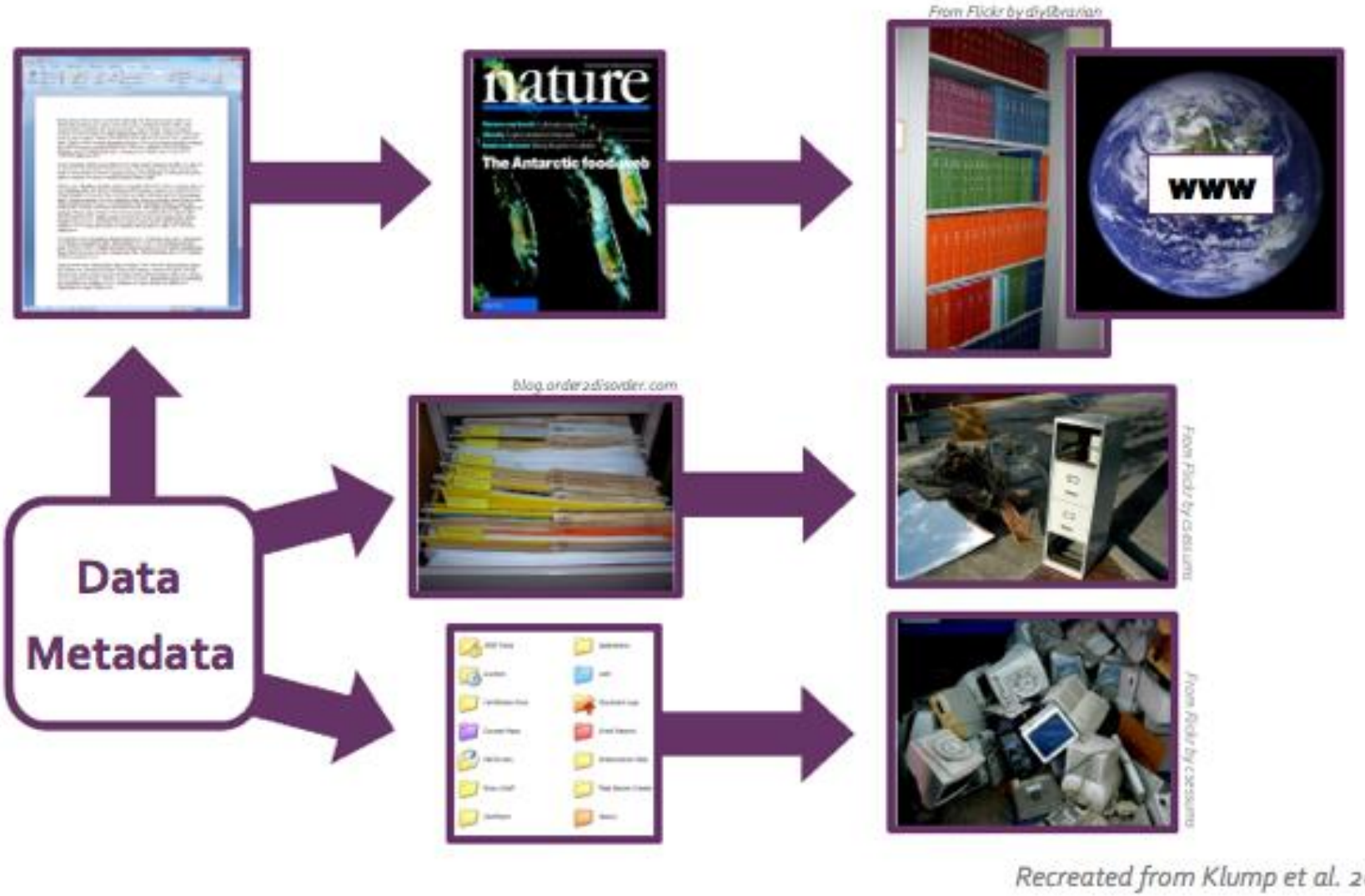
Data Management Workshop

May 2nd , 2017

Gatineau, Quebec

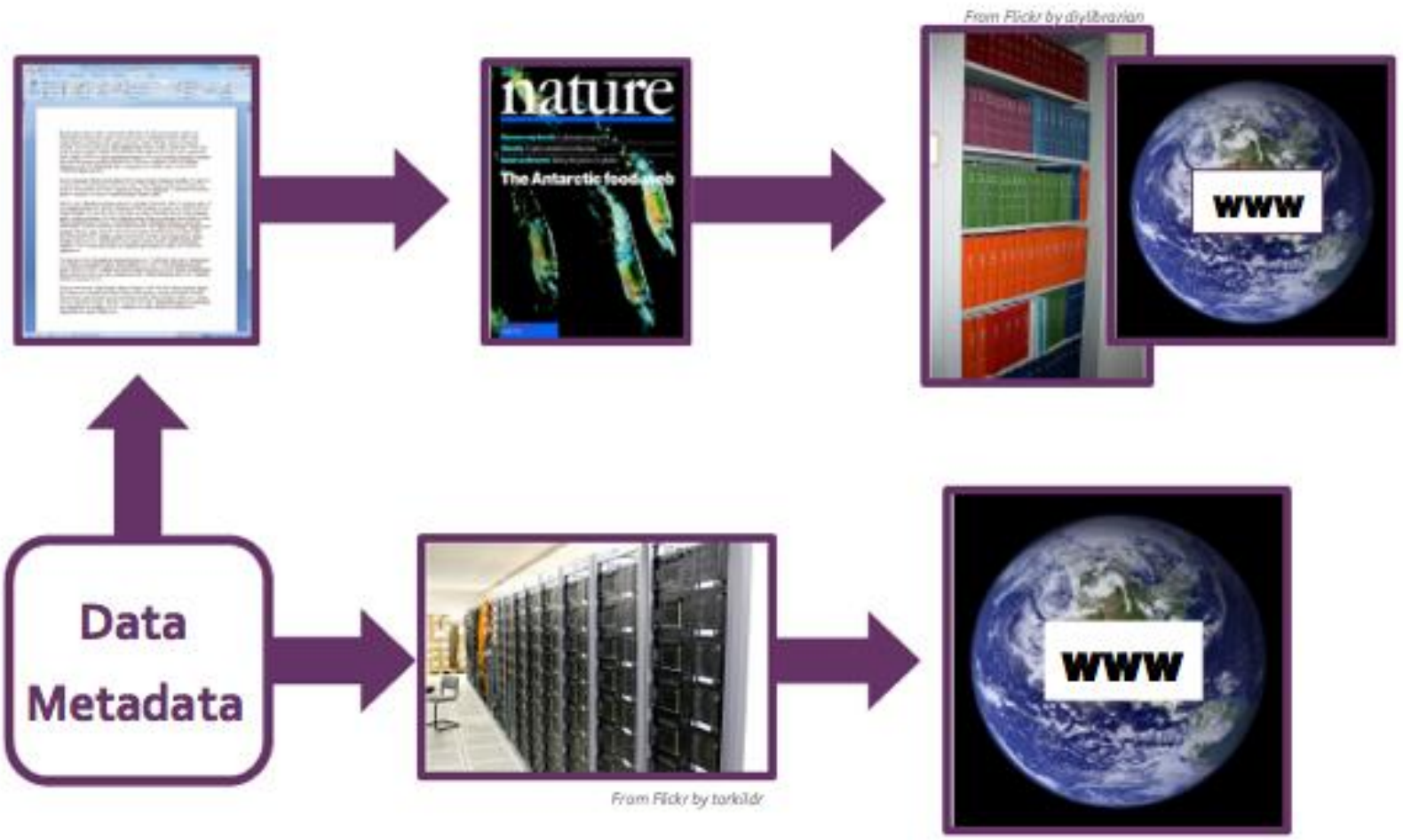
Where a majority of data end up now...

CHONe
CANADIAN HEALTHY OCEANS NETWORK



Recreated from Klump et al. 2006

Imagine if data were more accessible….



Recreated from Klump et al. 2006

Open Access and Publically Accessible Data

- Increases the **impact** and **visibility** of research
- Promotes **innovation** and potential **new data uses**
- Leads to new **collaborations** between data users and creators
- Maximizes **transparency** and **accountability**
- Encourages **improvement** and **validation** of research methods
- **Reduces cost** of duplicating data collection
- Provides important resources for **education** and **training**
- **NSERC requirements**

Data Sharing Stages

Describe	document the data content, character and process with metadata
Deposit	store the data in a location from which it can be accessed
Preserve	select storage formats and media with long term use in mind
Discover	publish information about the data so that others can find it

Data Sharing

All research data should be preserved in a **publicly accessible**, **secure** and **curated repository** for discovery and reuse by others.

[Dryad](#)

[Figshare](#)

[Biodiversity Data Journal](#)

[OBIS](#)

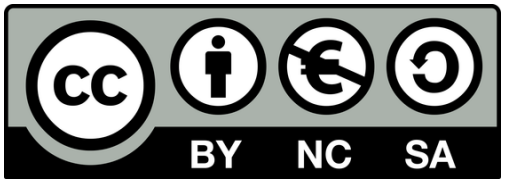
[GitHub](#)

[Zenodo](#)

Copyrights

Why use an open license?

- Facilitate data sharing and discovery
- Increase visibility of your data
- Advance science and public knowledge



Creative Commons Licenses:

CC0 (not a license, but a waiver)

CC BY (Attribution)

CC BY-ND (Attribution-NoDerivs)

CC BY-NC (Attribution-NonCommercial)

CC BY-SA (Attribution-ShareAlike)

Others:

- **Open Data Commons Open Database License** (ODbL)
- **MIT License** (for Software)

Acknowledgements

All users of research data should acknowledge, through citation and any other practices or standards relevant to their discipline(s), the source(s) of the data they are using and respect the terms and conditions under which these data were accessed.

We recommend the following acknowledgement when presenting your CHONe research:

"This research is sponsored by the NSERC Canadian Healthy Oceans Network and its Partners: Department of Fisheries and Oceans Canada and INREST (representing the Port of Sept-Îles and City of Sept-Îles). (NETGP 468437-14, CHONe Project #.#.#)"

Concerns About Data Sharing

Concern	Solution
inappropriate use due to misunderstanding of research purpose or parameters	
security and confidentiality of sensitive data	
lack of acknowledgement / credit	
loss of advantage when competing for research dollars	

Concerns About Data Sharing

Concern	Solution
inappropriate use due to misunderstanding of research purpose or parameters	metadata
security and confidentiality of sensitive data	metadata
lack of acknowledgement / credit	metadata
loss of advantage when competing for research dollars	metadata

Concerns About Data Sharing

Concern	Solution
inappropriate use due to misunderstanding of research purpose or parameters	***Abstract, Purpose, Attributes, Methods*** and ***Supplemental Information*** where needed
security and confidentiality of sensitive data	• the metadata does **NOT** contain the data • ***Use Constraints*** specify who may access the data and how
lack of acknowledgement / credit	specify a required **data citation** within the *Use Constraints*
loss data insight and competitive advantage when applying for research grants	**Open access** and **data discoverability** to promote research

Concerns About Data Sharing

CHOne
CANADIAN HEALTHY OCEANS NETWORK

Privacy laws

Sensitive data protection

Respect of traditional knowledge

Intellectual property laws

Solution!

Privacy:

- Institutional ethics review processes

e.g. Research Ethics Boards

- The rights and privacy of individuals who participate in research must be protected at all times.
- Data made available for broader use should be free of identifiers.

e.g. anonymization/de-personalization of data.

Name	ID
Angela Grant	1
Megan Bailey	2
Lucia Fanning	3

Tri-Council Policy Statement on the Ethical Conduct for Research Involving Humans, 2014

Solution!

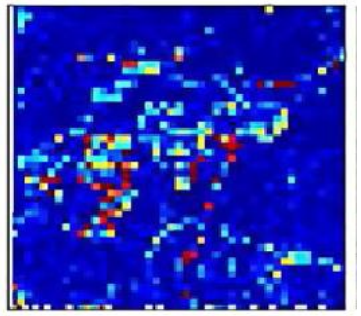
Sensitive data:

i.e. Location of endangered species or harvest data

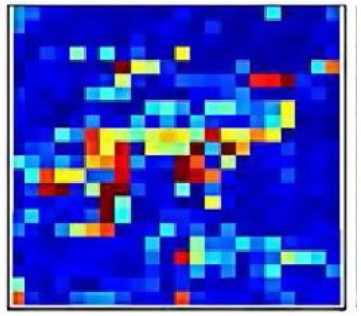
"data can be **embargoed** for a determined period of time or where possible, spatial and temporal **data aggregation methods**, e.g., the classification of fisheries data into a multi-class range (e.g., quantiles, equal interval breaks) may be applied to ensure protection of information prior to being shared publicly."

Abundance of Wolf Fish:

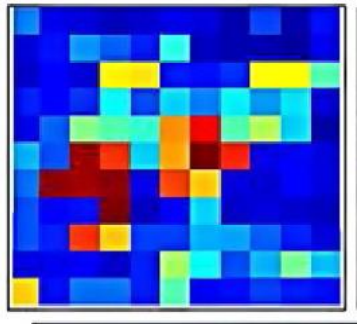
2m



20m



50m



Solution!

Traditional Knowledge:

"In the context of research involving Indigenous knowledge, data management principles based on the concepts of **respect**, **reciprocity**, and **responsibility** should be observed. This includes **appropriate engagement** of Indigenous people, communities or organizations throughout the entire data lifecycle, **formal attribution of contributed knowledge**, establishment of **informed consent** for use of knowledge and derived products, and the maintenance of contributor control of data and information resources. Required institutional ethics review processes (i.e. Research Ethics Boards) will guide data management, however **Indigenous communities or organizations may have specific practices or requirements in place**. It is the responsibility of researchers to familiarize themselves with and adhere to these practices and requirements."

The International Arctic Science Committee (IASC)

Solution!

CHOne
CANADIAN HEALTHY OCEANS NETWORK

Intellectual Property:

"It may be necessary to **delay sharing** research data for a period of time, in cases whereby institutions or researchers are applying for **patents** or developing **new applications** based on that data."

Well-Managed Data Can Result in Re-use, Integration, and New Science

CHONe
CANADIAN HEALTHY OCEANS NETWORK

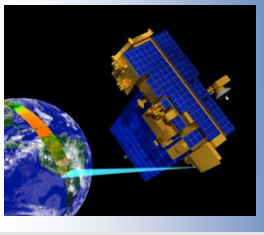
Bird Observations

Land Cover

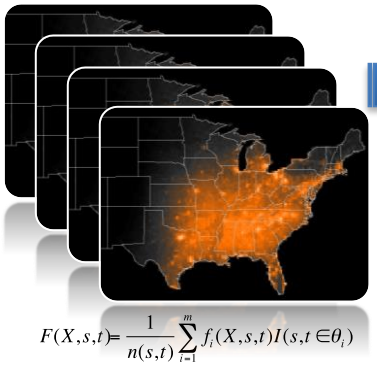
Meteorology

NASA

MODIS – Remote sensing data



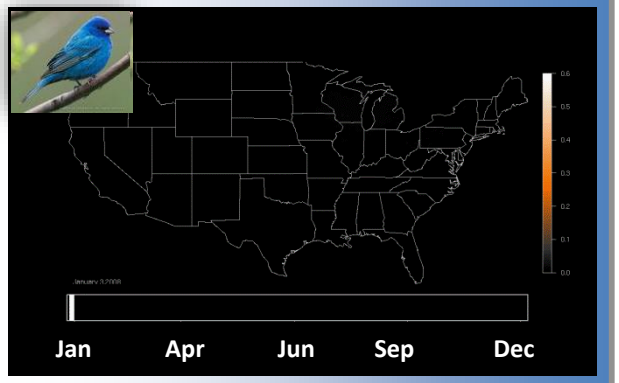
Model: Indigo Bunting locations during the year



$$F(X,s,t) = \frac{1}{n(s,t)} \sum_{i=1}^{m} f_i(X,s,t) I(s,t \in \theta_i)$$

Model results

Occurrence of Indigo Bunting



Potential Uses-

- Examine patterns of migration
- Infer impacts of climate change
- Measure patterns of habitat usage
- Measure population trends

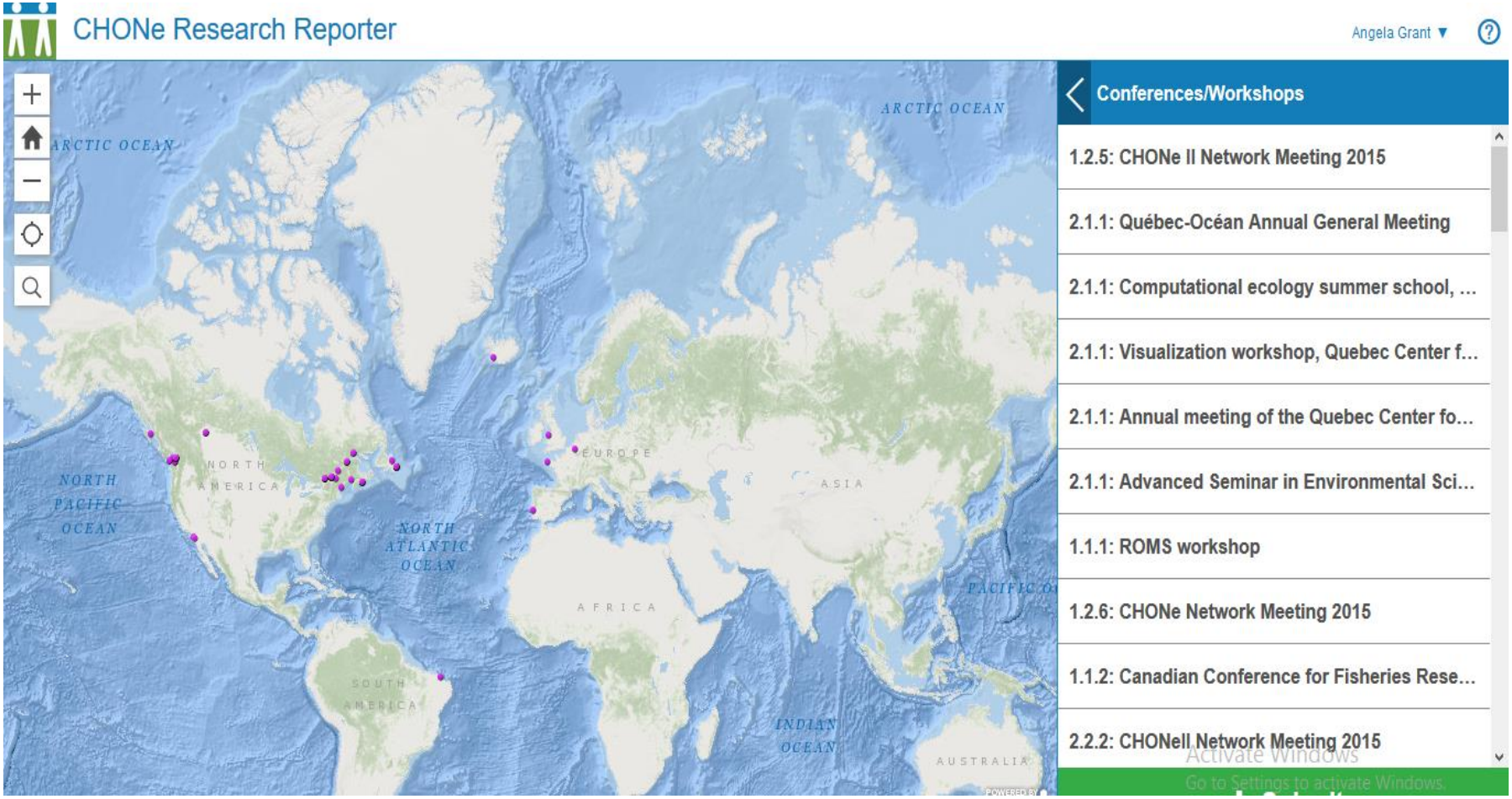
Slide courtesy of DataOne

Network Deliverables

Journal publications
Technical/policy reports
Books
Theses
Educational documents
Technical/analytical frameworks
New statistical/analytical techniques
Maps (pdf, raster, shapefiles..)
Models
Code (R, Python..)
Spreadsheets
Samples
Specimens
Video, Audio and Photos

Partnership meetings
Posters
Presentations
Public forums
Podcasts, blogs and videos
Participation in workshops and conferences and dicussions
Uptake of research by private enterprise
Contribution to management and policy decisions
Interviews with radio, tv or newspapers

CHONe Research Reporter



https://mun.maps.arcgis.com/apps/CrowdsourceReporter/index.html?appid=032ec4433b3e4e74bba51f4adecbbad9#