Meeting notes - “International Workshop on Geocoding in Global Databases”

25-26 February 2010 – Munich Re

1. Participants

Munich Re:
Angelika Wirtz
Markus Steuer
Petra Löw

University of Georgia, USA/ITOS/GIST:
Karen Payne

United Nations Development Programme/Global Risk Identification Programme (UNDP/GRIP):
Carlos Villacis
Veronica Grasso

European Commission, Joint Research Center, Global Disaster Alert and Coordination System (GDACS):
Andreas Hirner

Centre for Research on the Epidemiology of Disasters (CRED):
Regina Below
Femke Vos
Thomas Jakubicka

SwissRe:
Brain Rogers
Christina Schlenther

2. Agenda

Day 1	
9:30	Welcome and introduction; Overview of the meeting
Presentations 9:50 – 12:00	• UNDP/GRIP: Proposed Activities for the Enhanced Disaster Loss Data Component (10 minutes) • Presentation of GIST – Overview of projects (20 minutes) • Presentation of geocoding situation in data bases 2009 and 2010 onwards (Technical information and standards): - EM-DAT (20 minutes) - Munich Re NatCatSERVICE (20 minutes) - GDACS (20 minutes)
Workshop 13:30 – 15:30	• Examples of outputs - Geo-referenced flood events (CRED) - Spatial analysis of loss events (Munich Re) - GDACS database (JRC) + Discussion of the results
15:30 – 17:00	Discussion of a "perfect" geocoding world in - EM-DAT - Munich Re NatCatSERVICE - GDACS Discussion of pros and cons / Reasons why perfect geocoding of disasters is needed
Day 2	
Workshop 9:00 – 10:00	Future possible geocoding procedures in databases/data storage and retrieval
10:00 – 12:00	Development of a "best practice"
13:30 – 14:00	Capture of workshop results
14:00 – 14:30	Sources of geographical information
14:30 – 15:00	Wrap up and discussion of common paper "Best practice of Geocoding in Global Disaster Databases"

3. Report

Thursday, 25 February 2010

Introduction and overview of the meeting

Angelika W. welcomed all workshop participants and provided an overview of the aims of the meeting:

- Define the current geocoding state-of-art in disaster loss databases;
- Identify the needs and the end use of georeferenced data, as well as the challenges and possible future geocoding procedures in databases/data storage and retrieval;
- Explore a "Best practice of Geocoding in Global Natural Disaster Databases";
- Guidelines for georeferencing of disaster loss databases

Presentations

Carlos V. gave a presentation focusing on the integration, validation and interpretation of disaster information, taking the January 12, 2010 Haiti Earthquake as an example. The UNDP/GRIP proposed activities for the enhanced disaster loss data outcome area of GRIP were also presented. Among the activities planned for the next 3 years, GRIP will focus on:

- Development and field testing of guidelines, standards and training materials for disaster loss databases, on: hazard definition, loss characterization, inter-operability, quality control;
- Establishment of new National Disaster Observatories (NDOs); .
- Development of a core portal;
- Development of applications of disaster data to DRR strategies and actions.

Main outputs will be: guidelines and training modules, new NDOs, core portal, applications of disaster data to DRR strategies and actions. These activities will involve GRIP, CRED, Munich Re, LaRED, and GDACS, among others. .

Karen P. presented the ITOS Services for Humanitarian Mapping and the GIST portal, which comprises of downloadable databases for specific countries or global datasets. Furthermore, GIST provides custom IT support (an example is support to UNHCR providing Google images of Somalian IDP's). For spatially refining EM-DAT, GIST could provide polygons of the disaster footprints, as well as point locations derived from the descriptive strings present in EM-DAT. GIST holds polygons and static maps of several past disasters. As global administrative unit, GAUL would be the better choice since it is more complete, especially for the higher administrative levels (Admin0, Admin1, Admin2). For lower administrative levels ITOS Global map is somewhat more complete. Questions remain:
-Who are the end users of the georeferenced data?
-How can the georeferencing be maintained?
-How can EM-DAT georeferencing benefit from crowd-sourcing?

Femke V. presented the current state and objectives of georeferencing the EM-DAT database, and the development of a georeferencing protocol for future data. After an introduction in understanding disaster data in EM-DAT, the following issues were raised:

- typology of reported locations
- standardization of geographical information across database
- linking to dataset of sub-national administrative boundaries (i.e. GAUL).

The main questions were: how can we best identify disaster locations, what could be additional data sources, and how can we store and analyze geographical data on disasters?

Carlos V. raised a fundamental question on why georeferencing is needed, for both global and national databases. Different uses can be defined:
-Improved services for clients (Munich Re, SwissRe)
-Real-time information, potential damage estimates, validation of models (GDACS)
-Increased resolution of disaster data (CRED)
-
This also leads to the question of which level of resolution is needed for disaster data. The participants agreed on the consideration that the level of data resolution depends on which is the end use and which is the level (global, country, local) of the decisions that will be made based on this data.

Markus S. presented the georeferencing activities at the Munich Re NatCatSERVICE.
All natural catastrophes stored in this database are geocoded. The NatCatSERVICE has implemented a well-designed tool for georeferencing disasters. An automatic geocoder provides coordinates for locations. Coordinates are also checked with Google maps for validation. Since 2009/2010 natural catastrophes are being geocoded in more detail. Using the information on the affected area - area where property damage occurred and/or people were killed - and further information on the loss events (epicenter, landfall, main cause of loss etc.) more detailed analyses on natural catastrophe losses are done. Loss footprints with detailed information about the amount of overall and insured losses will be made, and the percentage of the damage that was caused by the different perils (flood, wind etc.) at a specific location will be visualized.

Andreas H. presented the current geocoding situation at GDACS. GDACS is a highly automatized system which provides disaster alerts, and some post-disaster analysis and outputs. GDACS provides near-real time information on earthquakes, cyclones, volcanic eruptions and floods, made available on its website and as email alerts to registered users. Disaster information is collected from multiple sources and automatically analyzed by a computer program to determine the likelihood of humanitarian intervention. Andreas H. highlighted the importance of interoperability of databases for data access and validation of models (such as GDACS impacts estimates, etc). GDACS expressed its interest in exploring ways to standardize the different database structures and facilitate information exchange between the databases. JRC has the technical skills to provide disaster footprints, for example by adding buffer zones to a disaster location. Two issues are to be addressed in the future:
-How do we associate single measurements with specific disasters?
-How do we identify disasters across various databases, institutions, companies?

JRC/GDACS has the technical capacities and is willing to provide its services to improve exchange of data across disaster databases. A visit of GDACS to CRED (or vice –versa) is foreseen.

Afternoon session: Workshop

The afternoon session of day 1 was dedicated to showing examples of outputs. Activities from EM-DAT (Thomas J.; Flood impacts), Munich Re (Markus S.; detailed spatial analyses of loss events) and GDACS (Andreas H.; Technical standards) were presented.
The discussion led again to several questions:

- How can we match information from different databases?
- How can we define different events (i.e.: storms and floods)
- Could the GLobal IDEntifier Number (GLIDE) help?

GDACS could benefit from the disaster classification[1] in order to enhance access and exchange of disaster data and information among organizations. Also, a GLIDE or similar disaster identifier, could benefit GDACS which is interested in sharing and exchanging disaster data with other global disaster databases.

Then, a discussion on the pros and cons of georeferencing followed. The following table shows an overview of the pros and cons.

Pros of geocoding	Cons of geocoding
Pricing	Time
Research / validation of models	Costs
Recreate / analysis of disasters and trends and predict losses for a better understanding of the phenomena	Uncertain if really used
Prevention and post-disaster management	Quality of end result ('real picture')
Resource allocation to areas	Consistency must be guaranteed, otherwise not useful
Conflict data / conflict management	
State of the art	
Allows in-house data analysis and outputs	

Focusing on the georeferencing of EM-DAT, the idea was proposed to develop georeferencing of EM-DAT step-by-step. Initially, the database structure could be maintained in MySQL, but extra tables for the storing of geographical information should be included. Likewise, the data entry interface ('EM-DAT Manager') could be adapted by including extra fields for geographic information. A standard list of additional data sources could be developed for obtaining more detailed geographic information per disaster event. The geocoder tool should be improved to have it ready to be used, including a visualization map.

[1] Disaster Category Classification and Peril Terminology for Operational Purposes, Common accord, 2009. Centre for Research on the Epidemiology of Disasters (CRED) and Munich Reinsurance Company (Munich Re).

Several geographical datasets were proposed as being useful for geocoding. Examples are JFR geomarketing (commercial), NGO providing open source dataset compilations (Thomas J.?), Global mapping initiative in Japan (Carlos V.). A list of useful datasets will be developed as an output of the meeting.

Friday, 26 February 2010

The second day of the Georeferencing Expert Working Group started with a synthesis of day 1. The need of having the disaster footprints on the one hand, and using a standardized dataset of administrative units to georeference disasters on the other hand, were discussed. The two groups of information were presented: information on the impact and scientific information.
A brief discussion on the GLIDE followed, introduced by a presentation from Carlos V. on the current situation of the GLIDE initiative. The challenges, shortcomings and possible solutions for the GLIDE were identified and discussed. The GLIDE governance issues will be discussed in the near future and GDACS offered to solve the technical issues, once the governance issues are solved.

A discussion on a best practice for georeferencing followed, taking EM-DAT as an example.
CRED took the opportunity to address specific questions to the meeting participants, in order to get feedback on how to proceed with georeferencing EM-DAT.
Main recommendations from the participants were:
-start simple;
-add additional fields in MySQL for geographical information (no transfer to PostgreSQL);
-adapt interface (EM-DAT Manager) for entry of geographical information;
-develop/improve geocoder tool, including a visualization map;
-each location should be entered and geocoded as its most precise type;
-location types should be flagged to indicate whether they are point coordinates or polygons (typology);
-precise guidelines per disaster type should be developed.
Overall, participants agreed that the current way of entering data in EM-DAT could be continued, and new geographical information should be added.

Markus S. presented a solution for storing geocoded data (see also figure 1):
Each specific location or best available geographical information per disaster event should be stored. With the help of the best available geographical data the higher levels of admin units can be identified.

> Example a) If you store the name of a city with the help of coordinates (X,Y), you get the information about all affected admin units by intersecting the point with all admin-layers (admin1, admin2, admin3 etc.).
>
> Example b) If the most detailed information is the admin2 unit, the higher-ranking admin 1 unit is identified by intersecting admin2 unit with the admin1-Layer.

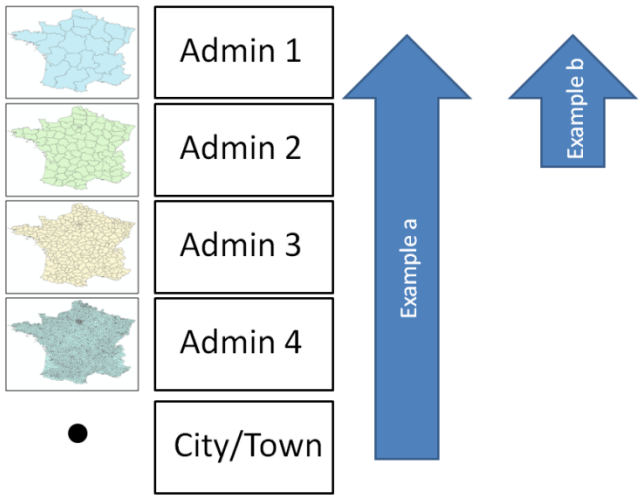

Figure 1

The steps of the proposed process in the EM-DAT database are:
Step 1: Add the requested fields in the data entry process;
Step 2: Define which unit to take for each specific disaster – distinguish the entering part from the analytical part;
Step 3: Identify weaknesses and strengths of geographical data set in order to have a better idea, and share experiences and knowledge.

The meeting action list to be addressed by the georeferencing expert working group is here synthesized:
*Creation of a list of useful public websites for georeferencing (e.g. public gazetteers, data sources,...);
*Development of the meeting report (Veronica G., Femke V.), including the pros and cons of georeferencing;

The expert working group membership will be kept as it is, , which includes the invited participants of the workshop, such as Munich Re, CRED, SwissRe, JRC/GDACS, ITOS/GIST, UNDP/GRIP. The final objective of the expert working group will be to write a guideline on the best ways of geocoding disaster data (joint document).