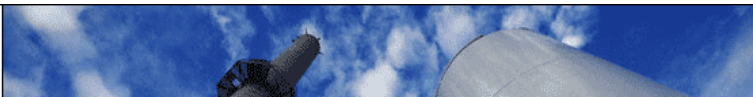
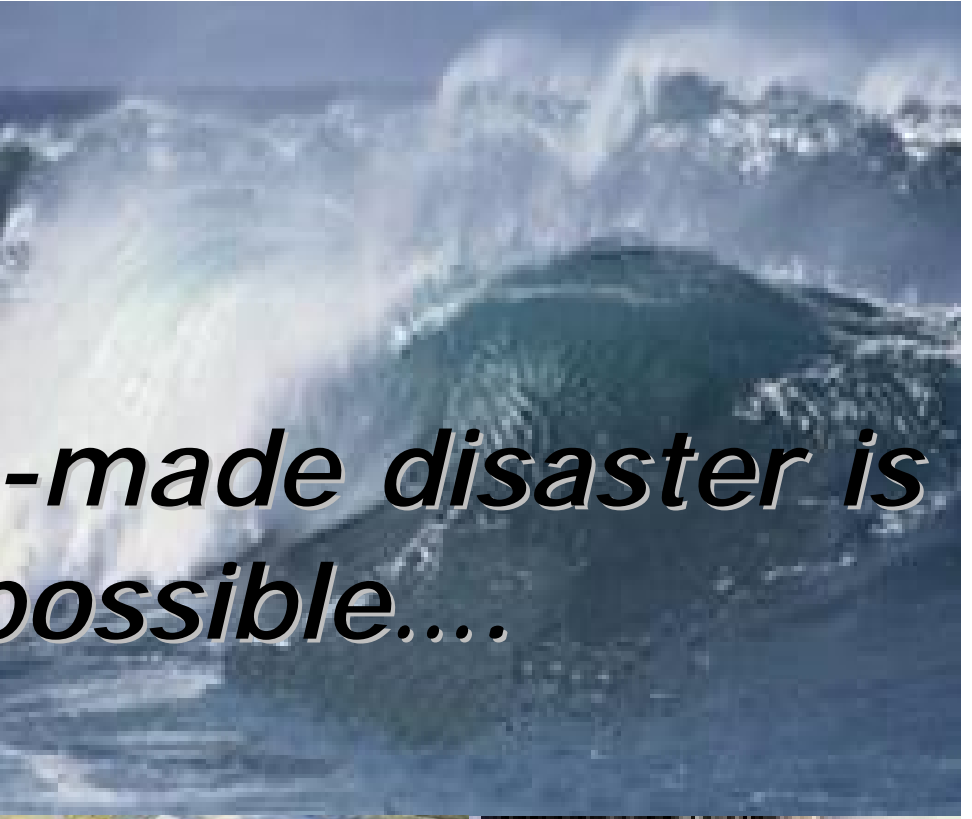


Self-deployed, web-based information aggregators for disaster-related information collection and broadcasting

Kostas Karatzas and Anastasios Bassoukos
Informatics Applications and Systems Group
Dept. of Mechanical Engineering
Aristotle University of Thessaloniki, Greece
kkara@eng.auth.gr, <http://isag.meng.auth.gr>



*A natural or man-made disaster is
always possible....*



Disaster management issues

- ⌘ Risk management is not yet a well organized discipline -> lack of unifying concepts
- ⌘ Unclear organizational responsibility for information generation
 - ⌘ lack of quality reporting -> lack of historical data
- ⌘ Incompatible information systems -> access to relevant data is not easy
- ⌘ Risks are handled in isolation
- ⌘ No clear methodology to handle inter-related risks
- ⌘etc....

Source EU

Disasters in the EU policy context

- ⌘ EU's 6th Environmental Action Programme & EU Sustainable Development Strategy
- ⌘ Civil Protection Community Action Program
 - **flood, fire, earthquake, landslides, marine pollution...**
 - **early warning, alerting the population, crisis management, emergency communication**
- ⌘ Development & Humanitarian aid, solidarity & cohesion funds, Common Foreign and Security Policy (CFSP)
- ⌘ Initiatives: INSPIRE, GMES, GEOSS ...

Source: EU

Disaster Management Cycle

Prevention and Mitigation

- Hazard prediction and modeling
- Risk assessment and mapping
- Regional/city Planning
- Structural non structural measures
- **Public Awareness** & Education

Preparedness

- Scenarios development
- Emergency Planning **maps**
- Training

Alert

- **Real time monitoring** & forecasting
- Early warning
- **Secure & dependable telecom**
- Scenario identification
- **all media alarm**

Response

- Emergency **telecommunication**
- Situational awareness, **crisis maps**
- Command control coordination
- **Information communication**
- Dispatching of resources
- Early damage assessment....

Post Disaster

- Lessons learnt
- Scenario update
- Socio-economic and environmental impact assessment

Reconstruction

- Spatial planning
- Re-establishing life-lines transport & **communication** infrastructure

Source EU

The IST Approach

- ⌞ To promote the development of cost-effective sustainable services
 - ⌞ Technology **integration** – solution driven
 - ⌞ Specific technological **developments**
 - ⌞ Market & user needs driven
 - ⌞ Focus on **generic solutions**
 - ⌞ **Re-usable** software components
 - ⌞ **Open source** software
 - ⌞ **Interoperability, scalability**
 - ⌞ Based on state-of-the-art scientific knowledge

Source: EU



Recent events have underlined the problems related to inhomogeneous information collection and updating under communication bottleneck conditions

Taken by STIA....

- ⌘ “**Low cost** information technology and **web-enabled**, location based services are driving demand for **readily available and accessible spatial data** (data pertaining to a physical earth location) for decision making in the public and private sector:
 - ⌘ Emergency 911 Response
 - ⌘ Evacuation Routing
 - ⌘ Land-use Planning
 - Crime and Law Enforcement
 - Transportation Planning
- ⌘
- ⌘ Data coordination can minimize duplication, reduce long-term costs, and streamline analysis and decision making for non-federal regional “customers”



Towards info aggregation and “unofficial” communication channels

- ⌞ During 9/11, the websites of mainstream news agencies were **flooded** by requests, forcing some to serve "low-resolution" versions of their webpages, while some effectively suffered a Distributed Denial of Service effect, rendering them unusable. Slashdot, on the other hand, a **communal weblog** aimed at technical users, and used to massive concurrent requests, **weathered the flood and provided timely updates**. During the same crisis, CNN used an **IRC channel** to transmit continuous coverage of the crisis.
- ⌞ The tsunami catastrophe of 12/2004 in Southeast Asia saw the, practically immediate, rise of a global community of **bloggers**, who collaborated via the **SEA-EAT blog** (TsunamiHelp), its associated **wiki**, and a **mailing list for contributors**. Updated constantly and being extremely **comprehensive** and fully **searchable**, SEA-EAT was dedicated entirely to providing news and information about **resources, aid, donations** and **volunteer efforts** related to the Tsunami disaster. Additionally, the Flickr.r folksonomy for **image sharing** was used to help with identifying missing persons.
- ⌞ All forms of **social technologies** contributed to disseminating information, during the recent London bombings. **Blogs** (personal, as well as communal, such as the Londonist) provided **constant coverage and perspective**, and were themselves **aggregated** by sites like Technorati; Wikinews provided first-hand reporting and constant news updates; Flickr.r was used extensively as a **photoreporting** tool, while news agencies and bloggers created maps of the bombing sites using the **Google Maps** and Google Earth technology.

TELECENTER FOR POST-TSUNAMI RECONSTRUCTION

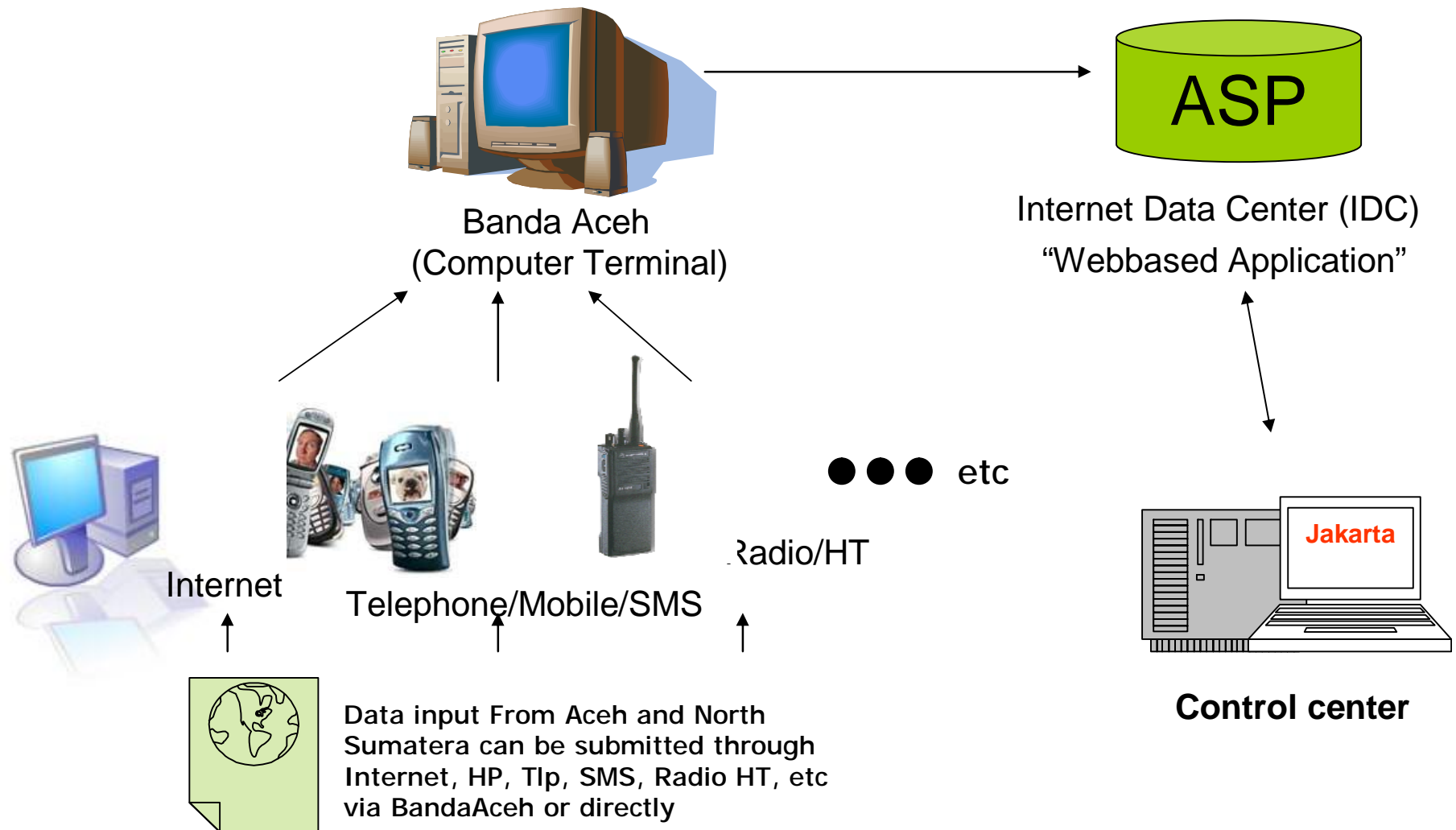


Program of APEC Telecenter Training Camp
Taipei, Taiwan
January 25th, 2005

Boni Pudjianto
Muslimin Kulle

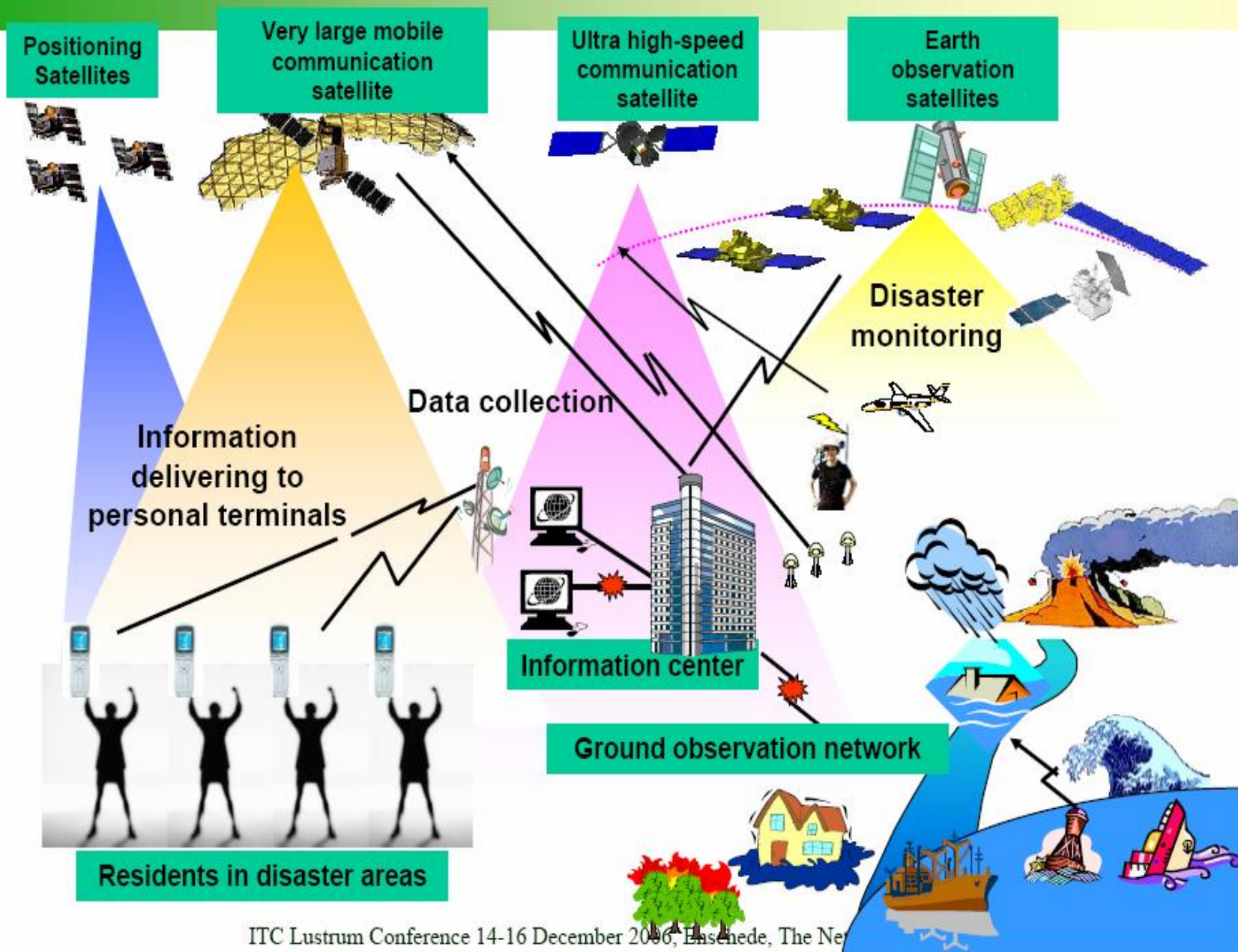
Ministry of Communication and Information
The Republic of Indonesia

Data Input Mechanism





Information Gathering and Warning System for Disaster and Crisis Management (JAXA Vision)



On this basis: Sketching of a possible solution

- ρ Information **hub**
- ρ Very easy to **set up**
- ρ Uses **proven** technologies
- ρ **Integrated** Content Management
- ρ Provides **easy connection** to forecasting models*

Information hub?

- ⌘ Opens a two-way communication (to/from public)
- ⌘ Integrated CMS: news, announcements, status
- ⌘ Aggregates information from multiple sources to a single public-facing site
- ⌘ Provides for a way to keep both public and emergency personnel informed (blogs, wikies, IRC feeds, photoblogs, etc)
- ⌘ *Allows for information management and rumor control*

How easy to set up? (1/2)

- ⌘ Should be as self-contained as possible
- ⌘ One single bootable LiveCD might be sufficient
- ⌘ Initial setup should not take more than 15-20 minutes by untrained staff.
- ⌘ Automatically set-up communication with predefined information resources.

How Easy to set up ? (2/2)

- ⌘ Incorporating new text input channels should not take much resources, expecting 10-15 minutes per ad-hoc source, and 30-40 minutes for data channels.
- ⌘ Prediction models:
 - ⌘ Provide interface (web services related) for framing the set-up, compilation and application of the model?
 - ⌘ Treated as web services that “broadcasts” information?
 - ⌘ can be fed from data channels directly, setup time ~15 min!

Information retrieval

- ⌘ Multiple, heterogeneous sources
 - ⌘ News sites
 - ⌘ Blogs
 - ⌘ News aggregators
 - ⌘ Government portals
 - ⌘ other
- ⌘ Sources that weren't designed with machine-readability in mind
 - ⌘ Screen scraping.
 - ⌘ both data and text/news
- ⌘ Dynamic source list
- ⌘ Multi-format

Information Processing

⌘ Aggregation

- ⌘ Collects information from many sites
- ⌘ Understands RSS
- ⌘ Has an HTML screen-scraping module
- ⌘ Can parse PDF files

⌘ Sorting

- ⌘ by priority, by keyword, etc

⌘ Filtering

- ⌘ by values, by keywords, by content, etc

⌘ Prediction model runs

⌘ Maybe custom processing

Information Presentation

- As news

- also machine-readable formats (RSS)

- As Maps

- possibly overlaid with data from heterogeneous sources

- As Graphs

- As notifications

- SMS, if applicable

Technologies

- ρ **Linux**
- ρ **Apache Tomcat**
- ρ **Java**
- ρ **Web-based interfaces**

Already tested

- ⌘ Platform uses Java and FOSS frameworks
 - ⌘ Implemented as a Servlet
 - ⌘ Remote management, administration over the web
 - ⌘ Uses Apache Turbine as the framework
 - ⌘ Provides scheduling services, templating, access control, database abstraction using Apache Torque, mail templates
- ⌘ Screen-scraping using XQuery
 - ⌘ Optional pass using JTidy to convert to XML DOM
 - ⌘ XQuery engine uses Saxon v8

Scenarios

- n Industrial accidents
- n Natural disasters
- n Terror attacks
- n Chemical transport accidents, flight and train accidents
- n as well as combinations in multi-crisis scenarios





On this basis....

Basic functions

- ⌘ Information aggregation
 - ⌘ collection of information
 - ⌘ location and the nature of the crisis
 - ⌘ type of crisis/disaster.
 - ⌘ cut across rigid information categories
 - ⌘ make information available via the web
- ⌘ Automatic categorization service
 - ⌘ Search engine technology
 - ⌘ hierarchical category classification system
 - ⌘ web crawlers
 - ⌘ find information from specific, predefined, sources (updated by domain experts)
- ⌘ users are notified on the basis of predefined preferences (profiling).
 - ⌘ Notification system sends email or SMS* when new articles match
 - ⌘ Configurable notification periods
 - ⌘ Multilingual
 - ⌘ Can separate official from unofficial sources (like blogs)

Matches a new entry against the ontology, multiple matches are possible, relevance is computed for each match, multilingual. Future goal: improve system by automatic evaluation of user feedback

Basic functions & operations (2/3)

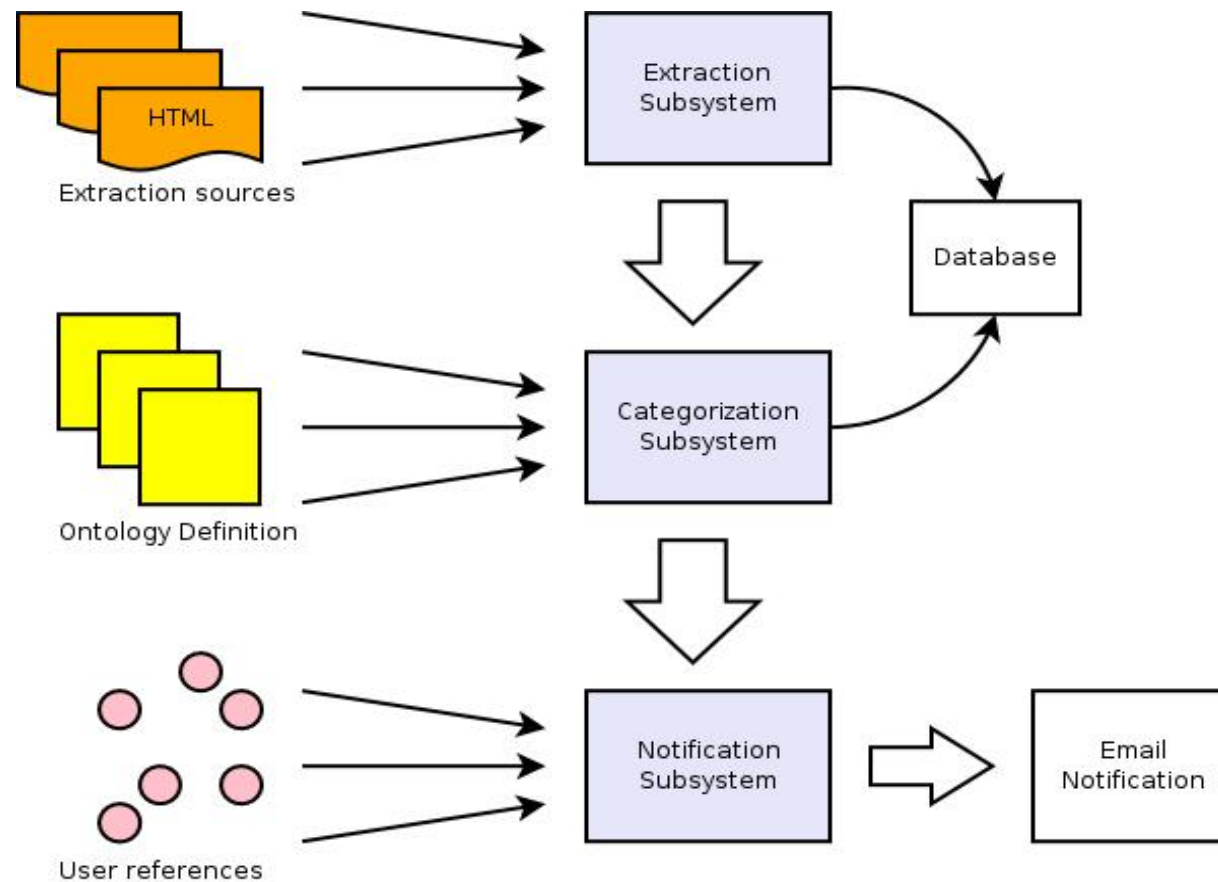
- ⌘ Information sources: n
- ⌘ Aggregation by way
 - n Templates
 - n typical RSS feeds
 - n other
- ⌘ emergency/crisis categories
 - n populated with relevant information for the specific emergency (type)
 - n relevant ontology?
 - n templates to be hand-coded for each individual non-structured information source (i.e. not offering RSS feeds).
 - ⌘ ***This means that, among other things, a large number of information resources will have to be “screened”, and then “encoded” into the application in advance. Consequence: necessity for a verification authority***

Domain experts define a ontology with the concepts of the problem domain . Definition contains rules that allow automatic categorization to the ontology

Basic functions & operations (3/3)

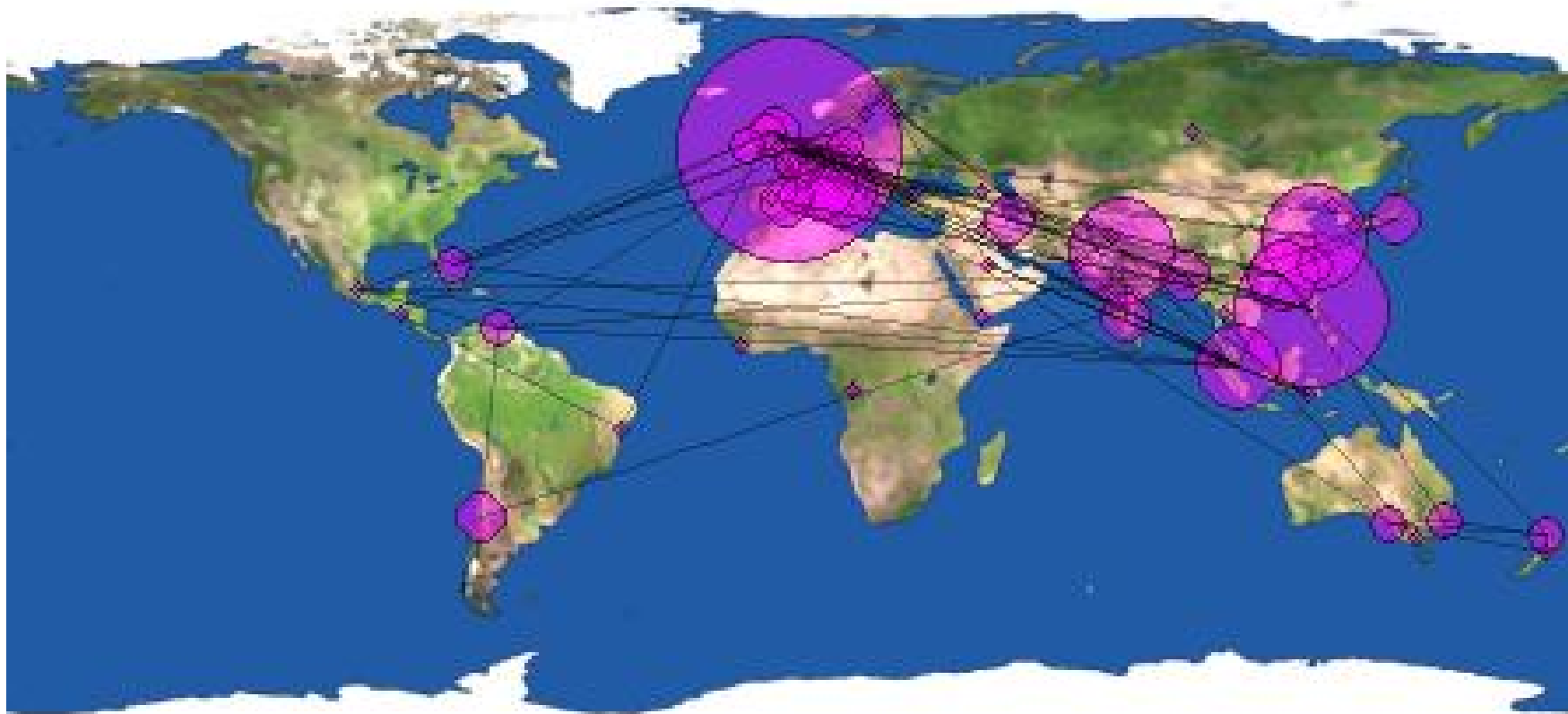
- ⌘ extraction subsystem will be responsible for extracting data and metadata from information sources,
 - ⌘ HTML pages,
 - ⌘ XML pages,
 - ⌘ processed data in the form of database sources.
- ⌘ extraction pipeline,
 - ⌘ feeds data to the categorization system,
 - ⌘ pre-defined user references,
 - ⌘ pass information to the notifications subsystem,
 - ⌘ where notifications will be generated and forwarded to the relevant information channels for delivery.
- ⌘ similarity index to be defined and developed, that would allow for extraction results to be sorted in terms of relevance to certain keywords.
- ⌘ The operator will be able to define the area of interest and thus the systems may “fetch” information like maps of the location and basic description on the nature of the event/incident. All this work will be conducted under the Communication channels / internet (RT2C).

Modular overview



Geographical mapping

- At a glance events on a global scale: Looking for city names in information sources



Some conclusions

Self-deployed, web-based information aggregators for disaster-related information collection and broadcasting are feasible for today's technology

Environmental information aggregators may support both horizontal and vertical information categorization

Alternative information management and dissemination means may provide important support under crisis

Future work: implementation of a pilot system!



Thank you!