# **AI4Good Hackathon, 2023**

[AI4Good Hackathon, 2023 1](#_Toc1069177820)

[About Red Cross 1](#_Toc2146612795)

[Problem statement 2](#_Toc1718575169)

[Title 3](#_Toc524941018)

[Context 3](#_Toc507513643)

[Broad Asks 4](#_Toc1352378018)

[Questions & Details 5](#_Toc1397297125)

[Data Section 7](#_Toc1069824735)

[Tools & Technologies 8](#_Toc838713155)

[Building Product 9](#_Toc283529656)

[Evaluation of the Product 10](#_Toc562142480)

[References 11](#_Toc141005439)

# About Red Cross

The International Federation of Red Cross and Red Crescent Societies (IFRC) is the world’s largest humanitarian network. The secretariat supports local Red Cross and Red Crescent action in more than 191 countries, bringing together more than 16 million volunteers for the good of humanity.

[About the IFRC | IFRC](https://www.ifrc.org/who-we-are/about-ifrc)

**Preparing for disasters saves countless lives, speeds up people’s recovery and saves money. The IFRC supports National Societies to continually improve their local preparedness and response capacity—ultimately preventing and reducing the impacts of disasters on communities.**

The need to prepare for a world of unexpected shocks has become clearer than ever. Epidemics, floods, storms, droughts and wildfires are all expected to become more frequent and severe, affecting hundreds of millions of people every year.

Red Cross and Red Crescent Societies around the world, supported by the IFRC, are working to improve their preparedness for disasters. This includes:

* Coordinating with national authorities & partners to know what risks and hazards to prepare for
* Training and equipping millions of volunteers as first responders to a [wide range of hazards](https://www.ifrc.org/what-disaster)
* Researching new technologies to improve their response
* Working with communities to understand the needs of those most at risk
* Setting up [early warning systems](https://www.ifrc.org/early-warning-early-action) so communities can take early action before a disaster hits

# Problem statement

Key areas: Natural Language processing, Machine learning, Dashboarding, Document processing and data extraction, Data engineering and visualization

### Title

How might we use & visualize past crisis data to better forecast our future response needs, to provide aid more quickly & efficiently when disaster strikes.

Context

* The [Disaster Relief Emergency Fund (DREF)](https://www.ifrc.org/happening-now/emergency-appeals/disaster-response-emergency-fund-dref) is a funding tool for small and medium disasters, that allow Red Cross & Red Crescent National Societies to obtain grants for the implementation of Emergency responses within their territories.
* The [Emergency Appeal (EA)](https://www.ifrc.org/emergencies/all) is a funding mechanism for large and complex disasters and crises that affect many people who will need long-term support to recover. These emergencies get donor specific grant funds and are typically emergencies that get media attention.

Recently the processes for a National Society to apply to the DREF has been added to [IFRC GO](https://go.ifrc.org/), structuring the data and information collected, enabling National Societies to access funds more efficiently, therefore providing aid more quickly to those struck by disaster. However historical data is hidden in previous narrative reports, inhibiting trend analysis and identification of recurrent crises.

Presenting this data will enable analysis over time for small to medium disasters, assisting in future strategic planning, such as identifying where funds have been allocated over time (to admin 1 or admin 2 levels), and how many people have been targeted and supported, to identifying repeatability of applications to the fund, or where and what dates the fund was accessed.

There are disasters in the world. A subset of those count with response from Red Cross/Red Crescent societies in country. A subset of those count with international assistance provided by the IFRC network. The subset of disasters attended by Red Cross/Red Crescent that count with international assistance provided by the IFRC network, is divided into two kinds: big emergencies ("Emergency Appeals"), and small emergencies ("DREF").

* For both kinds, there are reporting procedures. After the first days after the event/emergency, there is an initial report. Then, as the months go by, there are operational updates, and at the end of the response there is a "final report".
* Those final reports are the ones that we are interested in exploring with this exercise. The "emergency appeals" and the "DREF" processes are carried out by different teams, reflecting slightly different information needs, so it would be expected that there are differences between those two types of final reports.
* Also, although there are templates for the final reports, they DO change with time, and are adapted by the different regions/countries, so also there should be expected some variability in those terms.

### Broad Asks

Create one-stop solution to extract & derive insights from the appeal & appeal documents, thus building disaster monitoring & early warning system

1. Develop a **Natural Language Processing (NLP) pipeline, to extract and process key information** (such as location, date, hazard, targeted population, people supported etc.) from specific crisis reports (DREF and Emergency Appeals) via the IFRC GO API.
   1. Some of the mentioned attributes are already extracted from the “Appeal” document. <https://goadmin.ifrc.org/api/v2/appeal/>

Please hit to check the attributes. This can be readily used to construct structured data on demographics, funds asked, funds fulfilled, nature of the event, location, etc

* 1. **Participants should perform “*extraction of the operational strategy”***from various appeal documents, organize and present the information to drive consumable insights.

(refer Questions 2 down below for more details)

<https://goadmin.ifrc.org/api/v2/appeal_document/>

Participants will search for the term like **“Final report”** in the field **“name”**, case insensitive, in the document fields while hitting this API endpoint. This final report contains information on operational strategy (Section Questions: 2)

* 1. Programmatically download the appeal documents, read the pdf documents, extract and standardize information about operational strategy – usually tables or plain paragraphs of activities planned in different sectors such as Nutrition, Health, Shelter, etc . Few sample templates of appeals have been uploaded in ./Appeal Sample folder for your reference

1. Develop a **methodology to compute accuracy on the extraction process**, and to verify the extraction of information
   1. Participants should devise a strategy to demonstrate effectiveness of information extraction process
   2. Data thus extracted using smart & efficient NLP pipeline may be used by Red Cross Go. Should think about how to standardize information from various types of appeal documents
2. Develop a **tool to visualise the extracted and processed information** e.g., people supported by which sector, displaying which countries request funding most often, or which countries or regions request funding and when
   1. You could use “Appeal” (API) to get the various field related to the disaster, can be used create visualizations such as demographics, funds asked, funds fulfilled, nature of the event, location, etc
   2. Participants should also present and organise collected/extracted information from all appeals document (from point 1.2)
   3. Additionally, participants may supplement the information with additional mentioned external data. For e.g how was the country doing at the time of disaster, plot various indicators for the timeline
   4. It is slightly challenging to overlay GDELT or FEMA data, as disaster in the appeals may be relatively mid-sized to be reported in event registry platform. However, participants may use GDELT data to filter through socio-economic news occurring during the time of the event. Exploring, downloading and Integrating GDELT data is up to the participants; as this is a huge, curated datasets, one would need to filter datasets with some relevant criteria (Use of GDELT/FEMA dataset is good to have, but not mandatory)
3. **Pair your analysis** with s**upplementary data** provided (Please refer to section Building Product)
   1. UN and Governance indicators – easier to accommodate and swift through than working with GDELT/FEMA datasets. Refer 3.d above.

Please refer to **Data & Building Product** section for more details.

The output extracted data from the appeal documents can then be published on Red Cross Go platform and will be made available globally for all Red Cross/ Red Crescent' societies (future integration, requirement by Red Cross).

Blog:

* <https://ifrcgoproject.medium.com/digitalising-disaster-response-aa39c9e0be5>

### Questions & Details

**1. Is there a ready to go repository of documents and data sets that we can access, or do we need to crawl the API links?**

This is the API endpoint to retrieve the pdf documents We are **interested in the documents that contains 'final report' in their "name" attribute**, <https://goadmin.ifrc.org/api/v2/appeal_document/>

This is the API endpoint to find the country, type of disaster, date and budget related to the pdf documents, <https://goadmin.ifrc.org/api/v2/appeal/>

The "code" attribute in the “appeal” endpoint links to the "appeal" attribute in the “appeal\_document” endpoint.

<https://goadmin.ifrc.org/api/v2/appeal/>

Every appeal goes by aid in the results returned by the API Endpoint

{"aid": "17927",  
 "name": "India - Flood",  
 "dtype": {  
 "id": 12,  
 "name": "Flood",  
 "summary": "",  
 "translation\_module\_original\_language": "en"  
 },  
 "atype": 0,  
 "atype\_display": "DREF",  
 "status": 0,  
 "status\_display": "Active",  
 "code": "MDRIN028",  
 "sector": "New Delhi country cluster",  
 "num\_beneficiaries": 40000,  
 "amount\_requested": "539063.00",  
 "amount\_funded": "539063.00",  
 "start\_date": "2023-08-31T00:00:00Z",  
 "end\_date": "2024-01-31T00:00:00Z",  
 "real\_data\_update": "2023-09-01 07:10:01+00:00",  
 "created\_at": "2023-09-01 08:48:12.435928+00:00",  
 "modified\_at": "2023-09-06 06:51:03.950293+00:00",  
 "event": 6619,  
 "needs\_confirmation": false,

…..}

<https://goadmin.ifrc.org/api/v2/appeal_document/>

{  
 "created\_at": "2011-01-27T00:00:00Z",  
 "document": null,  
 "document\_url": "[https://go-api.ifrc.org/publicfile/download?path=/docs/appeals/09/&name=MDR61005fr.pdf](https://go-api.ifrc.org/publicfile/download?path=%2Fdocs%2Fappeals%2F09%2F&name=MDR61005fr.pdf)",  
 "appeal": "MDR61005",  
 "iso": null,  
 "description": null,  
 "id": 3424,  
 "name": "Final Report",  
 "translation\_module\_original\_language": "en"  
 }

Please refer to sample code in the folder to crawling API endpoint at ./API Appeal download Recipe.txt

Data API Links:

* <https://goadmin.ifrc.org/api/v2/appeal_document/>
* <https://goadmin.ifrc.org/api/v2/appeal/>

**2. Within the long list of data/information (attributes), are you able to point us to specific data you would expect/need us to extract so we can sharpen the focus of our participants: e.g., locational, date, type of hazard/disaster, targeted population, people supported etc.]?**

We had [this project](https://students.cs.ucl.ac.uk/2022/group3/index.html) with computer science students extract the location (admin1) of the documents. We would like to have by one part another iteration on this work (extraction of location - ideally going to admin 2, and providing an indicator of the certainty of the extraction)

<https://students.cs.ucl.ac.uk/2022/group3/index.html> should also give you sample extraction strategy previously developed.

***By another part the extraction of the operational strategy (what activities/distributions were planned vs done in each one of the sectors - they are usually presented as tables in the reports). Those are the main priorities. Examples of different sectors are Nutrition, Health, Shelter, etc. Reading the appeal document will give an idea about planned activities, effectiveness, impact created. Standardized Information thus collated from many many appeal documents (mapped back to appeal) will create a picture of how effectively Red Cross can address the appeal in the future and help study the impact and gap in the planned activities. This also helps in effective fund planning, appropriation and allocation***

Attached sample Excel document which shows a current structure, however additional columns would be required with the activities/distributions were planned vs done in each one of the sectors. (Please refer to data folder for the sample appeal document)

**3. Can you open up IFRC data portal for other relevant data?**

Participants could triangulate information with other sources ([GLIDE](https://glidenumber.net/), [ReliefWeb](https://reliefweb.int/), [EM-DAT](https://www.emdat.be/), [GDELT](https://www.gdeltproject.org/)), but it is important that it gets well documented in which points those sources were considered and how.

**Note:** Participants should refer to Data Section (Firm has obtained legal approval to the mentioned datasets)

Data Section

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **Data** | **Nature** | **Links** |
| **Red Cross Data** | **IRFC (International Federation of Red Cross and Red Crescent National societies)** | **Documents API**  **(Access sample will be shared)** | [**API to access “Appeal” document**](https://goadmin.ifrc.org/api/v2/appeal_document/)  **("name": "Final Report")** |
| [**API to access “Appeal” fields**](https://goadmin.ifrc.org/api/v2/appeal/) **(ready extracted attributes)**    **"code" attribute in the appeal = "appeal" attribute in the appeal document** |
| **Supplementary data** | | | |
| **Socio-economical** | [**UN data**](http://data.un.org/Explorer.aspx) | **Structured** | [**http://data.un.org/Explorer.aspx**](http://data.un.org/Explorer.aspx) |
| [**Governance Indicators**](http://info.worldbank.org/governance/wgi/Home/Reports) | **Structured** | [**http://info.worldbank.org/governance/wgi/Home/Reports**](http://info.worldbank.org/governance/wgi/Home/Reports) |
| **Event Registry** | **GDELT event data** | **Numerical + semi-structured text** | [**https://www.gdeltproject.org/about.html#termsofuse**](https://www.gdeltproject.org/about.html) |
| **Open FEMA (US)** | **API access** | [**https://www.fema.gov/about/reports-and-data/openfema**](https://www.fema.gov/about/reports-and-data/openfema) |

Note: ICEWS data has been removed – as this dataset can only be used for education and research purposes

Tools & Technologies

|  |  |
| --- | --- |
| **Tools & Technologies** | |
| **Development Environment - Google Collab** | [**Link**](https://colab.research.google.com/) |
| **Offline Development Environment - Anaconda (Python, JupyterLab, Spyder)** | [**Link**](https://www.anaconda.com/download) |
| **Visualization tool (Power BI Desktop or any other)** | [**Link**](https://powerbi.microsoft.com/en-in/downloads/) |
| **Personal Laptops Only (can be used in Office) | Collaboration on Wifi@Work, Setup Instructions are here** [**Link**](https://mysourcesocial.bnymellon.net/groups/wireless-network-access) | |
| **Microsoft Hackbox – Community Engagement & Global exposure to other Initiatives** | |

# Building Product

**What is expected?**

* Intelligent Document Processing for Key Information from Crisis reports with output data table aligned with existing structure with added attributes where relevant.
* App/Dashboard for crisis monitoring and early warning system
* Use raw materials & create your own version of system

Below are some recipes that you can use (you can build something completely new). **Recipes mentioned here are for your reference to get kickstarted**. Teams with the help of product managers are **free to explore their own idea of the product**, provided it covers basic functionality.

|  |  |  |  |
| --- | --- | --- | --- |
| Level | Nature |  | Sample Recipes |
| **Basic (minimum delivery)** | **Building document Intelligence**  **(**[**reference**](https://students.cs.ucl.ac.uk/2022/group3/index.html)**)** | Extract and Process Key information from crisis reports – appeal reports  visualize the extracted and processed information | * Information extraction and document processing * Global trends – bubble charts, heat world maps * Funding appropriation, fund allocation * People supported, and where (if possible disaggregated by sex/age/disability) * Flash crisis images (web download) * Word cloud – what are people talking about? * ….. |
| **Advanced (crossing benchmark)** | **Paired Exploration** | Overlay Supplementary data to enhance information | * Verification of extracted data/information * Sentiment & tone time graphs * Momentum indicators – spread of the event * Time series propagation of articles, news events * Flash images with annotations of impact (web download) * …… |
| **Paired exploration** | Overlay Supplementary data to enhance information (advanced) | * Overlay and comparisons with UN socio-economic data and areas reported to have an event occur * Clustering appeals on extracted information “Target”/“Done” * Flash related news popups & govt instructions/updates * **…..** |
| **Expert (Phenomenal)** | **Predictive**  **/Prescriptive** | Studying aftermath of the disaster | * Detecting and signaling if the event is going to become a major one; studying aftermath, and its impact * …. |

# Evaluation of the Product

|  |  |  |
| --- | --- | --- |
| Product 50% | Idea 25% | Demonstration 25% |
| Capability, Intuitiveness, Usability, Extendibility | Value proposition, Novelty, Maturity & Potential | Product snapshots, Paper (write-up) |

\*Evaluated by product and technical reviewers

Please do not forget to capture your journey to the final product with some awesome team photos!

**Filing next week, post hackathon:**

Paper describing component design, data sources & how they were used, analysis and insights, information extraction and methodology to evaluate the extraction process, final product snapshots, team collaboration, tool /tech stack used, challenges and future scope, anything else you would like to add.

Handover to Red Cross: Codepack, Paper, Product demo to Red Cross

***Note to participants***: We thoroughly understand it’s very challenging to develop a product in 1-3 days that can be absorbed by the Red Cross **readily**. At present, we are looking for **great ideas and functional products** surrounding reading appeal & appeal documents, extracting information, overlaying the supplementary data, finally visualizing information to derive insights from these appeals.

# References

Red Cross, [Homepage | IFRC](https://www.ifrc.org/)

Past work by students, [Team 3 (ucl.ac.uk)](https://students.cs.ucl.ac.uk/2022/group3/index.html)

API by Red Cross: <https://goadmin.ifrc.org/api/v2/appeal_document/> , [htttps://goadmin.ifrc.org/api/v2/appeal/](https://goadmin.ifrc.org/api/v2/appeal/)

[Downloading PDFs with Python using Requests and BeautifulSoup - GeeksforGeeks](https://www.geeksforgeeks.org/downloading-pdfs-with-python-using-requests-and-beautifulsoup/)

[HackBox (microsoft.com)](https://hackbox.microsoft.com/), community building platform

<https://ifrcgoproject.medium.com/digitalising-disaster-response-aa39c9e0be5>