A PAGES Floods WG core project:

Collaborative Flood Database for Multiple Archive Types

Albert-Ludwigs-Universität Freiburg

11-05-2017
OSM Pages, Zaragoza

Michael Kahle (1), Rüdiger Glaser (1), Pierre Francus (2), Pages Flood WG (3)
University Freiburg (1), Institut National de la Recherche Scientifique (2),
http://www.pages-igbp.org/ini/wg/floods/people (3)
Goal

- From the Floods WG page:
  - “Compile published data sets on floods for open-access archiving in order to facilitate the visibility of existing data and their inter-comparison”
  - (http://www.pages-igbp.org/ini/wg/floods/scientific-goals -> iii)

- Necessary: Common Data Structure across Multiple Proxy Types

If possible, use Sensor-Archive-Observation concept
- (Evans et al 2013: https://doi.org/10.1016/j.quascirev.2013.05.024)
Multiple Proxy Types

- Sediments
  - Lacustrine
  - Fluvial
  - Marine
- Speleothems
- Tree rings
- Historical documents
- Instrumental data
- ...

Source: http://www.pages-igbp.org/ini/wg/floods/intro
Meta Data Collection

- Lots of candidates for a common flood database across all types of archives, countries and time ranges
- Thanks to all contributors

Figure: Bruno Wilhelm, 2017
Minimal Data Structure

Clusters

ℹ️: Source
📍: Location
⏰: Time
💰: Classification
🔗: Reference

Figure: Michael Kahle 2017
Multiple Proxy Types

Differ by

- Time range covered (years back, seasons)
- Temporal resolution
- Location types covered
- Sensibility to floods (by extension or duration)
- Negative signals (droughts)
- Noise
- Additional information (causes, impacts, temperature, ...)

A PAGES Floods WG core project: Collaborative Flood Database for Multiple Archive Types
## Multiple Proxy Types

<table>
<thead>
<tr>
<th>Proxy</th>
<th>Location</th>
<th>Time range</th>
<th>Time resol.</th>
<th>Misc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sediments</td>
<td>River, Lake</td>
<td>10000s of years</td>
<td>1 year</td>
<td></td>
</tr>
<tr>
<td>Speleothems</td>
<td>Carst, Cave</td>
<td>1000s of years</td>
<td>1 year</td>
<td>Hydrol. Data, e.g. Discharge</td>
</tr>
<tr>
<td>Tree rings</td>
<td>Woods</td>
<td>500-1000 years</td>
<td>1 year, seasons</td>
<td>Precipitation</td>
</tr>
<tr>
<td>Historical documents</td>
<td>City, Village</td>
<td>500-1000 years</td>
<td>1 day, hour</td>
<td>Causes, Impacts</td>
</tr>
<tr>
<td>Instrumental measurements</td>
<td>Civilization</td>
<td>100s of years (1000s)</td>
<td>1 day, hour</td>
<td>Precise Magnitude</td>
</tr>
</tbody>
</table>
Sensor, Archive, Observation

Single archive approach (by Evans et al 2013):

- Environment
- Flood
- Temperature

Sensor → Archive → Observation → Conclusion

Tree → Wood → Ring Density → Flood Magnit.

Multiple archive approach:

- Precipitation
- Flood
- Temperature
- Nutrients

Precipitation → Cave → Stalagmite → Isotope ratios → Flood Magnit.

Tree → Wood → Ring width → Temperature

Person → Chronicle → Description → Flood Magnit.
Additional Data

- To use Sensor-Archive-Observation model more data is necessary
- Also documentation of whole process leading to flood data is desirable

- Example data was provided by members of the Floods WG
- ➔ More (partially optional) data fields are needed.
Common Data Structure

Clusters

ℹ️ : Source
📍 : Location
⏰ : Time
🧬 : Classification
🔗 : Reference

Figure: Michael Kahle 2017
File Format

- Human readable
  - Easy to create and edit

- Machine readable
  - Import and export to and from software tools

- Check for completeness and consistency
- Changes traceable
- ➔ Avoid binary format
File Format

https://xkcd.com/927/
Data Format: CSV

Mandatory fields are compact

<table>
<thead>
<tr>
<th>Institution</th>
<th>Proxy Type</th>
<th>Acquisition Date</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Resolution</th>
<th>Event Date</th>
<th>Uncertainty</th>
<th>Magnitude</th>
<th>URL Data</th>
<th>URL Publication DOI</th>
<th>Email Contact Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>University,</td>
<td>fluvial</td>
<td>2009-02-24</td>
<td>51,8619</td>
<td>6,1186</td>
<td>100</td>
<td>1926</td>
<td>180</td>
<td>3</td>
<td>w.h.j.toonen</td>
<td>Toonen, W.H.</td>
<td></td>
</tr>
</tbody>
</table>

BUT – optional data is huge

The table is vast and contains various fields such as:

- **Country Name**
- **Location**
- **Catchment**
- **Catchment Area**
- **Elevation**
- **Plain Type**
- **River Name**
- **Lake Name**
- **Lake Surface**
- **Max Depth**
- **Dating Method**
- **Dating Parameter**
- **Dating Unit**
- **Dating Value**
- **Dating Tolerance**
- **Sample Depth**
- **Zero Date**
- **Measure Measured**
- **Measured Parameter**
- **Measured Unit**
- **Measured Value**
- **Max Discharge**
- **Max Water Level Recurrence Year**
- **Cause**

Additional fields include:

- **Funding Agent**
- **LiPD Dataset**
- **URL Linked**
- **URI**
- **Probe**
- **UID Probe Std**
- **UID Contact**
- **UID Contact S**
- **Citation Short**

A PAGES Floods WG core project: Collaborative Flood Database for Multiple Archive Types
Data Format: LiPD

Mixture of json and csv files in zip

Advantage
- Covers: Sediments, Speleothems, Tree rings, Measurements on fixed points, …
- Well established file format with tools available
- Fine structured

BUT - Enhancements needed for
- Historical documents
- Standardize column names & Units

https://doi.org/10.5194/cp-12-1093-2016
N. P. McKay and J. Emile-Geay
## Proxy Archives vs Documents

<table>
<thead>
<tr>
<th>Proxy Archives</th>
<th>Historical Documents</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Probe on fixed location</td>
<td>One source often covers multiple locations, multiple observers</td>
</tr>
<tr>
<td>Mostly annually, periodically</td>
<td>Daily or even hourly resolution, randomly scattered samples</td>
</tr>
<tr>
<td>Time derived (counting, isotopes)</td>
<td>Time absolutely (calendar, clock)</td>
</tr>
<tr>
<td>Continuously recorded</td>
<td>Often only extremes recorded</td>
</tr>
<tr>
<td>May interfere with circumstances</td>
<td>Can distinguish circumstances</td>
</tr>
<tr>
<td>Causes and Impacts to guess</td>
<td>Causes &amp; Impacts often mentioned</td>
</tr>
</tbody>
</table>
LiPD – Documents Enhancements (1)

1st Possibility:
Extra table for
• Source data
• Location data (refines Geospatial)

Based on N. P. McKay and J. Emile-Geay
LiPD – Documents Enhancements (2)

2\textsuperscript{nd} Possibility:
Include tables into Proxy measurements
- Source data
- Location data

Based on N. P. McKay and J. Emile-Geay

Additional wish: Provide a version with json only (no csv files)
The final goal is to provide an online platform to store the flood data into hosted DB
Entries should be easy to add (e.g. by uploading csv or LiPD files)
Search function provided to search floods by archive type, time frame, location area and/or magnitude
Access data by downloading files or via API
Sustainable long-term availability desired
Advantage
- Historical Documents covered
- Database & Web Access running (operated by UB Freiburg)

Enhancement needed for
- Sediments
- Speleothems
- Tree rings
- Measurements on fixed points
Proxy Enhancements
Alternative Hosts

- Maybe other organizations or institutions can host the flood database
- Other Pages projects or WG may face the same issue
- LiPD on http://linked.earth would be very promising
- Long-term operation is crucial
Next steps

- Enhance LiPD to cover historical data as well
  - Add individual locations
  - Add individual sources and quotes
  - Standardize column names and units

- Agree on online platform and DB
  - Where to host
  - Find sustainable long-term strategy

- Agree on implementation
  - Who can join the work
  - How to finance development (Projects, etc.)
Vision

Search Events

- Query Text
  - Query Text

By Year

- 1812
- 1874

By Country

- None

By River

- None
  - Distance to River (km): km

Sort by

- None
  - Ascending

Institution | Proxy Type | Acquisition Date | Latitude | Longitude | Resolution | Event Date | Uncertainty | DOI | Email | Citation
--- | --- | --- | --- | --- | --- | --- | --- | --- | --- | ---
Uni Freiburg | Document | 2017-05-05 | 8,4578 | 42,1478 | 1 | 1852-11-23 | 1 | 3 | michael.kahle@geog |
Uni Freiburg | Tree rings | 2016-10-04 | 8,1234 | 43,457 | 2,5 | 1852 | 365 | 2 | michael.kahle@geog |
Uni Freiburg | Speleothems | 2017-01-14 | 7,9874 | 42,879 | 1,5 | 1847 | 365 | 3 | 10.6094/tambora.org/2016/c15
Uni Freiburg | Sediments | 2016-08-04 | 8,1234 | 42,573 | 2,5 | 1122 | 365 | 3 | michael.kahle@geog |

A PAGES Floods WG core project: Collaborative Flood Database for Multiple Archive Types
Thank you