Impact of Sediment Management Strategies on Deltas

Building a world-class delta protection measures database

O'DELL, J. (Joey)

Utrecht University, The Netherlands

ORCID® iD: 0000-0003-3953-8150

Figure 1 - Research Word Cloud

Table of Contents

1. Introduction 5
Aim 6
Objectives 6
2. Method 8
Data Management 8
Data Collection 12
Processing & Verification 15
3. Results 21
Data Dissemination 21
Data Assurance 28
4. Discussion 31
Future possibilities for automated processing 31
Data Platform Review 35
5. Conclusion 37
6. References 38
7. Appendices 40
Appendix I: Existing Databases Reviews 41
Appendix II: ESRI Academy Transcript 51
Appendix III: Manual Data Collection Flowchart 52
Appendix IV: Achievement of Research Objectives 53
Appendix V: Achievement of Educational Objectives 54
8. Copyright and Usage 55
Data Custodian 55

List of Figures

Figure 1 - Research Word Cloud 2
Figure 2 - Project Entity Relationship Diagram 9
Figure 3 - Hierarchy of Data Sources Flow Chart 13
Figure 4 - Verification of retrieved data against basemap (Ganges-Brahmaputra Delta, Bangladesh) 15
Figure 5 - Simple processing of levee line data (Gascoyne River, WA, Australia) 17
Figure 6 - Inference of levee feature using multiple data sources (Acheloos Delta, Greece) 18
Figure 7 - Example of interactive tooltip helptext in model 19
Figure 8 - Leveed Areas Processing Model 20
Figure 9 - Idealised distribution platform (single-copy design) 23
Figure 10 - Revised distribution platform 24
Figure 11 - Alternative URL redirection functionality within DataverseNL 24
Figure 12 - Example of interactive browsing experience in database (Ganges-Brahmaputra Delta, Bangladesh) 27
Figure 13 - Example of use of `NeedsReview` field 28
Figure 14 - Example of edu.nl link statistics 30
Figure 15 - Hypothetical processes for automated method 32
Figure 16 - Comparison of Australian hydrological data sources (Rockingham, WA, Australia) 34
Figure 17 - Submission form and notification workflow 36

1. Introduction

Human impact has ultimately altered the morphology of many deltas around the world, changed the risks of flooding and inundation, and affected the supply and removal of sediment from deltaic systems. Flooding has been, and continues to remain a substantial risk to those living in river deltas, and has a disproportionate impact compared to their non-deltaic counterparts due to a convergence of population density, environmental conditions, and low elevation (Edmonds, Caldwell, Brondizio, & Siani, 2020).

Methods to protect against flooding – including the building of levees - have been deployed in deltaic areas, but the impact of individual measures and larger levee systems has been poorly modelled. To investigate and model the impact of the presence or absence of such features in the future, it was necessary to collect information as to what methods exist, and where they have been implemented.

Levees are engineered solutions to control flooding and sedimentation in a wide variety of situations including deltas, generally taking the form of an elevated barrier constructed at the side of the river or the flood plain from various materials, ranging from earthen mounds to concrete and rebar cast sections. The implementation of this measure provides an increase to the cross-sectional area of the river, allowing greater volumetric capacity, and the resultant change to velocity is used to protect urban areas and farmland from flooding events by allowing greater water discharge. Sedimentation is also affected, as overbank deposits are inhibited by levee structures, and the adaptation of the river and delta system is then subsequently altered.

Despite levees being used for thousands of years, and first documented over 3,000 years ago in ancient Egypt, it has been unknown to exactly what extent flood-protection levees had been constructed globally, both in deltas and in general fluvial channels, as the mapping and documentation processes has varied greatly depending on the purpose and objectives of the organisation building them. This has in turn resulted in a patchy release of information on a country-to-country basis and has even varied substantially within regions of the same country.

Popular modelling and projection systems, including the Water Resources Institute AQUEDUCT flood application, and the PCR-GLOBWB 2.0 water balance modelling tool disregard levee features altogether, or use proxies and estimations that do not allow proper adaptation to flood risk. This limited a significant number of modern scientific studies, including those that provide global RSLR risk assessment; even modern modelling scenarios have been forced to assume the absence of flood protection due to the lack of an accurate global data set (Scussolini, et al., 2016; Eliander, et al., 2020).

Furthermore, decisions and policies based upon these outcomes are severely impacted, as they historically were unable to be 'ground-truthed' with levee data and realistic predictions. This has introduced a potential to cause both over and under investment in flood defence activities, which ultimately could have put lives at risk.

Various sources of flood-protection levee data have already been in existence, but these have been distributed across modalities, and did not share a common reference frame. Individual dataset curators have defined their own attributes, terminology, and coding schema. Altogether this has made simple comparison across even the same country difficult in some instances, and across national boundaries near impossible altogether.

Aim

The aim of the project was to build a common global dataset of flood protection measures, which could be used both by the scientific community and the wider delta stakeholders. This could then be cross linked to delta physics data, flooding model data, and other geospatial datasets to model and understand the changing deltaic environments in the future. As the dataset was a first-of-kind, there was no intention to make it a static entity, and instead there has been, and will continue to be the opportunity for the consumers of the data to contribute and better the dataset, as well as distributing it freely.

A recent paper by Özer, van Damme, & Jonkman (2020) ratifies this; through collating information on 18 existing levee and dam databases, Özer *et. al.* identified that 12 were not actively maintained, and only 3 were of actual levee features (compared to failure/flood data.). Most of the databases were also not publicly available, and this added further barriers to research.

Historic practices of keeping data under "lock-and-key" are prohibitive to progressive research in the field, and additionally are not in keeping with open-data principles that are increasingly common across the developed and developing worlds. By assimilating this data into a single database, and publishing it publicly online, this will remove the barrier to future research of delta protection and sedimentation. A key example is that climatic models that were previously run with assumed levee characteristics, or that negated levee features entirely, will be able to use the data to provide more realistic results.

Objectives

The project served not only to produce a high-quality research output in the form of a levee database, but additionally was designed to equip the author with skills and capabilities that could be applicable both within academia and in the wider world. These skills have been laid out clearly below, and are reviewed in Appendix IV and V.

Table 1 - Research and Educational Objectives

Research Objective	**Educational Objective**
Combine and harmonise existing national levee databases into a global database for deltas	Develop and refine data selection and processing skills to be able to accurately, and reliably, assess data for inclusion or exclusion.
Develop an algorithm and data standard for processing the flood protection data that can be applied to other deltas and used by others in the future	Learn industry-standard coding and processing methodologies to (partially-)automate data gathering and collation processes.
Contribute to the wider scientific community by making the database publicly available, to benefit future deltaic research	Understand the state-of-play for data dissemination systems, (meta-)data standards and platforms, and data best practice.

Table 2 - Comparison of existing delta & levee databases

Dataset Name	Year Published	Last Update	Publishing Organisation	Spatial Extent	Data Type	Release status	Update frequency	Licence	Comments
FLOPROS Database of FLOod PROtection Standards	2016	2020	Academic	Global	Theoretical Polygon	Public	Irregular, but ongoing	CC-BY 4.0*	Information on flood protection 'standards' for modelling, but no data on actual levee features
Delta Polygons	2019	2020	Academic	Global	Delta Polygon	Not yet released	Unknown	Unknown	Delta area according to a defined methodology, contains no levee data
Global Delta Change	2019	2019	Academic	Global	Watershed Polygon	Public	Irregular, but ongoing	CC-BY 4.0*	Watershed dataset containing modelling attributes
SAFElevee/ ILPD International Levee Performance Database	2019	2020	Academic	Global (incomplete)	Point (event location)	Public	No longer actively maintained	No database right asserted	Single point events detailing levee failures, testing events, and so forth. Includes historical data.
DANTE Database nazionale della AgriNature in TErra	2015	2020	Academic/ Governmental	Italy (incomplete)	Point (levee start/end)	Private	Regular, ongoing	Local	Levee monitoring and management system including start and end point data of levee feature, design flood data etc.
DWER-021 Department of Water and Environmental Regulation	2018	2018	Governmental	Australia (Western Australia (WA) only)	Levee Line	Public	No longer actively maintained	Local, compatible with open data	Basic levee feature line dataset, does not list construction attributes
Levee Banks	2013	2020	Governmental	Australia (South Australia (SA) only)	Levee Line	Public	Regular, ongoing	CC-BY 4.0*	Basic levee feature line dataset, does not list construction attributes
DLS Line Distinctive Land Surface	2020	2020	Governmental	Australia (New South Wales (NSW) only)	Levee Line	Public	Regular, ongoing	CC-BY 4.0*	Basic levee feature line dataset, does not list construction attributes
Coastal Levees	2011	2014	Governmental	Australia (Victoria (VIC) only)	Levee Line	Public	No longer actively maintained	CC-BY 4.0*	Basic levee feature line dataset, does not list construction attributes
EA AIMS Environment Agency Asset Information Management System	2019	2020	Governmental	United Kingdom (England only)	Levee Line	Public	Regular, ongoing	OGL 3.0 ✠, open data compatible	Comprehensive national dataset containing levees, coastal defences, and flood walls. Does not include area. Some information on construction attributes included.
EUROSION	2005	2016	Governmental	Continental Europe coastline (as at 2004)	Coastal Interface Line	Public	No longer actively maintained	CC-BY 4.0*	Coastal protection activities recorded at coastal interface line, limited to coastal data and from the EU as of 2004, no data on construction attributes
National Levee Database	2007	2020	Governmental	Mainland United States of America	Levee Line and Leveed Area	Public	Regular, ongoing	US Public Domain	Comprehensive national dataset containing levees and the area they protect, including data on construction attributes

* Creative Commons (CC) Attribution (BY) 4.0 Licence: https://creativecommons.org/licenses/by/4.0/
✠ UK Open Government Licence (OGL) 3.0, directly compatible with CC-BY 4.0: http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/

For further details, including citations of the datasets, source URLs, and the data behind this table, please see Appendix II

2. Method

Data Management

PLANNING

The data included in the database has been carefully curated both in terms of spatial extent and visualisation (directly visible to the user) but also in what was recorded in metadata and indexes (data that can be read by the user or the modelling system). The thorough curation and documentation process ensured that results remain repeatable and verifiable.

Before starting to enter data into a database the fields and tables were clearly defined – this reduced in-process modifications which could have affected data integrity, and also ensured that the data structure was documented and could be referred to during the project and in the future. Setting clear guidelines also ensured that future additions to the dataset are consistent and match the format already defined, which enabled easier future maintenance and reuse.

The database contained both self-created elements (the index and layer filers) and existing data from other sources, therefore by mapping out the relationship between these entities in an Entity-Relationship Diagram (ERD) unique field names were be clearly defined from the outset. This reduced the introduction of secondary, confusing field names (aliases) and allowed for a cleaner user interface as there was no need for repeated explanatory text to link aliases with field names.

Creating an ERD (see Figure 2, p.9) was the first step, however the ERD was reviewed at regular periods throughout the project to ensure it remained valid. The ERD was not published externally as part of the data package, but the key linking elements (e.g. use of primary/foreign keys to associate data without copying variables from other datasets) were noted in the metadata and the README.

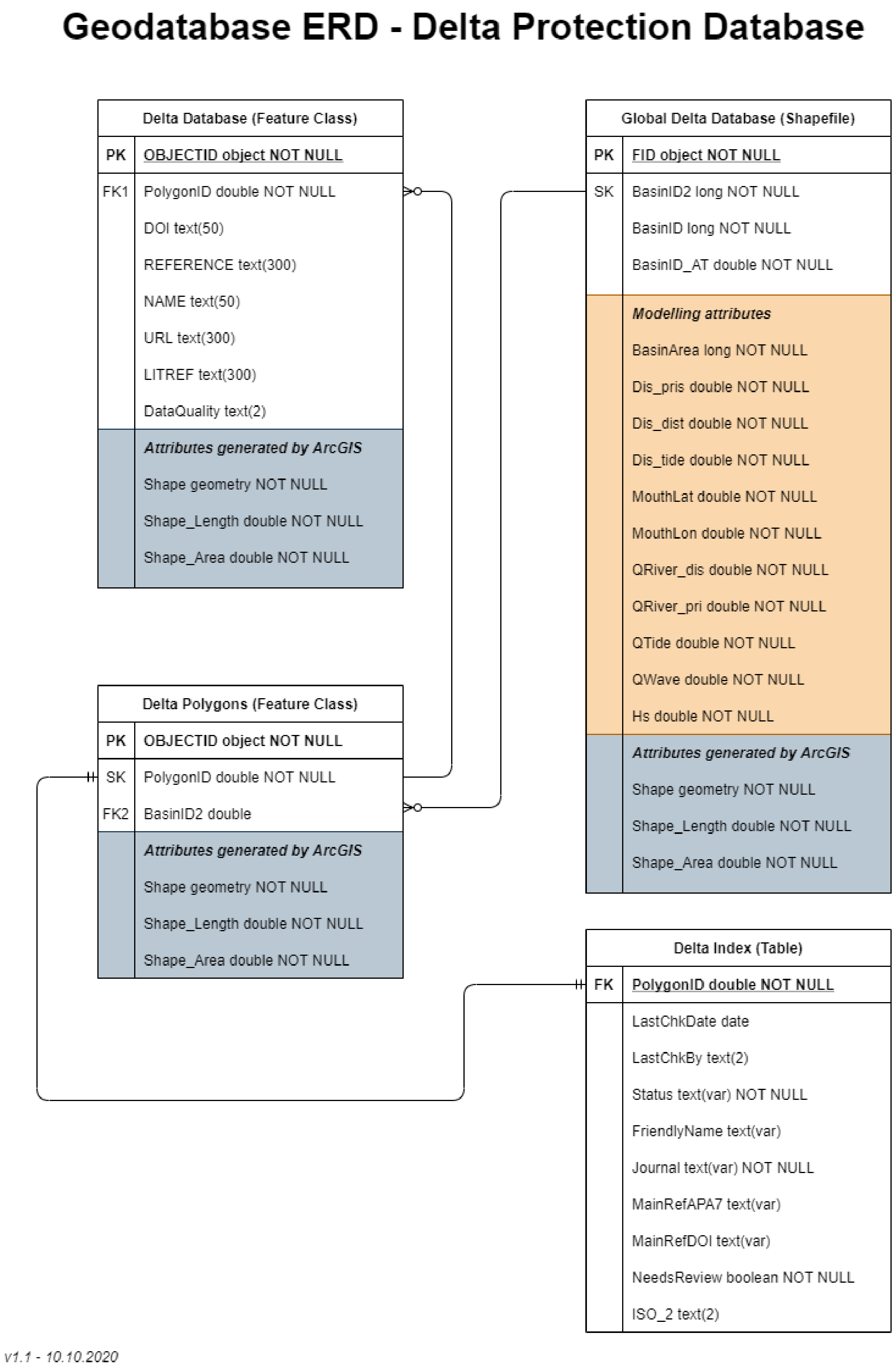


Figure 2 - Project Entity Relationship Diagram

M E T A D A T A

Even prior to the advent of mass data-sharing platforms and policies, a UN workshop by Kovács-Lang (2000) identified that proper metadata aided in increasing longevity, reducing entropy, and improving reuse and sharing of the dataset; metadata is key to making data "work".

Standards for geospatial metadata had been around for some time, with the US FGDC-CSDGM (Federal Geospatial Data Committee Content Standard for Digital Geospatial Metadata[1]) in existence since 1994, and the EU INSPIRE[2] directive from 2007. These standards underpin the discovery and flow of data across borders both physical and digital. However, there existed no single global standard, rather a select few standards who are used discretionally throughout the world. The most widely accepted of these is the ISO 19115 (and 19139) standards, from which INSPIRE was derived; despite popular use, FGDC-CSDGM was no longer recommended by the FGDC, who instead advise the use of an ISO standard (Federal Geographic Data Committee, n.d.).

With the global scale of this dataset, using a standard format of metadata enabled its discovery by users across the globe using a multitude of searching systems and algorithms. Users of the dataset are very likely to be in locations without access to specialist software, and may be in times of crisis, and so will have a need for data that could be found quickly and drawn into existing/open-source programs with ease.

Large dataset aggregators such as DataCite (datacite.org) actively combine information from thousands of repositories, and an increased focus in academic about making science open and accessible, metadata had a crucial part to play in future dataset discoverability. Regardless of the size or extent of a dataset, without accurate metadata it was very unlikely to be found. Metadata also encodes the units of measure, field descriptors, and other key information that users who are not familiar with the data would need to be able to decipher and use the dataset.

Within the European Community, the ISO standards have been formally adopted in legislation as the INSPIRE standard; this is built upon the already strong basis of ISO standards, and is the most interoperable choice, and so it was used as the metadata standard for this project. ArcGIS allowed for generation of metadata according to the INSPIRE standard natively within the system, however as part of the data quality exercise to be performed, the metadata was additionally validated through the INSPIRE Reference Validator[3] to ensure adherence to the full standard.

In all of the datasets reviewed in the production of this dataset, structured metadata was only published by the six of the areas - US, Australia (NSW, VIC, SA, WA), England – and with the exception of England and the US, the remainder were sparsely populated, and none were fully compliant to the INSPIRE standard. This required additional searching and manual notation of the syntax and format, which further delayed data use.

---

[1] US Federal Geographic Data Committee – Content Standard for Digital Geospatial Metadata - https://www.fgdc.gov/metadata/geospatial-metadata-standards/

[2] Article 5(1) of Council directive 2007/2/EC on establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) *Official Journal.* L 108/1

[3] INSPIRE Reference Validator © European Union - https://inspire.ec.europa.eu/validator/

REFLECTIONS FROM EXISTING DATASETS

Having reviewed the existing datasets, it was prudent to not only review the data they produced, but also to draw some conclusions and points of improvement ('lessons learned') from the work done by other academic and commercial projects. These were used in the definition of a methodology for research and for decisions regarding the onward dissemination of the data.

- **The data must be freely licenced in a manner that permits and encourages re-use, and the licence should be clearly defined in the metadata/documentation**

Licensing procedures and enquiries to data custodians only served to delay or dissuade further research. Data was to be made available under a widely used and understood licence which is compatible with the licences from the parent/source data.

The licence needed to be clearly displayed on the source website and within the (meta)data files, so that it can be easily understood whether any restrictions applied to usage.

- **Sources and reference should be both machine readable and intelligible to humans.**

Custom defined coding/reference schema served no purpose if no information was provided on how to interpret them or recreate the result. References to source material or other data must be in a way that is not tied to an individual hosting service, where possible, and a persistent identifier should be used.

- **Data should not be tied solely to the University or the specific publication, but should be available on a widely used service, including aggregators like DataCite.org**

Creating a custom website that was disconnected from data indexes seemed commonplace but served only to further isolate the data from possible users. This also introduced costly hosting and maintenance requirements. It was deemed essential that read-only data access was possible without a login.

- **The data platform must be able to be updated easily, with the minimal number of copies (ideally one) of the data in existence**

Publishing the data directly to a journal usually prevented further updates being made to the data, as it acted as a snapshot at time of the paper being submitted. A data repository compliant with the university and publisher's policy was to be used to allow a snapshot of the data at time of submission to be stored, but also allowed for future updates (versions) to be made so that data consumers always received the latest version. Where possible, data was to be held in one place.

There should also be a simple way for data users and reviewers to suggest amendments and reviews, as this was not present in any of the systems reviewed and resulted in data siloes.

- **Metadata is just as important as the data itself**

Although some of the government-services published metadata alongside their data, the quality and completeness of the metadata was generally poor. Metadata is key to data longevity and additionally ensured that data remained citable when re-used by additional parties. Additionally, issues faced with lack of understanding of coding/reference schema in the reviewed datasets could have been resolved by the better usage of metadata.

Data Collection

C O N S T R A I N T S

To build a global database, unsurprisingly it was necessary to collect information from all around the world, but there needed to be clearly defined extents for the data search to best utilise limited processing power and time limits. It was therefore decided from the outset that only areas which were reliably defined as deltaic were analysed for the purposes of the first version of the database. This does not rule out extension of the database in the future beyond delta areas.

The definition of a deltaic area has been inconsistent and was variably implemented in research papers. Edmonds *et. al.* (2020) thoroughly reviewed nine studies that defined area, and could not find a common method amongst them; they further expanded upon the differences and reasoning for the variances, and provided a documented method for delta shape definition using satellite imagery to define all areas where deltaic activity was occurring, or can be shown to have occurred in the past. This gave a worldwide, defined, and interactive set of delta 'polygons', which were used as the start point of data investigation.

Data would only be added if it fell within 100km of the bounding edges of the Delta Polygons, whether processed manually or through a geoprocessing model. This was defined to reduce the volume of data that was 'orphaned' (had no 'parent' delta) and additionally reduced work burden. As a variable, this could be changed in future iterations of the database if needed.

S E A R C H   H I E R A R C H Y

Priority was given to vector and geospatial data sources as these could be processed directly into the model, however these were not the only sources of data that could be used, and multiple data sources were consulted for the same area. An idealised process has been given in Figure 3, p.13.

However, as it was impossible to tell if a data source is complete due to the various temporal and geopolitical factors affecting data availability, a wide number of data categories were consulted for each area. Vector data was always compared to satellite imagery, and various sources were used to compile an accurate picture of the state of levee features per area.

R E V I S E D   S E A R C H   F O C U S

Initially, data collection began on a data-led approach to a target cluster of deltas (Mekong, Ganges-Brahmaputra, Rhine-Meuse-Scheldt, and Mississippi). This was key for the transfer of knowledge and the creation of the geoprocessing model using the National Levee Database as a baseline. As expertise was built up and knowledge of information sources increased, a more targeted approach was employed where data searching was done on a regional basis.

Although the target was never to complete full analysis of the 2,174 shapefiles in the Delta Polygons dataset, 1,277 (58.74%) had been reviewed at the point of writing this report. As the database is planned to be continually developed and enriched by the research hub and the wider community, the database was not considered fixed, but for the purposes of publishing the data in a journal, the decision was made to prioritise the remaining deltas based upon total area.

Where it was not possible to analyse a delta due to the time constraints, the relevant entry in the delta index was updated and the `NeedsReview` indicator was triggered. This was so that the data would be highlighted to users and consumers that it was incomplete.

Hierarchy of Data Sources Flow Chart - O'Dell, J. (Joey) - *Guided Research Data Selection v1.1 - 06/11/2020*

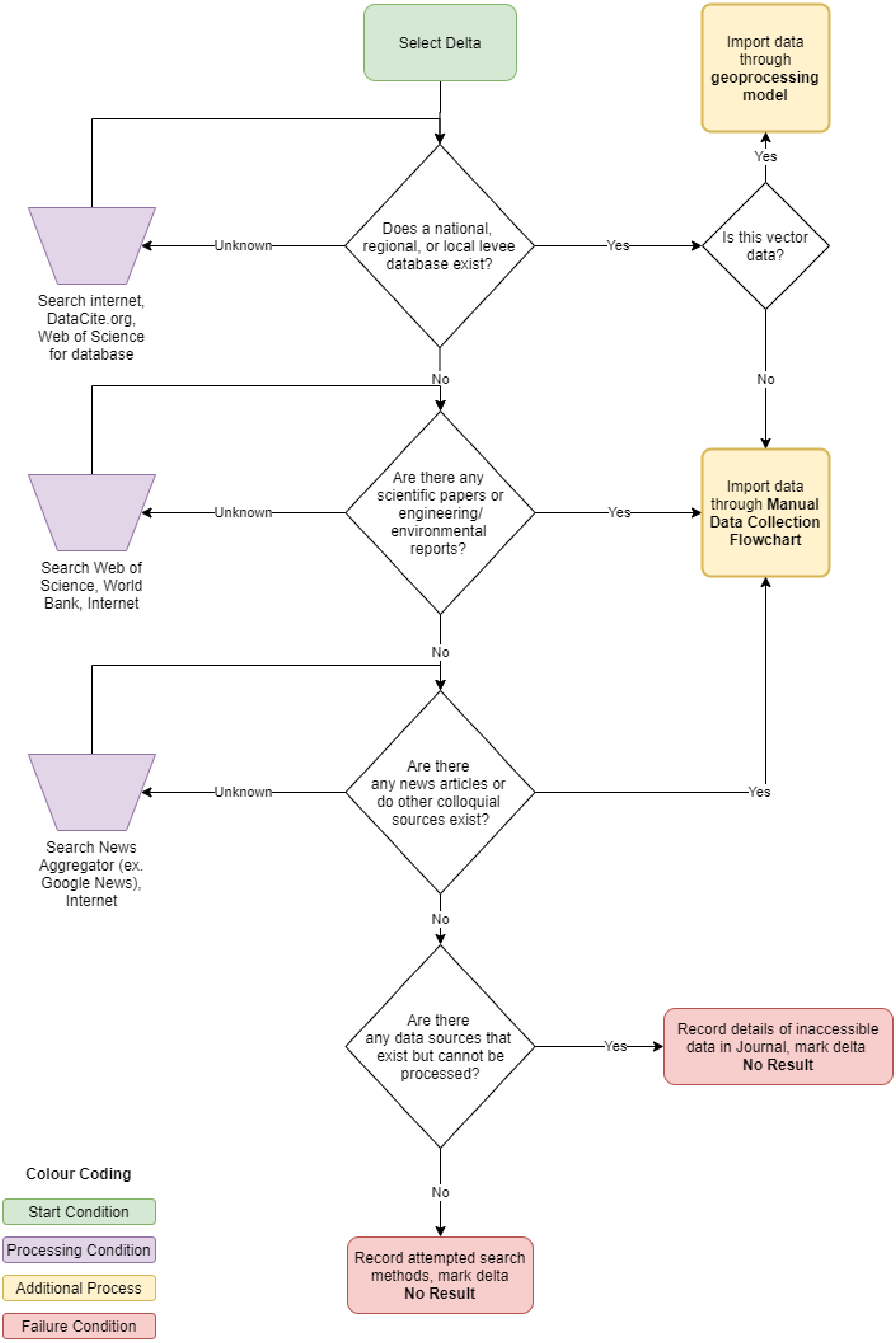


Figure 3 - Hierarchy of Data Sources Flow Chart

COMPREHENSIVE LITEREATURE SEARCH

Comprehensive literature search made use of aggregation and indexing services, including Web of Science™, DataCite and WorldCat®. For each location, targeted searches were made which looked for publications or other works using a standard set of keywords as well as numerous words that described the features to be searched for. Below is the key set of phrases that was used:

Table 3 - Initial search terms list

Location-sensitive	General search terms		
[River name]	Dyke	Levee	Revetment
[Delta name]	Dike	Flood	Defence
[Village name]	Seawall	Sea	Beach
[City name]	Coastal	Wall	Stopbank
[Country name]	Dredging	Channelisation	Channelization

Where it was known or discovered through research that another word existed - for example, the word ***tanggul*** meaning 'levee' or 'embankment' in Indonesian - this was also added for regionally appropriate searches.

In addition to this, a general internet search was also employed to scope out publications, reports, and other media that had not been indexed by the above systems. This was especially important for World Bank reports, but also for journals in a language other than English, as Google provided a direct translation service when searching websites and PDFs.

BASIC INTERNET SEARCH

Where it was impossible to deduce the name of the delta/river system through using the geolocate ("gazetteer") function in ArcGIS, or having inputted the coordinates into an online map (such as Google Maps) and searching local points of interest, then the comprehensive literature search was skipped (i.e. searches were not performed against DataCite, Web of Science) and instead only internet search providers (i.e. Google) were used.

Internet searching was however not limited to text results. Image searches often proved fruitful, for maps of dykes or levee systems at a very local level were sometimes found, and although a time-intensive search method, was useful for some areas where published data was sparse or difficult to find.

IMAGERY

Although a levee or feature may have been present at the time of writing a journal paper or publishing a database, it was likely that flood events and maintenance (or lack thereof) had occurred in the time since publication. As such, information gathered from disparate sources was confirmed using Maxar Vivid 0.5m (and where present, 0.36m) satellite imagery within the ArcGIS platform. This acted as part of the quality check process performed at the delta level.

Where the information could not be verified using satellite imagery, the appropriate data quality classification was applied to that entry to show that it was not verifiable from a second source.

The above steps have been documented as a generalised process in Appendix III.

Processing & Verification

The example in Figure 4 below shows the manual georeferencing process in operation. A raster image of a levee map from Bangladesh (World Bank, 2013) was viewed side-by-side (each image was displayed on a separate computer screen in use) with a current basemap. Key markers from the dataset were georeferenced using satellite and topographic maps (including OpenStreetMap).

The map has been suitably georeferenced to existing infrastructure, however the changing nature of the delta environment as reflected in the satellite imagery does not reflect the basemap used in the production of the report that the map was taken from. This was expected, as deltaic environments are susceptible to ongoing change and erosion (as well as aggradation) and therefore maps that varied by only a small number of years were expected to show such change.

Where variations, such as the outlines of the channel island edge as highlighted in the red circle, additional sources of basemap material were considered, and in the example given of Bangladesh, multiple sources of polder levee data were also used to mitigate cartographic differences. A best-attempt method of georeferencing was used to provide realistic data that can be verified.

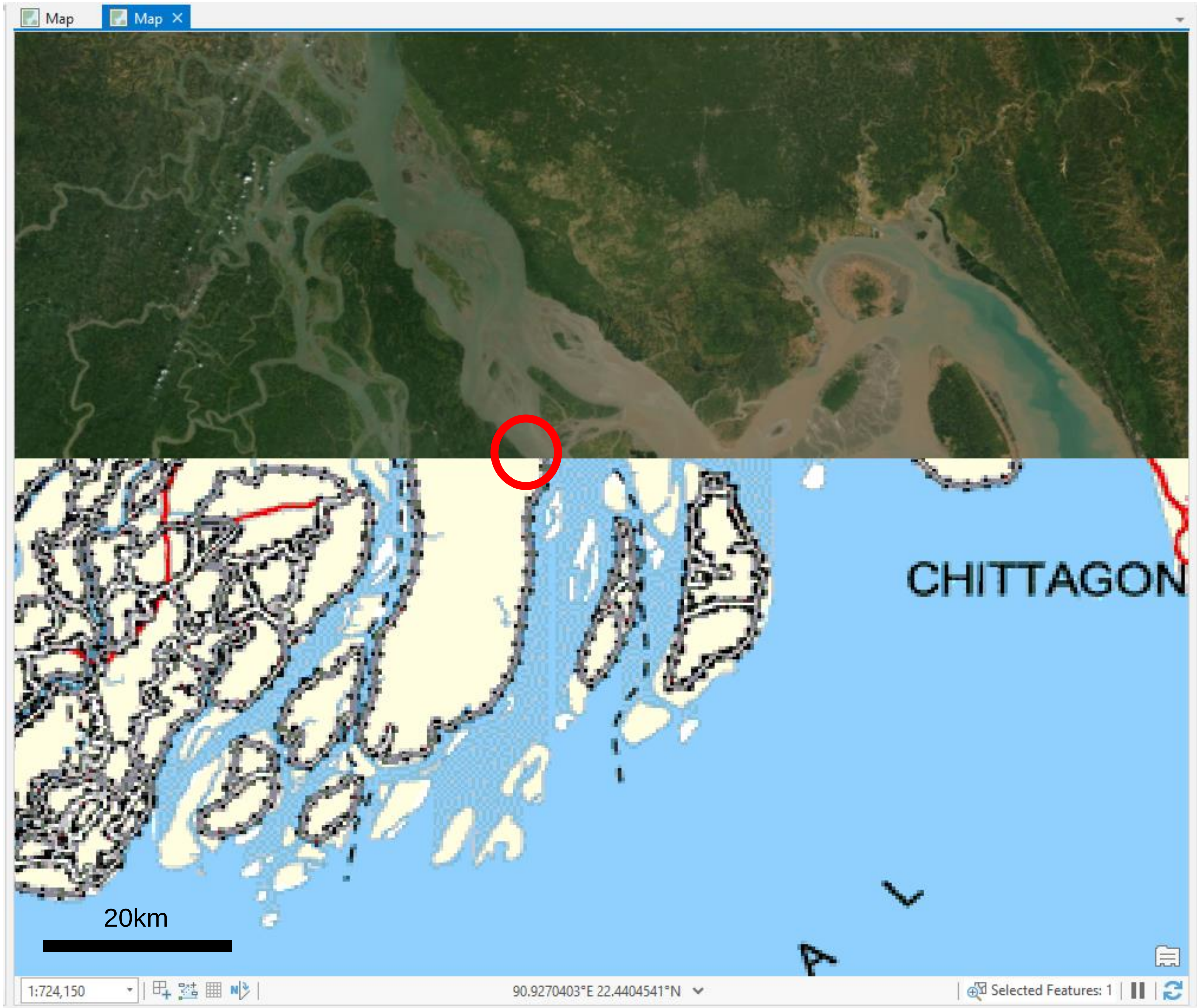


Figure 4 - Verification of retrieved data against basemap (Ganges-Brahmaputra Delta, Bangladesh)

PROBLEMS ENCOUNTERED

The following issues were identified during the manual georeferencing process:

Table 4 - Georeferencing problem resolution matrix

Problem	**Cause**	**Resolution**
Raster maps were sometimes extremely distorted when georeferenced accurately	Raster data used local projections which are not defined/no metadata	Used the highest resolution image available, and browsed alternative data sources
Raster maps did not align with modern basemaps	Basemap used by creator was old or custom made **Or** Basemap is of area at a different point in time and delta has changed shape	Compared using Open Street Map (OSM) for second opinion **Or** Reviewed historical satellite imagery **Or** If no positive results, recorded data quality flag 'C'

MANUAL METHOD

Where levee data was made available as line data, and no information on leveed area was found (such as the Australian levee datasets, but also some raster sources where only levee area was detailed) then human interpretation (manual/subjective interpretation) was required to decide the area protected by a levee. This interpretation made use of the following factors:

- Location of levees relative to rivers
- Topographical changes in the proximal area
- Presence of other hydrological features that inhibit levee function
- Best-effort estimation of levee protection

Where this was necessary, the following constraints were applied:

- Where the levee forms an enclosed shape, no vertices were to be added
- Where this was not possible, a minimal number of additional vertices were added, so that the leveed area remained realistic, but the outline of the levee features remained true to the source data
- Where the levee followed a coastline or other clearly defined feature (e.g. road, railway, sea wall), and this was confirmed in satellite imagery to have an elevation difference, this was sparingly used to ensure that the leveed area remained a complete shape
- Where a levee feature crossed a water body, two separate shapes were drawn but with reference to the same source dataset
- With respect to the age and purpose of the data, where contradictory data was received, it was reviewed together with the satellite imagery and the decision made was entered into the delta index.

Figure 5 & 6 on the following pages give examples of the levels of interpretation in use.

A – Levee feature visible in centre of image (brown line) Data: DWER (2020)

B – Traced polygon constructed solely from the vertices of the levee feature (blue) Data: DWER (2020)

Figure 5 - Simple processing of levee line data (Gascoyne River, WA, Australia)

A – Line data from EUROSION v2.1 dataset

Polygo	LastChkDa	L	Status	Friendly	Journal
2969	04/10/2020	JO	Processed	Achelous /	[04/10/2020, JO] Checked area using comprehensive literature search and Maxar Vivid 0.5m imagery. Coastal defences located within EUROSIONv2.1 and drawn in. Town of Missolonghi / Messolonghi is reported to be canalised but difficult to locate. Area to east of polygon has coastal defences confirmed in EUROSIONv2.1.1 however there appears to have been work undertaken since the release of EUROSION. Used satellite imagery to confirm presence of sea defences and re-drawn using satellite data. Unable to find any data on defences in bay area (21.2777989°E 38.3485453°N) and so remains blank. Evidence of new coastal works since EUROSIONv2.1.1 dataset and enhanced with satellite imagery, coastline altered.

B – Resultant polygons drawn in dataset and documented decision-making process in delta index table

Figure 6 - Inference of levee feature using multiple data sources (Acheloos Delta, Greece)

AUTOMATED METHOD (GEOPROCESSING MODEL)

The ArcGIS Pro model builder allowed for the creation of a model using building blocks (geoprocessing modules) to create a complete a process, as well as features to build an interactive user interface. This automates the process of selecting and importing vector data. Although vector data could have continued to be processed manually, creating a model using the Arc ModelBuilder allowed for three main functions:

- The same processes were applied to datasets quickly, and variables were set as parameters to allow for an interactive (point-and-click) interface for non-expert users
- The entire process was replicable and could be shared with another person in the form of a file, which improved process transparency
- The model could be run step-by-step, and the data generated at each stage was possible to check iteratively to ensure that data integrity was being maintained, and faults and problems could be analysed.

All of the above allowed for significant time savings when processing expansive datasets, as thousands of features could be processed in under an hour, whereas manually georeferencing data or tracing vector sources could take upwards of 3 hours per delta polygon.

Figure 8 (over page) demonstrates the layout of the model and the steps undertaken, using common geoprocessing tools as per ArcGIS Pro 2.6.2.

The below figure (7) is displayed when hovering over Fields in the user interface, as show in the figure on the next page. This prompted the user with any interface-specific features they may not be aware of, such as the need to replace certain formatting with a shortcode.

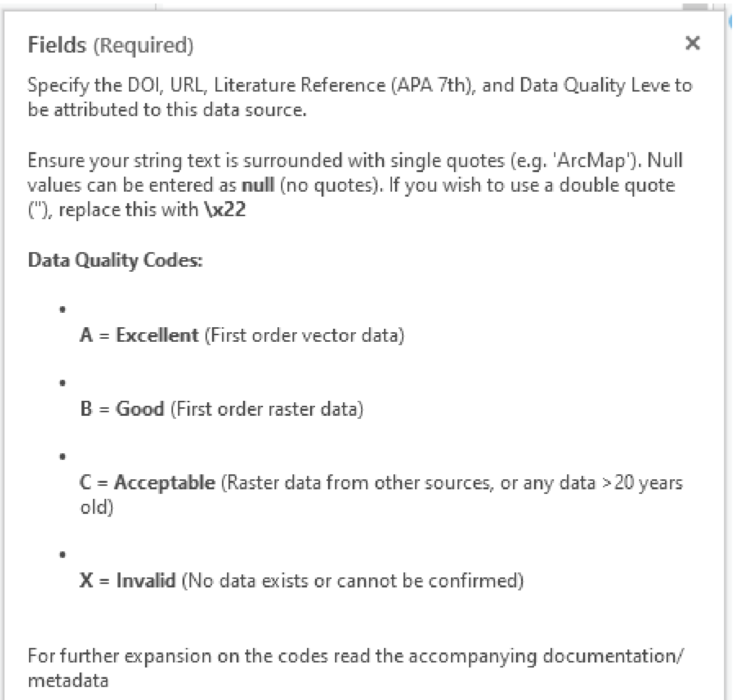


Figure 7 - Example of interactive tooltip helptext in model

Because of the implementation of interactive helptext, the model was designed to be used without full reference back to this report or the supporting documentation and reduced the need for a dedicated help file. This helped increase user autonomy and was designed to reduce model queries/emails after publication.

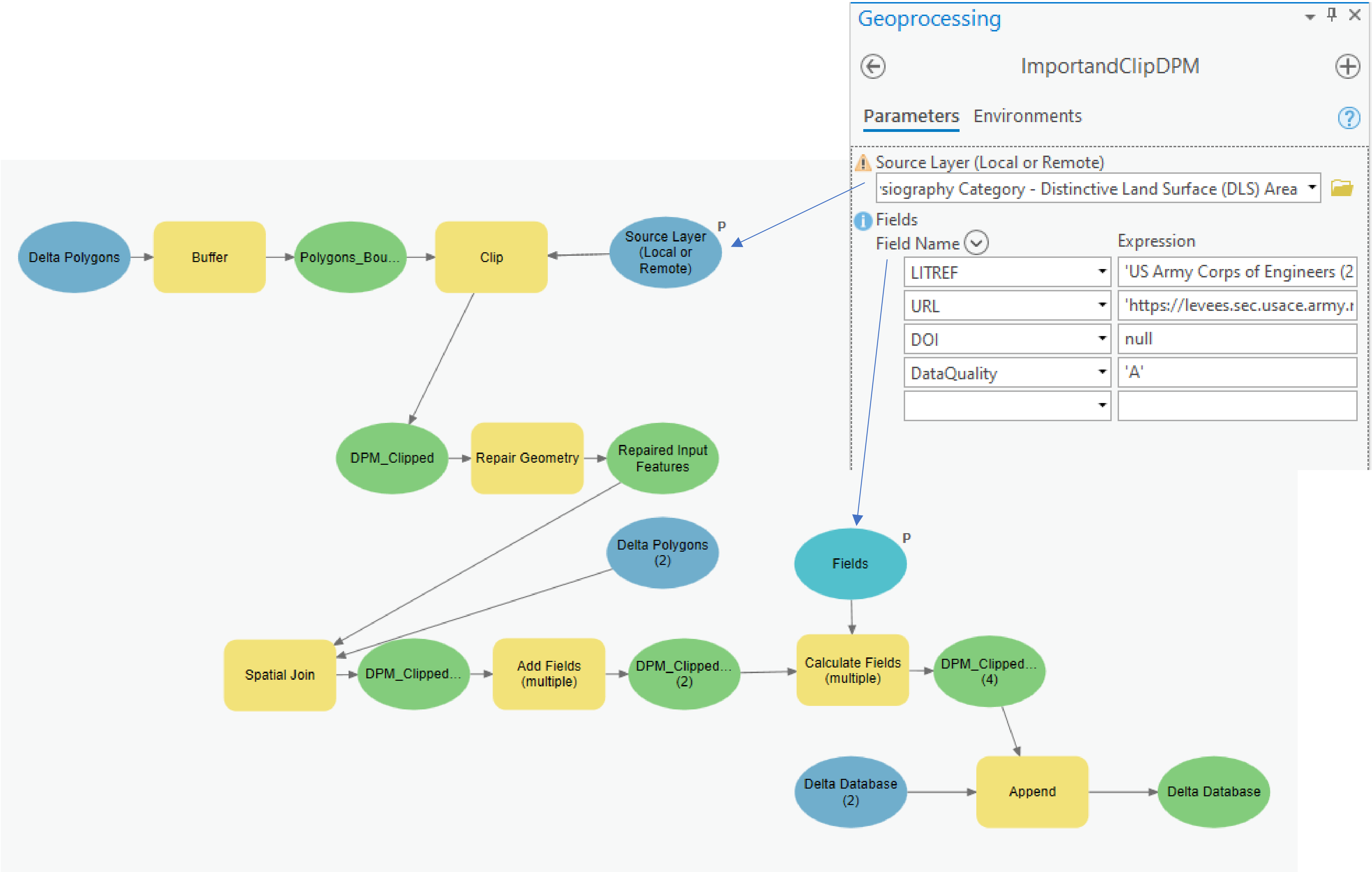


Figure 8 - Leveed Areas Processing Model

3. Results

Data Dissemination

P L A T F O R M S P E C I F I C A T I O N

The project was focussed not only on the creation of a dataset, but also ensure that the resultant database could be shared digitally on an open platform that can be interrogated by members of the public, researchers, etc.

The criteria for the choice of platform were set as follows:

Table 5 - Platform Specification Criteria

Criterion	Importance	Reason
Free of cost at point of use	Essential	Dataset is being developed for re-use and dissemination and has been produced at a public university. Therefore, the data should be able to be consumed by all.
Integrated map viewer	Preferable	The data will be more readily consumed if it can be viewed on the same platform. But it can be acceptable to have it hosted just as a download.
Version control or archival function	Essential	The dataset will be revised and improved during the lifespan of this project and beyond. It is necessary to be able to update the dataset without losing historic data.
Data sharing/export function	Essential	It must be possible to access a copy/export of the dataset in its entirety for other users to be able to rework/edit/analyse the data.
WMS facility	Preferable	Allows consumption of data without downloading, which is preferred by some professional users of GIS platforms
Data platform is well established with a global network of users	Essential	The data must be easily found, and the users must not be geographically limited – i.e. it must not be a platform that is only used by Dutch universities
Integration into existing geocatalogues	Preferable	Being able to search for the data directly within the GIS programme will increase uptake and use.
Platform must be available in English	Essential	This aids global consumption.
Other viewing languages should be possible in the platform	Preferable	Although most users will download the data and open in a local programme which should be in their local language, it would be beneficial to have a small number of interface language choices.
Persistent identifier	Essential	A persistent identifier (preferably a DOI) should be used so that links to the database can be relied upon
Integrated with the storage platform	Preferable	Storing a single copy of the dataset, rather than one on the mapping platform and another on the data drive.

The analysis of each platform performed against the criteria set out in Table 5 (p. 21) can be seen in Table 6. The following research data platforms were discounted from analysis for the following reasons:

- Dryad – https://datadryad.org/ - Provided similar service to OSF but with US$120 data curation fee
- Zenodo – https://zenodo.org/ - Indexed to OpenAIRE, which OSF is also indexed to
- Figshare – http://figshare.com – 5GB file limit

P L A T F O R M   D E C I S I O N   M A T R I X

Table 6 – Data Platform Decision Matrix

Platform	ESRI ArcGIS Online	DataverseNL	Open Science Foundation	PowerBI	Harvard WorldMap	GeoNode	Yoda	SURFdrive
Deployment Type	Cloud	Cloud	Cloud	Cloud	Cloud	UU Hosted	UU Hosted	Cloud
Cost to project	None – borne by department licensing costs	Free	Costs being implemented end 2020 for storage	None for standard use Cost for Azure storage platform	Free	? - Local IT infrastructure, domain, hosting, RDM authorisation, IT architect time etc.	EUR4 per TB per month	None – cost borne by SURF
Cost to user	Free	Free	Free	Free	Free	Free	Free	Free
Versioning	Yes	Yes	Yes	Not directly – need to use SharePoint or ArcGIS versioning	No	Yes	Yes	Yes
Map Viewer	Yes	No	No	Yes	Yes	Yes	No	No
Global User Base	Yes	Yes- via DataCite	Yes	Yes	Yes	No	No	No
Geocatalogue Integration	Yes, extensive	No	No	Yes, limited	No	No	No	No
Multiple Languages	Yes (global)	Yes (NL/EN)	Yes	Yes (global)	No	Yes (local configuration)	Yes (NL/EN)	Yes (NL/EN)
Datacite indexed	No	Yes	Yes	No	No	No	No	No
Accessibility	Free to consume Account to edit/version	Free to consume Account to edit/version	Free and open access to consume Account to edit/version	Free to consume Account to edit/version	Free and open access to consume Account to edit/version	Free to consume Possible to configure locally requirements for editing	External users must be granted access. No access without account	Restricted to users within SURF federation (Netherlands)
Data security	Load distribution and global data backups within platform. No delete restoration possible.	Data Archiving and Networking Service (DANS) national data security policy. Hosted by DANS and backed up within the Netherlands. Data versioning and recovery possible	Data backed up by COS. Data is encrypted at rest and snapshots are encrypted also. COS has a preservation fund to enable the data to be accessible for 50+ years in the event the organisation closes	Data backed up by Microsoft to off-site locations. Delete restoration possible (30 days) in SharePoint	No documented data security policy	Data security defined by project scope and UU IT At least to meet minimum standard of UU policy, backups by UU IT.	Data backed up by UU IT. Encrypted replicas stored offsite by EUDAT Delete restoration possible (30 days)	Data backed up nationally by SURF federation Delete restoration possible (30 days)
System integration	Directly integrates with ArcGIS / ArcMap Publishable as an ArcGIS server layer	No direct integration to ArcGIS platform. Can be connected to OSF	Integrates with git commit function No direct integration to ArcGIS API available to access data	Standard: None Azure: Integrates with ArcGIS Pro (not public)	No direction integration to ArcGIS platform.	No direct integration to ArcGIS platform.	No direction integration to ArcGIS platform.	No direction integration to ArcGIS platform.
Interoperability (incl. OGC Compliance)	Layers can be separately published as open standards layers (WMS/WFS) if needed	Not possible to publish as WMS/WFS	Not possible to publish as WMS/WFS	Not possible to publish as WMS/WFS	Map can be published as WMS/WFS but without *getcapabilities* *½ Mark*	Layers can be published as WMS/WFS/OpenAPI	Not possible to publish as WMS/WFS	Not possible to publish as WMS/WFS
Timescale to release	Immediate – publish from ArcGIS	Short (1-2 days) – publish manually	Short (1-2 days) – publish manually/via git	Short (1-2 days) – publish manually	Short (1-2 days) – publish manually	Long (>1 month) – Would need to procure hosting and arrange IT/Dept authorisations	Short (1-2 days) – publish manually	Short (1-2 days) – publish manually
Access recoverable by UU	Yes – UU GEO ICT can recover access	Yes – UU RDMS can recover access	No	Yes – UU ITS can recover access	Yes – UU ITS can recover access	No	Yes – UU ITS can recover access	Yes – UU ITS can recover access
Persistent Identifier	None	DOI	DOI	None	None	None	Persistent Keys	None
Criteria Met (/16)	14	13	10	10	8.5	7	7	7
Chosen Application	Map Viewing Platform	Research Data Storage	Not chosen	Not chosen	Not chosen	Not chosen	Not chosen	Not chosen

PLATFORM DESIGN

It was decided to set up the ArcGIS project as a File Geodatabase, as this was a solid base point for easy transition of data in the future and for the storage of vast disparate datasets (Childs, 2009). This also enabled easy transport of the geodatabase and made backups easier to roll-back. Compressed 'snapshots' of the dataset were taken at regular intervals and stored in different storage locations (physically located on-campus) as per best practice.

Not only in keeping with general data management practice, but also to reduce data overhead in the future, it was identified that there should exist only a single copy of the data published to the public. By removing multiple copies of the same dataset, the administrative burden in maintaining these updated was removed, and the likelihood of the latest version of the data always being available increased. The idealised solution was initially proposed in the following diagram:

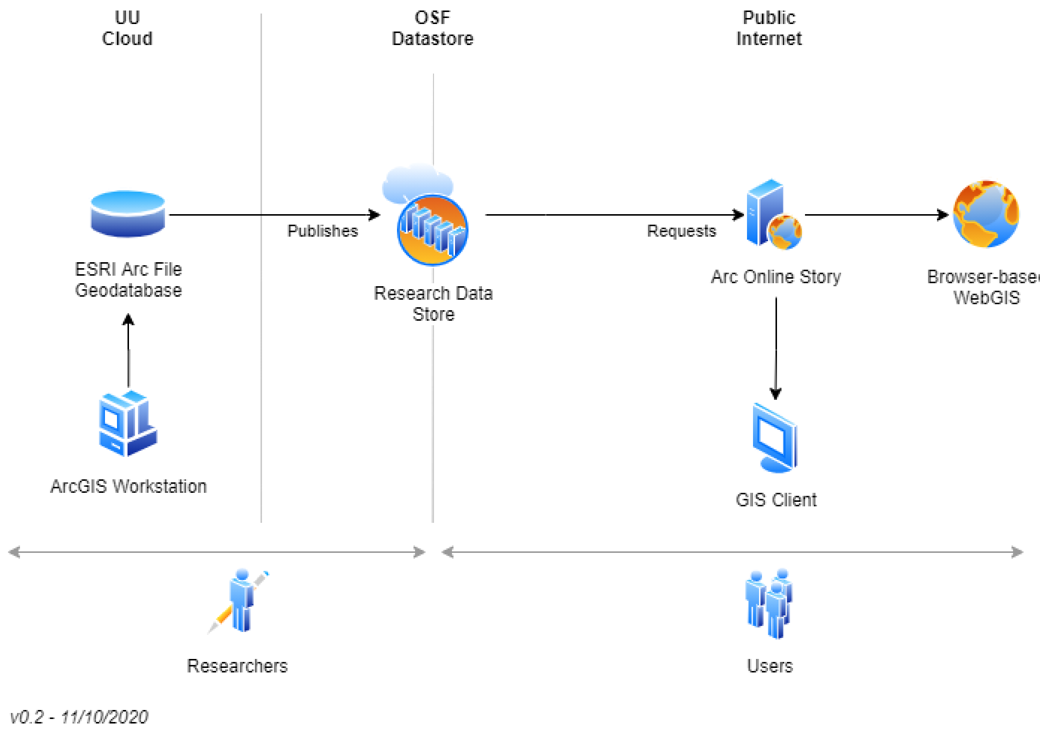


Figure 9 - Idealised distribution platform (single-copy design)

However, after investigation with the functionality of the ArcGIS platform, it became apparent that adding a file on an outside hosting service into an ArcGIS Online Story/Layer required a persistent URL. None of the data sharing platforms reviewed were able to offer this feature as they offered versioning and distributed storage, so that file URLs were not direct links to the file but to a script that returned the location of the latest version of the file.

Additionally, as the data was being published in an academic journal, the journal's data policy had to be taken into account. This required data to be published on a FAIR-aligned data repository (see FAIR Checklist, p.28) that issued a DOI. As this had already been covered in the data platform decision matrix, this was not a concern.

Therefore, a revised platform design was created. This centred around the use of the ArcGIS Online dataset, but where a copy at each revision is saved to the research data store. This store was issued a DOI and contains metadata and descriptive text to identify the 'live' layers on ArcGIS Online. The hosted layers could be edited directly in ArcGIS Online without the need for a full Arc licence, but also in ArcGIS Pro if more intense processing was needed.

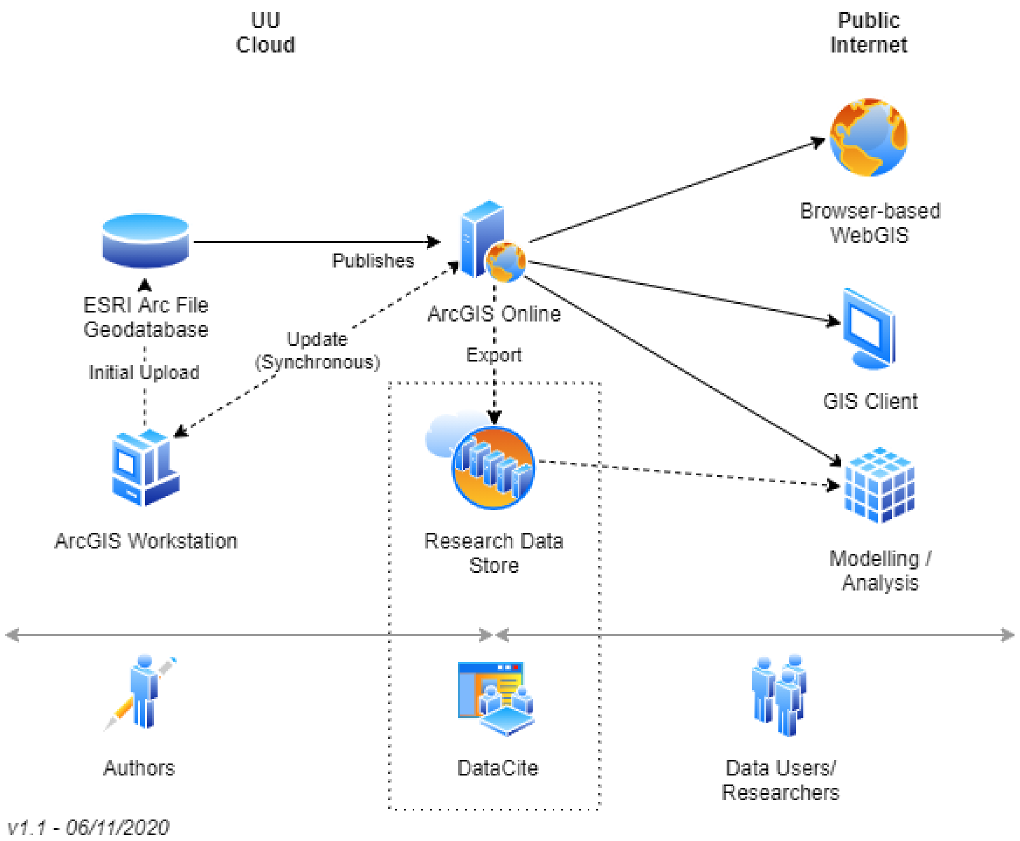


Figure 10 - Revised distribution platform

Data was released directly to the ArcGIS Online cloud environment from the workstation. From here, the following files were downloaded to give a true snapshot of the data state, and this is then published on DataverseNL. This was then to be used for the journal submission.

- Keyhole Markup Language (KML) Zipped (*.kmz) – *DEPROMdb_KML.zip*
  Chosen as it provided a single file which could be imported into various WebGIS and non-standard platforms (e.g. Google Earth Engine) and complied to an open standard sponsored by the Open Geospatial Consortium.
- ESRI File Geodatabase (*.gdb) - *DEPROMdb.gdb.zip*
  Richest dataset for replication, largest file size, but as a proprietary file system it required ArcGIS to open.

Self-directed downloads in other forms (GeoJSON, CSV, File geodatabase) were available from the ArcGIS Online service page. The Delta Index was published on DataverseNL as a CSV.

At the time of writing the ArcGIS Online platform had no built-in system for persistent identifier (i.e. DOI) creation, and the process proposed by Wright (2018) identified the need for an external provider to register a DOI. DataverseNL allowed for allocation of a DOI, and redirection to an 'Alternative URL' which could be accessed from the DOI thus linking the two services together and identifying they were related.

Figure 11 - Alternative URL redirection functionality within DataverseNL

D A T A   D I C T I O N A R Y

The data dictionary (Table 7) is provided as an insert on page 26

Custom fields that were added to the dataset, and which were not automatically generated by the GIS platform, have been described in the data dictionary.

Field names entered wholly in capitals (ex. `LITREF`) were those which were permitted to be populated within the geoprocessing model and so have been indicated as such for clarity during the model building phase.

Where a field present in the attributes was from another data source, this was defined, however any attributes (such as modelling attributes in the global delta database) which were not foreign keys were not. This table has also been encoded into the metadata file provided with the dataset, and as a separate 'Maintenance Guide'.

L O N G E V I T Y

In line with the best practice guidance from the University Research Data Management Support (RDMS) service, a clear end-point for the data was defined before data collection began. This included curation and maintenance of (meta)data after the end of the research project, and how continued access to the data was ensured.

Primary data storage was arranged through accounts tied to the university infrastructure (SharePoint) and is stored as a sub-folder of the research hub's environment so that access automatically passed to the research hub members.

Further information on data storage rationale has been provided on page 22.

Table 7 - Data Dictionary

Field	Description	Status	Source Dataset
BasinID2	A globally unique identifier for the sub-basin from the Global Delta Database	Foreign Key	Global Delta Database (Nienhuis, et al., 2020)
DataQuality	Classification of the data quality of each individual entry's source material	Attribute	Table 9, p.29
DOI	The Digital Object Identifier for the source material for the specific deltaic protection feature	Attribute	n/a
FriendlyName	Human-readable name for the delta area, if known	Attribute	n/a
ISO_2	ISO 3166-1 alpha-2 country code for base-level grouping of deltas, where the delta covers two geopolitical territories, the country with the largest coverage of the delta polygon is given.	Foreign Key	ISO Standard 3166-1 alpha-2
Journal	Diary-style log of processes undertaken, and decisions that were made at a delta level including alterations to methodology	Attribute	n/a
LastChkBy	A two-letter user ID used by the person or process that last performed an action on this dataset. This is not centrally maintained and is user-generated.	Attribute	n/a
LastChkDate	The date that the information was last "checked" (i.e. updated)	Attribute	n/a
LITREF	Literature reference to APA 7th standard for the source material for the individual entry	Attribute	n/a
MainRefAPA7	Literature reference to APA 7th standard for the source material for the polygon (master reference)	Attribute	Per the Journal entry for this area.
MainRefDOI	The Digital Object Identifier for the source material for the polygon (master reference)	Attribute	n/a
NAME	Where present, the name or title of the feature or area protected where the data has been imported from a spatial database/file	Attribute	Per each individual shapefile's LITREF
NeedsReview	Boolean value to identify an entry where additional checks/verification are due or where insufficient information has been processed	Attribute	n/a
PolygonID	The unique ID of the polygon in the reference dataset that the specified relates to ("parent" or "host" delta)	Foreign Key	Edmonds *et. al.* (2020)
REFERENCE	Where present, the unique reference, item or object ID from the source dataset where the data has been imported from a spatial database/file	Foreign Key	Per each individual shapefile's LITREF
Status	Processing status of the delta area according to the following choices: Not Processed | Pending | Processed | No Result	Attribute	n/a
URL	The URL of the source material for this entry	Attribute	Per each individual shapefile's LITREF

WEB INTERFACE

Although the dataset was primarily designed for onward use in other research, best practice in open science dictated that the data had to be made available for secondary uses beyond the initial scope of the system, and as such there needed to be an adequate way to interrogate the data and metadata using a web browser. The ArcGIS Online platform allowed publishing a layer as a web layer which could be consulted as a web app.

The pop ups were modified using Arcade expressions to include a direct link to the DOI, via the DOI resolver service. The DOIs were not stored in the attribute table as a full URL so that any changes in the resolver service could be implemented per map, or where the resolver service is accessible via an additional proxy (for example as at the University). This allowed increased customisation, and where a local/organisation specific system is implemented, capability to reduce the number of click-throughs required to gain access to data.

As the web app resides fully within the ArcGIS Online platform, it also provides instant access to data changes. Changes in data layers are dynamically fed into the web app, and this reduces the administrative burden in releasing new versions of the dataset.

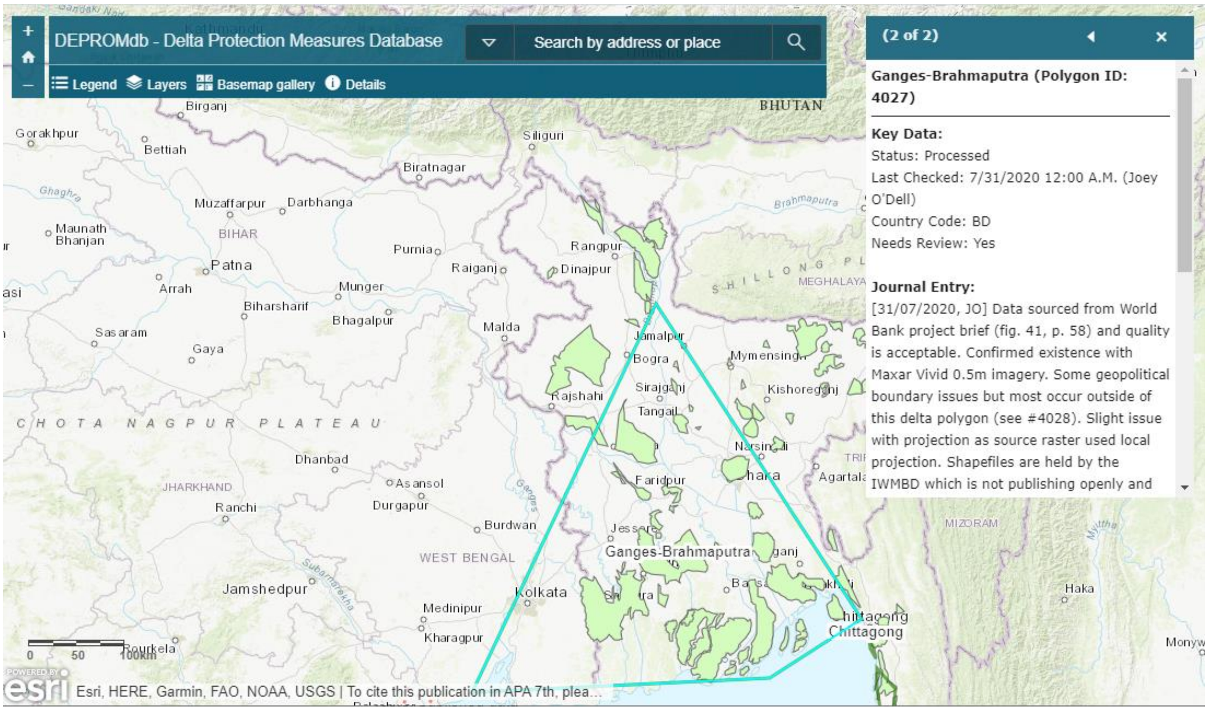


Figure 12 - Example of interactive browsing experience in database (Ganges-Brahmaputra Delta, Bangladesh)

Data Assurance

FAIR CHECKLIST

The EUDAT FAIR principles for open data and open science (Jones & Grootveld, 2017) were reviewed, and the data was found to meet the FAIR standard required for publishing in an academic journal. The dataset was evaluated in the following manner against the checklist:

Table 8 - FAIR criteria assessment

Findable	The data has a persistent identifier (DOI) There is detailed metadata for all aspects of the dataset The data repository is indexed to a data repository (DataCite) The metadata records at all points of publishing specifies the persistent identifier
Accessible	Resolving the DOI returns the DataverseNL page where data can be downloaded The data is accessible through multiple standards (HTTP, WMS, WFS, SSH) There are no authentication and authorisation steps for general access Metadata is always accessible, even when the data is not published
Interoperable	The data is provided in various common formats, including the open KML format Metadata is authored to the INSPIRE standard References are provided both within the data and the metadata
Reusable	The data has been quality checked and is well described The attributes are relevant and datasets with further attributes are linked The person and date of creation and alteration are recorded in the index A descriptive journal is provided to show changes and decision making

REVIEW INDICATOR

Whilst collecting data, data was discovered that was not possible to directly include in the database, or data was confirmed to exist but could not be accessed for reasons pertaining to licensing restrictions, lack of appropriate rights, or the absence of dataset reproduction permission.

As such, there was a polygon-level attribute of `NeedsReview` (Boolean, Y/N) which has been used to prioritise delta polygons for review when there was a possible additional data source or outstanding query. This was intended to be used in combination with the `Journal` field, as in Figure 13. The narrative showed that data has been found but could not be entered, so the `NeedsReview` flag was set to Y (True). This also allowed for symbology layers to be created in the web interface so that users and viewers of the data could be alerted to the need to revise the information.

Using the review flag also triggered a link to the submission form to be displayed at the top of the pop-up window, inviting users to contribute data to this area. This was designed to increase user-directed change and ownership of the dataset once released.

Journal	MainRefAPA7	MainRefDOI	NeedsReview
[16/08/2020, JO] Comprehensive literature search and Maxar Vivid 0.5m imagery. Comprehensive dataset available from the Danish govt (http://miljoegis.mim.dk/) however little information on flood defences. Further investigation shows that some measures exists but getting data is difficult (http://nordress.hi.is/wp-content/uploads/2016/12/NORDRESS_30-nov-2016_Thorsten_Piontkowitc.pdf). There is a sea wall/flood defence present but again difficult to find the information (https://northsearegion.eu/media/4433/poster_fair_ringkplus-bing_final.pdf) [03/10/2020, JO] Reviewed area against EUROSIONv2 dataset, no result.	NULL	NULL	Y

Figure 13 - Example of use of `NeedsReview` *field*

QUALITY CLASSIFICATION

Although good scientific method has been applied to data collection, and well evidenced reasoning has been stored at a polygon-level for the deltas, there was still a need for data quality to be machine-readable. Therefore, the following data quality groupings were generated to allow automated processing of data by source class, in addition to the linking of the source item.

Table 9 - Data quality assessment criteria

Category	Quality Class	Data Source/Reasoning
A	Excellent	Vector data First-order data source (i.e. scientific papers, governmental geospatial data, original publication) Spatially complete (with respect to geopolitical boundaries) Existence verifiable with satellite imagery
B	Good	Raster data (suitably georeferenced) First-order or re-cited/modified (original inaccessible) but published within a scientific or governmental publication Existence verifiable with satellite imagery
C	Acceptable	Raster data (loosely georeferenced) Conjectural or non-scientific source (ex: newspaper) Source >20 years of age, regardless of type Existence (partially) verifiable with satellite imagery
X	Invalid	Data inaccessible (blocked, hidden, non-existent) Known irrecoverable issues with data quality Could not confirm existence using satellite imagery Temporary or reactive measures only (ex: sandbags)

Category *X* was included for completeness and for future use (in regard to the receipt and processing of externally submitted data), however in the final output from this project, no data was published that met that category.

VALIDATION

Data validation was performed in the following order:

- Checked for duplicate `PolygonID` entries in the index, and consolidated duplicate entries
- Checked for orphan polygons where the linked `PolygonID` could not be found in the index dataset
- Verified there were no delta polygons where a join to the delta index table on `PolygonID` returned null
- Checked that the `BasinID2` was within an acceptable range (-1 to 7602906)
- Checked that where a delta polygon did not fall within a global delta dataset polygon, the `BasinID2` attribute was set to `-1`.
- Verified there were no delta polygons where a join to the global delta dataset on `BasinID2` returned null
- Metadata validation was performed by exporting the metadata as an XML file from ArcGIS and verifying it through the INSPIRE validator[4]

[4] Available as a web application: https://inspire.ec.europa.eu/validator/

STEWARDSHIP

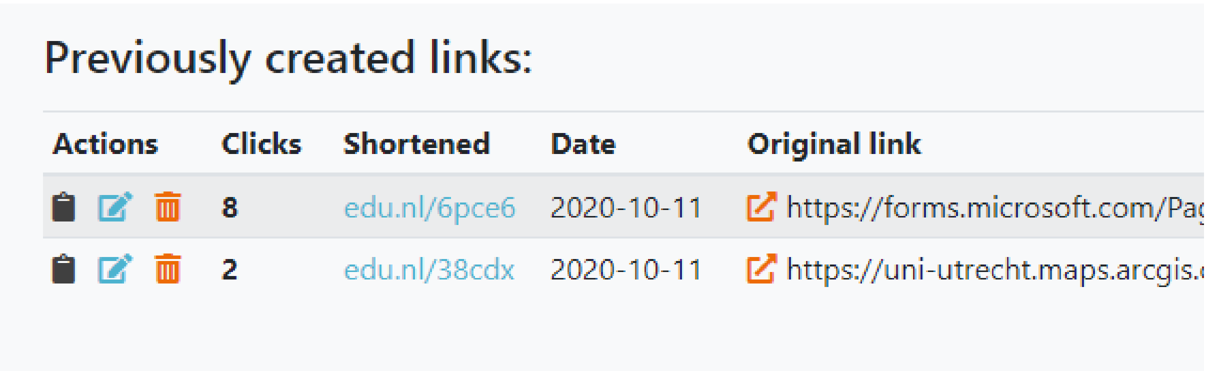
Given the lasting nature of the data created, and the intention for it to be possible to contribute to and enhance the dataset, there was a verifiable need to ensure that data remained accessible and attributable beyond the initial project phase.

The selected data storage medium (DataverseNL) provided a publicly available data store with persistent identifiers (DOIs) and helped to fulfil several requirements as outlined by the University's Research Data Management Service (RDMS). It is also RDMS who would be able to regain control of a dataset or arrange withdrawal in the event of loss of access or an issue that required suspension of the dataset. Additionally, as it was hosted by the Data Archiving and Networked Services (DANS) it ensures that access could be recovered should there be multiple breaks of the administrative chain within the faculty or wider university.

However, simply publishing data without appointing anyone to maintain the data (or who can facilitate withdrawal in the future) could be careless. Therefore, the dataset as a whole is set to be maintained by the WCFD research hub within the university by Dr Jaap Nienhuis, and access to the ArcGIS Online platform could be controlled by the department's ICT function (GEO-ICT).

Additionally, in line with common best practice advice from the Berkley Library (How to Write a Good Documentation, 2020) a README file was generated and included in the research data to act as a way finder for data users. The readme file was formatted in Markdown[5] so that it could be viewed by humans, but also interpreted and displayed on a website if necessary.

Any links provided in documentation were created using the edu.nl link shortener. This allows for control on the destination even after the document has been published, as the destination of the short URL could be edited after it has been released. This also provided an additional benefit that anonymous traffic information (number of hits) can be retrieved to monitor dataset utilisation.



Previously created links:

Actions	Clicks	Shortened	Date	Original link
	8	edu.nl/6pce6	2020-10-11	https://forms.microsoft.com/Pa
	2	edu.nl/38cdx	2020-10-11	https://uni-utrecht.maps.arcgis.

Figure 14 - Example of edu.nl link statistics

[5] Markdown is a text formatting (markup) language designed for readme files. Markdown © 2004 John Gruber, released under a pseudo-BSD open source licence : https://daringfireball.net/projects/markdown/syntax

4. Discussion

Although the project was designed as an ongoing effort, and the database had thus far been a positive outcome, it was prudent to consider the "*what else*", and the "*what next*."

Just in the same way as the world continues to spin, new and different ways of using the data were to become possible once the data was published. However, by looking at some possible development opportunities, it was possible to see how the data could find further and more advanced use in the future.

Further to this, the implementation of the data platform was also retrospectively reviewed against the existing data platforms that had been examined. This ensured that there were no features which had been omitted or overlooked.

Future possibilities for automated processing

For importing area data, the geoprocessing model built relied on data which was already an area/polygon. For large-scale datasets, such as the US National Levee Database, this data was readily available and could be processed using the geoprocessing model documented before.

However, it was known that these datasets were unique. As reviewed by Özer *et. al.* (2020) the vast majority of databases consisted of extent data related only to the levee feature. Some such as DANTE only contained geospatial data on the start and end point of the levee feature.

If it were possible to supply a geoprocessing model with data on levee points, and the hydrological features that they referred to, it could result in the production of an automated levee line feature.

As this seemed to be a working methodology, Australia was selected to test a basic work-through on process. This was for the following reasons:

- Vector data at national and regional scale for surface water bodies was freely available[6] in several formats. This allowed testing of model function at variety of scales.
- Vector data at a regional level was available for embankment/levee features from each respective territory.
- A study area was found outside of the delta polygons at Rockingham[7] which contained a network of levees which were adequately covered by the national water body dataset.

The following conditions were defined for the automated process:

- Levees had to be located near a water course, natural or otherwise
- Rivers could be banked one side, or both sides, or not at all
- The levees could be made of different materials, constructed to different heights, or have differing attributes
- There could be multiple levee elements which constructed a singled leveed area
- Leveed areas should not cross into the water body

Generating leveed areas from the levee line data was hypothesised as the following operations, and the model could possibly be adapted to use line or point data and produce line or area output. This would have enabled the model to remain versatile for data from many different subsets.

[6] https://www.ga.gov.au/scientific-topics/national-location-information/national-surface-water-information
[7] Bounding Box: East:115.99°E, West:115.81°E, North: 32.16°S, South: 32.43°S

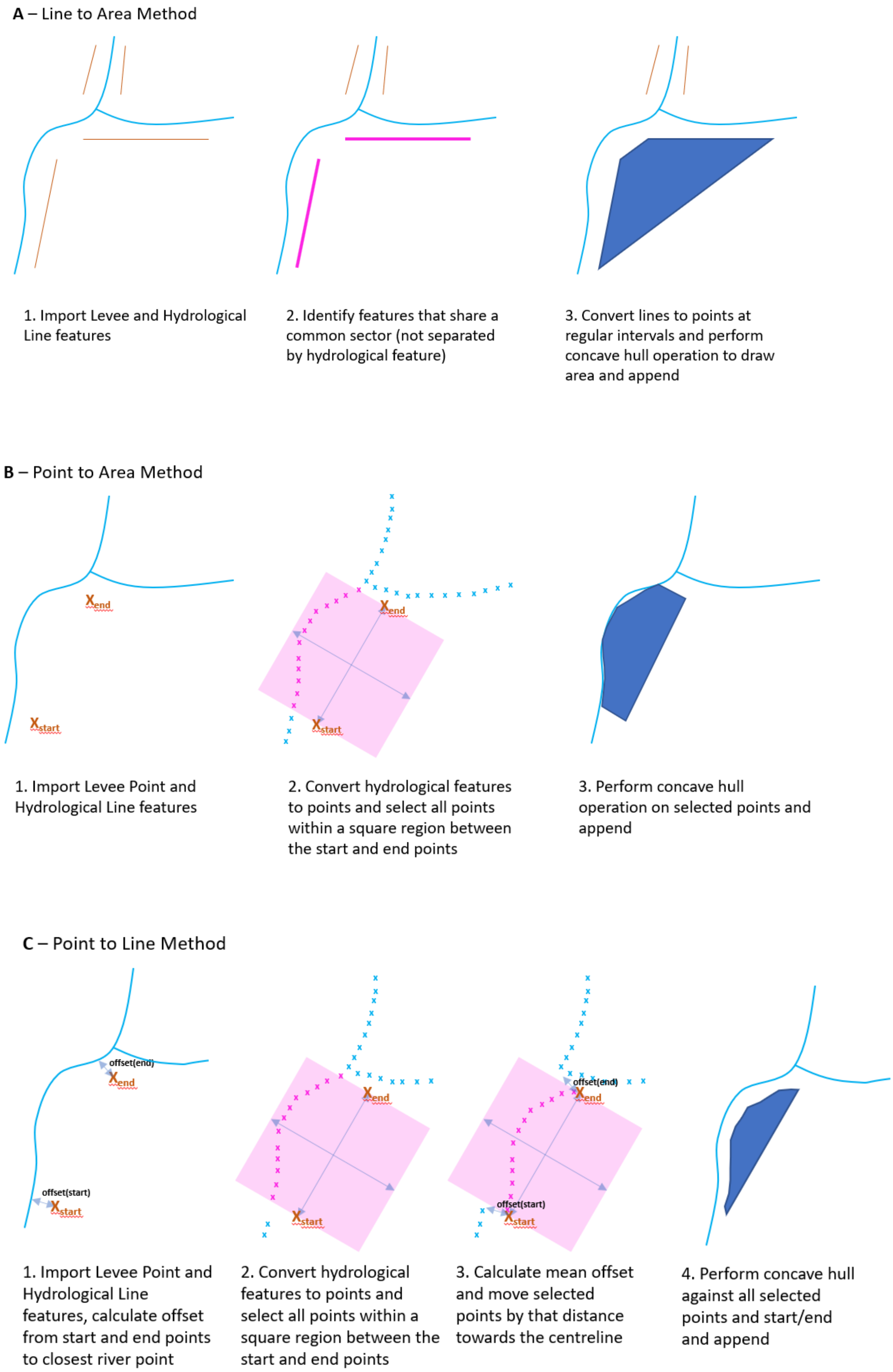


Figure 15 - Hypothetical processes for automated method

To confirm these hypotheses against real-world data, the previously discussed area at Rockingham was consulted. Using both levee and hydrological introduces a large possibility for error and as such, simply trusting the publisher was not a viable option. Different datasets are compiled for very different reasons, and the needs and the uses for this data varied.

Although at this stage this was for viability purposes only, it helped further interest into the practicality of the model and whether this was at all feasible.

Figure 16 (over page) indicates how data from similar sources can vary greatly. On the right is a national-scale dataset of waterways from Geospatial Australia[8]. On the left is the DWER-031[9] regional dataset from the Department of Water and Environmental Regulation of Western Australia.

By observing the area circled in yellow, the variance was seen – the waterway (light blue for national data, dark blue for regional data) much more closely followed the satellite imagery on a regional scale. There could be numerous reasons for these errors to be introduced:

- Data used as a national scale was created at a low resolution, or from lower resolution imagery. Metadata usually confirmed the accuracy and image source.
- Information on regional and local scale features, and changes made to them, may not have propagated sufficiently to the national level
- There may have been no need for data at this resolution (inter)nationally, or there may have been no budget to create or maintain it.
- Methods for analysing and processing the data will have varied from each institution and based upon the primary use of the data. Some datasets were point-driven polylines whereas others were traced outlines of features using satellite imagery.

Therefore, with due consideration for the variability of data within one country, collecting the data for this kind of task was extremely intensive and fell well outside of the remit of this project. And where this data was available, it was generally the same locations where detailed spatial datasets of levee information had already been made public.

However, this did not rule out further development of the problem. Indeed, combining this technique with other data (i.e. a high-resolution Digital Elevation Model [DEM] generated with LiDAR imagery) to identify levee extent in both length and width) could produce a high-quality output. Steinfeld *et. al.* (2013) presented a semi-automated method for identifying earthen levees on floodplains in Australia using a DEM. This was particularly relevant, as unregistered (possibly illegal) levees in Australia have been known to have an unknown impact on flood management (Brewster, 2020) and so being able to deduce them from remotely sensed data could have been greatly beneficial for active flood management.

[8] https://www.ga.gov.au/scientific-topics/national-location-information/national-surface-water-information
[9] https://catalogue.data.wa.gov.au/dataset/hydrography-linear-hierarchy

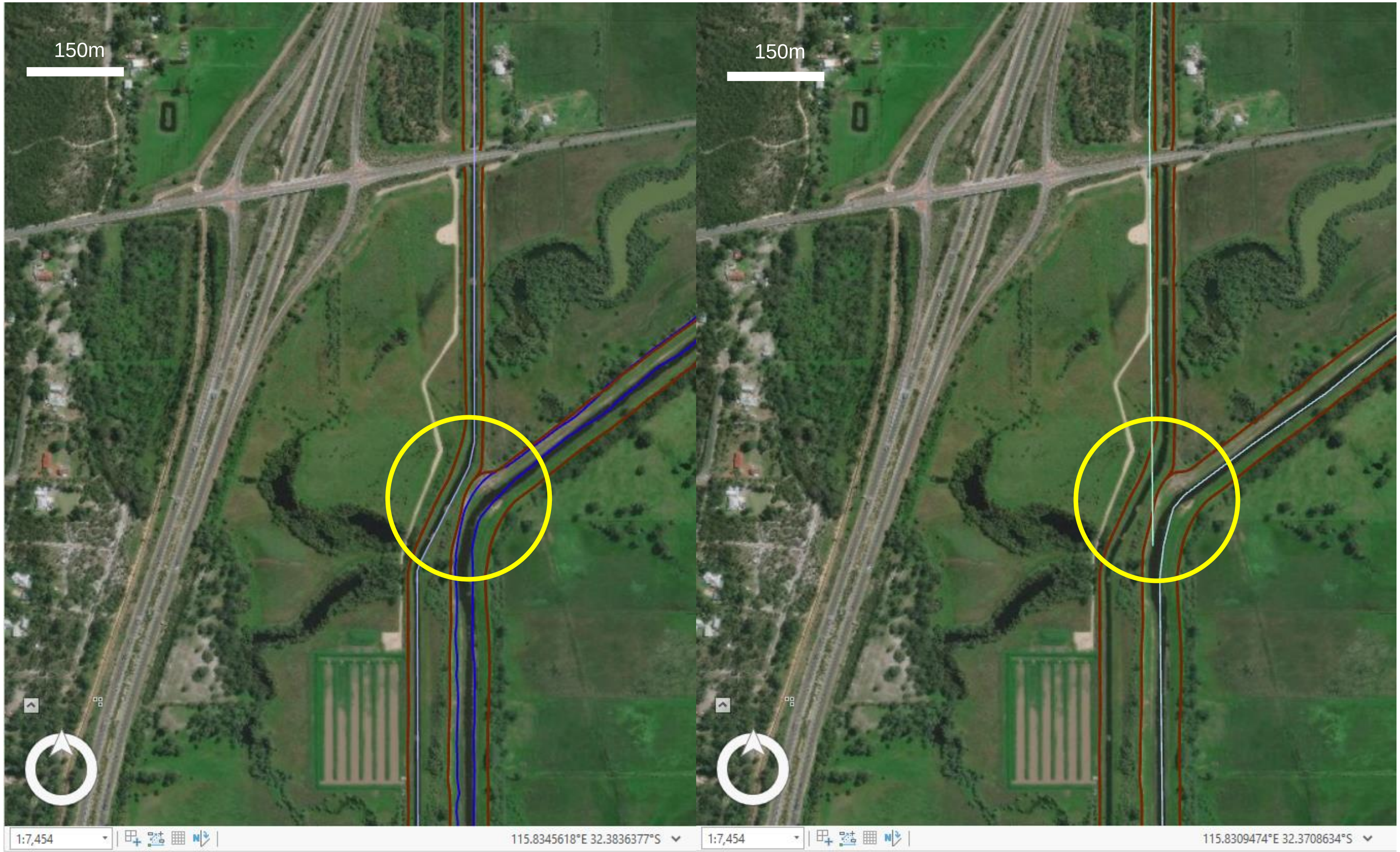


Figure 16 - Comparison of Australian hydrological data sources (Rockingham, WA, Australia)

Data Platform Review

Aside from the Data Platform Decision Matrix (Table 6, p.22) which reviewed the data platforms against the initially defined criteria, the information collected during the research of existing datasets was also pertinent to the choice of platform. By learning from the mistakes of the previous platforms (p.7), it was hoped that this dataset would not fall victim to the same errors.

Table 10 - Reflective review of data platform

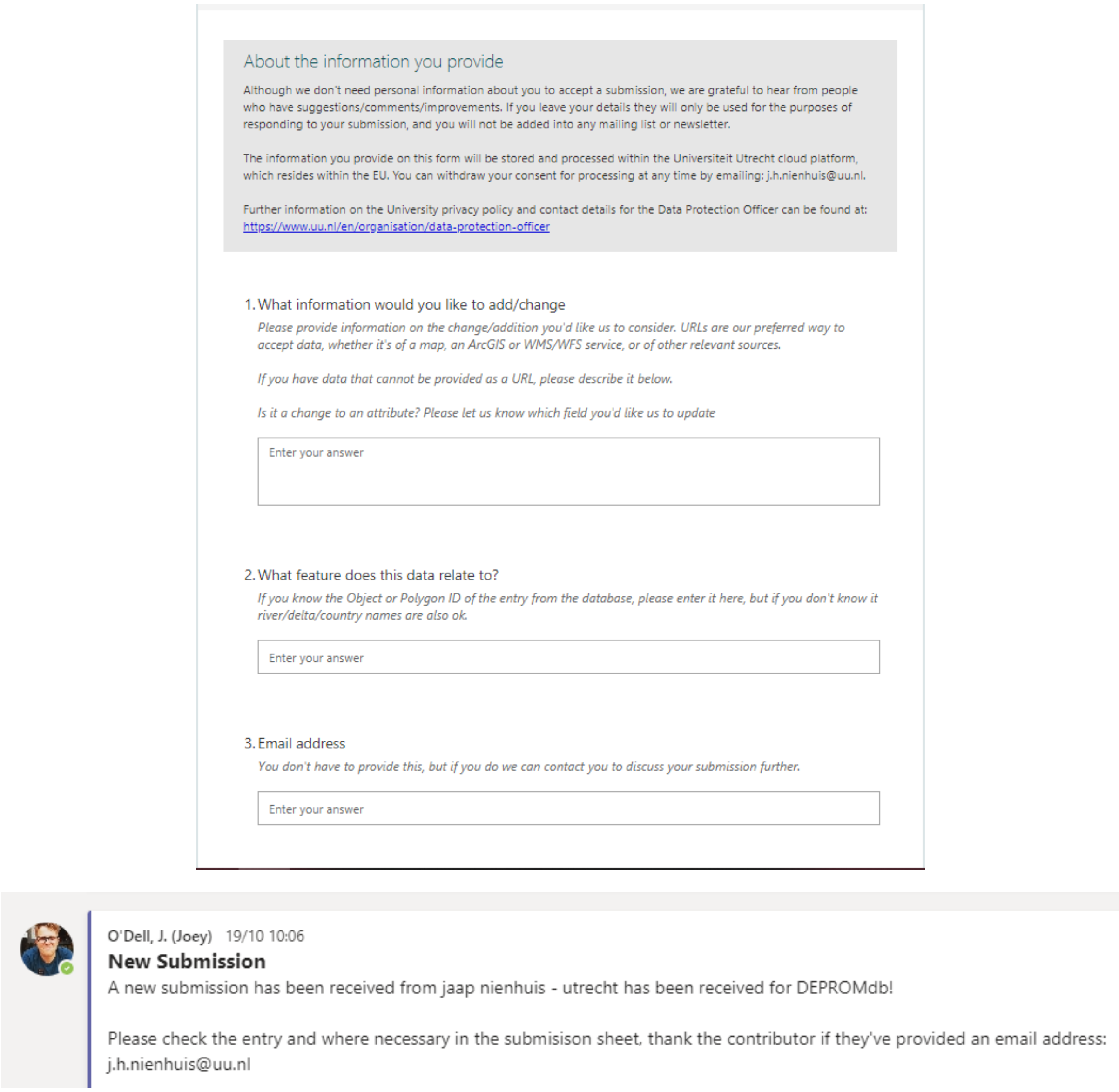
Reflection	Response to reflection
The data must be freely available, and a license that encourages re-use/onward publication should be used and clearly defined in the dataset/files	- Data released under CC-BY licence and a journal was selected that matched the licence - Licence clearly documented in metadata, in publication, and in file descriptions/README
Sources and reference should be both machine readable and intelligible to humans.	- Literary source provided in APA 7th and permanent DOI - Avoided use of any coding schema, instead used human-readable titles for attributes and file names, and descriptions encoded in metadata
Data should not be tied solely to the University or the specific publication, but should be available on a widely used service, including aggregators like DataCite.org	- Data published on ArcGIS Online, linked to a global ArcGIS geocatalogue, and DataverseNL, which are indexed to DataCite - ArcGIS Online controlled by the department, DataCite is controlled by DANS - No login required to access or download data
The data platform must be able to be updated easily, with the minimal number of copies (ideally one) of the data in existence	- ArcGIS Online and DataverseNL both supported versioning whilst holding static identifiers (URLs/DOIs) - The scientific journal submission was made with reference to the DOI - Submission form developed for freely accepting additions, alterations, and amendments
Metadata is just as important as the data itself	- Great care was taken to author detailed metadata to the INSPIRE standard - Additional supporting information on licenses, data curation, and a data dictionary were compiled and released in the dataset

U S E R - D R I V E N   C H A N G E

Accepting submission of source datasets was the first step to enabling continuous improvement, and through the integration of a submission form to the published material, it encouraged an ongoing effort to keep the database alive and valid. As uncontrolled edits were not a welcome concept in such a dataset (Özer, van Damme, & Jonkman, 2020) there needed to be a controlled manner for the public consumers of the data to report problems.

Discussions continued past the production of this report into source datasets from parties around the globe, but additionally the release of the dataset into a web viewing platform offered a unique opportunity to encourage edits and updates from viewers and users. The below form was triggered in the browsing window, but was also stated in the metadata, to encourage submissions. It had been coded so that it causes a notification workflow in Microsoft Teams for the project.

Additionally, designing such a form required careful consideration for the EU's General Data Protection Regulation (GDPR) so as to collect the least amount of data necessary to process the submission, and to give clear directions (as visible in the top section of the form) on the data subject's rights and ability to remove consent for processing. This is given in the figure below:



About the information you provide

Although we don't need personal information about you to accept a submission, we are grateful to hear from people who have suggestions/comments/improvements. If you leave your details they will only be used for the purposes of responding to your submission, and you will not be added into any mailing list or newsletter.

The information you provide on this form will be stored and processed within the Universiteit Utrecht cloud platform, which resides within the EU. You can withdraw your consent for processing at any time by emailing: j.h.nienhuis@uu.nl.

Further information on the University privacy policy and contact details for the Data Protection Officer can be found at: https://www.uu.nl/en/organisation/data-protection-officer

1. What information would you like to add/change

Please provide information on the change/addition you'd like us to consider. URLs are our preferred way to accept data, whether it's of a map, an ArcGIS or WMS/WFS service, or of other relevant sources.

If you have data that cannot be provided as a URL, please describe it below.

Is it a change to an attribute? Please let us know which field you'd like us to update

Enter your answer

2. What feature does this data relate to?

If you know the Object or Polygon ID of the entry from the database, please enter it here, but if you don't know it river/delta/country names are also ok.

Enter your answer

3. Email address

You don't have to provide this, but if you do we can contact you to discuss your submission further.

Enter your answer

O'Dell, J. (Joey) 19/10 10:06

New Submission

A new submission has been received from jaap nienhuis - utrecht has been received for DEPROMdb!

Please check the entry and where necessary in the submisison sheet, thank the contributor if they've provided an email address: j.h.nienhuis@uu.nl

Figure 17 - Submission form and notification workflow

5. Conclusion

The database had clearly highlighted the need for "joined-up thinking" in the world of levee documentation and research. Individual projects such as DANTE/INLED and the ILPD have formed to solve a specific task and so had made their data uniquely suited to the task, which made onwards reuse difficult, and therefore limited the benefit that the data could have had. No dataset is ever complete, and it was planned that amendments and ongoing updates to the database were to be data-driven.

Taking a closer look at the lack of unified data, it was possible to speculate many reasons why (lack of: infrastructure, funding, interest are a few examples) but thought also centres on the fact that levees have been considered a sensitive matter affecting public safety, and that by releasing information on the location of levees, information on their failures became easier to consume (Özer, van Damme, & Jonkman, 2020). Even when levee data was available, the area which it protects may not have been publicly available information, such as for the UK AIMS dataset and DANTE/INLED. A possible reason for not making "leveed area" data public was to reduce the influence of this upon the costs of insurance, as the National Levee Database has done so in the United States of America (National Research Council, 2013, pp. 68, Box 5-1).

Closing the system completely off to the public for access, such as that that was undertaken by the China Levee Project Information Management System (CLPIMS) in China, despite the extensive documentation of the system (Zhang, Ye, Shen, Mei, & Wang, 2018). Understandably keeping this data in a controlled environment meant that it could not be used for purposes that it was not intended. This also ensured that the data was constantly at the standard defined by the creator, and interpretations and onward uses were controlled. However, each dataset is siloed and became isolated; this did not mean that data which is outside of the remit of the curator was never included in the dataset. The example of Australia was again used, as the privately built levees had never been documented by the relevant departments as they were outside of the legislative remit of their function.

By producing a dataset that is compliant to open science and open data standards (i.e. INSPIRE, EUDAT FAIR) but that also followed creative and sharing licensing standards (Creative Commons), it is now possible to run more realistic modelling and climatic response scenarios, therefore improving the application and outcome of future research.

6. References

Barbetta, S., Camici, S., Maccioni, P., & Moramarco, T. (2015). National Levee Database: monitoring, vulnerability assessment and management in Italy. *EGU General Assembly 2015. 17*, pp. EGU2015-10170. Vienna: Geophysical Research Abstracts. Retrieved from https://meetingorganizer.copernicus.org/EGU2015/EGU2015-10170.pdf

Berkley Library, University of California. (2020, July 9). *How to Write a Good Documentation*. Retrieved September 7, 2020, from Managing Research Data: https://guides.lib.berkeley.edu/how-to-write-good-documentation

Brewster, K. (2020, May 28). NSW water officials knew decades of unmeasured floodplain harvesting by irrigators was illegal. *The Guardian*. Retrieved November 1, 2020, from https://www.theguardian.com/australia-news/2020/may/29/nsw-water-officials-knew-decades-of-unmeasured-floodplain-harvesting-by-irrigators-was-illegal

Childs, C. (2009). The Top Nine Reasons to Use a File Geodatabase. (M. Pratt, Ed.) *ArcUser*(Spring 2009), pp. 12-15. Retrieved August 12, 2020, from https://www.esri.com/news/arcuser/0309/spring2009.html

Directorate-General for Environment (European Commission). (2004). *Living with Coastal Erosion in Europe - Sediment and Space for Sustainability (Catalogue number KH-59-04-952-EN-C).* (P. Doody, M. Ferreira, S. Lombardo, I. Lucius, R. Misdorp, H. Niesing, . . . M. Smallegange, Eds.) Luxembourg: Office for Official Publications of the European Communities. Retrieved from https://op.europa.eu/s/ooik

EADS S&DE. (2003). *EUROSION D2-3-3 Dataset Structure Description V3-3*. Tolouse: Publication Office of the European Communities. Retrieved October 10, 2020, from https://www.eea.europa.eu/ds_resolveuid/EDT59CHNA7

Edmonds, D. A., Caldwell, R. L., Brondizio, E., & Siani, S. (2020). Coastal flooding will disproportionately impact people on river deltas. *Nature Communications, 11*, 4741. doi:10.1038/s41467-020-18531-4

Eliander, D., Couasnon, A., Ikeuchi, H., Muis, S., Yamazaki, D., Winsemius, H. C., & Ward, P. J. (2020). The effect of surge on riverine flood hazard and impact in deltas globally. *Environmental Research Letters, 15*, 104007. doi:10.1088/1748-9326/ab8ca6

European Commission. (2016, June 28). EUROSION - Geomorphology, Geology, Erosion trends and Coastal defence works (DAT-62-en). (ver2.1). European Environment Agency. Retrieved October 2020, from https://www.eea.europa.eu/ds_resolveuid/DAT-62-en

Federal Geographic Data Committee. (n.d.). *Selecting a Geospatial Metadata Standard*. Retrieved August 28, 2020, from https://www.fgdc.gov/metadata/selecting-a-geospatial-metadata-standard

Jones, S., & Grootveld, M. (2017, November 24). *How FAIR are your data?* doi:10.5281/zenodo.1065990

Kovács-Láng, E. (2000, November 4). Metadata and their Importance to Information Technology. Vácrátót, Hungary: UN-FAO (Food and Agriculture Organization of the United Nations).

Retrieved August 18, 2020, from GTOS Central and Eastern Europe Terrestrial Data Management and Accessibility Workshop: http://www.fao.org/3/ac665e/ac665e08.htm

National Research Council. (2013). Insuring Properties Behind Levees. In *Levees and the National Flood Insurance Program* (pp. 63-96). Washington, D.C., United States of America: The National Academies Press. doi:10.17226/18309

Nienhuis, J. H., Ashton, A. D., Edmonds, D. A., Hoitink, A. J., Kettner, A. J., Rowland, J. C., & Törnqvist, T. E. (2020). Global-scale human impact on delta morphology has led to net land area gain. *Nature, 577*, 514-518. doi:10.1038/s41586-019-1905-9

Özer, I. E., van Damme, M., & Jonkman, S. N. (2020). Towards an International Levee Performance Database (ILPD) and Its Use for Macro-Scale Analysis of Levee Breaches and Failures. *Water, 12*(1), 119. doi:10.3390/w12010119

Scussolini, P., Aerts, J. C., Jongman, B., Bouwer, L. M., Winsemius, H. C., de Moel, H., & Ward, P. J. (2016). FLOPROS: an evolving global database of flood protection standards. *Natural Hazards and Earth System Sciences*(16), 1049-1061. doi:10.5194/nhess-16-1049-2016

Steinfeld, C. M., Kingsford, R. T., & Laffan, S. W. (2013). Semi-automated GIS techniques for detecting floodplain earthworks. *Hydrological Processes, 27*(4), 579-591. doi:10.1002/hyp.9244

World Bank. (2013). *Report No: 74482-BD - Project Appraisal Document [..] To the People's Republic of Bangladesh for a Coastal Embankment Improvement Project Phase-I.* Disaster Risk and Climate Change Unit. World Bank. Retrieved July 31, 2020, from http://documents1.worldbank.org/curated/en/748531468209052823/pdf/744820PAD0P1280disclosed060701300SD.pdf

Wright, D. (2018, October 8). *Making Story Maps Citable (e.g., with Digital Object Identifiers)*. (ESRI, Inc.) Retrieved September 5, 2019, from ArcGIS Blog: https://www.esri.com/arcgis-blog/products/story-maps/sharing-collaboration/making-story-maps-citable-e-g-with-digital-object-identifiers/

Zhang, B., Ye, Y.-w., Shen, X.-z., Mei, G., & Wang, H.-x. (2018). Design and implementation of levee project information management system based on WebGIS. *Royal Society Open Science*, 180625. doi:10.1098/rsos.180625

7. Appendices

Appendix I: Existing Dataset Reviews p. 41

Appendix II: ESRI Academy Transcript p. 51

Appendix III: Manual Data Collection Flowchart p. 52

Appendix IV: Achievement of Research Objectives p. 53

Appendix V: Achievement of Educational Objectives p. 54

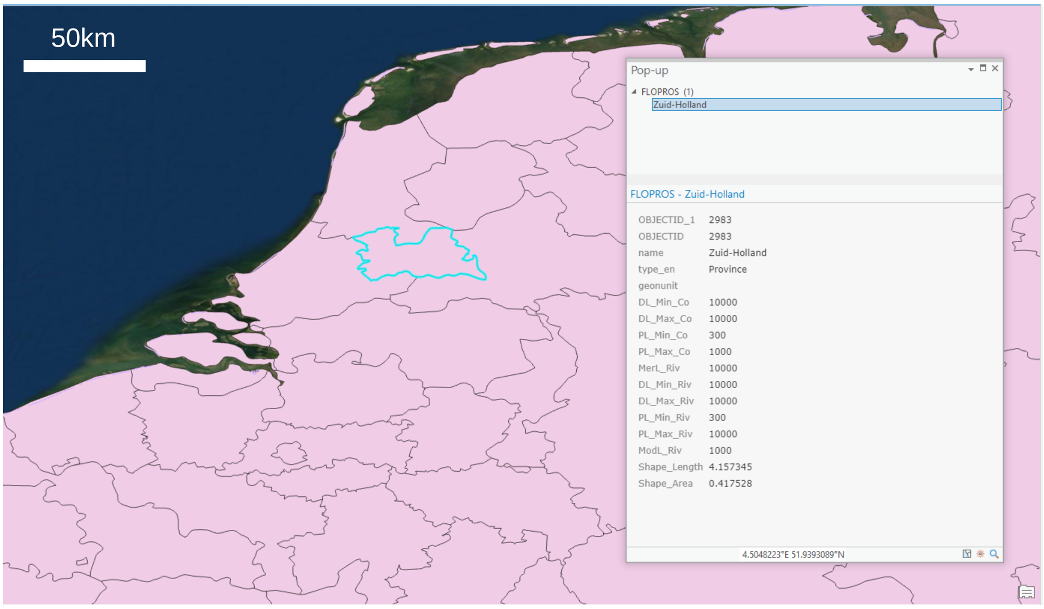
Appendix I: Existing Databases Reviews

F L O P R O S

Site/Resource Link: doi:10.5194/nhess-16-1049-2016
Licence: CC-BY 4.0 International

The database of **FLO**od **PRO**tection **S**tandards (*FLOPROS*) presented by Scussolini *et. al.* (2016) although not itself a database of levees, does include three important layers of flood management dat. The dataset was mostly used as a reference against data release and format, instead of being source material for the new database. The layers in the FLOPROS dataset are:

- **Policy Layer** Regulatory and (local) governmental policy on flood protection
- **Design Layer** Data on physical flood protections for a given area
- **Model Layer** Calculation of level of protection through validated modelling



Overview of FLOPROS dataset – Netherlands/Benelux Region - Data/shapefile from Scussolini et. al. (2016)

The data within FLOPROS is aggregated by NUTS 2[10] level as a polygon. An example in the figure above is the data for Provincie Utrecht, Netherlands (NUTS 2: NL31).

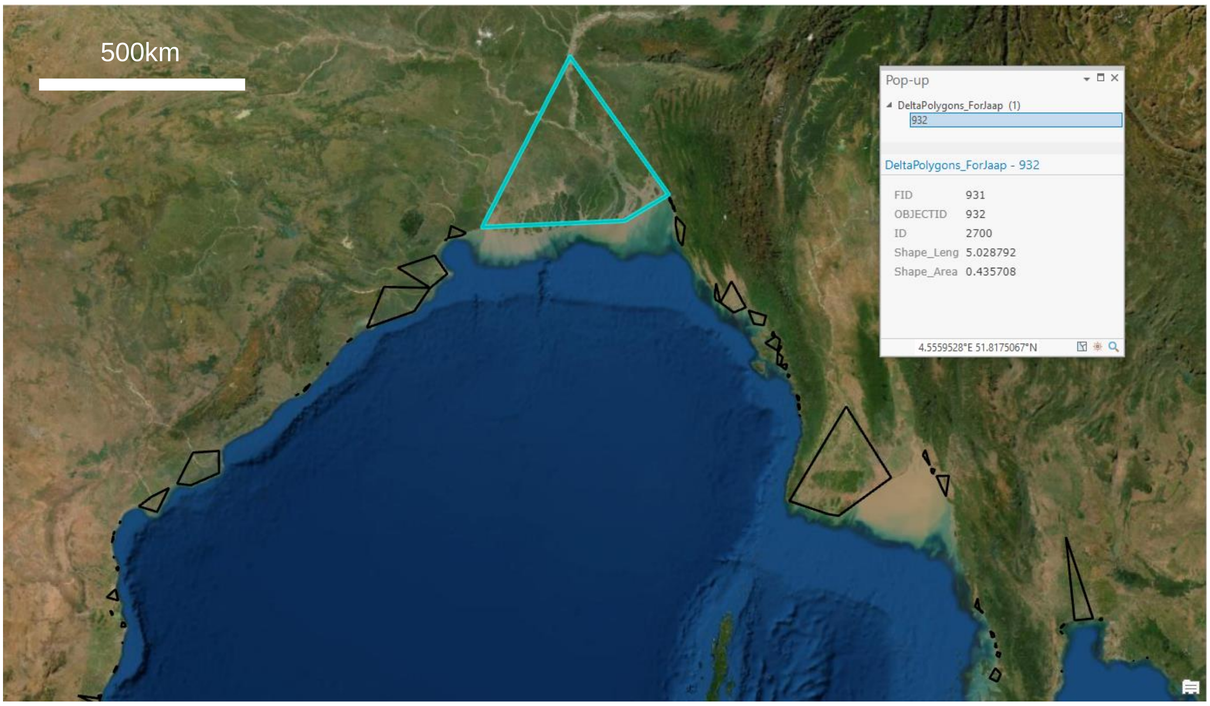
Benefits	**Limitations**
Global coverage, rich dataset on flood return period protection	No actual information on flood defences or areas protected by flood defences
Simple dataset that is easy to load, and consists of one data layer with the merged attributes	Data can only be processed spatially, no defining characteristics that can be interpreted from attributes only
Data itself is easy to update and can be released a shapefile	Method of distribution makes revisions difficult to release, as journal holds 'snapshot' of data that cannot be overwritten or updated
Data is published in a journal that is indexed to various data sources (such as DataCite.org)	No easy way to add contributions, must find email address of author

[10] NUTS – Nomenclature of territorial units for statistics – EU-level analysis of country subdivisions. Level 2 = region/provincial level. More information: https://ec.europa.eu/eurostat/web/nuts/background

D E L T A   P O L Y G O N S

Site/Resource Link: [doi:10.1038/s41467-020-18531-4](https://doi.org/10.1038/s41467-020-18531-4)
Licence: CC-By 4.0 International *(Open Access Publication)*

Edmonds, Caldwell, Brondizio, & Siani (2020) ratify previous discourse on the topic of a replicable definition of delta area, and produced a dataset of c. 2,000 delta 'polygons', which have vertices hand drawn from satellite imagery to define the lateral extents of the deltas, including historical deltas. This dataset is not to be taken as a verbatim definition of the existence or absence of a delta, however by following a rigid and repeatable method, has produced a clearly definable starting point for analysis of delta sedimentation analysis.



Delta Polygons Dataset – Ganges-Brahmaputra Delta, Bangladesh - Edmonds et. al. (2020)

The example in the figure above of the Ganges-Brahmaputra Delta shows the existence of the larger delta polygon for the main delta. The extents (vertices) match the satellite imagery, but as discussed, the lines drawn between the vertices do not follow coastline/delta extent.

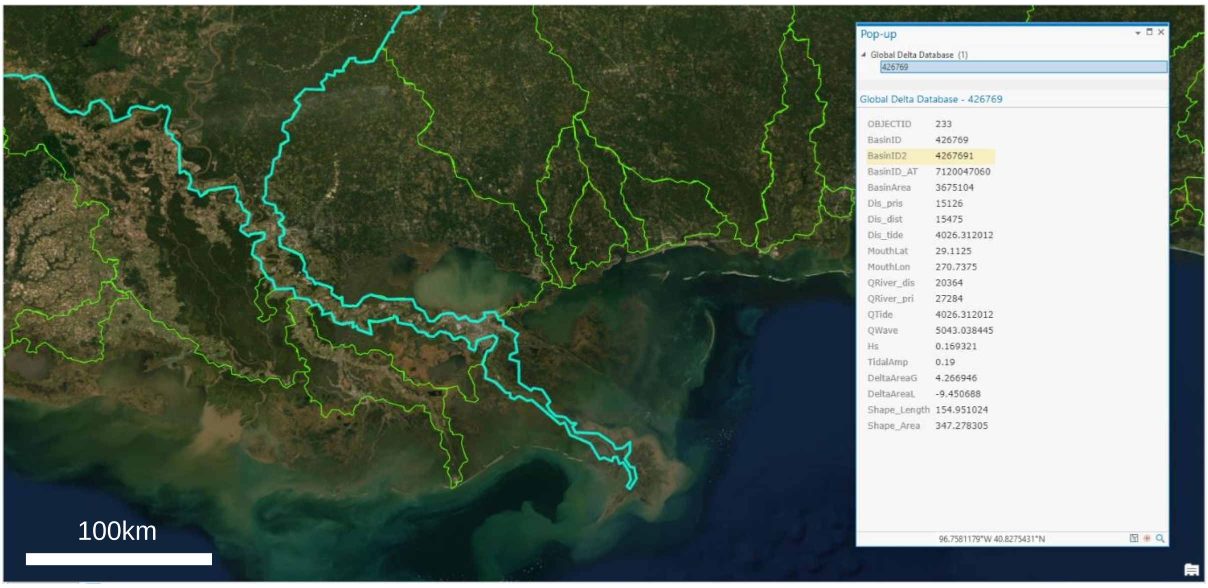
However, the global extent of this dataset allows it to be used as a focus point for the initial data collection exercise – so to best prioritise time and effort. However, these polygons are not intended to be the sole focus of this databases extent. Initial work will focus within a boundary of these shapes, but as the dataset is contributed to in the future, levees outside of the areas can also be included to expand the application of the dataset. This also allows for revisions to the polygons to be included in future work.

Benefits	**Limitations**
Global coverage of delta area using a clearly defined and ratified method	Levee data is not included in this dataset
Dataset is not confined by regional/territory grouping as FLOPROS and cross geopolitical boundaries.	The shape of the polygon does not directly represent the shape of the delta, moreover its clearly defined extents.

GLOBAL DELTA CHANGE DATABASE

Site/Resource Link: [doi:10.17605/OSF.IO/S28QB](doi:10.17605/OSF.IO/S28QB)
Licence: CC-BY 4.0 International

This database, which is supplementary to the paper by Nienhuis *et. al.* (2020) forms a global body of delta basins and sub-basins. This also then presents delta modelling attributes as discussed in the accompanying paper that can be used to model flooding events and tidal input.



Global Delta Change Database – Mississippi Delta, USA - Nienhuis et. al. (2020)

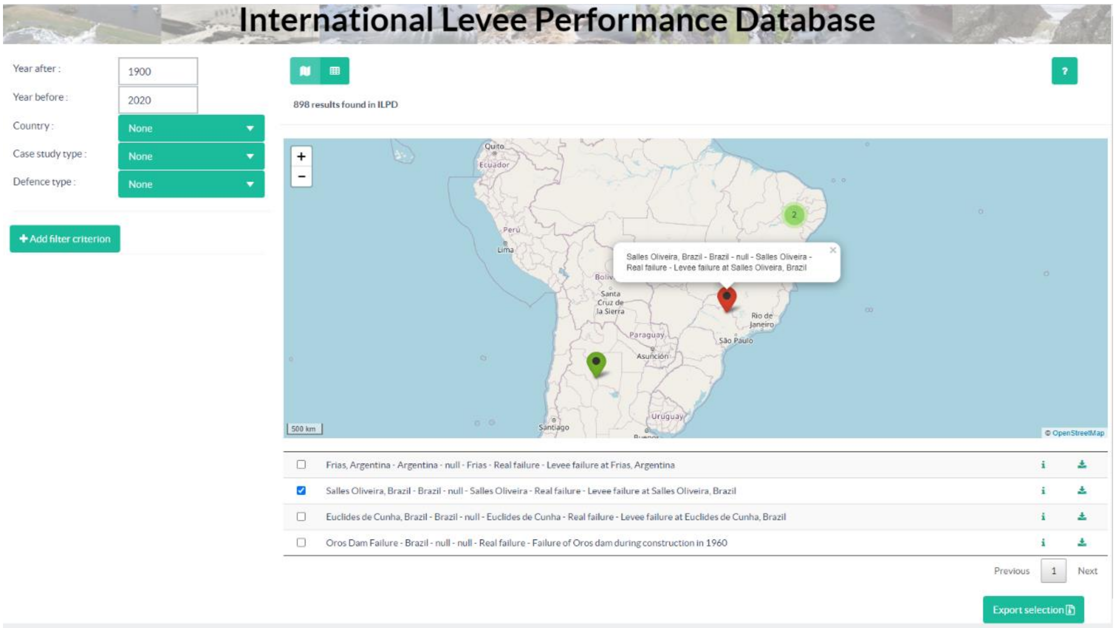
This dataset is of limited use for the data collection aspect of the project but is useful for onward linking and integration of the dataset to encourage maximum re-use of the data.

Benefits	**Limitations**
Global coverage	Levee data is not included in this dataset
Presence of delta modelling characteristics	Mouth extent data is single point coordinate
Assigns a globally unique ID to each (sub-)delta	No metadata or readme published with the data and so difficult to interpret for someone who is not directly involved in the project.
Extensive sub-delta formation results in situations where proximal levee features resolve to a different sub-delta.	Boundary edges (i.e. coastlines) are fixed at time of release and will need updating as these change.
Data is published in an indexed data repository (OSF) as well as an academic journal	

SAFELEVEE / ILPD

Site/Resource Link: https://leveefailures.tudelft.nl/
Licence: No database right asserted

Özer, van Damme, & Jonkman (2020) put forward an International Levee Performance Database (ILPD) as part of the SAFELevee project at TU Delft. This consists of a website where information on levee failure and testing events is mapped and information is downloadable as point data however it does not show spatial extent or limit of levee features and coverage is sporadic.



International Levee Performance Database web interface – South America - Özer et. al. (2020)

Unfortunately, although the website displayed information properly, the referencing/download system was completely inaccessible and generated empty folders. Additionally, there is no classification system for data quality and so some data points were of considerable age but were still included at the same relevance as data from the current century.

Furthermore, the website states that the data is made "*publicly available online*" and that "*The ILPD is not owner of the data*" but fails to apply any form of licensing, open or otherwise which could be seen as a barrier to re-use of the data.

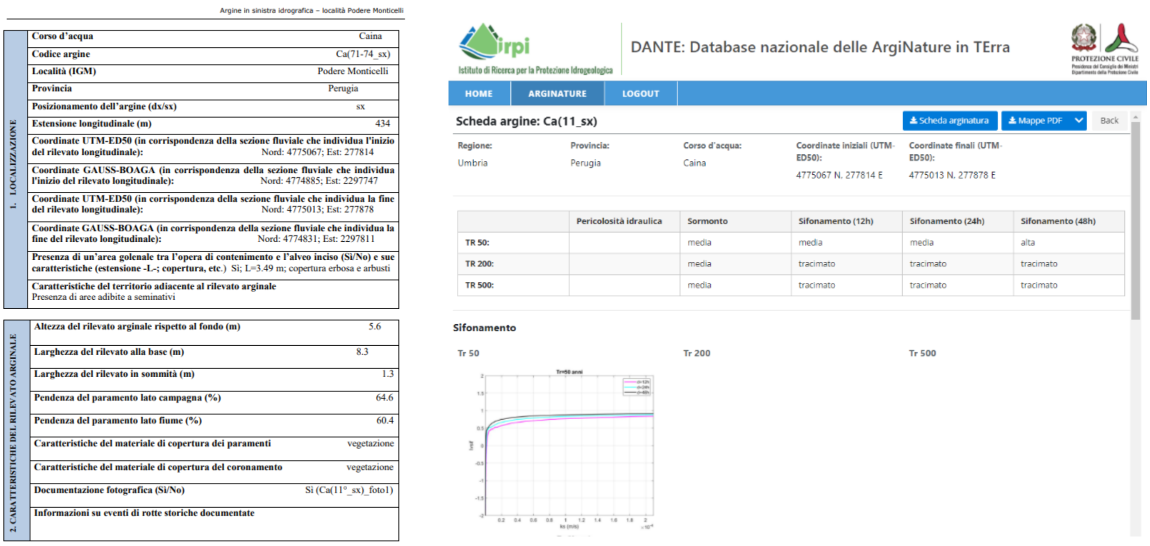
Benefits	**Limitations**
Rich dataset which can be plotted for information on levee failures	SAFElevee projected closes end 2020 and maintenance is not likely to continue past then.
Extensive information across temporal ranges about levee existence/management	Focus is on dam (i.e. inland) levees and mostly on failure only, whereas this project is focussed on coastal levees initially.
Professional website with good interaction potential if maintained properly	Data is far from complete as discussed in the paper announcing the dataset, and some data accuracy/assurance tasks do not appear to be have been performed.
Produced in collaboration with industry and the International Commission on Large Dams (ICOLD)'s Levees and Flood Defences group (LFD-EurCOLD, https://lfd-eurcold.inrae.fr/)	Data download is difficult for someone who is not directly involved with the project to interpret and use, and information on source works is hard to decipher.

D A N T E

Site/Resource Link: https://dante.irpi.cnr.it/
Licence: Data not freely available, access upon request only

The *Database nazionale della AgriNature in TErra* (DANTE) is an actively developing national levee database for Italy. Previously known as the *Italian National LEvee Database* (INLED) when first proposed by Barbetta, Camici, Maccioni, & Moramarco (2015), this database has since grown to include additional regions of Italy. The dataset contains:

- Start and End location of the levee features in UTM-ED50 eastings/northings
- Data on construction type, length, and land usage categories
- River, region, and system name



DANTE/INLED web interface - Barbetta et. al. (2015)

The figure above shows the output from DANTE. On the left is a documentary output of an embankment system, and on the right is the web interface to each individual levee element. The database is still in active development Return period (Tr) data is included as a graphical representation and a textual indicator of response (average, overflowed).

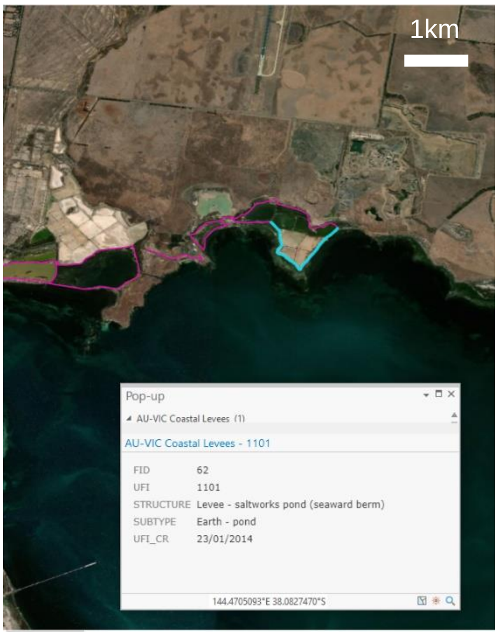
Benefits/Limitations (DANTE/INLED)

Benefits	Limitations
Incredibly rich dataset with ongoing maintenance	Only published in Italian, reliance upon external translation tools to use the database where Italian is not spoken
Slick, easy to use interface with user control/granularity of view/edit functions.	Coverage still limited to certain areas where the IRPI has data. Not fully national coverage.
Interface is being extended to include maintenance and hazard data, increasing usability and cross-discipline applicability.	Start and end point of levees do not allow accurate mapping of levee extent.
Active development and management by a research body of the Italian Government.	Data only accessible upon request/licensing, not freely published.

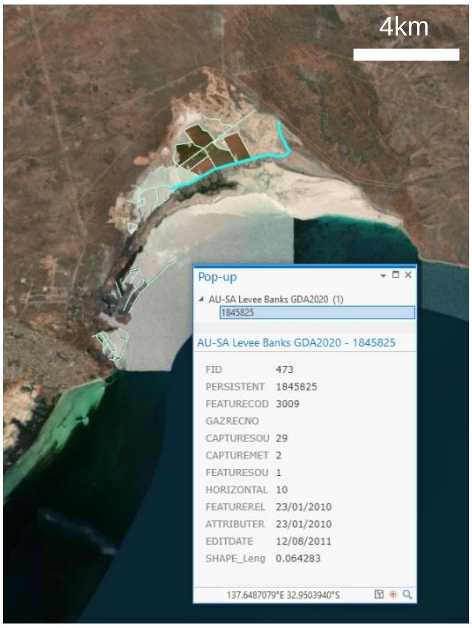
AUSTRALIAN OPEN DATA

The individual regions of Australia publish information on levees using a common geocatalogue format per state. Although these are held by the respective state, they are covered by a general directive from the central government[11] to make this data open and accessible.

State	Dataset	Site/Resource Link	Licence
WA	DWER-021	https://catalogue.data.wa.gov.au/tr/dataset/fpm-levee-banks	Local
SA	Levee Banks	https://data.sa.gov.au/data/dataset/levee-banks	CC-BY
NSW	DLS Area	https://data.nsw.gov.au/data/dataset/1-7c5fe5ae67604a348f0c95847224fa66	CC-BY
VIC	Coastal Levees	https://discover.data.vic.gov.au/dataset/coastal-levees	CC-BY



A



B

Australian Levee Datasets - A) Victoria Coastal Levees (Avalon, VIC), B) South Australia Levee Banks (False Bay, SA)

Benefits	Limitations
Datasets are constantly maintained by the government departments and are generally freely available	Each individual dataset is released by the relevant governmental body despite the existence of a national data catalogue
As data is created at a state level, the data is generally of very good quality and it is possible to directly contact the team responsible for managing the data	Data is not aggregated in a central platform, and so each dataset must be accessed and processed.
Availability of a state data catalogue makes locating additional information easy and enables further use of data	Licensing for the WA (DWER-021) dataset is confusing and required additional consultation with the data custodian
	Lack of national approach means different datasets have different attributes, schemas, methodology, and quality

[11] https://www.pmc.gov.au/public-data/open-data

Site/Resource Link: https://data.gov.uk/dataset/cc76738e-fc17-49f9-a216-977c61858dda
Licence: Open Government Licence (OGL) 3.0[12] *(CC-BY Compatible)*

The Environment Agency publishes the Asset Information Management System Spatial Flood Defences dataset which uniquely contains information not only on directly owned and controlled assets, but also on assets which are managed or inspected by the agency.

A

objectid	62409
PROT_TYPE	coastal
LENGTH	1591.6
shape	Polyline
st_lengthshape	1591.60084
asset_id	8387
frms_id	2516
asset_ref	054CBSEAP0304C03
asset_name	Happisburgh to Winterton Sea Wall - Crinkle Hill to Bramble Hill
sub_type	wall
pri_purp	flood_risk_management
sec_purp_1	flood_risk_management
sec_purp_2	flood_risk_management
toe_level	2.41
psra_site	yes
asset_main	environment_agency
asset_oper	environment_agency
asset_own	environment_agency
act_cond	2

B

Environment Agency AIMS – Winterton-on-Sea, Norfolk, UK - A) Feature Layer, B) Item Attributes

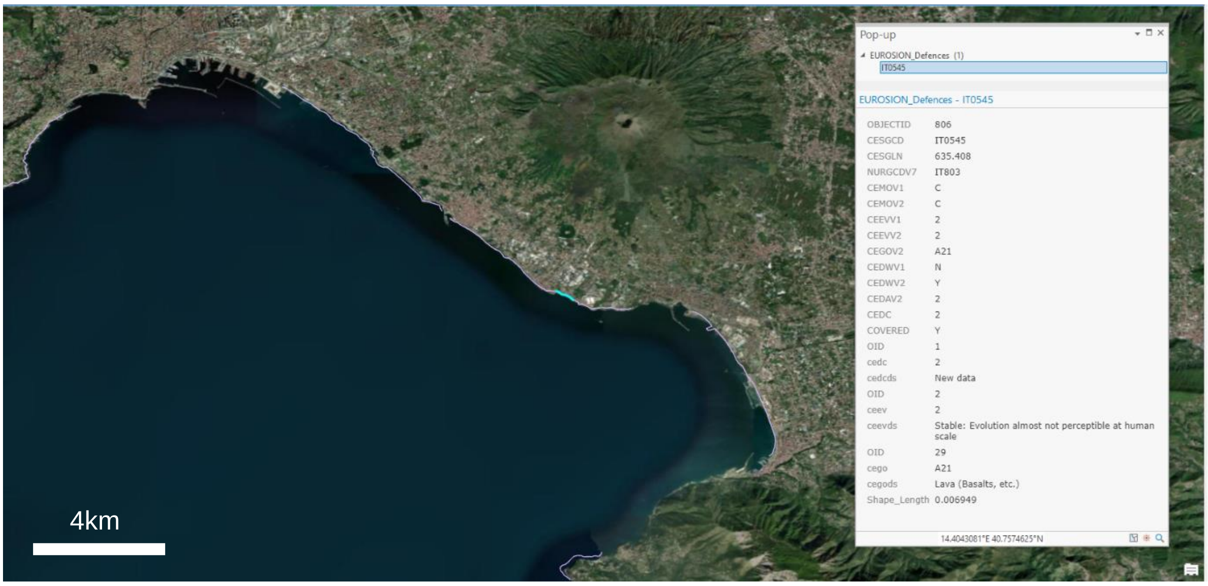
Benefits	**Limitations**
Spatial data of high quality with detailed attributes on feature and type of defence	Despite being on a UK data platform, it only covers the agency's remit in England.
Data extends beyond directly owned/managed assets to also include those which are (or should be) inspected by the agency	Attribute data is in local format and difficult to interpret if not familiar with dataset.
Licensed under the Open Government Licence 3.0 and so is freely reusable	Streaming service restricted functionality (not possible to select features, cannot query attributes directly as held in a separate service)
Regularly updated streaming data set and provided with detailed metadata and point of contacts	

[12] Open Government Licence: http://www.nationalarchives.gov.uk/doc/open-government-licence

E U R O S I O N

Site/Resource Link:	https://www.eea.europa.eu/data-and-maps/data/geomorphology-geology-erosion-trends-and-coastal-defence-works
Licence:	CC-BY 4.0 International

From the European Union's EUROSION project (Directorate-General for Environment (European Commission), 2004), analysis of the coastline of the European Communities as performed and this resulted in a database of coastal features including attributes on whether they actively protected/defended, the material they were composed of, and other key data attributes. This dataset was further revised in 2006 and 2016 to result in EUROSIONv2.1 (European Commission, 2016)



EUROSIONv2.1 dataset output showing joined attributes – Bay of Naples, Italy - European Commission (2016)

The data was initially created 2002-2004 (v1), and then revised in 2006 (v2) with small corrections to the projection data in 2016 (v2.1) and as such the data is aged. There have been changes to EU state membership since 2004 which means this no longer covers the entire continental landmass. Therefore, it is best used as a baseline where information is not possible to locate after extensive internet and literature review. The nature of the data being constrained only to coastal features restricts leveed area inference, as coastal works may not cover the entire land area.

Benefits	**Limitations**
Pan-European dataset covering all coastlines	Data only covers coastal interface/shoreline and so is limited to 'presence/absence'
Spatial dataset in vector format with rich attribute tables on material, defence, erosion characteristics etc.	Age of data means that changes since 2002-2004 are unlikely to be fully incorporated despite 2016 update (v2.1)
European Environment Agency open data published under CC-BY licence.	Some countries were not members of the EU in 2004 and so are not present.
Backwards-referenceable data with detailed documentation and references	Data requires a considerable amount of work to make functional in a modern GIS
	Confusing and difficult to navigate file structure made data inaccessible

From reviewing the EUROSION dataset (p.48) it was clear that the method utilised was extremely complex, and resulted in a dataset that required its own 104-page document explaining the data structure. This resulted in a very powerful database, but one that was difficult to use despite the extensive documentation. Considerable amounts of information that would now be encoded in metadata or data attribute fields were encoded into the file names using a local schema.

Example of EUROSION database complexity (EADS S&DE, 2003)

4.2.1.1 Overview of the database structure

CLEUER100kV1

Within the framework of EUROSION project, to follow the GISCO naming conventions, the name proposed for the EUROSION shoreline coverage is: CLEUER100kV1, which means:

- CL for CoastLine (layer name)
- EU for EUrope (georeference)
- ER for EuRosion (source)
- 100K for scale (scale)
- V1 version

4.2.1.2 Description of the Arc Attribute Table of the CLEUER100kV1 coverage

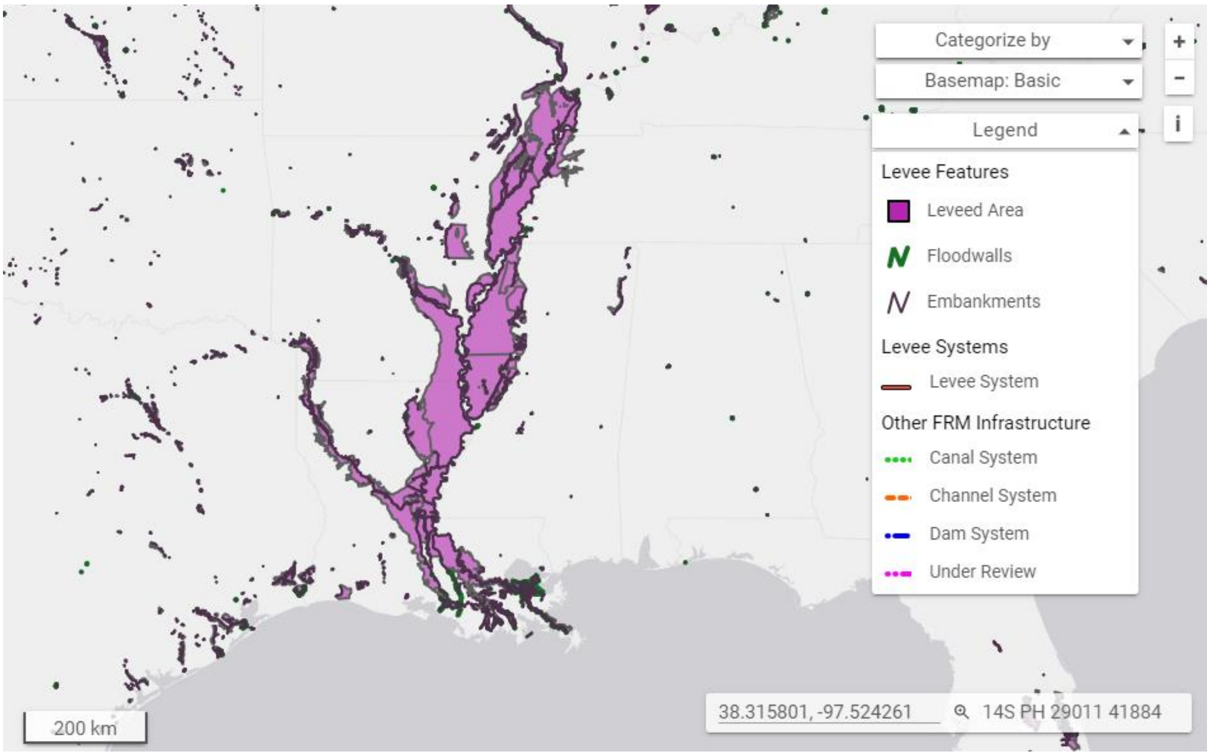
Attribute Name	**Type**	**Description**
FID	OID	Unique Object Identifier
Shape	Geometry	Polyline
CLCESGCD	String	Coastal Erosion SeGment CoDe. This is the identifier of every coastal segment. It is composed of 2 letters representing the country followed by a sequential number.

The above was a key factor in deciding the schema and structure for the database as described in the main body of the report. However, it is important to note that the INSPIRE European metadata directive did not exist at the time of the database creation and so it is likely that most of the documentation created then could now be included as a metadata file or within the dataset itself. Furthermore, advances in computing technology since EUROSION was published in 2005 are likely to have rendered some of the norms and processes of the time obsolete.

U S A C E   N A T I O N A L   L E V E E   D A T A B A S E

Site/Resource Link: https://levees.sec.usace.army.mil/
Licence: Public Domain

The USACE (US Army Corps of Engineers) National Levee Database (NLD) is aggregated from municipal and governmental contributors to derive a single layer for levee features (lines) and levee areas (polygons) which can be accessed in a number of streaming formats. The data is publicly available and as US governmental data is within the public domain, so is freely re-usable. Attributes are homogenised and metadata is published which supports the data and the services provided.



USACE NLD Leveed Areas and Levee Features web viewer – Mississippi, United States – USACE NLD (2020)

Given the excellent quality and regular maintenance of the dataset (the updates feed identifies on average one update per month) it is to be considered the gold standard in data dissemination for this kind of activity, and is actively used and reused across the academic and professional community. This includes the impact the data has on house insurance prices in the US through defining leveed and non-leveed areas (National Research Council, 2013).

Benefits	Limitations
US-wide dataset containing levee lines and leveed areas	Only available as a streaming service, cannot be downloaded
Detailed construction attributes including levee height, material, year of construction etc.	Data collated from various sources, impossible to query data quality
Open access website with multiple data sources, including streaming via many standards, and ArcGIS Service (SDE)	
Homogenous and comparable attributes across the entire country	

Appendix II: ESRI Academy Transcript

My Learning Transcript

Joey O'Dell

Created October 19, 2020

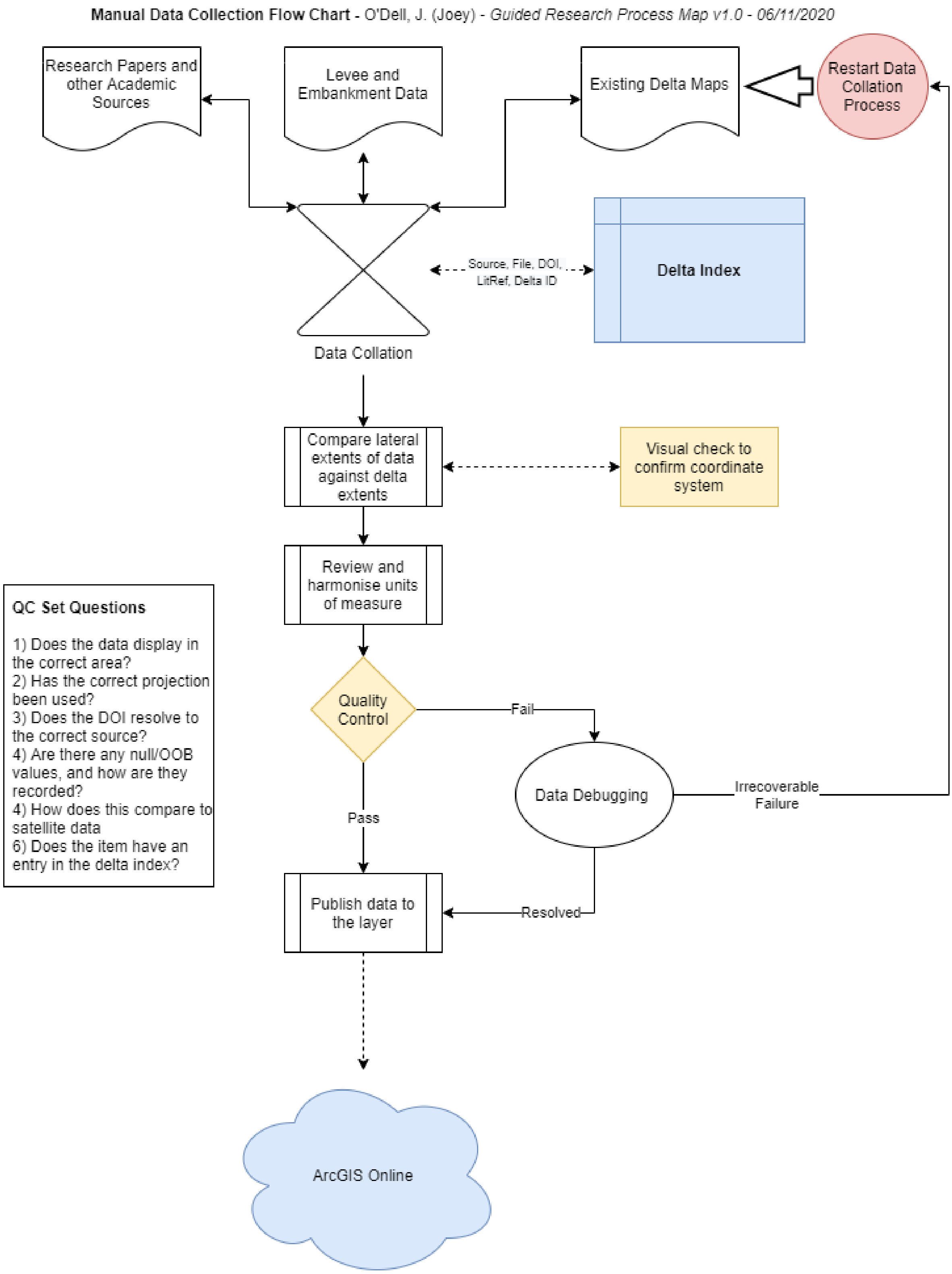
2020

ArcGIS Online: Administration Basics	Video	Watched September 20, 2020	Score: 100
Changing Technology: Real-Time/Live Data	Video	Watched September 20, 2020	Score: 100
ArcGIS Online: Sharing Basics	Video	Watched September 20, 2020	Score: 100
ArcGIS Online: Analysis Basics	Video	Watched September 20, 2020	Score: 100
ArcGIS Online: Data Basics	Video	Watched September 20, 2020	Score: 100
Creating and Sharing GIS Content Using ArcGIS Online	Web Course	Completed September 20, 2020	Score: 80
Getting Started with ArcGIS Pro	Web Course	Completed September 19, 2020	Score: 100
Understanding Spatial Relationships	Web Course	Completed September 19, 2020	Score: 100
Introduction to Spatial Data	Web Course	Completed September 19, 2020	Score: 100
Getting Started with Geoprocessing	Web Course	Completed September 5, 2020	Score: 100
Getting Started with the Geodatabase	Web Course	Completed September 5, 2020	Score: 100
Basics of Geographic Coordinate Systems	Web Course	Completed August 30, 2020	Score: 80
Building Geoprocessing Models Using ArcGIS Pro	Web Course	Completed August 30, 2020	Score: 90
GIS Basics	Web Course	Completed August 30, 2020	Score: 100
Creating and Editing Metadata in ArcGIS	Web Course	Completed August 30, 2020	Score: 100

Page 1 of 1

esri

Appendix III: Manual Data Collection Flowchart



Appendix IV: Achievement of Research Objectives

Research Objective	Review
Combine and harmonise existing national levee databases into a global database of flood-protection in deltas	• Conducted research into existing databases and analysed the needs and requirements for a global database • Collated vector and raster data and imported them into a single cohesive dataset
Develop an algorithm and data standard for processing the flood protection data that can be applied to other deltas	• Created data collation workflows • Developed a clear set of criteria for data quality standards to define selection criteria for new data sources • Created a geoprocessing model to process vector data
Contribute to the wider scientific community by making the database publicly available, to benefit future deltaic research	• Documented and annotated the key functions of the data, model and tools in metadata, as well as producing a README and tool-tip interactive help text • Selected a data dissemination platform that provide free access to the data in several formats, and upheld the University's data management policies • Contacted the publishers of recent papers who identified the need for this dataset and invited contact to further develop the dataset • Ensured that there was an appointed custodian that would remain available past the end of this individual project

Appendix V: Achievement of Educational Objectives

Educational Objective	Review
Develop and refine data selection and processing skills to be able to accurately and reliably include acceptable data	• Independently searched diverse sources using all available tools and correctly decided when to contact external dataset holders. • Documented the decision-making process fully and included reasons for negative/unsuccessful outcomes to further support research in this area as well as to support own thoughts and ideas.
Learn industry-standard coding and processing methodologies to (partially-)automate the data gathering and collation processes.	• Developed an Entity Relationship Diagram to control database design and identify scope for inter-linking of data • Learned the basic syntax and functionalities of the Arcade[13] language to undertake simple data manipulation in ModelBuilder and ArcGIS Online • Inspected model data in Python output and gained an understanding of the different ways that Python can be integrated into ArcGIS
Understand the state-of-play for data dissemination systems, (meta-)data standards and platforms, and data best practice.	• Set out clear requirements for a data dissemination platform and critically evaluated candidate systems • Adapted to changes in data platforms and requirements of data consumers to ensure a high-quality product • Produced detailed and validated metadata for the shapefile and tool in compliance with the EU INSPIRE Directive for geospatial metadata.

[13] Arcade is a proprietary expression language written by ESRI for data manipulation in the ArcGIS suite. © 2020 ESRI, Inc. Further information: https://developers.arcgis.com/arcade

8. Copyright and Usage

The report and dataset are an original piece of work by the author. The resultant work is released under the CC-BY 4.0 (International) licence.

Resources produced throughout this research project includes data from various sources, reference to the source of each shapefile is included within the metadata of the individual polygon and in the Delta Index which can be found within the dataset.

Maps throughout this report were created using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license. Copyright © Esri. All rights reserved. For more information about Esri® software, please visit www.esri.com.

ORCID® is a registered trademark and the ORCID logo and iD icon are trademarks of ORCID, Inc., and have been used strictly within the terms of the ORCID Brand Guidelines for Members and Partners (as at 18 June 2020)

Research word cloud (Figure 1, p.2) created using WordClouds.com © Zygomatic 2020

Figures 4, 5, 6, 16, and those in Appendix I use the World Imagery basemap containing imagery from Earthstar Geographics, CNES/Airbus DS, and Maxar Technologies © Esri, HERE, Garmin, USGS, NGA and are used under the terms of the Esri Master Licence Agreement

DANTE output has kindly been made possible by Silvia Barbetta, Istituto di Ricerca per la Protezione Idrogeologica (IRPI), Italy.

EA AIMS output © Environment Agency copyright and/or database right 2020. All rights reserved. Data source: https://data.gov.uk/dataset/cc76738e-fc17-49f9-a216-977c61858dda

Rear cover photo (Chiloé, Chile) by Jaume Galofré on Unsplash

Data Custodian

The report is held by Universiteitsbibliotheek Utrecht and can be accessed through the student theses archive (https://studenttheses.library.uu.nl/) and the DANS NARCIS service (http://narcis.nl) after the board-levied embargo expires.

The permanent identifier (DOI) for the research dataset is: doi:10.34894/2WZ0S9

Should the data subsequently be published in a journal or similar format, the DataverseNL page will be updated, which can be consulted via the DOI above.

The appointed data custodian is Dr. ir. Jaap Nienhuis (Universiteit Utrecht)

Comments and queries regarding the method, data, or publication can be addressed to the author using the **Dataset Contact** function within DataverseNL.

This report is released under a Creative Commons Attribution 4.0 International licence

For access to project data, visit:
doi:10.34894/2WZ0S9

To view the dataset in ArcGIS Online, visit:
https://edu.nl/38cdx