This document provides data descriptions for the datasets submitted to the TERN - AEKOS SHaRED data platform under the title:

‘**Spatial analysis data for 'Lines in the sand: quantifying the cumulative development footprint in the world’s largest remaining temperate woodland**'

By Keren Raiter

keren.raiter@research.uwa.edu.au / keren.raiter@me.com / kerengila@gmail.com

This data accompanies the article ‘Lines in the sand: quantifying the cumulative development footprint in the world’s largest remaining temperate woodland’ Published in Landscape Ecology. DOI: 10.1007/s10980-017-0558-z (In press as at August 2017). Please read the article for a description of the context, study area, and methods.

**The datasets**

data in csv format:

1. development footprint by sample area
2. gww 20 km grid
3. disturbance by patch
4. mapped versus digitised tracks
5. edge effect scenarios

shapefiles:

1. Great Western Woodlands boundary
2. sample areas
3. compiled dataset of mapped linear infrastructure (extending beyond gww boundary by ~100 km; note there is a more detailed metadata document inside this folder)
4. linear infrastructure footprints
5. digitised tracks
6. digitised hub infrastructure
7. edge effect zones

**Column explanations and data descriptions**

data in csv format:

1. **disturbance by sample area**

This dataset provides the calculated development footprints for the 24 sample areas assessed in the study, including the different infrastructure types (roads, railways, mapped tracks, unmapped tracks which have been manually digitised in the study using aerial imagery and hub infrastructure such as mine pits and waste rock dumps, also manually digitised in the study). Also contains some key covariables assessed as potential explanatory variables for development footprint in the region-wide analysis.

The region-wide modelling of development footprint found strong positive effects of mining project density and pastoralism, as well as a highly significant negative interaction between the two. At low mining project densities, development footprints are more extensive in pastoral areas, but at high mining project densities, pastoral areas are relatively less developed than non-pastoral areas, on average.

- column A) sa: = sample area (1-24).

- column B) area.km: area of the sample area calculated in square kilometres.

- column C) category: one of eight categories based on four levels of mining intensity (no mines or low, medium or high mining intensity) and pastoral tenure status.

- column D) pastoral: pastoral status.

- column E) m.intensity: mining intensity = a numerical equivalent to the above mining levels (3 = high mining intensity, 2= medium mining intensity, 1= low mining intensity and 0 = no minedex mining projects in that sample area). Used as predictor variable.

- column F) mprojects: number of MINEDEX mining projects in the sample area

- column G) m.density: number of MINEDEX mining projects in the sample area divided by the area of the sample area (in km2)

- column H) dtown: distance to nearest town in kilometres

- column I) dwb: distance to edge of the wheatbelt (in kilometres)

- column J) dpolys\_m2: area of hub infrastructure (polygonal disturbance; i.e. not linear) in spare metres

- column K) footprint: total development footprint in square kilometres

- column L) footprint.prop: the proportion of the sample area accounted for by the development footprint (i.e. development footprint divided by area)

- columns M and N) the main and other minerals being targeted by MINEDEX projects within the sample area.

- columns O to S) the area (part of the development footprint) accounted for by each of: main roads, sealed roads (not including main roads), unsealed roads (other than tracks), and mapped tracks. All of these are from various existing mapped layers, with the linear features buffered by average width of each linear infrastructure type for each sample area. Please note that in some cases these will slightly overlap and therefore the calculated sum of these individual infrastructure types might not equal the total development footprint (column K)

- column T) unmappedtracks.m2: development footprint of unmapped tracks (identified by digitizing aerial imagery)

- column U) linear.m2: total of development footprint of all linear infrastructure (mapped and unmapped)

- column V) prop.lin: proportion of total development footprint that consists of linear infrastructure.

1. **gww 20 km grid**

This dataset provides data for the 20x20 km grid placed over the whole Great Western Woodlands and used for the regional estimation of development footprint, linear infrastructure density, and linear infrastructure type based on the region-wide analysis. Data is for each cell in the grid and provides the total length of roads in that grid cell, MINEDEX mining projects, pastoral status, etc.

This dataset was used to project the data from the 24 study areas across the whole of the Great Western Woodlands and calculate region-wide estimates of development footprint and linear infrastructure lengths.

- column A) id: the ID for the grid cell. not useful other than for identifying individual cells

- column B) rdlengths: the total length of linear infrastructure within that grid cell

- column C) mines: the total number of registered Minedex projects in that grid cell

- column D) rd\_density: density of linear infrastructure (total length / total area)

- column E) pastoral status

- column F) minedens\_100km2: the density of mining projects; number of Minedex projects per 100 km2

- column G) category: one of eight categories based on four levels of mining intensity (no mines or low, medium or high mining intensity) and pastoral tenure status

- column H) shape\_length: the perimeter of the grid cell in metres. not used in the analysis

- column I) shape\_area: the area of the grid cell in m2. Used for standardising mining project density etc.

1. **disturbance by patch**

This dataset provides the data for each patch for the analysis of patch-level drivers of development footprint, which was performed to gain further insights into the effects of other landscape variables that what could be gleaned from the region-wide analysis. For this analysis, we divided sample areas into polygonal ‘patch types’, each with a unique combination of the following categorical covariables: pastoral tenure, greenstone lithology, conservation tenure, ironstone formation, schedule 1 area clearing restrictions, environmentally sensitive area designation, vegetation formation, and sample area.

For each patch type (n=261), we calculated the following attributes: number of mining projects, number of dead mineral tenements, sum of duration of all live and dead tenements, type of tenements (exploration/prospecting tenement, mining and related activities tenement, none), primary target commodity (gold, nickel, iron-ore, other), distance to wheatbelt, and distance to nearest town.

These variables are self- explanatory, together with reference to the descriptions of the other datasets. Potentially ambiguous terms defined as follows: dt=dead tenement. sched\_1 = schedule 1 clearing restrictions. lt = live tenement. cons-estate = conservation estate. ten\_yrs = total number of years that (both living and dead) tenements have existed over that area. expl\_ten= exploration tenement. disturbed area = development footprint. dist\_perc = the percentage of the patch that has been disturbed i.e. is under development footprint.

1. **mapped versus digitised tracks**

This dataset provides mapped and unmapped track widths, measured using high-resolution aerial imagery at at least 20 randomly-generated locations within each of 24 sample areas. Pastoral tenure and mining intensity for each sample area are included for analysis purposes.

This data was analysed as follows: we used a t-test to test for a difference between mapped and unmapped track width, conducted data exploration as per (Zuur et al. 2009), and modelled track widths using linear mixed models with ‘lme4’ package in R. We created a global model containing the following fixed variables: mapped/unmapped status; mining activity level for the relevant sample area, and pastoral status. Sample area identity was included as the random effect in all models after testing for its significance.

We used the ‘dredge’ function in ‘MuMin’ package to model all possible subsets of the global model and rank them based on AICc values. The optimal model included only mapped/unmapped status as a fixed effect, and the other top-ranking model also included a positive effect of pastoral tenure on track width. Mapped tracks were found to be on average ~1 m wider than unmapped tracks (p < 0.001) (Figure A2.1). Average widths of mapped and unmapped tracks were 6.06 m (s.e. 0.15 m) and 4.92 m (s.e. 0.10 m) respectively. No effect of mining activity was included in the top-ranking models.

- column A) type = type of track; either mapped or unmapped

- column B) width = width in metres. This was the response variable

- column C) sa = sample area (1-24). used as random factor in the analysis

- column D) mining = mining level (categorised as ‘no.mines’ or ‘low’, ‘medium’ and ‘high’ mining activity)

- column E) m.intensity = mining intensity = a numerical equivalent to the above mining levels (3 = high mining intensity, 2= medium mining intensity, 1= low mining intensity and 0 = no minedex mining projects in that sample area). Used as predictor variable.

- column F) pastoral = pastoral tenure status; either grazed (pastoral tenure) or ungrazed (not pastoral tenure). Tested as possible predictor variable.

- column G) category: category based on both mining intensity and pastoral tenure status

1. **edge effect scenarios**

Hypothetical edge effect zones were created, based on effect zones gleaned from the literature and arranged under three scenarios, to reflect potential risks of offsite impacts in areas adjacent to development footprints observed (see appendix 3 of article).

The calculated proportion of the entire GWW within edge effect zones varied from ~3% under the conservative scenario to ~35% under the maximal scenario. Within the range of development footprints observed in this study, the proportion of a landscape that lies within edge effect zones increases hyperbolically with the number of mining projects, and approaches 100% in the maximal scenario, 60% in the moderate scenario, and ~20% under the conservative scenario.

- column A) sa: number of sample area (1-24)

- column B) area: area of sample area

- column C) pastoral: 1 = grazed (pastoral tenure); 0=ungrazed (not pastoral tenure)

- column D) mprojn: number of MINEDEX mining projects in the sample area

- column E) mproj1: number of MINEDEX mining projects in the sample area plus one (used to calculate log in next column)

- column F) logproj: log (base ten) of number of MINEDEX projects in the sample area plus one

- column G) footprint\_km: total development footprint in square kilometres. Please note: there is a slight discrepancy between the total footprint calculation for the individual sample areas in this dataset compared to the footprint areas given in the disturbance by sample area spreadsheet. This is because they were calculated with datasets created using slightly different projections of the shape of the earth’s surface. This spreadsheet was created using an Australia-wide Albers projection while the first was created using MGA zone 51.

- column H) conservative\_km: total area affected under the conservative scenario

- column I) moderate\_km: total area affected under the moderate scenario

- column J) maximal\_km: total area affected under the maximal scenario

- column K) footprint.p: proportion of sample area taken up by the area disturbed by the direct footprint

-column L) conservative.p: proportion of sample area taken up by the area disturbed by the direct footprint and the edge effect zones under the conservative scenario

- column M) moderate.p: proportion of sample area taken up by the area disturbed by the direct footprint and the edge effect zones under the moderate scenario

- column N) maximal.p: proportion of sample area taken up by the area disturbed by the direct footprint and the edge effect zones under the maximal scenario

**shapefiles:**

1. **Great Western Woodlands boundary**
2. **sample areas** (layer file shows sample areas by category). We used stratified random sampling to distribute 24 circular sample areas, each 25 km in diameter, among the 8 mining and pastoral categories. We used circular sample areas to minimise the edge-to-area ratio of the sample areas and therefore maximise the extent to which the sample areas reflected the category represented rather than the adjacent landscape.
3. **linear infrastructure extending beyond gww boundary by ~100 km**. This is a dataset compiled from 23 different sources that represents the most comprehensive spatial dataset for the GWW available at the time of publication, to KR’s knowledge. However, it does contain a number of different sources of error and should not be considered to necessarily reflect an updates, accurate dataset (note there is a more detailed metadata document inside this folder).
4. **linear infrastructure footprints**. Linear features buffered by average width of that linear infrastructure type for each sample area. Linear features include paved roads and railways, unpaved roads, mapped tracks, and unmapped tracks (digitized from aerial images in this study).
	1. prr = all paved roads and railways (grouped together as there were not many features in this category)
	2. ur = unpaved roads
	3. mt = mapped tracks
	4. ut = unmapped tracks
5. **digitised tracks** All linear infrastructure that hadn’t already been mapped in #8 above. Manually digitised from high-resolution aerial images in this study.
6. **digitised hub infrastructure** Development footprints of all non-linear (i.e. polygonal) anthropogenic disturbance, including mine pits, waste rock dumps, mining camps and accommodation villages,dams, and other cleared areas, manually digitised from high-resolution aerial imagery in this study.
7. **edge effect zones** Polygons created by creating buffers around the development footprint as described in Appendix 3 of the article. These zones around the direct development footprint represent offsite impact risk for each type of infrastructure, using a hypothesized set of risk buffers. These were based on edge effect distances reported in the literature for species and processes from around the world. Three scenarios are represented: a conservative, moderate, and maximal scenario.

We gratefully acknowledge support from the Gledden Postgraduate Research Scholarship, the Australian Research Council Centre of Excellence for Environmental Decisions, The

Wilderness Society, Gondwana Link, the Natural Environmental Research Program

Environmental Decisions Hub, and the Great Western Woodlands Supersite, part of Australia’s Terrestrial Ecosystem Research Network. We thank Ophir Levin, Julia Waite, Brad Desmond, and Rachel Omodei for assistance in digitising the unmapped development footprint. We also thank Fiona Westcott for her assistance in ground-truthing, Cliffs Natural Resources for in-kind 500 support in the field, and Amanda Keesing (Gondwana Link), Judith Harvey (DPaW) and Katherine Zdunic (DPaW) for their assistance in supplying spatial information. Ashley Sparrow, Richard Forman, Andrew Bennett, John Bissonette, and two anonymous reviewers provided valuable reviews that improved this manuscript. All fieldwork was carried out under Department of Parks and Wildlife Regulation 4 lawful authority CE003548.