Limitations on Storage and Computation in the use of High Resolution Digital Elevation Models

Hartcher, M.G., Carlin, G.D., Cook, F.J.

CSIRO Land and Water, Long Pocket Laboratories, Indooroopilly, QLD. 4068.

Keywords: *Digital elevation models, spatial resolution, modelling, processing, computation, storage.*

EXTENDED ABSTRACT

Digital elevation models (DEM's) were compared at cell resolutions of 250 metres, 25 m, 5 m, and 1 metre between the Murray-Darling basin (Approximately 1 million sq km area), the Burdekin catchment (Approximately 130,000 sq km area) in North Queensland, and the Bowen-Broken (Approximately 9500 sq km) which is a sub-catchment of the Burdekin. Figure 1 shows the location of the study regions.

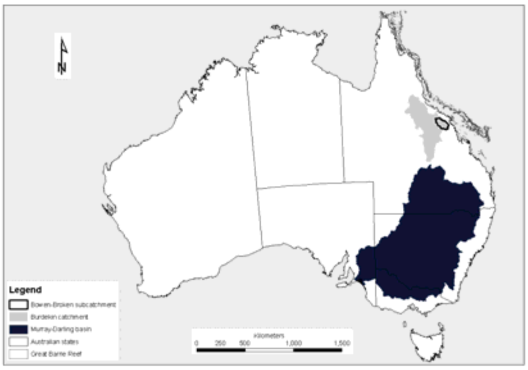

Figure 1. The Australian continent overlayed with the Murray-Darling basin, Burdekin catchment, and Bowen-Broken sub-catchment.

DEM's are an important source of land surface parameters such as for catchment modelling and agent-based modelling applications. However, DEM's covering wide areas in Australia are typically at a fairly coarse scale, i.e. 250m. This resolution does not allow for accurate depictions of finer scale landscape processes, especially when analysis is focused down to the scale of a grazing paddock. Higher resolution DEM's are therefore required in order to capture surface parameters such as surface slope within paddocks, so that more accurate models of surface water, ground water, and sediment and nutrient transport can be developed.

The DEM's were compared in terms of their size and required disk storage, as well as the number of cells they contained. The results highlighted some limitations of using higher resolution DEM data by demonstrating that an increase in the resolution of a DEM, e.g. from 250 m to 1 m, will ultimately create a file that is significantly larger, i.e. 62 500 times larger, and has 62 500 times more points to process.

Even if there is enough space to store the DEM data, many models and/or software packages are limited or not able to handle these datasets due to their sheer size. However, some strategies and techniques are discussed which may allow for analysis of high resolution DEM data, i.e. 1m, even across areas as large as the Murray-Darling Basin (MDB).

The limitations of computing resources as platforms for storage and processing the various resolutions of data for each area, from the desktop level, through to more costly infrastructure such as major data servers are discussed. An assessment of the most viable coupling of spatial resolution and coverage area is explored, in light of determining an optimal combination for a given level of computing resources, and with the assumption that funding for developing such datasets is available.

Techniques such as disaggregating the whole-of - basin/catchment/sub-catchment into more manageable areas for processing, such as catchments/sub-catchments or regular tiles, are discussed as well as methods for determining spatial variance, coupled with context based classifications, which can help focus the acquisition of high resolution in only those areas where it adds value.

The utilisation of high-end computing infrastructure, such as the CSIRO's 'Water Resources Observation Network' (WRON) data centre and processing facility, are also discussed. It is shown that such infrastructure is necessary for the provision of adequate resources to store, backup, archive, and process high resolution DEM's over expansive areas.

1. INTRODUCTION

Digital elevation models (DEM's) are being utilised increasingly for deriving various land surface parameters required in modelling. Catchment models and Agent-Based models (ABM's), which also incorporate biophysical attributes into agent behaviour, both require DEM's for delineating hydrological boundaries and for extraction of various surface parameters.

Catchment modelling applications such as SedNet utilise the revised universal soil loss equation (RUSLE) for estimating surface erosion prior to modelling sediment transport throughout a given catchment (Prosser, *et al.* 2001). A key surface parameter used in this equation is surface slope. SedNet also requires a DEM as input in order to create an initial catchment configuration. SedNet then delineates sub-catchments and a drainage network based on the modeller specifying a catchment area threshold and minimum stream tributary length.

Understanding and identifying the major processes involved in the delivery of sediment and nutrients to streams, the critical areas of erosion potential, and the major contributors of sediment and nutrients to the coast, are required for sustainable catchment management (Bartley *et al.* 2004). The use of detailed spatial representation of hillslope erosion and unique methodologies for deriving erosion parameters provides a unique opportunity to gain insight into the sensitivity of a stream system to the effects of landscape processes (Kinsey-Henderson *et al.* 2003). Catchment modelling results have shown that there can be significant uncertainty in results which may be due to the spatial resolution of input data (Hartcher and Post, 2005).

ABM's (Heckbert *et al.* 2005) use DEM's to define catchment boundaries and stream networks, and to obtain average surface slope for grazing properties and paddocks (Heckbert *et al.* 2005; Carlin *et al.* 2007). An ABM is a computational model often used to create dynamic systems that simulate social agents using simple rules that often result in emerging complex behaviour (Carlin *et al.* 2007). As research strives to investigate land use change at finer scales, e.g. grazing paddocks, there is a greater requirement for finer scale DEM's which will improve the accuracy of such derived parameters (Carlin, *et al.* 2007).

SEPIA has a collection of software agents (objects) that mimic the real world behaviour of land managers for sugar cane, tree fruits (banana), and beef cattle (grazing) producers. SEPIA also incorporates the biophysical world that our land managers interact with. The land manager agent behaviour results in the enactment of one of a number of possible land-use strategies. The effect of these land-use decisions in turn has a possible effect on biophysical conditions at the paddock scale, a resulting outcome for agent financial payoffs associated with agricultural production, and a potential raised level of environmental gratification derived from the state biophysical world (Carlin *et al.* 2007).

The biophysical world within SEPIA is modelled at the paddock scale using the surface and groundwater hydrology OOP models (Carlin *et al.* 2007). This is achieved by linking the paddock spatial data file to a paddock agent that also instantiates the surface and groundwater hydrology OOP models. The paddock then executes the surface and groundwater hydrology models at a daily time step to potentially produce run-off and groundwater. The run-off and groundwater is then moved through the catchment via the contours, reaches, pour points and sub-catchments (Carlin *et al.* 2007).

Analysis on non-point pollution model sensitivity has shown that cell size selection is not an arbitrary choice and should be based on the scale necessary to capture the spatial variability (Vieux *et al.* 1993). As grid-cell sizes increase, stream meanders are short-circuited with the shortened stream lengths causing sediment yield to increase by as much as 32% (Vieux *et al.* 1993). Analysis of spatial variability, focused on surface infiltration parameters, in distributed hydrological modelling, showed that there is a critical resolution for analysis, and that if a more coarse resolution is utilised the results are erratic and contain large errors (Farajalla and Vieux 1995).

This paper will highlight the limitations of using higher resolution DEM data by demonstrating that an increase in the resolution of a DEM from 250 m to 1m ultimately creates a file that is 62 500 times larger, i.e. 250 m x 250 m, and will be more costly and difficult to acquire due to security and availability issues. In addition, there may be a number of data layers which will be spatially coincident to the DEM and these also need to be managed and utilised in each modelling framework.

Even with the assumption that funds are available for the acquisition and development of such data, there is a need to find enough storage to manage the DEM's and associated data layers. The capacity to manage such large data sets will vary greatly between small organisations with limited infrastructure and large organisations such as the CSIRO who have invested in robust computing infrastructure, such as the WRON (Water Resources Observation Network) computing infrastructure, with large storage volumes, high performance processing facilities, fibre-optic links, and adequate backup/archiving facilities.

However, even when facilities exist to manage very large data sets, many models and or software packages may not be able to process these datasets due to their sheer size. An assessment of a viable coupling of spatial resolution and area is examined in light of determining an optimal combination of spatial resolution and coverage area, given a particular level of available computing infrastructure. Issues such as the cost per unit area, the value added by acquiring high resolution data, analysis of spatial variance and dynamics in the landscape in relation to selecting an appropriate resolution, and some exploration of methods for managing high resolution data over large areas such as the Murray-Darling Basin in Australia, are also discussed.

2. METHODS

The methods presented here will illustrate the variation size of using higher resolution DEM data and the computational hurdles that need to be addressed. This study has focused on a comparison of DEM's between a sub-catchment (Bowen-Broken), whole catchment (Burdekin), and river basin (Murray-Darling), with each being an order of magnitude larger in area.

2.1. Study Areas

Bowen-Broken Sub-catchment

The Bowen-Broken is an important sub-catchment of the Burdekin catchment covering an area approximately 9 500 km², in North Queensland, Australia. The landforms found within the catchment are relatively complex due to variations in the underlying geology and geomorphic processes over time (Isbell and Murtha 1972.). The Bowen-Broken sub-catchment has long been considered a 'hot spot' area in terms of sediment and nutrient loss in the Burdekin Catchment (Prosser *et al.* 2001; Brodie *et al.* 2003). The Bowen-Broken catchment has recently been the focus for sediment modelling (Bartley, *et al.* 2004) and agent-based modelling applications (Heckbert *et al.* 2005) as well as paddock scale modelling of surface and groundwater flow (Carlin *et al.* 2007). Figure 2 shows the Bowen-Broken sub-catchment with grazing properties used in the SEPIA model, and a SedNet modelling stream network overlayed.

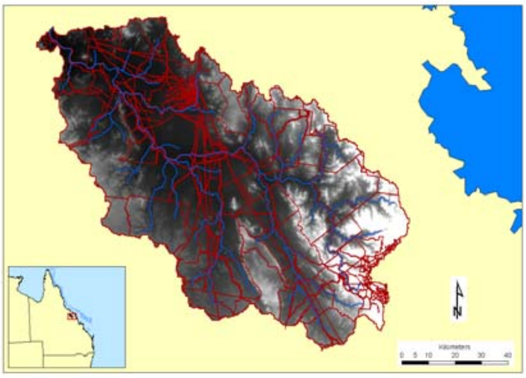

Figure 2. The Bowen-Broken catchment showing properties and SedNet streams overlayed on a 250 m DEM.

Burdekin Catchment

The Burdekin catchment is located in North Queensland, Australia, and covers an area of approximately 130 000 km². While grazing is the dominant land use occurring in the catchment, the lower Burdekin contains the most productive sugar cane growing region in Australia and also includes crops such as fruit, vegetables, hemp, and plantation timber. Figure 3 shows the Burdekin catchment 250 m DEM.

Grazed catchments such as the Burdekin are complex systems, often with considerable variation in grazing pressure, and diverse topography, soils, rainfall, and vegetation cover (Prosser *et al.* 2002). In recent years the Burdekin has been the focus of sediment and nutrient modelling, and has also recently been used as a case study for agent based modelling using SEPIA (Heckbert *et al.* 2005).

Murray-Darling Basin

The Murray-Darling basin (MDB) covers an area of approximately 1 million km² and includes four Australian states and the Australian Capital Territory. As might be expected with such an extensive coverage area, the MDB is highly variable in its topography, land use, and population density. The MDB is currently the focus of an assessment of sustainable water yields by the CSIRO, which includes analysis of surface and, ground water, and environmental sensitivity. Figure 4 shows the MDB in relation to South Eastern Australia.

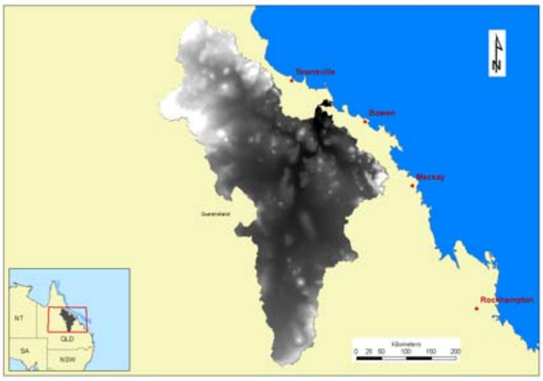

Figure 3. A 250 m DEM of the Burdekin River catchment.

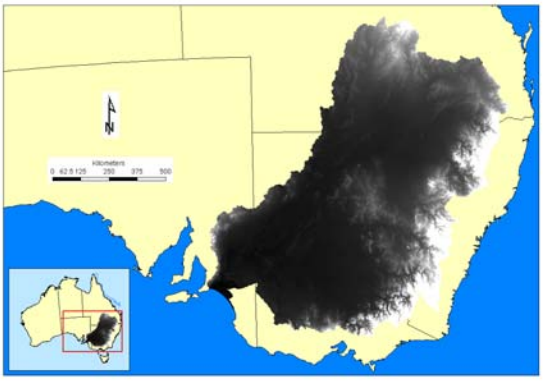

Figure 4. A 250 m DEM of the Murray-Darling basin.

2.2. Digital Elevation Models

The sizes of the different DEM's were calculated for 32 bit continuous floating point grids stored in ESRI GRID format. A 250 m resolution DEM exists for continental Australia. This DEM was re-sampled with ArcGIS version 9.2 software package to create DEM's of 25 m and 5 m. At the time of creation there was not enough storage space to resample 1 m resolution grids so the number of cells and file sizes were calculated by multiplying the number of cells and size by 62 500, i.e. an increase of 250 x 250 cells.

3. RESULTS

The Bowen-Broken results (Table 1) show that an increase in resolution from 250 m to 5 m produces a file size of approximately 2.5 Gigabytes (GB) with approximately 695 million cells. If we increase the resolution to 1m the result is a file size of 62.5 GB with approximately 17.4 billion cells.

The Burdekin results show that the initial 250 m grid has a file size of 21 Megabytes (MB) with approximately 5.5 million cells. This increased to 51.5 GB and approximately 14 billion cells when a 5 m DEM was generated. The calculation for a 1 m DEM resulted in a file size of approximately 1.3 Terrabytes (TB) with approximately 346 billion cells.

Table 1 The number of cells and disk storage required for DEM's ranging from 250 m to 1 m for the Bowen-Broken, Burdekin, and MDB.

Bowen-Broken (Area = 9 500 km²)

	Approximate No. Cells	Approximate Size (MB)
250	2.78×10^{5}	1
25	2.78×10^{7}	106
5	6.95×10^{8}	2590
1	1.74×10^{10}	62500

Burdekin (Area = 130 000 km²)

	Approximate No. Cells	Approximate Size (MB)
250	5.53×10^{6}	21
25	5.53×10^{8}	2060
5	1.38×10^{10}	51500
1	3.46×10^{11}	1312500

Murray-Darling (Area =1 064 600 km²)

	Approximate No. Cells	Approximate Size (MB)
250	3.24×10^{7}	124
25	3.24×10^{9}	12400
5	8.11×10^{10}	310000
1	2.03×10^{12}	7750000

The Murray-Darling results show that the initial 250 m grid size was 124 MB with 32.4 million cells. An increase in resolution to 5 m resulted in a file size of 310 GB with approximately 81 billion cells. The increase in resolution to 1 m resulted in a file size of 7.75 TB with 2 trillion cells.

4. DISCUSSION

4.1. Data Storage

A key limitation in the use of DEM's is the sheer size resulting from an increase in the resolution of a DEM, from 250 m to 1m. For the Bowen-Broken area the resulting file sizes, for DEM's up to 5 m in resolution, does not significantly affect storage with most desktop PC's providing 100 GB of hard disk storage as a standard. However, the 1 m grid may be too large for a stand alone PC but should be accommodated by most centralised data servers which often provide between 1 and 5 TB of storage.

The Burdekin data showed a significant increase in file size and cells compared to the Bowen-Broken. While the 250 m and 25 m grids were manageable by most desktop PC's, an increase to 5 m would probably be too large for a PC, but would be managed by most centralised data servers. However, the 1 m grid may be a significant burden on most data servers around the 5TB size and may require the resources of a major data centre, with significant storage volumes, backup and archiving services, fast processing, and fibre optic links.

The Murray-Darling basin results showed that the 250 m and 25 m grids could be accommodated by most standard PC's. An increase to 5 m resolution would exceed most PC storage limits, but should be accommodated by a standard data server. However, the increase in resolution to 1 m would not be manageable on a standard data server and, would require the resources provided by a major data centre such as the CSIRO WRON facilities.

4.2. Data Processing

The increase in resolution from 250 m to 1 m also means that there are 62 500 times more points to process. The 1 m Bowen-Broken, which has some 695 million cells, will significantly hinder computation times. Some memory dependant modelling packages, such as SedNet, will still be able to process grids of this size.

The increase in resolution from 250 m to 25 m for the Burdekin saw an increase from 5.5 million to approximately 14 billion cells, yet the recent versions of SedNet, for instance, have been able to model the Burdekin at a resolution of 25 m. However, an increase in resolution to 1 m will significantly increase computation times with some 346 billion cells to process. Programs such as SedNet would be unable to process such grids.

The increase in resolution from 250 m to 25 m for the Murray-Darling basin should still be managed by programs such as SedNet with approximately 3.2 billion cells to process. When the resolution increases to 5 m, the resulting 81 billion cells will exceed the limits of memory. However, the increase to 1 m, with some 2 trillion cells will definitely exceed the limits of most software modelling packages, and would strain the resources even for data centres such as the WRON. However, there may be strategies which can be employed in order to facilitate storage and processing of these high resolution grids.

4.3. Archiving Strategies

As previously mentioned, the large file sizes resulting from 1 m resolution grids for areas such as the Burdekin or MDB would be a serious strain on resources even for a major data centre. However, it may not be necessary to store the entire 1 m DEM for either of these regions. An archiving strategy could be employed whereby the entire grid is initially backed up to tape storage media. Prior to the backup a grid would be clipped into smaller sections, such as tiles or topographic units. It would only be necessary to retrieve and store those sections which are required for analysis and which can be stored without impeding other work. The set of results for a single section could then be archived and removed from the server prior to the next section being retrieved from tape and processed. While this will require some organisation, and may take days or even weeks to process, it would at least provide the capability to analyse large coverage areas at a 1 m resolution. In addition, any opportunity to compress data or to utilise formats yielding smaller files sizes should be explored.

4.4. Spatial Variance Techniques

An alternative to providing a high resolution DEM over an entire region is to only capture the resolution required to analyse a particular area. Often when a fine scale DEM is used the result will contain little more information than is in a coarser scale resolution. It may also be possible to use some selection process to decrease the amount of information that needs to be stored. Such problems have been studied in a number of contexts and various algorithms developed (MacQueen 1967, Saeed *et al.* 2003, Dimitrova *et al.* 2005), which allow an objective means of creating clusters from the original data sets.

With such an approach the higher resolution data may only be acquired for areas with greater spatial variance. Analysis of spatial variance and/or topographic complexity across a region may be used to develop, for example, flight paths for minimising the area required and cost of acquisition of high resolution elevation data using airborne scanners such as LIDAR.

Some regions within the MDB, for example, are extremely flat with as little as 1 m variation in elevation across a 20 km area. The cost of acquiring high resolution data for such areas would not be justified as there would be very little, if any, improved value in the data. However, areas such as the Burdekin river delta, which is very flat but highly dynamic, may still require high resolution data. In such areas small changes in elevation may be significant, e.g. small changes in elevation may be the difference between salt water intrusion and fresh water.

5. CONCLUSION

While high resolution DEM's are an important source of land surface parameters, they can easily overburden available computing resources for large areas such as the Burdekin catchment or MDB. However, these data sets can be utilised for smaller catchments or sub-catchments, such as the Bowen-Broken, although some consideration still needs to be made concerning storage methods and processing times. If addition, spatially coincident grids, containing additional parameters, e.g. soils, are utilised in the same modelling context, then these will also need to be considered as part of any storage and computational strategy.

If high resolution DEM's are to be utilised, across large areas, then it is necessary to move to more robust computing infrastructure. Through the utilisation of resources such as the CSIRO's WRON computing infrastructure, along with strategies involving data archiving and piecemeal processing, it is possible to carry out analysis across very large areas at resolutions as fine as 1 m.

Techniques such as disaggregating the whole-of -basin/catchment/sub-catchment into more manageable areas for processing, such as catchments/sub-catchments or regular tiles, will need to be utilised even with the use of major data centres, as demand for such resources is often high. Also, the use of data compression and automated archiving should be utilised in order to minimise disk storage and facilitate data backup and retrieval for the user.

In addition, methods for determining spatial variance in elevation, which can help focus the acquisition of high resolution data, in only those areas where it adds value, also need to be examined. However, spatial variance alone will not provide adequate context for identifying priority areas. The significance of slight changes in elevation may be quite high in more dynamic locations such as the Burdekin river delta, where a 2 m change in elevation may be the difference between salt water intrusion and fresh water, compared to very flat areas in less dynamic locations such as deserts or very flat clay pans.

The use of ancillary data, which can provide context in classifying areas in need of high resolution data, should be examined. The application of a context based classification to a sub-catchment/catchment/basin area may significantly reduce the amount of high resolution data needing to be acquired, therefore reducing cost of acquisition of data, storage volumes, and providing greater utility for performing catchment modelling and landscape analysis.

6. ACKNOWLEDGEMENTS

This project is supported by the CSIRO Land & Water 'Water Resources Observation Network' WRON Water Accounting Project. The authors also thank Dr David Post and Mr Trevor Dowling from CSIRO division of Land and Water for reviewing this paper, and Mr Joseph McMahon, an Honours student working for CSIRO Land and Water, for valuable input.

7. REFERENCES

Bartley, R., Hartcher, M., Henderson, A., Chen, Y. (2004), The Application of the SedNet Model to the Bowen Catchment, Queensland: assessment of sediment and nutrient loads at a sub-catchment scale under different grazing scenarios. CSIRO Land and Water Client Report, November 2004.

Brodie, J., McKergow, L.A., Prosser, I.P., Furnas, M., Hughes, A.O., Hunter, H. (2003), Sources of Sediment and Nutrient Exports to the Great Barrier Reef World Heritage Area, ACTFR report number 03/11 for CSIRO.

Carlin, G.D., Cook, F.J., Crop, R., Hartcher, M.G., Smajgl, A., Hodgen, M., Heckbert, S. (2007), Using a Catchment Contour Approach for Simulation of Ground and Surface Water behaviour within Agent based Modelling Platforms, Submitted to MODSIM07.

Dimitrova, E. S., J.J. McGee and R.C. Laubenbacher, (2005), Discretization of time series data. arxiv.org/pdf/q-bio/0505028

Farajalla and Vieux (1995), Capturing the Essential Spatial Variability in Distributed Hydrological Modelling: Infiltration Parameter. Hydrological Processes, Vol. 9, 55-68.

Hartcher, M. G. and Post, D. A. (2005). Reducing uncertainty in sediment yield through improved representation of land cover: application to two sub-catchments of the Mae Chaem, Thailand. MODSIM 2005 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2005, 1147-1153. http://www.mssanz.org.au/modsim05/proceedings/papers/hartcher.pdf

Heckbert, S.R., Smajgl, A., Hartcher, M.G., Carlin, G. (2005) Land use modelling for policy impact assessment in the Great Barrier Reef region. Australian Journal of Agricultural and Resource Economics (Submitted).

Isbell, R.F. and Murtha, G.G., (1972) Burdekin-Townsville Region, Queensland Resources Series, Landforms. Geographic Section, Department of National Development, Canberra.

Kinsey-Henderson, A.E., Prosser, I., Post D.A., (2003), SubNet – Predicting Sources of Sediment at Sub-catchment Scale Using SedNet, MODSIM Conference, Townsville, 2003, 590-595.

MacQueen, J., (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the 5Th Berkeley Symposium of Mathematical Statistics and Probability, 1, 281-297. University of California Press, Berkley, CA.

Post, D. A. and Hartcher, M. G. (2006). Evaluating uncertainty in modelled sediment delivery in data sparse environments: Application to the Mae Chaem catchment, Thailand. In Sivapalan. M. (ed.). Prediction in Ungauged Basins: Promises and Progress, IAHS Publication 303, 80-89. IAHS Press, Wallingford UK.

Prosser, I., Rustomji, P., Young, W., Moran, C., Hughes, A., (2001) Constructing River Basin Sediment Budgets for the National Land and Water Resources Audit. CSIRO Land and Water Technical Report 15/01 http://www.clw.csiro.au/publications/technical2001/tr15-01.pdf

Prosser, I.P, Moran, C.J., Lu, H., Scott, A., Rustomji, P., Stevenson, J., Priestly, G., Roth, C.H. and Post, D. (2002) Regional Patterns of Erosion and Sediment Transport in the Burdekin River Catchment, Queensland. CSIRO Land and Water Technical Report 5/02, Canberra. http://www.clw.csiro.au/publications/technical2002/tr5-02.pdf

Saeed, A., V. Sharov, J. White, et al., (2003), TM4: a free, open-source system for micro-array data management and analysis. BioTecniques, 34, 374-378.

Water Resources Observation Network [WRON] (http://wron.net.au/)