

Scoping a Geospatial Repository for Academic Deposit and Extraction

<http://edina.ed.ac.uk/projects/grade>

Anne Robertson
EDINA

*EDINA National Data Centre
University of Edinburgh*

*Digital Repositories: Dealing with the Digital Deluge
Manchester University, 5th and 6th June 2007*

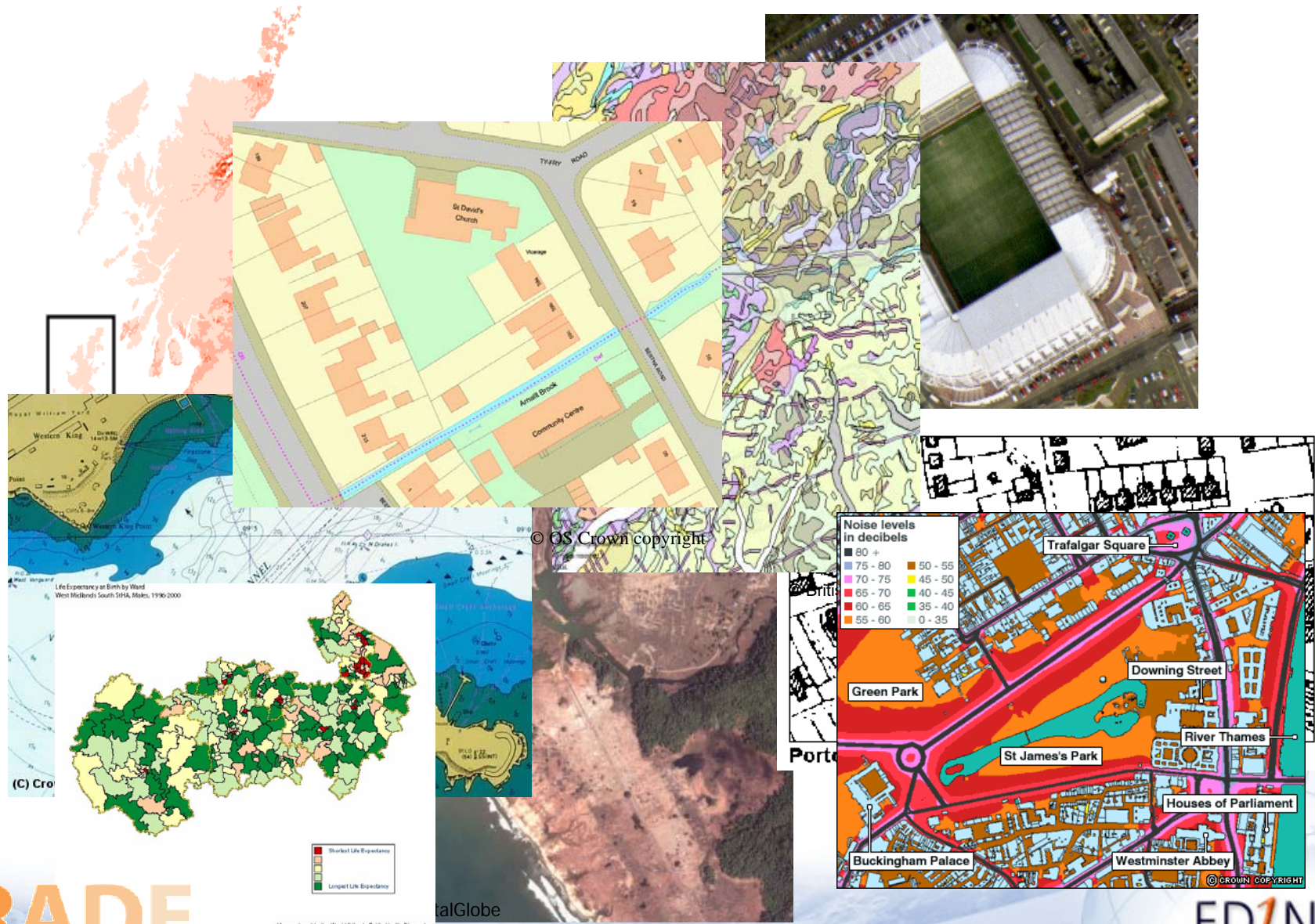
GRADE



Background - GRADE project introduction

- June 2005 – April 2007
- Investigating and reporting on the technical and cultural issues surrounding the **reuse** of geospatial data
- Investigative in nature, not building a geospatial repository
- Particular focus on sharing and reuse of **derived** geospatial data
- EDINA leading GRADE with consortium partners:
 - AHRC Research Centre for Studies in Intellectual Property and Technology Law, School of Law, Edinburgh University
 - National Oceanography Centre, Southampton University
 - Variety of other associate partners including
geospatial data creators/consumers at Nottingham, Kingston, Strathclyde, Edinburgh universities

Background - Geospatial data



Rationale - why geospatial data sharing?

Digimap

- subscription service offering various geospatial datasets for download
- for this academic year so far (1 Sep 06 to date)
 - 31,000 users from 148 institutions
 - 387,713 data downloads
 - Survey of 550 users, >40% said 'yes' I've made new data

“ Research is not merely a consumer of geospatial data but is also a **producer...**

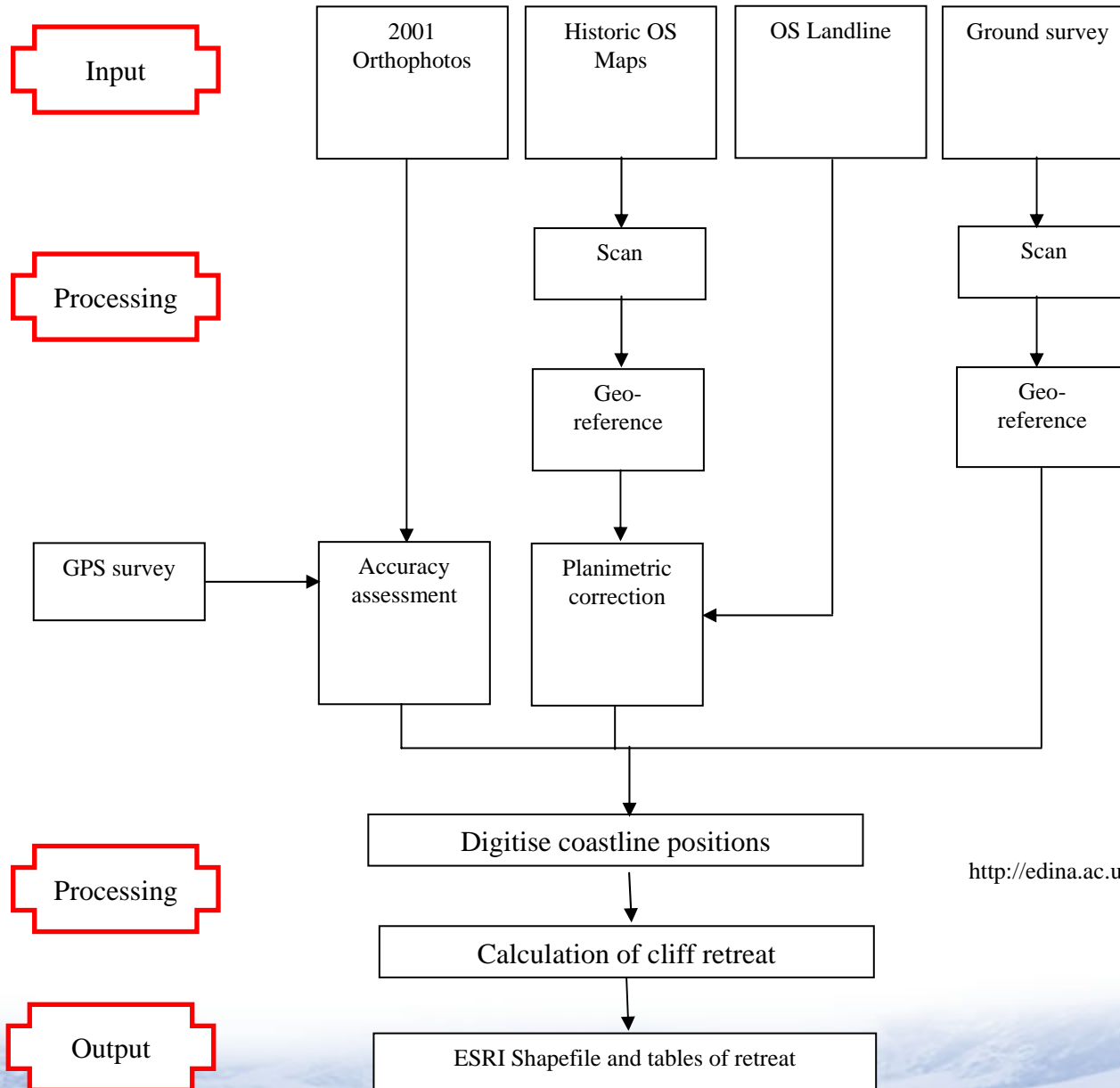
mechanisms that help researchers share their own data sets **should be more actively supported**, such as is the case with source repositories across the scientific and research community. ”

*JISC Geospatial Working Group
Draft Vision Document April 2006*

GRADE Project Work Packages

- discrete work packages
 - Digital rights issues - when we consider the reuse of derived geospatial data concerns over data ownership, IPR and copyright are commonplace
 - Investigate and make an assessment of informal mechanisms for geospatial data sharing
 - Establish user based evidence for the requirements and functionality of a repository capable of managing geospatial data
 - Role of institutional repositories for geospatial data sharing/reuse

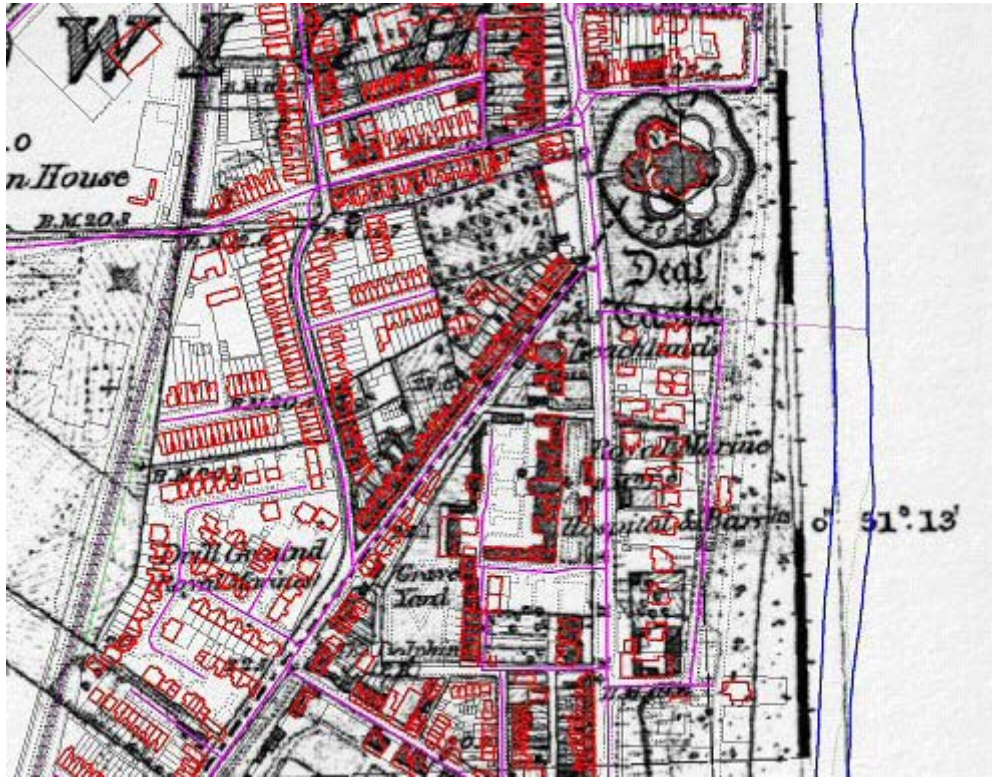
Derived Data Example



Source: Use case provision of derived geospatial data for GRADE project
<http://edina.ac.uk/projects/grade/usecasecompendium.pdf>

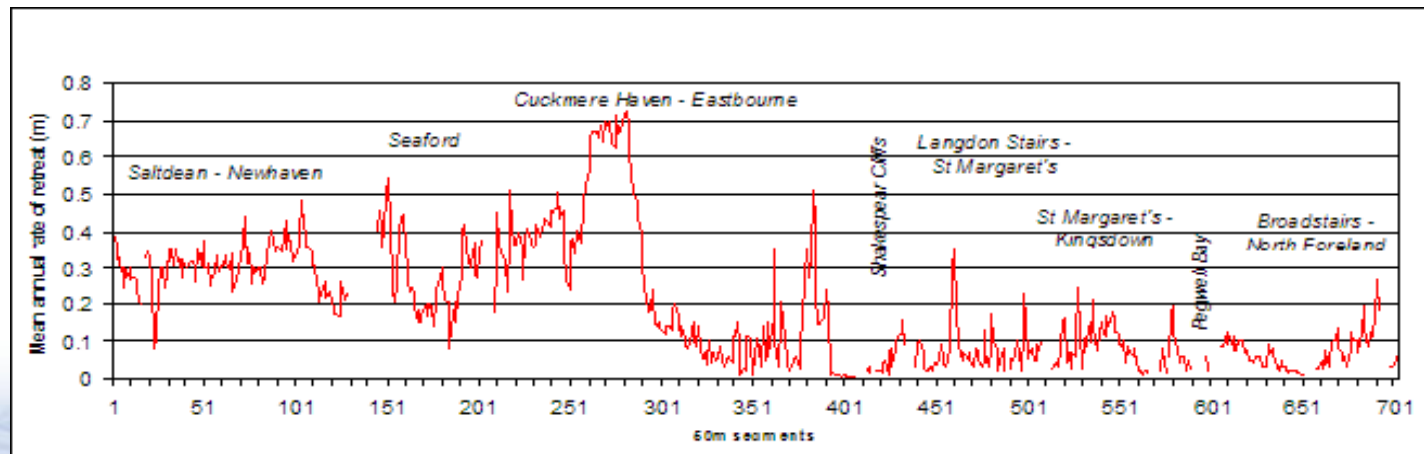


Derived Data Outputs



Scanned Ordnance Survey First Edition Maps overlain with Ordnance Survey Landline to illustrate rectification control using building outlines and road alignments.

Mean annual rate of chalk cliff retreat along the East Sussex and Kent frontage for the period 1870s to 2001



Findings: WP on Legal Issues

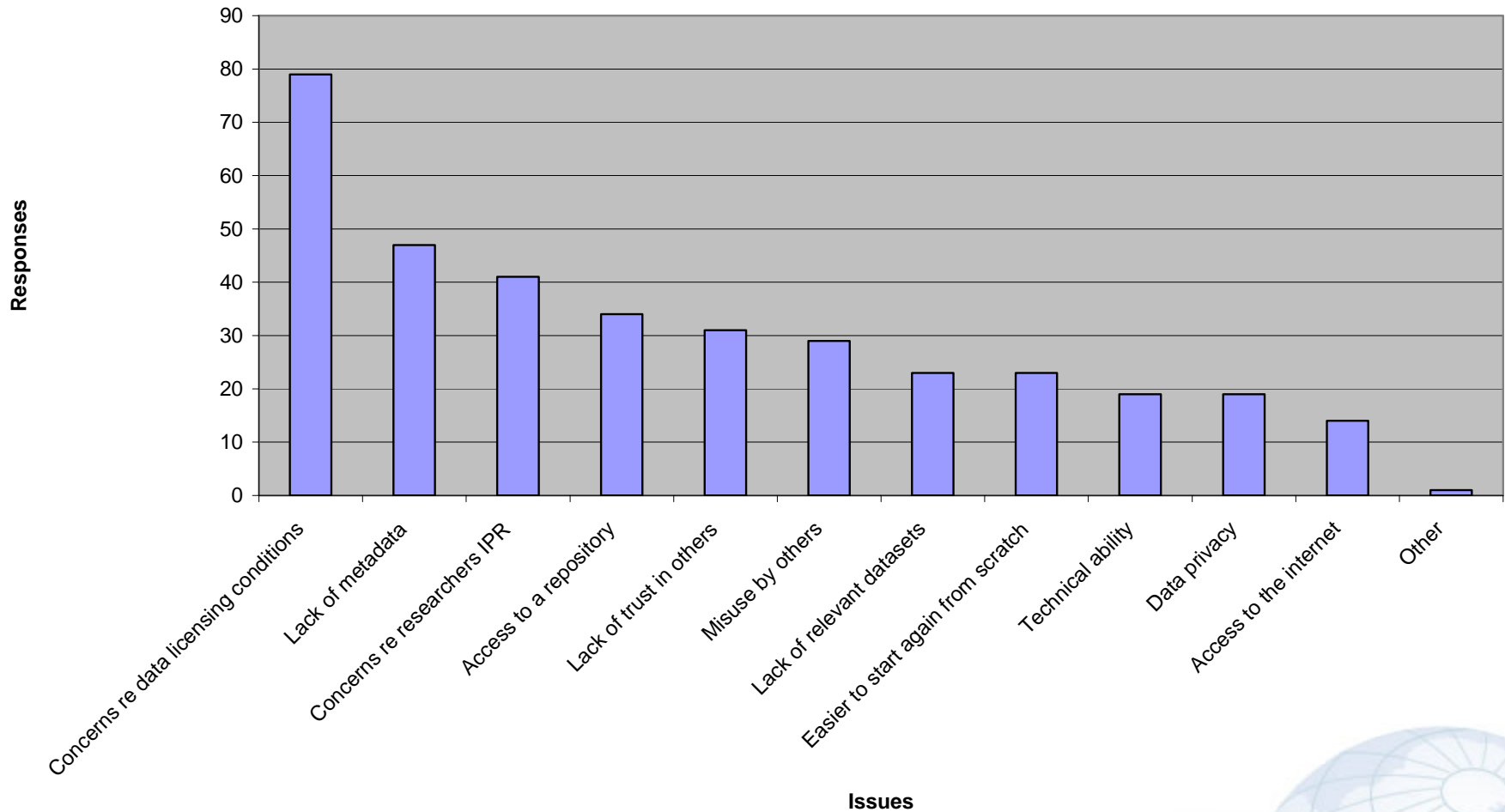
- Work led by AHRC Research Centre Studies in Intellectual Property and Technology Law
- Ground-breaking report on legal findings for a licensing framework
 - Copyright does not subsist in geospatial data rather EU Database Directive is appropriate law
 - A licensed user of a geospatial database can use an insubstantial part of the database as they see fit (this includes redistribution of derived data)
 - Where a researcher or teacher deposits outputs which comprise only an insubstantial part of the contents of a source database for which they are a lawful user, no permissions are necessary from the original maker and no attribution in relation to the data is needed. (even if substantial part, still no permissions required)
- Work shared with key groups (JISC Collections, JISC IPR Advisory Group, JISC GWG, OGC geoDRM)

Findings: WP on current data sharing practices

- Anonymous questionnaire survey
- Over 100 respondents
- Findings
 - Over 78% responses said had shared data, many qualifying that it happened on a frequent basis
 - Most common method of sharing was via CD/DVD and email attachment
 - Top mechanisms for finding geospatial data: Google followed by EDINA services, US geospatial clearinghouses, word of mouth
 - Most common barrier to sharing relates to concerns over breaking licensing conditions, lack of quality metadata and protection of individual's IPR
 - What would make reuse and sharing of geospatial data easier? less restrictive licensing and national geospatial data repository

Findings: WP on current data sharing practices

Figure 2: Barriers to sharing geospatial data



Findings: WP on current data sharing practices

- Interviewed our 4 associate partner sites
 - Lack of departmental/institutional policy on data management – sharing/reuse/long term access.
 - Poor metadata practices
 - Need to capture post graduate data
 - Current method of accessing data is to contact researcher created data
 - Data sharing via people networks
 - Data audit at 4 sites, 900 data sets that could be made available for reuse



Findings – WP on Repository Functionality

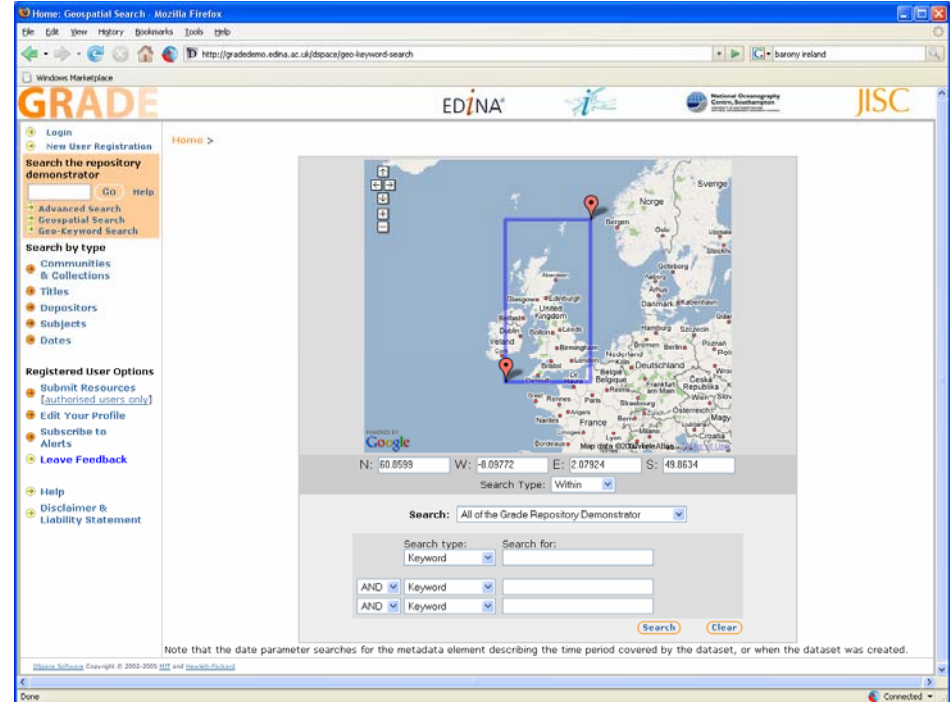
Demonstrator repository to gather user-based evidence

- Searching for data
 - Drawing box on map
 - Clicking a recognisable area on a map (e.g. county boundary)
 - Place name
 - Post-code
- Assessing fit for purpose
 - Dublin core metadata elements sufficient
 - View extent of data on map & thumbnail view of data & data format

The screenshot shows the GRADE Repository Demonstrator website. The browser window title is "Home: Home - Mozilla Firefox" and the address bar shows "http://gradedemo.edina.ac.uk/dspace/index.jsp". The website header includes the GRADE logo, EDINA logo, National Oceanography Centre logo, and JISC logo. The main content area is titled "Welcome to the GRADE Repository Demonstrator" and contains two main sections: "Search & Download" and "Submit". The "Search & Download" section features a map of Ireland with a magnifying glass over a specific area, and a "Search" button. The "Submit" section features a map of Ireland with an envelope icon, and a "Submit" button. The footer of the website includes the text "Most recent upload: Raw Tracklog of Gower Peninsular Footpath Survey" and "Last Modified: 27/03/2007".

Findings: WP on Repository Functionality

- Data Download
 - Directly to desktop for use within GIS
 - Full geospatial metadata (human readable)
- Data Upload
 - Ability to upload packaged project data
 - All respondents prepared to provide geospatial metadata (e.g. data format, coord system)
 - Automatic creation of deposit metadata esp min, max x,y
 - Assistance with licensing statement



- Daily search activity
- 140+ datasets and 100+ registered users
- Requests for access from many sectors, data hungry

Findings WP on Institutional repositories

- Baseline audit of existing institutional repositories
- Survey findings confirmed geospatial data not currently within IRs
 - * Many IRs dealing with publication outputs only (pay lip service to inclusion of data)
 - * IR software don't have ready made metadata schema for data, (if IR software vendors developed a Dataset plug-in it is possible that Institutional Repositories would have already been challenged to manage them)
 - * Simply haven't been offered geospatial data yet
 - * Survey suggests would be willing to consider managing geospatial data if offered but treat it in the same way as any other data
- Where designated data centres exist it is unlikely that Institutional Repositories will be the archive of choice for datasets, but there are many disciplines where there is no formal data archive available, or the data centre has a strict scoping on size and subject of datasets they accept. In these cases rather than datasets existence being hidden on a personal pc, it is possible IRs may have a role to play

SWOT analysis of IRs dealing with geospatial data

STRENGTHS of an IR dealing with geospatial data

- One repository – less administrative and technical overhead
- Linking text, datasets, images easier within one environment
- Showcase for **all** institutional research
- IR Software - Open Access – interoperability – visibility
- Software based on International Standards
- Metadata skills provided by Information community
- Formal Dataset citation
- Supports Citation analysis and metrics for research funding and personal promotion for data generators/managers

OPPORTUNITIES of an IR dealing with geospatial data

- Contribute to the design of an IR dataset metadata module
- To offer a data archive (where non exists)
- Treats 'orphan' datasets not accepted by DCs
- Enhancement of IR staff skills
- Showcase in one digital repository of **all** research output
- Ready host when dataset deposit is mandated
- Integration – joined up research
- Additional Funding opportunities from e-Research projects
- Input to the Data citation model
- Data and Information communities working together
- Collaboration between disciplines
- Dataset harvesting from IRs to Data Centres

WEAKNESSES of an IR dealing with geospatial data

- Software not designed to cope with data
- No IR metadata schema for datasets yet
- IR staff without Data Processing skills
- IRs do not quality control content
- IRs not involved in production of information products
- Storage – Preservation (all media types)
- OA culture not yet extended to data altho OEDC, EU and some Research Councils etc. mandate deposit of data emanating from funding

THREATS of an IR dealing with geospatial data

- Turf war between IRs and DCs
- Will funding follow to IRs
- Will funding stream for data management reduce?
- Too large an undertaking for IRs
- Data lost in publication 'bucket'
- 'Thematic' datasets distributed
- No migration/preservation policy
- Datasets fall 'between stools'

GRADE: Summary of main findings

- Repositories do have a part to play in facilitating the sharing and reuse of research-generated geospatial data
- A significant degree of informal geospatial data sharing occurs because of the lack of any formal mechanisms
- Community desire for a mechanism to legitimately share and reuse geospatial research data
- Main barriers to more formal geospatial data sharing within the community are:
 - perceived complexity of licensing and digital rights issues surrounding data (re)use in the UK
 - lack of quality metadata
 - concerns over the protection of depositors intellectual property

Main Findings contd.

- Institutional repositories do not manage any geospatial content (and would not be capable of meeting the needs of those working with geospatial data currently)
- Geospatial data creators would support data reuse BUT not necessarily (at present) within an OA IR. More fine grained sharing mechanisms are preferred i.e. data sharing amongst peer group networks defined by the depositor

Thank you

Anne Robertson, EDINA

<http://edina.ac.uk/projects/grade>

a.m.robertson@ed.ac.uk

GRADE

