Library vs publisher vs archive: managing the UK topographic record





Kimberly C. Kowal Curator of Digital Mapping The British Library kimberly.kowal@bl.uk

Jonathan Holmes NGD & Corporate Data Manager Ordnance Survey jonathan.holmes@ordnancesurvey.co.uk

10 June 2014

Open Repositories 2014

Non-Print Legal Deposit Regulations

•By law, a copy of every UK print publication must be given to the British Library by its publishers, and to five other major libraries that request it. This system is called legal deposit and has been a part of English law since 1662.

•As of 6 April 2013, legal deposit also covers material published digitally and online, so that the Legal Deposit Libraries can provide a national archive of the UK's non-print published material, such as websites, blogs, e-journals and CD-ROMs.





CAMBRIDGE UNIVERSITY LIBRARY



UK Legal Deposit Libraries

- The British Library
- Bodleian Library, Oxford
- Cambridge University Library
- National Library of Scotland
- National Library of Wales
- Trinity College, Dublin



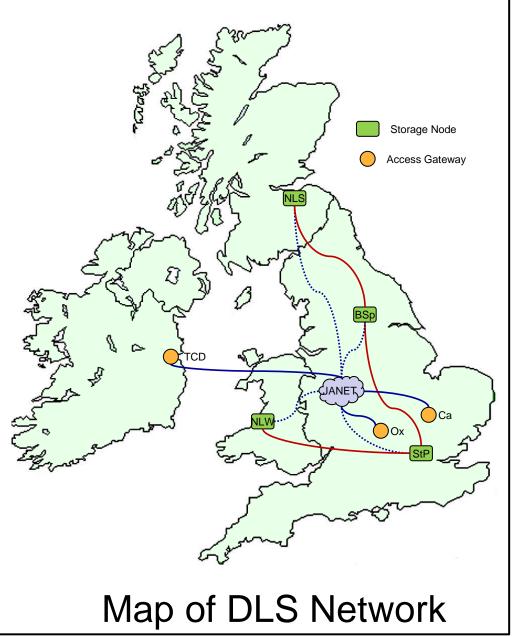


Bodleian Library

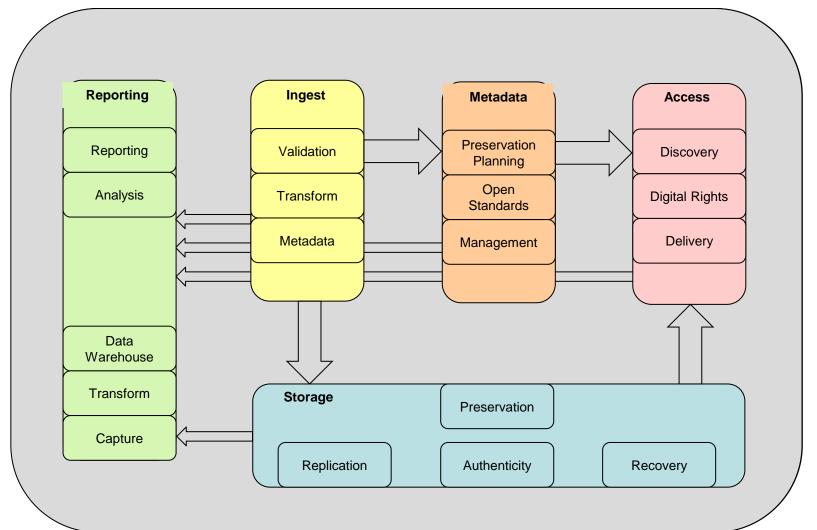


Digital Library System Mission

- Provide a secure place to store digital material created and acquired by the BL, forever
- Acquire Digital Legal Deposit Material on behalf of the other Legal Deposit Libraries and share it with them
- Ensure that all material is permanently findable and accessible
- Provide the widest possible access whilst respecting the wishes of rights holders
- Make Digital 'Business as Usual' at the BL
 3

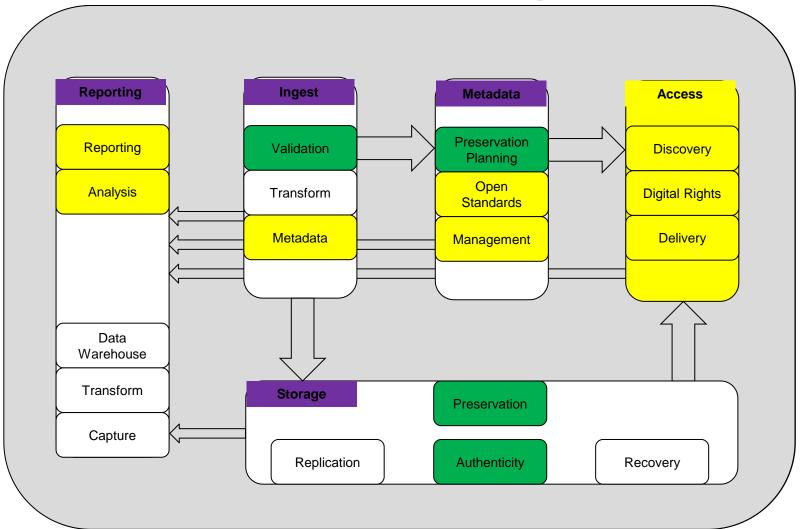


DLS Software Components



Access system Preservation actions Ingest to Digital Library System

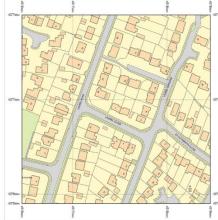
Componentsspread



Ordnance Survey (OS) digital data

Commercially valuable, large scale mapping of GB, source scales 1:1,250; 1:2,500; 1:10,000

- 1998-2007 Land-Line (ntf)
- Land-Form Profile (dtm)



- 2006- onwards OS MasterMap Topography Layer[®] (gml)
- 2006- onwards OS MasterMap[®] Integrated Transport Network[™](ITN) Layer



"In the sprit of long-term archiving"

Obligations and Restrictions:

Required secure, user rights management framework to control application and data access Access limited to readers inside the deposit libraries Printing – A4 extracts for non-commercial users Digital copying for the Library (*long term preservation*) No downloading Usage and security reporting

Geospatial datasets: Access

Software 1 - standalone application with data stored in each LDL, 1999-2006

Software 2 - networked customised web mapping application, hosted and managed, 2006-2013

Cooperative arrangement with formal LDL approval •European Union procurement - 'restricted procedure' - Jan - May 2006

Software 3 - networked customised web mapping application, hosted and managed 2014-2019

Cooperative arrangement with formal LDL approval

•European Union procurement - 'restricted procedure' - Jan - Aug 2012

The system was to provide a generic and cost-effective infrastructure designed to:

 Be flexible and scalable, accepting additional data of various types and from numerous sources

• Take advantage of open source technologies, which are cheaper, more effective, and widely used

 Hold metadata about map content in a more open, accessible form, allowing for easy transfer into other repository systems

Allow controlled access to accommodate legal deposit regulations

 Reflect developments in web-mapping technologies and improvements to the map viewing interface



Geospatial datasets: Preservation



Scalable Preservation Environment

- SCAPE: EU project to develop tools and services to for large-scale digital repositories
- Research Datasets testbed: Identification, validation and checksumming of a complex corpus
 - A tool to create/check fixity of files
 - A tool to identify the files in the corpus
 - A tool to validate files adhere to the format specifications, where appropriate
- Integration of SCAPE toolset with DLS



Geospatial datasets: Digital Library System

Metadata + Ingest Creating a new Content Stream for Geospatial Data

- The Library uses the Metadata Encoding and Transmission Standard (METS) schema to convey the metadata required to manage the digital objects held within the Digital Library System (DLS).
- **DLS Strategic Ingest:** Metadata Extension Repository is meant to cover all content streams
- Defining a Content stream:
 - mapping onto the SI METS profile.
 - creation of data model
 - developing software to create METS file and ingest



Geospatial datasets: Digital Library System

Ingest + Metadata Creating a new Content Stream for Geospatial Data

- What makes it different from other library materials?
- Usage unfamiliar
- Diverse formats, complex relationships
- Unfamiliar descriptors
 - Coordinate reference systems
 - Cartographic representation
 - Topology (data structures)
 - Project files, data packaging

How do the data producers handle managing data in a repository

- The Ordnance Survey
- The National Archives

Over to Jonathan...

Background



- Ordnance Survey
 - Founded in 1791
 - Responsibility for mapping Great Britain (England, Scotland, Wales)
 - Currently a Government Trading Fund
- The National Archives (TNA)
 - Formerly the Public Records Office

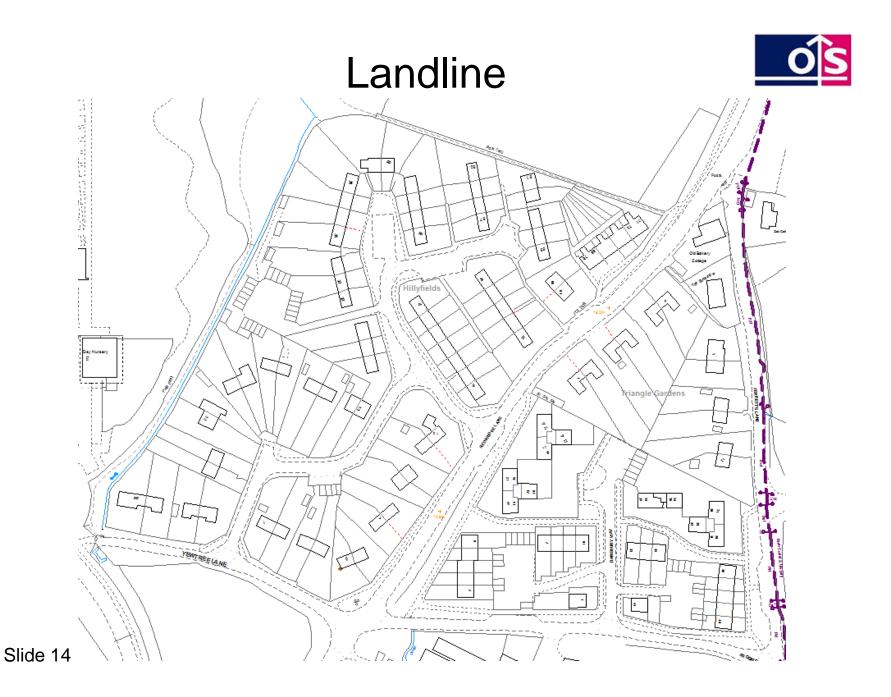


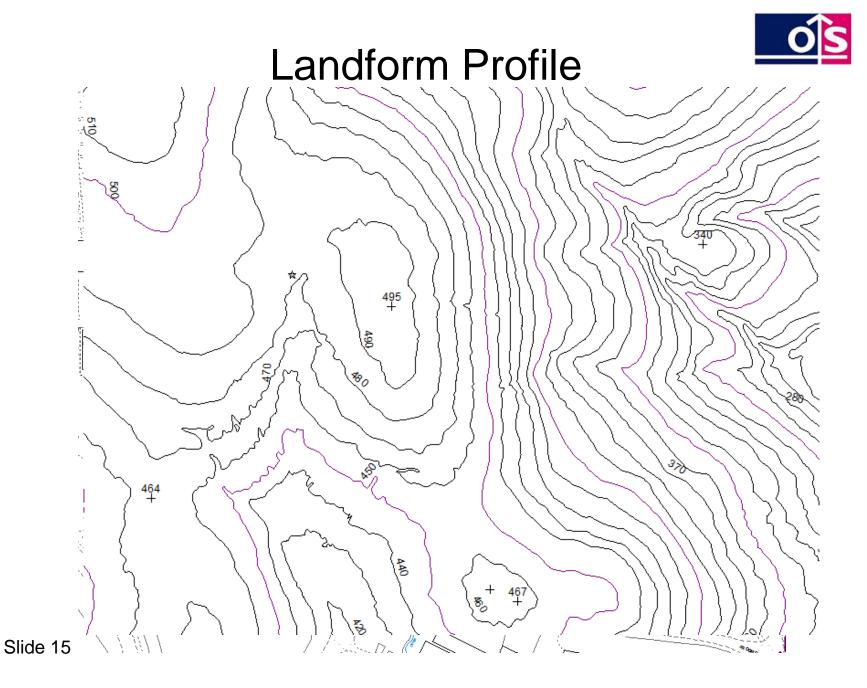
Founded in 1838 to reform the keeping of government and court records

Background



- In 1999 Ordnance Survey and The National Archives signed an agreement for Ordnance Survey to become the first Electronic Place of Deposit (POD)
- Stores Landline, Landform Profile and OS MasterMap







OS MasterMap



Why Ordnance Survey became a PoD

70 07TL2038 210000620000000100000620000000100000520000000238000000000001% 72 00 73 00 00000000 000000019901203B 18 74 002002080820041208LSDS 051994v1.0 0000000 00000000 00000000 0000000081% 75 00WARDS IN MID BEDFORDSHIRE DISTRICT OPERATIVE 01/05/2003\N000000008 76 23000004 0030 7010702 0000000% 77 210000020010100000044760 096900036730 086970040720 088420048960 098340045840 1% 78 00098486045254 098930045110 099470045460 099680045420 100000045315 0% 79 23000028 0030 3940908 0000000% 80 210000020022051485088440 051560088440 051590088450 051610088470 051540088610 1% 81 00051540088690 051550088980 051590089430 051640089760 051670090120 0516801% 82 00090270 051680090430 051620090830 051550091370 051480092100 051390093110 1% 83 00051290094100 051200095140 050930097830 050800099230 050740099770 0507401% 84 00100000 0% 85 23000030 0036 3970224 0000000% 86 210000020011038682091196 038687090556 040280090170 041200089970 042240089700 1% 87 00043430089350 044470089070 045260088870 046170088670 047160088500 0482871% 88 00088290 0% 89 23000031 0030 5940908 0000000% 90 210000020002036140088350 036120087757 0% 91 23000036 0030 7031010 0000000% 92 2100000020008001200056930 000800055410 000600054610 000390053890 000320053660 1% 93 00000250053440 000150053180 000001052934 0% 94 23000037 0030 5940908 0000000% 95 210000020002054123057657 054550057260 0% 96 23000045 0030 5940908 0000000% 97 2100000020017091837082304 093190082630 093950082790 094230082870 094500082920 1% 98 00094790082950 095120082980 095390083000 095870082990 096440082970 0970501% 99 00082960 097490082930 097760082920 098200082920 098690082970 099260083020 1% .00 0010000083100 0% .01 23000046 0030 5940908 0000000% .02 210000020063090540056280 090570056390 090630056500 090780056660 090870056770 1% .03 00090970056880 091100056980 091250057060 091420057130 091530057180 0916101% .04 00057210 091700057220 091980057160 092150057130 092320057110 092520057110 1% .05 00092710057140 093010057220 093140057250 093260057280 093400057330 0935001% .06 00057360 093610057380 093710057370 093820057360 094050057370 094230057360 1% .07 00094370057340 094510057290 094790057100 094920057010 095040056920 0953901%

Why Ordnance Survey became a PoD

- <osgb:descriptiveGroup>Network Or Polygon Closing Geometry</osgb:descriptiveGroup>
- cosgb:descriptiveTerm>Polygon Closing Link</osgb:descriptiveTerm>
- <osgb:nonBoundingLine>true</osgb:nonBoundingLine>
- <osgb:physicalLevel>50</osgb:physicalLevel>
- <osgb:physicalPresence>Closing</osgb:physicalPresence>
- <osgb:polyline>
- <gml:LineString srsName='osgb:BNG'>
- <gml:coordinates>403999.020,411237.110 404000.000,411237.260 </gml:coordinates>
- </gml:LineString>
- </osgb:polyline>
- </osgb:TopographicLine>
- </osgb:topographicMember>
- <osgb:topographicMember>
- <osgb:TopographicLine fid='osgb1000002102987241'>
- <osgb:featureCode>10113</osgb:featureCode>
- <osgb:version>1</osgb:version>
- <osgb:versionDate>2011-05-03</osgb:versionDate>
- <osgb:accuracyOfPosition>1.0m</osgb:accuracyOfPosition>
- <osgb:changeHistory>
- <osgb:changeDate>2011-05-03</osgb:changeDate>
- <osgb:reasonForChange>New</osgb:reasonForChange>
- </osgb:changeHistory>
- <osgb:descriptiveGroup>Network Or Polygon Closing Geometry</osgb:descriptiveGroup>
- <osgb:descriptiveTerm>Polygon Closing Link</osgb:descriptiveTerm>
- <osgb:nonBoundingLine>true</osgb:nonBoundingLine>
- <osgb:physicalLevel>50</osgb:physicalLevel>

Place of Deposit System



- Written in-house using Visual Basic and Unix
- Simple system
- Ingests a snapshot of the live product
- Two backups are made of the data
 - One stored locally and one remotely
- File structure is also simple
 - Product
 - ≻Year
 - ➤ 100km grid square

Place of Deposit System



- Data specification and product specifications are stored in the system
- Metadata very limited
- There is no viewing tool. Data has to be extracted and loaded into a GIS

Ordnance Survey – Place of Deposit



Ordnance Survey Place of Deposit	
Ingest Dat	a
Create CRC Va	lues
Check CRC V	lues
Exit	
	V2.1 Mar 201;

System use



- Data is used internally
- Used externally in response to Freedom of Information requests
- By 2019 Need to start to make data archived for TNA available for viewing publicly
- Under the terms of the Act very limited copying for the purposes of:
 - Private study;
 - Non-commercial research, criticism, review and news reporting, (provided that any copy is accompanied by the following acknowledgement ©Crown Copyright. Reproduced by permission of Ordnance Survey®;

or for

- Parliamentary or judicial proceedings.
- How we do this is yet to be determined

Challenges for Ordnance Survey

- Only some of the products are stored in the PoD
- Many of the other products are strewn across drives, networks and servers or in cupboards.
- Much of this data has a short / medium term use internally but it is not stored securely.
- Storage and maintenance in the medium term requires a more sophisticated system than we currently have.
- What do we do with the data longer term?
- Should the archiving processes of Ordnance Survey and the British Library be linked with some sort of accession plan?

Conclusion



- OS, TNA and the LDLs are three separate bodies with different but similar obligations. There are clear synergies between them and there are clearly opportunities for the three organisations to work together.
- Should the same data be held in more than one place for more than one purpose? This seems inefficient.
- Can this data just be held once for the nation?