

National Geospatial Digital Archive

Greg Janée

Outline

- Two preservation misadventures
- Digital preservation problems
- Genesis of NGDA
- Project approach/philosophy
- What it will mean to be a data provider

Domesday Book— 1086



¶ Dunelmie ten' edric' de lefeldde . t . r . e . pmo man' . e' m'
ten' . hobe' malet . Ine' . ii . car' t'z . m' . i . mare abstulit alia .
¶ sep' . i . car' indnio . Ine' . xii . bord' . m' . ii . ¶ xxxiii . franci .
hoes' . de xlac' t'z . ¶ omne' Consuetudine' redat' huic manerio .
¶ Ine' c . ¶ xxx . burgenses . ¶ m' . cc . ¶ xxxv . ¶ paupes hoēs
c . quat' uiginti . ii . min' . Ine' . i . ecclā . m' . iii . ¶ reddunt . iii .
lib' . ¶ x . sol' . ¶ int' totū ualē . l . lib' . e' lx . mill' allectium .

de dono . Et t . R . E . redat' . x . lib' . ¶ e' adhuc ten' roeb' de uallib;
¶ i . ac' t'z . uat' . viii . d' . Et norman . i . ac' ten' . quat' . ii . sol' . ¶ viii . d' .
¶ godric' . i . ac' . uat' . viii . d' . ¶ hoc ten' de . R . malet . ¶ Gislebe'
blundus . quat' uiginti hoēs ten' . ¶ redat' . iii . lib' . ¶ viii . mill' allec
tium de eodē . R .

Domesday Book— 1986



Guardian Unlimited Books | News | Digital Domesday Book lasts 15 years not 1000

http://books.guardian.co.uk/news/articles/0,6109,661585,0 Google

News

Digital Domesday Book lasts 15 years not 1000

Robin McKie and Vanessa Thorpe
Sunday March 3, 2002
[The Observer](#)

It was meant to be a showcase for Britain's electronic prowess - a computer-based, multimedia version of the Domesday Book. But 16 years after it was created, the £2.5 million BBC Domesday Project has achieved an unexpected and unwelcome status: it is now unreadable.

The special computers developed to play the 12in video discs of text, photographs, maps and archive footage of British life are - quite simply - obsolete.

As a result, no one can access the reams of project information - equivalent to several sets of encyclopaedias - that were assembled about the state of the nation in 1986. By contrast, the original Domesday Book - an inventory of eleventh-century England compiled in 1086 by Norman monks - is in fine condition in the Public Record Office, Kew, and can be accessed by anyone who can read

Meanwhile, back at NASA...

- 1976
 - Viking probes go to Mars
- 1999
 - USC neurobiologist Joseph Miller asks for data
 - tapes coded “in a format so old that the programmers who knew it had died”
 - works off of paper records

Preservation issues

- Physical
 - media
 - systems
- Contextual
 - format
 - semantics
 - authenticity
- Legal
 - copyright

Project genesis

- NDIIPP
 - Library of Congress, 2000
 - \$100M
 - <http://www.digitalpreservation.gov/>
- NGDA
 - UCSB (MIL) & Stanford (Branner Library)
 - \$2.6M, 3 years
 - archive geospatial data on a national scale
 - <http://www.ngda.org/>



http://www.ndga.org/



Google



The Registry & Sanctioning Association Dedicated Exclusively
to the **Nigerian Dwarf Goat**

Important News!

[NDGA News Bulletin](#)



[Click to subscribe to Lets Talk NDGA](#)

[Full website reviews](#)
[Finding great goats websites fast.](#)

[New Merchandise for Sale](#)

Some philosophy

- Archival has to be cheap & easy
 - must be distributed
 - but reality is little incentive, no funding
- Archive definition:
 - offers access now & in the future
 - no mandatory services beyond simple access
- Policy separated from mechanism
- Archive includes data semantics
 - key differentiator from text, audio, video

Philosophy, cont.

- Curatorial, not archeological approach
 - assumption: content comes in discrete, self-contained chunks
- Preservation by format definition, archival & association
 - support for derivative forms, services
- Must support *long-term* preservation
 - need to migrate archive itself



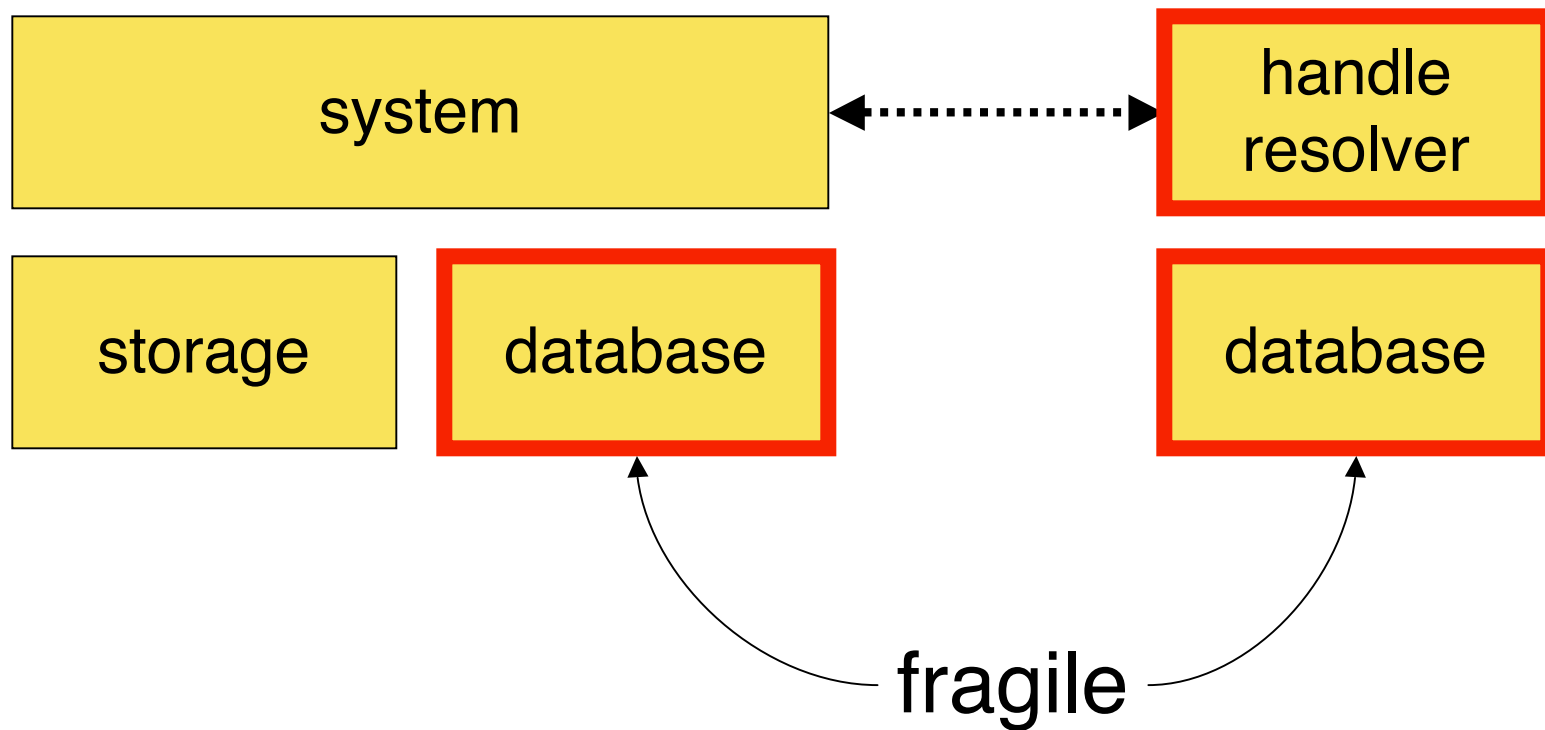
MIT Media Lab
Stewart Brand, "How Buildings Learn," p. 53



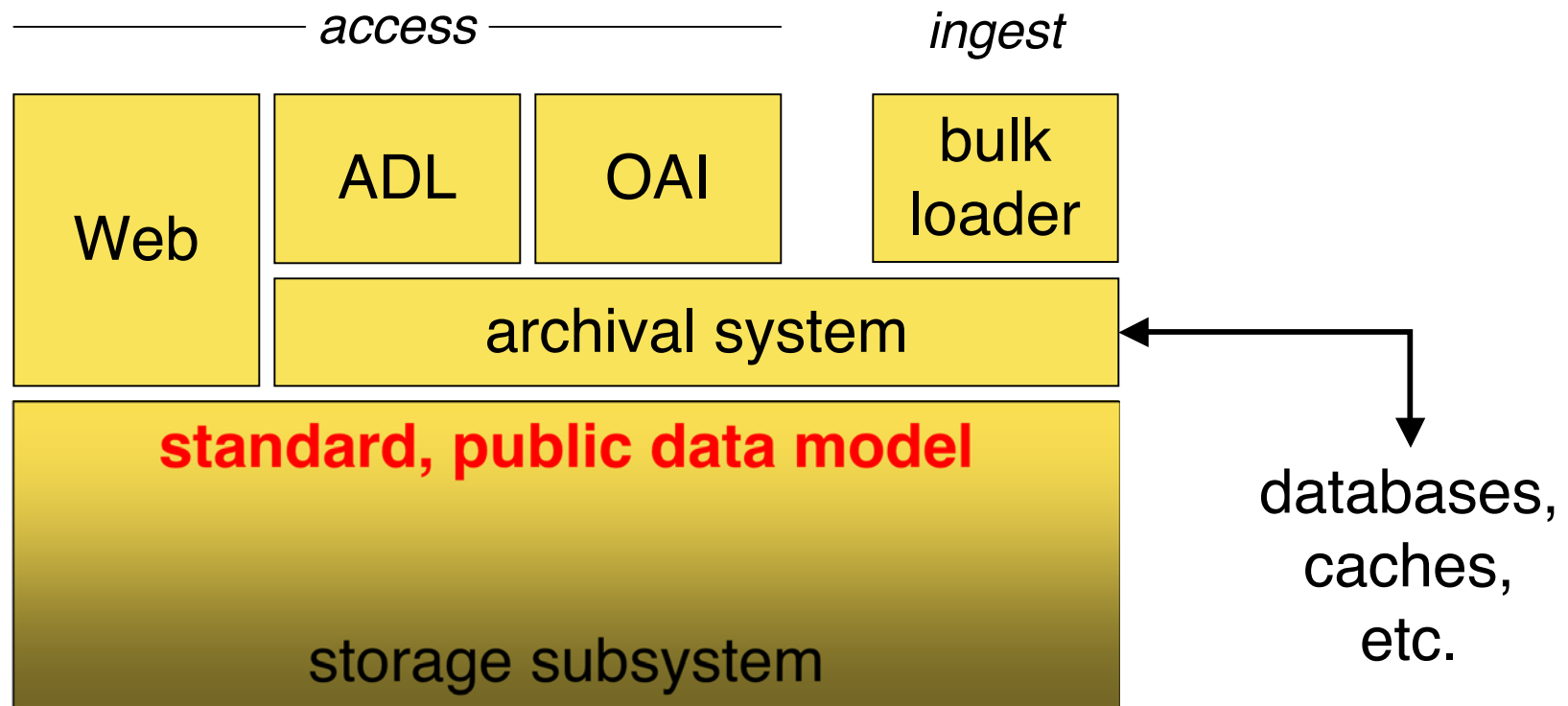
1990 - Author Fred Hapgood wrote in 1993 of Building 20, "The edifice is so ugly that it is impossible not to admire it, if that makes sense; it has ten times the righteous nerdy swagger of any other building on campus." (*Up the Infinite Corridor*, New York: Addison-Wesley)

MIT Building 20
Ibid., p. 26

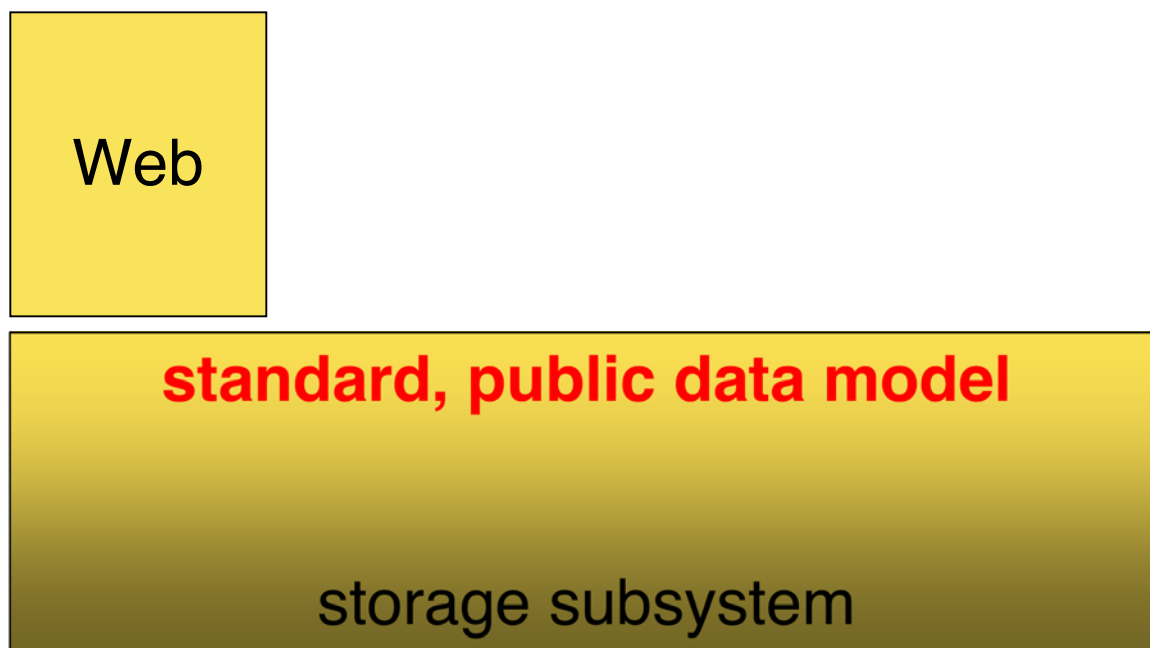
Typical repository architecture



NGDA architecture



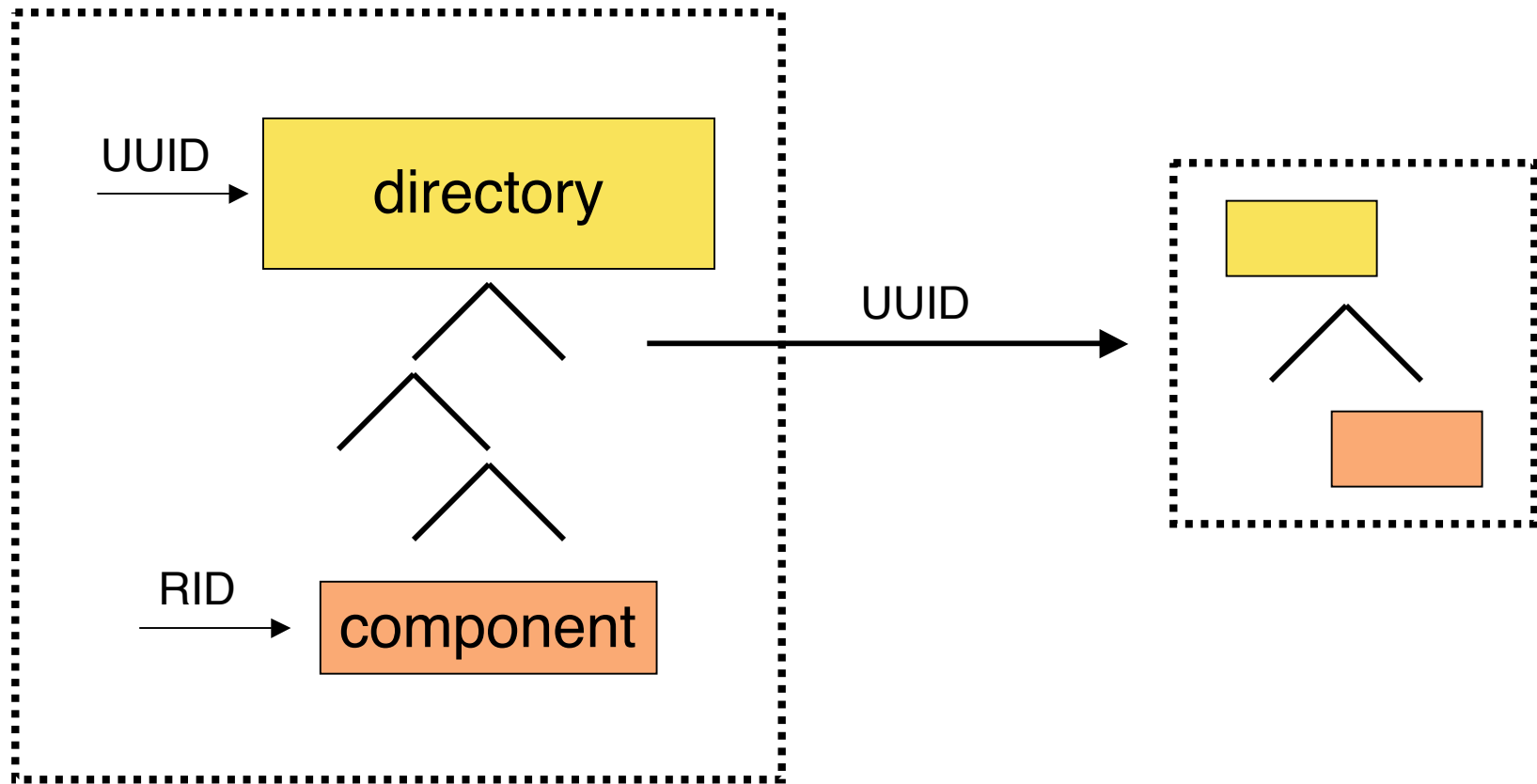
Post-NGDA architecture



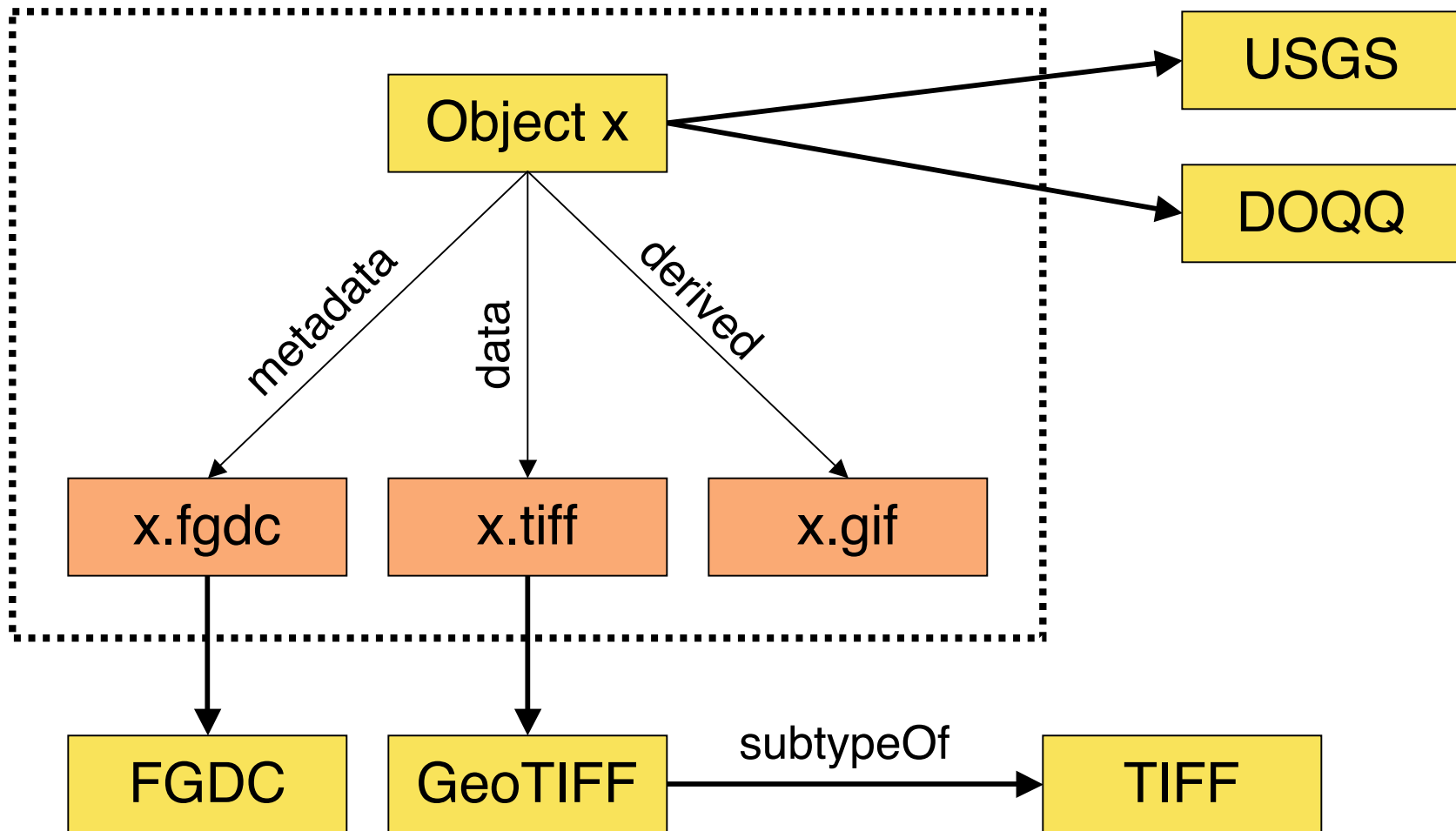
Storage system requirements

- Req's:
 - associate UUIDs/RIDs with bitstreams
 - retrieve global/local bitstream by UUID/RID
 - determine (parent) UUID of any bitstream
 - list all UUIDs
- Satisfied by:
 - any filesystem
 - tag URIs for UUIDs
 - tag:library.ucsb.edu,2005: *identifier*

Archival objects



Example



Object types

- Data, other content
- Format definitions
- Semantic definitions
- Providers
- Organizational structures
 - collection
 - series
 - ingest session

Archive-provider agreement

- Defines
 - common structure of objects to be ingested
 - necessary validations
 - associations to other objects
 - assumes pre-loading of semantic definitions
 - policies, rights, etc.
- Represents choke point
 - requires human evaluation