ELSEVIER

Space Policy 17 (2001) 35–48

Space Policy

www.elsevier.com/locate/spacepol

Earth observation data archiving in the USA and Europe

Ray Harris*, Nicola Olby

Department of Geography, University College London, 26 Bedford Way, London WC1H 0AP, UK

Abstract

The growing number of Earth observation satellites are producing exponential increases in the amount of data produced. These data sets contain valuable information on the state of the environment to accurately monitor and predict the impact of global change. Global change investigations require access to long-term data. This paper is a review of the current archiving practices of NOAA, NASA, SPOT, ESA and the DLR. It is demonstrated that despite recognition at a regional and organisational level of the importance of data archiving much work remains to be done. Funding responsibility is identified as a particularly problematic area of data archiving: it is an issue along with the need for easy accessibility and adequate metadata that will need to be resolved to ensure that future generations have access to adequate Earth observation data archives. © 2001 Elsevier Science Ltd. All rights reserved.

1. Introduction

1.1. Earth observation data archives

Earth observation satellites provide an important and unique source of information for studies of the Earth's system. There are currently over 45 Earth observation satellite missions operating and around 70 missions, carrying 230 instruments, planned for operation in the next 15 years by the world's civil space agencies. These future missions will create vast amounts of data, much of which will need to be preserved.

Data from Earth observation satellites are often the only source of information available on the condition and dynamics of large parts of the Earth's surface and atmosphere at appropriate spatial and temporal scales. Monitoring activities related to climate and environmental change are important. Global change investigations have created new challenges for researchers, data managers and decision makers because of their multi-disciplinary nature, their emphasis on data sharing and exchange, and the need for long-term data sets [1,2].

This is one reason why the long-term archiving of such data has become essential to environmental monitoring and the scientific modelling of terrestrial processes. Data preservation is required both for the immediate needs of the user community as a record of the Earth system, yet it also has an undefined value for future generations since the applications by future generations are unknown [3]. Indeed, as noted by Thomas Holm, of the USGS EROS Data Center:

> Experience has shown that after a period of time (for example, 10 years) the value (not necessarily measured in economic terms) of the archived remotely sensed data increases as it becomes an indelible historical record of the condition and configuration of the landscape [4].

1.2. Future generations

The world community is taking some action to safeguard the interests of future generations through international environmental treaties, protocols and conventions. For example, the United Nations Educational, Scientific and Cultural Organization (UNESCO) has adopted a Declaration on the Responsibilities of Present Generations Towards Future Generations. The declaration underlines the opinion that present generations should consider the needs and interests of future generations [5].

The aim of the Declaration is to ensure a viable future for future generations. The preamble states that:

> The fate of future generations depends to a great extent on decisions and actions taken today, and that present day problems, including poverty, technological and material underdevelopment, unemployment,

* Corresponding author. Tel.: +44-20-7679-4283; fax: +44-20-7679-4293.

E-mail address: r.harris@geog.ucl.ac.uk (R. Harris).

0265-9646/01/$ - see front matter © 2001 Elsevier Science Ltd. All rights reserved.
PII: S0265-9646(00)00052-7

exclusion, discrimination and threats to the environment must be solved in the interests of both present and future generations.

Of the 12 articles in the declaration, articles 4 and 5 are particularly related to Earth observation.

Article 4: Preservation of life on Earth

The present generations have the responsibility to bequeath to future generations an Earth which will not one day be irreversibly damaged by human activity. Each generation inheriting the Earth temporarily should take care to use natural resources reasonably and ensure that life is not prejudiced by harmful modifications of the ecosystems and that scientific and technological progess in all fields does not harm life on Earth.

Article 5: Preservation of the Environment

In order to ensure that future generations benefit from the richness of the Earth's ecosystems, the present generations should strive for sustainable development and preserve living conditions, particularly the quality and integrity of the environment.

The present generations should preserve for future generations natural resources necessary for sustaining human life and for its development.

Given these responsibilities the destruction of valuable or potentially valuable Earth observation data could be considered environmental vandalism. This also highlights one of the major problems associated with data archiving, the fact that preservation of data is based on perceived need, which may well alter in the future. Who is to say whether data currently considered to contain little value will not be re-evaluated in the future. A prime example of this was the requirement to re-examine many years of satellite data once depletion of the ozone layer had been discovered [6].

1.3. Value

A key value of Earth observation data archives is to increase the usefulness of current data by setting it against the background of historical data. Earth observation data archives are an essential part of the process of analysing the past so that we can understand the present and thereby predict the future. The answers to the question of why to keep Earth observation data can be summarised as:

- To reconstruct the past.
- To obtain insight into and understanding of the past.
- To develop models of environmental processes.
- To predict the future.

Major environmental programmes such as the International Geosphere Biosphere Program (IGBP) and the Global Ocean Observing System (GOOS) rely on archives of Earth observation data. Archives of Earth observation data are becoming of increasing value in assessing adherence by nations to the Kyoto protocol because both base year data and change data are required. Within Europe Earth observation data archives contribute to the assessment of policies on agriculture, environment and global change by allowing the examination of changes over time. There is a significant potential for the European Union, for example, to exploit Earth observation data archives in its plans for enlargement (by analysing land management in eastern Europe) and in regional development (by assessing investment in less favoured areas). In both cases Earth observation data archives allow the evaluation of environmental change and environmental management processes which then provide a scientific basis for policy decisions.

1.4. Data archiving

Data archiving is a problematic area within Earth observation data policy, and raises many issues such as the amount of data and processed products to archive, the data preservation temporal scale and the associated metadata questions. Moreover, since much of the data can be regarded as a public good, the issue of who is responsible for data archiving costs is a long standing and complex one.

Many of the issues mentioned above are not directly compatible with the objectives and goals of those organisations assigned the responsibility of archiving, operating in an environment of ever tightening budgets, and the fact that transgenerational policies are not presently, nor likely to be universally, implemented. Yet data archiving requires coordinated, long-term policies and guaranteed long-term funding. This compounds the problems associated with data archiving since the requisite resources necessary to create and run an efficient data archive are often not in place.

Today most large Earth observation satellite organisations operate data archives on varying scales. Table 1 presents a list of a number of organisations which operate data archives at global, regional and national scales.

1.5. Purpose of the paper

This paper will discuss the issues of Earth observation data archiving, focusing on the experience of selected organisations in the USA and Europe that have responsibility for Earth observation data archiving. Different

Table 1
Examples of organisations operating Earth observation data archives

	Name
World	World Data Center A for Remotely Sensed Land Data
Regional	European Space Agency
	EUMETSAT
National	Australian Centre for Remote Sensing
	Canadian Centre for Remote Sensing
	France Centre Nationale d'Etudes Spatiales/SPOT Image
	Germany DLR-DFD
	Indian National Remote Sensing Agency
	Japan National Space Development Agency
	Russia Space Research Institute
	US NASA
	US NOAA

attributes are emphasized for each of the organisations examined, namely:

- NOAA, showing the operational character of the data systems and archives.
- NASA, showing the geographical distribution of archives.
- SPOT, showing the level of control over the archive and the need for continuous maintenance.
- ESA, showing the size of the ERS SAR archive and considerations of long-term funding.
- DLR-DFD, showing the accessibility of the archive.

The paper does include a large number of acronyms. The acronyms are listed and explained in the appendix.

2. Archiving issues

2.1. Introduction

This section discusses the issues which are important in Earth observation data archiving policy, namely the length of data preservation, funding responsibility, data access, level of processing and metadata. The main challenges are also assessed.

2.2. Length of data preservation

Data archiving raises issues related to the length of time and amount of data which should be archived. All data could potentially be preserved indefinitely through periodic transcription. However, operators of Earth observation satellite systems usually make archiving provisions during the lifetime of a mission and for some specified period after the mission has been completed, although this is not always the case [7]. For example ESA's Envisat data policy states that "ESA will ensure that the logical and physical organisation of its Envisat product archives will be such that they can be maintained beyond the lifetime of the Envisat mission, with reconditioning as necessary" [8]. The CEOS satellite data exchange principles are wider in their coverage in stating that "Preservation of all data needed for long-term global change/climate and environmental research and monitoring is required" [9].

Given the undeniable trend for the continued rapid growth of Earth observation data acquisition it is important that data management practices and policies are constructed to cater for these increasing data inputs. In a world released from the confines of technological and economic realities all data could be archived in the same format indefinitely. Indeed in a study into the preservation of electronic materials the argument was raised that if a judgement has been made to archive data it should never be destroyed since it is not possible to judge what may be needed in the future and it is impossible to predict the future usefulness of the material [10]. This position is unrealistic in terms of an explicit data policy in the Earth observation sector because of the economic constraints and the large data volumes involved.

In most cases the statements of organisations in the Earth observation sector concerning archiving are rather vague. While the importance of archiving is supported, few explicit conditions are provided on how long the data should be preserved, for example:

- NASA Earth Science Enterprise Statement on Data Management: "All data required for long term global change research shall be archived" [11].
- USGCRP Data Management for Global Change Research Policy Statement: "Preservation of all data needed for long term global change is required" [12].

In the current context of rapid technological change, "long-term" may actually mean "short-term". For example, at the Archival Workshop on Ingest, Identification and Certification Standards (AWIICS) *long-term* was defined as "long enough to have technology become an issue": in recent years this period is becoming as short as five years [13].

Making explicit statements on archiving data over a long time period (i.e. 50–100 years) is likely to be a rarity given the funding arrangements of many space organisations reliant upon government funding. It is unrealistic for a government whose guaranteed term in office is short to make transgenerational policies on any issue. Few governments or institutions are going to make the long-term monetary commitments required to maintain an Earth observation data archive indefinitely despite the importance of the data that it may hold. Similarly, for commercial operators long-term financial commitment is unlikely to be guaranteed because of the turbulent nature of the economic environment, which raises the issue of whether commercial organisations with an explicit

motivation to make a profit should have to face the financial burden of funding an archive for what could be considered to be a public good.

2.3. Data archiving funding responsibility

The issue of who should carry the financial burden of data archiving is a difficult issue to resolve. The ultimate users of the preserved data are often not in a position to make financial contributions to the costs of data preservation since they are undefined. Clearly primary users of the data will emerge but the timescale is difficult to predict accurately. However, what is clear is that if funding arrangements are not put in place for the long-term archiving of data it will be these undefined future users who will ultimately suffer the consequences.

Few would argue against the importance of archiving Earth observation data, yet when it comes down to allocating budgets for this activity few organisations are prepared or able to make the long-term financial commitments necessary.

2.4. Data access

Access to Earth observation data archives is as important as the preservation of the data. Making data easily accessible to users poses one of the most challenging and expensive aspects of archiving. There is little value in an organisation making the large financial commitments necessary to manage a data archive if the archive becomes nothing more than a repository for the data. The potential value contained in Earth observation data sets necessitates that there is open and easy access to the data.

Access to data has been improved by advances in technology. The development of CD-ROM disks for data storage has increased the accessibility of remote sensing data to a wider customer base in addition to making remote sensing data distribution easier [14].

2.5. Level of processing

The level of processing to archive is an important issue since it has implications for the future use of the data. Level 0 data is generally the preferred level to archive since it is the rawest form of data, enabling re-processing of the data when new applications and new algorithms are developed or when failures in the processing chain are detected. An example of this is the ozone hole over the Antarctic which was discovered by ground measurements. Re-processing Total Ozone Mapping Spectrometer (TOMS) data revealed an error in the threshold set in the original algorithm resulting in the ozone hole going undetected. Subsequent processing of the satellite data has revealed important information on the time sequence and spatial extent of ozone depletion [15,16].

2.6. Metadata

Metadata consist of information that characterises the data. The availability of metadata for long-term data preservation is essential to the usefulness and reliability of the data in the future, in particular to ensure that errors in the data are apparent to future users, otherwise potentially valuable data could become meaningless. Proper use and analysis of remote sensing data requires an understanding of how the data were obtained. Currently, despite the efforts of several organisations including the European Commission, the adoption of active metadata use is weak.

The US National Research Council strategy for recording metadata used for global climate change research is encapsulated by the 20-year test, posing the question of whether someone 20 years from now would be able to find the target data sets and then fully understand and use the data with the aid of the documentation archived with the data set [16].

2.7. Main challenges

A major challenge for Earth observation data archiving is how to determine the value of the archive. Is the value the current sales price of all the archived data and products added together? Is the value the cost of maintaining the archive? Or is the value outside the market expression of what can be sold? In addition to economic value there are other ways of defining value, including social, environmental, scientific, humanitarian and security value.

Costanza et al. argue that, when monetised, the value of the world's ecosystem services and natural capital is of the order of US$33 trillion per annum, while global gross national product is around US$18 trillion per annum [17]. The definition of ecosystem services adopted by Costanza et al. includes many ecosystems for which archives of Earth observation data are vital to measure their extent and their changes over time, for example climate, environmental disturbance, erosion control and sediment retention, food production and the production of natural raw materials such as lumber. Table 2 is an extract from the information presented by Costanza et al. and shows for each of five biomes the value of the ecosystem services (in 1994 US$ $ha^{-1}\,yr^{-1}$) provided by climate regulation, environmental disturbance regulation and food production. Earth observation data archives can contribute significantly to the provision of information to determine the actual characteristics of these biomes, but little or no research has been published in Earth observation on this matter. It would be interesting to estimate the value of Earth observation data archives in relation to the categories of ecosystem services in Table 2, including forecasts of future value as ecosystems change with global change.

Table 2
Examples of average global value of ecosystem services. Extracted and adapted from Costanza et al. [17]

Biome	Area (million ha)	Ecosystem services (1994 US$ ha^{-1} yr^{-1}) Climate regulation	Disturbance regulation	Food production	Total US$ (millions)
Estuaries	180		567	521	195,840
Coral reefs	62		2750	220	184,140
Tropical rainforests	1900	223	5	32	494,000
Temperate/boreal forests	2955	88		50	407,790
Wetlands	330		4539	256	1,582,350

The use of Earth observation data archives to analyse the past so that we can understand the present and thereby predict the future is also of value for the social and humanitarian uses of Earth observation. Global climate change appears to be currently characterised by significant regional extreme events, for example flooding, droughts, rapid soil erosion and violent winds. Earth observation data archives provide the ability to determine the likely limits for such dramatic regional climate changes. The value of Earth observation data archives to humanitarian causes cannot be determined solely by the sale of data from archives. There is a value for civil society which goes beyond a willingness or an ability to pay for data in the short term [18].

A second major challenge for Earth observation data archiving policy is who pays for the archives in the medium and long term. As with so much Earth observation data policy, the organisational responsibility for Earth observation data archives comes down to the question of *who pays?* A number of possibilities for organisational responsibility are given below.

1. At present the main organisations responsible for Earth observation data archives are the mission management organisations, for example SPOT, ESA, EUMETSAT and NOAA. Typically, these organisations have budgets for each mission, but have to find recurrent budgets for archive maintenance and management. The incentive for these organisations to maintain their data archives given obsolete equipment, obsolete media and high costs is rather restricted. We return to this point later in the paper with the example of SPOT.
2. Climate and environmental monitoring are being presented with firmer targets. The Framework Convention on Climate Change, and related Kyoto protocols, combined with the impacts of climate change provide a rationale to justify environmental organisations increasing their responsibility for medium- and long-term Earth observation data archives in support of their own missions.
3. As digital technologies improve more generally then Earth observation can take advantage of these technologies rather than inventing its own methods. Earth observation data are part of the broad category of digital geographic data and therefore one can envisage geographic organisations adopting Earth observation data sets as part of larger geographic information archiving responsibilities. The Digital Earth initiative at NASA's Goddard Space Flight Center illustrates a step in this direction.
4. Major public libraries, such as the new British Library in London, maintain precious texts for future use. These institutions could, in principle, take an interest in Earth observation data archives, particularly as libraries change their role to be information providers rather than book and journal facilities.

These possibilities suggest that Earth observation data archiving is not a question just to be left to Earth observation data suppliers. The responsibility for maintaining important and sometimes unique data on the environment is broader than simply the Earth observation sector.

3. Data archives in the USA

3.1. Legal context

It was determined by the US Land Remote Sensing Policy Act of 1992 that it was "in the best interest of the United States to maintain a permanent, comprehensive government archive of global Landsat and other land remote sensing data for long-term monitoring and the study of the changing global environment". This was followed in 1996 by the release of the National Space Policy which directed the Department of the Interior, through the US Geological Survey (USGS), to maintain a national archive of land remote sensing data and other surface data as appropriate, and to make such data available to US government and other users. Thus establishment of the National Satellite Land Remote Sensing Data Archive was authorised [4].

In addition to the specifically remote sensing legislation, US government agencies also have responsibilities

within broader government policies, in particular for Earth observation the Paperwork Reduction Act of 1995[1] which codifies the information dissemination provisions contained in OMB (Office of Management and Budget) Circular A-130 [19].[2] Circular A-130 provides uniform US government-wide information resource management policies. It starts from the premise that government information is a valuable resource, and the free flow of information between government and the public is essential to a functioning democratic society, fosters excellence in scientific research and provides for the effective use of government funds. In order to maximise the usefulness of government information resources, systematic attention to the management of government records, in this case Earth observation data, is a necessary condition. OMB Circular A-130 contains a series of protocols to guide US government agencies: protocols relevant to Earth observation data archiving include planning and records management, providing information to the public, and dissemination management.

3.2. NOAA

NOAA is particularly interesting in relation to Earth observation data policy because it is an operational organisation. Its remit and its ethos are to provide a regular flow of environmental information to the public, government departments and the private sector. Therefore its orientation is to the long term. NOAA is responsible for the US weather satellites, and in the future for a wider range of operational Earth observation data supplies. Fig. 1 presents a summary of the growth in the NOAA data archive resulting from the major environmental observing systems, primarily Earth observation, in the period 2000–2014. The data volume at the beginning of the period is about one million Gbytes, rising to about 13 million Gbytes by the year 2014. Much of this growth is made up of new data sets, including NOAA's responsibility for archiving data from the US Earth Observing System (EOS), the civilian operation of the DMSP system, new polar orbiting satellite missions and in situ weather radar observations.

NOAA operates three National Data Centers, the National Oceanography Data Center (NODC), the National Geophysical Data Center (NGDC) and the National Climatic Data Center (NCDC). The NODC is the US repository and distribution facility for global ocean data. The NODC also operates the World Data Center-A, Oceanography. NGDC is the national repository for geophysical data, providing a wide range of science data, services and information. The NCDC is the world's largest active archive of weather data, and also operates

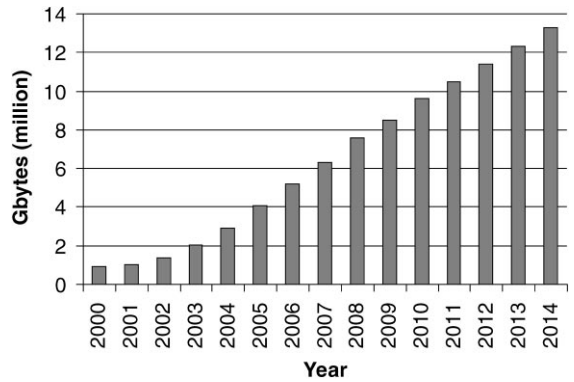

Fig. 1. The expected growth of NOAA's environmental data archives, 2000–2014. *Source*: Zevin [19].

the World Data Center-A, Meteorology. Together these three data centres comprise the NOAA National Data Centers (NNDC).

The NNDC's funding is part of NOAA's long-term base funding from the US Congress. Moreover it is specified by US law that some data, for example the US weather data, must be archived at NOAA's data centres. The data held at the NNDCs are thus guaranteed to be archived indefinitely. US government records cannot be destroyed without a formal process specified by law.

In line with US government policies, NOAA employs the following guidelines concerning the dissemination of its data sets.

- Full and open access to the data.
- No copyright on the data and no restrictions on distribution.
- Data provided at the cost of reproduction, except for US academia which may receive data free of charge or in exchange for other environmental data or products.
- Dissemination of data follows international guidelines, within the scope of US government regulations, e.g. WMO Resolution 40.

The exponential growth of data illustrated in Fig. 1 plus the full and open access position noted above does present challenges for archive data management. Table 3 summarises the stages involved in NOAA's environmental data management, from planning and collection through to storage and migration of the data onto new media. Table 3 gives illustrations of where NOAA is able to manage existing and new Earth observation (and related) data sets, and where it will need additional investment in infrastructure to carry out the data management functions. The key message from Table 3 is that while the early stages of data acquisition are resourced, the longer-term archiving issues are not.

[1] PL 104-13, section 3606(d).

[2] Web site http://www.whitehouse.gov/OMB/circulars/a130/a130.html.

Table 3
The status of NOAA environmental data management[a]

Data sets	End to end environmental data management functions							
	Planning	Collect or rescue	Ingest	Metadata and catalogue	Calibrate and validate	Store	Access	Migrate
Historical								
In situ observations	✓	✓	✓	✓	✓	✓	○	○
COOP/USHCN	✓	✓	✓	✓	✓	✓	○	○
GHCN	✓	✓	✓	✓	○	○	○	○
CARDS/COADS	✓	✓	✓	✓	○	○	○	○
Modernisation								
DMSP	✓	✓	✓	✓	○	○	○	×
POES	✓	✓	✓	○	○	○	×	×
ASOS	✓	○	○	○	○	○	×	×
NEXRAD	✓	○	○	○	×	×	×	×
GOES	✓	○	○	○	×	×	×	×
Future								
New in situ observations	✓	×	×	×	×	×	×	×
EOS[b]	○	×	×	×	×	×	×	×
NPP	○	×	×	×	×	×	×	×
NPOESS	○	×	×	×	×	×	×	×

Key: ✓, Can do with current resources; ○, Needs incremental resources; ×, Requires substantial additional resources.
[a] *Source*: Zevin [19].
[b] Archiving after the DAAC stage.

3.3. NASA

3.3.1. Earth Observing System

The Earth Observing System (EOS) is a programme which encompasses multiple spacecraft and interdisciplinary science investigations to provide a 15-year data set of key parameters and advances in scientific knowledge needed to understand global climate change. To fulfil this task, NASA is launching a series of EOS satellites which will take measurements important to Earth science research. Smaller, more narrowly focused satellite missions will provide additional information, for example the Vegetation Canopy Lidar (VCL). An unprecedented volume of data will be produced by these missions and will be stored, processed and distributed by NASA's EOS Data and Information System (EOSDIS) [11].

The EOS Data and Information System (EOSDIS) will process and archive most Earth Science Enterprise data, making it one of the largest civilian information systems. A major role of EOSDIS is to provide archive and distribution services for NASA's Earth science data and related data from other sources. EOSDIS data sources include current holdings of existing data centres, new products to be generated from current and future satellite missions, remote sensing data generated by international partner missions, data from the instruments on the EOS and international partner satellites, and data from ground-based studies by NASA and other organisations [20].

On a daily basis the EOSDIS Core System (ECS) must process over 480 Gbytes of raw data and produce 18,000 product instances with an aggregate size of 1600 Gbytes. ECS data archives will grow at the rate of two Tbytes of data per day and will become above a petabyte in size in the early years of the 21st century [21].

3.3.2. Archive centres

Currently EOSDIS consists of nine Distributed Active Archive Centers (DAACS) which were selected based on their existing institutional Earth science discipline and research expertise, infrastructure and commitment. DAACs are located at four NASA centres.

- Goddard Space Flight Center (GSFC).
- Jet Propulsion Laboratory (JPL).
- Langley Research Center (LaRC).
- Marshall Space Flight Center (MSFC).

DAACS are also located at five non-NASA data centres.

- Alaska SAR Facility (ASF) at the University of Fairbanks.
- US Geological Survey (USGS) EROS Data Center (EDC).
- National Snow and Ice Data Center (NSIDC) at the University of Colorado.
- Department of Energy (DOE) Oak Ridge National Laboratory (ORNL).
- Earth Science Information Network (CIESIN).

Fig. 2 shows the geographical distribution of the DAACs.

As components of EOSDIS, DAACS are institutions that generate EOS standard data products and carry out NASA's responsibilities for data archiving, distribution and management. Many sites also carry related data products [22]. This section of the paper gives information on the scale of the archiving responsibilities of three of the nine DAACs, namely the Goddard Space Flight Center, the EROS Data Center and the National Snow and Ice Data Center.

3.3.2.1. GSFC DAAC. The Goddard Space Flight Center (GSFC) DAAC is a repository of Earth science information on the global biosphere, atmospheric dynamics and the upper atmosphere, with links to other geophysical data sets archived at GSFC. Table 4 indicates the size of data holdings at the GSFC DAAC in 1996 and planned for 2001. The table shows the variety of environmental categories included, the wide variety of different data types and the different ages of Earth observation data in these archives.

3.3.2.2. EROS Data Center. The Earth Resources Observation System (EROS) Data Center (EDC) was established in the early 1970 s as a data management centre for the US Geological Survey (USGS). The original role of the EDC was to receive, process and distribute data from the Landsat satellites. In addition, the EDC stored, processed and distributed other types of satellite, aircraft and cartographic data. When the National Satellite Land Remote Sensing Data Archive was established [23], the role of the EDC was made more formal. Table 5 shows the data holdings in the EDC which contribute to EOSDIS. The focus is on general purpose missions with a land orientation. While substantially lower in data volume than the GSFC DAAC, the data quantities at the EDC still present a sizeable challenge for archiving and

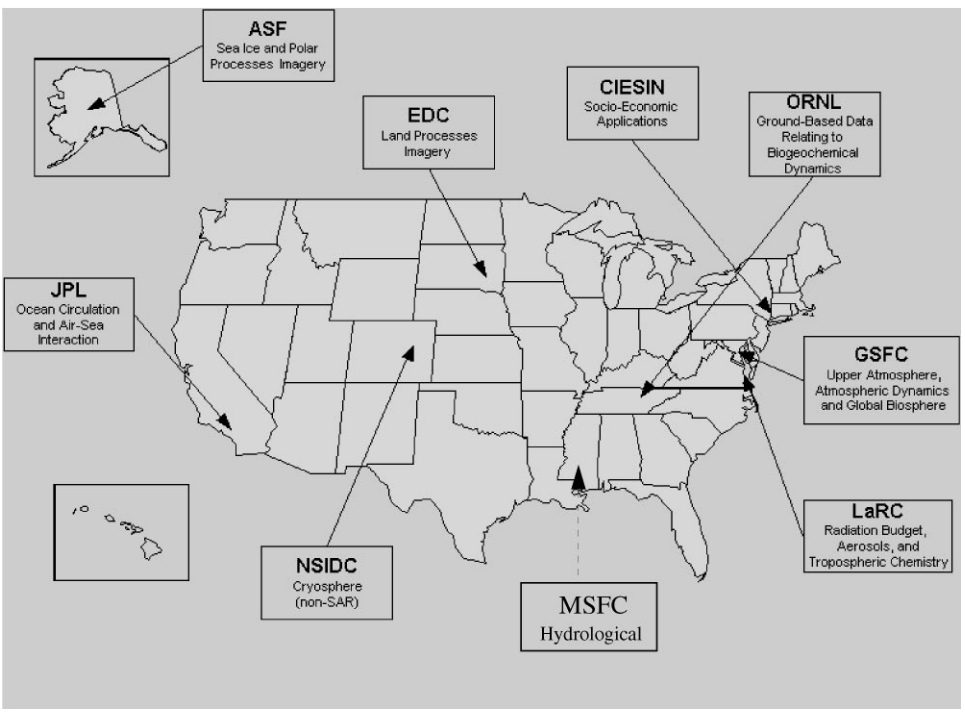

Fig. 2. The geographical distribution of the EOSDIS DAACs. *Source*: Schwaller [20].

Table 4
A summary of data archives at the GSFC DAAC

Data type	Data volume 1996 (Gbytes)	Data volume 2001 (Gbytes)	Sensors
Land biosphere	1384	306,500	AVHRR, MODIS
Hydrology	10	16,000	SSM/I, GOES, TRMM
Ocean biology	757	4300	CZCS, SeaWiFs, OCTS
Atmospheric dynamics	1250	4050	TOVS, TOGA-COARE, 4-D assimilated data
Upper atmosphere	102	440	UARS, TOMS

Table 5
A summary of data archives at the EROS Data Center DAAC that contribute to EOSDIS

Data type	Data volume 1996 (Gbytes)	Data volume 2001 (Gbytes)	Sensors
Global land products	3540	6720	AVHRR
Aircraft scanner	164	1754	
Landsat	538	1102	MSS, TM, supporting AVHRR
Radar data	800	1047	Seasat, SIR-B, SIR-C, SRTM, AirSAR
Digital chart and elevation data	1	800	

Table 6
A summary of data archives at the National Snow and Ice Data Center DAAC

Data type	Data volume 1996 (Gbytes)	Data volume 2001 (Gbytes)	Sensors
Polar products	140	402	AVHRR
Surface brightness temperature	80	137	SSM/I, ESMR, SMMR
Ocean and ice altimetry	15	39	Seasat, Geosat

for archive management, not least because of the age of some of the data sets (particularly Landsat, Seasat, AVHRR and SIR-B). In addition to civil, digital Earth observation data, the EDC also has more than 880,000 declassified intelligence satellite photographs.

3.3.2.3. National Snow and Ice Data Center. The National Snow and Ice Data Center (NSIDC) DAAC provides information on snow and ice processes. It aims to improve the understanding of the properties, characteristics and contexts of snow and ice, and their significance for human activity [24]. The challenge for data archiving and management, as illustrated in Table 6, are quite different for the NSIDC DAAC. The data sets are smaller than at GSFC or EDC, are more specialised and are generally focussed on a more specialised user community. The specialisation does raise questions about the duration of archive responsibility for these important data sets.

4. Data archives in Europe

4.1. European context

In 1997 a study for the European Commission investigated the archive status of 12 European Earth observation data archives [6]. The study found that the data were stored on a project or mission basis with no long-term policies for their maintenance, with the exception of EUMETSAT which does have a long-term data preservation policy. This section of the paper examines the Earth observation data archiving experience of the two European organisations with the largest archives, SPOT and ESA, and also looks at the experience of one European country's data archiving centre, the German Remote Sensing Data Centre.

4.2. SPOT

SPOT satellites have been in operation since 1986. By the end of 1999 SPOT had in its central catalogue a total of 7,483,285 scenes, of which 53% were stored at the two primary European receiving stations of Toulouse and Kiruna. The remaining 47% of scenes were stored at the 21 other receiving stations around the world [25]. The total data stored in all SPOT receiving stations is approximately 225,000 Gbytes, of which over 130,000 Gbytes are stored under SPOT's central control in Toulouse and Kiruna. Because SPOT has direct control only over the scenes recorded at Toulouse and Kiruna its active archive policy concerns these data sets.

A major challenge for SPOT is to maintain an accessible data archive in the face of obsolescent data storage media. During the period 1986 to 1995 SPOT data were stored on 25,000 High Density Digital Tapes (HDDTs). These tapes have a lifetime of 10 years, so SPOT sought an alternative storage medium in the form of optical tapes. This offered considerable storage advantages, firstly because one optical tape could store the equivalent data of 220 HDDTs, and secondly because the lifetime of the optical tapes is over 100 years. Unfortunately the CREO optical tape drive machines are no longer built, so the tapes cannot be read reliably. From 1996 to 1999 SPOT turned to D1 cassettes as storage media. Each D1 cassette holds the equivalent data of 10 HDDTs, but again they have a lifetime of only 10 years. So, in mid-1999 SPOT began to store all the SPOT data received at Toulouse and Kiruna onto a robotic archive system. This robotic system includes within it the ability to copy data

automatically onto new media as the media ages. This enables automated management of the SPOT data archive, at least for the data received at Toulouse and Kiruna.

SPOT is in the process during the year 2000 of transcribing its data archive onto the robotic storage system. Only the relatively cloud-free data are being retained. The criteria which determine the data to retain vary with geographical region. The list below gives the criteria. The list uses the segment as the basic amount of data to retain: the segment is a continuous strip of data imaged by one SPOT camera and has a length ranging from 60 to 500 km.

Polar regions: The segment must contain at least 30% of data with less than 25% cloud.

Equatorial regions: The segment must contain at least one quarter of a SPOT scene with less than 10% cloud.

Temperate regions: The segment must contain at least 20% of data with less than 10% cloud.

Semi-arid regions: The segment must contain at least 50% of data with less than 10% cloud.

Arid regions: The segment is not transcribed if it has more than 30% of the data with more than 25% cloud.

The effect of applying these criteria will be to save 40–50% of the SPOT data received at Toulouse and Kiruna. Already SPOT controls only about half of all the data recorded worldwide: the cloud filtering procedure will reduce this to approximately one quarter of all the SPOT data ever recorded.

4.3. ESA

ESA receives a wide array of Earth observation data. Its main responsibility is for the ERS satellites, and through Earthnet ESA also receives or has received data from Landsat, J-ERS, SPOT, NOAA, SeaWiFS and MODIS. SAR data from the ERS satellites is received at 34 stations around the world. Table 7 shows the size of the ERS SAR data archives held in 1996 and 2000 at ESA and non-ESA stations. The total of ERS SAR data received has now grown to about 350,000 Gybtes, of which ESA has direct control over about one quarter of the SAR data. Almost as much SAR data is received outside of the ESA participating states than is received by ESA.

For its historical data ESA is transcribing the data onto new media. Landsat TM data held on HDDTs are being transcribed onto D1 cassettes, and there are efforts to transcribe old (1978–1982) Landsat MSS data by cleaning and baking the tapes to recover the data which would otherwise be lost.

ESA operates the largest archive of Earth observation data in Europe, with over 400,000 Gbytes in total, but it has no clear mandate for the long-term preservation of data. The total amount of data held in the archive is growing exponentially and this growth rate is expected to continue. ESA plans to support the cost of maintaining its ERS data holdings for a period of 12 years after the end of ERS-2 operations. The Processing and Archiving Facilities (PAFs) will be responsible for maintaining these data during this period. The PAFs will be supported by ESA, although funding has only been allocated until 2003 [6]. ESA has allocated funding through the Envelope Programme to maintain an archive of Envisat data for a period of 5 years after the end of the ERS mission. After this point ESA anticipates that funding through the Envelope Programme will continue, at least for the short term, until other sources of funding can be found to ensure long-term preservation of the Envisat data holdings.

4.4. German Remote Sensing Data Centre

The German Remote Sensing Data Centre (DFD) is part of the German Aerospace Centre (DLR). DFD develops and operates ground segments for international Earth observation programmes, including data reception, processing, archiving and distribution activities [26]. DFD operates fully automated robot archives in Oberpfaffenhofen and Neustrelitz. The storage capacity can be expanded up to 1 Tbyte, and the technology will permit efficient conversion to future, more powerful storage systems as they become available by simply replacing existing storage media and drives [26].

Table 8 lists the Earth observation data holdings of the DLR in 1999. Except for the ERS SAR tandem data sets, the table does not contain reference to the DFD's holdings of single ERS data nor of Landsat data, which are

Table 7
SAR data acquisitions from the ERS-1 and ERS-2 missions[a]

Station type	Data volume mid-1996 (Gbytes)	Data volume mid-2000 (Gbytes)
ESA stations	70,215	97,475
National stations of ERS participating states	99,638	161,964
Foreign stations	67,876	89,708
Total	237,729	349,147

[a] *Source*: ESA DOSTAG papers, 1996 and 2000.

Table 8
A summary of data held in the robotic archive at the German Remote Sensing Data Centre, DLR-DFD, Oberpfaffenhofen[a]

Data type	Data volume 1999 (Gbytes)	Sensors
Radar	6,600	X-SAR, Seasat, ERS SAR tandem
Land surface	4,000	AVHRR
Atmosphere	1,100	GOME
Geostationary	400	Meteosat
Altimetry	100	ORB, ALT, PRARE

[a]*Source*: DFD [26] and Noack [3].

both kept in a manual archive. By the year 2005 it is estimated that the DFD's data archive will store between 80 and 300 Tbytes of data [27].

In contrast to the USA, Germany has no national data archiving policy, although data archiving is performed at the national level. The archiving activities of the DLR differ fundamentally from those identified in the US by the examples of NASA and NOAA since the data policy of the DLR is mission driven, with data being archived by agreement between the mission partners. The role of the DLR is not to archive data on a continuous basis from a series of satellites, but to archive data from certain experimental missions, such as X-SAR or for particular campaigns [6].

5. Discussion

Earth observation data volumes are significant now. With the introduction of new Earth observation systems the data archiving challenges will grow in complexity as well as in size. As NOAA has shown by Table 3, the concentration of effort in data archiving is largely on the early stages of the process such as data collection and archive ingest. It is the later stages that are particularly poorly developed and which need greater resources. The cost of maintaining Earth observation data in archives grows with age while, unlike a good wine, users expect the selling price of archived data to decrease with age.

The storage media and the archiving equipment in many organisations are becoming obsolete and unreliable. SPOT's solution is to use a dedicated, robotic archive which has its own capacity to copy data to new media as the storage media age. This should ensure that the data are technically accessible. At the same time SPOT acknowledges that its primary market is for data less than one year old. The sales profile for SPOT data reveals the following characteristics.

- 62% of requests are for data less than 1 year old.
- 82% of requests are for data less than 3 years old.
- 88% of requests are for data less than 5 years old.

For a commercial company such as SPOT Image, this raises the question of the desirability of maintaining an archive which costs money for little sales return. On the other hand, a pubic good argument would support the case for retaining all the SPOT data for future generations.

Earth observation systems are no longer governmental-only systems. This paper concentrates on government-sponsored systems and their archives because these are the largest Earth observation data archives and there is a clear public policy dimension to their activities. However, with the development of Space Imaging, Orbimage, EarthWatch and other very high spatial resolution systems, plus the growing interest in public–private partnerships in Earth observation, there will be more and more data that are only archived by the private sector. In their early days of operation the private sector systems will be collecting data to populate their own databases and so be able to offer a good range of data to their customers. Should private sector Earth observation data be considered a good in which governments have an interest in the sense that they have an interest in the state of the environment in the long term, or should the private sector be left to its own devices? If data archives are adopted by organisations that are not Earth observation mission managers then the scope to take on a wider range of Earth observation data will probably exist.

This paper has shown the different formal contexts of Earth observation data archiving approaches in the USA and Europe. The USA has a firm basis for archiving with laws on remote sensing, OMB Circular A-130 and the National Satellite Land Remote Sensing Data Archive. This has provided a sound basis for the development of national policy on Earth observation data archiving, and for individual actions such as the data purge alert by the USGS. By contrast the situation in Europe is weak, with archiving issues left to the Earth observation mission managers and to relevant national organisations such as the German Remote Sensing Data Centre. The piecemeal approach will allow valuable data sets to be lost in the future.

6. Conclusions

There is no doubting the importance of the long-term archiving of Earth observation data. Global change requires access to long-term data sets for research which in turn reinforces the need for data archiving. The management of archives therefore forms a vital part of global change research [31]. A problem here is that the judgement of the value of Earth observation data archives is only weakly developed. The value is commonly expressed either as the sales value of the Earth observation data or as the wider and very general contribution that Earth observation can (potentially) make to society. One of the

recommendations of a project on Earth Observation Data Policy and Europe (EOPOLE) is directly relevant to the question of the value of Earth observation data archives:

> The value of Earth observation data is not recognised fully because the value is largely judged in market terms rather than in other value terms. A greater recognition of the wider value of Earth observation data, including its public good value, is likely to help the sector significantly. The transition to sustainable Earth observation may not be achievable without a consideration of the humanitarian, scientific, security and environmental value of the data and information services provided. Therefore, there should be further research on how to capture explicitly the value of Earth observation data beyond the simple price that a user will pay [28].

In the past inadequate attention was paid to the maintenance of long-term data sets, often with low priority being given to long-term data management compared to that allocated to initial data collection. This situation is changing, particularly in the US where legislation has been put in place to enforce the preservation of data [29].

Some of the more pressing issues in data archiving relate to the length of time data should be stored and managed, the amount of data to be archived and which bodies are going to finance the activity. The issues of length of time and who pays are likely to remain unresolved because they are long-term issues and as yet adequate long-term policies are not in place. The finance issue is likely to become more contentious in the future as the inverse relationship between the exponential growth in data volumes coupled with ever tightening financial budgets widens. The financial issues concerning data archiving are perhaps the most difficult to resolve because of the long-term nature of the financial commitment required and the fact that the return on investments made will not be immediate, or necessarily measured in economic terms.

The current situation bears witness to a dichotomy between the US and Europe on the issue of data archiving. To ensure an appropriate data archiving infrastructure is put in place the US has developed legislation to ensure the preservation of important data sets. In Europe the situation is less clear, owing in part to the large number of actors involved and the increased difficulty in forming a consensus on the issue of data archiving.

Arguments for the need for long-term data archiving often reduce to moral obligations in the fact that we, the existing population, have a moral duty to provide future generations with the necessary means to manage their environment. There is no disagreement that this forms a strong argument, but it does not offer a solution to how these activities will be funded. Working towards legislation that supports the establishment of Earth observation archives would provide great benefits for future generations. Getting legislative bodies to realise that the value of an Earth observation archive increases with time will be the challenge for us all [30], but this is the path that appears to be the most promising to safeguard valuable environmental data sets.

Acknowledgements

The authors gratefully acknowledge the support from the Framework Programme IV of the European Commission for the EOPOLE project, and thank the EOPOLE team for their stimulating discussions. They also thank the University of London Intercollegiate Research Services (ULIRS) for providing facilities and the anonymous referee for valuable comments.

Appendix. List of acronyms

ALT	Altimeter on ERS
ASF	Alaska SAR Facility
ASOS	Automated Surface Observing System
AVHRR	Advanced Very High Resolution Radiometer
AWIICS	Archival Workshop on Ingest, Identification and Certification Standards
CARDS	Comprehensive Ocean-Atmosphere Data Set
CEOS	Committee on Earth Observation Satellites
CIESIN	Earth Science Information Network
COADS	Comprehensive Aerological Reference Data Set
COOP	Cooperative Observing Network
CZCS	Coastal Zone Colour Scanner
DAAC	Distributed Active Archive Center
DFD	German Remote Sensing Data Centre
DLR	German Aerospace Centre
DMSP	Defense Meteorological Satellite Program
DOE	US Department of Energy
EDC	EROS Data Center
Envisat	ESA Environmental Satellite
EOS	Earth Observing System
EOSDIS	Earth Observing System Data and Information System
EROS	Earth Resources Observation System
ERS	European Earth Resources Satellite
ESA	European Space Agency
ESE	Earth Science Enterprise
ESMR	Electrically Scanning Microwave Radiometer

EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
GHCN	Global Historical Climatology Network
GOES	Geostationary Operational Environmental Satellite
GOME	Global Ozone Mapping Experiment
GSFC	Goddard Space Flight Center
HDDT	High Density Digital Tape
J-ERS	Japanese Earth Remote Sensing Satellite
JPL	Jet Propulsion Laboratory
Landsat	US Land Satellite
LaRC	Langley Research Center
MODIS	Moderate Resolution Imaging Spectroradiometer
MSFC	Marshall Space Flight Center
MSS	Multispectral Scanner on Landsat
NASA	National Aeronautics and Space Administration
NCDC	National Climatic Data Center
NEXRAD	Next Radar of NOAA
NGDC	National Geophysical Data Center
NNDC	NOAA National Data Centers
NOAA	National Oceanic and Atmospheric Administration
NODC	National Oceanographic Data Center
NPOESS	National Polar-orbiting Operational Environmental Satellite System
NPP	NPOESS Preparatory Program
NSIDC	National Snow and Ice Data Center
OCTS	Ocean Color and Temperature Scanner
OMB	Office of Management and Budget
ORNL	Oak Ridge National Laboratory
PAF	Processing and Archiving Facility
POES	Polar-orbiting Operational Environmental Satellite
PRARE	Precise Range and Range Rate Equipment
SAR	Synthetic Aperture Radar
SeaWiFS	Sea-viewing Wide Field-of-view Sensor
SIR	Shuttle Imaging Radar
SMMR	Scanning Multichannel Microwave Radiometer
SPOT	Satellite Pour l'Observation de la Terre
SRTM	Shuttle Radar Topography Mission
SSM/I	Special Sensor Microwave/Imager
TM	Thematic Mapper on Landsat
TOGA	Tropical Ocean Global Atmosphere
TOGA-COARE	TOGA-Coupled Ocean Atmosphere Response Experiment
TOMS	Total Ozone Mapping Spectrometer
TOVS	TIROS Operational Vertical Sounder
TRMM	Tropical Rainfall Mapping Mission
UARS	Upper Atmosphere Research Satellite
UNESCO	United Nations Educational, Scientific and Cultural Organization
USGCRP	United States Global Change Research Program
USGS	United States Geological Survey
USHCN	US Historical Climatology Network
VCL	Vegetation Canopy Lidar
WDC	World Data Centre
WMO	World Meteorological Organisation
WWW	World Wide Web
X-SAR	X-band Synthetic Aperture Radar

References

[1] McKay D, Henderson D. Data management in Canada: issues for global change research, 1999. Available at www.usask.ca/library/gic/vln2/henderson/henderson.html.

[2] Tateishi R, Hastings D. Global environmental databases, International Society of Photogrammetry and Remote Sensing, Hong Kong, 2000. 233pp.

[3] Noack W. Experience with on-line catalogues and ordering systems at the German Remote Sensing Data Centre. Earth Observation Data Policy and Europe (EOPOLE) project, European Commission ENV4-CT97-0760, Workshop 3, July 1999. 6pp.

[4] Holm T. National Satellite Land Remote Sensing Data Archive A national asset, USGS EROS Data Center, 1999. Available at: http://www4.nas.edu/cger/besr.nsf/.

[5] UNESCO. Declaration on the responsibilities of the present generation towards future generations. Adopted on 12 November 1997 by the General Conference of UNESCO at its 29th session. Available at http://unesdoc.unesco.org/ulis/ged.html.

[6] Hubbard N, Wisman V. The long-term archiving of European Earth observation satellite data: current status and guiding principles, CEO/282/1998 issue 1.0. European Commission, Joint Research Centre Ispra, 1998. 28pp.

[7] Zenker S. Data policy aspects of archiving: technical background, the EDDA study and related topics. Earth Observation Data Policy and Europe (EOPOLE) Project, European Commission ENV4-CT97-0760, Workshop 1, January 1999, 1998. 7pp.

[8] ESA. Envisat data policy. Approved at the 66th Meeting of the ESA Earth Observation Programme Board, February 1998.

[9] CEOS. Satellite data exchange principles in support of global change research. Adopted at the Sixth CEOS Plenary Meeting, London, 9–10 December 1992.

[10] Haynes D et al. JISC/NPO studies on the preservation of electronic materials: responsibility for digital archiving and long term access to digital data, 1999. Available at www.ukoln.ac.uk/services/papers/bl/jisc-npo67/digital-preservation.html.

[11] NASA. NASA Earth Science Enterprise statement on data management, 1999. Available at: http://www.earth.nasa.gov/visions/data-policy.html.

[12] USGCRP. Policy statements on data management for global change research, 1999. Available at: http://www.gcrio.org/USGCRP/DataPolicy.html.

[13] NASA. Archival workshop on ingest, identification and certification standards (AWIICS), 1999. Available at: http://SSDO.GSFC.NASA.GOV/NOST/ISOAS/GWIICS/.

[14] Tahu G, Baker J, O'Connell K. Expanding global access to civilian and commercial remote sensing data: implications and policy issues. Space Policy 1998;14:179–88.

[15] Goede APH, Tanzi CP, Aben I, Burrows JP, Weber M, Perner D, Monks PS, Llewellyn-Jones D, Corlett GK, Arlander B, Platt U, Wagner T, Pfeilsticker K, Taalas P, Kelder H, Piters A. GODIVA, a European project for ozone and trace gas measurements from GOME. Advances in Space Research 2000;26(6): 951–4.

[16] Harris R. Earth observation data policy. Chichester: Wiley, 1997. 155pp.
[17] Costanza R, d'Arge R, de Groot R, Farber S, Grasso M, Hannon B, Limburg K, Naeem S, O'Neill RV, Paruelo J, Raskin RG, Sutton P, van den Belt M. The value of the world's ecosystem services and natural capital. Nature 1997;387:253–60.
[18] Williams J. Micro solutions to macro problems in the information society. In: From data to information, Proceedings Remote Sensing Society, Cardiff, p. 333–40.
[19] Zevin S. NOAA's data archives: data policy implications. Presentation to Earth Observation Data Policy and Europe (EOPOLE) Project, European Commission ENV4-CT97-0760, Workshop 5, January 2000.
[20] Schwaller M. Science Data Plan for the EOS Data and Information System. Document version 4, NASA, Washington DC, 1996.
[21] Clinard R. NASA's Earth Observing System Data and Information System (EOSDIS): an integrated system for processing, archiving and disseminating high-volume Earth science imagery and associated products, 1996. Available at: http://edhs1.gsfc.nasa.gov/waisdata.
[22] ORNL. ORNL DAAC Data Center Document, 1999. Available at: http://www-eosdis.ornl.gov/DATACENTER_DOCS/ornl_datacenter_document.html.
[23] USGS. About the EROS Data Center, 1999. Available at: http://edcwww.cr.usgs.gov/content_about.html.
[24] NSIDC. The snow & ice distributed active archive DAAC, 1999. Available at: http://www-nsidc.colorado.edu/NASA/GUIDE.
[25] Delclaux P. Data archiving and management in the SPOT system. Presentation to Earth Observation Data Policy and Europe (EOPOLE) Project, European Commission ENV4-CT97-0760, Workshop 5, January 2000.
[26] DFD. German Remote Sensing Data Center, 1999. Available at: http://www.dfd.dlr.de/dfd/brochure.html.en.
[27] DLR. The DFD data library, 1999. Available at: http://www.dfd.dlr.de/archive/index.html.en#arch.
[28] EOPOLE. Earth observation data policy and Europe. Final Report, ENV4-CT98-0760, European Commission, Brussels, 2000.
[29] Shaffer L. US data policy for Earth observations from space. In: Guyenne TD, Hunt JJ, editors. Space in the service of the changing Earth, vol. III, ESA SP-341. Noordwijk: ESTEC, 1992. p. 1477–81.
[30] Faundeen J. Digital data archives for Earth observation: an American perspective, 1999. Available at: http://www2.ssd.rl.ac.uk/public/eeos/abstracts.html.
[31] Harris R, Olby N. European perspectives on trends in public sector Earth observation data policies. Space Policy 2000;16:45–59.