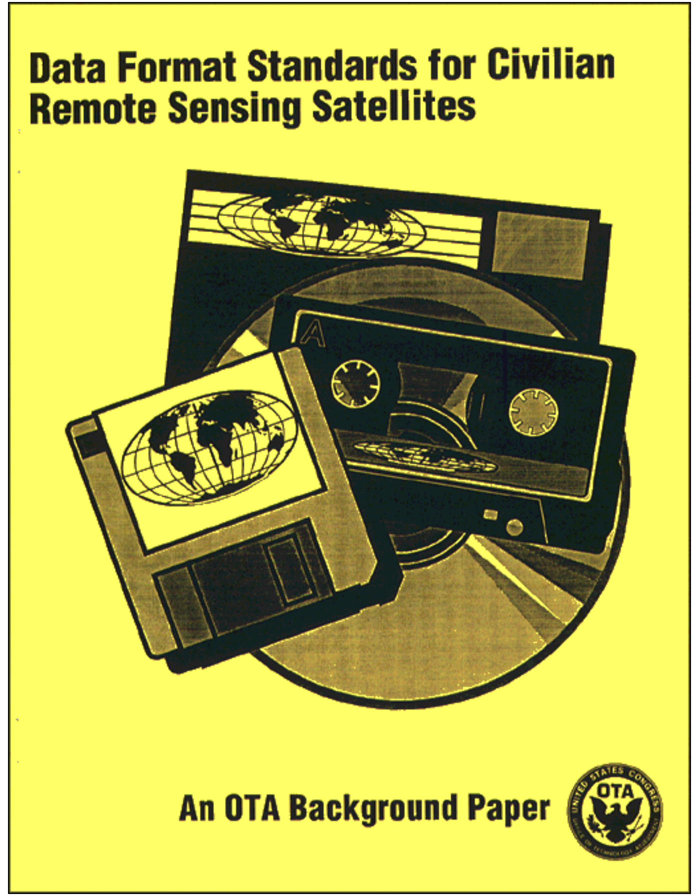
Data Format Standards for Civilian Remote Sensing Satellites

May 1993

OTA-BP-ISC-114
NTIS order #PB93-192290



The views expressed in this background paper are not necessarily those of the Board, OTA Advisory Council, on individual members thereof.

Earth Observation Systems Advisory Panel

Rodney Nichols, *Chair*
New York Academy of Sciences
New York, NY

James G. Anderson
Professor
Department of Chemistry
Harvard University
Cambridge, MA

D. James Baker
Director
Joint Oceanographic Institutions, Inc.
Washington, DC

William Brown
President
ERIM Corporation
Ann Arbor, MI

Ronald Brunner
Professor
Center for Public Policy Research
University of Colorado
Boulder, CO

Joanne Gabrynowicz
Associate Professor
Department of Space Studies
University of North Dakota
Grand Forks, ND

Alexander F.H. Goetz
Director
Center for the Study of Earth from Space
University of Colorado
Boulder, CO

David Goodenough
Chief Research Scientist
Pacific Forestry Center
Forestry Canada
Victoria, BC

Donald Latham
Corporate Director
Loral Corporation
Reston, VA

Cecil E. Leith
Physicist
Lawrence Livermore National Laboratory
Livermore, CA

John H. McElroy
Dean of Engineering
The University of Texas at Arlington
Arlington, TX

Molly McCauley
Fellow
Resources for the Future
Washington, DC

Earl Merritt
Vice President, Research
Earthsat Corporation
Rockville, MD

Alan Miller
Director
The Center for Global Change
University of Maryland
College Park, MD

Raymond E. Miller
Professor
Department of Computer Science
University of Maryland
College Park, MD

Kenneth Pederson
Research Professor of International Affairs
School of Foreign Service
Georgetown University
Washington, DC

David T. Sandwell
Geological Resources Division
Scripps Institute of Oceanography
La Jolla, CA

Dorm Walklet
President
TerraNOVA International
Los Altos, CA

Albert Wheelon
Consultant
Los Angeles, CA

NOTE: OTA appreciates and is grateful for the valuable assistance and thoughtful critiques provided by the advisory panel members. The panel does not, however, necessarily approve, disapprove, or endorse this background paper. OTA assumes full responsibility for the background paper and the accuracy of its contents.

OTA Project Staff—Data Format Standards for Civilian Remote Sensing Satellites

Lionel S. Johns, *Assistant Director, OTA*
Energy, Materials, and International Security Division
(through February 1993)

Peter Blair, *Acting Assistant Director, OTA*
Energy, Materials, and International Security Division
(from February 1993)

Alan Shaw, *International Security and Commerce Program Manager*

Ray A. Williamson, *Project Director*

Brian McCue

Contractor

Madeline Gross

Administrative Staff

Jacqueline R. Boykin

Louise Staley

Workshop-Formats, Media and Standards for Civilian Remotely Sensed Data October 2, 1992

Donald Artis
Director
CETEC-ZC-B
U.S. Army Topographic
Engineering Center

Jill Butler, Major, USA
Central Imagery Office

Kenneth Daugherty
Chief Scientist
Defense Mapping Agency

Jim Frey
Program Manager
NPIC
National Exploitation Laboratory

D. Brian Gordon
Chairman
Tactical & Military Multispectral
Requirements Working Group
Defense Intelligence Agency

William Kennedy
Hughes STX

Lon Naylor
NPIC
Exploitation Support Group

NOTE: OTA appreciates and is grateful for the valuable assistance and thoughtful critiques provided by the workshop participants. The panel does not, however, necessarily approve, disapprove, or endorse this background paper. OTA assumes full responsibility for the background paper and the accuracy of its contents.

Data Format Standards for Civilian Remote Sensing Satellites

Introduction

Earth data—positional, topographic, climatological, meteorological, man–made features, and changes over time in all of these—are increasingly important to the military. They form the heart of navigation, intelligence, combat information, situational awareness, weapon guidance, damage assessment, and training systems. Analysis indicates that great advantages may accrue from being able to integrate these functions. However, being able to do that integration in a routine and efficient way will require either generating databases in one format (or a very few compatible formats) or developing equipment to convert quickly and routinely among various formats. Today, neither of these conditions exists. Data exist (and are gathered) in a variety of diverse formats. This is the natural consequence of a situation in which–in the absence of a central plan for integrating collectors, processors, and users–format commonality has not been a major design factor. The format for any one system is usually chosen to meet the specific needs of that system and is driven by the technology available at the time it is created. Often the factors that dictate format are compelling (e.g., continuity of data provided to long-standing clients, or optimization of satellite power budget). In many cases, altering format would necessitate costly, extensive changes in hardware and software systems. Finally, setting standards is difficult when technology moves quickly, and is particularly hampered by the notoriously cumbersome MILSPEC process.

The Office of Technology Assessment (OTA) was asked by the Senate Armed Services Committee to investigate the plethora of formats for remotely sensed Earth data. At a workshop held on October 2, 1992, participants discussed the pros and cons of standardizing formats for remote-sensed data, the question of how many different standards are needed, with what else should remote-sensed data be compatible, and who should set the standards. The discussion closed with the participants’ views as to what actions could be taken by Congress. Although the attendees (see also the attached list) were from the national security community, the discussions concerned primarily the use of data from civilian Earth remote sensing satellite systems: that is systems such as Landsat that were built and operated primarily to provide data to the scientific and civil commercial sectors. Data from these systems are bought and extensively used by the military and intelligence communities. The need to integrate data from military-unique systems as well only complicates the situation.

It was clear from the workshop that data from difference sources are largely provided in different formats and on different media. Moreover, formats, media, and recording hardware for the same system have often changed over time, leading to situations where old data is virtually inaccessible. Creating order out of the current chaos will be a very difficult proposition, and is likely to cause problems for many suppliers and users of data. Nevertheless, the payoff for doing so may be sufficiently great to warrant a major effort in that direction.

Levels of Compatibility

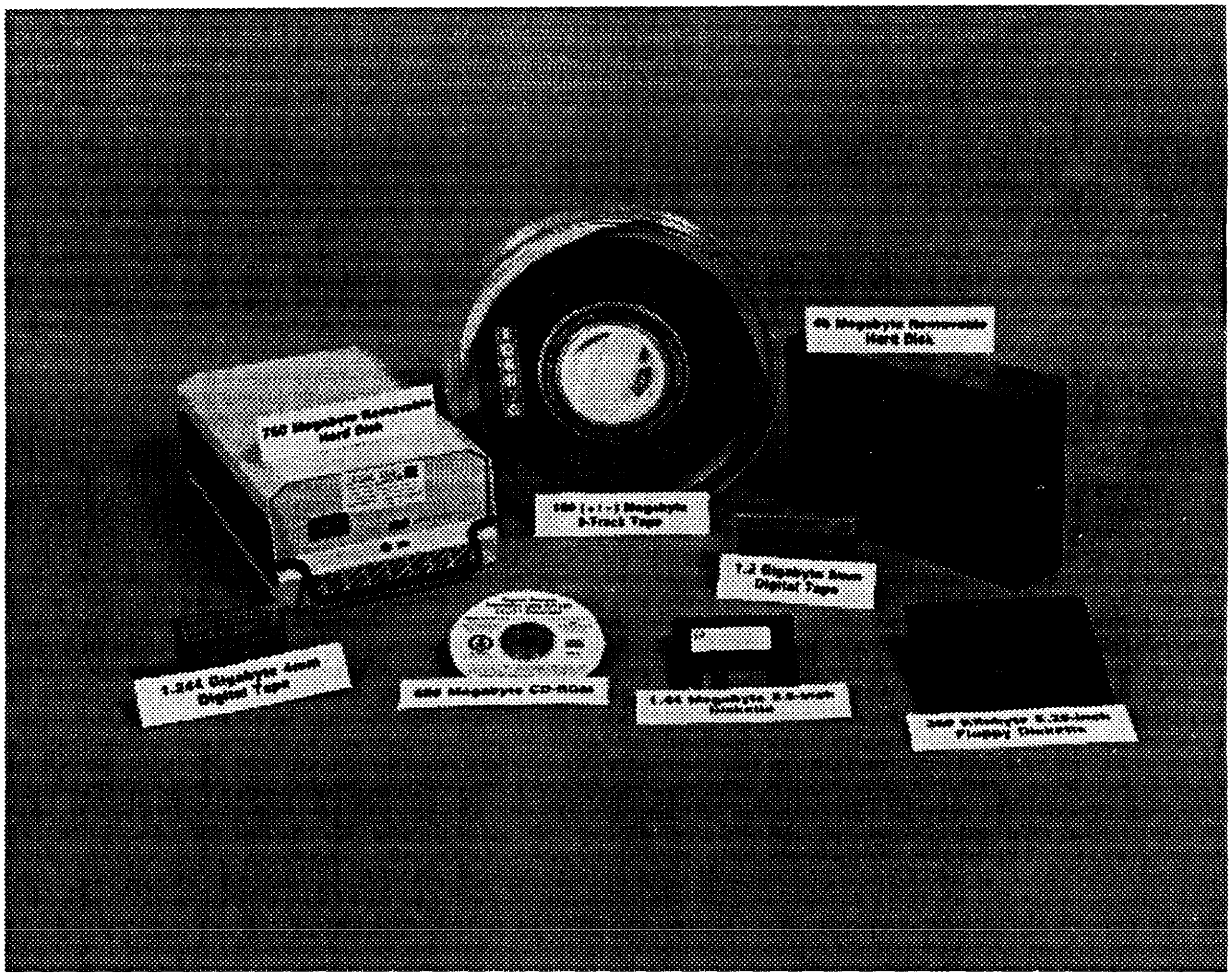
Most civilian remote sensing satellite systems, such as the U.S. Landsat and the French SPOT, use digital image transmission to send their data to Earth. Customers, while they may receive a photographic image for reference’s sake, obtain the data on computer tapes. No standard format for these tapes exists—each satellite system uses a unique format for its data, and this multiplicity of formats can lead to difficulty for users. Indeed, standards do not even prevail within a given system: the different Landsat ground stations distribute their data in slightly different formats. Moreover, every compression scheme[1] is, in fact, a format.

---

1 Data "compression" schemes shorten messages by capitalizing on their redundancies. Imaged scenes, for example, tend to contain sizable patches of identical shade or color. A transmission scheme that specifies the

Participants identified three levels of compatibility—hardware, software, and data. To take examples from the now-familiar world of personal computers, two users of IBM-compatible machines have hardware compatibility; if they both use Lotus 1-2-3 they have software compatibility; and if they each set up their Lotus spreadsheets in the same way they have data compatibility. Alternatively, they can have hardware compatibility without both owning IBM machines if both machines are “IBM-compatibles”; they can have software compatibility without using Lotus if they agree to use Lotus-compatible spreadsheet packages; but the only way to have data compatibility is to stick to the same data format. Also, and perhaps somewhat surprisingly, software compatibility can be had without hardware compatibility if the correct communications protocol is used. These ideas carry over into the rest of the computer world, including the collection and use of remotely sensed Earth data. As will be discussed below, problems resulting from incompatibilities can be resolved through the use of appropriate conversion software. The routine use of conversion software is an alternative to compatibility, not a form of it.

color of each picture element, saying “blue, blue, blue, red, green, green, green,” will be less efficient than one that uses compression to say, “2 blues, 1 red, 3 greens. ”



A common form of hardware incompatibility is that of the media used to store data. The picture shows a frustrated user's exhibit of the many media on which he received remotely sensed Earth data during Operation Desert Storm.

Source: United States Army Topographic Engineering Center.

There is no shortage of horror stories regarding data and data formats. In the Persian Gulf War, for example, Apache helicopters had the terrain of the United States hardwired[2] into their navigational systems and thus could not be loaded with Persian Gulf data. Similarly, British Tornado fighter-bombers were hardwired for Europe. The JSTARS aircraft had data for the right

2 That is to say, built in rather than written in.

place, but for the wrong time: they were loaded with data of 1985-1987 vintage and lacked any provision for updating the data.

The Present Situation

To a large degree, data formats (especially the formats of "raw" data) are determined by the sensors that originate the data. Different sensors, e.g., electro-optical cameras and SARs, would be hard-pressed to transmit data in the same format.[3] Thus the data are "born" in incompatible formats, and any compatibility must be forced on them later.

Use by some vendors of proprietary data formats, for which they retain the source code,[4] is another source of incompatibility. While these formats are legitimately proprietary in that they relate to the vendors' trade secrets about how their products work, the use of proprietary formats weds the user to the vendor: the vendor must enter any new or updated data that arise because the user cannot do so on his own. Such practices can actually result in the de-standardization of data, such as the DMA maps scanned into the video memories of aircraft map-display systems. U.S. aircraft went to the Persian Gulf with three sets of such formerly-identical data, rendered into different formats.

DoD contracting regulations stymie attempts to arrive at standardized data formats and technologies. Data are acquired at the command[5] level, and the commands thus write independent

3 It is intended that EOSDIS data formats be sensor-independent, but workshop participants felt that complete sensor-independence would not be attained.

4 "Source code" is the computer program as written by a programmer. It could be easily read by another programmer. To be read and run by a computer it must be "compiled," a nearly irreversible predigestion step that renders the program palatable to the computer but unfit for human consumption.

5 The military's 10 "unified and specified commands" are the highest levels of geographic and functional aggregation present in the U.S. Armed Forces.

contracts and wind up with up-to-date, incompatible systems bought for a good price. They have no formal means by which to coordinate a standard with each other, and scant leverage to enforce an informally coordinated standard on vendors should they arrive at one.

As noted above, the different Landsat ground stations distribute their data in slightly different formats, perhaps because they were built by different contractors. The format of Landsat data has changed over time, and even the passage of a mere 3-5 years has been enough to render some data inaccessible. Multispectral (MSS) data originally collected by Landsats 1-3 are in a different format from those collected by Landsats 4 and 5. There are no fewer than 5 different Tm[6] formats In some cases, the format is standard while the meaning of the data is not: brightness, for example, may be expressed in the same numbers and recorded in the same format but the same number may, in different data sets, refer to different absolute levels.

France's SPOT system is more uniform, perhaps only because it is newer. There are two SPOT formats now, and a new one forthcoming.

The Landsat-SPOT translation is the classic case of incompatibility in the civilian Earth remote sensing community today. In one sense, it is quite difficult because the formats were developed completely independently of each other and thus have nothing in common. In another sense, it is quite easy because it arises so often that computer programmers have come to grips with it and produced software packages that are easy to use. In preparing to conduct the workshop, OTA staff questioned several experts on important questions of compatibility and were puzzled that the question "How hard is it to merge Landsat and SPOT data?" elicited responses ranging from "That's very difficult indeed" to "We do it on a daily basis. " Colloquy at the

6 Thematic Mapper.

workshop revealed the resolution of this paradox: the software to perform this merging is difficult to write, but once it has been written it is easy to use.

The newly created Central Imagery Office is seen by some as the Government's vehicle of deliverance from the current surfeit of non-standard data formats, and indeed the CIO is at work on a National Imagery Transmission Format Standard.[7] However, a participant expressed the sentiment that until CIO has "a budget and a building" its pronouncements will carry little weight.

Pros and Cons of Standardization

Some may feel that the surfeit of different formats—and even media–creates an open-and-shut case for standardization. Others, seeing standardization as analogous to U.S. adoption of the metric system—compellingly efficient in principle, horrendously difficult in practice-say, "We can't change everything now. " Finally, some may say that standardization is a fine idea, but really means "Let's standardize on my format. "

In the absence of standards, each organization develops a format that best fits its own needs. This format will, of course, differ from that of any other organization. Each user discounts the value of standardization, if only because it will not pay off until some vague time in the future. Eventually, the need will arise for one organization to use another's data. Then the call for compatibility will be heard.

Making two or more organizations format their data in the same way has obvious advantages, and resistance may seem to be rooted purely in inertia It is important to keep in

7 This standard actually predates CIO, and has been in the works for 8 years.

mind, however, that compliance with the standard has a greater cost than "just" the rewriting of existing software: it forces the organization out of a format custom-tailored to its own needs and into a one-size-fits-all format representing a major compromise among a number of users. It can also act as a brake on innovation, and can cause the community to miss out on new technologies or even new sources of data. For this reason, no standards can be permanent—it will at some point become outdated and need replacement, triggering a new debate.

Thus careful thought should precede any shift to a standard format. Standardization just isn't appropriate for some users. For example, the downlink from the satellite itself is only received by a few stations. The format of the downlinked data is a reflection of the sensor hardware aboard the satellite and is carefully designed to wring the greatest possible capability out of the scant power, bandwidth, and transmission time available to the satellite.

EOSAT'S standardization of data formats was viewed as "a questionable model" by workshop participants, who held that the Government committed errors in this connection. EOSDIS,[8] not only because of its intended sensor-independence but also because of its sheer magnitude, may become a de facto standard in the near future. If it does so, it may even begin to mold the shape of future data, ruling out any data that cannot be expressed in an EOSDIS format.

Conversion

An option open to the user who wants to retain his own data format is to do so, converting data to or from the officially sanctioned standard format whenever he communicates with the outside world. The sufficiently sophisticated user could thus show a standardized face to

8 The Earth Observation Satellite Data and Information System.

the world while internally operating in whatever idiosyncratic way was needed to get the job done. At present, instances of compatibility are the exception, and conversion software is widely used. Some organizations even retain programmers whose primary job is to write custom software to bridge compatibility gaps as they arise. Such an organization is likely to say, “We don’t have much of a problem with non-standardization of remotely sensed Earth data,” because they have the problem well in hand. However, such a statement ignores the payroll cost of retaining the programmers and the opportunity cost incurred through allocation of the programmers’ efforts to the solution of avoidable problems.

Another option open to the user who likes his own format is to use it both internally and with the outside world, and let others bear the cost of conforming to it. In this practice lies the genesis of today’s profusion of data formats. Its attraction is that it allows the user to create a format that is optimal for him: the costs of conversion are laid at the other users’ feet or, in economists’ terms, externalized. As a frequent bearer of this externalized cost, the Government could, for the purposes of letting contracts, figure its own conversion costs into the true cost of each bidder’s proposed system. An even greater cost would be born if lives were lost as the result of an inability to transmit or display data properly.

With What Else Should Remotely Sensed Earth Data Be Compatible?

Remotely sensed Earth data should be formatted with a view toward compatibility with each other and also with image processing tools, the terrain data in models and simulations used in weapons development and crew training (such as SIMNET), and the object data (representations of tanks, aircraft, etc.) in these models and simulations. They should also be compatible with a variety of data not obtained by remote sensing, such as cartographic and geodetic data obtained by in situ measurement, and even with such non-image data as 150 years' worth of in situ weather data with which EOS data will need to be compared by the global warming community.

Time-Domain Compatibility

One workshop member, in answer to the question posed in the heading of the preceding section, said simply, "the future. " Consideration needs to be given to compatibility with future systems as well as the spectrum of present-day ones. The anecdote (recounted above) about the lack of any provision to update the data in JSTARS is not unique. The legendary rate of technical progress in the computer industry will quickly make any standard into a limitation: familiar rules of thumb hold that hardware doubles in capability every 9 months, and software lags by 5 years. Data formats of necessity lag software, and standards will lag the formats, so almost any standard will greatly confine future developers. At present, for example, our telephone standards are shaped by the requirements of voice transmission. Yet a large and growing fraction of telephone traffic (30°/0, expected to be 70°/0 by the year 2000) is data, not voice.

Consideration also must be given to compatibility with the systems of the past. The software in image processing work stations, for example, ought to be able to deal with data from archives as well as with data from a satellite currently in orbit. NASA's archives of satellite pictures of other planets provide a cautionary example in this regard: much of the "take" has never been looked at and is in formats not accessible to any present-day system and not sufficiently well documented that any system can readily be configured to read them. It is said that Government cryptologists have been brought in to decipher the data, as if it were some clandestine transmission,

How Many Different Standards Are Needed?

Would it really be feasible to have a single standard? Should there be a standard for producers and another for consumers, with the transformation accomplished by some single entity? Perhaps, as in so many things, the separate services will assert that each of them, because of its unique requirements, needs to have its own standard. More substantively, the final product will not be a single standard, but a family of standards. Some will be for archiving, others for transfer, and so on. Different standards will apply to data of different qualities—one standard for extrapolation-quality data, another for presentation-quality data, and so on again. Presentation-quality data could logically be in vector format for some applications and in raster-scan format for others, so more than one standard is necessary. Standards for transfer will vary according to the transmission medium and the time-vs.-bandwidth tradeoff it presents-there could be one standard for transmission by fiber-optic link and another for transmission by 2400 baud telephone line.

With what else should remotely sensed data be compatible? These data are not used in a vacuum–should their format be compatible with other data, such as terrestrially acquired Earth data, other than Earth data, or non-image data used by the same users? Who are the users we care about–mapmakers, modelers, simulators, guidance-system designers . . . ? In the end, what is needed is not only a family of standards for data formats, but a standard for interoperability between nonstandard formats and a well-thought-out program of standard maintenance.

At the very minimum, all formats could be required to use a standard header that would give some facts about the origin of the data but which would mainly contain an indicator of what format (standardized or otherwise) the coming data were in.

Finally, there will always be some users who demand the raw bit stream exactly as it came out of the satellite or sensor. For them, there can be no standardization (other than that stemming from possible standardization of sensors) because, as mentioned above, the format of data so close to the sensor is determined by the characteristics of the sensor itself.

Who Should Set The Standards?

A variety of organizations are potential setters of imagery standards: CIO, NASA NIST (The National Institute of Standards, formerly the National Bureau of Standards), DoD, and the international standard-setting bodies, to name a few. The CIO could help considerably in this regard. Industry will favor standards and will eventually create them if Government does not, but will scrutinize proposed Government standards carefully for advantages to particular vendors. On a worldwide basis, the United States can hope and expect to exert strong influence over the rest of the world, not only through market leadership but through good example. Internally, a particular company may sooner or later come to dominate the market and thus verge on becoming a de facto standard, but one should recognize that the success of the company in the market place may be despite their data format instead of because of it, and therefore one should avoid jumping to the conclusion that the leading data format (if any) is the best.

Those new to the standards business are sometimes surprised that there needs to be a specification for standards—on the face of it, this seems to be another round of wasted Governmental motion. Yet such an effort is needed, to ensure that the standards address the same issues, use the same vocabulary, and will be understood in the same ways. Today, for example, confusion sometimes arises because the term “null character” means “ “ to some, “O” to some others, and" "to others still.

Issues for Congress

Is this standard to be a law? Should its formulator have regulatory authority? Where in the budget will standardization be paid for? Where in the budget will the savings become evident? Ought the conversion cost inflicted on the Government by the production of non-standard data be legislatively decreed as a cost of the contract, to be used in choosing vendor? Ought Congress to ban the writing of contracts in which the Government acquires data written in a proprietary format?