PFG 2015 / 6, 0449–0456
Stuttgart, December 2015

Article

Development of an ISO-Standard for the Preservation of Geospatial Data and Metadata: ISO 19165

Wolfgang Kresse, Koszalin, Poland & Joan Masó Pau, Barcelona, Spain

Keywords: long-term preservation, archiving, geospatial data, ISO, OGC, metadata

Summary: Most of the paper maps produced a century ago are still very accessible in cartographic libraries preserved by the producer. It is our present obligation to guarantee the preservation of digital geospatial data today and allow for digital cartographic accessibility one century into the future. In addition, there is an increasing demand for older maps that goes beyond pure historical interest motivated by the study of dynamic problems such as impacts of the climate change, human activities and sustainability. The long-term preservation of large volumes of geospatial data in a uniform way still remains an unsolved question. A systematic solution has been demanded by National Mapping and Archival Agencies in Europe and North America. One year ago the ISO/TC 211 "Geographic information / Geomatics" published a New Work Item Proposal (NWIP) named ISO 19165 "Preservation of digital data and metadata" accompanied by a Working Draft document. The proposed standard is built upon the principles laid down in the ISO 14721 "Open Archival Information Systems" and upon the data model of the ISO 19115-1 "Metadata – Part 1: Fundamentals". This article reports on the specialization of both standards for the purpose of archiving of geospatial data and asks for contributions to the ISO 19165 under development.

Zusammenfassung: *Entwicklung einer ISO-Norm für die Archivierung von Geodaten: ISO 19165.* Papierkarten aus dem frühen 20. Jahrhundert sind meist heute noch gut zu gebrauchen. Inwieweit man digitale Geodaten von heute in 100 Jahren nutzen kann, bleibt abzuwarten. Zudem ist ein gestiegener Bedarf an historischen Karten zu beobachten. Die Langzeitarchivierung von umfangreichen Geodaten in einheitlicher Form bleibt vorerst eine ungelöste Aufgabe. Eine systematische Lösung wurde von nationalen Vermessungsbehörden und Archiven in Europa und Nordamerika gefordert. Vor einem Jahr veröffentlichte daher das ISO/TC 211 „Geographic information / Geomatics" ein New Work Item Proposal (NWIP) unter dem Namen ISO 19165 „Preservation of digital data and metadata" verbunden mit einem bereits fertiggestellten Working Draft Dokument. Die neue Norm baut auf den in der ISO 17421 „Open Archival Information System" dargelegten Prinzipien und dem Datenmodell der ISO 19115-1 „Metadata – Part 1: Fundamentals" auf. Dieser Beitrag berichtet von der Spezialisierung der beiden ISO-Normen für Zwecke der Archivierung von Geodaten und ruft zur Beteiligung bei der weiteren Entwicklung der ISO 19165 auf.

1 Introduction

Looking back a century, or even before, we find a lot of geographic information created at that time that is still accessible and usable in the form of paper maps. Paper has proved itself a good medium for long term preservation if it is looked after with appropriate care. It is easy to handle and printed information in graphical form is immediately accessible. However, in today's digital world, geographic information is produced and distributed predominantly in digital formats that have a very strong reliance on technology to both store and access data. If we look one century ahead it is quite difficult to see how the preservation of digital data will be assured and how the information we generate today will be accessed (Rönsdorf et al. 2013).

In addition, there is increasing demand for older and superseded data to support historical and temporal analyses related to change in Earth's natural and human landscape (Library of Congress 2010), to know the im-

www.schweizerbart.de
DOI: 10.1127/pfg/2015/0000
1432-8364/15/0000 $ 2.00

pact of human activities and eventually to extract lessons for a sustainable development that does not go irreversibly beyond planetary boundaries (Steffen et al. 2015). The exponential growth of studies done as a result of the free availability of the complete Landsat series is another example of the monetary value of well preserved Earth observation archives (Wulder et al. 2012). Consequently, national mapping agencies and Earth observation data producers (including remote sensing space agencies and in-situ campaigns) are challenged to preserve the history of geospatial data before new updates take place.

After some years of a short history of geospatial digital production, we are now realising that long-term preservation of large digital geospatial datasets in a uniform way is a still unsolved question. Though the problem has been known for a while it has been mitigated by producer's comprehensive storage policies and, so far has been compensated by the fast development of storage media towards cheap and voluminous units. However, currently the growth of data is faster than the growth of storage media, particularly due to constant inputs to remote sensing datasets.

Remote sensing agencies have realized this issue and they already had applied data preservation policies and procedures to their data archives. The European Space Agency (ESA) Preserved Data Set Content (LTDP 2012) and the NASA Earth Science Data Preservation Content Specification (NESDPCS 2012) are two examples of success stories in the remote sensing community. A systematic solution has been demanded also by National Mapping and Archival Agencies in Europe and North America. In 2003 the International Organization for Standardization (ISO) published a first preservation standard developed by a group of agencies running imaging space-borne sensors: the ISO 14721 "Open Archival Information System" (OAIS). Though the document was developed by space agencies, the specification is generic enough and has been applied to digital libraries around the world.

Solutions commonly applied in libraries to preserve printed documents (including OAIS) cannot be directly applied because geospatial data possesses a number of peculiarities that do not agree with or even contradict common archival principles. Some of the data are never final like cadastral data and thus in principle could never be archived. Other data has legal or ownership restrictions which keep their validity and must be acknowledged by an archive. Topologically structured data features are fully interlinked and thus cannot be separated into packages without modifying their structure. Geospatial Information Systems design principles often force the data models to be divided into thematic layers and each one considering geospatial data, metadata, symbolization, printing, etc. that are often stored individually. In addition, geospatial data are often linked with non-geospatial information increasing the storage capacities even more. Service-Oriented Architectures (SOA) produce maps on-the-fly sometimes related with highly dynamic real time phenomena which are gone shortly after the call, thus, hardly impossible to be preserved. Though some of the cases may rarely require archival, in others can open new forms of studying human evolutions. For example, the CartoBD company is storing all georeferenced twits believing that they will allow for studies of human activities in the future. These emerging cases illustrate the differences and potentialities that archival of digital geospatial data can provide. The authors assume that a preservation case for every example could be perceived. Finally, large data volumes of the imagery domain tend to exhaust the available long-term storage capacities.

The ISO/TC 211 "Geographic information / Geomatics" is ISO's Technical Committee that pools all standardization projects for geographic information. This is the reason why the national mapping agencies of Europe, the Open Geospatial Consortium and others have asked ISO/TC 211 to take the action. Though several other topics have been standardized already, an ISO-standard that addresses the specificities of the preservation of geospatial information is still missing. Many works of administrative units require ISO-standard as a fundament while non-ISO-standard are not accepted e.g. for calls for tenders, in particular in the U.S..

The development of the ISO 19165 "Preservation of geospatial data and metadata" has started a year ago. In June 2015, the project

team discussed the first Working Draft which is built upon the ISO 14721 and a proposal of the Open Geospatial Consortium. This proposal was prepared by the Catalan Cartographic and Geologic Institute aiming at the definition of a minimum set of metadata for preservation by extending the existing ISO 19115-1 "Metadata – Part 1: Fundamentals".

The intention of this article is disseminate the work that the ISO/TC 211 is doing to cover the need of a geospatial preservation standard toward a future official ISO 19165.

2 Preliminary Works

The most detailed standard that addresses preservation in a generic way is the ISO 14721 "Open Archival Information System". This ISO standard has been initiated by the NASA and developed in cooperation with the ESA and other space agencies (ISO 14721:2012). A fundamental element of this ISO-standard is the Information Package (IP) which contains the data (content information) together with their metadata (preservation description information). This concept has been adopted by the new ISO 19165. The definition of new metadata elements will be restricted to those not already defined in the ISO 19115-1.

The European National Mapping and Cadastral Agencies published 16 principles for the archival of geospatial data under the title GI+100 (Rönsdorf et al. 2013). Four important principles sound:

- Archiving of digital Geographic Information begins at the point of data creation, rather than at the point of withdrawal from active systems (1).
- Be selective and decide what to archive and what to lose (3).
- Consider preservation timeframes of 1, 10, 100 years (4).
- Geographical data should be preserved in a way that non geo-specialists can handle (8).

In a similar way the U.S. Library of Congress (2010) demands appraisal and selection of geospatial data because of the limited resources that most organizations have for preservation.

An international group of library experts published under the leadership of the U.S. a data dictionary for preservation of metadata, the PREMIS data model (PREMIS 2012). Ramapriyan & Moses (2012) prepared the data preservation content specification of NASA's Earth sciences division. The U.S.-states North Carolina, Kentucky, Montana, Utah, started together with the Library of Congress GeoMAPP in 2010. The GeoMAPP effort aimed to address the preservation of "at risk" digital geospatial content such as land parcels, zoning, roads, and jurisdictional boundaries which change regularly. Existing copies of these data are often at risk of being overwritten when updates or changes are made (GeoMAPP 2010). Higgins (2008) presented a Curation Lifecycle Model developed at the Digital Curation Centre of the University of Edinburgh, United Kingdom. Engel et al. (2013) of the land survey administration of Baden-Württemberg, Germany, discuss the archival with a focus on appropriate data formats. A methodology for the preparation of documentation evaluation and access proposals (PAAD) is prepared by the Catalonian National Commission on Access, Evaluation and Selection of Documents (CNAATD 2015).

A broad discussion on aspects of the preservation of digital cartography can be found in Jobst & Gartner (2011). Moe & Longhenry (2013) discuss technical aspects of the preservation of the USGS aerial film archive.

3 Working Draft ISO 19165

Since the year 2010 a group of the European National Mapping and Cadastral Agencies requested an ISO-standard for the preservation of geospatial data. The National Standardization Body of Germany (DIN) prepared a New Work Item Proposal that was accompanied by a Working Draft document. In the autumn of 2014 the members of the ISO/TC 211 "Geographic information / Geomatics" approved the new project with no no-votes. This was an indication of the strong demand for ISO-standard, now numbered ISO 19165.

3.1 *Scope*

According to the scope-section this standard sets the rules for the long-term preservation of digital geospatial data. These data include metadata and other ancillary data that are necessary to fully understand and rebuild the archived digital environment.

Geospatial data are preserved as a geospatial archival information package. This standard defines its details. A geospatial archival information package will be fully self-explanatory and will allow a future reconstruction of the dataset without external documentation.

3.2 *Terms and Definitions*

The first Working Draft contains 37 terms and definitions that were mostly adopted from the ISO 14721 "OAIS" and the ISO 19115-1 "Metadata – Part 1: Fundamentals".

3.3 *Summary of the ISO 14721*

The Information Package (IP) is central to the OAIS. The standard distinguishes further between: the input, the Submission Information Package (SIP); the storage, the Archival Information Package (AIP); and the output, the Dissemination Information Package (DIP) (Fig. 1). It does not provide any implementation or format for the IP.

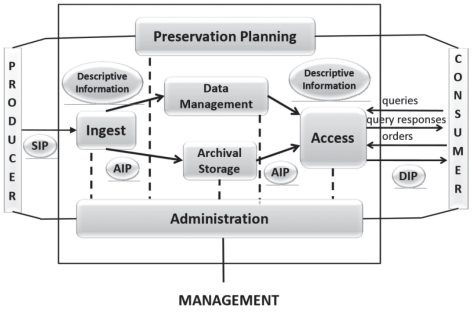

Fig. 1: Model of the Open Archival Information Standard (OAIS). SIP = Submission Information Package, AIP = Archival Information Package, DIP = Dissemination Information Package (ISO 14721:2012).

In addition, the ISO 14721 defines a number of responsibilities of the organization that operates an OAIS archive. Those include a sufficient control of the information needed to ensure long-term preservation, as well as copyright implications, intellectual property and other legal restrictions.

Every archive needs a designated community which is able to understand the archived information without needing special resources, such as the ones available to the experts who produced the information.

No matter how well an OAIS maintains its current holdings, it will eventually need to migrate much of its holding to different media. Digital migration is defined to be the transfer of digital information, while intending to preserve it.

Three major motivators are seen to drive digital migration of AIPs within an OAIS. These are:

- Improved cost-effectiveness: The rapid pace of hardware and software evolution provides greatly increasing storage capacities and transfer bandwidths at reducing costs
- New consumer-service requirements
- Media decay

3.4 *Preservation Principles of the ISO 19165*

3.4.1 Prioritization

The exponential growth of the data volume prevents a full archival of all data. Consequently, only a selected subset can go for a long-term archive. A temporal classification may follow the proposed categories 1 year, 10 years, and 100 years.

3.4.2 Data formats

Today, all geospatial data are stored in commonly accepted specialized data formats. Those formats have a specific structure and include metadata. Some of the formats are standardized by ISO and/or IEC (International Electrotechnical Commission), others are de-facto standards. In 2014, the format description document (FDD) database of the US Congress Library contained 334 format de-

scriptions, 34 of them are geospatially related (Library of Congress 2011). A more comprehensive list of formats can be found in the GDAL/OGR read and conversion open source libraries (GDAL 2015).

It is almost impossible to recommend a subset of formats since each format is used in its context. Instead, a geospatial dataset should be archived together with a fully documented data format specification. For practical reasons the documentation of the data format can be delegated to a format-registry.

3.4.3 Database

Many geospatial data are object-structured and stored in databases (sometimes known as geodatabases). In order to preserve this structure, the full database content should be transferred to the archive, which demands an archiving strategy that allows a persistent understanding of the technology for accessing this dataset.

3.4.4 Properties of geospatial data

Geospatial data often have a large number of attributes of which only a few are relevant for a specific mapping product. For instance, a 1:100,000 map does not show most of the road details, such as traffic lights. In the case of a limited archival space, this observation leads to another method of data reduction: assumptions about the future potential use of the data. Based on this assumption, a decision is needed which of the properties and which of their details will be archived or dropped. As a result, not all properties of the geospatial data are archived. Only those are maintained which are required to create one of the assumed products.

3.4.5 Level of aggregation

Often the same geospatial data exists at several servers with different levels of aggregation and processing. This means that prior to archiving it needs to be specified which levels of detail are required for archival. In some cases, raw data could be too dependent on the sensor technology or software licenses, and a further elaborated product could be easier to use and better for storage. In other cases, aggregated products can be derived from raw data with no or with little costs. Then, preserving the processing algorithms with the raw data could be the right choice.

3.4.6 Gold copy

The totality of all methods can never guarantee a full recovery of the data after a very long period of time. In order to increase reliability, a separate copy version of the 100 year data preservation should be established in an open format, file based repositories, avoiding databases, or other complex environments. Often, this copy is called a gold copy.

3.4.7 Intellectual Property Rights

As mentioned in the introduction, the regard of intellectual property rights imposed on the data is crucial. Authoritative geospatial data often possess legal restrictions that are written in license agreements. Licences for geospatial data usage need to include a special clause authorizing future curators the rights for preservation actions including archive, media migration and redistribution among future users of the archived data.

3.4.8 Time

According to an archiving rule the incoming data should have lost their relevance for the governmental work. However, many geospatial data are never obsolete or are continuously updated such as cadastral data. These kinds of geospatial datasets never become mature for archiving. The ISO 19165 defines a number of methods to overcome this problem.

3.4.9 Archiving package for geospatial data registries

The standard should define, as one of its central components, the elements of the AIP. This package should be ready to be shared with other organizations, including those outside the geospatial community. Studies or other forms of research are needed to find out the user's requirements.

The packaging format should be built upon existing standards. During the first project team meeting an example was presented: the Open Packaging Convention defined by Microsoft. This packaging format is standardized as ISO 29500-2 and ECMA-376 and it has a reference implementation in the geospatial world (MMZX format) introduced in the MiraMon GIS.

The ISO 19165 will define specialized versions of the IPs named Geo-SIP, Geo-AIP, and Geo-DIP. Their special properties include, lossless compression, cartographic series support (a manageable regional size, e.g. series of tiles of 10 km x 10 km), and a container for information regarding geometry (vector and raster), attributes, topology, metadata, quicklooks and recommendation on how to symbolize the data.

3.5 *Metadata built upon ISO 19115-1*

A geospatial dataset is always linked to a set of metadata. The metadata should be archived in a way that allows an undoubtedly reference between both data and metadata. Preservation requires that more emphasis is put on metadata. In the long term, when the producer of the dataset is no longer available metadata could become almost the only source of additional information about the preserved dataset.

A number of metadata models have already been proposed for the preservation of geospatial data such as the well-known prototyped PREMIS metadata model. It defines entities, among which are Intellectual entity (intellectual property of the dataset), Digital Objects (dataset), Agents (person or organization involved in the life of the dataset), and Rights (permissions pertaining to the dataset) (PREMIS 2012).

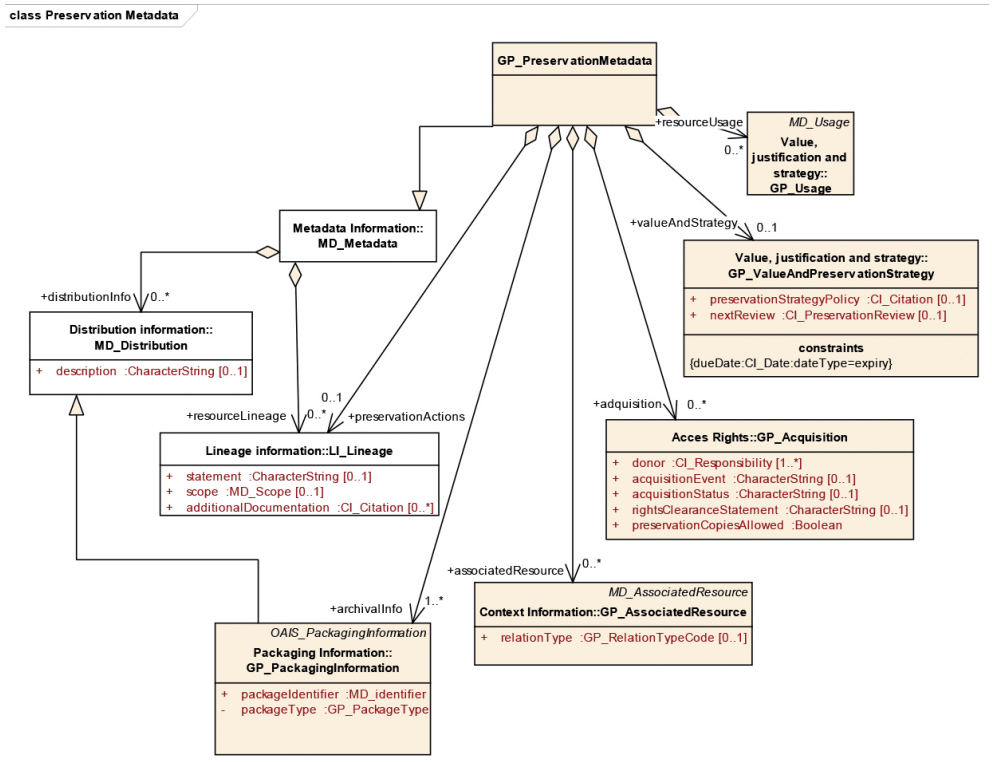

Fig. 2: Top-level classes of the metadata model for preservation (reddish) and the related classes of the ISO 19115-1 (white).

ISO 19115-1 "Metadata – Part 1: Fundamentals" is the metadata standard developed by the ISO/TC 211. The former version (ISO 19115) has been in place for more than a decade and is implemented in numerous environments worldwide. The first revision (ISO 19115-1) is descendent compatible to ISO 19115. It seems natural that the metadata for preservation follow the ISO 19115-1 principles.

ISO 19115-1 defines a comprehensive set of metadata for topics such as distribution, lineage, data formats, spatial and content representation, legal and security constraints, identification, usage, and coordinate reference systems. ISO 19165 will build upon those packages of the ISO 19115-1 metadata model, but define new specific "preservation classes" when the required information cannot be encoded in the ISO 19115-1 model. In addition, for preservation, and within the current ISO 19115-1, it is important to clarify about how to deal with data and metadata identifiers and how to include links to important documentation such as technical specifications, mission concepts, mission definition, data model descriptions. Fig. 2 shows the top-level class of the ISO 19165 model "GP_PreservationMetadata" which itself is a specialization of the MD_Metadata class. This class is shown in the upper left part of the diagram.

The additions to the ISO 19115-1 metadata model are: preservation actions (following the metadata class LI_Lineage), value, use and justification, acquisition, relations to other resources (associated resources), and preservation package information.

Preservation actions are processes that were done by the curators and not by the originally responsible parties of the dataset with the objective of preserving the data, e.g. documenting a media migration. *Value, use and justification* deals with documenting the administrative, legal, evidentiary, research, historically recognized value of the resource and the justification for preserving it, e.g. documenting the legal mandate to preserve the dataset for 10 years for legal purposes. It also includes geospatial services usage statistics, e.g. documenting the number of times the dataset was visualized in a web map service, as another means of justifying the importance of the dataset, the preservation criteria review dates, including the eventual decision of discontinuing the preservation of the resource. *Acquisition* element deals with how, when and with what legal constraints the dataset was given to a preservation body, e.g. including specific licenses agreements and restrictions to package redistribution. *Relations to other resources* reuses the MD_AssociatedResources and includes a PREMIS (PREMIS 2012) relation taxonomy, e.g. the dependency that a geological dataset has to a reference topographical map. *Preservation package information* lists all the parts (files and documents) that together form the AIP, e.g. enumerating all files covering the data values, additional data tables, the metadata, the product specifications, the format specifications, the symbols used in the visualization, and a quicklook in PDF format.

The acronym "GP" is a proposed shortcut that identifies the preservation standard within the ISO/TC 211 family of standards.

4 Next Steps of the Development Process

According to the ISO regulations, the expert's contributions shall be provided on the Working Draft level. This will last at least for the next one year. The following Committee Draft level initiates comments from the Technical Committee 211 "Geographic information / Geomatics". The later stages such as Draft International Standard are reserved for the ISO central secretariat to polish the standard on the editorial level.

The authors, who also chair the project team of the ISO 19165, ask for contributions which refine the proposed model and which eventually allows for the definition of more detailed procedures for a stable and save preservation of geospatial data.

References

CNAATD, 2015: http://cultura.gencat.cat/ca/departament/estructura_i_adreces/organismes/dgpc/temes/arxius_i_gestio_documental/cnaatd/accio_de_la_cnaatd/index.html (5.8.2015).

Engel, H., Engelmeier, S., Grams, G., Gültlinger, M., Häussermann, M., Hermann, C., Hoss, H., Keitel, C., Koch, E., Lang, R., Naumann, K.,

Schurer, H., Spohrer, M., Strietter, O., Witke, T., Wöhrle, M. & Wütherich, T., 2013: Archivierung von Geobasisdaten im Kontext der Gesamtüberlieferung des Vermessungswesens. – Landesamt für Geoinformation und Landentwicklung Baden-Württemberg, Stuttgart.

GDAL, 2015: Supported raster data formats: http://www.gdal.org/formats_list.html, Supported vector data formats: http://www.gdal.org/ogr_formats.html (30.9.2015).

GeoMAPP, 2010: Geospatial Multistate Archive and Preservation Partnership. – http://www.geomapp.net/default.htm (5.8.2015).

Higgins, S., 2008: The DCC Curation Lifecycle Model. – The International Journal of Digital Curation **3** (1): 134–140, http://www.ijdc.net/index.php/ijdc/article/download/69/48 (4.8.2015).

ISO 14721, 2012: Reference Model for an Open Archival Information System (OAIS). – http://public.ccsds.org/publications/archive/650x0m2.pdf (4.8.2015).

Jobst, M. & Gartner, G., 2011: Structural Aspects for the Digital Cartographic Heritage. – Jobst, M. (ed.): Preservation in Digital Cartography – Archiving Aspects, Springer.

Library of Congress, 2010: Appraisal and Selection of Geospatial Data. – 23 p., Library of Congress, Washington D.C., USA.

Library of Congress, 2011: Sustainability of Digital Formats Planning for Library of Congress Collections. – http://www.digitalpreservation.gov/formats/index.shtml (30.9.2015).

LTDP, 2012: Long Term Data Preservation, Earth Observation Preserved Data Set Content, LTDP/PDSC. – 80 p., European Space Agency.

Moe, D. & Longhenry, R., 2013: Metrically Preserving the USGS Aerial Film Archive. – Photogrammetric Engineering & Remote Sensing **2013:** 225–228.

NESDPCS, 2012: NASA Earth Science Data Preservation Content Specification. – Earth Science Data and Information System Project, 423-SPEC-001, 30 p., https://earthdata.nasa.gov/files/423-SPEC-001_NASA ESD_Preservation_Spec_OriginalCh01_0.pdf (30.9.2015).

PREMIS, 2012: Data Dictionary for Preservation Metadata, version 2.2, July 2012 (premis-2-2.pdf). – http://www.loc.gov/standards/premis (5.8.2015).

Ramapriyan, H.K. & Moses, J.F., 2012: NASA Earth Science Data Preservation Content Specification. – 30 p., NASA, 423-SPEC-001, Greenbelt, MD, USA.

Rönsdorf, C., Mason, P., Gerber, U., Bos, M., Shaon, A., Naumann, K., Kirstein, M., Samuelsson, G., Rantala, M., Kvarteig, S. & Stössel, W., 2013: GI+100: Long term preservation of digital Geographic Information – 16 fundamental principles agreed by National Mapping Agencies and Archives. – http://www.eurosdr-ireland.net/archiving/GI+100%20-%2016%20EuroSDR%20Archiving%20Principles_V3%201.pdf (4.8.2015).

Steffen, W., Richardson, K., Rockström, J., Cornell, S.E., Fetzer, I., Bennett, E.M., Biggs, R., Carpenter, S.R., de Vries, W., de Wit, C.A., Folke, C., Gerten, D., Heinke, J., Mace, G.M., Persson, L.M., Ramanathan, V., Reyers, B. & Sörlin, S., 2015: Planetary boundaries: Guiding human development on a changing planet. – Science **347** (6223), 1259855.

Wulder, M.A., Masek, J.G., Cohen, W.B., Loveland, T.R. & Woodcock, C.E., 2012: Opening the archive: How free data has enabled the science and monitoring promise of Landsat. – Remote Sensing of Environment **122:** 2–10.

Addresses of the Authors

Prof. Dr. Wolfgang Kresse, Politechnika Koszalińska, ulica Śniadeckich 2, PL 75-453 Koszalin, Tel.: +48-94-34-86-728, e-mail: wolfgang.kresse@tu.koszalin.pl

Dr. Joan Masó Pau, CREAF, Fac. Ciencies, Edifici C, Universitat Autònoma de Barcelona, Tel.: +34-93-581-1771, Fax: +34-93-581-4151, e-mail: joan.maso@uab.cat

Manuskript eingereicht: August 2015
Angenommen: Oktober 2015