GeoERA
INFORMATION
PLATFORM

Deliverable

Report on new legislation covering access/open access, etc.

Authors and affiliation:

Mary Carter (GSI), Christopher Luton (UKRI)

E-mail of lead author:
mary.carter@gsi.ie

Version: 1.0

Deliverable Data		
Deliverable number	D10.3	
Dissemination level	Public	
Deliverable name	Report on new legislation covering access/open access, etc.	
Work package	WP10, IPR and data policy issues	
Lead WP/Deliverable beneficiary	UKRI	
Deliverable status		
Submitted (Author(s))	3/6/2020	Mary Carter
Verified (WP leader)	15/6/2020	Chris Luton
Approved (Coordinator)	16/6/2020	Jørgen Tulstrup

GENERAL INTRODUCTION

This document looks at data archiving and storage with a particular emphasis on legislation and how it affects the processes. Guidance for geospatial data preservation is discussed, as well as International Standards which apply to the area.

EXECUTIVE REPORT SUMMARY

Archiving and storage of data has changed markedly in the digital era. With digital data it is important that data creators consider how data is to be preserved and archived. Most Geological Surveys are Government organizations who have been mandated to archive data by a National Archives Act. Various models have been developed over time staring with the Open Archival Information System (OAIS) framework which has been adopted and implemented widely. They have also developed into international standards such as ISO 19165-1:2018 Geographic information — Preservation of digital data and metadata, which is associated with ISO19115-1:2014/Amd.1:2018(en) Geographic information — Metadata.

Through various recent EU Directives culminating in the Open Data Directive the EU have tried to provide a comprehensive legal framework to make data reusable, promoting transparency, with great benefits for the citizens of Europe

TABLE OF CONTENTS

1   INTRODUCTION

This document looks at data archiving and storage with a particular emphasis on legislation and how it affects the processes. Guidance for geospatial data preservation is discussed.

1.1   Data Archives
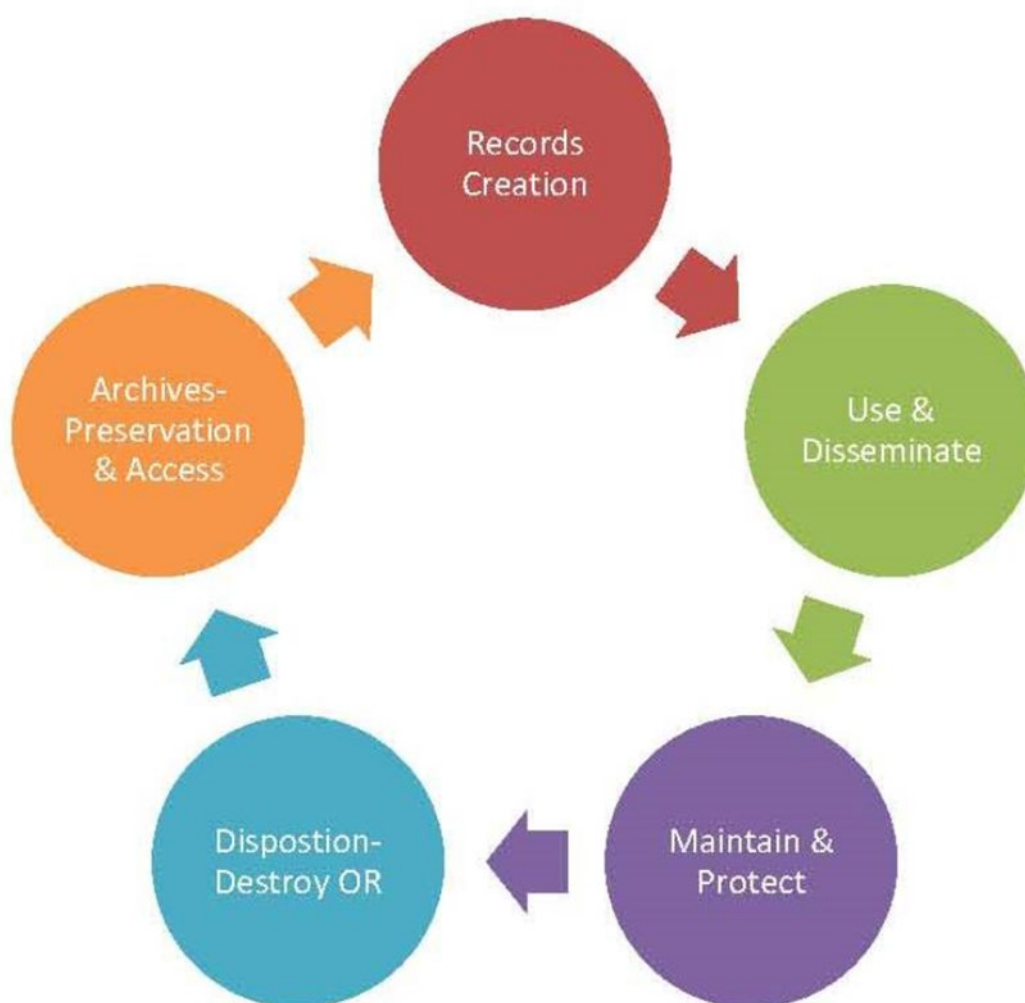
Figure 1 Records & Information Lifecycle
From https://slcoarchives.wordpress.com/2012/04/13/managing-records-now-for-the-future/

Data archiving is the management of static information moving it from production systems to an archive repository. Data preservation is a part of the data and information management lifecycle (see figure 1). Therefore it is important to contemplate the possibility of preservation when any data is being created. With paper data such as maps and documents the creator regularly has no input in the process of preservation, but with digital data it is different and preservation has to be considered by the creator.

1.2   Metadata

Metadata is defined as the data providing information about one or more aspects of the data; it is used to summarize basic information about data which can make tracking and working with specific data easier. Some bodies distinguishes among three types of metadata: descriptive, structural, and administrative. Descriptive metadata is typically used for discovery and identification, as information to search and locate an object, such as title, author, subjects, keywords and publisher. Structural metadata is about the structure of multipart objects, and administrative metadata provides information to help manage a resource such as how it was created, file type and who can access it.

When distinguishing between data and metadata, the main difference is that data is simply the content that can be a description of something, reading, measurements, observations, report anything. On the other hands, metadata describes the relevant information about the data. But, metadata is considered as processed data.

According to Article 5(1) of INSPIRE Directive 2007/2/EC, Member States are required to ensure that metadata are created for the spatial data sets and services corresponding to the themes listed in Annexes I, II and III, and that those metadata are kept up to date.

Since the Community institutions and bodies in most cases have to integrate and assess spatial information from all the Member States, **INSPIRE** recognises the need to be able to gain access to and use spatial **data** and spatial **data** services in accordance with an agreed set of harmonised conditions.

EU Government bodies are required to use the Portal at the following location:-

https://inspire-geoportal.ec.europa.eu/

The most current Geospatial Information Metadata standard is ISO19115-1:2014/Amd.1:2018.

Metadata is a fundament part of the data life cycle, and makes the data discoverable in large portals which organizations create to allow access to data of scientific value. Therefore data needs to be well described in metadata by the creator. Metadata is added during the archiving process in accordance with International archiving standards

2   LEGISLATION AND POLICY

2.1   National Archives

Most countries have a National Archives act. In Ireland, for instance, this is The National Archives Act 1986, and Regulations 1988. The Act operates within a legal framework for the management of government records that includes other statutes relating to areas such as Data Protection, Freedom of Information, and legislation relating to specific areas of work such as cultural heritage, policing, revenue and social protection, among many others. Any record in either paper or electronic format made or received and held by a Department of State in the course of its business is subject to the terms of the National Archives Act, 1986 and must be managed accordingly. The National Archives Act, 1986 only applies to departmental records over 30 years old.

2.2   GDPR

The General Data Protection Regulation (GDPR) stems from a European initiative. It came into force on May 25, 2018, and was designed to modernise laws that protect the personal information of individuals. It also boosts the rights of individuals and gives them more control over their information. It also protects publishers, product developers and subsequent users of datasets in that they have a process by which they can ensure the use of personal data does breach their rights. It is something anyone releasing personal information/data needs to be aware of and ensure personal rights are protected.

It sets out seven principals as regards personal data:

- Lawfulness, fairness and transparency
- Purpose limitation
- Data minimisation
- Accuracy
- Storage limitation
- Integrity and confidentiality (security)
- Accountability

2.3   Copyright

2.3.1   Main Directives

Copyright is partially harmonised throughout the EU by the Copyright Term Directive, the Information Society Directive and the Directive on Copyright in the Digital Single Market, but there are still some variations by country.

2.3.2   Directive on Copyright in the Digital Single Market 2019 - Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC

This Directive is aimed at making it easier for producers of content to negotiate deals with internet platforms for the use of their work. It makes internet providers responsible for content uploaded to their site.

The directive also makes it easier for research by text and data mining, but only where organisations have bought copyright and can only be carried out for scientific research and not for profit making organisations. Copyright holders are allowed to use technical security and integrity measures to protect their infrastructures (e.g. by limiting the speed or quota reserved to a certain IP address to bulk download the entire collection of data).

Teachers and students in formal education environments should be able to use digital copyright material except where the copyright owner is selling a licence to their content. Libraries and museums should be able to make copies of works for Cultural Heritage purposes, and where works are unavailable but still protected by copyright.

2.4   Licensing

The standard way to release information is by way of licence, even if it is a free licence. Most Government bodies insist on it as a prerequisite, because a licence contains some of the elements aimed at protecting both a supplier of information and a user. Even simple things like the parties to the licence are identified, and this helps to give an end-user the confidence permission have been granted for appropriate usage. A licence might also contain some terms and conditions a user must sign up to, ensuring correct and legal access. A term, for example, might be a simple acknowledgement as to the originator of data.

Creative Commons is one of the more popular licenses in use in Europe. https://creativecommons.org/

It is translated into most languages in the world and takes the form of one of six licences dependent of the type of use envisaged. I offers a standardised way of granting copyright permissions and ensure attribution while allowing others to copy, distribute and make use of data and information.

In the United Kingdom, the UK Government Licensing Framework (UKGLF) provides a policy and legal overview of the arrangements for licensing the use and re-use of public sector information both in central government and the wider public sector. It sets out best practice, standardises the licensing principles for government information, mandates the Open Government Licence (OGL) is a UK-based "open access" licence as the default licence for Crown bodies and recommends OGL for other public sector bodies.

The UKGLF and OGL form part of the UK Government's drive to open up access to publicly held information and datasets, promoting transparency and enabling wider economic and social gain.

2.5   Open Data Directive

The Directive on open data and the re-use of public sector information provides a common legal framework for a European market for government-held data (public sector

information). It is built around two key pillars of the internal market: transparency and fair competition.

The Directive on open data and the re-use of public sector information, also known as the 'Open Data Directive' (Directive (EU) 2019/1024) entered into force on 16 July 2019. It replaces the Public Sector Information Directive, also known as the 'PSI Directive' (Directive 2003/98/EC) which dated from 2003 and was subsequently amended by the Directive 2013/37/EU (more about the process of this revision).

The review process that led to the adoption of the Open Data Directive was launched in 2017, when the European Commission opened a public online consultation on the review of Directive 2013/37/EU, fulfilling the periodic review obligation prescribed by the Directive. Building on the results of the consultation, together with an extensive evaluation of the Directive and an impact assessment, a proposal for a revision of the Directive was adopted by the European Commission on 25 April 2018. On 22 January 2019, negotiators from the European Parliament, the Council of the EU and the Commission reached an agreement on the revision proposed by the Commission. When adopted in June 2019, the Directive was renamed as the Open Data and Public Sector Information Directive and will make public sector and publicly funded data re-usable.

The new PSI Directive, which extends previous legislation, focuses on the economic aspects of the re-use of information rather than on access to information by citizens. It is similar to the INSPIRE Regulation, which countries were again under a legal obligation to enter into national laws, within a set date – ensuring that metadata standards were applied throughout Europe. It encourages the Member States to make as much information available for re-use as possible. It addresses material held by public sector bodies in the Member States, at national, regional and local levels, such as ministries, state agencies and municipalities, as well as organisations funded mostly by or under the control of public authorities (e.g. meteorological institutes).

Since its 2013 revision, content held by museums, libraries and archives also falls within the scope of application of the Directive.

The Directive covers written texts, databases, audio files and film fragments; it does not apply to the educational, scientific and broadcasting sectors.

Once fully transposed on the national level (which has now happened across Europe with a fixed date for implementation in EU countries), the new rules will:

- Stimulate the publishing of dynamic data and the uptake of Application Programme Interfaces (APIs).
- Limit the exceptions which currently allow public bodies to charge more than the marginal costs of dissemination for the re-use of their data.
- Enlarge the scope of the Directive to:
    - data held by public undertakings, under a specific set of rules. In principle, the Directive will only apply to data which the undertakings make available for re-use. Charges for the re-use of such data can be above marginal costs for dissemination;
- research data resulting from public funding – Member States will be asked to develop policies for open access to publicly funded research data. New rules will

also facilitate the re-usability of research data that is already contained in open repositories.

- Strengthen the transparency requirements for public–private agreements involving public sector information, avoiding exclusive arrangements.

In addition, the Open Data Directive requires the adoption by the Commission (via a future implementing act) of a list of high-value datasets to be provided free of charge. These datasets, to be identified within a thematic range described in the Annex to the Directive, have a high commercial potential and can speed up the emergence of value-added EU-wide information products. They will also serve as key data sources for the development of Artificial Intelligence.

Member States have to transpose Directive (EU) 2019/1024 by 16 July 2021.

The Directive introduces the concept of high value datasets, defined as documents the re-use of which is associated with important benefits for the society and economy. They are subject to a separate set of rules ensuring their availability free of charge, in machine readable formats, provided via Application Programming Interfaces (APIs) and, where relevant, as bulk download. The thematic scope of high value datasets is provided in an Annex to the Directive.

The thematic categories of high-value datasets, as referred to in Article 13(1) of the Directive, are:

1. Geospatial
2. Earth observation and environment
3. Meteorological
4. Statistics
5. Companies and company ownership
6. Mobility

Within these limits, and with the assistance of a Committee on open data and the re-use of public sector information composed of Member States representatives, the Commission will adopt in 2021 a list of specific high value datasets by way of an implementing act, following an impact assessment.

2.6 Objectives of EU Copyright Directives

Digital technologies have radically changed the way creative content is produced, distributed and accessed. Member States are adapting the EU copyright rules to new consumer behaviours in a Europe which values its cultural diversity.

Copyright ensures that authors, composers, artists, film makers and other creators receive recognition, payment and protection for their works. It rewards creativity and stimulates investment in the creative sector. 33 sectors of the EU economy are considered copyright-intensive, accounting directly for over 7 million jobs, or 3% of employment in the EU.

EU actions have led to more harmonised protection of rightholders, lower transaction costs and greater choice for users of content, notably through:

- a European regulatory framework for copyright and related rights;

- the promotion of inclusive and dynamic stakeholders dialogues on copyright and related issues, to seek views, concrete experience and contributions from all interested parties; and
- a leading role in international negotiations and discussions on copyright and related issues.

3   FRAMEWORKS

GSPs will need to ensure they use one of these types of frameworks.

3.1   OAIS

The Consultative Committee for Space Data Systems (CCSDS) was formed in 1982. It has been developing recommendations for data and information systems standards to promote interoperability and cross organizational support among cooperating space agencies. CCSDS 650.0-M-2 is the reference model for Open Archival Information Systems (OAIS) and the latest version was published in 2012. The ISO equivalent is 14721:2012 Space data and Information transfer systems – Open Archival Information systems (OAIS) – Reference model

Although it's primary focus is on digital information it does accommodate material such as digital maps. There are three primary roles Producer, Management and Consumer.
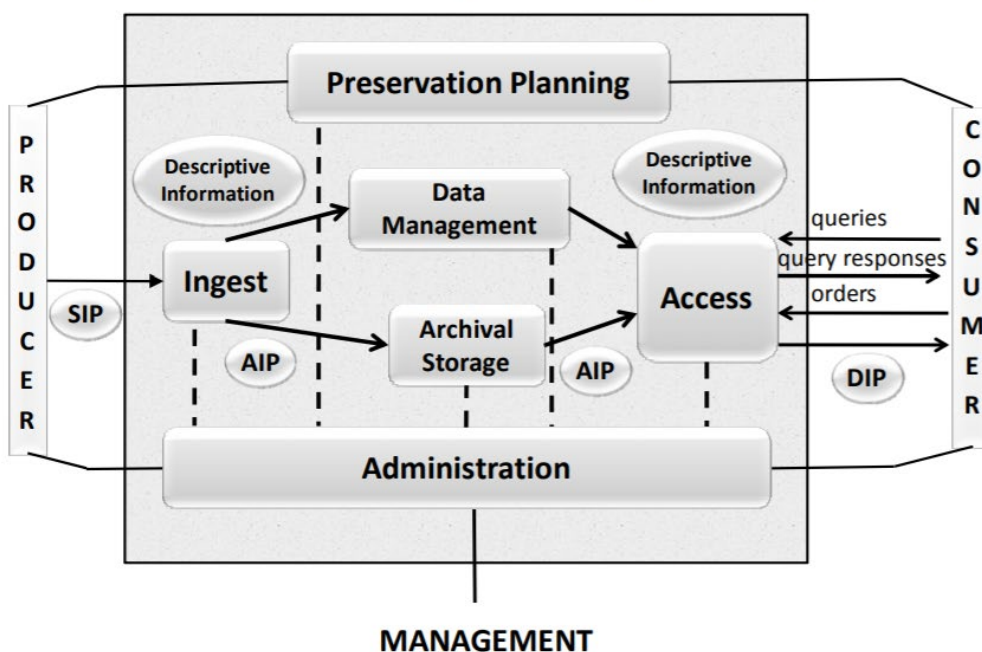

Figure   Oasis Functional Entities *source CCSDS(2012)*

There are six functional entities in the model

INGEST provides the services and functions to form the interface between the creator and the preserver. The data is received as a submission interface package (SIP) and

an archival Information package is created (AIP) which contains the SIP and the preservation description information (PDI)

ARCHIVAL STORAGE provides services and functions for the storage and retrieval of AIPs

DATA MANAGEMENT provides services and functions to manage administrative information about the records and descriptive information about the archived material.

ADMINISTRATION provides the services and functions to control the operation of the entire archive system. This includes include soliciting and negotiating submission agreements with Producers, auditing submissions to ensure that they meet Archive standards, and maintaining configuration management of system hardware and software

PRESERVATION PLANNING provides the services and functions to the long term maintenance of records as they go through migration and updates. It ensures records are consistently available and maintains authenticity. This is where prototype migration plans are tested.

ACCESS provides the services and functions that allow consumers to access the records, and it creates dissemination information packages (DIP) to deliver records to consumers.

3.2 European Long Term Data Preservation Common Guidelines

In 2006 a consolidated European LDTP Guidelines was produces by the European Space Agency, when they adapted OAIS to their context. The guidelines are contained in eight themes.

1. Preserved Data Content Definition and Appraisal
2. Archive Operations and Organization
3. Archive Security
4. Data Ingestion
5. Archive Maintenance
6. Data Access and Interoperability
7. Data Exploitation and Reprocessing
8. Data Purge Prevention

Preserved Data Content Definition and Appraisal is analogous to a SIP and an AIP in the OAIS reference model. In order to make sure data is fully usable in the future it not only needs the associated metadata, but also 'associated knowledge' such as information on intellectual property, access restrictions, specifications etc.

Archive Operations and Organization One of the guidelines includes the adoption of the OAIS reference model. This consists of all daily activities needed to run and monitor the archival system.

Archive Security encompasses a comprehensive information security framework to guarantee archive security.

Data Ingestion encompasses the services and functions that, according to OAIS standard accept Submission Information Packages (SIPs) from data producers, prepare

Archival Information Packages (AIPs) for storage, and ensure that the Archival Information Packages and their supporting descriptive information are stored in the archive system.

Archive Maintenance consists of all the activities aimed at guaranteeing the integrity of the archived data. Data integrity assures that the archived data are complete and unaltered through loss, tampering or data corruption. Archive maintenance refers to the storage of equipment, media and hard disk arrays in secured and environment controlled rooms, and a set of defined activities to be performed on a routine basis, such as migration to new systems and media, in accordance to the technology and consumer market evolution, data compacting and data format/packaging conversion.

Data Access and Interoperability corresponds to the activities that make the data archive content visible and available to users. Interoperability is related to the possibility of accessing data in a common and standardized way despite the intrinsic differences between the data sets on one hand and the accessed systems on the other hand.

Data Exploitation and Reprocessing This theme covers activities related to the exploitation of archived data by data processing and reprocessing, regeneration or enhancement of the catalogues (e.g. through data mining), integration of new services (e.g. through service work-flow orchestration) and quality assessment of the products and services.

Data Purge Prevention Data purging is any loss of the archived data, and this theme sets out procedures to minimise any loss.

3.3 GI+100: Long term preservation of digital Geographic Information — 16 fundamental principles agreed by National Mapping Agencies and State Archives

EuroSDR is a not-for-profit organisation linking National Mapping and Cadastral Agencies with Research Institutes and Universities in Europe for the purpose of applied research in spatial data provision, management and delivery. These principals were compiled by the EuroSDR Archiving working group and sets out the path and describes what needs to be done now to future-proof the investment government agencies around the world have made in creating digital Geographic Data. It was approved by the EuroSDR Board of Delegates on 30th May 2013

1. Archiving of digital Geographic Information begins at the point of data creation, rather than at the point of withdrawal from active systems
2. All organisations must have a maintained Archiving Policy
3. Be selective and decide what to archive and what to dispose of
4. Consider preservation timeframes of 1, 10, 100 years
5. Migration or emulation is inevitable in the medium and long term. Be prepared and choose which properties to preserve in advance
6. The output of the archival planning process should also be preserved over the long-term to accommodate future preservation requirements
7. Archiving is not back-up. You must also back-up your archive
8. Geographical data should be preserved in a way that non geo-specialists can handle
9. Information objects should be self-contained and independently understandable

10. Keep the gold copy version of the 100 year data archive in open, file based repositories, not in databases, nor other complex environments

11. Consider keeping a graphical representation alongside the logical representation of the data

12. Restrict the number of formats and encodings to a widely agreed set of open, simple and well documented file formats

13. Prefer simple data models and schemas over complex ones

14. Keep the access mechanism for archived data simple. Focus on basic current user requirements – an archival viewing system does not need to be a fully functioning GIS

15. Ensure effective management and quality assurance of the metadata associated with your data

16. Make some assumptions about future use, but don't be too restrictive

3.4 ISO 19165-1:2018 Geographic information — Preservation of digital data and metadata

ISO19165-1 defines the requirements for the long-term preservation of digital geospatial data. The standard sets rules for the long term preservation of data and metadata. It uses OAIS which does not include any implementation or format for the information package (IP). Also a lot of the elements were found to be metadata so ISO19115 was extended to for 19115-1, and 19165 became a specialization of the geospatial metadata covered in 19115-1. A geospatial dataset is always connected to a set of metadata. It also has a Value Strategy to document criteria for selecting what to preserve. ISO19165 addresses geospatial data. Its data structures, data formats, and Intellectual Property rights.

4 EXAMPLES

4.1 Archiving of Federal Geodata Switzerland

First the legal basis was set out with Federal archiving legislation and then specific Geoinformation Act 2008. A Handbook was written in 2016, drawn up as part of "Project Ellipse", for the purpose of operational implementation of archiving of geodata in the Federal Office of Topography, swisstopo. Swisstopo also participated in the EuroSDR archiving group which produced "GI+100: Long term preservation of digital Geographic Information — 16 fundamental principles agreed by National Mapping Agencies and State Archives" at 3.3 above.

Conservation and archive planning (CAP) was conducted as part of project Ellipse. This involved defining the conservation periods for long term availability and evaluating the geodata for archival value. Authorities update CAP once a year.

Archiving in swisstopo is integrated with the Federal Spatial Data Infrastructure (FSDI), and is integrated with other Swiss federal authorities, including the Swiss Federal Archives (SFA). Data archiving at the SFA is based on the OAIS reference model, as well as all geodata archiving.

5. CONCLUSIONS

Changing from methods used in paper data archiving required a change of mindset where one has to consider how the data will be archived from the beginning. Directives are only now catching up with the ways that digital data are used. The Directive on Copyright makes it easier to research by the new methods of text and data mining while still protecting the rights of the Copyright holder. The Open Data Directive replenishes and replaces the Public Service Information Directive. With the help of all the new EU directives, many governments (extending also outside the EU) are using or replicating this new legislation to open up access to publicly held data and information, thereby delivering a better service to citizens. We are looking towards a future where we can extract the maximum economic advantage, building a value driven economy.