

D- EITI Open Data Concept

Content

Foreword	1
1. <u>Open Data in the D-EITI process</u>	2
2 <u>Ten D-EITI Open Data principles</u>	4
1 <u>Completeness</u>	4
2 <u>Primary sources</u>	4
3 <u>Timely availability</u>	4
4 <u>Easy access</u>	4
5 <u>Machine readability</u>	5
6 <u>Non-discrimination</u>	5
7 <u>The use of open standards</u>	5
8 <u>Licensing</u>	6
9 <u>Persistence</u>	6
10 <u>Free use</u>	6
3. <u>Recommendations for implementation</u>	7
Glossary	8
Annex	12

Foreword

With this Concept, the multi-stakeholder group (MSG) of the Deutschland Extractive Industries Transparency Initiative (D-EITI) sets out the principles for the implementation of the Open Data principle within the framework of the D-EITI. The Concept should only be used for data and information which is public, for data within the context of the current legal basis and for data the publication of which has been decided by the MSG within the framework of the D-EITI. No provisions for the scope and the process of reporting within the framework of the D-EITI result from the Concept itself. The implementation of the Concept and the use of the Open Data principles mainly take place together with the publication of the annual D-EITI report. No obligations for the reporting institutions and companies to prepare open data arise from the Concept.

The current definition from the [Open Data Handbook](#) is used for Open Data:

“Open data is data that can be freely used, re-used and shared by anyone – the only restriction concerns the obligation of naming the author.”

The reasons for the use of open data in the context of D-EITI are described in the first chapter of the Concept. Chapter 2 provides an overview of the ten principles for the implementation of open data within the framework of the D-EITI. The third chapter contains a list of measures to which the MSG has agreed; they are designed to support the implementation of the D-EITI Open Data Concept and to ensure its success. At the end of the Concept is a Glossary with explanations of the key terms in the context of Open Data, plus the Annex, which contains the text of the Open Data requirements from the International EITI Standard.

This present version of the D-EITI Open Data Concept was unanimously adopted at the 7th MSG meeting of the D-EITI on 21.9.2016.

1. Open data in the D-EITI process

Germany started the implementation of the “Extractive Industries Transparency Initiative” – EITI in 2014. The [German EITI](#) is preparing a relevant annual report, which will be published from 2017 on. In addition to context information about the German extractive sector, it also compares the most important financial flows of extractive companies with government agencies.

The MSG controls the implementation of the EITI in Germany and it decided on the [first objective](#) for the EITI implementation in Germany at its meeting of 06.10.2015. This decision was approved in the presence of Mr. Uwe Beckmeyer, who is the Special Federal Government Commissioner for EITI Implementation in Germany and Parliamentary State Secretary at the Federal Ministry for Economic Affairs and Energy (BMWi). The first objective is:

“To produce timely reports that are understandable, accessible to the general public and based on a transparent, open and innovative EITI process in Germany.”

This objective reflects the [International EITI Standard](#), the updated 2016 version of which has significantly strengthened the demand for open access to EITI reporting data. EITI Requirement 7 addresses the result and impact of the reporting – and it not only instructs the implementing countries to print the EITI report and publish it online (7.1a); in points 7.1b and c, they are also urged to

“agree upon a clear strategy with regard to the access, sharing and re-use of EITI data. The implementing countries will be encouraged to publish their EITI data under an open license and to advise users that the information may be used without prior consent”;

and “to place the EITI report online in an Open Data format (.xlsx or.csv) and make it generally known that the report is available.”

EITI Requirement 7.2 and the EITI Open Data Directive also encourage the implementing countries to design access to EITI-relevant data in such a way that it can be compared with publicly-available data and understood as fully as possible.

Open Data is therefore an essential prerequisite for conforming to the International EITI Standard – and at the same time it is an important contribution towards the achievement of the objectives the German MSG has set for itself.

Open Data in the context of the D-EITI can support a transparent presentation of government and companies' activities in the extractive sector and also make a positive contribution to:

- accountability
- good governance
- public debates and dialogues
- efficiency and acceptance of the management.

This D-EITI Open Data Concept sets out the basic principles for dealing with the data which is collected within the framework of the D-EITI and which is intended for public reporting purposes. For each reporting period, the MSG decides which data should be earmarked for D-EITI reporting – and consequently the data that will be published in the format of open data in compliance with the requirements of this Concept. In both cases, publication takes place within the framework of the publication of the D-EITI report. In the case of information relating to operating and business secrets of companies, care is also taken to ensure that these are not published in an unauthorised manner, without the approval of the MSG and of the (individual) companies involved.

In April 2016, Germany announced that it would join the Open Government Partnership (OGP). The OGP is an international initiative that was launched in 2011 by the US and Brazilian governments. 70 countries are now involved with the initiative, the aims of which are to create more transparency, accountability and citizens' participation. To qualify for membership, a state must first meet certain requirements; then it must create a national action plan together with civil society and implement the plan in a timely manner. As in other OECD countries (e.g. USA or France), the national EITI implementation in Germany can also contribute to the national action plan, providing a positive impetus to the work of the federal government. The implementation of the principle of Open Data offers the chance to make the D-EITI into an Open Data Lighthouse Project in Germany.

2. Ten D-EITI Open Data principles

The following ten D-EITI Open Data principles are based on the [overriding principles](#) pertaining to open government information. These principles were issued by the [Sunlight Foundation](#) in 2007 and are regarded as being the standard for Open Data. They were adapted for the German EITI process and are subject to the applicable legal provisions.

1 Completeness

The records published by the D-EITI should be as complete as possible. They should represent the entire scope of what is documented within the framework of the D-EITI process. All the raw data of a record should be made available to the public if this can be achieved without violating any laws (e.g. the Federal Data Protection Act, the Copyright Act, or the Freedom of Information Act); however, data which is for the protection of personal or otherwise confidential data according to the law, as well as data which the MSG decides is not to be published will not be made available to the public. Metadata that describes the raw data and provides explanations about it should also be provided, together with formulas and explanations about the calculation of the data. This will enable users to fully understand the content of the available D-EITI information and to examine each data element in maximum detail.

2 Primary sources

Records published by the D-EITI should be primary sources. Ideally, this includes the originally-acquired information, details on how the data was acquired and the original source documents that document the survey. Public dissemination enables users of D-EITI data to verify that the information has been properly acquired and recorded accurately.

3 Timely data availability

Records published by D-EITI should be available to the public within a reasonable period of time. If at all feasible, information acquired by the D-EITI should be published as soon as it is acquired and compiled. Priority should be given to data the usefulness of which is time-dependent.

4 Easy access

Records published by the D-EITI should be made as accessible as possible. Accessibility is defined as the “ease with which information can be obtained”. Some hurdles for automated,

electronic access include data access only being permitted after entry forms have been completed and systems that require browser-based technologies (e.g. Flash, JavaScript, cookies and Java applets). In contrast, an interface for users that allows the immediate downloading of all stored data (also known as “Bulk Access”) in one single procedure and the offers to acquire specific data via a programming interface (API) make information much more accessible. Another important aspect is searchability, i.e. the ability to easily find and download content. The D-EITI data portal should ensure that such ease of access always exists.

5 Machine readability

Programs and machines can cope with certain types of document formats much better than other formats – for example, machines find information that is disseminated in the well-known PDF format very difficult to analyse. Information should therefore be stored in established file formats that can be easily read by machines. Exemplary formats for this are xls, csv and xml file formats. D-EITI files should be accompanied by documentation which pertains to the format and explains how to use it in relation to the data.

6 Non-discrimination

Non-discrimination refers to who can access the data and how it is accessed. In the broadest sense, non-discriminatory access means that every person has access to the data at all times, without having to identify themselves or having to provide justification for their actions. The D-EITI data portal should also ensure the existence of this non-discriminatory access.

7 The use of open standards

The demand for the use of jointly-developed (“open”) standards refers to the ownership of the formats used. For example, if only one company makes a program that can read a file in which the data is stored, access to this information depends on the use of the processing program of that particular company. A program like this is sometimes available for the general public free of charge, but a fee may also be involved. Microsoft Excel, for example, is a fairly common spreadsheet program and it costs money to use it. However, there are often freely-available formats on the market that can access the data without the need for a software license. If these cost hurdles are eliminated, the data will be available to a larger group of potential users. We recommend the csv file format, or TSV/JSON format.

8 Licensing

Imposing “Terms of Use”, the obligation to disclose names, restrictions on dissemination etc. are all obstacles for the public use of data. Maximum openness therefore means that public information should be clearly marked as the work of the D-EITI, but it should be made available in the public domain without any usage restrictions. This allows the use of the [CC BY 4.0 license](#), or the [Datalizenz Deutschland Version 2.0](#), which we recommend.

9 Persistence

(Data) persistence is defined as a situation where specific information can still be found over a long period of time. Records published by the D-EITI online should persist, meaning that they should be permanently available in online archives. Information is often updated, changed or removed without any information being given on the changes that have been made; or information is made available as a data stream, but is not archived anywhere at all. In order to promote good usage by the public, information that is put online should have appropriate version control and be permanently archived online. The D-EITI data portal should ensure that data persistence will be available.

10 Free use

All data provided by the D-EITI should be available free of charge.

3. Recommendations for implementation

The MSG of the D-EITI has agreed on the following measures to ensure the successful implementation of the D-EITI Open Data Concept:

- In the case of data that the MSG has earmarked for publication, we ensure the availability of the machine-readable data in both granular and aggregated forms on the D-EITI data portal.
- We encourage publication of the metadata associated with the D-EITI data on GovData.
- We promote the creation of an open register on the beneficial owners of the D-EITI reporting companies, in order to highlight the exemplary function of the D-EITI's Open Data aspirations within Germany and also on an international level within the framework of the EITI.
- In the long term, we want to ensure the implementation of interoperability for the future international standard of the international EITI Board.
- We strive to achieve an international exchange of experiences on the implementation of Open Data in the context of the EITI.
- We promote data skills in the preparation and use of open data in the digitisation process within the individual participating organisations, while at the same time actively assisting authorities, companies, NGOs, media, etc.
- We want to achieve the long-term integration of the D-EITI process into public administration processes – this will ensure timely data, good data quality, reusability and cost effectiveness; and we try to depict the federal structure in the best possible light.

Glossary

Aggregated data

Aggregated data results from the consolidation of individual values into larger units. The level of detail of individual data submitted in reports by companies or by government agencies within the framework of the D-EITI is determined by the International EITI Standard and by the decisions made by the MSG.

Bulk access

Bulk access refers to the downloading of all the stored data in one single procedure.

CC BY 4.0 license

This license allows the users of the data to

- share it – the material may be reproduced and further circulated in any format or medium.
- process it – the material may be remixed and changed. It may also be expanded for any purpose, even for commercial reasons.

The licensor cannot revoke these freedoms as long as users adhere to the licensing conditions.

The conditions for sharing and processing the data are:

- Disclosing of names: Users must provide adequate copyright and rights information, add a link to the license and specify whether or not changes were made. This information may be given in any reasonable manner, but not if the impression is created that the licensor is only supporting the user or the usage.
- No further restrictions: Users may not use any additional clauses or technical methods that legally prohibit others from doing anything that the license allows.

Further information at: <https://creativecommons.org/licenses/by/4.0/deed.de>

Cookies

A cookie is a small file that is stored on the PC of a visitor to the website. Cookies store information such as the user's preferred language and other personal settings. The next time

the user visits the website, this information is transmitted back to the site, enabling the display of individual and user-tailored information.

CSV file format

CSV stands for comma-separated values. It describes the structure of a text file for storing or exchanging data that has a simple structure. The file name extension is “.csv”. Tables or a list of various lists of different lengths can be displayed in.csv files.

Data licence Deutschland Version 2.0

The data license Deutschland Version 2.0 is an “open license” and therefore complies with all the Open Data legal requirements. More information on this licence is available at <https://www.govdata.de/dl-de/by-2-0>.

Flash

Flash by Adobe is a platform for the programming and display of multi-media and interactive content. The platform can display, animate and manipulate vector graphics, raster graphics and video clips.

Granular data

Granular data is the display of individual values with the greatest possible level of detail. The level of detail of individual data submitted in reports by companies or by government agencies within the framework of the D-EITI is determined by the International EITI Standard and the decisions made by the MSG.

GovData

GovData is the data portal for Germany at <http://govdata.de>. The portal offers uniform, centralised access to the administrative data of federal, state and local authorities. The aim of GovData is to have the data in one place, making it easier to find and use.

Interface

An interface is a ‘common border’ between two separate components of a computer system. The purpose of the interface is the exchange of information, which can take place between software, computer hardware, human beings and combinations of these. The term interface also partly refers to the user interface or GUI (graphical user interface).

Interoperability

Interoperability refers to the connectivity of the data to future developments in the International EITI Standard, to achieve a level of cooperation that is as seamless as possible. The purpose of interoperability is to exchange information in an efficient and usable manner, or to make it available to the user.

Java Script/Java Applets

JavaScript is a scripting language that was originally designed for dynamic HTML in web browsers. A Java applet is a computer programme that is created using Java technology. It typically runs in a web browser.

JSON

JavaScript Object Notation, JSON for short, is a compact data format in an easily-readable text format for the purpose of data exchange between applications.

Open data

Open data is data that can be freely used, re-used and shared by anyone – the only restriction is that the author must be named.

Metadata

Metadata is data that contains information about the characteristics of records or documents (but not about the data itself) – it categorises the data, for example and describes its license.

Primary data or raw data

Primary data is data that is directly acquired from a data survey. Primary data exists when the acquired data has neither been accumulated nor evaluated. It is also known as raw data. In individual cases, the data may already have been edited to make it manageable or to meet legal requirements like data protection and statistical confidentiality – data like this is called basic data. In the context of the D-EITI, the MSG determines which primary data is to be requested or delivered for the report.

Primary sources

Data should be published directly from the original sources (primary sources) and enriched with relevant information concerning the collection and compilation of the data. This allows third parties to track the data creation. In the context of the D-EITI, primary sources are the various companies and government agencies that have been requested to provide data in accordance with a decision made by the MSG.

Programming interface (API – application programming interface)

A programming interface – or more specifically an interface for programming applications (often shortened to API) – is part of a programme made available by a software system of other programmes. Its purpose is to connect to the system in question.

Source documents

Data and documents that have been collected at their source.

Raw data

See primary data.

TSV file format

The TSV file format stands for “Tab Separated Values” and is created and used by many spreadsheet programs. The contents of these files can display text and mathematical, scientific or statistical data separately in rows and columns. The file format is similar to CSV files, but it has a different separator.

XLS file format

XLS is a proprietary file format of the Microsoft Excel spreadsheet program.

XML file format

Extensible Markup Language or XML is a markup language for displaying hierarchically-structured data in the form of text files.

Annex

EITI Requirement 7.2: Access to data¹

The multi-stakeholder group is encouraged to make EITI Reports machine readable, and to code or tag EITI Reports and data files so that the information can be compared with other publicly available data by adopting Board-approved EITI data standards. As per Requirement 5.1(b), the multi-stakeholder group is encouraged to reference national revenue classification systems, and international standards such as the IMF Government Finance Statistics Manual. The multi-stakeholder group is encouraged to:

- a) Produce brief summary reports, with clear and balanced analysis of the information, ensuring that the authorship of different elements of the EITI Report is clearly stated.
- b) Summarise and compare the share of each revenue stream to the total amount of revenue that accrues to each respective level of government.
- c) Where legally and technically feasible, consider automated online disclosure of extractive revenues and payments by governments and companies on a continuous basis. This may include cases where extractive revenue data is already published regularly by government or where national taxation systems are trending towards online tax assessments and payments. Such continuous government reporting could be viewed as interim reporting, and as an integral feature of the national EITI process which is captured by the reconciled EITI Report issued annually.
- d) Undertake capacity-building efforts, especially with civil society and through civil society organisations, to increase awareness of the process, improve understanding of the information and data from the reports, and encourage use of the information by citizens, the media, and others.

¹Compare EITI Standard 2016, Page 30

Open Data directive²

Preamble

1. This policy contains recommendations on open data in implementation of the EITI. It builds on lessons emerging from national level implementation and emerging international best practice.

2. The EITI Principles declare that “a public understanding of government revenues and expenditure over time [can] help public debate and inform choice of appropriate and realistic options for sustainable development” (EITI Principle 4). The EITI Standard therefore requires EITI Reports that are “comprehensible, actively promoted, publicly accessible, and contribute to public debate” (EITI Requirement 7.1). Improving the accessibility and comparability of EITI data is essential to realise these objectives.

Open Data Objectives

3. Open EITI data can increase transparency about what government and business activities, and increases awareness about how countries’ natural resources are used and how extractives revenues are levied and spent, providing strong incentives for that money to be used most effectively.

4. Open data promotes accountability and good governance, enhances public debate, and helps to combat corruption. Providing access to government data can empower individuals, the media, civil society, and business to make better informed choices about the services they receive and the standards they should expect. Open data, can also be a valuable tool for government in improving policy making and sector management.

5. Free access to, and subsequent re-use of, open data are of significant value to society and the economy.

Open Data in EITI Implementation

6. EITI Implementing countries are encouraged to:

a) orient government systems⁹ towards open data by default. It is recognised that there is national and international legislation, in particular pertaining to intellectual property, personally-identifiable and sensitive information, which must be observed;

²See [English Standard](#) P. 46-47, German Standard 2016 P. 45-46

- b) ensure that this data are fully described, so that users have sufficient information to understand their strengths, weaknesses, analytical limitations, and security requirements, as well as how to process the data;
- c) release data as early as possible, allow users to provide feedback, and then continue to make revisions to ensure the highest standards of open data quality;
- d) release data under an open license that allows users to freely obtain and easily re-use it;
- e) share technical expertise and experience with other countries to maximise the potential of open data;
- f) work to increase open data literacy and encourage people, such as developers of applications and civil society organisations that work in the field of open data promotion, to unlock the value of open data;
- g) ensure that data is interoperable with national and international standards, including adopting data standards approved by the EITI board and additional guidance provided by the EITI secretariat;
- h) where possible use unique identifiers to link data across years of reporting or different sources;
- i) work towards mainstreaming the creation of open data for EITI into government systems to ensure timeliness, data quality, reuse and cost effectiveness;
- j) provide data in granular, machine-readable formats.