📇 DPGAlliance / **dpg-resources** Public

<> Code   ⊙ Issues **1**   ⇅ Pull requests **2**   ▶ Actions   📖 **Wiki**   ⊘ Security   📈 Insights

5. Documentation

Jump to bottom

Bolaji Ayodeji edited this page on May 16 · **19 revisions**

> ⓘ **Note**
>
> **Indicator Requirement**: "Digital public goods require documentation of the source code, use cases, and/or functional requirements."

For this indicator, you must provide detailed documentation of your digital solution that will enable anyone unfamiliar with the project to understand how to use, deploy, and modify it. The sections below are guidelines for each DPG category.

Jump to DPG Category ↓

- 🖥️ Open Software
- 📊 Open Data
- 🤖 Open AI System
- 📚 Open Content

Open Software

For Open Software solutions, documentation should include guides, technical specifications, functional requirements, etc., that would allow a technical person unfamiliar with the software to launch and run it. The documentation must show the following aspects (non-exhaustive list):

- How to install the software (local environments, testing, code runs etc).
- How to fork the software (forking, patching, contributing upstream and downstream).
- How to deploy the software as a user.
- Any additional context (both technical and non-technical) that could help a user or a developer navigate through the software.

▶ **Click here to view list of common sections or types of software documentation.**

Open Data

For Open Data solutions, documentation must include enough descriptive information to ensure easier use. Data that has been well documented is recognizable, comprehensible, and usable in the future. You should record your data at each stage of the research or data collection process.

▶ **Click here to view the list of recommended aspects of datasets that should be documented.**

Open AI System

For Open AI System solutions, the following documentation is required:

▼ **Data**

This document provides extensive information about the dataset(s) utilized in the creation and implementation of the AI system. You can use any datasheet template, however, the following information must be provided in the datasheet for assessment:

Field Name	Description
Basic Information and Overview	Dataset name/identifier, version and date, creator/maintainer, use cases, and other general details.
Technical Details	Data provenance, data dictionary, data schema, unique identifiers, crosswalks to ontologies or vocabularies, data quality, and limitations.
Dataset Composition and Characteristics	Data instances, number of instances, data format, data fields/features, labels/target variables (if applicable), data splits (if applicable).
Data Collection and Preprocessing	Data sources, collection process, data cleaning and preprocessing steps, and data labelling.

> [!NOTE]
>
> You can submit the mandatory information mentioned above as part of any other document if it already exists in it (e.g., a model card). You should also consider including the following information (but they are optional):
>
> - Maintenance plan, update frequency, error reporting and handling, and support for contributions.

> - License and terms of use, distribution mechanism, data retention, and consent.

▼ Code

The documentation should include guides, technical specifications, functional requirements, etc., that would allow a technical person unfamiliar with the software components of the AI system to launch and run it. The documentation must show the following aspects (non-exhaustive list):

- How to install the software (local environments, testing, code runs etc).
- How to fork the software (forking, patching, contributing upstream and downstream).
- How to deploy the software as a user.
- Any additional context (both technical and non-technical) that could help a user or a developer navigate through the software.

▼ Click here to view list of common sections or types of software documentation.

- **Overview**: This briefly introduces what the software does, how it works, and who it is for.
- **Architectural Diagrams**: This shows the structure, components, and relationships of the software using visual diagrams and descriptions.
- **Technology Stack**: This lists the technologies and dependencies used in the software, as well as their versions and compatibility.
- **Installation Guide**: This explains how to install and run the software in different environments, such as local or production.
- **User Guide**: This teaches the end-users how to use the software and may include a FAQ section.
- **Release Notes**: This section follows [semantic versioning](#), and [records the changes and updates](#) for each version of the software.
- **Contributing Guide**: This section provides guidelines on contributing and participating in the software project.

> [!TIP]
>
> For all the examples mentioned above, you can explore these open source [documentation templates](#) from The Good Docs Project. The templates and accompanying guides will help you create quality documentation faster and easier.

▼ Model

This document accompanies an AI system, offering transparent reporting on its functionality, development process, and intended application. It is a vital resource for various stakeholders, providing comprehensive information about the model's capabilities and limitations. The fundamental objective of a model card is to foster transparency and accountability throughout the AI system's lifecycle by making essential details readily available. You can use any model card template, however, the following information must be provided in the model card for assessment:

Field Name	Description
Model Overview	Name, version, date, developer, description, and contact information.
Intended Use	Primary intended uses, intended users, and out-of-scope applications.
Performance Metrics	Key quantitative evaluation metrics, accuracy, precision, recall, and other relevant performance indicators.
Limitations	Known weaknesses, failure modes, and potential biases were identified.

> [!NOTE]
>
> You can submit the mandatory information mentioned above as part of any other document if it already exists in it (e.g., a datasheet). You should also consider including the following information (but they are optional):
>
> - Research paper, finetuned from, and other resources.
> - Details about the model's architecture and parameters.
> - Evaluation data, disaggregated performance metrics across various subgroups, and intersectional analyses.
> - Quantitative analyses, uncertainty estimates, confidence intervals, and model interpretability insights.
> - Information about the carbon footprint associated with training the model.

> [!TIP]
>
> Kindly find listed below example model card templates you can consider:
>
> - Model Cards for Model Reporting
> - Hugging Face Model Card Template
> - FAU's Master's Model Card Template
> - Google Model Card Toolkit
> - Model-cards-and-datasheets

▼ **Risk Assessment**

This document provides information on how risk is considered in the development of the AI system, fostering transparency and accountability. You can use any risk assessment template, however, the following information must be provided in the risk assessment template for assessment:

Field Name	Description
Proportionality	Impact on people and vulnerable groups, engagement with stakeholders, principles followed, etc.
Bias and Fairness	Steps to monitor, mitigate, and address biases, fairness assessment, model thresholds, etc.
Risks and Harms	Validation tests, misuse or unintended use, ethical considerations, guardrails, etc.
Mitigations	Accuracy evaluation, model validation and quality assurance, robustness and security, oversight and control, etc.
Transparency	Model explainability, logic, and decision-making, user information, tagging AI-generated content, etc.

> [!NOTE]
>
> You can submit the mandatory information mentioned above as part of any other document if it already exists in it (e.g., a datasheet, model card, etc.).

> [!TIP]
>
> Kindly find listed below example risk assessment templates you can consider:
>
> - AI Risk Assessment Template
> - Google Secure AI Framework (SAIF)
> - NIS AI Risk Management Framework (AI RMF)
> - Microsoft Responsible AI Impact Assessment Template
> - ISO 23894 - AI Guidance on Risk Management
> - TMC3 AI Impact Assessment Template

Open Content

For Open Content collections, this should include all relevant/compatible apps, software, or hardware required to access the content collection, as well as instructions regarding how to use it. A good way to provide evidence of this is to provide:

- A link to the section of your user guide that explains how users can access the content.
- A link where you state any technical requirements for accessing the content.

> 💡 **Tip**
>
> Here's a collection of [extra resources and helpful links](#) curated by the DPGA and the DPG community you can explore or contribute to.

[Digital Public Goods (DPGs)](#) are open-source software, open data, open AI systems, and open content collections that adhere to privacy and other applicable laws and best practices, do no harm, and help attain the Sustainable Development Goals (SDGs). If you have any questions regarding the DPG application process or anything else, you can ask directly to the [DPG Community](#) for guidance or [send us an email](#); we're available to help you.

▾ **Pages** 12

[Find a page...]

▸ **Home**

▸ **0. General Information**

▸ **1. SDG Relevance**

▸ **2. Open Licensing**

▸ **3. Clear Ownership**

▸ **4. Platform Independence**

▾ **5. Documentation**

      Jump to DPG Category ↓

    Open Software

    Open Data

    Open AI System

    Open Content

▸ **6. Mechanism for Extracting Data**

▸ **7. Privacy & Applicable Laws**

▸ **8. Open Standards & Best Practices**

▸ **9. Do No Harm by Design**

▸ **Contributing Guide**

Clone this wiki locally

https://github.com/DPGAlliance/dpg-resources.wiki.git ⧉