OECD.AI work promoting interoperability of AI risk management frameworks

IGF Policy Network on AI
Meeting #4

Celine Caira
Economist/Policy Analyst
18 July 2023
The OECD AI Principles (2019)

5 values-based principles for trustworthy, human-centric AI

- Benefit People & Planet
- Human rights, values & fairness
- Transparent & explainable
- Robust, secure & safe
- Accountable

5 principles for national policies, for AI ecosystems to benefit societies

- AI research & development
- Data, compute, technologies
- Policy & regulatory environment
- Jobs & skills, labour transitions
- International cooperation & measurement

https://oecd.ai/ai-principles
Jurisdictions are moving from PRINCIPLES to PRACTICE through various policies (laws, regulations, standards) and programs.

How can we ensure these initiatives are INTEROPERABLE with one another?
What does interoperability mean?  

Some examples..  

“…the ability of various privacy regimes, or legal frameworks, to work together to facilitate transborder data flows while ensuring the consistent protection of these data”

- OECD “Mapping commonalities in regulatory approaches to cross-border data transfers”, 2021

“…a pragmatic arrangement to promote policy coherence in the context of a shifting regulatory environment and multiple privacy frameworks and data regulations (e.g. data localisation requirements)”

- OECD Going Digital Toolkit Policy Note, 2021
OECD work promoting interoperability in AI governance

What is it? Defining an AI system
What does it do? Understanding the AI system lifecycle
How can it be used? Classifying AI systems
What are the risks? Mapping possible risks and incident reporting
Going back to basics: 
OECD AI System Definition (OECD, 2019)

“An AI system, is a machine-based system that is capable of influencing the environment by producing an output (recommendations, predictions or decisions) for a given set of objectives.

i) perceives environments through data or input;
ii) abstracts these perceptions into models;
iii) uses the models to formulate options for outcomes.”
AI system lifecycle stages

**ECONOMIC CONTEXT**
- Plan & design
- Operate & monitor
  - Actors include system operators

**DATA & INPUT**
- Collect & process data
  - Actors include data collectors & processors

**AI MODEL**
- Build & use model
- Verify & validate
  - Actors include developers & modellers

**PEOPLE & PLANET**
- Use or impacted by
  - Actors include end-users & stakeholders

**TASK & OUTPUT**
- Deploy
  - Actors include system integrators

**DATA & INPUT**
- Collect & process data

**AI MODEL**
- Build & use model

**PEOPLE & PLANET**
- Use or impacted by

**ECONOMIC CONTEXT**
- Plan & design
- Operate & monitor
Why classify AI systems?

A variety of systems and policy implications
OECD Framework for Classifying AI systems

Each AI framework dimension has its own properties and attributes…

…and involves specific actors
Linking the classification & AI system lifecycle actors

**Framework dimensions**
- People & Planet
- Economic Context
- Data & Input
- AI Model
- Task & Output

**Actors include**
- End-users & stakeholders
- System operators
- Data collectors & processors
- Developers & modellers
- System integrators

**Lifecycle stage**
- Use or are impact by
- Plan & design
- Operate & monitor
- Collect & process data
- Build & use
- Build & validate
- Deploy
Goal: enable international interoperability in AI risk management by identifying common guideposts

➢ **MAPPING** of existing and developing core standards, frameworks and guidelines for AI risk management to the top-level interoperability framework.

➢ **STOCKTAKING** of commonalities and differences in concepts and terminology between initiatives and conduct a gap analysis.

➢ **TRANSLATING** analysis into good practice to inform development of due diligence guidelines for responsible business conduct in AI.

➢ **ANALYSING** the alignment of certification schemes with OECD responsible business conduct (RBC) and AI standards.

➢ **DEVELOPING** an interactive online tool to help compare frameworks and navigate existing methods, tools and good practices for identifying, assessing, treating and governing AI risks.
# High-level mapping of select risk management frameworks to the Interoperability Framework

<table>
<thead>
<tr>
<th>OECD INTEROPERABILITY FRAMEWORK</th>
<th>GOVERN</th>
<th>DEFINE</th>
<th>ASSESS</th>
<th>TREAT</th>
</tr>
</thead>
<tbody>
<tr>
<td>OECD DDG</td>
<td>TRACK</td>
<td></td>
<td>IDENTIFY &amp; ASSESS</td>
<td>CEASE, PREVENT &amp; MITIGATE</td>
</tr>
<tr>
<td>ISOO 31000</td>
<td>MONITORING &amp; REVIEW</td>
<td>COMMUNICATION &amp; CONSULTATION</td>
<td>RECORDING &amp; REPORTING</td>
<td>LEADERSHIP &amp; COMMITMENT</td>
</tr>
<tr>
<td>NIST AI RMF</td>
<td></td>
<td>GOVERN</td>
<td>MAP</td>
<td>MEASURE</td>
</tr>
<tr>
<td>EU AIA</td>
<td>Post-market monitoring system and regular systemic updating</td>
<td>Communication of residual risks, accuracy, conformity, serious incidents</td>
<td>N/A</td>
<td>Documentation, record keeping, traceability</td>
</tr>
<tr>
<td>HUDERIA</td>
<td>Iterative requirements</td>
<td>N/A</td>
<td>Stakeholder Engagement Process (SEP)</td>
<td>N/A</td>
</tr>
<tr>
<td>IEEE 7000-21</td>
<td>N/A</td>
<td>Transparency management process</td>
<td>Ethical values elicitation and prioritisation</td>
<td>N/A</td>
</tr>
<tr>
<td>ISO/IEC Guide 51</td>
<td>Validation &amp; documentation</td>
<td>N/A</td>
<td>N/A</td>
<td>Validation &amp; documentation</td>
</tr>
</tbody>
</table>
High-level mapping of select risk management frameworks to the Interoperability Framework

- Key risk management frameworks generally aligned with the 4 top-level steps of the Interoperability Framework (Define | Assess | Treat | Govern).
- Most of the differences between frameworks relate to the GOVERN function.
Further reading
THANK YOU

celine.caira@oecd.org

FOLLOW US: @OECD-AI