Archival Values
and Artificial Intelligence

Presented by

Jim Suderman

to the

InterPARES Trust AI International Symposium
27 October 2022
Tinajo, Lanzarote, Spain

InterPARESTrust AI: Comparative Ethics study

Contributes to ITrustAI Project purposes by:

- Helping determine benefits and risks of using AI technologies
- Helping frame archival concepts and principles for appropriate inclusion in the development of AI systems

Study has two phases:

1. Comparative review of ethical principles of the two communities
2. Review of ethical issues identified in ITrustAI Project studies

PRINCIPLED ARTIFICIAL INTELLIGENCE:

Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI

Source: Jessica Fjeld, *et al*, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI, Berkman Klein Center, 2020

Evolving Archival Ethical Codes

Society of American Archivists

2005 Code of Ethics for Archivists: establishes standards for the archival profession.
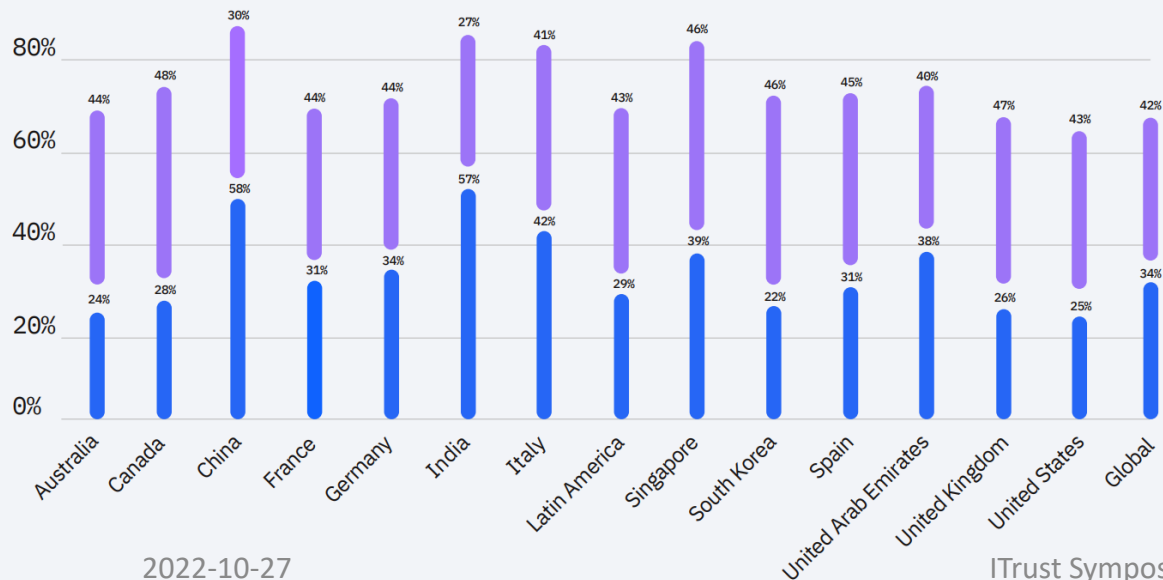
2020 Code of Ethics for Archivists: outlines behaviors and characteristics as aspirational principles to be considered in the creation of trusted archival organizations.

Michael Cook connects the development of ethical codes in the archival community in the 1990s with the emergence of technical standards for record keeping.

AI systems: rapid, pervasive adoption

- 35% (of 7,500) companies have adopted AI technology, up from 31% in 2021

- Key challenge: Trustworthy AI
  - Bias reduction
  - Tracking performance variation
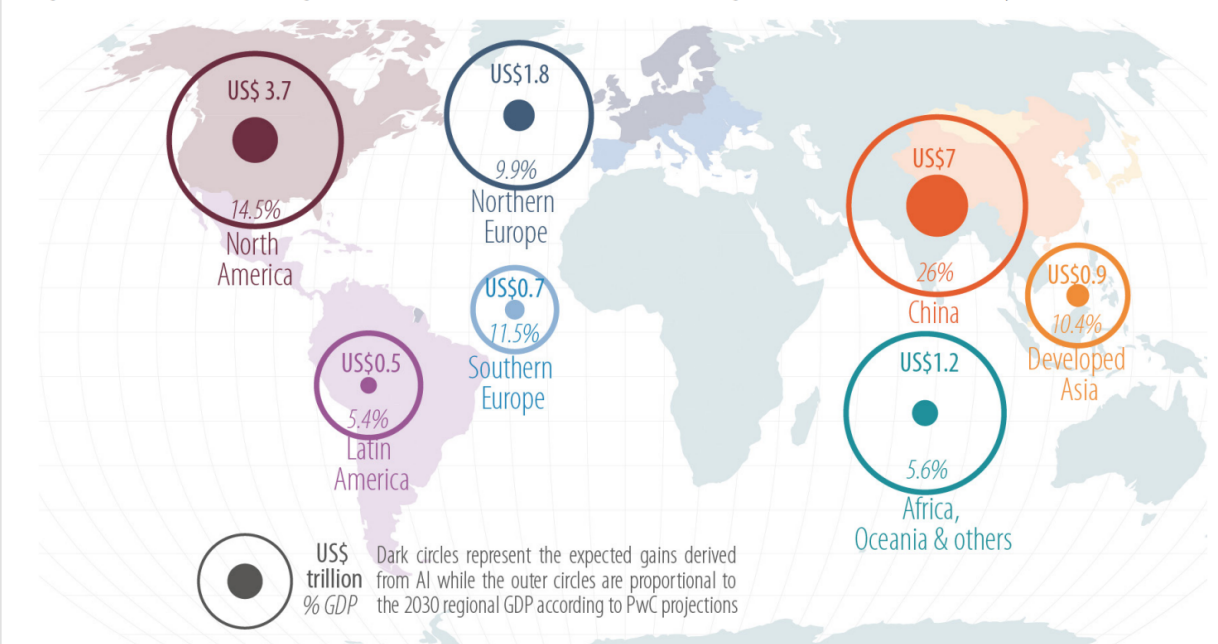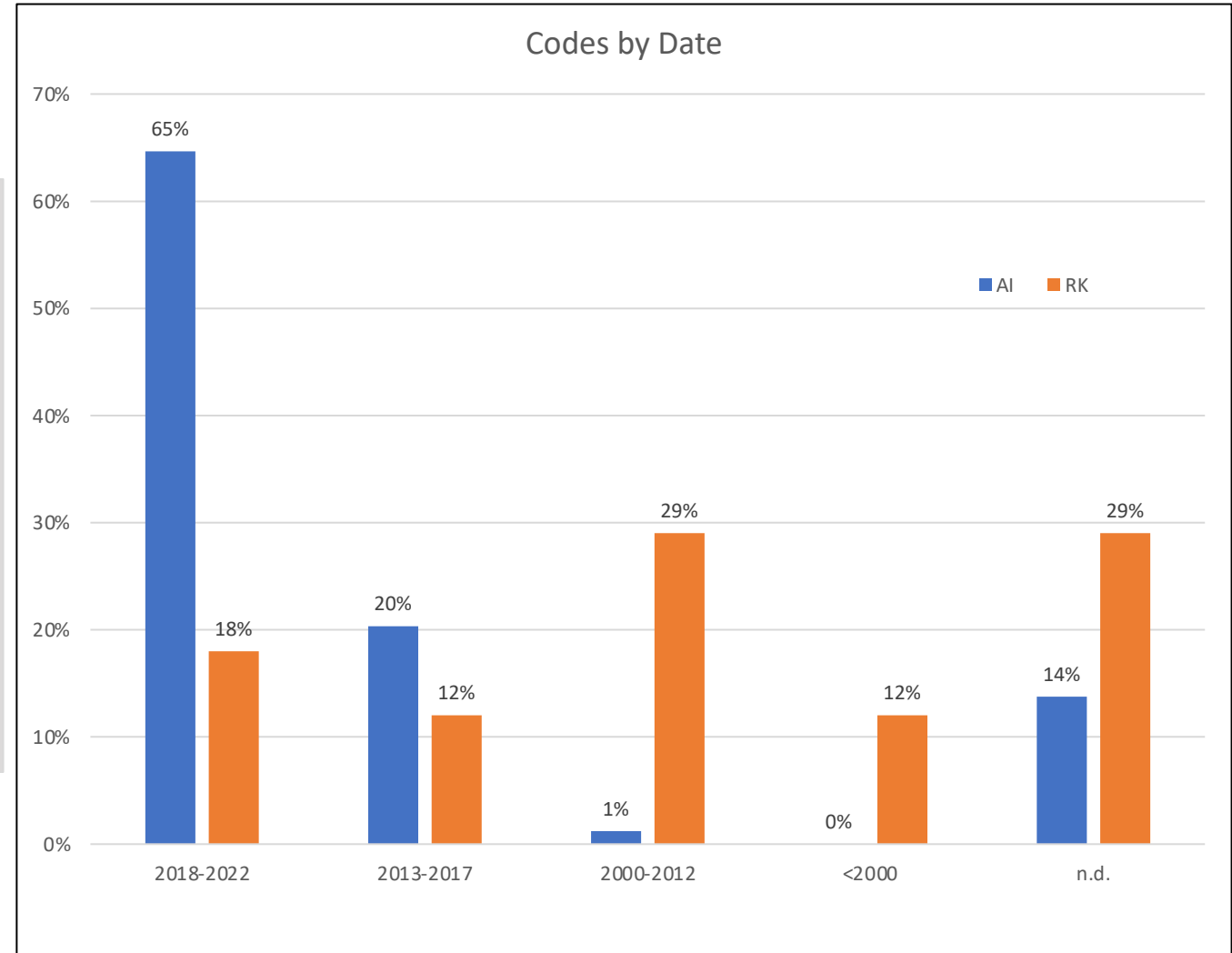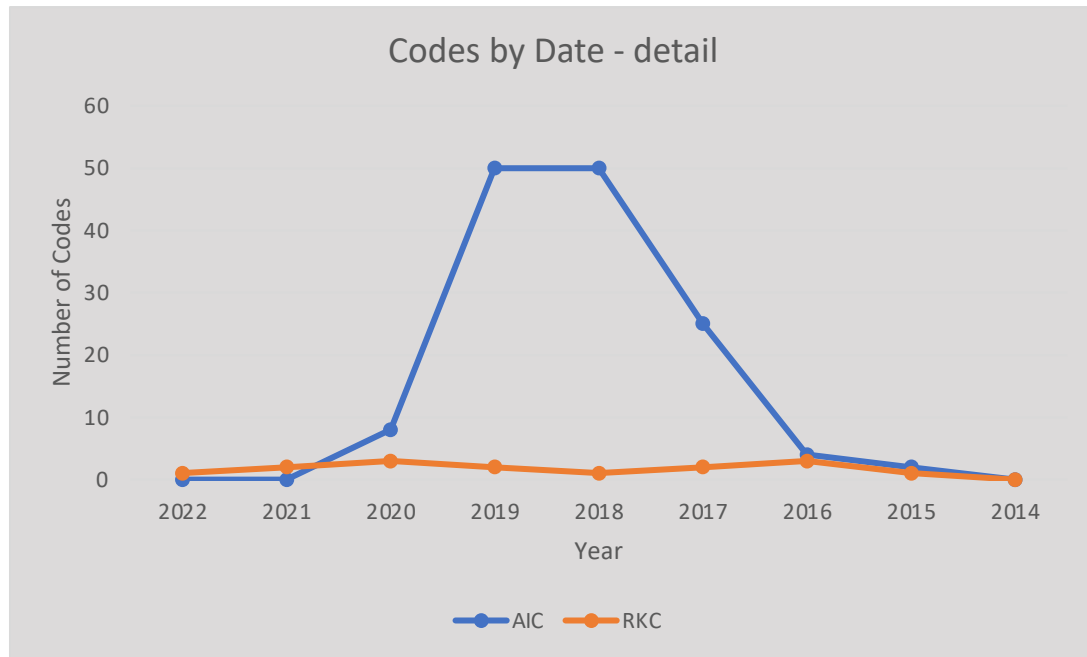 - Explainability of AI-powered decisions

Source: IBM Global Adoption Index 2022, p. 4

Figure 2 – Expected gains from AI in the different regions of the world by 2030

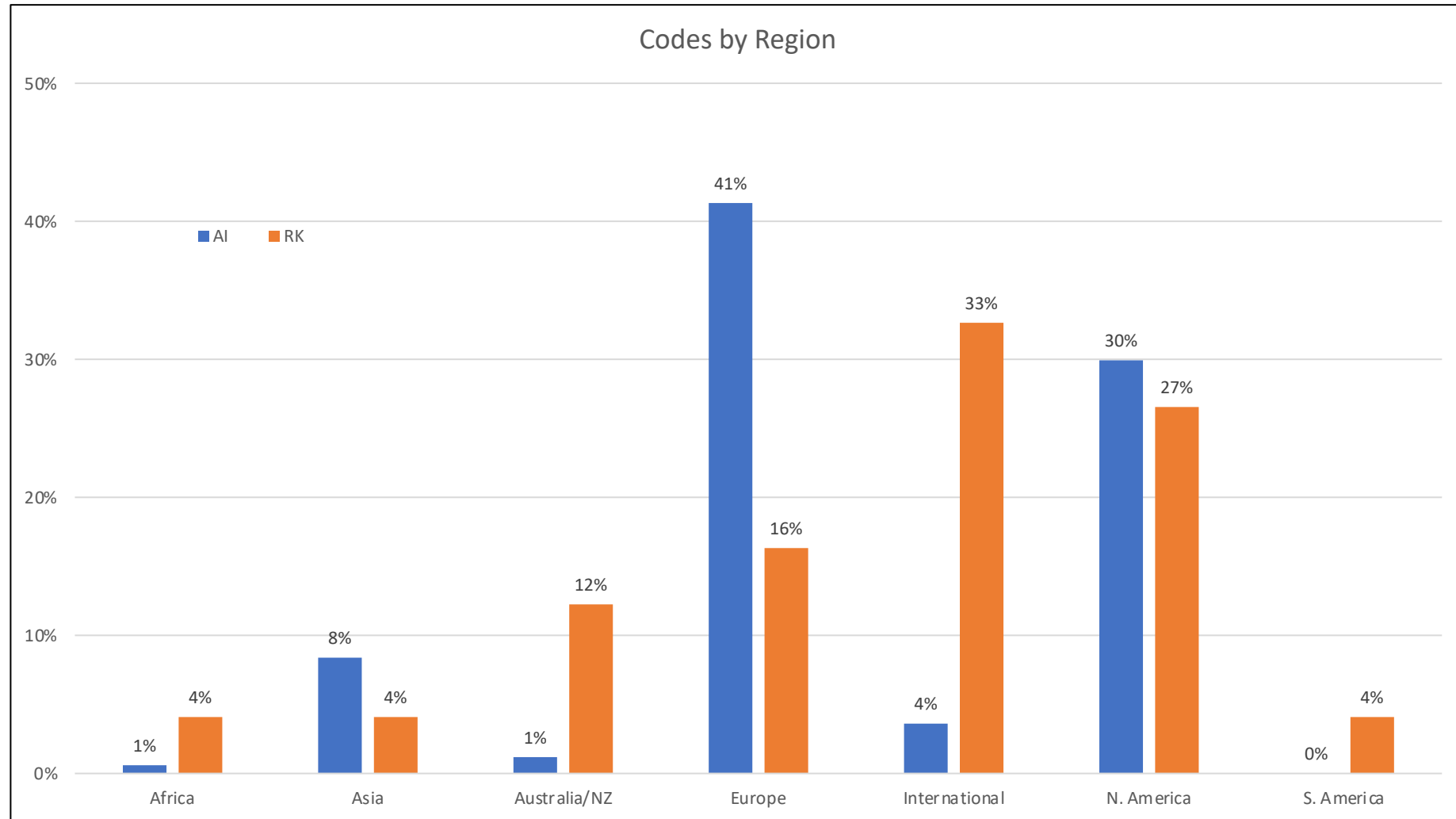Source: The macroeconomic impact of artificial intelligence, PwC, 2018.

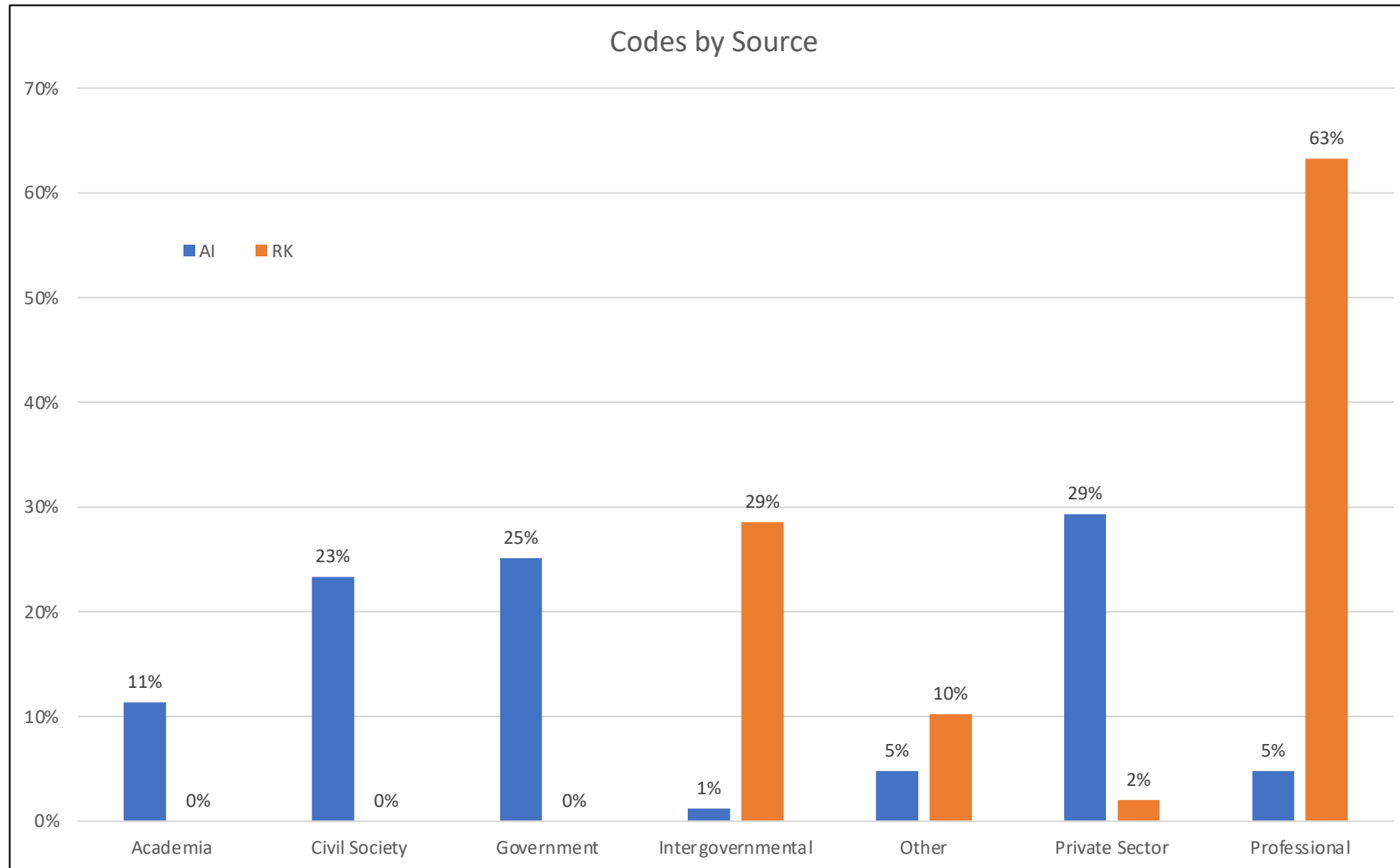# Ethical Codes by the Numbers: Date

Codes by Date - detail

Codes by Date

Ethical Codes by the Numbers: Region

Codes by Region

Ethical Codes by the Numbers: Source

Codes by Source

Ethical themes from the two communities

AI Themes (Fjeld study)

Accountability ———————————— Accountability

Safety and Security

Transparency and Explainability ———— Transparency

Fairness and Non-discrimination

Professional Responsibility ————— Professionalism

Human Control of Technology

Privacy ———————————————

Promotion of Human Values

RK Themes [working list – ITrust study]

Accountability

Preservation/Duty of Care

Transparency

Access

Professionalism

Social Benefit

Confidentiality

Advocacy

Collaboration

PRINCIPLED ARTIFICIAL INTELLIGENCE:
Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI
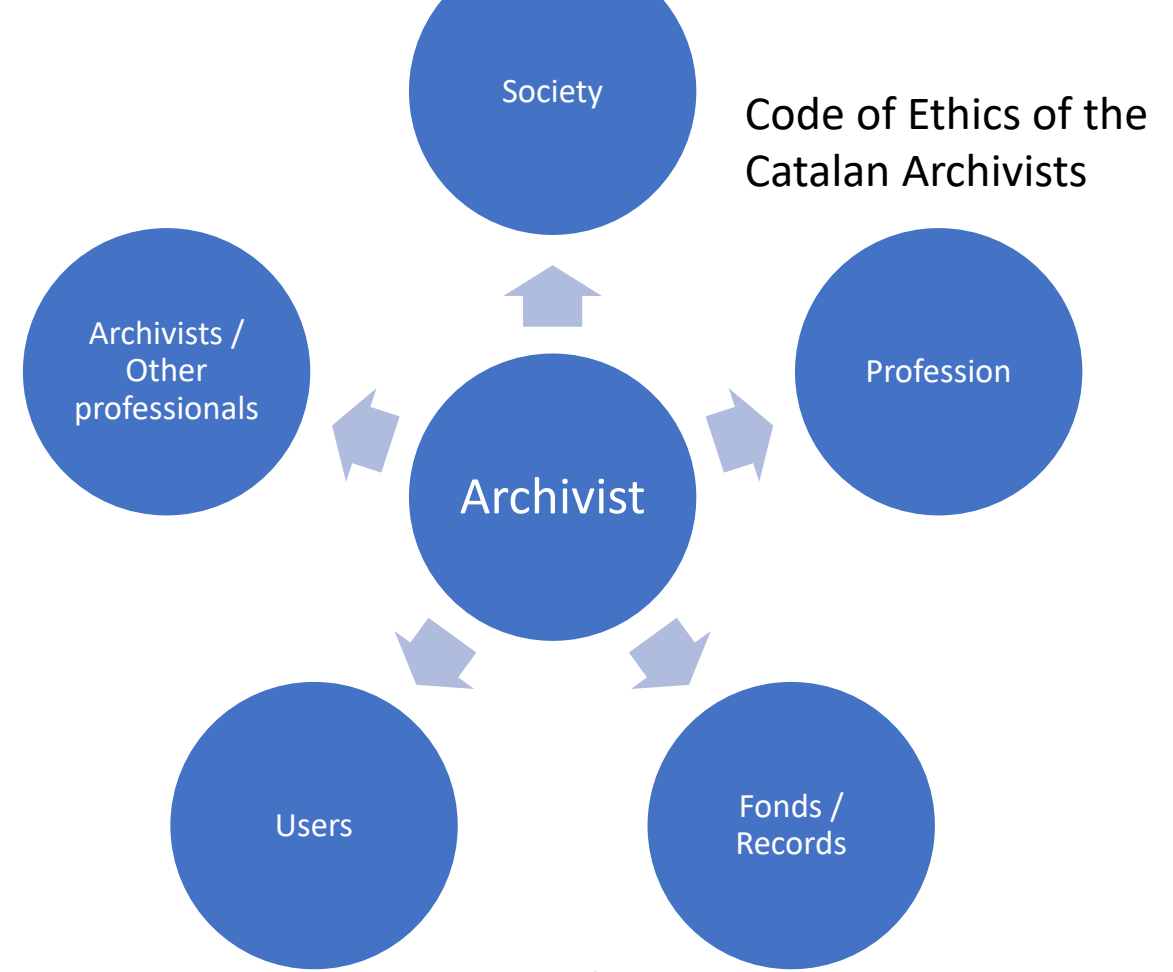
InterPARES TrustAI

Accountability Theme

Most common of 9 AI principles

- Impact Assessments

- Recommend new regulations

- Evaluation and Auditing

- Liability and Legal Responsibility

PRINCIPLED ARTIFICIAL INTELLIGENCE:
Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI

AI System Lifecycle

| Design | Monitoring | Redress |

Society

Code of Ethics of the Catalan Archivists

Archivists / Other professionals

Profession

Archivist

Users

Fonds / Records

InterPARES TrustAI

Accountability: stakeholders

AI		RKC
• **Users**	• Developers	• **Users**
• **Society**	• Implementers	• **Society**
• **Regions**	• Regulators	• **Regions**
• Marginalized groups	• [Beneficiaries]	• Professionals
• Human rights org'n	• States	• Public authorities
• Independent experts	• Independent auditors	• Profession
• Universities	• Internal review boards	• Institutions
• Research institutes	• Judicial authorities	• Individuals
• Industry		• Colleagues
• Policy makers		• Family

PRINCIPLED ARTIFICIAL INTELLIGENCE:

Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI
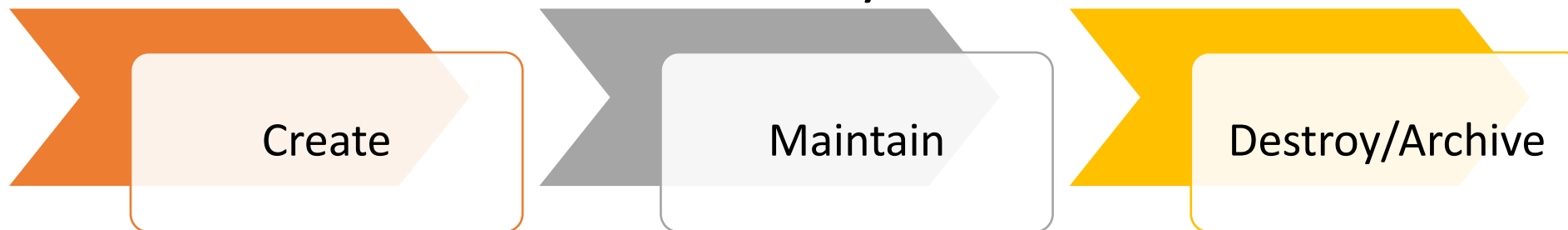
InterPARES
Trust AI

Accountability: AI Development Lifecycle

PRINCIPLED ARTIFICIAL INTELLIGENCE:

Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI

Design	Monitoring	Redress
Verifiability and Replicability	Evaluation and Auditing Requirements	Remedy for Automated Decision
Impact Assessment	Creation of a Monitoring Body	Liability and Legal Responsibility
Environmental Responsibility	Ability to Appeal	Recommends Adoption of New Regulations

Record Lifecycle

Create   Maintain   Destroy/Archive

InterPARES Trust AI

Preservation Theme

Principles

- Records
- Record collections
- Record values
- Record context
- Record provenance

Values

- Authenticity
  - Identity
  - Integrity
- Cultural
- Evidential
- Informational
- Historical
- Scientific

Open Data

Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness).

Concluding reflections

- Ethics: con dientes o sin dientes?

- Common commitment to accountability
    - Work in either field done badly can cause societal harm
  - Redress
    - AI principles may place a greater emphasis on redress
    - Archival principles may move in this direction due to emerging pressures
  - Impact Assessments
    - Practical process for assessing risk of harm

- How will codes in the AI field evolve after the "explosion" (2017-2019)?

Thank you!

Jim Suderman

InterPARES Trust AI Project