NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

Czech Academic and Research
Discovery Services
CARDS

Domain Specific Metadata Profiles

Petra Černohlávková, Jakub Klímek, Hana Vyčítalová, Kristýna Vojtíšková

National EOSC CZ Conference 2025: Repo builders workshop

3 December 2025, Ostrava

Co-funded by the European Union

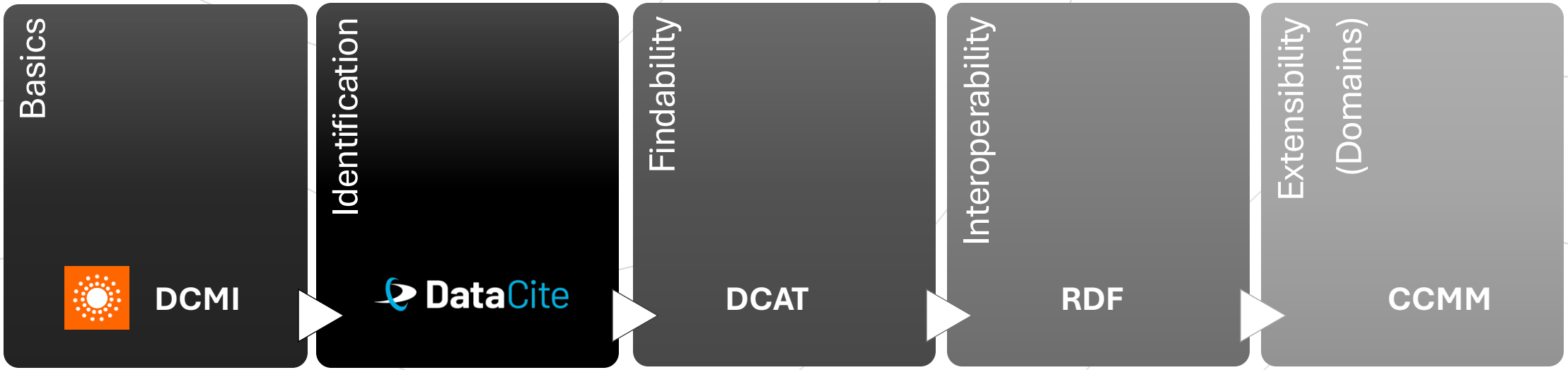
MINISTRY OF EDUCATION, YOUTH AND SPORTS

National Czech Programme | eosc

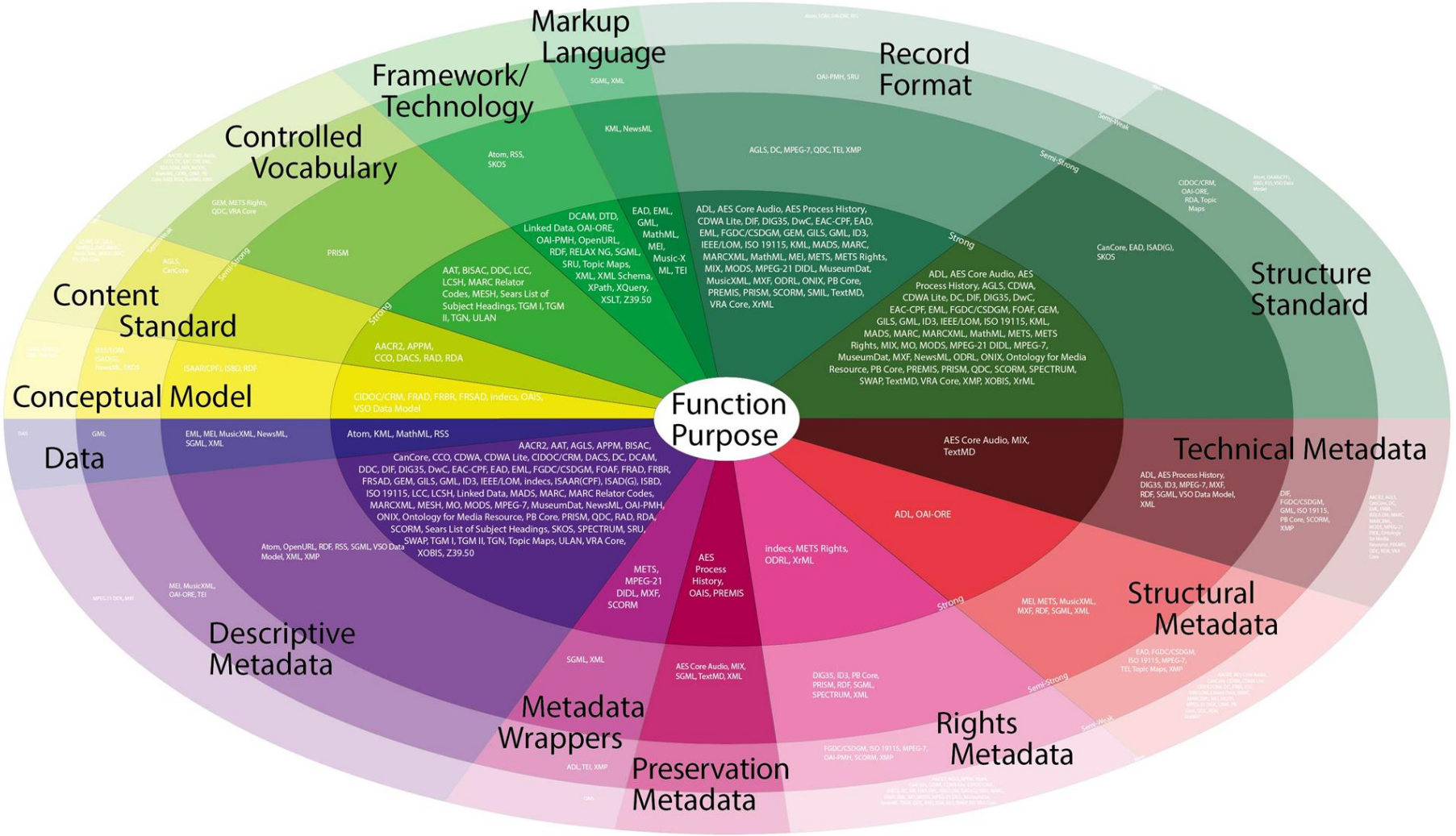
CCMM

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

- **Czech Core Metadata Model for Research Data (CCMM)**
  - universal schema for Czech research community
  - https://www.ccmm.cz/
  - recommended base for data desription in the Czech Republic
  - stems from / Aligns with DCAT-AP, DataCite, Dublic Core
  - ensures national interoperability across **NRP** and **NMA** infrastructures



More metadata



Source: http://jennriley.com/metadatamap/

What should the model do….

…fulfill your community needs & the FAIR principles

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4: (Meta)data are registered or indexed in a searchable resource

- A2. Metadata are accessible, even when the data are no longer available

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2.(Meta)data use vocabularies that follow FAIR principles
- I3.(Meta)data include qualified references to other (meta)data

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
- R1.1. (Meta)data are released with a clear and accessible data usage license
- R1.2.(Meta)data are associated with detailed provenance
- R1.3. (Meta)data meet domain-relevant community standards

What‘s the purpose of the model

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

- Data description and context
- Data identification
- Data management
- Searching
- Interoperability
- Archiving, preservation
- Re-usability
- Evaluation

Do you know what are your data?

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

Different objects might have different needs

Víno (láhev vína)

Název: Château Margaux 2018

Odrůda: Cabernet Sauvignon

Ročník: 2018

Obsah alkoholu: 13,5 %

Objem: 0,75 l

Region: Bordeaux, Francie

Certifikace: AOC (Appellation d'Origine Contrôlée)

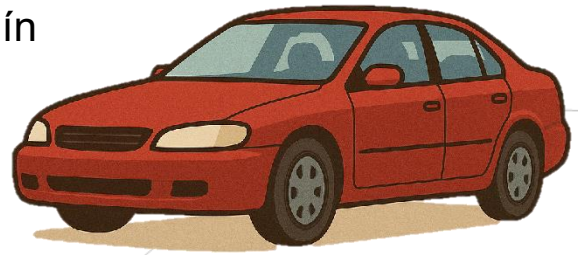
Chuťový profil: ovocné tóny, jemné taniny

Auto

Značka a model: Škoda Octavia Combi

Rok výroby: 2022

Typ motoru: 1.5 TSI benzín



Výkon: 110 kW

Barva: Modrá metalíza

Počet dveří: 5

VIN: unikátní identifikátor vozidla

Spotřeba: 6,2 l/100 km

Výbava: Navigace, adaptivní tempomat

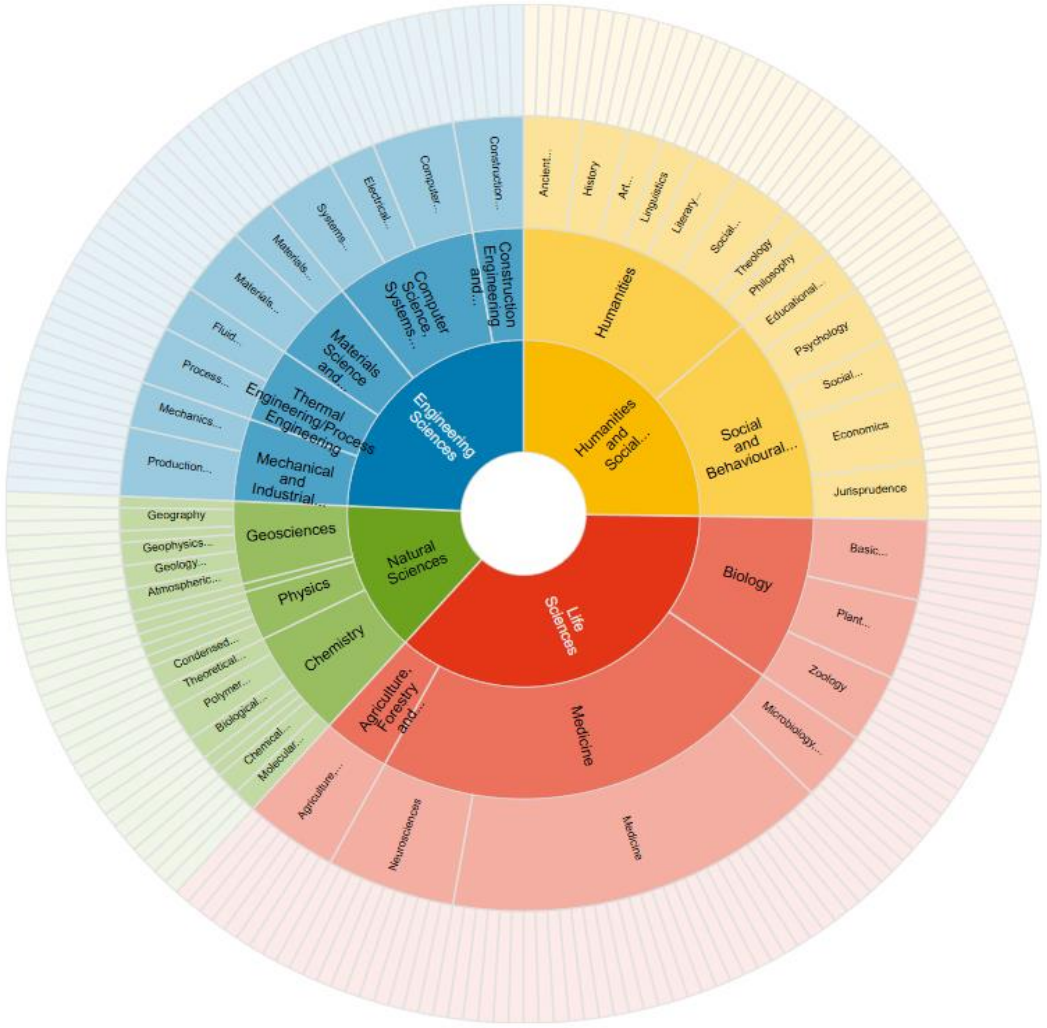
Source: AI-generated.

Understand the Domain Community

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

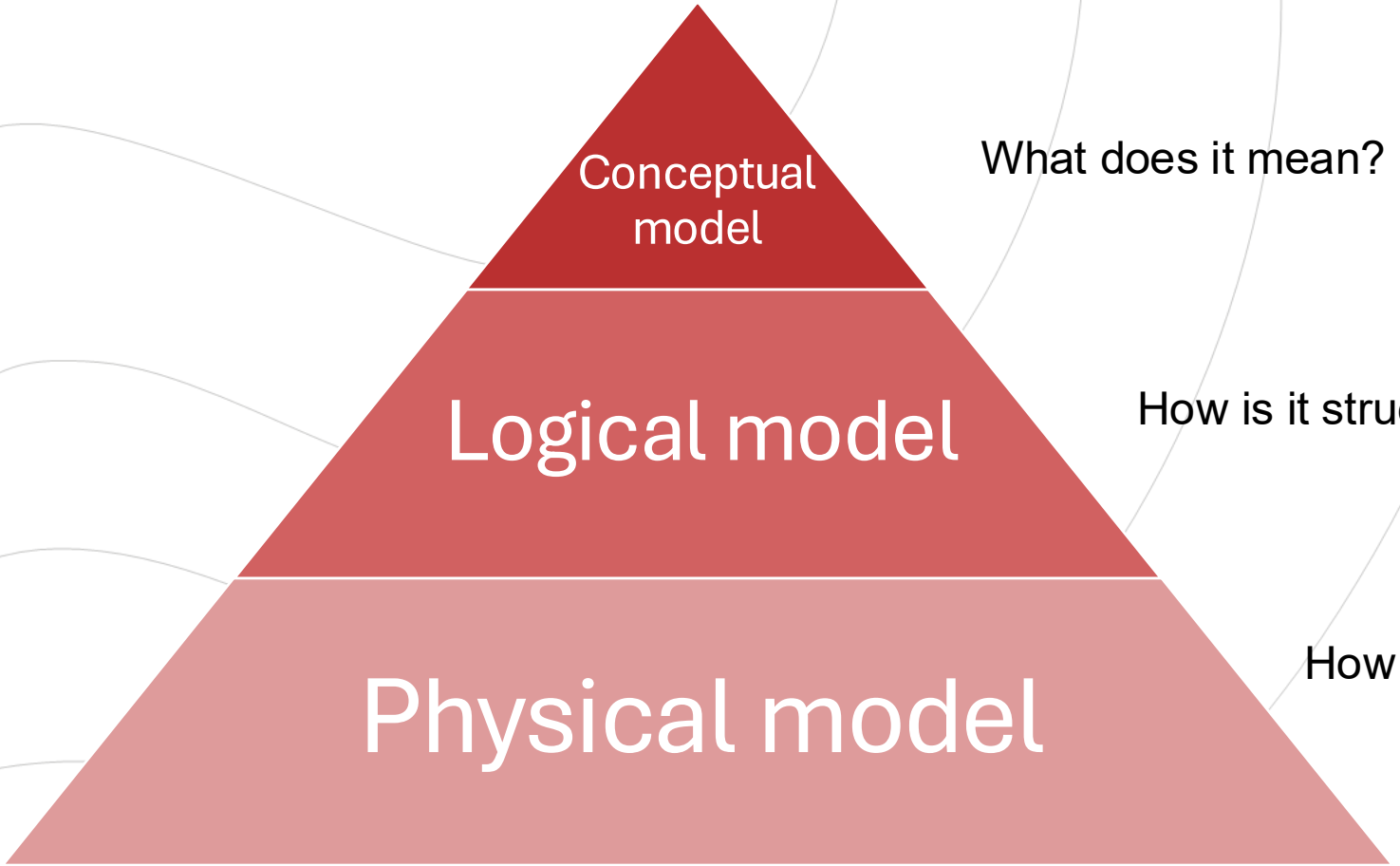
- What **disciplines** and **datatypes** are in scope?
- Who is the **target community**?

- Identify what the **community needs** for their data description
- Use interviews, domain-specific (meta)data examples
- Review existing repositories and their metadata models, e.g. via Re3data.org



Source: www.re3data.org/browse/by-subject

(Meta)data modelling



What does it mean?

UML diagram, RDF classes, properties, vocabularies

How is it structured?

XML Schema, RDF/OWL, JSON schema, logical ER

How is it implemented?

SQL table, Turtle, XML file, JSON(-LD), YAML

Roadmap of metadata

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

About to build repository:

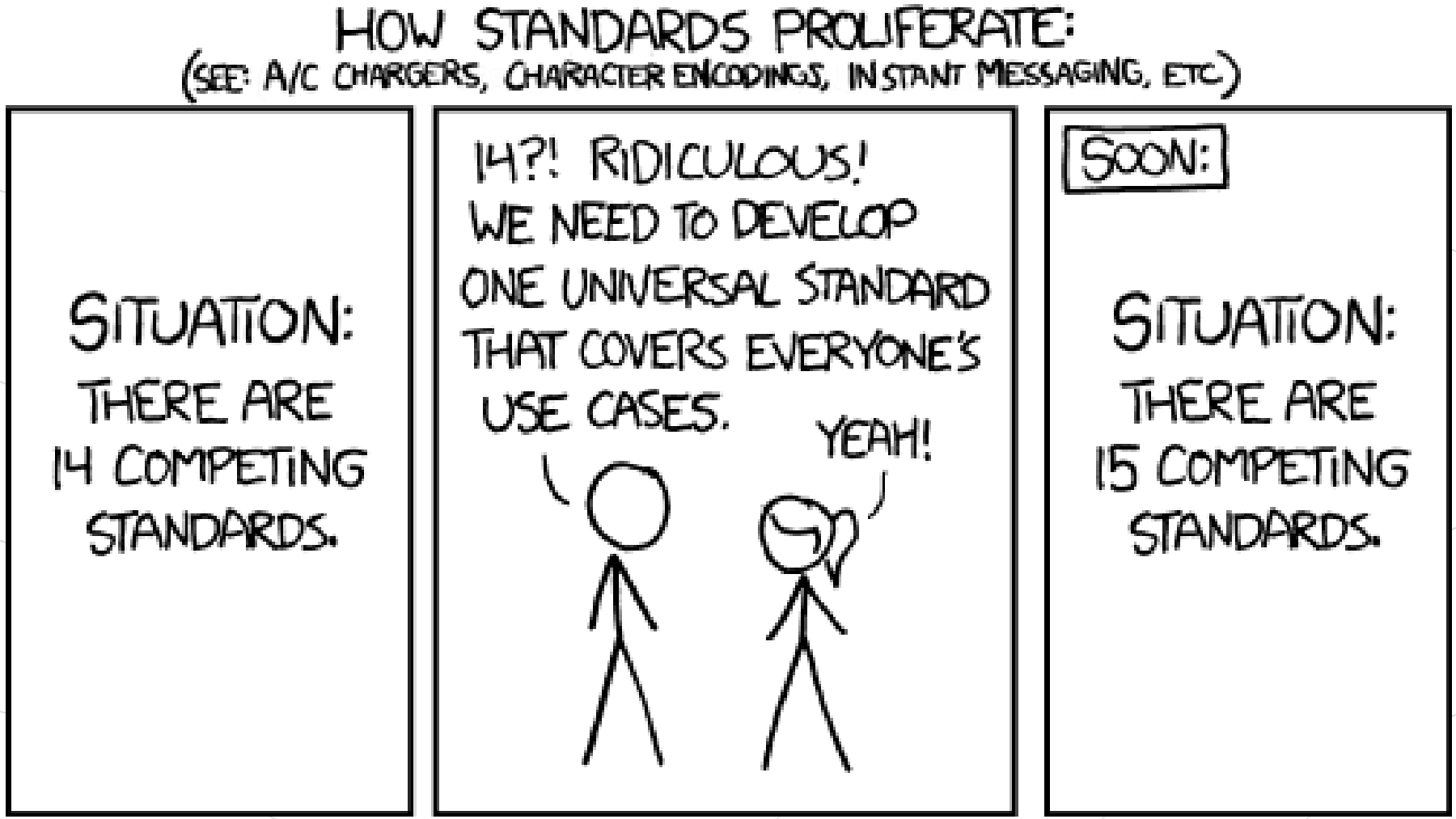
I need to define metadata model → Selection of metadata model → Implementation of model to the repository

Next:

Registration to the National Catalogue of Repositories (NKR) → Connection to aggregators i.e. NMA and OpenAIRE → Review of the model

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

Selection of the metadata model



Source: https://imgs.xkcd.com/comics/standards.png

What is the metadata/application profile?

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

- Selection of terms from other models (vocabularies and other profiles) to be used in a specific **context**

- Why? I need to share how the repository will work with metadata
- To whom? Human (community) and Machine

- Artefacts making the profile = That‘s my language to do so
  - **human** readable specification
  - Diagram of the profile
  - **machine** readable specification
  - Relations to other specifications and terms
  - Xml schema (other schema depending on the needs)
  - Exmaples of values
  - Controlled vocabs

§ 2.3 Agent

Profiled class IRI(s)	foaf:Agent
IRI	https://model.ccmm.cz/research-data/Agent
Label	Agent
Definition	Any entity carrying out actions with respect to the entities Catalogue and the Catalogued Resources.
Profiles	class profile Agent (https://mff-uk.github.io/specifications/dcat-ap#Agent) Definition: *Any entity carrying out actions with respect to the entities Catalogue and the Catalogued Resources.* class profile Agent (https://mff-uk.github.io/specifications/dcat-dap#Agent) class Agent (foaf:Agent)

KEEP ALL CONSISTENT!

Explore Existing Metadata Standards

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

Focus on:

1. Domain specific schemas
   - Full adoption
   - Adoption + editing, extending with domain elements

2. Universal schemas
   - Full adoption
   - Adoption + editing, extending with domain elements
     - CCMM
     - DataCite
     - DCAT-AP-CZ
     - DCAT-AP
     - Dublin Core

Explore Existing Metadata Standards

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

Sources to explore:

- Metadata Standards Catalog (available at https://rdamsc.bath.ac.uk/)
- Fairsharing.org (available at https://fairsharing.org/)
- Re3data.org (available at https://www.re3data.org/)
- MSCR Registry (available at https://mscr-release.2.rahtiapp.fi/)

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

FAIRsharing.org
standards, databases, policies

RDA
RESEARCH DATA ALLIANCE

Metadata „model“ analyses

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

Does it meet all your needs?

- Description granularity, rich metadata
- Community adoption/needs
- Interoperability with CCMM and other identified models/systems
- Crosswalk options
- SW technical options

Source: https://pixabay.com

YES	PARTIALLY	NO
• Reuse directly	• you found model to build your profile upon	• keep searching or extend general model

- Does it have the conceptual and logical layers?
- Does the license allow you to use it?

YES, fitting existing model

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

We choosed fitting existing model

YES	PARTIALLY	NO
• Reuse directly	• you found model to build your profile upon	• keep searching or extend general model

- Analyse this metadata model
- Revise:
  - Cardinalities
  - Formats
  - Controlled vocabularies
- If the format is suitable with repository software → ready for implementation
- If the format is **not** suitable with repository software → convert to right format

PARTIALY fitting existing model

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

**We choosed existing model,
but we need to do some minor extensions**

- Analyse selected metadata models
- Choose the main model
  - Focus on the choice of the elements from the analyzed models for extensions

YES	PARTIALLY	NO
• Reuse directly	• you found model to build your profile upon	• keep searching or extend general model

How can you do the extension?

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

Depends on what you have

Vocabulary of the existing model in RDF Schema

There isn´t a vocabulary in RDF Schema

YES	PARTIALLY	NO
• Reuse directly	• you found model to build your profile upon	• keep searching or extend general model

Use Dataspecer

Convert it to RDF Schema & use Dataspecer

Create the extensions manually or with a partial support

A tool for management of semantic data specifications

Dataspecer is a tool for management of semantic data specifications, from vocabularies and application profiles to derived technical artifacts such as data schemas, validation rules, APIs and application prototypes.

Read documentation | See specifications

Go to demo instance

Reusable vocabularies

Define new classes, properties and their relationships as well as their

Application profiles for narrower contexts

Combine terms from existing

Semantically compliant technical artifacts

Use your domain ontology or

NO, lack of domain specific models

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

- Extend CCMM with domain specific attributes

YES	PARTIALLY	NO
• Reuse directly	• you found model to build your profile upon	• keep searching or extend general model

- Use Dataspecer for creating the profile
- Use exports of CCMM on GitHub to start with
  - https://www.ccmm.cz/
- Protect your work by committing the progress on any git platform if working in a shared environment

Finalize the profile

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

- Export from Dataspecer or create profile manualy
- Model validation - with defined needs and community should be in place
- Publication: GitHub or similar platforms, www
  - Example: CCMM GitHub https://github.com/techlib/CCMM
- Define license

BUT…

Source: https://pixabay.com

Implementation to the repository

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

- Which **software** from the **NRP platform** do you use (Invenio, DSpace, ARL)?
  - Technical options (metadata format, hierarchy, vocabularies…)
  - Consultation with the NRP repository system specialists
- Affects physical model
- It may also affect logical model even though it shouldn‘t
- Testing and validation with defined needs and the community should be in place

Published your profile 1.0.0

→ Register to NKR

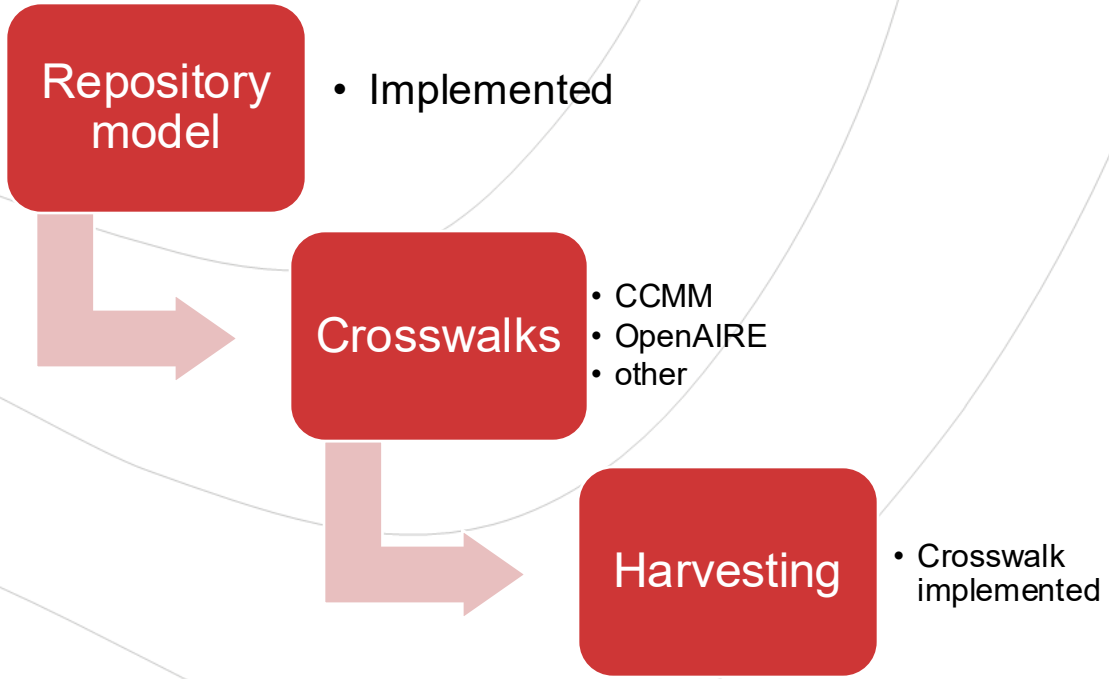
→ Verify interoperability – export your metadata

Source: https://pixabay.com

Interoperability in practice

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

- Map to CCMM for NMA indexing (aggregation and harvesting)
- Create and test export formats for OAI-PMH (XML) and other API
- Some crosswalks will be provided by metadata support



Mapping (crosswalks)

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

- Looking for existing mappings https://mscr-release.2.rahtiapp.fi/
  - Mappings from NTK will be created
    - CCMM x OpenAIRE
    - CCMM x RIV
    - CCMM x DataCite
- New crossswalks register to NKR that will push them to MSCR
  - https://mscr-release.2.rahtiapp.fi/

OpenAIRE

DataCite
CONNECTING RESEARCH,
ADVANCING KNOWLEDGE

Maintenance, Versioning, and Documentation

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

- Community feedback
- Keep version control (e.g. SemVer „MAJOR.MINOR.PATCH.“ 1.0.0., 1.1.0)
- Document changes and rationale (e.g. using Github or other systems)
- Plan for regular updates (and publish information)
  - Watch the changes and new versions in used schemas!

You are not alone

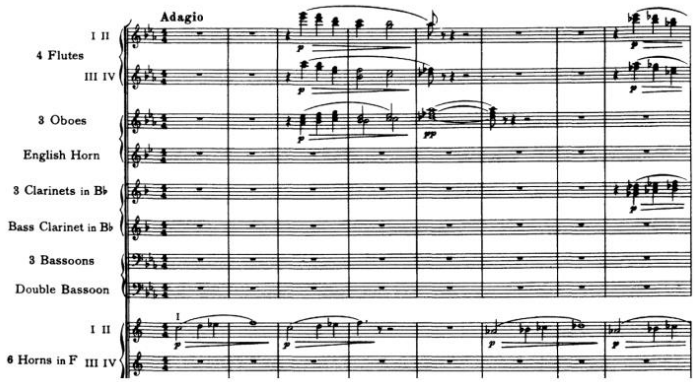
NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

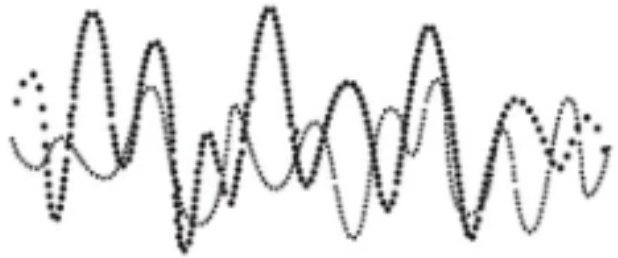
- **Consultations**
  - managers of existing repositories
  - NRP repository system specialists
  - NTK metadata department (metadata@techlib.cz)
  - Repository specialist of the used SW

Source: https://pixabay.com

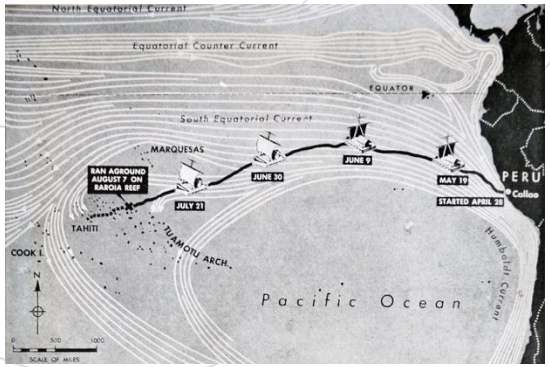
Goal Seeking balance

- Detailed but still possible to fill
- Vocabularies – help or limit
- What have to part of the model, when reference is good enough
- Interoperability with disciplinary precision

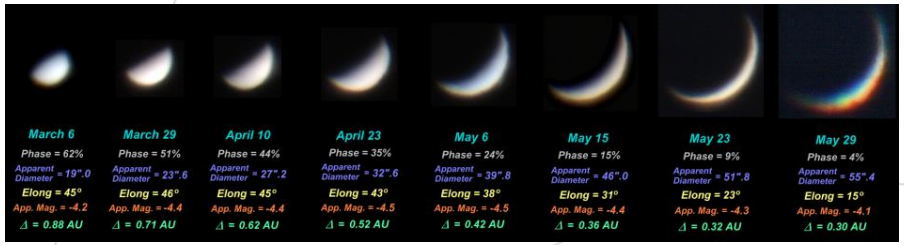


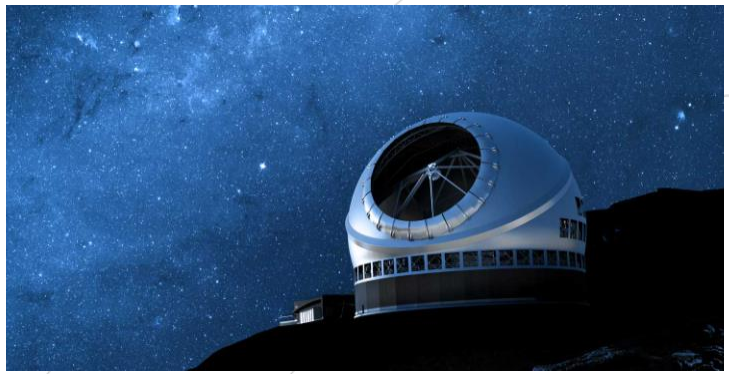


Gustav Holst, Opus 32, 2nd Movement
Source: Sydney Symphony Orchestra, 2023

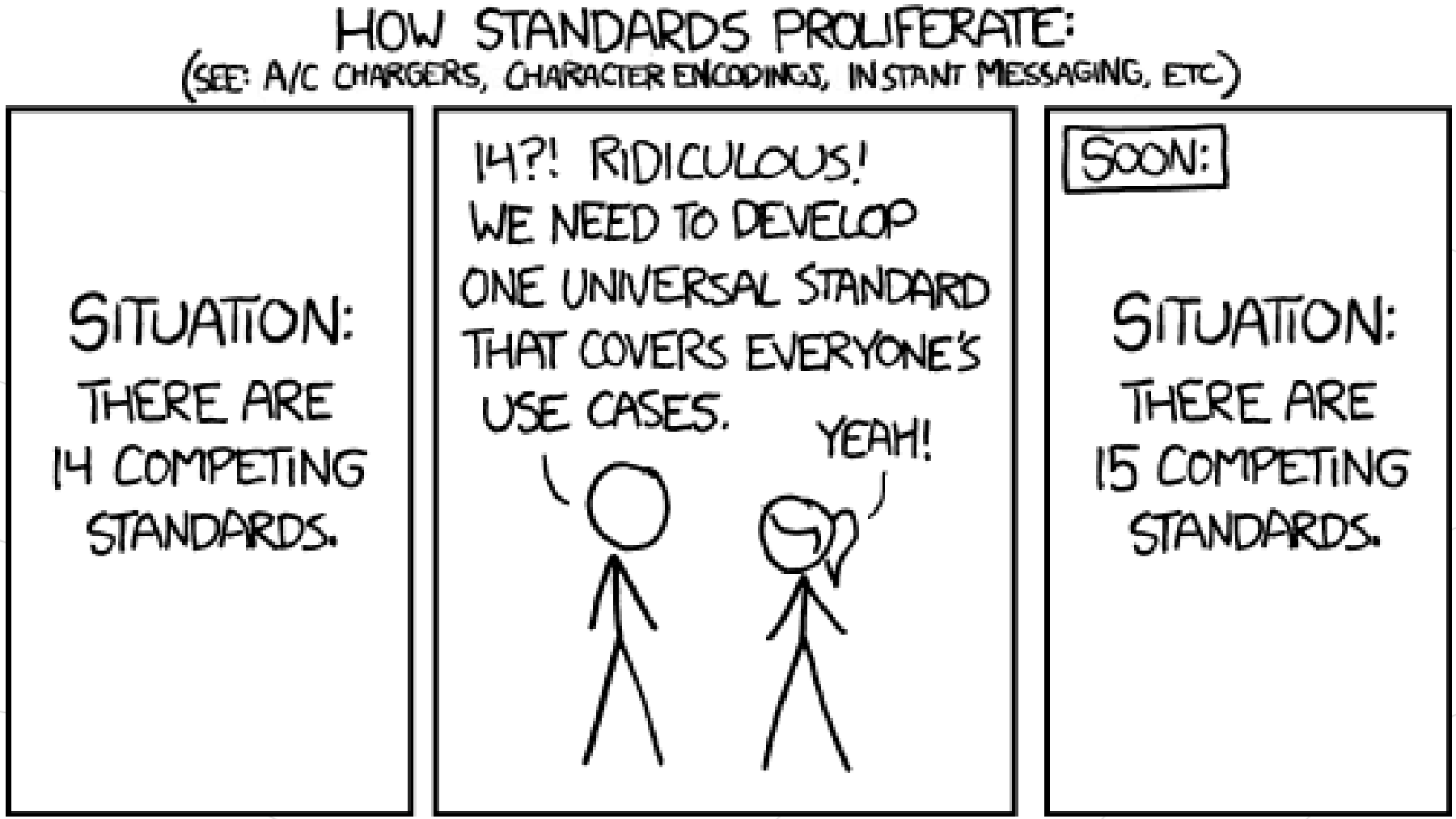


Kon Tiki Expedition and Polynesian Navigation,
Source: Kontiki Museum in Oslo, 2016





TMT Laser Guide Star Facility, Hawai'i
Source: www.tmt.org, 2025



Source: https://imgs.xkcd.com/comics/standards.png

Resources and Support

NTK
50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

- **NTK Metadata Unit:** metadata@techlib.cz
- **CCMM documentation** and **issue submissions:** https://github.com/techlib/CCMM
- **Dataspecer:** https://dataspecer.com/

www.ccmm.cz/en/

https://techlib.github.io/CCMM/en/

Thank you!