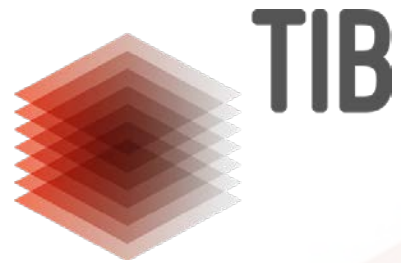

LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY



A Practical Case Study About Metadata

M. Lindlar

Online dpc Briefing Day, 30th April 2020

“Building a Digital Future : Challenges & Solutions for
Preserving 3D Models”

What I'd like to talk about today

Setting the Scope

- Who is TIB
- What is the use case for 3D?
- Information types and file formats

Some Definitions

- Preservation metadata / PREMIS
- Technical metadata

Preservation Metadata for Architectural 3D

- Descriptive metadata – *buildm*
- Technical metadata – *e57m* and *ifcm*
- What can I use techMD for?

Conclusion and Outlook

Who is TIB?

- German National Subject Library for **Science** and **Technology**, **Architecture**, **Chemistry**, **Computer Science**, **Mathematics** and **Physics**
- University Library of the Leibniz University Hannover



- Founded 1959
- Annual budget (incl. project funding) 47 Mio. Euro
- Staff size: 536

- Holdings: 9.2 Mio media units
- 56.750 journal titles (42.900 electronic)
→ 60 % in national sole possession
- 76.7 Mio electronic docs (19.3 Mio pay-per-view)
- 209 km shelves
- Competence centre for non-textual materials



DURAARK project – Architectural 3D Data



DURAARK (DURABLE Architectural Knowledge)

FP7 – ICT – Digital Preservation (STReP)

February 2013 – January 2016



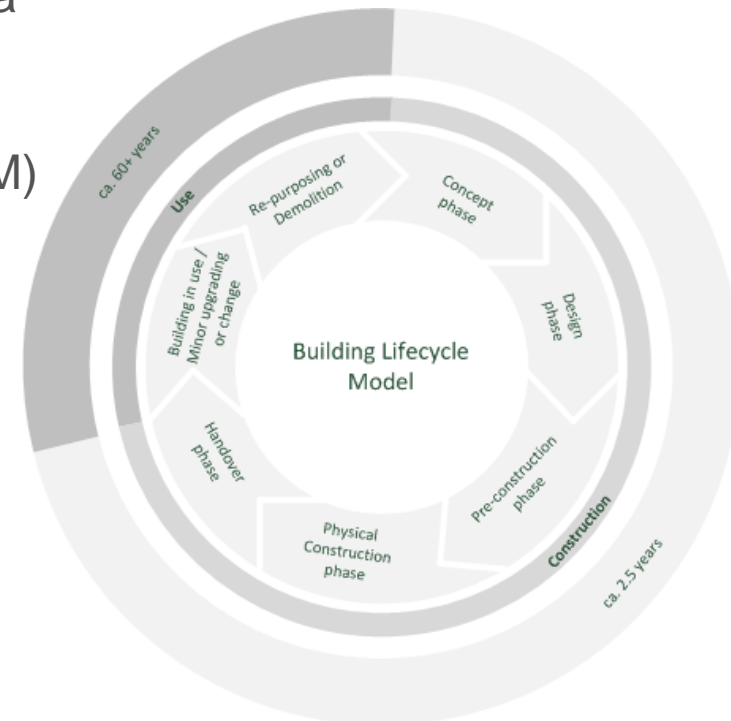
Goal: Develop methods and tools for digital curation and preservation of 3D building data (BIM, point cloud scans), metadata, related knowledge & web data

TIB Use Case:

Preservation of Building Information Models (BIM) & point cloud scans for

- Facility Maintenance
- Cultural Heritage

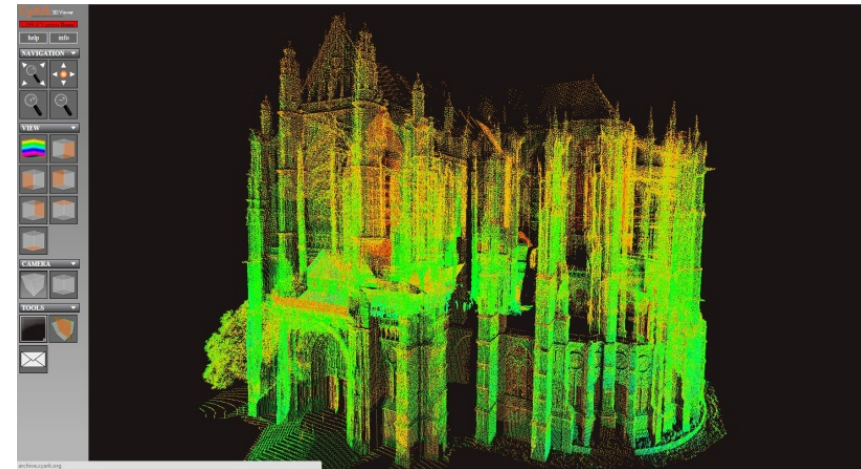
Results were incorporated into OAIS compliant digital preservation system (TIB Digital Archive)



Point Cloud Scans and e57

Point Clouds

- Set of points in a 3D coordination system
- Describe external surfaces of a scanned object
- Document a building or structure „as-is“
- Are inevitably tied to temporal and spatial aspects



<http://archive.cyark.org/exterior-cathedral-of-beauvais-3dviewer>



ScanCopter by FaroLabs

E57

- Openly standardized file format (E57 – ASTM E2907-11 Standard)
- Supported by many scanner & software vendors
- Open reference implementation of supporting software available (libE57) www.libe57.org

Building Information Models and IFC

BIM

- Covers entire design-to-construction process (incl. project planning, cost, part specifications, ...)
- Documents a building / structure „as-planned“ / „prescriptive representation“ → may deviate from the as-is-state
- Moves towards „as-is“ state for facility maintenance

IFC – Industry Foundation Classes

- ISO Standard within an ISO Standard
- STEP ISO 10303 application profile
- IFC – ISO16739-1:2018
- As standardized data exchange format supported by most software vendors for CAD

3D CAD

Geometry along X-Y-Z axes



4D CAD

Schedule time



5D CAD

Cost-related information



6D CAD

Energy and sustainability



7D CAD

Facility management

What is Preservation Metadata?

Preservation Metadata

NOUN UNCOUNTABLE /prezə(r)'veɪf(ə)n/ metə'deɪtə/

Information required to ensure the long-term usability of a digital object.

What is it?

Who created it?

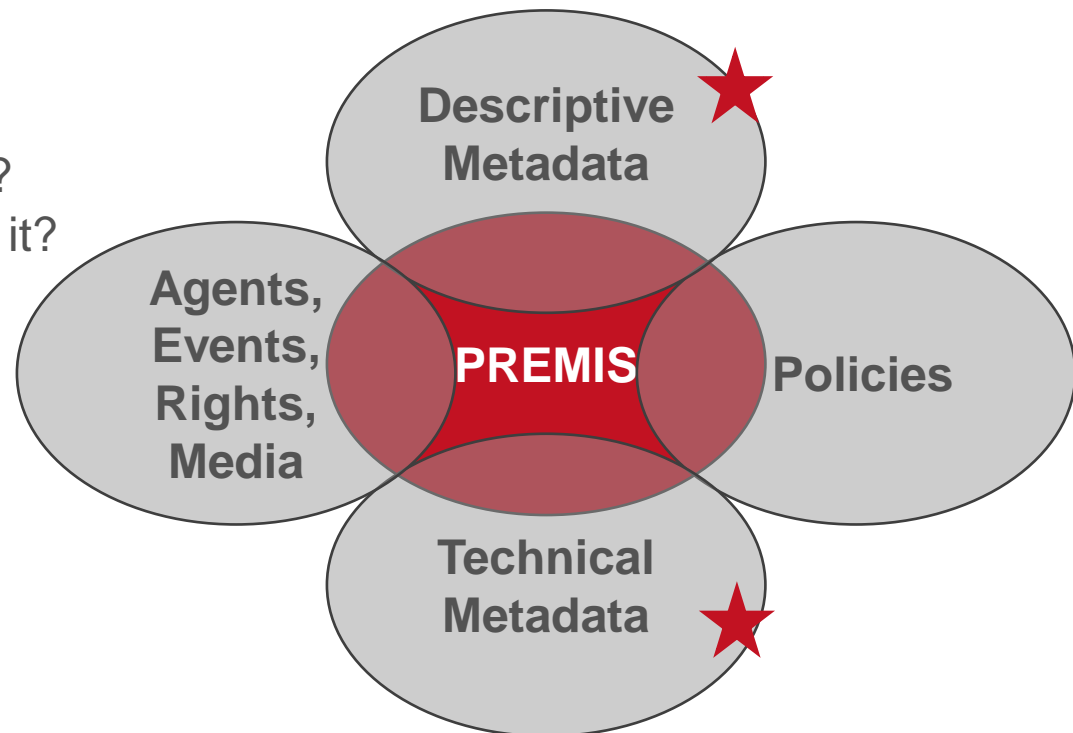
No, what is it ... really?

What am I allowed to do with it?

What am I supposed to do with it?

What did you do with it !?

Who should be able to understand it?



What is PREMIS?

PREMIS

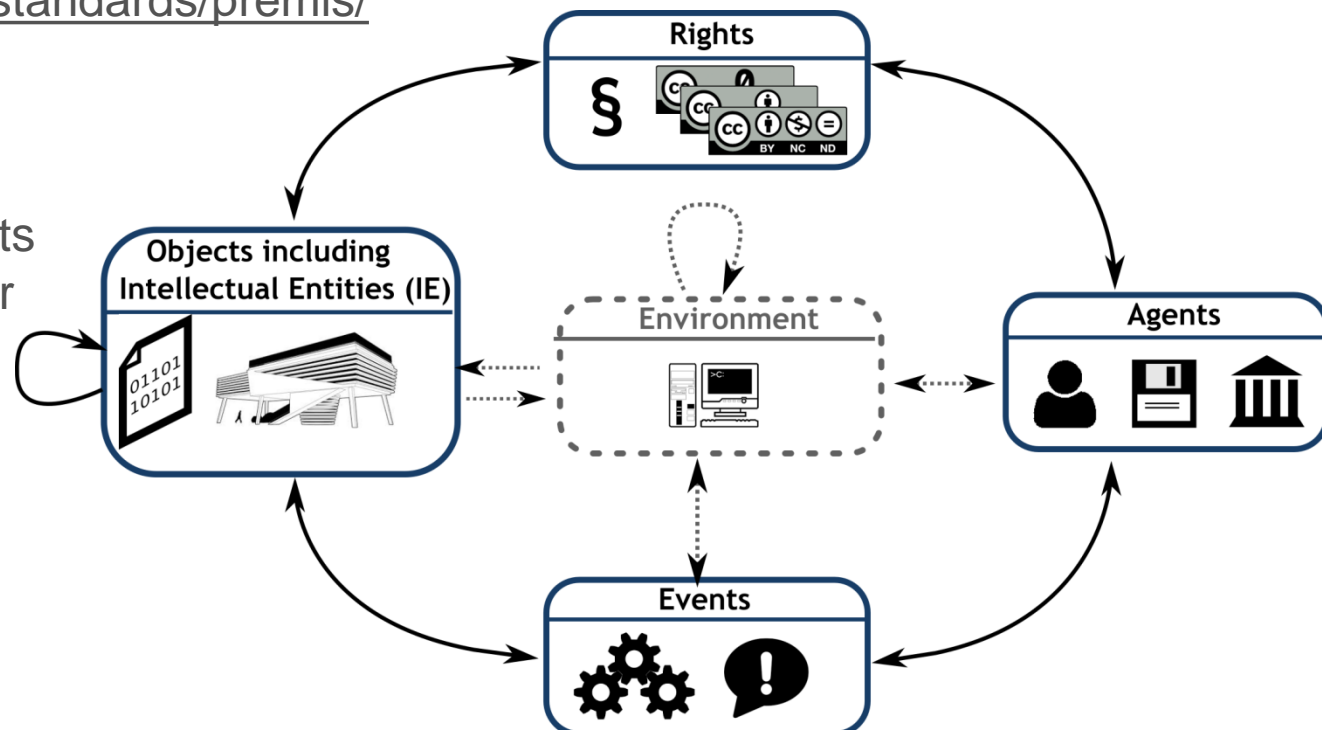
NOUN UNCOUNTABLE /'premis/

De facto standard for *preservation metadata*. Implementation agnostic data dictionary with accompanying documentation, an XML schema and an OWL ontology.

<https://www.loc.gov/standards/premis/>

Main Entities:

- Objects
(IE & Environments are objects in their own right)
- Rights
- Agents
- Events



What is technical metadata (techMD)?

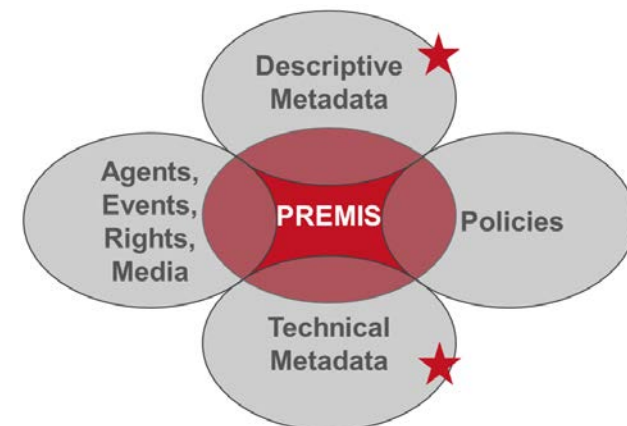
Technical Metadata

NOUN UNCOUNTABLE /'teknɪk(ə)l/ metəˌdeɪtə/

Physical (rather than intellectual) characteristics of digital object. Closely tied to file format.

Technical Metadata in PREMIS?

- recognizes importance of techMD
- recognizes that techMD specification requires expertise
- includes extensibility mechanism incl. semantic unit
objectCharacteristicsExtension to use external techMD schemas



Examples for other techMD schemas

TechMD Schemas

- MIX „NISO Metadata Images in XML“
- audioMD and videoMD
- textMD



Other sources

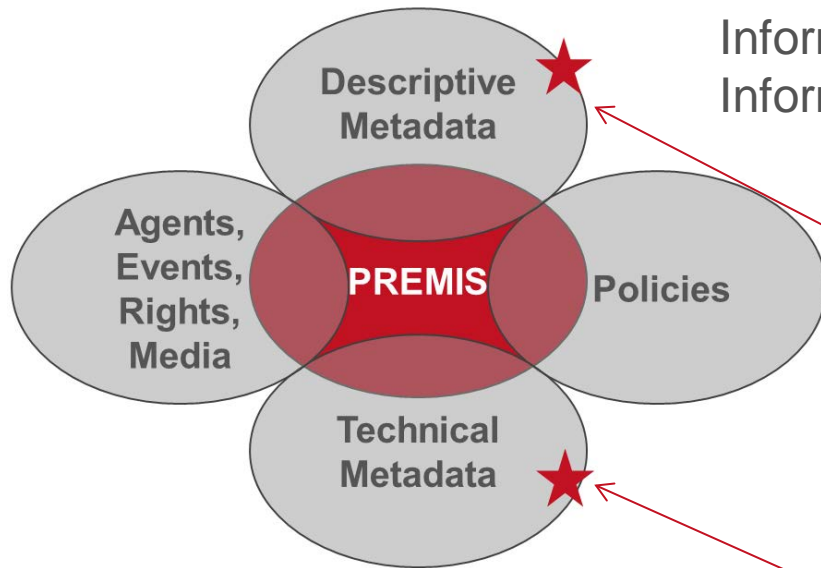
- Tool-based output (e.g., JHOVE for pdf)
- Standards-based output (e.g., TIFF Tags)



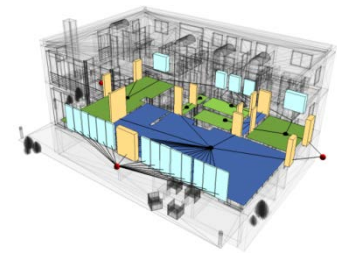
Sample output (Excerpt of NisoImageMetadata for TIFF)

```
ByteOrder: little_endian
CompressionScheme: uncompressed
ImageWidth: 2961
ImageHeight: 4746
ColorSpace: RGB
ICCProfileName: sRGB IEC61966-2.1
ReferenceBlackWhite: 0, 255, 0, 255, 0, 255
DateTimeCreated: 2013-08-31T11:54:01
ScanningSoftware: SRZ ProScan V3.3
Orientation: normal
SamplingFrequencyUnit: inch
```

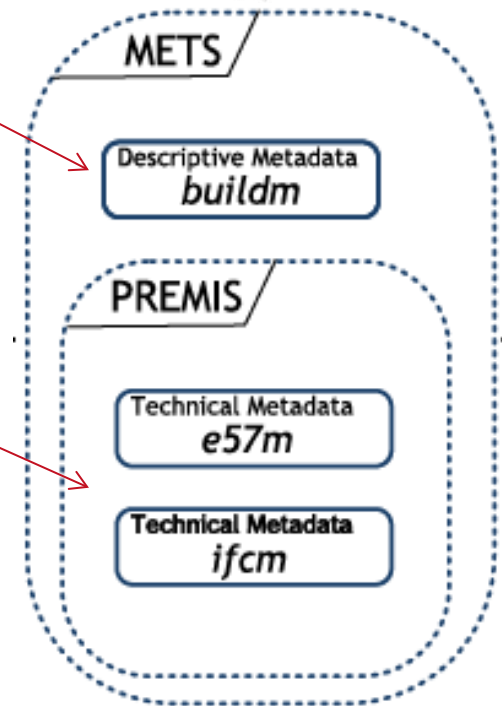
Metadata about architectural 3D objects



Information about the analog object
 Information about the digital object



File-format specific metadata;
 Different attributes for BIM objects („plans“) and point-cloud scans



Descriptive Metadata – buildm

Schema

<https://github.com/DURAARK/Schemas>
(xsd and rdf serializations)

Data Dictionary

<https://doi.org/10.5281/zenodo.1115511> p. 45 – 58

Based on

- CARARE v2.0
- MIT Facade PIM
- Rec. of Historic Buildings & Monuments Commission for England
- PROBADO3D MD Core
- Dublin Core

See Deliverable for crosswalk

buildm	3.1 Draft	Historic Build-ings ⁵⁵	MIT FACADE PIM ⁵⁶	PROBADO3D meta-data core	Dublin Core ⁵⁷	CARARE
DigitalObject:creator			dcterms:creator	CONTRIBUTOR	dc:creator (digital)	Digital Resource: Actors
DigitalObject:filename		file name of raw data	dcterms:title	TITLE / MODELFILE	dc:title (digital)	Digital Resource: Appellation
DigitalObject:dateCreated		date of capture	dcterms:created	DATES	dc:date (digital)	Digital Resource: Created
DigitalObject:isPartOf			dcterms:isPartOf	RELATION		Digital Resource: Is Part Of
DigitalObject:hasPart					RELATION	Digital Resource: Has Part
DigitalObject:Description				MODELDESCRIPTION	dc:description (digital)	Digital Resource: Description

buildm Entities: Data Object and Physical Asset

buildm – descriptive metadata for architectural
3D objects

Data Object section

*Information about the digital object
(scan, plan or any other digital object
representing the Physical Asset)*

mandatory information

(identifier, creator, filename, date created)

optional information

(e.g., license, relation to other
digital objects)

Physical Asset section

*Information about the physical
structure, e.g. address, architect,
construction year*

mandatory information

(identifier, name, latitude, longitude)

optional information

(e.g., architect, modification date function,
description)

buildm – minimal set of information

Mandatory – PhysicalAsset

- PhysicalAsset:identifier
- PhysicalAsset:name
- PhysicalAsset:latitude
- PhysicalAsset:longitude

Mandatory – DigitalObject

- DigitalObject:identifier
- DigitalObject:creator
- DigitalObject:name
- DigitalObject:dateCreated

- minimal descriptive information required to manage digital assets over long term
- Specific to the use cases:
 - architectural 3D data
 - facility maintenance
 - historic building data
- partially extractable from digital objects



*... schema can
be easily adapted to
YOUR use case*

Technical Metadata – e57m and ifcm

Schemas

<https://github.com/DURAARK/Schemas>
(xsd and rdf serializations)

Data Dictionaries

<https://doi.org/10.5281/zenodo.1115511>
e57m – p. 60 - 88
Ifcm – p. 88 - 98

Based on

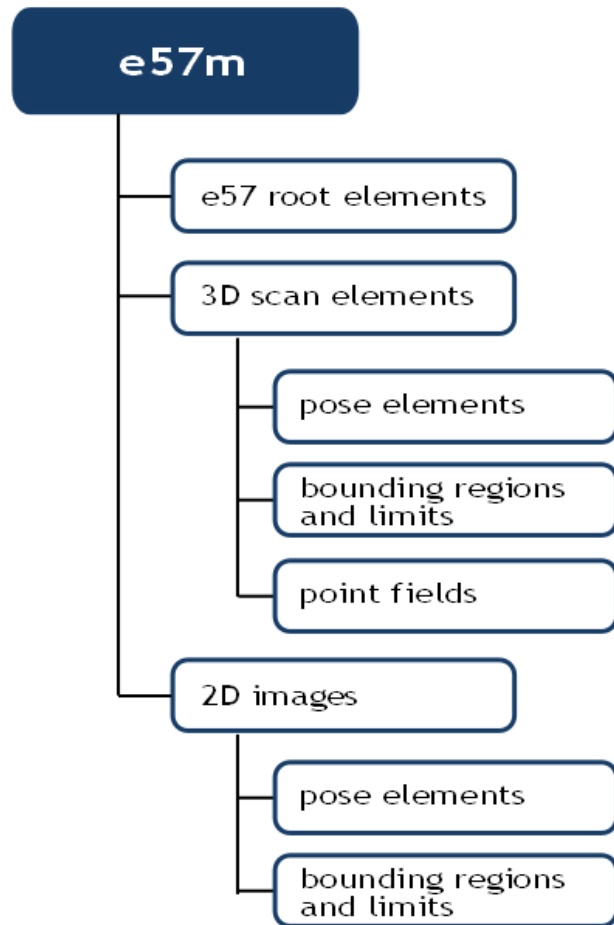
- What's in the file formats
based on openly available specs
- *What is already extractable*
based on libe57 (<http://www.libe57.org/>) and
IfcOpenShell (<http://ifcopenshell.org/>)
- *What use case owners need*
Conducted workshops and surveys



IfcOpenShell

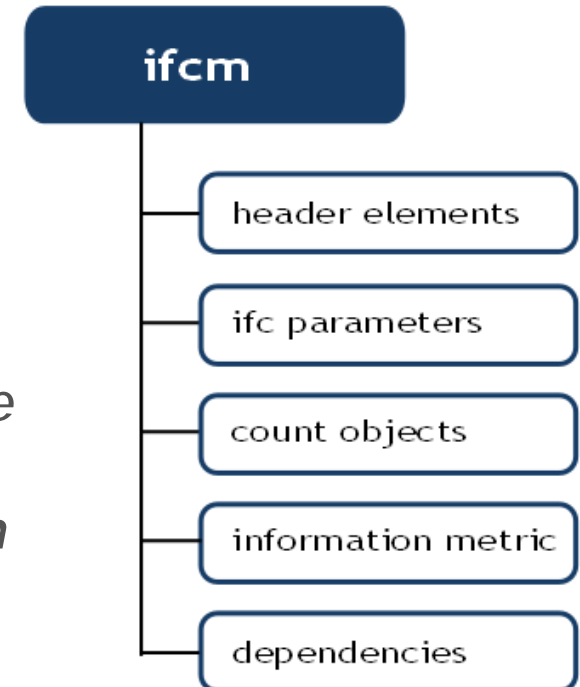
the open source ifc toolkit and geometry engine

Technical Metadata – e57m and ifcm

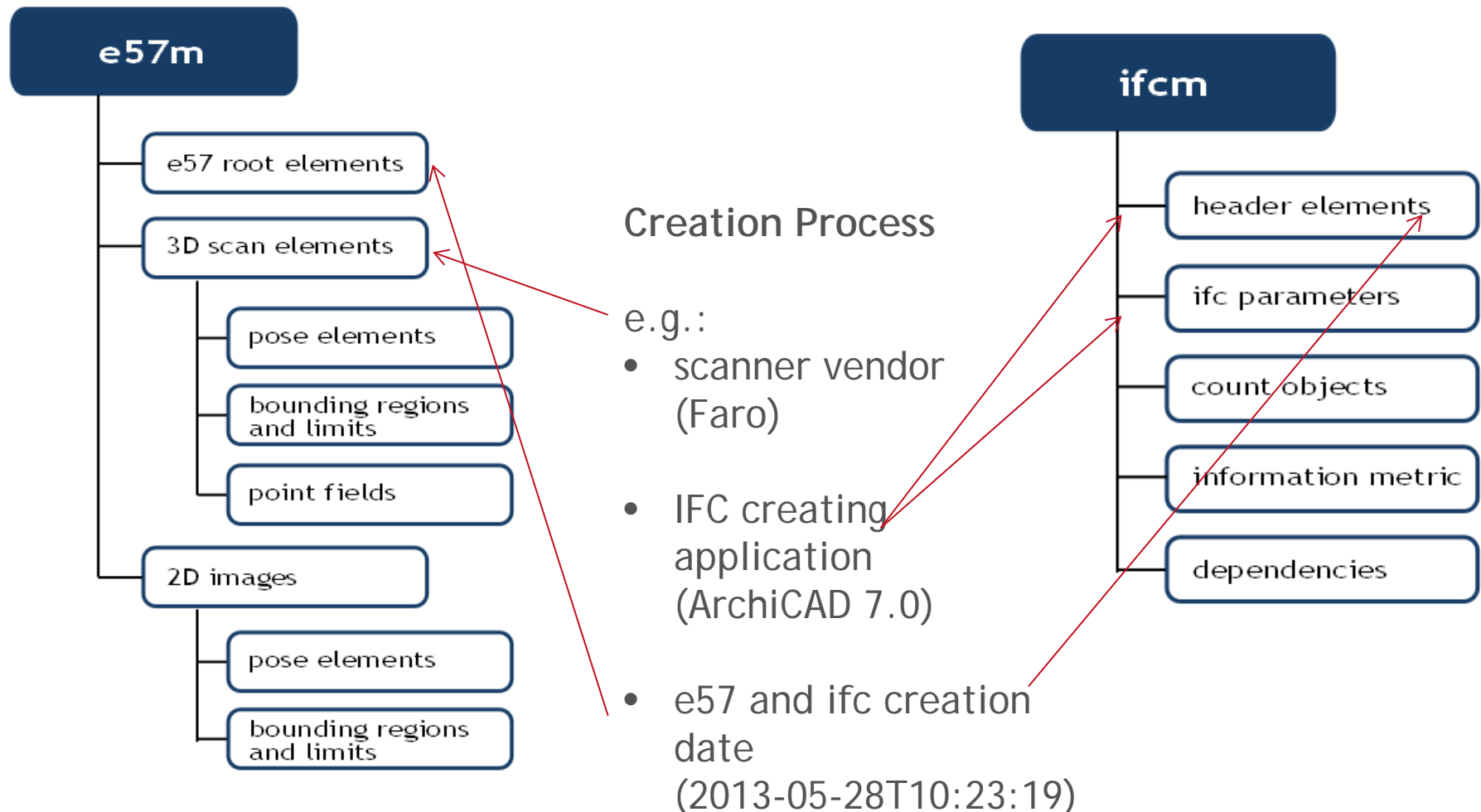


Technical metadata include information about

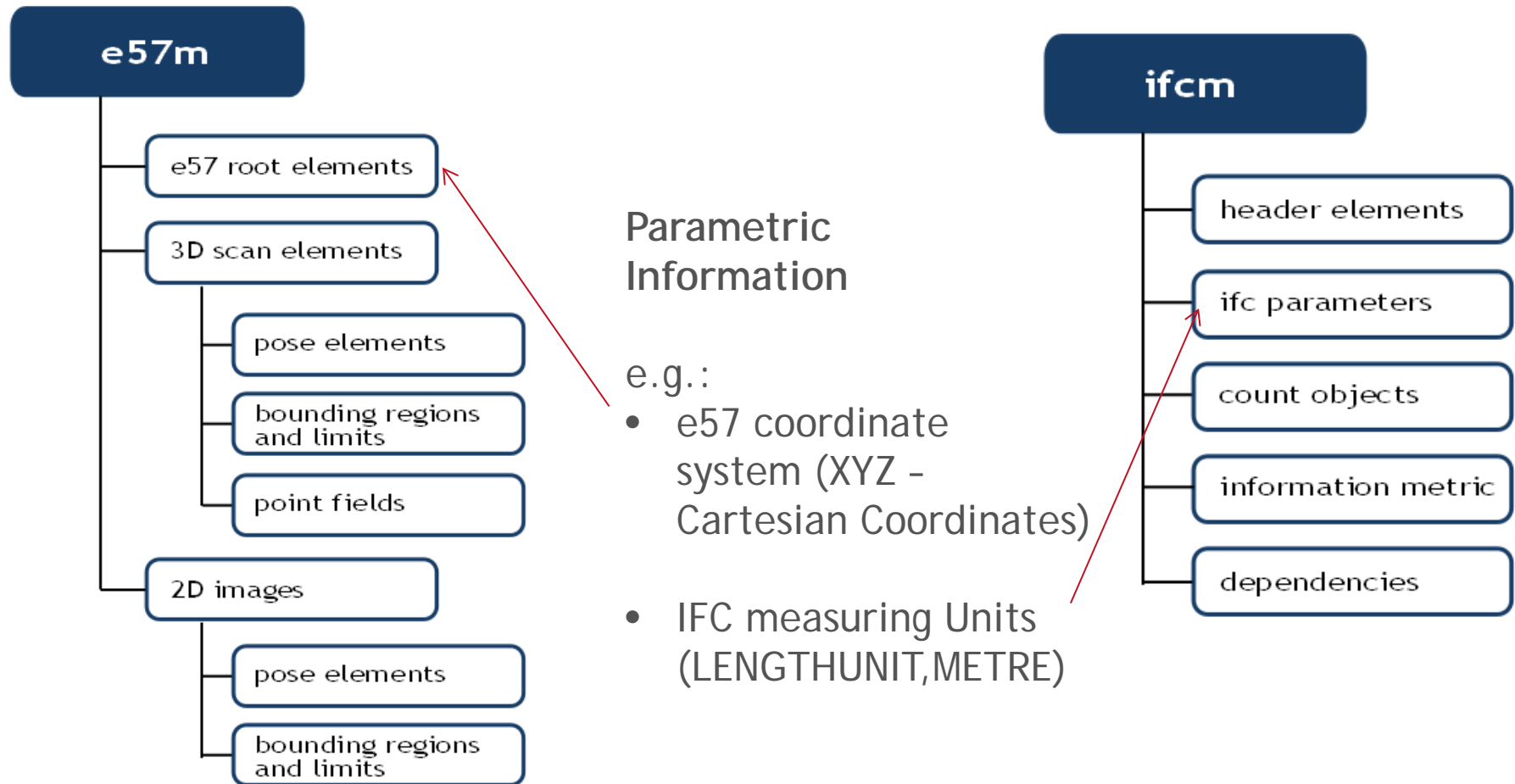
- *Digital Provenance*
- ***Information Scale***
- *Information Depth*



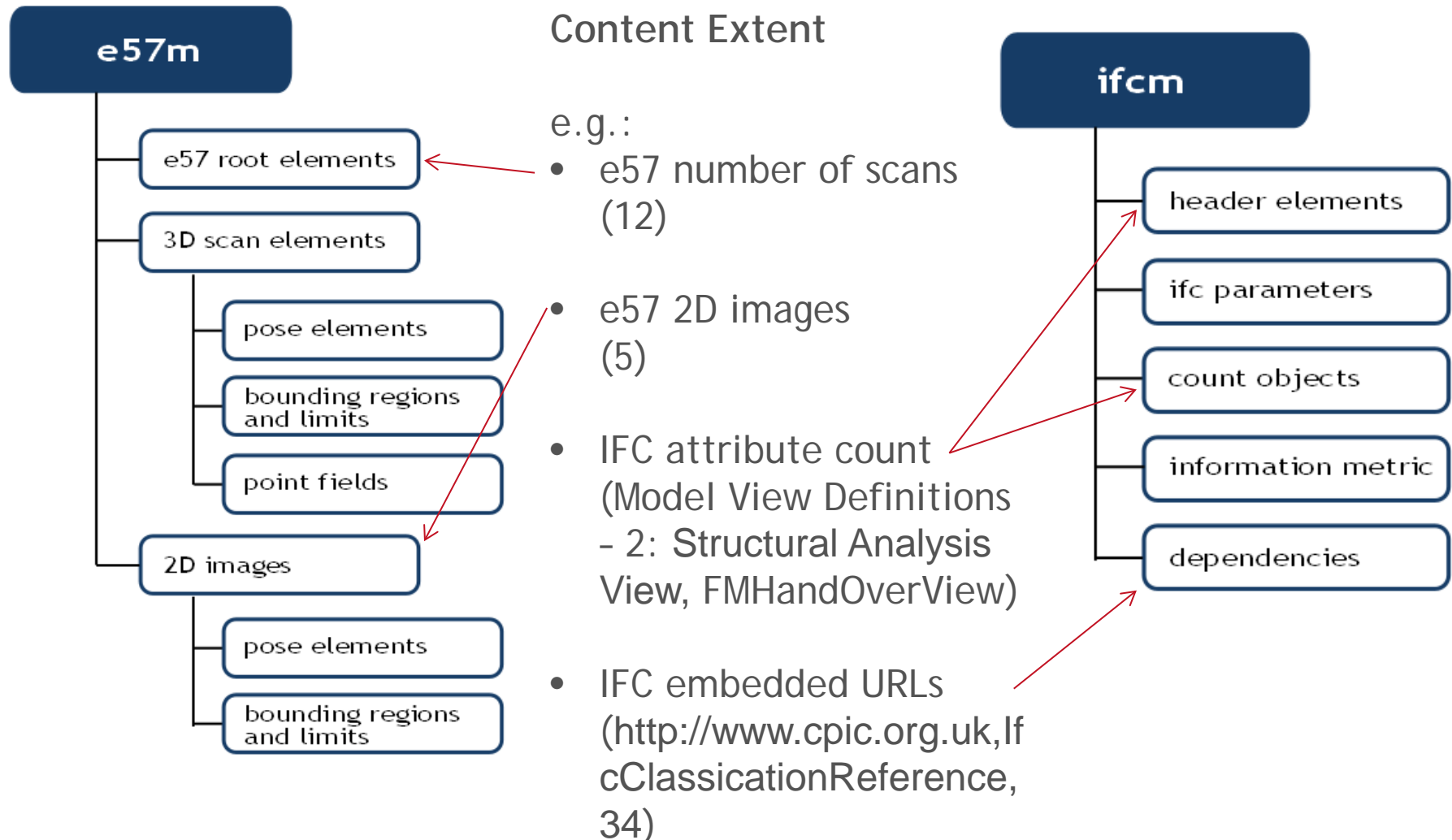
Technical Metadata – Digital Provenance



Technical Metadata – Information Scale



Technical Metadata – Information Depth



Technical Metadata – What do I need it for !?

Collection profiling

- Ground truth about data you have in your repository

Further risk analysis

- to group together objects that share a risk (e.g., dependency on external resources linked via URLs)

Preservation planning

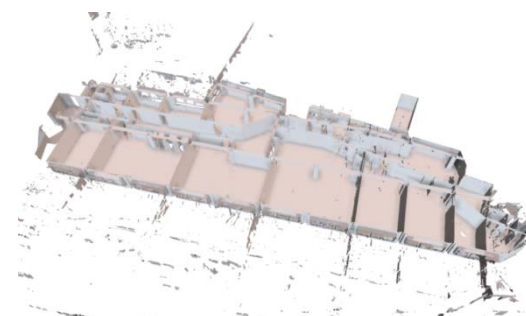
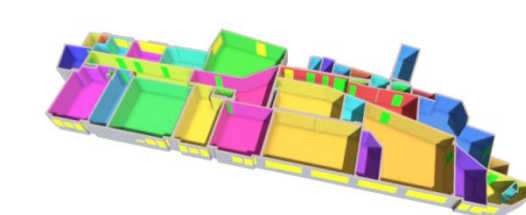
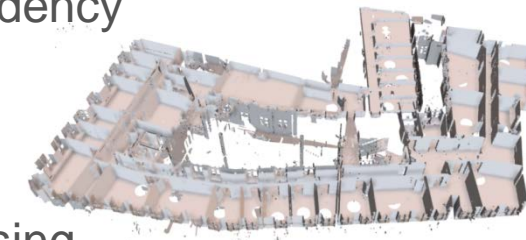
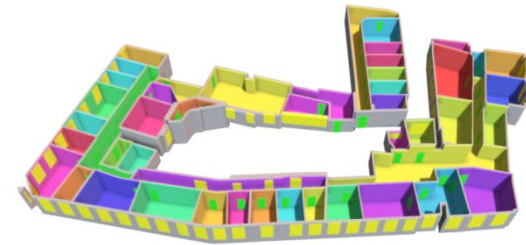
- to build a set of objects for further preservation processing (e.g., extraction of 2D images from e57 files)

Preservation action

- Knowledge of creating software allows to target correct rendering environment in case of emulation

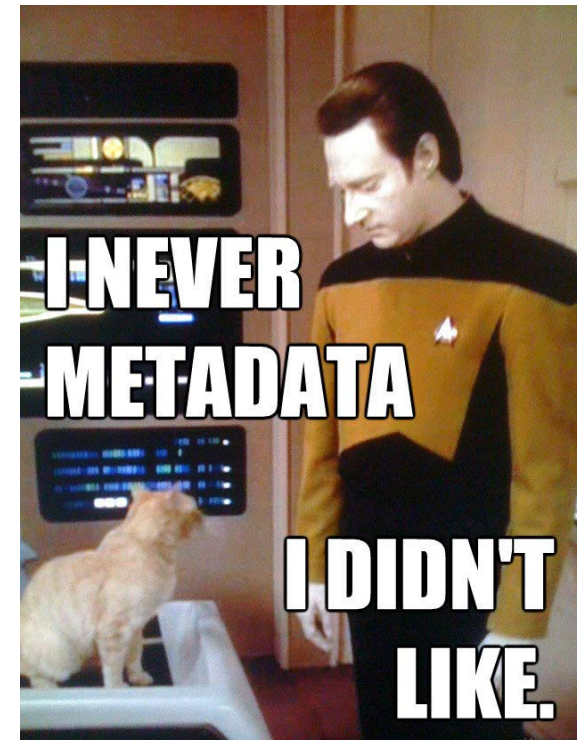
File Format Validation

- techMD extraction is about 2/3rds of the way towards validation

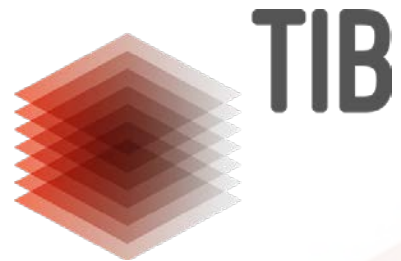


Conclusion and Outlook

- Schemas and processes established through *buildm*, *e57m* and *ifcm* are still valid for TIB today
- Number of models in TIB Digital Archive comparatively low (120 objects of 2 million total)
 - Requires easy to maintain workflow with best possible knowledge of data
- Looking to extend techMD schemas to support further file formats / 3D content types
 - Interested in possibility of super-class for (architectural) 3D content
- Gain better understanding of IFC validation
 - Reported problems even with roundtripping
 - Due to vast degrees of IFC implementation IFC validation is a huge undertaking



LEIBNIZ INFORMATION CENTRE
FOR SCIENCE AND TECHNOLOGY
UNIVERSITY LIBRARY



Thank you!
Questions? Comments!

Contact:

M. Lindlar – TIB Hannover



Michelle.lindlar@tib.eu



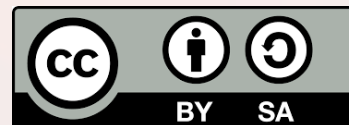
0511 762 19826



Lindlarm



mickylindlar



Further Resources

TIB Digital Preservation Wiki

<https://wiki.tib.eu/confluence/display/lza/Digital+preservation+at+TIB>

Buildm, ifcm and e57m schemas

<https://github.com/DURAARK/Schemas>

E57m and ifcm metadata extrcators

<https://github.com/DURAARK/duraark-metadata>

DURAARK Repository

<http://duraark.eu/data-repository/>