InPT-Dat
Metadata for applied plasma physics and plasma medicine

Markus Becker, Steffen Franke, Lucian Paulet

June 2018

GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung

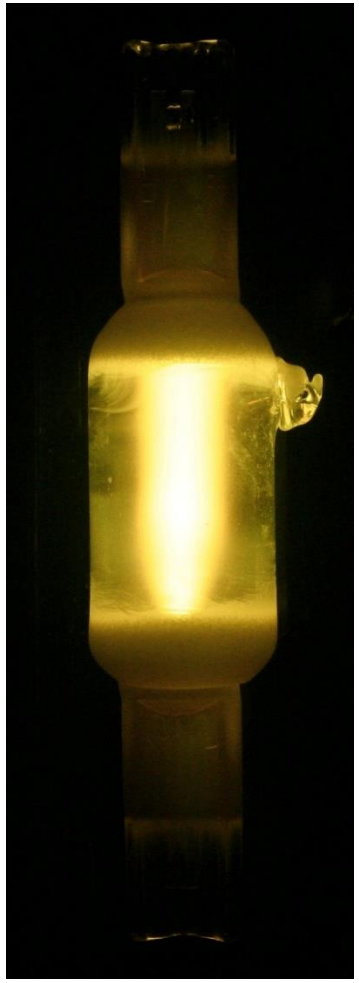
Leibniz-Gemeinschaft

VON DER IDEE ZUM PROTOTYP

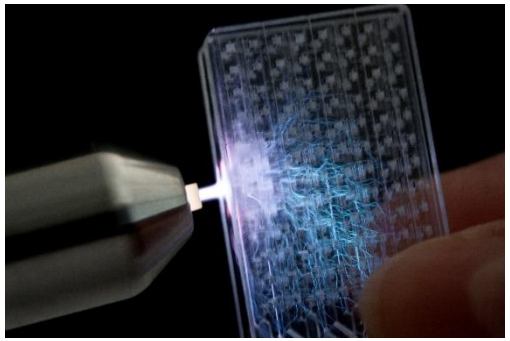
Motivation

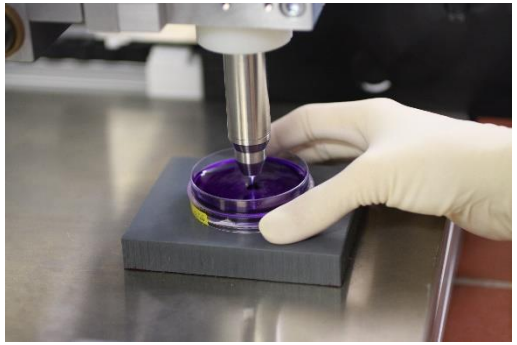
- Metadata are important
- DKAN relies on DCAT metadata standard
- Dublin core included
- Subject specific meta data must be evaluated.
- The motto is:
  As little as possible – as much as necessary.

Metadata in plasma physics

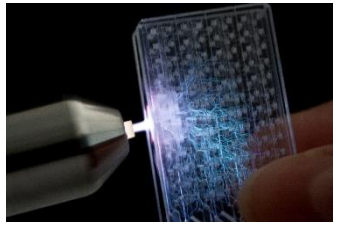


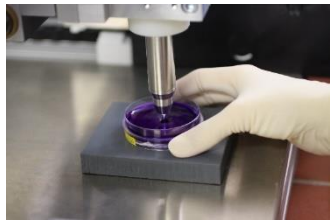
- No geo location, no critical temporal specification, no well defined coverage! (no periodicity, …)
- No distinct resource file type (e.g. jpg), no distinct resource data type (e.g. SEM image).
- However, a clear structure all (physical) plasma research can be related to: source – medium – target.
- Examples for plasma applications: lighting, switching, welding, surface modification, etching, thin film deposition, decontamination, wound healing, …



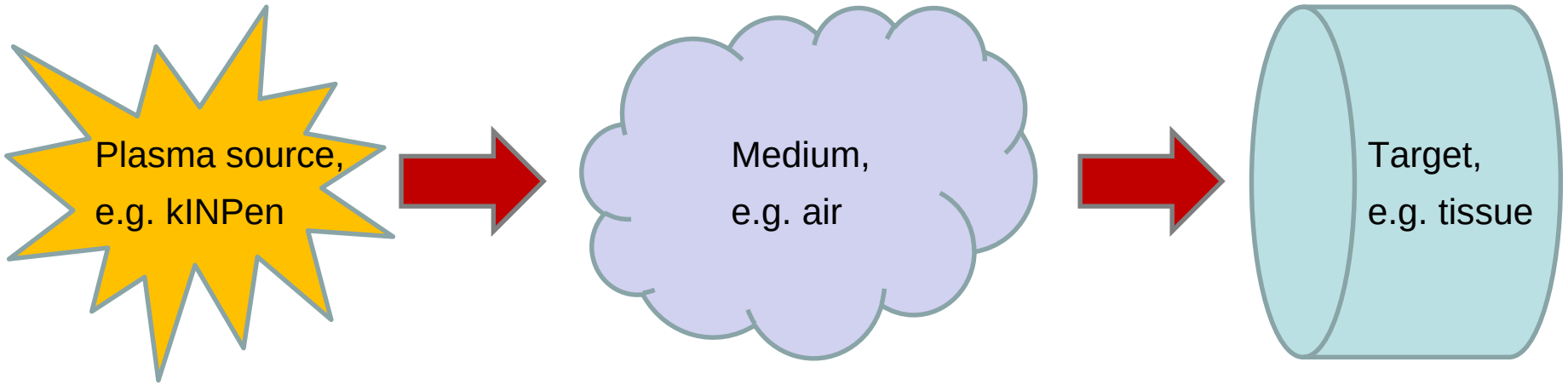


Metadata in plasma physics





- No geo location, no critical temporal specification, no well defined coverage! (no periodicity, …)
- No distinct resource file type (e.g. jpg), no distinct resource data type (e.g. SEM image).
- However, a clear structure all (physical) plasma research can be related to: source – medium – target.



Subject specific metadata standards

- http://www.dcc.ac.uk/resources/subject-areas/physical-science

Physical Science

Materials Science Geography Geology Geoscience Crystallography Solar physics Astrophysics Molecular biology Multi-disciplinary Biochemistry Meteorology Space science Remote Sensing Chemistry Nuclear and Particle Physics Physics Bioinformatics Astronomy

Metadata Standards

AVM - Astronomy Visualization Metadata
A standard defining discovery metadata for fully rendered astronomical imagery.

CIF - Crystallographic Information Framework
An extensible standard file format and set of protocols for the exchange of crystallographic and structured data.

CSMD-CCLRC Core Scientific Metadata Model
A study-data oriented model that captures high-level information about scientific studies and th they produce, primarily tailored for the physical sciences.

FITS - Flexible Image Transport System
Used by the astronomy community to originally describe telescope images, but is now a family to describe multi-dimensional data including spatial, spectral and temporal dimensions and the distortions that may be present.

- IVOA - Resource Metadata for the Virtual Observatory

Collection and service content metadata

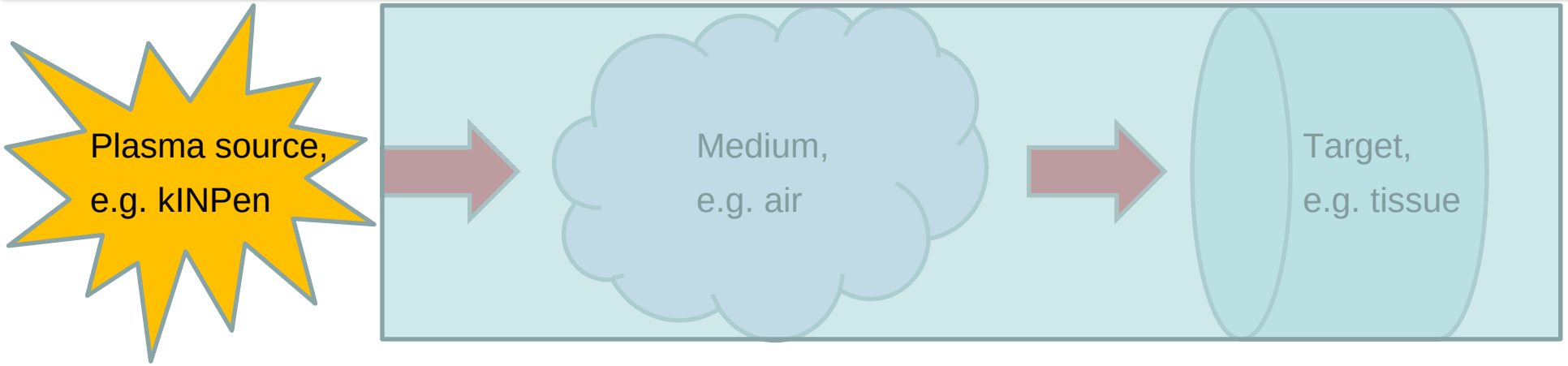
Facility	Apache Point Observatory, Sloan 2.5-m Telescope
Instrument	Five-band clocked CCD camera
Coverage.Spatial	PositionInterval FK5 145.17 –1.25 235.9 1.25 PositionInterval FK5 250.71 52.15 267.0 66.29 PositionInterval FK5 350.43 –1.25 359.99 1.17 PositionInterval 0.0 –1.25 56.37 1.17
Coverage.RegionOfRegard	0.0001
Coverage.Spectral	Optical
Coverage.Spectral.Bandpass	u', g', r', i', z'
Coverage.Spectral.MinimumWavelength	400.e-9
Coverage.Spectral.MaximumWavelength	850.e-9
Coverage.Temporal.StartTime	1999-12-25
Coverage.Temporal.StopTime	2001-07-15
Coverage.Depth	3.e-6
Coverage.ObjectDensity	6.e4
Coverage.ObjectCount	2.e7
Coverage.SkyFraction	0.01

InPT-Dat Metadata: Dublin Core Terms (dct schema)

Label	Field (schema.element.qualifier)	Content
Description	dct.description	Abstract or summary.
Group	dct.publisher	Department
Title	dct.title	
Topic	dct.subject	From taxonomie of subjects (topics)
Licence	dct.rights.licence	e.g. Creative Commons Attribution
Authors	dct.contributor.creator	
Permanent Identifier (DOI)	dct.identifier.doi	DOI of the dataset
Permanent Identifier (URI)	dct.identifier.uri	Uniform Resource Identifier (URL).
Is supplementing (referencing)	dct.relation.references	Reference to published paper
Rights	dct.rights	Access level
Language	dct.language	Language used

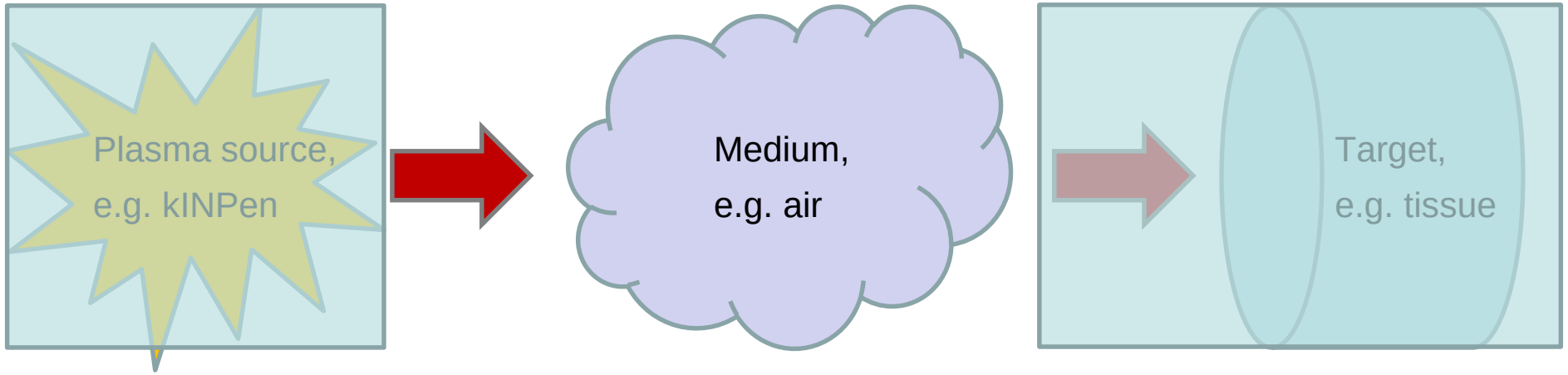
InPT-Dat Metadata: Subject specific metadata schema!

Label	Field (schema.element.qualifier)	Content
Plasma source name	plasma.source.name	Name of the plasma source. (Preferably from a list of names plasma sources.)
Plasma source application	plasma.source.application	Application the plasma source is intended for. (Might be more than one. Preferably from a list of named plasma applications.)
Plasma source properties	plasma.source.properties	Properties of the plasma source. (power, current amplitude, current waveform, frequency, gas, …)
Plasma source procedure	plasma.source.procedure	Procedure to prepare the plasma source. This field should also be used to described the whole procedure including medium and target. That is a (standardized) procedure to treat a medium (if relevant) and act on a target (if necessary).



InPT-Dat Metadata: Subject specific metadata schema!

Label	Field (schema.element.qualifier)	Content
Medium name	plasma.medium.name	Medium name the plasma source is acting on or operated in (e.g. water, dry air). The medium is an optional meta datum and must be given only if the action of the plasma on a target is mediated by some substance without presence of a plasma.
Medium properties	plasma.medium.properties	Properties of the medium, like humidity (air), distilled water, …
Medium procedure	plasma.medium.procedure	Standard procedure to prepare the medium (pre-treatment).

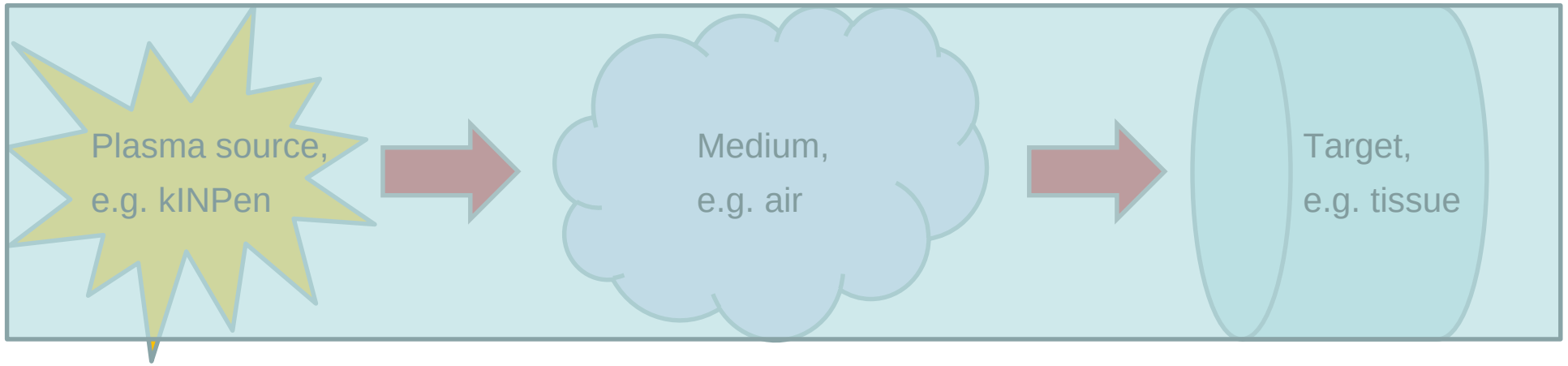


InPT-Dat Metadata: Subject specific metadata schema!

Label	Field (schema.element.qualifier)	Content
Target name	plasma.target.name	Target name the plasma source is acting on either directly or mediated by the above named medium. Can be omitted if only the characterization of a plasma source is intended.
Target properties	plasma.target.properties	Properties of the target (SiO2, polymer, bacteria).
Target procedure	plasma.target.procedure	Standard procedure to prepare the Target (pre-treatment).

InPT-Dat Metadata: Subject specific metadata schema!

Label	Field (schema.element.qualifier)	Content
Resource file type	plasma.resource.filetype	Which file types are saved with this dataset (pdf, jpg, ascii, proprietary file types, ...).
Resource data type	plasma.resource.datatype	Which kind of digital data are saved with this dataset (report/pdf, SEM image/jpg, cfu-plot/ascii, Proteomics, 2D Optical Emission Spectroscopy, ...). Which additional metadata should be given with certain data types?
Resource range	plasma.resource.range	In which range the resource is valid.
Resource quality	plasma.resource.quality	Data quality score



Summary

- Is the choice of plasma metadata reasonable?
- Are they consistent with respect to given standards DublinCore/DataCite? Metadata missing? Metadata doubled?
- Controlled vocabulary vs. free text? Thesaurus vs. hierarchical taxonomy? (plasma sources, plasma applications, standard procedures, …)
- Metadata for datasets vs. metadata for resources?
- **Comments on <plasma> metadata schema are welcome!**

Outlook

- Metadata harvesting as well as registration with DataCite to receive DOI.
- Metadata mapping to meet the standards.

Contact

Leibniz Institute for Plasma Science and Technology
Address: Felix-Hausdorff-Str. 2, 17489 Greifswald
Phone: +49 - 3834 - 554 300, Fax: +49 - 3834 - 554 301
E-mail: welcome@inp-greifswald.de, Web: www. leibniz-inp.de