

# AG Formaterkennung

## Ziele und Aufgaben

- Ziele und Aufgaben
  - Mission Statement
  - laufende Arbeitspakete der nestor AG
    - AP1: Tool Registries
    - AP 2: Testsuites
    - AP 3: Register und Kontaktbörse für obsoleszente Speichertechniken (»R.O.S.T.«)
    - AP 4: Format Registries
    - AP 5: Formatvalidierung
    - AP 6: Best Practices in der Identifizierung
  - laufende Aktivitäten der nestor AG
  - Ansprechpartner
  - Mitglieder
  - Protokolle der Treffen
  - Literatur und nützliche Weblinks

## Mission Statement

Ausreichend detaillierte Kenntnisse zu Datei- und Datenträgerformaten bilden die Grundlage der Risikoabschätzung und Bestandserhaltung für die digitale Langzeitverfügbarkeit. Wegen der Fülle von Formaten weisen die vorhandenen Hilfsmittel jedoch große Lücken auf. Gleichzeitig ist es schwierig, einen Überblick über die Vielfalt bewährter und neu entstehender Werkzeuge zu behalten.

Die nestor-AG Formaterkennung zielt deshalb darauf ab, grundlegende Informationen zu den Themen Formaterkennung, Charakterisierung und Validierung zu sammeln und zu präsentieren. Die AG leistet einen Beitrag zur Verbesserung der Formaterkennung und zum Ausbau der bestehenden Formatdatenbanken (Format Registries). Sie richtet sich an alle Personen und Institutionen, die sich mit der Langzeitverfügbarkeit und -archivierung beschäftigen (v. a. Archive, Bibliotheken, Forschungseinrichtungen und Museen). Diesen Gruppen soll der Einstieg in das Thema Formaterkennung erleichtert werden. Weiterführende Links und Literaturangaben sollen zudem eine vertiefende Beschäftigung mit Fragen der Formaterkennung ermöglichen. Zu diesem Zweck wird die AG auf bereits bestehende Informationsangebote verweisen, wie etwa Tool-Sammlungen (z.B. COPTR) und Dateisammlungen zu verschiedenen Formaten (wie z.B. von Archivemata und dem ArchiveTeam), und möchte dazu anregen, sich am weiteren Ausbau existierender Sammlungen mit eigenen Beiträgen oder Aktualisierungen zu beteiligen. Obsoleszente aber noch funktionstüchtige Abspielumgebungen und Speichermedien der beteiligten Institutionen werden verzeichnet und Erfahrungen darüber ausgetauscht. Die Vernetzung mit anderen nestor-AGs (insbes. AGs E-Akte, AV-Materialien, Digitale Bestandserhaltung) wird dort angestrebt, wo inhaltliche Schnittpunkte zu erwarten sind.

**Aufgaben:** Die nestor-AG Formaterkennung sammelt und präsentiert grundlegende Informationen zu den Themen Formaterkennung, Charakterisierung und Validierung. Die Ergebnisse der AG sollen als erste Anlaufstelle für alle Personen dienen, die sich mit der Risikobewertung in der Bestandserhaltung befassen.

Die AG wird bereits vorhandenen Infrastrukturen wie der Toolsammlung COPTR (Community Owned digital Preservation Tool Registry) zuarbeiten und eine Dateisammlung ausgewählter Formate erstellen. Ein Verzeichnis von verfügbaren Abspielumgebungen für obsoleszente Speichermedien wird aufgebaut und der Erfahrungsaustausch gefördert. Die Vernetzung mit anderen nestor-AGs ist bei Querschnittsthemen geplant.

### Ziele 2015-2017:

- 2015: Die Unter-AGs konstituieren sich und schärfen das Profil der Arbeitspakete.
- ab 2016: Informationen werden fortlaufend gesammelt, aufbereitet und zur Diskussion bereitgestellt.

## laufende Arbeitspakete der nestor AG

### AP1: Tool Registries

Das Arbeitspaket strebt die Erarbeitung einer Liste vorhandener Formaterkennungs-, Formatvalidierungs- und Formatcharakterisierungstools an, einerseits um Einsteigern einen schnellen Überblick über Werkzeuge zu geben, andererseits um auf neue bzw. unbekanntere Tools aufmerksam zu machen. Da es bereits Tool Registries wie COPTR gibt, soll keine eigene Seite mit Toolbeschreibungen aufgebaut werden, sondern vielmehr die vorhandene Tool Registry COPTR durch Zuarbeiten der nestor-AG unterstützt werden. Die hier nachfolgende Übersicht dient nur dem zentralen Einstieg in das Arbeitspaket und der Dokumentation der Arbeitsergebnisse. Das Arbeitspaket wird federführend von Claire Roethlisberger-Jourdan betreut; Zuarbeiten erfolgen auch durch weitere Mitglieder der AG.

### AP 2: Testsuites

Ziel dieses AP ist, ein öffentlich publizierbares Sample für Dateiformate zu erstellen, das für Tests und Tests von Tools genutzt werden kann. Wenn möglich können offen publizierbare Beispiele für Dateiformate hinzugefügt werden. Federführend ist Andre Müller, die Zuarbeit erfolgt auch durch weitere Mitglieder der AG.

## AP 3: Register und Kontaktbörse für obsoleszente Speichertechniken (»R.O.S.T.«)

[siehe Unterseite](#)

## AP 4: Format Registries

Dieses Arbeitspaket umfasst die praktische Zuarbeit zu Format Registries wie beispielsweise PRONOM. Die AG wird durch eigene Beiträge die Datenbestände der Format Registries erweitern. Zudem wird sie eine deutschsprachige Anleitung zur Erstellung von Formatsignaturen für PRONOM erarbeiten, um so die Kenntnisse und die Motivation zur Mitarbeit auch außerhalb der AG zu stärken. Federführend ist Martin Hoppenheit vom LA NRW.

## AP 5: Formatvalidierung

Der Fokus dieses Arbeitspakets wird zurzeit geschärft. Federführend sind Yvonne Friese (ZBW) und Stefan Hein (DNB).

## AP 6: Best Practices in der Identifizierung

Das Arbeitspaket strebt die Erarbeitung verschiedener Best Practices bei der Identifizierung an, einerseits um Einsteigern einen schnellen Überblick über die verschiedenen Szenarien zu geben, andererseits um auf Probleme und Lösungsmöglichkeiten aufmerksam zu machen. Das genaue Profil wird zurzeit geschärft.

## laufende Aktivitäten der nestor AG

Kooperation mit der Open Preservation Foundation (OPF) Document Interest Group (DIG)

Umfrage im deutschsprachigen Raum: Welche Identifizierungs- und Validierungstools werden verwendet?

Aufbau und Ausbau eines Registers und Kontaktbörse für obsoleszente Speichertechniken (»R.O.S.T.«)

## Ansprechpartner

Name, Vorname	Einrichtung	Kontakt
Tunnat (geb. Friese), Yvonne	ZBW	<a href="mailto:y.tunnat@zbw.eu">y.tunnat@zbw.eu</a>

## Mitglieder

Name	Einrichtung	Kontakt
Willems, Laura	Bundesarchiv	<a href="mailto:l.willems@bundesarchiv.de">l.willems@bundesarchiv.de</a>
Jacobs, Rainer	Bundesarchiv	<a href="mailto:r.jacobs@bundesarchiv.de">r.jacobs@bundesarchiv.de</a>
Rausch, Joachim	Bundesarchiv	<a href="mailto:j.rausch@bundesarchiv.de">j.rausch@bundesarchiv.de</a>
Kramski, Heinz Werner	DLA Marbach	<a href="mailto:Heinz.Werner.Kramski@dla-marbach.de">Heinz.Werner.Kramski@dla-marbach.de</a>
Hein, Stefan	DNB	<a href="mailto:S.Hein@dnb.de">S.Hein@dnb.de</a>
Steinke, Tobias	DNB	<a href="mailto:T.Steinke@dnb.de">T.Steinke@dnb.de</a>
Müller, Andre	GESIS Köln	<a href="mailto:andre.mueller@gesis.org">andre.mueller@gesis.org</a>
Klindt, Marco	Konrad-Zuse-Zentrum Berlin	<a href="mailto:klindt@zib.de">klindt@zib.de</a>
Röthlisberger-Jourdan, Claire	KOST Bern	<a href="mailto:Claire.Roethlisberger@kost.admin.ch">Claire.Roethlisberger@kost.admin.ch</a>
Hoppenheit, Martin	Landesarchiv Nordrhein-Westfalen (LAV NRW)	<a href="mailto:Martin.Hoppenheit@lav.nrw.de">Martin.Hoppenheit@lav.nrw.de</a>
Meckel, Konrad	Landesarchiv Thüringen (LATH)	<a href="mailto:Konrad.Meckel@la.thueringen.de">Konrad.Meckel@la.thueringen.de</a>
Kortyla, Stephanie	Sächsisches Staatsarchiv	<a href="mailto:Stephanie.Kortyla@sta.smi.sachsen.de">Stephanie.Kortyla@sta.smi.sachsen.de</a>
Romeyke, Andreas	SLUB Dresden	<a href="mailto:romeyke@slub-dresden.de">romeyke@slub-dresden.de</a>
Röhrle, Mario	Staatliche Akademie der Bildenden Künste, Stuttgart	<a href="mailto:mario.roehrle@abk-stuttgart.de">mario.roehrle@abk-stuttgart.de</a>
Alfredo Wutschka	ZB MED Köln / Bonn	<a href="mailto:wutschka@zbmed.de">wutschka@zbmed.de</a>
Silke Reich	Hebis	<a href="mailto:sreich@hebis.uni-frankfurt.de">sreich@hebis.uni-frankfurt.de</a>
Fabienne Hartwig (geb. Metzger)	Kreisarchiv Esslingen	<a href="mailto:Hartwig.Fabienne@Ira-es.de">Hartwig.Fabienne@Ira-es.de</a>

Julia Krämer-Riedel	Stadt Köln Historisches Archiv	<a href="mailto:julia.kraemer-riedel@stadt-koeln.de">julia.kraemer-riedel@stadt-koeln.de</a>
Christian Treu	Sächs. Informatik Dienste	<a href="mailto:christian.treu@sid.sachsen.de">christian.treu@sid.sachsen.de</a>

nestor Mailingliste [nestor-formaterkennung@lists.dnb.de](mailto:nestor-formaterkennung@lists.dnb.de)

## Protokolle der Treffen

1. Treffen: 12.05.2014, 11:30 - 16:00 Uhr an der DNB in Frankfurt a.M., [Protokoll](#)
2. Treffen: 4.11.2014, 11:00 - 16:00 Uhr an der Gesis in Köln, [Protokoll](#)
3. Treffen: 25.03.2015, 11:00 - 16:00 Uhr beim Deutschen Literaturarchiv in Marbach, [Protokoll](#)
4. Treffen: 24.08.2015, 11:00 - 16:00 Uhr am Konrad-Zuse-Institut in Berlin, [Protokoll](#)
5. Treffen: 9.05.2016, 11:00 - 16:00 Uhr an der ZBW in Kiel, [Protokoll](#)
6. Treffen: 20.09.2016, 11:00 - 16 Uhr an der TIB in Hannover, [Protokoll](#)
7. Treffen: 14.03.2017, 11:00 - 16 Uhr am LAV in Duisburg, [Protokoll](#)
8. Treffen: 27.06.2017, 14:30 - 18:30 Uhr in der ZBW Kiel, [Protokoll](#)
9. Treffen: 19.03.2018, 11:00 - 15:00 Uhr an der GESIS, Köln, [Protokoll](#)
10. Treffen: 06.11.2018, 11:00 - 16:00 Bundesarchiv in Koblenz, [Protokoll](#)
11. Treffen: 27.03.2019, 11:00 - 16:00 ZB MED in Köln, [Protokoll](#)
12. Treffen: 05.11.2019, 9:00 - 15:00 in Kiel
13. Treffen: voraussichtlich März 2020 in Hannover (10:00 - 16:00 ?)

## Literatur und nützliche Weblinks

- Validierungstool [KOST-Val](#)
- [KOST: Formaterkennung und -validierung](#)
- [Pronom](#)
- [Google Forum von Pronom](#) (User und Entwickler)
- Die [Zukunft von JHOVE](#) und JHOVE 2 und mögliche Weiterentwicklungen
- [Ange Albertini](#): Funky File Formats (Vortrag): [Vortragsvideo](#) / [Folien](#) / [Demodateien](#)
- [Ange Albertini](#): Preserving arcade games ([Vortrag](#)) (weniger zu Formaten, mehr zur Archivierung allgemein)