



# Enhancing Reproducibility in Precipitate Analysis: A FAIR Approach with Automated Dark-Field Transmission Electron Microscope Image Processing

Ghezal Ahmad Jan Zia<sup>1</sup> · Thomas Hanke<sup>2</sup> · Birgit Skrotzki<sup>1</sup> · Christoph Völker<sup>1</sup> · Bernd Bayerlein<sup>1</sup> 

Received: 27 October 2023 / Accepted: 28 November 2023  
© The Author(s) 2024

## Abstract

High-strength aluminum alloys used in aerospace and automotive applications obtain their strength through precipitation hardening. Achieving the desired mechanical properties requires precise control over the nanometer-sized precipitates. However, the microstructure of these alloys changes over time due to aging, leading to a deterioration in strength. Typically, the size, number, and distribution of precipitates for a quantitative assessment of microstructural changes are determined by manual analysis, which is subjective and time-consuming. In our work, we introduce a progressive and automatable approach that enables a more efficient, objective, and reproducible analysis of precipitates. The method involves several sequential steps using an image repository containing dark-field transmission electron microscopy (DF-TEM) images depicting various aging states of an aluminum alloy. During the process, precipitation contours are generated and quantitatively evaluated, and the results are comprehensively transferred into semantic data structures. The use and deployment of Jupyter Notebooks, along with the beneficial implementation of Semantic Web technologies, significantly enhances the reproducibility and comparability of the findings. This work serves as an exemplar of FAIR image and research data management.

**Keywords** Automated image analysis · FAIR research data management · Reproducibility · Microstructural changes

## Introduction

High-strength aluminum alloys, which are used for example in aerospace and automotive applications, obtain their strength by precipitation hardening within a heat treatment process [1]. During this process, tiny (nanometer-sized) precipitates are formed which serve as obstacles for dislocation motion in these alloys. The microstructure of precipitation-hardened aluminum alloys is of crucial importance for their strength since only materials with a targeted microstructure (i.e. size, number, and distribution of precipitates) achieve the desired strengths. However, this

optimized microstructure may change during operation (a process called aging), because the components are used at temperatures close to or even exceeding the hardening temperature (e.g., in the case of radial compressor wheels in turbochargers). This process leads to a deterioration of strength (and also hardness) [2] over time and is undesirable. Today, such effects are not considered in existing lifetime prediction models because quantitative data on the changes are generally lacking, which would, however, be a prerequisite for representation in models. While the changes in mechanical properties can be investigated and quantified with comparatively little effort, microstructural characterizations for the evolution of the precipitates require much more time (as well as technical specialization) and are accordingly rarely performed in detail. An example of one of the comprehensive studies on the temporal evolution of hardening phases for different temperatures is the work of Rockenhäuser et al. [3, 4, 5]. They investigated the coarsening processes of the rod-shaped hardening phase (hereafter referred to as S-phase) in aluminum alloy EN AW-2618A (referred to as 2618A in the following) at high temperatures. For this purpose, specimens were aged (i.e., annealed) at 190 °C for up to 25,000 h and

---

✉ Ghezal Ahmad Jan Zia  
ghezal-ahmad.zia@bam.de

✉ Bernd Bayerlein  
bernd.bayerlein@bam.de

<sup>1</sup> Bundesanstalt für Materialforschung und -prüfung (BAM),  
Unter den Eichen 87, 12205 Berlin, Germany

<sup>2</sup> Fraunhofer-Institut für Werkstoffmechanik (IWM),  
Wöhlerstraße 11, 79108 Freiburg im Breisgau, Germany

the microstructure was extensively characterized based on DF-TEM imaging technique. The aim of the study was to determine a geometric parameter relevant to the strength of the material (in this case, the rod radii) for each aging condition and to describe the evolution of the radii distribution and the average radii with aging time. The results can be incorporated into models for predicting the service life of the alloy 2618A.

The use of a manual approach as described in the study by Rockenhäuser et al. is common in materials science, but has a number of disadvantages:

- High time and labor requirements, especially for large datasets or when analyzing specimens with complex structures.
- Susceptibility to subjectivity.
- Dependence on the experience and skills of the operator.
- Transfer of knowledge about the exact way of evaluation is critical, as often not documented in sufficient detail.

This can lead to variability and thus uncertainty in the results, which affects the accuracy and reliability of further analysis (e.g., in the subsequent use of the data in life-cycle prediction models). Automated methods, on the other hand, can potentially be more efficient, objective, and reproducible, but include their own limitations and potential sources of error.

An alternative to the manual analysis described is provided by automatable image analysis tools running in Jupyter Notebook environments. Jupyter Notebooks allow collaborative development and usage, support multiple programming languages, and can be customized for specific use cases [6]. The image data to be analyzed can be published in a central repository, such as Omero,<sup>1</sup> which can be accessed by a Jupyter Notebook script via an application programming interface (API). Image files loaded into the notebook can then be processed through a preconfigured sequence of image processing, feature segmentation, and analysis steps. By accessing the documented script and the specific input parameters used, the image analysis process and the output data become reproducible. This procedure is similar to most imaging techniques and their associated quantification of relevant image information. Initially, the image file is converted and processed, followed by the segmentation of selected image features, typically using a thresholding approach. In the final step, these features are measured and quantitatively evaluated [7, 8].

Imaging analysis techniques play a crucial role in characterizing materials in materials science and engineering. The increasing interest in robot-assisted high-throughput

techniques with high levels of automation and image data generation necessitates the development of advanced and automatable image analysis routines as an integral part of the process [9]. Moreover, more advanced image and data management practices are needed to keep up with these advances [10]. Therefore, in the context of FAIR (Findable, Accessible, Interoperable, Reusable) data management, it is increasingly important to present image data streams and resulting tables of values in standardized formats to ensure comparability [11, 12]. Semantic Web technologies offer suitable solutions for many emerging digital challenges [13]. Unlike static relational databases, data and important meta-data for process reproducibility can be flexibly made available in triple stores<sup>2</sup> for reuse and retrieval. This is achieved by annotating and transforming data into RDF (Resource Description Framework) instances, known as triples,<sup>3</sup> using ontologies. Homogenizing complete materials data along with relevant source information into high-quality data structures adhering to FAIR principles facilitates comparability, which serves as the foundation for knowledge discovery and retrieval [14, 15].

Our work aims to make a significant contribution to the field by presenting an understandable and accessible example of S-phase precipitation analysis based on DF-TEM images of different aging states of the aluminum alloy 2618A. The primary objective of this work is to elucidate the efficiency of collaborative tools, incorporating automatable and modular digital image processing, and analysis workflows, alongside the integration of Semantic Web technologies. The underlying proposition of our work is the potential of utilizing semantic representations of valuable measurement data in conjunction with automated image processing pipelines, quietly alluding to the prospect of adhering to FAIR principles in a scalable manner. Through exemplary digital and semantic representations of this specialized characterization method, we demonstrate the inherent attributes of reproducibility, reliability, and accessibility in processes and generated data. Implicitly, we put forth effective solutions to surmount extant challenges and actively promote digital transformation within the scientific community. Envisaging the extraction of intricate, specialized insights from distributed data, we strive to render invaluable inherent knowledge searchable and readily available for extensive analysis by a diverse scientific audience.

<sup>1</sup> <https://www.openmicroscopy.org/omero/>.

<sup>2</sup> [https://www.w3.org/2001/sw/wiki/Category:Triple\\_Store](https://www.w3.org/2001/sw/wiki/Category:Triple_Store).

<sup>3</sup> <https://www.w3.org/RDF/>.

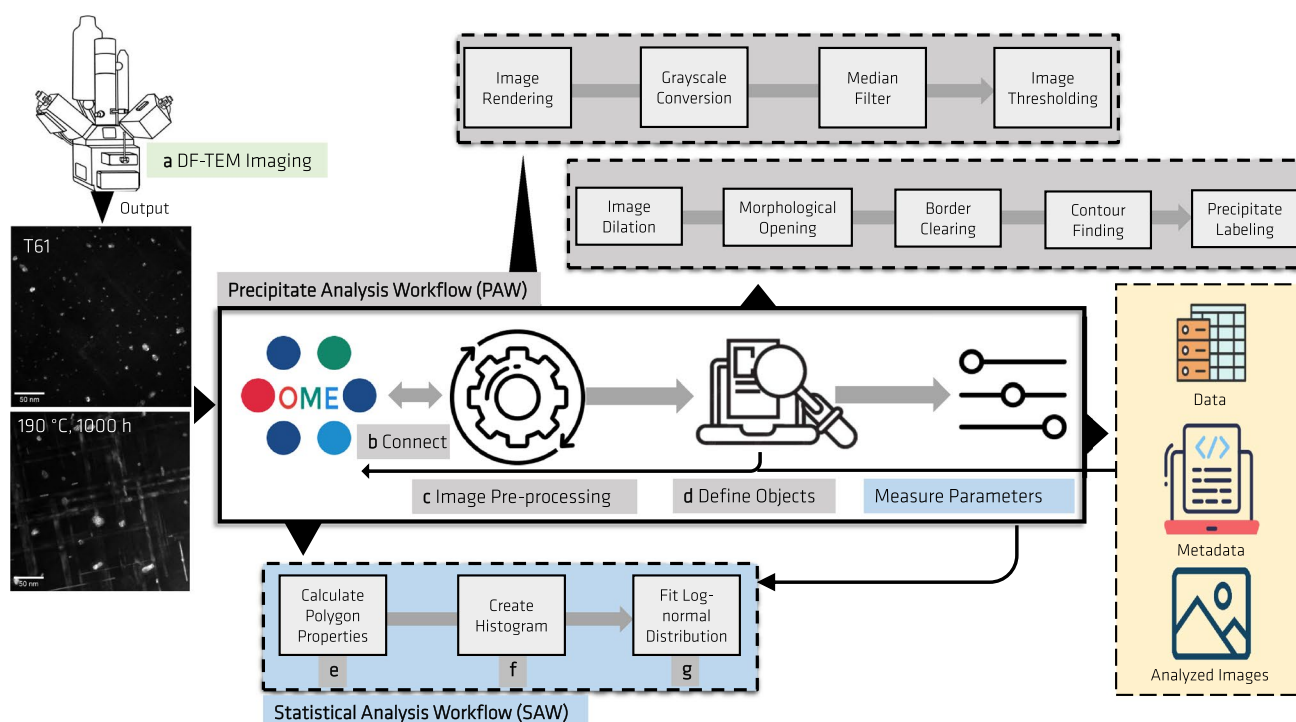
**Table 1** Configuration parameters for various methods

| ID of directory in Omero | ID of image   | Filter    | Threshold  | Morphological operation |
|--------------------------|---------------|-----------|------------|-------------------------|
| T61 (63)                 | Stelle1 (128) | Median    | Otsu       | Dilation                |
| .                        | .             | Gaussian  | Additive   | Erosion                 |
| .                        | .             | Sobel     | Binary     | Opening                 |
| .                        | .             | Laplacian | Multilevel | Closing                 |

## Precipitate Analysis Workflow (PAW)

To illustrate our findings, Fig. 1 showcases the architecture of the DF-TEM image analysis processing pipeline. The PAW was carefully designed to facilitate modularization, providing the capability of customization at each stage of the pipeline. This flexibility enables users to tailor the pipeline to their specific requirements and preferences, as exemplified in Table 1.

For instance, users have the option to select alternative



**Fig. 1** Schematic overview of DF-TEM imaging and the subsequent image processing and analysis pipeline, incorporating both PAW and SAW components: **a** DF-TEM imaging (see “TEM Imaging”) (DF-TEM images include 50 nm scale bar for size reference). The PAW

includes: **b** Omero server connection, **c** image pre-processing, and **d** object definition. The SAW comprises: **e** Polygon properties calculation, **f** histogram creation, and **g** log-normal distribution fitting

## Results

In the following sections, the two main components of our automated image processing and precipitation analysis pipeline, the Precipitate Analysis Workflow (PAW) and the Statistical Analysis Workflow (SAW), are presented and described in detail. The last subsection addresses the creation of two ontologies, the PAW Ontology (PAWO) and the SAW Ontology (SAWO), that semantically represent these workflows and enable RDF instance creation.

thresholding algorithms, such as Otsu’s method, and fine-tune associated parameters to achieve precise thresholding results. Similarly, various filtering techniques, such as the median filter, can be chosen with filter parameters adjustable to attain desired image smoothing or noise reduction effects. Furthermore, users can customize object expansion or connectivity criteria by modifying the size and shape of the structuring element, thus configuring the dilation process to their needs.

Additionally, the pipeline seamlessly manages directories of microscopy images on the Omero server, automatically organizing them into a structured hierarchy of directories and filenames for input and output.

In the following, we provide a detailed description of the three key components of the PAW.

### Establishing a Connection to Omero

To initiate the process, the OmeroConnect class, sourced from Omero tools [16] is employed to establish a connection with the Omero server, which manages image data. The primary pipeline begins by requesting the user's Omero credentials, comprising their username and password. The script utilizes these credentials to establish a connection with the Omero server, subsequently retrieving a list of datasets.

For each dataset, the script [17] iterates through the images, extracting pertinent metadata such as aging temperature, aging time, specimen name, and more. This information is vital for subsequent analysis.

### Image Pre-Processing

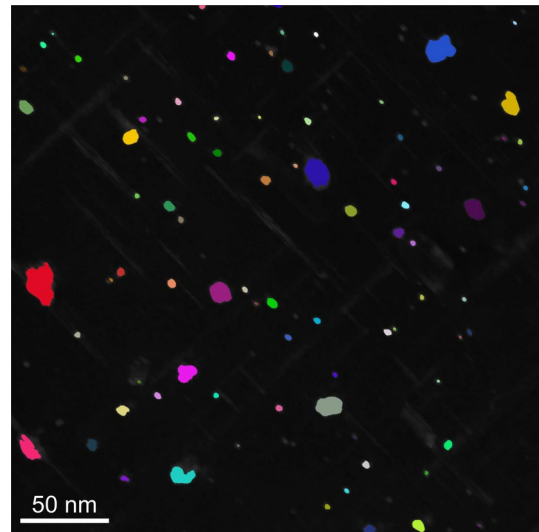
Image pre-processing is a fundamental step in the analysis pipeline, involving several key operations to optimize the image for further analysis. The primary objective of this phase is to generate a visually enhanced representation of the image, highlighting specific regions of interest.

In this context, rendering techniques are applied to focus on the central  $z$ -section and the initial time point. The rendering process is accomplished using the *renderImage()* function, which effectively extracts the desired image slice for subsequent analysis.

Following rendering, the image undergoes grayscale conversion using the OpenCV library,<sup>4</sup> transforming it into a single-channel grayscale format. This simplification is essential for enhancing the efficiency of various image analysis algorithms, as they tend to perform better on grayscale images [18].

Crucially, image pre-processing involves a series of critical parameter adjustments, including the application of a median filter, the selection of an appropriate thresholding method, and consideration of size thresholds. This crucial step encompasses several procedures aimed at accurately identifying precipitates within the grayscale image.

- *Median filtering*: to reduce noise in the image, a median blur filter is applied using OpenCV.
- *Thresholding*: Otsu's thresholding method is utilized to convert the filtered image into a binary image. This



**Fig. 2** Detected precipitates in an exemplary image from the T61 dataset: labeled mask of precipitates with assigned distinct colors

method automatically calculates an optimal threshold value to distinguish between foreground and background pixels [19].

### Define Objects

The process of defining objects serves as the pivotal function within the pipeline, with its significance stemming from its role in parameter utilization for fine-tuning the detection algorithm. This critical phase involves the incorporation of various parameters, including dilation, dilate kernel size, and morphological opening. The following sequential operations are executed on the thresholded image to effectively identify precipitates:

- *Dilation*: using a kernel, the thresholded image undergoes dilation, connecting nearby pixels to form complete objects.
- *Morphological operations*: the application of morphological opening aids in the removal of small objects or noise by performing erosion followed by dilation. The erosion step erodes the boundaries of objects, effectively eliminating small or thin structures. Subsequently, the dilation step expands the remaining structures, restoring the original shape and size of larger objects. Applying dilation before morphological opening enhances the removal of small objects while preserving the integrity of larger structures in the image. However, the effectiveness of this approach may vary depending on the specific image characteristics and desired outcomes.
- *Clear border*: the clear border operation is employed to remove any remaining objects near the image border.

<sup>4</sup> <https://opencv.org>.

- *Contour finding*: leveraging OpenCV's *findContours()* function, the contours of the remaining precipitates are identified. Contours with an area smaller than the specified threshold are recognized as noise and subsequently removed.
- *Precipitate labeling*: the resulting image is labeled by assigning a unique label to each precipitate. Furthermore, the number of detected precipitates is determined by counting the total number of labels. To visualize the labeled mask, distinct colors are assigned to each label, creating a plot to display the image (see Fig. 2).
- *Regions of interest (ROIs)*: the contours of the detected precipitates are uploaded to the Omero server as polygons using the *create-omero-roi-polygons* function from the Omero-tools library [16]. ROIs<sup>5</sup> enable convenient access and visualization of the detected precipitates on the Omero platform.

### Create Documentation Metadata

For each processed image, a record is generated to document all variables and utilized resources. This resulting table is then exported in comma-separated values (CSV) format [20]. Furthermore, it is annotated with a Mat-O-Lab tool CSVToCSVW [21], utilizing the CSV on the Web (CSVW) vocabulary from W3C.<sup>6</sup> The derived metadata document, available in JSON-LD<sup>7</sup> format, serves to define the structure and content of the resulting table. This not only facilitates the proper interpretation of the table but also allows for linking its content to other semantic data. Additionally, it provides the functionality to convert the entire table into RDF [22], utilizing the standard output configuration.<sup>8</sup>

To enhance the documentation even further, a knowledge graph is linked to individual records within the table. This graph is created using Draw.io<sup>9</sup> in conjunction with the OntoPanel plugin, which is elaborated upon in “[Ontology Creation and Application](#)”. The graph provides detailed explanations of the steps taken to achieve the results. Entities highlighted in yellow, as depicted in Fig. 5 are mapped to the columns of the result table by defining specific mapping rules. This mapping is established using the Mat-O-Lab tool MapToMethod [23] and is required only once for a given CSV table configuration. It can be reused as long as the CSV configuration remains relatively stable. The resulting

mapping file [24], available in YARRRML<sup>10</sup> format, is input into the Mat-O-Lab RDFConverter tool [25], which ultimately combines all the data into a singular metadata document [26].

### Summary

The Precipitate Analysis Workflow (PAW) comprises three primary components. Initially, it establishes a connection to the Omero server to acquire image datasets and essential metadata for subsequent stages. Next, it conducts image pre-processing, optimizing the images through techniques such as rendering, grayscale conversion, and thresholding in preparation for precipitation analysis. Finally, the workflow defines objects by applying operations such as morphological processing, border clearing, and contour finding while also creating ROIs crucial for further processing in the Statistical Analysis Workflow (SAW). All the steps taken, along with the data variables used and the resources (images, ROIs), are documented using the Mat-O-Lab toolchain elements, culminating in a single metadata document [26].

### Statistical Analysis Workflow (SAW)

In the SAW the systematic evaluation of detected precipitates is a critical step. This process involves the calculation of various polygon properties, the creation of histograms, fitting log-normal distributions, and ultimately, the representation of results in a structured and semantically annotated format. This comprehensive approach ensures the reproducibility and accessibility of the analysis results. In the following, the details of the SAW are highlighted for each stage of the process for a clearer understanding (see Fig. 1) or view the source code [27].

### Polygon Properties Calculation

The first step of the SAW is the calculation of various properties of detected precipitates. This process involves the retrieval of relevant data, including image information, ROIs, and specimen characteristics like aging temperature and aging time. SPARQL<sup>11</sup> query language is leveraged for data retrieval, and the collected data is structured into a data frame for further analysis.

Additionally, another SPARQL query is executed to obtain information about the physical size of the images, specifically their length and width. This physical size information is crucial for converting the relative coordinates of precipitation polygons into real units (e.g., nanometers, nm). This conversion is essential for subsequent analysis,

<sup>5</sup> <https://docs.openmicroscopy.org/ome-model/5.6.3/developers/roi.html#roi-attributes-and-simple-children>.

<sup>6</sup> <https://www.w3.org/ns/csvw>.

<sup>7</sup> <https://json-ld.org/>.

<sup>8</sup> <https://w3c.github.io/csvw/csv2rdf/>.

<sup>9</sup> <https://draw.io>.

<sup>10</sup> <https://rml.io/yarrrml/>.

<sup>11</sup> <https://www.w3.org/TR/sparql11-query/>.



including area calculations, centroid determination, radius computation, and other properties.

Next, the shape and size of the polygons are characterized by fitting ellipses to them using a weighted covariance matrix. Various properties of the fitted ellipses, such as major and minor axes, orientation, height, and width, are computed based on the polygon coordinates and centroid. Additionally, the area (spatial extent in the image) and radius are determined for each polygon. To reduce artifacts, polygons with an area of less than 1.0 nm are filtered out.

To gain insights into the shape, size, and geometric characteristics of the precipitates, various properties are computed. The labeling function is employed to obtain precipitate characteristics. The computed values are stored alongside metadata from various PAW steps, such as the disk radius used in median filtering, in a tabular data file in CSV format.

The computed polygon features for each dataset are then organized into a data frame, and the results table is saved as a CSV file for various purposes.<sup>12</sup>

### Histogram Creation

Following the calculation of polygon properties, the SAW proceeds to create cumulative radii distribution functions based on the determined radii of S-phase precipitate polygons at different aging states (see Fig. 3). These histograms provide insights into how precipitate density changes as aging progresses. In these plots, the bars represent the number of detected precipitates normalized by the total number of detected precipitates for each aging state. The data indicates that the precipitate density decreases as aging progresses.

### Log-Normal Distribution Fitting

After creating histograms, log-normal distributions are fitted to the data. These distributions are represented as lines on the histograms and show good agreement with the experimental data. The fitting results include parameters such as the determined median radius ( $r_m$ ), geometric standard deviation ( $\sigma_{\text{geo}}$ ), standard error, and chi-square values. These results, along with the calculated average radius ( $r_a$ ) (see Sect. “Determination of Average Particle Radius” for definition), are again exported in CSV format [28].

### Create Documentation Metadata

Metadata for the results table [28] is generated analogously to the results of the PAW algorithm, as described in sect.

“Create Documentation Metadata”. The CSV file is annotated with CSVToCSVW [21], resulting in CSVW metadata [29] and transformation into RDF [30]. The creation of the mapping file [31] for the SAW knowledge graph (Fig. 6) is done with MapToMethod [23]. All data is then consolidated using RDFConverter [25] into the final RDF document [32].

In addition, the SAW reads and analyzes the original data from Rockenhäuser et al. [3, 4], reproducing the cumulative radius distribution functions with the log-normal distributions using the analysis procedure described above. The results are compared for exemplary datasets in Fig. 3, and a comparative analysis in Fig. 4 shows the average radius ( $r_a$ ) as a function of aging time for each dataset. This allows for a comparison between the automatically generated precipitation data and the results of the manual analysis.

This comprehensive analysis workflow provides a foundation for the creation and application of ontologies, as described in the next section, to semantically represent and enhance the understanding of the entire process. Moreover, the SAW’s design ensures the reproducibility of the obtained precipitation data from the ROIs without the need for additional (proprietary) software or extra tools, making it highly accessible and self-contained within the script.

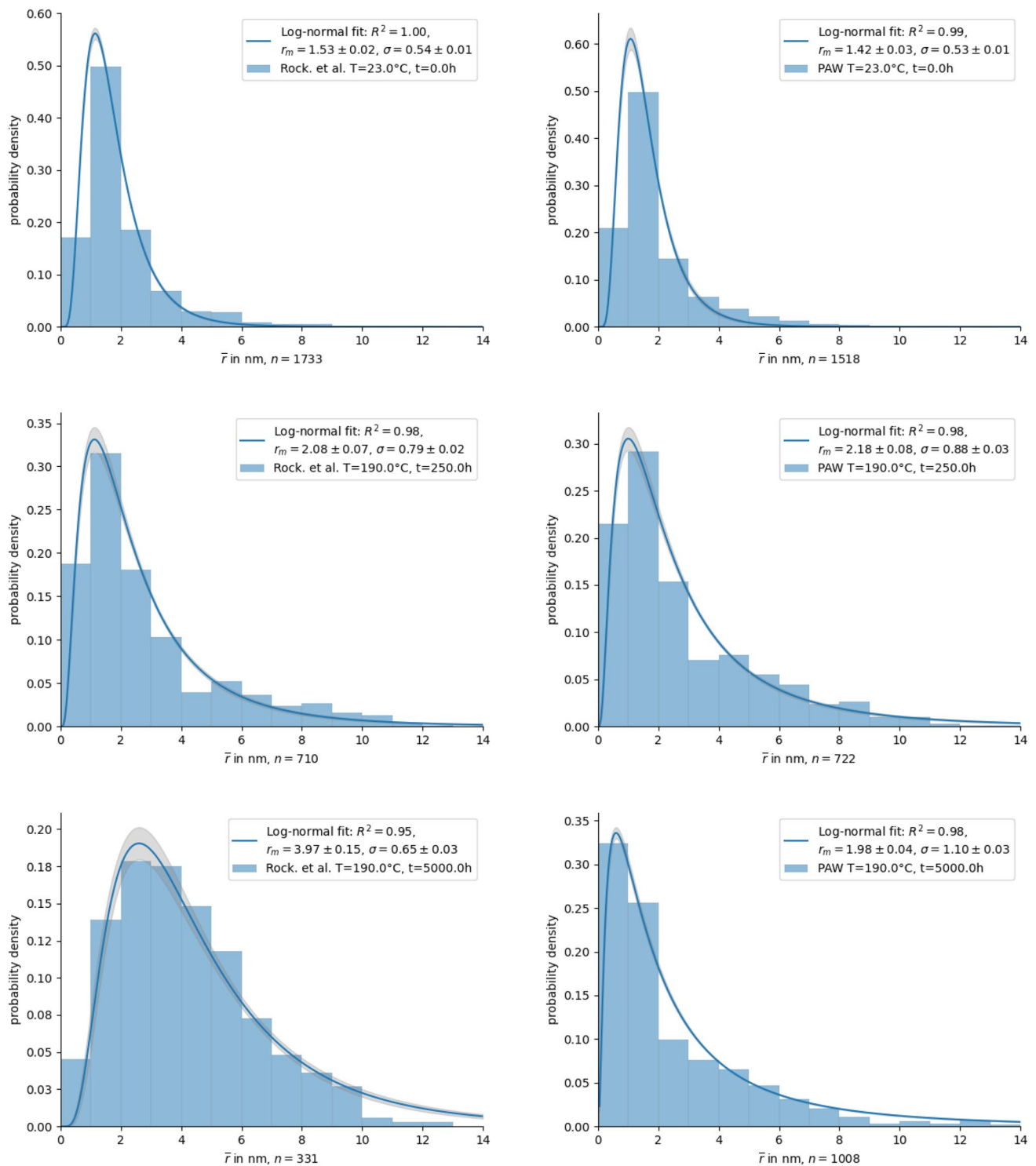
### Ontology Creation and Application

The enhancement of (meta)data for precipitate analysis and the organization of data rely on two ontologies: the Precipitate Analysis Workflow Ontology (PAWO) and the Statistical Analysis Workflow Ontology (SAWO).

The PAW is represented as a sequence of process steps, semantically modeled using the Platform MaterialDigital Core Ontology (PMDco) [33] process classes and object properties (Fig. 5). As a comprehensive mid-level ontology for the materials science and engineering domain, the PMDco facilitates semantic representations of process chains, processes, and materials (meta)data. Relationships between individuals are defined by object properties. For instance, *co:subordinateProcess* signifies subordinate steps, and *co:nextProcess* denotes the next process step in the chain. A crucial metadata aspect of the PAW is the execution date, associated with *co:characteristic*. Input data and parameters for specific sub-processes, such as median filter disk radius, threshold method, kernel size, and threshold area size, are represented with *co:input*. These value-related entities, belonging to the superclass *co:ValueObject* in the PMDco, are essential for process and result reproducibility.

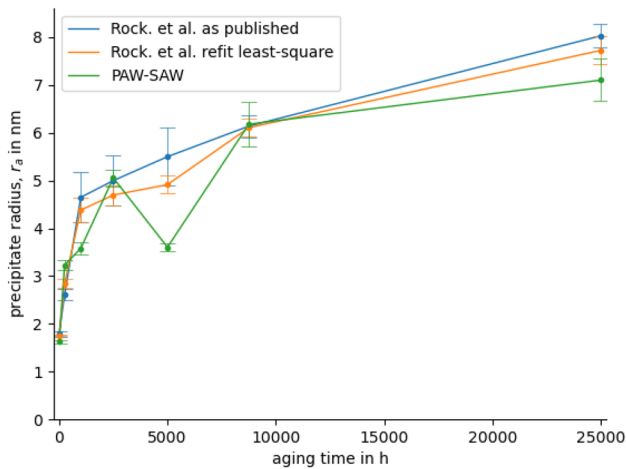
The formal representation of the SAW and its associated entities is also realized through the PMDco (Fig. 6). Different workflow steps are represented as instances of *co:AnalysingProcess*. A sub-process calculates polygon properties and generates three output value objects: Centroid, polygon area, and equivalent radius, modeled using

<sup>12</sup> <https://github.com/BAMresearch/DF-TEM-PAW/tree/main/polygons>.



**Fig. 3** Exemplary comparison of fitted radii distributions: manually obtained results by Rockenhäuser et al. [3, 4] (left) and automatically obtained results using PAW-SAW (right). While there is a close agreement in the detected number of precipitates, their sizes, and statistical values for both the T61 (top row) and the 190 °C—250 h

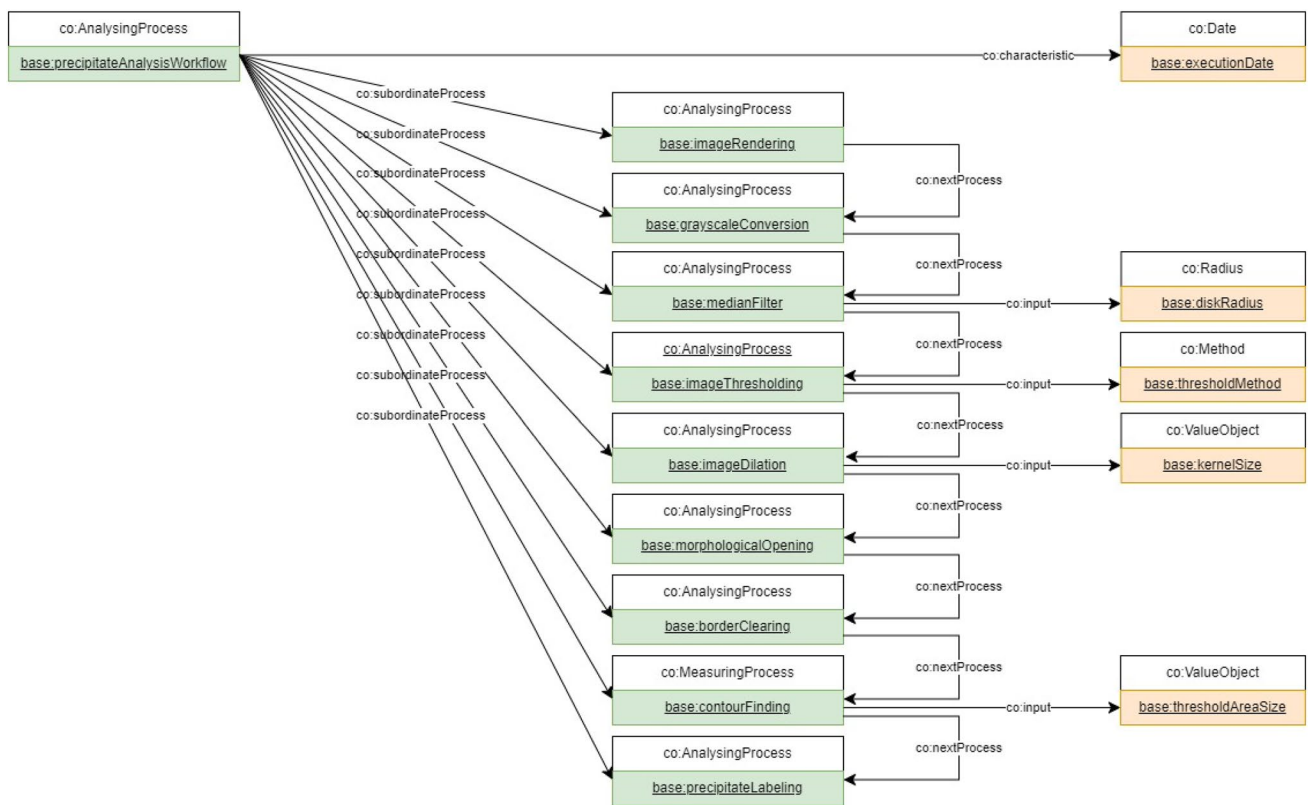
(middle row) datasets when comparing the manual and automatic methods, significant discrepancies emerge in the 190 °C—5000 h (bottom row) dataset. These variances are primarily attributed to image artefacts in the automatic analysis, which lead to an increased detection of small features (see Fig. 7)



**Fig. 4** Average precipitate radius ( $r_a$ ) as a function of aging time at a temperature of 190 °C. The lines represent the relationship between the initial T61 condition and the aged specimens, while the error bars encompass both statistical and systematic errors associated with size measurements in the DF-TEM images

the *co:output* object property. The equivalent radius also serves as input (*co:input*) for plotting the radius distribution histograms. In this sub-process step, three value objects x-max, x-min, and bin width-also act as input parameters. In the final workflow part, the log-normal fit is performed. Additional instances of the SAWO contain specific values and units, such as average radius and mean radius, to provide precise data for further statistical analysis.

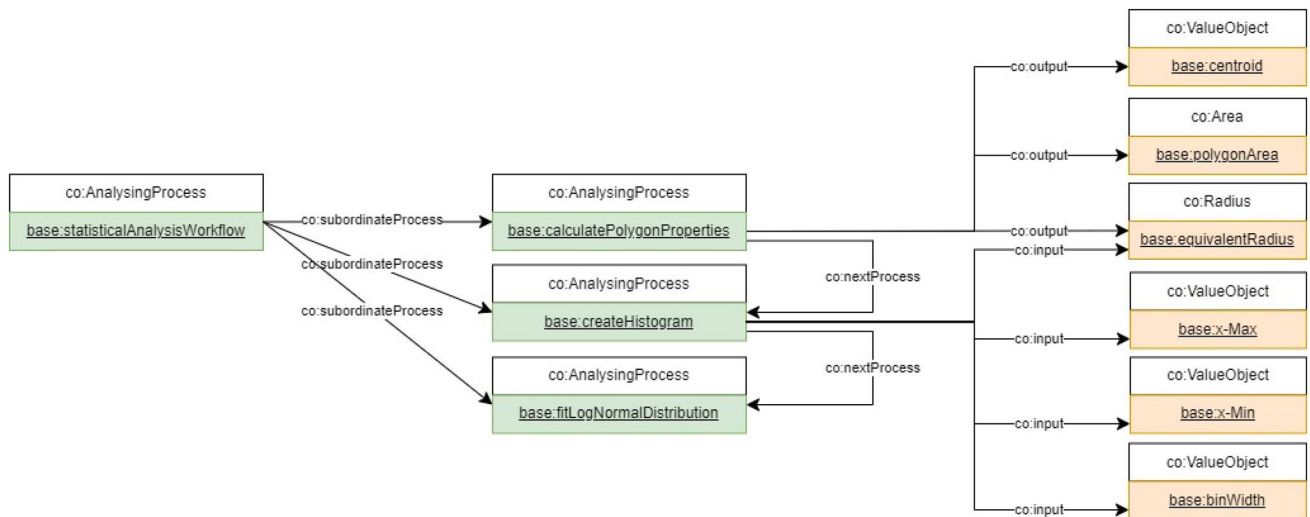
These ontologies serve as semantic representations of the analysis workflows implemented in the scripts. They define the characteristics, measurements, and processes involved in the analysis, ensuring the transformation of unstructured result data into structured RDF triples. This enhancement improves the organization and accessibility of precipitate quantification results by presenting them in a machine-readable and semantic format. This, in turn, enables smoother integration, sharing, and correlation with other datasets. To retrieve data from the RDF graph, SPARQL querying language formulations are executed within the script, facilitating the extraction of pertinent information for analysis,



**Fig. 5** Process chain modeling with the PMDco: this figure depicts the sequence of subordinate process steps within the PAW, highlighting relevant classes and properties (*co:subordinateProcess* and *co:nextProcess*) in green. Value objects and their subclasses, repre-

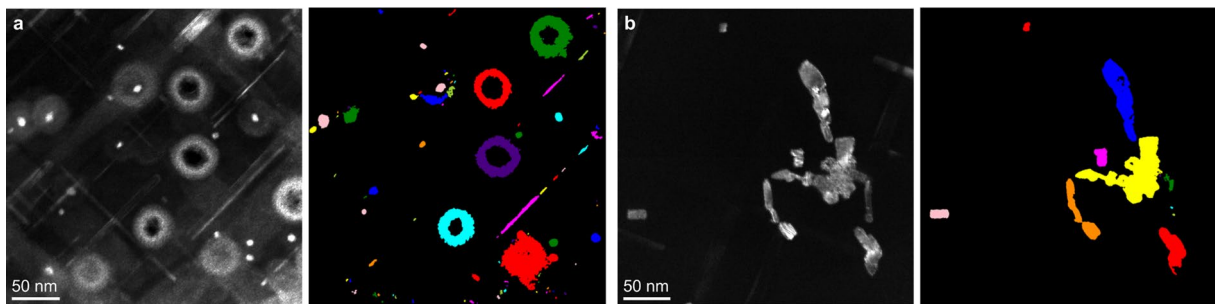
senting (meta)data, are depicted in apricot. Notably, essential meta-data, including the execution date, is included. Additionally, inputs for sub-processes are modeled to guarantee the reproducibility of the process and resulting data





**Fig. 6** SAW process chain modeling with the PMDco: the figure illustrates the sequence of sub-processes in the SAW. Outputs from sub-processes, such as the equivalent radius, serve as inputs for the

subsequent processes, facilitating the creation of histograms. This modeling approach ensures a seamless data flow and continuity throughout the workflow



**Fig. 7** Image artifacts and precipitate clusters in aged specimens: **a** illustration of brighter halo-shaped image artifacts that impact image segmentation, arising from specimen preparation. **b** Formation of precipitate clusters, observed after an extended aging duration

including image features, polygon properties, and fitting results.

The incorporation of ontologies and semantic representations makes a substantial contribution to the systematic analysis of S-phase precipitates and facilitates the linking and comparison of results. In the forthcoming discussion section, we delve into the validation of our results and assess areas where further improvement is possible.

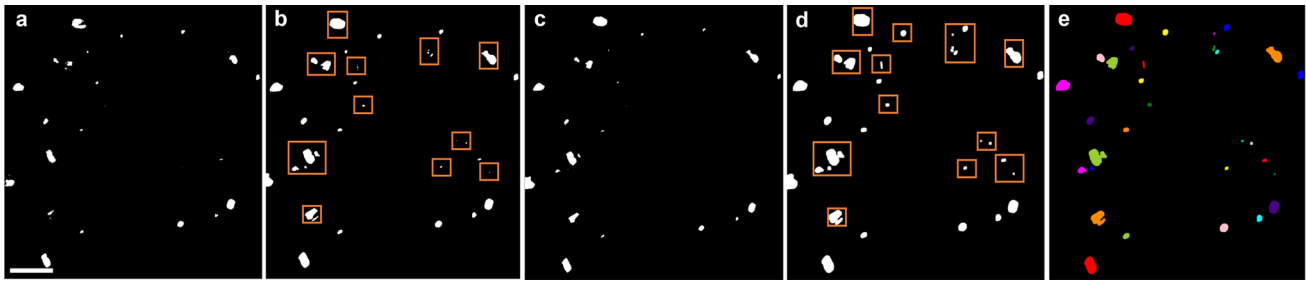
## Discussion

### Results Validation

Our integrated image analysis pipeline, which incorporates both the PAW (described in Sect. “[Precipitate Analysis Workflow \(PAW\)](#)”) and SAW (outlined in Sect. “[Statistical Analysis Workflow](#)”) digital workflows, effectively

detects S-phase precipitates and accurately determines their radii. We have achieved a noteworthy level of agreement when comparing the automatically generated results with those obtained manually by Rockenhäuser et al., especially for datasets such as T61 and aging times of 250, 1000, and 2500 h at 190 °C. This underlines the consistency and reliability of our automated analysis pipeline in these specific cases, affirming its capability for precise characterization.

However, challenges arise when analyzing datasets with extended aging times, specifically the 5000 and 25,000 h aging times at 190 °C. These challenges primarily stem from image artifacts resulting from TEM specimen preparation (as illustrated in Fig. 7a) and the formation of precipitation clusters (as seen in Fig. 7b). These complexities introduce intricacies into the quantitative analysis, leading to discrepancies in average precipitate radii ( $r_a$ ) specific instances. These challenges emphasize the limitations of



**Fig. 8** Comparison of thresholding techniques and morphological operations for precipitation detection: comparison between **a** adaptive and **b** Otsu thresholding methods. The automated threshold method (Otsu) consistently provides more reliable results, retaining a higher number of precipitates. Effects of **c** erosion and **d** dilation operations in precipitate detection. Erosion leads to the loss of precipitates due

to the removal of boundary pixels (highlighted in orange squares), while dilation is preferred for retaining precipitates and enhancing their visibility. **e** The final output image after applying the selected thresholding and morphological operations (inserted scale bar in **a** is 50 nm for reference)

automated analysis when confronted with intricate image artifacts and evolving precipitate behaviors.

Our image analysis process for the DF-TEM image datasets (outlined in Table 2 entails the detection of S-phase precipitates represented as polygons for each aging state. This process involves up to two specimens and multiple images. Subsequently, the SAW is used to determine the precipitate radii. The imaged precipitates are approximated as cylindrical rods with a radius ( $r$ ), but determining the precise length ( $l$ ) proves challenging due to low contrast. Consequently, our quantitative analysis focuses on the precipitate radius, presented as cumulative radii distributions normalized by the total number of detected precipitates in each aging time.

To analyze these radii distributions, we employ a log-normal distribution as the fitting function, characterized by the particle radius ( $r$ ), the median particle radius ( $r_m$ ), and the geometric standard deviation ( $\sigma_{\text{geo}}$ ). Detailed information about this methodology can be found in [34], Sect. “**Determination of Average Particle Radius**” and in studies by Rockenhäuser et al. [3, 4]. Figure 3 illustrates the comparison between the fitted radii distributions of the automatically obtained PAW-SAW results (right side) and the manually obtained results by Rockenhäuser et al. (left side).

While the comparison between the manually and automatically generated results demonstrates strong agreement for the T61 (i.e.  $t = 0$  h), 190 °C—250 h, 190 °C—1000 h, and 190 °C—2500 h datasets, it becomes more challenging for the 190 °C—5000 h and 190 °C—25,000 h datasets due to specific reasons. In the former case, the image data contains numerous artifacts generated during the TEM specimen preparation, particularly during the electropolishing process (see Fig. 7a). In the latter case, the observed discrepancy is attributed to the formation of precipitate clusters after very long aging times (see Fig. 7b). Rockenhäuser et al. overcame these artifacts through manual processing in the image analysis, which required additional effort.

Furthermore, we compared the values of  $r_a$ , derived from  $r_m$  and  $\sigma_{\text{geo}}$  using Eq. (2), with those determined by Rockenhäuser et al. in Fig. 4. The blue curve shows the data of Rockenhäuser et al. using the log-normal fit of the measured radius distributions, the orange curve represents the result of a least square fit, and the green curve represents the results of the automated evaluation. As mentioned earlier, the agreement is lower for the 5,000 h and 25,000 h aging times than for the other times due to the artifacts described above. However, the trend is accurately represented.

## Room for Improvement

Our research into automatable DF-TEM image analysis for characterizing S-phase precipitation in aluminum alloy 2618A has yielded promising results and valuable insights. The development and validation of our pipeline, consisting of PAW and SAW, have demonstrated its reliability and reproducibility in obtaining accurate data from image signals. By comparing our results with manually obtained data, we have established its efficacy in delivering consistent outcomes, further reinforcing its utility for precise characterization.

However, despite the significant progress achieved with our automated image analysis pipeline, certain challenges persist when processing challenging raw image data, as shown in Fig. 7. Overlapping precipitates, variations in their orientations within the aluminum matrix, or image artifacts generated during specimen preparation and TEM imaging present difficulties for automatic detection by the PAW. In contrast, Rockenhäuser et al. addressed these challenges through time-consuming manual interventions during image processing.

As demonstrated in Fig. 4, the reevaluation of the original dataset created by Rockenhäuser et al. resulted in differing characteristic radius evaluations. This discrepancy can be attributed to the use of different minimization algorithms

in the fitting routine. The reevaluation employed least-square fitting using the Levenberg–Marquardt algorithm, while the authors of the Rockenhäuser publications used an undocumented algorithm. This highlights the importance of robust data documentation to reduce effort and enhance reproducibility.

In our future research endeavors, we will focus on leveraging machine learning (ML) to enhance the analysis of S-phase precipitation detection in DF-TEM imaging. Our plan is to develop an ML model trained with a dataset of annotated DF-TEM images, incorporating details such as the size and shape of the precipitates. The ultimate goal is to enable the model to identify and measure precipitates in captured images, making the analysis process more efficient. Furthermore, the possibilities of employing deep learning techniques for precipitation detection analysis will be explored. Deep learning, a subset of ML, excels at finding patterns in data and often provides superior predictive performance compared to traditional ML algorithms [35]. We believe that deep learning algorithms could reveal insights into precipitation detection analysis that may surpass what conventional ML algorithms can achieve. Therefore, we intend to investigate and apply these methods in this context.

To enhance the performance of the PAW, we have evaluated various techniques, including adaptive thresholding and the Otsu method for precipitate detection. The Otsu method consistently provided more accurate and reliable results compared to adaptive thresholding (see Fig. 8a, b), leading us to adopt it for further analysis. We also examined the effects of erosion and dilation during morphological operations. While erosion may be useful in specific scenarios, we found that dilation offered greater advantages for our purposes. Dilation allowed us to capture more precipitates and improve their visibility without losing valuable information (see Fig. 8c, d). The combination of the Otsu method for thresholding and dilation for morphological operations is shown in Fig. 8e. This output demonstrated significant improvements in precipitate detection and the highlighting of their features for further analysis.

Nevertheless, there is room for further improvement. Optimizing the PAW through the application of ML approaches to process image data with high accuracy and reliability [36] is a desirable objective. This could lead to more precise insights into changes in precipitate shape with increasing aging time, resulting in a better understanding of the actual form and morphology of precipitates. Currently, precipitates are approximated as circular, but ML techniques could capture their true shapes.

In addition to our technical achievements, we have successfully addressed the critical issue of data reusability by implementing a comprehensive pipeline concept. Our pipeline seamlessly integrates Omero as an image repository, incorporating metadata and a semantic representation of

PAW and SAW using well-defined ontologies. This approach has significantly improved data interoperability and accessibility, aligning with FAIR (Findable, Accessible, Interoperable, Reuseable) principles and promoting open science practices. By contributing to the advancement of scientific research as a whole, we believe our work has a positive and lasting impact.

To further enhance the developed ontologies, we recognize the importance of making the generated semantic data even more interoperable. One promising avenue for achieving this goal involves integrating the ongoing development of the Electron Microscopy Glossary.<sup>13</sup> In this initiative, the scientific community collaboratively contributes relevant class definitions. By incorporating community-driven efforts and feedback, we aim to continually improve the effectiveness and utility of our ontologies, fostering broader adoption and impact in the scientific community.

Our work demonstrates the potential for FAIR research data management and scalability through the incorporation of collaborative tools, automated digital workflows, and Semantic Web technologies. By transparently publishing the Jupyter Notebooks and ontologies, we have made our approach readily adoptable by the scientific community, encouraging reproducibility and fostering a more accessible scientific landscape.

In conclusion, this work offers valuable contributions to the materials science community, facilitating the extraction of specialized information from complex datasets and enabling extensive analysis. We firmly believe that our work will positively impact the adoption of automated analysis techniques and semantic representations, leading to enhanced scientific exploration and knowledge dissemination within the broader research community. The combination of technical advancements and the promotion of open and collaborative practices will drive innovation and pave the way for more sophisticated materials research in the future.

## Methods

### Sample Material States

The coarsening processes of the S-phase in aluminum alloy 2618A at high temperatures were investigated in a previous study by Rockenhäuser et al. [3]. In this study, samples were aged at 190 °C for up to 25,000 h, and the microstructure was extensively characterized afterward. The T61 initial condition included a solution annealing at 530 °C for 8 h followed by quenching in boiling water and aging at 195 °C for 28 h.

<sup>13</sup> <https://helmholtz-metadaten.de/de/em-glossary-initiative>.

**Table 2** Overview of DF-TEM image datasets used in this study

| ID | Datasets                  | Specimens | Images |
|----|---------------------------|-----------|--------|
| 1  | T61                       | 2         | 12     |
| 2  | 190°C_250h_state_DF_TEM   | 1         | 11     |
| 3  | 190°C_1000h_state_DF_TEM  | 1         | 12     |
| 4  | 190°C_2500h_state_DF_TEM  | 2         | 17     |
| 5  | 190°C_5000h_state_DF_TEM  | 2         | 23     |
| 6  | 190°C_8760h_state_DF_TEM  | 2         | 19     |
| 7  | 190°C_25000h_state_DF_TEM | 2         | 21     |

## TEM Imaging

The images presented and analyzed in this work were produced using a JEM-2200FS TEM with a field-emission gun operating at 200 kV. The S-phase precipitates form as rods in three orthogonal directions ([001], [010], and [100]) within the Al-matrix. Consequently, TEM samples were oriented in the [001] direction for DF-TEM investigations. Figure 1 in [3] shows a selected area diffraction pattern of the sample area with clearly visible reflections of the oriented Al-matrix (bright spots). The rod-shaped precipitates induce thin, weak lines (called streaks) in between the bright matrix reflections. An aperture, as indicated by the white circle in the figure, was employed to isolate these streaks, thereby imaging solely the precipitates, excluding the matrix. This aperture configuration ensures the capture of all three precipitate orientations, thus rendering them visible in the DF-TEM image. In these images, particularly on the far left side of Fig. 1 in this manuscript, spots exhibiting bright contrast correspond to precipitate rods oriented along the [001] direction, penetrating the image plane (upper DF-TEM image, T61). The lath-shaped contrasts observed in the bottom DF-TEM image (190 °C—1000 h) result from rods oriented orthogonally to the incident electron beam, i.e., their rod axis lies parallel to the image plane. Notably, rods aligned along the [001] direction, with their axis orthogonal to the image plane, exhibit more pronounced contrast compared to those oriented orthogonally to the [001] beam direction. This distinction arises because the electron beam traverses the entire length of the former, while it only crosses the diameter of the latter, thereby generating less contrast.

The DF-TEM images were acquired manually. Ensuring reliable and correct imaging conditions depends on the skills and experience of the microscope operator. The contrast in DF-TEM images can depend on the bending of the sample and the sample thickness. This is also the case in the alloy studied but was not overly pronounced. To minimize this effect, the TEM foil was carefully aligned, and images were taken in areas of sufficiently high magnification to ensure uniform local thickness and orientation.

To quantitatively evaluate a minimum of 300 precipitates per material state, we analyzed up to 23 images representing different specimen positions (see Table 2).

Each dataset corresponds to a specific material state, and for each state, a sufficient number of S-phase precipitates were imaged using the DF-TEM technique. Since the rod-shaped S-phase forms in three perpendicular crystallographic directions, an easy-to-analyze direction was (preferably) imaged (i.e., one rod axis perpendicular to the image plane). The S-phase then appears bright against a dark background. For simplicity, the precipitates were assumed to be cylindrical in the analysis and the radius associated with the area was determined. The DF-TEM images were available in the dm3 data format, which is a 16-bit raster image format for electron microscopy images. It contains metadata about the TEM process itself, such as the CCD camera of the microscope, the exposure time, and a timestamp. For more detailed information on the materials, methods, and software-based image analysis, please refer to Rockenhäuser et al. [3, 4]. The corresponding dataset was published on Zenodo [5].

## Determination of Average Particle Radius

Particle sizes often follow a log-normal distribution [37], and they are described by the following equation [34]:

$$n(r) = \frac{1}{\sqrt{2\pi}r\sigma_{\text{geo}}} \exp\left(-\frac{\left(\ln\left(\frac{r}{r_m}\right)\right)^2}{2\sigma_{\text{geo}}^2}\right) \quad (1)$$

Here,  $n(r)$  represents the normalized number of precipitates per radius size class,  $r$  is the particle radius,  $r_m$  is the median particle radius, and  $\sigma_{\text{geo}}$  is the geometric standard deviation. This equation effectively fits the measured radii distributions, where measured values are normalized to the total number of precipitates. In the context of a log-normal distribution, it is essential to note that the average radius  $r_a$  is not the same as the mean radius. Instead, it is calculated using the following equation [38]:

$$r_a = r_m \exp\left(\frac{\sigma_{\text{geo}}^2}{2}\right) \quad (2)$$

These average radii ( $r_a$ ) are subsequently plotted against aging time. The data can also be utilized to describe coarsening processes using suitable models [e.g., Eq. (3)]. Detailed evaluations and analyses of these processes were performed and discussed in previous publications [3, 4].

## Modeling Particle Coarsening

The coarsening, also referred to as Ostwald ripening, of the S-phase (i.e., its evolution over time) can be quantitatively described using various models. The first quantitative equation for spherical precipitation was introduced by Lifshitz and Slyozov [39] and Wagner [40]:

$$r^3 - r_0^3 = k(t - t_0) \quad (3)$$

In this equation,  $r$  represents the mean radius at time  $t$ ,  $r_0$  is the mean radius at the start of coarsening ( $t_0$ ),  $t$  is the coarsening time,  $t_0$  is the start time of coarsening, and  $k$  is a constant. It is important to note that certain assumptions underlie this model, and these are elaborated upon in previous publications [3, 4].

## Jupyter Notebook Environment and Python Packages

The digital workflows for image analysis and data statistics are developed and created within the Jupyter Notebook environment [6], utilizing various Python libraries.<sup>14</sup> For image processing and manipulation, OpenCV<sup>15</sup> was employed, which provides functions for all the tasks mentioned. The NumPy<sup>16</sup> and Pandas<sup>17</sup> data frames are utilized for performing numerical operations. Additionally, functionalities offered by the scikit-image<sup>18</sup> library are leveraged.

The Omero Python package<sup>19</sup> is integrated for interaction with the Omero server, enabling access and retrieval of image data. Additionally, *shapely*,<sup>20</sup> Python package for computational geometry, used to handle the polygon data. Shapely supports the analysis of planar geometric objects, facilitating tasks such as finding polygon properties.

For the analysis of precipitates and the generation of distribution plots, the Python package (*lmfit*) for *Non-Linear Least-Squares Minimization and Curve Fitting*<sup>21</sup> is employed. Lmfit offers a convenient interface for addressing nonlinear optimization and curve fitting problems, allowing for parameter definition, model construction, and the

execution of fitting processes. In this context, Lmfit is used for fitting log-normal distributions to the histogram data.

Visualizations, including bar charts and error bars, are created using the matplotlib<sup>22</sup> plotting library. Additionally, the rdflib<sup>23</sup> package is employed for RDF data handling and graph-based representations, which are essential for semantic data integration.

## Image Repository

The Open Microscopy Environment image repository<sup>24</sup> is employed to store and organize DF-TEM images using its server application. Omero supports a wide variety of different file formats and provides static Uniform Resource Locator (URL) for each image. Data from this secure, central repository can be accessed through the Flask REST API service.<sup>25</sup> Omero's functionality also enables the processing of created precipitation contours for quantitative analysis.

## Ontology Creation and Application

Process graphs for the semantic representation of the PAW (Sect. “*Precipitate Analysis Workflow (PAW)*”) and the SAW (Sect. “*Statistical Analysis Workflow (SAW)*”) are created using Ontopanel and its plugins [41]. The PROV Ontology (PROV-O)<sup>26</sup> and PMD Core Ontology (PMDco) [33]<sup>27</sup> serve as upper-level ontologies. Entities from the QUDT<sup>28</sup> and OME<sup>29</sup> are also incorporated. These created ontologies are leveraged in Jupyter Notebook environments to perform local SPARQL<sup>30</sup> queries for subsequent data processing.

**Acknowledgements** B.B., B.S., and T.H. thank the German Federal Ministry of Education and Research (BMBF) for financial support of the project Innovation-Platform MaterialDigital ([www.materialdigital.de](http://www.materialdigital.de)) through project funding FKZ no: 13XP5094E (BAM) and 13XP5119A-G (KupferDigital).

Additional funding provided by Bundesanstalt für Materialforschung und -prüfung (BAM) and Fraunhofer Group Materials and Components (MATERIALS) in the context of the Mat-O-Lab project is gratefully acknowledged.

The authors would like to thank Jürgen Olbricht for the active exchange and valuable discussions.

**Author Contributions** G.Z. was involved in the conceptualization, design, and technical implementation of the PAW and SAW. He also played a significant role in creating visualizations and contributed significantly to the writing, review, and editing of the manuscript. T.H.

<sup>14</sup> <http://www.python.org>.

<sup>15</sup> <https://opencv.org/>.

<sup>16</sup> <https://numpy.org/>.

<sup>17</sup> <https://pandas.pydata.org>.

<sup>18</sup> <https://scikit-image.org/>.

<sup>19</sup> <https://docs.openmicroscopy.org/omero/5.6.0/developers/Python.html>.

<sup>20</sup> <https://shapely.readthedocs.io/en/stable/manual.html>.

<sup>21</sup> <https://lmfit.github.io/lmfit-py/>.

<sup>22</sup> <https://matplotlib.org/>.

<sup>23</sup> <https://rdflib.readthedocs.io/en/stable/index.html>.

<sup>24</sup> <https://www.openmicroscopy.org>.

<sup>25</sup> <https://flask.palletsprojects.com/en/2.2.x>.

<sup>26</sup> <https://www.w3.org/TR/prov-o/>.

<sup>27</sup> <https://github.com/materialdigital/core-ontology/>.

<sup>28</sup> <https://github.com/qudt/qudt-public-repo>.

<sup>29</sup> <https://gitlab.com/openmicroscopy/incubator/ome-owl>.

<sup>30</sup> <https://www.w3.org/TR/sparql11-query/>.



contributed to the conceptualization, design, and technical implementation of both the PAW and SAW. Additionally, he made important contributions to ontology creation for the work and visualization tasks. B.S. provided valuable input during the conceptualization phase and made substantial contributions to writing, reviewing, and editing the manuscript. C.V. actively participated in the conceptualization process, providing valuable ideas. B.B. provided valuable supervision and guidance throughout the research work, encompassing the conceptualization, design, and technical implementation of the PAW and SAW. He played a key role in ontology creation, and visualization tasks, and contributed significantly to writing, reviewing, and editing the manuscript. All authors have read and approved the final manuscript.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Code and data availability** All underlying code and all datasets used and analyzed during this research work are available and can be accessed at this public code repository [42]. All TEM Images used are published on a public Omero Image Service [43]. The source code of the tools used for metadata transformation and annotation are available here [21, 23, 25].

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Polmear IJ (2006) Light alloys—from traditional alloys to nanocrystals. Oxford Elsevier/Butterworth-Heinemann
- Rockenhäuser C, von Hartrott P, Skrotzki B (2023) Brinell-hardness data (HBW 2.5/62.5) of aluminum alloy EN AW-2618A after different aging times and temperatures. *Data Br* 46:108830. <https://doi.org/10.1016/j.dib.2022.108830>
- Rockenhäuser C, Schriever S, van Hartrott P, Piesker B, Skrotzki B (2018) Comparison of long-term radii evolution of the S-phase in aluminum alloy 2618A during ageing and creep. *Mater Sci Eng, A* 716:78–86. <https://doi.org/10.1016/j.msea.2018.01.033>
- Rockenhäuser C et al (2021) On the long-term aging of S-phase in aluminum alloy 2618A. *J Mater Sci* 56:8704–8716. <https://doi.org/10.1007/s10853-020-05740-x>
- Rockenhäuser C, Skrotzki B (2023) Radii of S-phase Al<sub>2</sub>CuMg in Al-alloy EN AW-2618A after different aging times at 190°C. Zenodo. <https://doi.org/10.5281/zenodo.7625259>
- Kluyver T et al (2016) Jupyter Notebooks—a publishing format for reproducible computational workflows. IOS Press
- Russ JC, Neal FB (2016) The image processing handbook, 7th edn. CRC Press
- Duval L, et al (2014) Image processing for materials characterization: issues, challenges and opportunities. In: 2014 IEEE international conference on image processing (ICIP), pp 4862–4866. <https://doi.org/10.1109/ICIP.2014.7025985>
- Yang L et al (2021) High-throughput methods in the discovery and study of biomaterials and materiobiology. *Chem Rev* 121:4561–4677. <https://doi.org/10.1021/acs.chemrev.0c00752>
- Pelkie BG, Pozzo LD (2023) The laboratory of babel: highlighting community needs for integrated materials data management. *Digit Discov* 2:544–556. <https://doi.org/10.1039/D3DD00022B>
- Wilkinson MD et al (2016) The FAIR guiding principles for scientific data management and stewardship. *Sci Data* 3:1–9. <https://doi.org/10.1038/sdata.2016.18>
- Bayerlein B et al (2022) A perspective on digital knowledge representation in materials science and engineering. *Adv Eng Mater* 24:2101176. <https://doi.org/10.1002/adem.202101176>
- Valdestilhas A, Bayerlein B, Moreno Torres B, Jan Zia G-A, Muth T (2023) The intersection between semantic web and materials science. *Adv Intell Syst* 5:2300051. <https://doi.org/10.1002/aisy.202300051>
- Zhou J, Hong X, Jin P (2019) Information fusion for multi-source material data: progress and challenges. *Appl Sci* 9:1–18. <https://doi.org/10.3390/app9173473>
- Soedarmadji E, Stein HS, Suram SK, Guevarra D, Gregoire JM (2019) Tracking materials science data lineage to manage millions of materials experiments and analyses. *NPJ Comput Mater* 5:1–9. <https://doi.org/10.1038/s41524-019-0216-x>
- Hanke T (2023) Python script to interact with Omero server API. [https://github.com/BAMresearch/DF-TEM-PAW/blob/main/omero\\_tools.py](https://github.com/BAMresearch/DF-TEM-PAW/blob/main/omero_tools.py)
- Jan Zia G-A, Hanke T (2023) Jupyter notebook running the PAW. [https://github.com/BAMresearch/DF-TEM-PAW/blob/main/PAW\\_precipitate\\_detection.ipynb](https://github.com/BAMresearch/DF-TEM-PAW/blob/main/PAW_precipitate_detection.ipynb)
- Kanan C, Cottrell GW (2012) Color-to-grayscale: Does the method matter in image recognition? *PLoS One* 7:1–7. <https://doi.org/10.1371/journal.pone.0029740>
- Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern* 9:62–66
- CSV (2023) Table containing variables and resource locations used by the PAW. [https://github.com/BAMresearch/DF-TEM-PAW/blob/main/detection\\_runs.csv](https://github.com/BAMresearch/DF-TEM-PAW/blob/main/detection_runs.csv)
- Kröcker B, Fechner R, Hanke T (2023) Generates JSON-LD for various types of CSVs, it adopts the vocabulary provided by W3C at CSVW to describe structure and information within also uses QUDT units ontology to lookup and describe units. <https://github.com/Mat-O-Lab/CSVToCSVW>
- CSVW (2023) metadata document containing the result table of the PAW in turtle format. [https://github.com/BAMresearch/DF-TEM-PAW/blob/main/detection\\_runs.ttl](https://github.com/BAMresearch/DF-TEM-PAW/blob/main/detection_runs.ttl)
- Hanke T (2023) Tool to map content of JSON-LD documents (for example output of CSVToCSVW (<https://github.com/mat-o-lab/csvtocsvw>) to entities in knowledge graphs by creating mapping rules with conditions in YARRRML (<https://rml.io/yarrml>) format). <https://github.com/Mat-O-Lab/MapToMethod>
- YARRRML (2023) mapping file linking result table records to PAW graph representation. [https://github.com/BAMresearch/DF-TEM-PAW/blob/main/detection\\_runs-map.yaml](https://github.com/BAMresearch/DF-TEM-PAW/blob/main/detection_runs-map.yaml)
- Kröcker B, Valdestilhas A, Fellenberg H, Hanke T (2023) It is a service for joining and converting meta data documents based on YARRRML mapping files to RDF, optionally a validation can be conducted using SHACL shapes. <https://github.com/Mat-O-Lab/RDFConverter>
- Joined RDF (2023) dataset containing all records of the PAW algorithm enriched by knowledge graph representing the steps taken in the algorithm. [https://github.com/BAMresearch/DF-TEM-PAW/blob/main/detection\\_runs-joined.ttl](https://github.com/BAMresearch/DF-TEM-PAW/blob/main/detection_runs-joined.ttl)

27. Jan Zia G-A, Hanke T (2023) Jupyter Notebook running the statistical analysis algorithm. [https://github.com/BAMresearch/DF-TEM-PAW/blob/main/analyse\\_precepitates\\_data.ipynb](https://github.com/BAMresearch/DF-TEM-PAW/blob/main/analyse_precepitates_data.ipynb)
28. CSV (2023) Table output of the SAW containing variables and resource locations used to fit the histograms as well as the fit parameters and statistics. [https://github.com/BAMresearch/DF-TEM-PAW/blob/main/fit\\_results.csv](https://github.com/BAMresearch/DF-TEM-PAW/blob/main/fit_results.csv)
29. CSVW (2023) Metadata document explaining the result table of the SAW. [https://github.com/BAMresearch/DF-TEM-PAW/blob/main/fit\\_results-metadata.json](https://github.com/BAMresearch/DF-TEM-PAW/blob/main/fit_results-metadata.json)
30. CSVW (2023) metadata document containing the result table of the SAW in turtle format. [https://github.com/BAMresearch/DF-TEM-PAW/blob/main/detection\\_runs.ttl](https://github.com/BAMresearch/DF-TEM-PAW/blob/main/detection_runs.ttl)
31. YARRRML (2023) mapping file linking result table records to SAW graph representation. [https://github.com/BAMresearch/DF-TEM-PAW/blob/main/fit\\_results-map.yaml](https://github.com/BAMresearch/DF-TEM-PAW/blob/main/fit_results-map.yaml)
32. Joined (2023) RDF dataset containing all records of the SAW enriched by knowledge graph representing the steps taken there. [https://github.com/BAMresearch/DF-TEM-PAW/blob/main/fit\\_results-joined.ttl](https://github.com/BAMresearch/DF-TEM-PAW/blob/main/fit_results-joined.ttl)
33. Bayerlein B et al (2023) PMD Core Ontology: achieving semantic interoperability in materials science. *Mater Des* 237:112603. <https://doi.org/10.1016/j.matdes.2023.112603>
34. Du Q, Holmedal B, Friis J, Marioara CD (2016) Precipitation of non-spherical particles in aluminum alloys part II: numerical simulation and experimental characterization during aging treatment of an Al-Mg-Si alloy. *Metall Mater Trans A* 47:589–599. <https://doi.org/10.1007/s11661-015-3196-6>
35. Moraux A, Dewitte S, Cornelis B, Munteanu A (2021) A deep learning multimodal method for precipitation estimation. *Remote Sens*. <https://doi.org/10.3390/rs13163278>
36. Banko L, Maffettone PM, Naujoks D, Olds D, Ludwig A (2021) Deep learning for visualization and novelty detection in large X-ray diffraction datasets. *NPJ Comput Mater* 7:104. <https://doi.org/10.1038/s41524-021-00575-9>
37. Underwood EE (1973) Quantitative stereology for microstructural analysis. Springer, pp 35–66
38. Corlett WJ (1957) The lognormal distribution, with special reference to its uses in economics. *J R Stat Soc Ser C* 6:228–230. <https://doi.org/10.2307/2985613>
39. Lifshitz IM, Slyozov VV (1961) The kinetics of precipitation from supersaturated solid solutions. *J Phys Chem Solids* 19:35–50. [https://doi.org/10.1016/0022-3697\(61\)90054-3](https://doi.org/10.1016/0022-3697(61)90054-3)
40. Wagner C (1961) Theorie der Alterung von Niederschlägen durch Umlösen (Ostwald-Reifung). *Z Elektrochem, Ber Bunsenges Phys Chem* 65:581–591. <https://doi.org/10.1002/bbpc.19610650704>
41. Chen Y et al (2022) Ontopanel: a tool for domain experts facilitating visual ontology development and mapping for FAIR data sharing in materials testing. *Integr Mater Manuf Innov* 11:545–556. <https://doi.org/10.1007/s40192-022-00279-y>
42. Jan Zia G-A, Bayerlein B, Hanke T (2023) Code repository of the PAW and SAW. <https://github.com/BAMresearch/DF-TEM-PAW>
43. Jan Zia G-A, Bayerlein B, Hanke T (2023) OMERO project containing the TEM images. <https://omero.matolab.org/webclient/?show=project-51>