

Feature I

Das Knipping-Projekt, Teil III

Valeria Jana Schwanitz und August Wierling

Einführung

Dies ist der dritte und letzte Teil eines Features über das Knipping-Projekt, das handschriftliche Wetteraufzeichnungen aus Japan digitalisiert, vereinheitlicht und wissenschaftlich einordnet. Die Daten aus den Jahren 1872–1878 sammelte das OAG-Mitglied Erwin Knipping (1844–1922). Als Pionier der frühen Meteorologie in Japan stellten wir ihn bereits ausführlich in Teil I¹ vor. Über Jahre hinweg beobachtete er das Wetter mit den damaligen Instrumenten, dokumentierte akribisch Methoden, Einheiten und Messungenauigkeiten. Teil II² beleuchtete die heutige Nutzung seiner Daten, den Einfluss des Klimawandels auf Tokyo und Prognosen bis 2100. Nun zeigt Teil III, wie moderne Datenwissenschaft vorgeht, um historische Datensätze wie die Knippings zugänglich zu machen.

Wie digitalisierte Daten Wissen über die Zeit tragen – am Beispiel der historischen Daten von Knipping

Tabellen als kulturelle Errungenschaft

Betrachtet man die Tabellen, die Knipping zur Dokumentation seiner meteorologischen Daten nutzte, erkennt man schnell: Es handelt sich um monatliche Einträge (siehe Abb. 1). Die Überschrift nennt Monate wie Juli, August, September, und die Spalten tragen Beschriftungen wie in Zeile 12: „Die Zahl der Gewitter war ...“. Sofort wird klar, was die Einträge in den Zellen der Tabelle bedeuten. Man versteht intuitiv, warum einige Elemente hervorgehoben sind – durch Kursivdruck, Fettdruck und Unterstreichungen – oder warum Symbole verwendet werden, etwa Punkte oder Anführungszeichen. Solch eine tabellarische Ordnung kann man ohne Übertreibung als kulturelle Errungenschaft bezeichnen. Sie hat sich über Jahrhunderte entwickelt, und viele Standards sind längst selbstverständlich. Man braucht kaum darüber nachzudenken. Neben den Zellinhalten finden sich ergänzende Überschriften, Fußnoten und Randnotizen. So lässt sich eine Fülle an Wissen vermitteln, ohne viel erklären zu müssen. Die meisten verstehen etwa sofort, was ein kleines „ten“ hinter einer Zahl wie „20“ bedeutet.

1 OAG-Notizen 01/2026, S. 34–44. Kostenfrei einzusehen unter: <https://oag.jp/books/notizen-januar-2026/>

2 OAG-Notizen 03/2026, S. 10–19. Kostenfrei einzusehen unter: <https://oag.jp/books/notizen-maerz-2026/>

Die Digitalisierung historischer Aufzeichnungen verlangt, dass Mensch und Computer dieselbe Sprache sprechen. Die französische Künstlerin Barbara Bellier hat dies in der Abbildung 2 anschaulich dargestellt. Zunächst gilt es zu verstehen, welche Informationen Knipping damals vermitteln wollte. Konkret: Was wollte er sagen? Und wie haben die Leser seiner Zeit das verstanden? Wissen richtet sich immer an eine Zielgruppe, deren Hintergrund und Denkweise bestimmen, wie die Information präsentiert wird. Knipping sprach vor allem Meteorologen an – seine Hauptadressaten. Seine Beiträge in den *OAG-Mitteilungen* fanden schon damals weltweit Beachtung in Fachzeitschriften, wie zahlreiche Zitate und Verweise zeigen. Diese Leser hatten eine spezifische Art, die Daten zu lesen und zu deuten. Doch solche Kontexte ändern sich oft stärker, als uns bewusst ist. Ist Ihnen aufgefallen, dass Knippings Tabelle keine Einheiten für den Barometerstand des Monats (Zeile 1) oder die höchste Wärme des Monats (Zeile 5) nennt? Wer Hektopascal oder Grad Celsius vermutet, irrt. Tatsächlich handelt es sich um Pariser Linien und Réaumur (siehe auch Feature I).



Abb. 2: Die Karikatur der französischen Künstlerin Barbara Bellier verdeutlicht, dass Menschen und Computer dieselbe Sprache sprechen müssen, um vielfältige und große Datenmengen auszuwerten. Die Bildrechte liegen bei den Autoren.

Ein weiterer Aspekt der bisherigen Überlegungen zur Zielgruppe verdient Beachtung. Im digitalen Zeitalter sollen Daten auch für eine sekundäre Leserschaft verständlich und nutzbar sein. Das dient nicht nur der Transparenz oder Verbreitung von Wissen, sondern soll auch breitere Nutzergruppen und neue Anwendungen erschließen. Das ist der große Vorteil der Digitalisierung und der automatisierten Datenauswertung. Das erfordert neue wissenschaftstheoretische Ansätze. Wir stecken mittendrin in der Entwicklung, sind aber noch lange nicht am Ziel.

Keine Dokumentation, so gut sie auch sein mag, ist fehlerfrei. Schauen wir erneut auf Abbildung 1. Unter Eintrag Nummer 7 sind drei Zeilen hervorgehoben – sie enthalten einen Druckfehler. Für die Monate Juli, August und September stehen dort drei Niederschlagswerte: die gesamte Regenmenge in Kubikzoll sowie die jeweiligen Anteile von Regen und Schnee. Da Schnee im Sommer in Tokyo äußerst unwahrscheinlich war und

ist – was der Strich im Schneefeld bestätigt –, müssten der erste und der zweite Wert übereinstimmen. Das trifft auf August und September zu, doch im Juli weichen die Werte voneinander ab. Die Tabelle enthält also einen klaren Druckfehler.

Eine häufige Fehlerquelle sind vor allem Einträge, die auf Berechnungen basieren. Monatliche Mittelwerte waren schwer zu ermitteln, da das Teilen durch 31 Tage ohne Taschenrechner oft zu Fehlern führte – sei es beim Kopfrechnen oder beim Nachschlagen in damals gängigen Tabellenwerken. Ein weiteres Beispiel zeigt Zeile 9: „Hieraus ist die mittlere Wind-Richtung berechnet ...“. Knippings Tabelle verrät nicht, wie der Mittelwert aus den acht Windhäufigkeiten der verschiedenen Richtungen berechnet wurde. Unsere Recherchen ergaben, dass man die sogenannte Lambertsche Formel anwandte. Diese spielt allerdings heute keine Rolle mehr. Sie nicht mehr zu verwenden, ist das Ergebnis einer langen Diskussion unter Meteorologen. Was soll eine mittlere Windrichtung überhaupt aussagen? Angenommen, der Wind wehte fast immer aus Nordwest. Dann sollte der Mittelwert dies auch ergeben. Doch ein nordwestlicher Mittelwert kann auch entstehen, wenn der Wind nie aus Nordwest kam, sondern gleich oft aus Norden und Westen. Wegen dieser Unklarheit verwendet man heute eine Windrose, die Windstärken und Häufigkeiten durch Längen darstellt und so das Problem löst.

Die Herausforderung in der Digitalisierung von Daten liegt darin, Fehler richtig zu behandeln und implizites Wissen zu nutzen, um sie möglichst automatisiert zu erkennen. Manchmal lassen sich Fehler korrigieren, indem man aus anderen Informationen Schlüsse zieht, etwa durch das Prüfen von Summen. Doch oft geht das nicht. Welche der Regenmengen ist denn nun die richtige? Dann bleibt die Frage, was man stattdessen überträgt und welche Folgen die Lücke hat. Ähnlich schwierig verhält es sich mit unleserlichen Stellen: Häufig lassen sich etwa 3, 6 und 8 nur schwer auseinanderhalten.

Abb. 1 weist zudem auf ein sprachliches Problem hin: Knipping sprach von „Regensumme“, meinte aber aus heutiger Sicht „Niederschlagsmenge“. Ob die Begriffe synonym sind, hängt vom Kontext und vom Sprachgebrauch der jeweiligen Zeit ab. Umgangssprachlich mögen sie noch immer dasselbe bedeuten, doch wissenschaftlich ist „Niederschlag“ der Oberbegriff für Regen, Schnee, Hagel und Ähnliches. Wobei jede Epoche ihren eigenen sprachlichen und wissenschaftlichen Standards folgt. Die Digitalisierung der Daten verlangt, dass der Wandel dieser Standards bekannt und automatisch berücksichtigt wird.

Allerdings bleibt immer noch ungewiss, ob wir dann alle Informationen besitzen, um den damaligen Wissensstand von Knippings Leserschaft vollständig zu rekonstruieren. Bisher haben wir lediglich sichergestellt, die Tabelleneinträge korrekt zu interpretieren – einschließlich der Entscheidungen bei Fehlstellen oder Druckfehlern – und Vertrauen darin entwickelt, dass die Einträge weitgehend stimmen und sorgfältig erstellt wurden. Doch das garantiert keineswegs, dass Knippings Daten korrekt erhoben wurden. Ein Beispiel dafür ist seine eigene Feststellung, dass das Barometer über fünf

Jahre falsch kalibriert war, was zu fehlerhaften Messungen führte (siehe Feature I). Solche Fehler fallen oft erst auf, wenn man sie mit Alternativen vergleicht, etwa mit anderen Messreihen oder Simulationsergebnissen.

Im Fall von Knipping können die Daten des Japanischen Bergamtes (永田町 鑛山寮 氣象觀測表) für den Zeitraum 1872–1876 und ab 1875 die des Kaiserlich Meteorologischen Instituts als Vergleich herangezogen werden. Dabei muss man jedoch bedenken, dass die Messungen nicht an denselben Orten und Zeiten stattfanden – die Wetterstationen lagen etwa drei Kilometer auseinander. Das mag nach Haarspalterei klingen, hat aber Folgen, zum Beispiel im Fall von Taifunen. Die gemessene Regenmenge kann sich an beiden Orten stark unterscheiden, je nach Zugbahn der recht schmalen Regenbänder. Zudem landet bei starkem Wind weniger Niederschlag im Messgerät, was zu erheblichen Abweichungen führen kann. So war es nicht verwunderlich, dass der Vergleich der Datensätze vor allem während der Taifun-Saison größere Unterschiede aufweist. Die Aufgabe besteht daher darin, diese Abweichungen zu analysieren und ihre Ursachen zu klären. Hinzu kommen instrumentelle Messfehler, Ableseungenauigkeiten und natürliche statistische Schwankungen. Die Frage ist, wie stark eine Abweichung sein darf, damit die Daten noch als valide gelten. Die Luftdruckwerte des Bergamtes weichen so stark von den zwei anderen Datensätzen ab, dass man ihnen misstrauen sollte.

Vom Menschen zum Computer

Das sorgfältige Prüfen historischer Daten und ihres Kontexts bildet, wie erwähnt, den ersten Schritt zur Digitalisierung. Dabei kommt man nicht ohne die tiefgehenden Überlegungen eines Menschen aus. Noch scheitert der Computer an diesem Schritt, aber Algorithmen können unterstützen, etwa bei Rechenprüfungen. Denn wie gelingt der Übergang zur vollständigen Digitalisierung von Daten und Kontext, sodass Computer die Informationen für verschiedene Zielgruppen verarbeiten können? Wie bringt man Computern bei, was Menschen denken und gedacht haben? Wie formuliert man eine Anweisung, die es dem Computer erlaubt, auf das komplex vernetzte Wissen aus Beobachtungen, Kontext und Hintergrundinformationen jederzeit zuzugreifen und es in seine Interpretation einzubeziehen?

Die Antwort lässt sich an einem Beispiel verdeutlichen. Machen wir einen Ausflug in die Linguistik. Wir betrachten, wie semantische Verfahren helfen, Wissen zu vernetzen. Zuerst analysieren wir, wie Menschen eine Frage lesen, Wissen verknüpfen und Antworten finden. Anschließend vergleichen wir dies mit dem Vorgehen einer Maschine. Die Frage, die wir betrachten, lautet: Was bedeutet ein Temperaturunterschied von 5 Grad Celsius zwischen Mittag und Abend?

Ein Mensch liest die Buchstaben, erkennt Wörter und versteht ihre grammatikalischen Beziehungen. Er verknüpft diese Informationen mit seiner Erfahrung. Vielleicht denkt er: „Ah, die Temperatur wird um 5 Grad sinken.“ Aus seinem Erfahrungsschatz

schließt er: „Ein Pullover reicht wohl nicht für heute Abend, ich brauche eine Jacke.“ Der Mensch nutzt also einen klaren Messwert (Temperatur), kombiniert ihn mit seiner Erfahrung und schließt daraus die Bedeutung der Veränderung.

```

ex:obs001 a sosa:Observation ;
  sosa:observedProperty ex:airTemperature ;
  sosa:hasFeatureOfInterest ex:station_Knipping ;
  sosa:madeBySensor ex:thermometer_K1 ;
  sosa:resultTime "2026-05-14T12:00:00+09:00"^^xsd:dateTime ;
  sosa:hasSimpleResult "21.1"^^xsd:decimal ;
  qudt:unit unit:DEG_C .

ex:obs002 a sosa:Observation ;
  sosa:observedProperty ex:airTemperature ;
  sosa:hasFeatureOfInterest ex:station_Knipping ;
  sosa:madeBySensor ex:thermometer_K1 ;
  sosa:resultTime „2026-05-14T18:00:00+09:00“^^xsd:dateTime ;
  sosa:hasSimpleResult „16.1“^^xsd:decimal ;
  qudt:unit unit:DEG_C .

```

Abb. 3 zeigt, wie ein Computer Temperaturmessungen der Messstation „Knipping“ verarbeitet. Die einzelnen Zeilen definieren verschiedene Informationen und verknüpfen sie über Internetseiten wie das Beispiel für „sosa:Observation“ in Bildschirmfoto 4 verdeutlicht.

Abb. 3 zeigt Schritt für Schritt, wie ein Computer die Temperaturdifferenz begreift. In der ersten Zeile steht, dass es sich um eine Beobachtung handelt (englisch: Observation). Die zweite Zeile präzisiert, dass die beobachtete Größe die Lufttemperatur ist. Die dritte und vierte Zeile ergänzen, dass die Temperatur an der Station „Knipping“ mit dem Thermometer „K1“ gemessen wurde. Danach folgen der genaue Zeitstempel für die Mittags- und Abendwerte sowie die gemessenen Temperaturen von 21,5 und 16,1 Grad Celsius, jeweils als Dezimalzahlen. Am Ende wird die Einheit, Grad Celsius, angegeben. So funktioniert das Grundprinzip. Entscheidend ist, dass hinter jeder dieser Informationen eine klar definierte Webseite steht. Diese nennt man persistenten Identifikator (englisch: *persistent identifier*), wie etwa die bekannte DOI für Dokumente (englisch: *digital object identifier*). Damit erkennt die Maschine die einzelnen Bestandteile. Um sie nun in einen sinnvollen Zusammenhang zu bringen, werden die Internetseiten verknüpft, ähnlich wie wir in einem Satz Subjekt, Prädikat und Objekt verbinden. So entsteht ein Wissensnetz – riesig und beliebig erweiterbar. Zum Beispiel ließe

sich die Temperaturdifferenz mit Daten für Kleidungsempfehlungen verknüpfen oder mit zusätzlichen Informationen zu Erwin Knipping.

The screenshot shows a web browser window displaying the W3.org page for the 'sosa:Observation' class. The page is titled '4.3.22 sosa:Observation' and includes the following content:

- IRI:** <http://www.w3.org/ns/sosa/Observation>
- a OWL Class**
- Observation** - Act of carrying out an (Observation) [Procedure](#) to estimate or calculate a value of a property of a [FeatureOfInterest](#). Links to a [Sensor](#) to describe what made the [observation](#) and how; links to an [ObservableProperty](#) to describe what the result is an estimate of, and to a [FeatureOfInterest](#) to detail what that property was associated with.
- Example** The activity of estimating the intensity of an Earthquake using the Mercalli intensity scale is an [Observation](#) as is measuring the moment magnitude, i.e., the energy released by solid earthquakes.
- Restrictions**
 - [sosa:madeBySensor](#) **EXACTLY 1**
 - [sosa:madeBySensor](#) **ONLY** [sosa:Sensor](#)
 - [sosa:usedProcedure](#) **ONLY** [sosa:Procedure](#)
 - [sosa:hasFeatureOfInterest](#) **EXACTLY 1**
 - [sosa:hasFeatureOfInterest](#) **ONLY** [sosa:FeatureOfInterest](#)
 - [sosa:observedProperty](#) **EXACTLY 1**
 - [sosa:observedProperty](#) **ONLY** [sosa:ObservableProperty](#)
 - [ssn:wasOriginatedBy](#) **EXACTLY 1**
 - [ssn:wasOriginatedBy](#) **ONLY** [ssn:Stimulus](#)
 - [sosa:phenomenonTime](#) **EXACTLY 1**
 - [sosa:hasResult](#) **MIN 1**
 - [sosa:hasResult](#) **ONLY** [sosa:Result](#)
 - [sosa:resultTime](#) **EXACTLY 1**

At the bottom of the restrictions section, there is a link: [\[Hide additional SSN axioms\]\[Back to module overview and examples\]\[Back to log\]](#)

Abb. 4: Es zeigt, wie die Angabe, dass es sich um eine Beobachtung handelt – in der Zeile unter „sosa:Observation“ (siehe Abb. 3) – im Internet präzisiert wird.

Abb. 4 zeigt die Internetseite hinter der Computeranweisung „sosa:Observation“. Sie liefert die Information, dass es sich bei der zu interpretierenden Angabe um eine Beobachtung handelt. Auf der verbundenen Internetseite ist die computerlesbare Definition von Beobachtung ausführlich beschrieben und logisch mit anderen Informationen verknüpft. So erfahren wir, dass eine Beobachtung ein Prozess ist, der einen Wert schätzt oder berechnet und zu einem Ergebnis führt. Konkret: Jede Beobachtung hat dabei mindestens einen Wert ([sosa:hasResult](#) MIN 1) und kann nur zu einem bestimmten Zeitpunkt gemessen werden ([sosa:phenomenonTime](#) EXACTLY 1). Diese Einträge sind ebenfalls mit Links zu Internetseiten versehen und führen zu weiteren Themen wie „Sensor“ und so weiter und so fort. Erst durch diese klar definierten Verknüpfungen kann eine Maschine Informationen logisch verbinden und so „verstehen“. Der Vollständigkeit halber sei erwähnt, dass es sich bei den Computeranweisungen um RDF handelt, einen Standard zur Beschreibung von Daten im Internet. RDF steht für „Resource Description Framework“.

Nun haben wir alles beisammen, um nachzuvollziehen, wie sich Wissen aus Daten digital weitergeben lässt: Wissensbausteine werden klar definiert und in logische Zusammenhänge gebracht. Aussagen zu einem Objekt werden gebündelt, wodurch ein Kontext entsteht, der über die bloße Darstellung von Tabellen wie denen Knippings

hinausgeht. Durch komplexe Wissensnetzwerke wird der gleichzeitige Zugriff auf alle über viele OAG-Hefte verteilte Tabellen als auch auf weitere Tabellenwerke, Aufzeichnungen, Protokollierungsvorschriften und Bücher ermöglicht. In unserem Beispiel zeigt sich diese Bündelung in der Anweisung „ex:obs001“, gleich zu Anfang von Abb. 3. Sie bedeutet, dass alle folgenden Informationen im Objekt „obs001“ zusammengefasst werden sollen. Dieses Objekt steht wiederum in Verbindung mit anderen, etwa mit „Person Erwin Knipping“, wie es auf Wikidata definiert ist (s. Abb. 5).

The screenshot shows the Wikidata entry for Erwin Knipping (Q11291134). The page header includes the Wikidata logo and a search bar. The main content area displays the item name and a table of labels in different languages. Below the table, there is a 'Statements' section with an 'instance of' statement.

Language	Label	Description	Also known as
default for all languages	Erwin Knipping	–	
English	Erwin Knipping	German meteorologist (1844–1922)	Erwin Rudolf Theobald Knipping
German	Erwin Knipping	deutscher Meteorologe und Kartograf	Erwin Rudolf Theobald Knipping
French	Erwin Knipping	No description defined	
Bavarian	Erwin Knipping	No description defined	

Statements

Instance of: human (edit)

+ 0 references

+ add reference

+ add value

image (edit)

Abb. 5: Das Bildschirmfoto zeigt einen kleinen Ausschnitt vom Eintrag über Knipping in wikidata.

Wikidata, eine freie und offene Wissensdatenbank, sammelt strukturierte Daten und macht sie sowohl Menschen als auch Maschinen zugänglich. Als Teil der Wikimedia-Familie wird sie von der Wikimedia Foundation betrieben. Im Eintrag zu Knipping – genauer gesagt zu Objekt Q11291134 – erfahren wir in mehreren Sprachen, dass er ein deutscher Meteorologe ist, ergänzt durch seine Lebensdaten. Damit ist klar, um wen es sich handelt. Die Webseite verlinkt zudem auf die Wikipedia-Seite zu Knipping, was

den Kontext weiter vertieft. Bei unserer Recherche fiel uns übrigens auf, dass es noch keinen Wikidata-Eintrag³ zur OAG gibt. Wer möchte, kann diesen jederzeit erstellen oder erweitern. Das semantische Web lebt schließlich von Freiwilligenarbeit.

Weltweite Institutionen koordinieren und standardisieren inzwischen systematisch. Institutionen wie das World Wide Web Consortium (W3C) stehen im Zentrum. Tim Berners-Lee, der Erfinder des Webs, gründete es bereits 1994. Auch nationale Organisationen tragen zur Wissensorganisation bei. In Japan führt die Japan Science and Technology Agency (JST) die Standardisierung und Interoperabilität von Daten an. Europa hingegen arbeitet weniger zentralisiert: Das Europäische Komitee für Normung (CEN) legt allgemeine Standards fest, CENELEC konzentriert sich auf elektrotechnische Normen und ETSI entwickelt Telekommunikations- und digitale Standards, Internetprotokolle und Kommunikationsinfrastrukturen. Im Forschungsbereich setzt Europa auf einen koordinierten, regelbasierten Ansatz mit gemeinsamen Standards und institutionellen Rahmenwerken. Projekte wie die European Open Science Cloud zielen darauf ab, Daten grenzüberschreitend einheitlich zugänglich und nutzbar zu machen. Dafür gibt es zum Beispiel verpflichtende Vorgaben für die Bearbeitung von Daten in Forschungsprojekten. Konkret müssen die sogenannten FAIR-Prinzipien angewendet werden. FAIR steht dabei für findbar, zugreifbar, kompatibel und wiederverwertbar (*F-Findable, A-Accessible, I-Interoperable* und *R-Reusable*). Japan verfolgt einen anderen Weg und setzt stärker auf Vernetzung mit der Industrie und ist weniger normativ. Für Forschungsprojekte sind zum Beispiel die FAIR-Standards nur empfohlen und nicht verpflichtend.

Fazit

Die Digitalisierung historischer Daten erfordert umfangreiche Verknüpfungen von Wissen und Kontext. Nur so können Computer Informationen verstehen und für verschiedene Zielgruppen nutzbar machen. Knippings meteorologische Aufzeichnungen zeigen, wie komplex dieser Prozess ist – und wieviel Potenzial in der Verbindung von Mensch und Maschine steckt.

Dr. Valeria Jana Schwanitz, Diplom-Physikerin, promovierte Diplom-Wirtschaftsphysikerin, ist Professorin an der Western Norwegian University of Applied Sciences mit dem Lehr- und Forschungsgebiet „Klimawandel und Energieökonomie“. Sie ist regelmäßig privat und beruflich in Japan unterwegs. Sie war DAAD-Stipendiatin und Stipendiatin der Japanese Association of University Women sowie Gastwissenschaftlerin an den Universitäten Kyoto und Osaka, an der Universität Ryūkyū sowie am Okinawa Institute of Science and Technology (OIST). Während eines Postdocs am Potsdam-Institut für Klimafolgenforschung hat sie mit multi-komponenten Klimamodellen gearbeitet und war am SSP-Szenarioprozess beteiligt.

³ Einen ausführlichen Wikipedia-Eintrag über die OAG gibt es bereits. (Anmerkung der Redaktion)

