



Open Power System Data

A free and open data platform for power system modelling.

Gefördert durch:
 Bundesministerium
für Wirtschaft
und Energie
aufgrund eines Beschlusses
des Deutschen Bundestages

Erster Workshop des Projekts *Open Power System Data*

DIW Berlin, 15. Oktober 2015

Protokoll

www.open-power-system-data.org

1. Einführung und Überblick (Lion Hirth)

- Input: siehe Folien

2. Datentypen im Einzelnen: Vorstellung der Arbeitsgruppen (Projektteam)

- Bestandsdaten konventionelle Kraftwerke und Wasserkraft
- Bestandsdaten Windkraft, PV, Biomasse; Wetterdaten
- Zeitreihen: Einspeisung Windkraft und PV, Last
- Inputs: siehe Folien

Diskussion

- Nachfragen: Bietet OPSD auch Zeitreihen zu
 - Wasserkraft?
 - Kraftwerksverfügbarkeit?
 - Wärmeauskopplung?
- Scigrid: Offenes Netzmodell für Europa, Netztopologie, veröffentlichen ihre Daten, würden ihre Skripte zur Verfügung stellen

3. Technisches Konzept (Ingmar Schlecht)

- Input: siehe Folien

Diskussion

- UTF-8 oder ISO-8859-1?
 - Anmerkung: UTF-8
- Wer möchte Excel zusätzlich zu CSV?
 - 20-30%
- Welche Linebreaks?
 - Mehrzahl nutzt Windows
- bezeichnet ein Zeitstempel Anfang, Mitte oder Ende der Stunde?
 - zu klären; Thema für Zeitreihen
- Was passiert bei Änderung der Datenformate?
 - Anpassung des Skripts notwendig;
- Vorschlag: Ein Skript, das jeweils ein aktuelles Datenpaket beim TSO herunterlädt
- Können Nutzer das Skript auf dem Server auslösen?
 - Nein
- Welche Rohdaten werden benutzt? Werden Daten rückwirkend neu runtergeladen? Wird man an die Rohdaten kommen, ohne den Umweg über Transparency-Plattformen?
 - Script lädt direkt vom Bereitsteller (ÜNB). Daher ist eine Diskrepanz möglich, wenn z.B. ein ÜNB rückwirkend Veränderungen an den Daten vornimmt. Ursprünglich heruntergeladene Daten sollen jedoch auf dem OPSD-Server vorhanden bleiben; derzeit ist allerdings nicht klar, ob diese auch extern angeboten werden können.

- Ist eine Validierung vorgesehen?
 - Ja, Datenbereinigung durch das Skript
- Anmerkung: Déja-Vu von Vorgehen bei Fernerkundung, GEOSS, Thema Linked Open Data
 - Antwort Ingmar Schlecht: Daten haben dann oft nicht das richtige Format für Modellierer
 - Rücksprache noch mit Enipedia wegen IT-Lösungen zu Linked Open Data
- Vorschlag: (Konzept) OPSD könnte als Broker arbeiten und z.B. TSOs eine Möglichkeit bieten ihre Daten hochzuladen
- Anmerkung: In den CSVs sollten wir unterscheiden zwischen 0-Werten (Null) und N/A (missing values).

4. Rechtliches Konzept (Lion Hirth)

- Input: siehe Folien

Diskussion

- Frage zu commercial use: Fällt das Benutzen von Zeitreihen für Beratung unter commercial use?
 - Lion Hirth: Sehr wahrscheinlich ja
 - Anmerkung: Intel definiert commercial use darüber, ob Leute damit arbeiten, die dafür bezahlt werden.
 - Anmerkung Lion: derzeit wird niemand belangt, aber das könnte auch noch in mehreren Jahren passieren.
- Eigener Erfahrung eines Teilnehmenden: Datenbereitstellung ist häufig durch Gesetze eingeschränkt. Wichtiger Beitrag von OPSD: Strategien entwickeln, wie man damit umgeht
 - Lion Hirth: NIP hat sich damit auch noch nicht beschäftigt, hofft auf Gesprächsprozess.
- Anmerkung: Es ist wichtig eine Sensibilität bei den Datenprovidern zu schaffen -> Erneuerung der Lizenzen?
- Anmerkung (SciGrid?): Ein Rechtsgutachten erstellen, welches den Datenprovidern den juristischen Prozess vereinfacht.
- Anmerkung: Bei den Fraunhofer Energiecharts werden keine Rohdaten zur Verfügung gestellt, vermutlich wegen rechtlicher Bedenken

5. ENTSO-E Transparency (Lothar Rausch, Öko-Institut, ENTSO-E Transparency Core User Group)

- Input: siehe Folien

6. Drei Arbeitsgruppen

Anzahl Teilnehmer: Gruppe a: 18, Gruppe b: 6, Gruppe c: 14

a. Bestandsdaten konventionelle Kraftwerke und Wasserkraft (Clemens Gerbaulet, Friedrich Kunz, Wolf-Peter Schill)

Präsentation von Wolf-Peter Schill:

- Biomasse und Biogas werden hier nicht berücksichtigt (--> Gruppe b)
- Für DE im Detail, dabei BNetzA-Liste als Grundlage, ergänzt durch Informationen aus UBA-Kraftwerksliste
- Für andere Länder zunächst aggregierte Daten, später nach Möglichkeit zusätzliche kraftwerksscharfe Daten
- Darstellung aktueller Kraftwerksparks, aber keine Daten für vergangene Jahre
- Nutzung offizieller und möglichst valider Datenquellen

Beispielskript von Clemens Gerbaulet:

- Vorstellung eines Beispielskripts inkl. Output.
- Frage zu Windkraftanlagen im Output
 - müssen noch rausgefiltert werden.
- Kommentar zu matching: Problematik von Identifiern
 - nicht immer konsistent; Herausforderung, verschiedene Listen zusammenzuführen
- Anmerkung zu Standorten/Geodaten, Wirkungsgrade (ich weiß leider nicht wer das gesagt hat): Wir verfügen über diese Informationen
 - Wolf-Peter Schill: wir haben das alles auch schon einmal gemacht; die Idee ist, das alles noch einmal gut dokumentiert zu machen; vorhandene Daten gerne zum abgleichen benutzen
- Frage: wie geht ihr mit Metadaten um? Wo werden die hingeschrieben?
 - Clemens Gerbaulet: Quelle wird im Jupyter-Skript vermerkt, Quellen werden in der json Datei auch noch extra aufgeschrieben
- Frage: Wie funktioniert das dann mit der Versionierung?
 - Mit Github.
- Wenn es zwei verschiedene widersprüchliche (und dennoch offizielle) Quellen für ein Datum gibt, sind wir uns noch nicht einig wie wir damit umgehen
- Anmerkungen zum Wirkungsgrad:
 - Unklare Definition / Betriebspunkte; es gibt daher nicht den einen "korrekten" Wirkungsgrad; evtl. könnte immer der höchste oder der niedrigste verfügbare Wert verwendet werden
 - Darüber hinaus bestehen erhebliche Herausforderungen bei der Recherche; möglicherweise könnten Datenlücken durch statistische Verfahren aufgefüllt werden; diese könnten wiederum in den Skripten angelegt werden
- Frage nach anderen technisch-ökonomischen Parametern, z.B. Ramping-Parameter, Mindestlast, Stilllegungen etc.?
 - Wolf-Peter Schill: Ramping-Parameter sind meist Hilfskonstrukte von Modellierern, lassen sich teilweise gar nicht empirisch erheben; zudem sehr unterschiedliche Anforderungen verschiedener Modelle → definitiv out of scope
 - Anmerkung Friedrich Kunz: Wir sind uns nicht sicher wie man mit den Technologie Daten umgeht
 - Datenrecherche: z.B. Internetrecherche, oder HiWis bei Kraftwerken anrufen lassen?
 - Anregung von Teilnehmern: man könnte eine Art Vorschlag mit entsprechenden Parametern erstellen (Ramping, Mindestlast etc.), als eine Art Empfehlung des OPSD-Teams → halte ich persönlich für nicht praktikabel (WS)!
- Anmerkung: Tabelle mit angenommenen Parametern aus anderen Publikationen wäre gut

- Wolf-Peter Schill: Widerspruch zur Philosophie des Projekts. Eigentlich wollten wir nur robuste Daten nehmen und keine Parameter annehmen. Man könnte aber eine ergänzende Linksammlung auf der Homepage anlegen.

(Ende Beispielskript)

- Anmerkung: Historie beim Kraftwerkspark wäre doch super für back-testing
 - Wolf-Peter Schill: Wie viele Jahre wären wichtig?
 - Teilnehmer: z.B. bis 2010 zurückgehen; vor Fukushima wäre relevant. Ist einmaliger Aufwand, großer Nutzen
 - Wolf-Peter Schill: In BNetzA-Liste (und teilweise UBA-Liste) schon enthalten: Inbetriebnahme, teilweise Revision, Abschaltung von Kraftwerken, allerdings nichts was vor 2011 passiert ist.
 - Friedrich Kunz: Überlegung: BNetzA anschreiben und nach alten Datensätzen fragen. Keine Zusicherung
- Hinweis: Man könnte BNetzA-Datensatz verbessern: z.B. Energieträger, Turbinenkonfiguration nur semi-gut. Bitte an Qualität der Daten arbeiten.
 - Wolf-Peter Schill: Hauptenergieträger sind angegeben. Es ist schwierig das besser zu wissen.
 - Konsens: alle Energieträger-Spalten der BNetzA-Liste mitführen, allerdings die ersten beiden nicht übersetzen
- Frage Wolf-Peter Schill: Welche Typen von Kraftwerken sind denn relevant für euch? Wie scharf guckt ihr drauf? Was aggregiert ihr?
 - Keine klare Antwort → die meisten Teilnehmer scheinen aber recht umfassend zu aggregieren (z.B. GuD nicht getrennt als GT und DT)

Abfrage welche Infos relevant sind:(slide 7)

zu ID:

- Sollen kleine Kraftwerke ohne offizielle ID eine ausgedachte bekommen? Sollen disaggregierte Kraftwerke mehrere neue IDs bekommen?
 - Anmerkung: Soll nachvollziehbar sein
- Hinweis: NEP-Kraftwerksnummern könnten eine hilfreiche Ergänzung sein
 - auch weitere IDs könnten eine sinnvolle Ergänzung sein, z.B. EIC Codes
 - Ob das jemand braucht wird nicht direkt beantwortet.
- Frage: Wie geht man damit um, wenn sich z.B. die BNetzA-Nummern ändern?
 - Wolf-Peter Schill: sollte eigentlich nicht passieren, sonst echtes Problem. Sollte im Auge behalten werden, BNetzA ggf. direkt drauf ansprechen

zu Unternehmen: werden einfach übernommen

zu Kraftwerksname: werden übernommen

zu Standort Kraftwerk:

- es werden noch 3 Geokoordinatenspalten hinzugefügt. (Längen- und Breitengrad, sowie eine maschinenlesbare Spalte)
- Frage Wolf-Peter Schill: gibt es weitere Anforderungen?
 - Keine.

zu Blockname: wird übernommen

Problem: Wenn mehrere Jahre angegeben sind: wie sollen wir damit umgehen? Wie macht ihr das?

- Textspalte mit Eintragung, daneben möglicherweise weitere Spalte mit Integer als Mittelwert
- Retrofit als Extrainfo in separater Spalte

Energieträger:

- Frage Wolf-Peter Schill: nur Hauptenergieträger benutzen (Rest raus)? Wer braucht die Unterscheidung zwischen versch. Gasen? wie macht ihr das?

- Ein Teilnehmer: Erstmal alles bereitstellen was in der Ursprungsliste steht, dann auswählen. Sonst schaut man doch immer wieder in die Originalquelle
- Problem: Energieträger nicht schlüssig kategorisiert, sehr viele Möglichkeiten -> ggf. deutsche Info mitnehmen (definitiv nicht übersetzen)
- Anmerkung Friedrich Kunz: wir schmeißen keine Infos weg, man kann immer zur Rohinfo zurück kommen, wir suchen hier Prioritäten
- Wolf-Peter Schill: In der csv sind die dann allerdings nicht drin. Nur in den Rohdaten (Primärquelle im Netz - aber unklar, ob archivierte Rohdaten vom OPSD-Server öffentlich angeboten werden können, s.o.)

zu Vergütungsfähigkeit EEG: betrifft hier nur Wasserkraft

zu einem merge von UBA- und BNetzA-Listen:

- Frage: wer benötigt konkrete Daten zur Wärmeauskopplung von KWK-Kraftwerken?
 - ca. 10 Hände
- Frage zu UBA-Liste: stehen da auch Kraftwerke die Industriegwärme bereitstellen drin, oder fallen die durchs Raster?
 - grundsätzlich ja, aber diverse Einzelfälle deuten zumindest auf unklare Abgrenzung hin
- Bruttoleistung in UBA-Liste, Nettoleistung in BNetzA-Liste
 - Rückfrage zu BNetzA: tatsächlich Nettoleistung
 - Wolf-Peter Schill: werden beide benötigt? Antwort: grundsätzlich ja
 - In UBA-Liste nur > 100MW; Anmerkung: Vorschläge sammeln wie man damit umgeht

zu Pumpspeichern und (möglicherweise kaskadierenden) Reservoirs:

- Interessiert Einzelleistung oder Aggregation?
 - Einzelleistung wäre interessant; aber keine artifizielle "eigene" Aufteilung
 - aggregierte Pumpleistung in BNetzA-Liste soll nicht direkt aufgeteilt werden, sondern erstmal so lassen
 - Turbinentyp und Größe des Sees auch wichtig.
- Wolf-Peter Schill: was ist offizielle Quelle von Speicherbecken?
- Ingmar Schlecht: weitere Tabelle von Becken und Pumpen
 - Größe/Einheit: Wasser (m³), weil Energie dann von der Turbine abhängt
 - Anmerkung: Tabellenformate sollten so sein, dass andere Länder mit übernommen werden können
 - Anmerkung: weitere Spalte(n): hinter welchen Reservoir hängt die Pumpe/Turbine; in der Liste angeben, welche Anlagen mit welchen zusammenhängen
- Frage: Sollen Pumpen und Turbinen differenziert werden?
 - Differenzierung von Maschinentypen von Turbinen/Pumpen ist datenseitig schwierig
- Anmerkung: Pumpleistung wäre noch wichtig anzugeben (ist in der Regel nicht gleich Turbinenleistung)
 - Anmerkung: In DE weniger wichtig als im Alpenraum und Skandinavien. Da ist es allerdings unrealistisch alle Pumpen zu finden.

Sonstiges:

- Frage Wolf-Peter Schill: "Elektrisches Inland" vs. politische Grenzen? Frage im Projektkontext möglicherweise nur für eine Übergangszeit relevant; wenn ihr von DE spricht, gehört Luxemburg dazu und Österreich?
 - Wortmeldung: Luxemburg in DE, Österreich eigene Region
- Verknüpfungspunkt (Schaltanlage) in BNetzA-Liste ziemlich unsystematisch, genauso ist Netz oder Umspannebene des Anschlusses inhomogen
- Anmerkung: wäre grundsätzlich super wenn die Plattform fremde Erweiterungen aufnehmen würde
 - Grundsätzlich durch Veröffentlichung auf Github (?)

- Allerdings sind Erweiterungen durch Dritte zunächst einmal nicht vorgesehen
- Anmerkung: Modularität wäre vorteilhaft, falls man nicht alles braucht
 - Schwierig, da dann nicht mehr alles in einem Skript vorhanden. Verschachtelung von Skripten problematisch
- Frage der restlichen Lebensdauer / Laufzeiten (insbesondere Stilllegungszeitpunkte) von Kraftwerken
 - Wolf-Peter: wollen wir grundsätzlich nicht mit reinnehmen. Evtl. könnte man NEP-Daten dafür heranziehen.
- Anmerkung Casimir: Zuordnung zu den präqualifizierten Regelleistungspools (gibt eine öffentliche Liste der präqualifizierten Regelleistungspools. Fraglich ist aber, welches Kraftwerk genau zu welchem dieser Pools gehört)
 - Wolf-Peter Schill: weit über der Grenze von validen Daten
 - Friedrich Kunz: Evtl. ginge es über eine nachträgliche Verschneidung der Liste mittels der Namen der Firmen.
 - Kraftwerke und nicht Firmen sind präqualifiziert
 - Frage: könnte man das als Modul bereitstellen?
 - Anmerkung Berit Müller: man sollte das mit der open model initiative verbinden
- Anmerkung: "Modulierbarkeit" ist wichtig (gemeint waren wohl Parameter zu Ramping und Mindestlast)
 - Wolf-Peter Schill: technische Grunddaten werden bereitgestellt, jeder kann sich dann selbst etwas zu Ramping und Mindestlast ausdenken
- Stimmungsbild: wie geht ihr um mit GuDs, was braucht ihr?
 - Generell: braucht man nicht getrennt.
 - Teilnehmer: getrennt wäre schon gut.
- Gibt es Interesse an eine Disaggregation von Kohle-Gas-Kombiblöcken? → weitgehend nicht.
- Information zu Industriekraftwerk ja/nein ist relevant
- Heizkraftwerke/Wärmeauskopplung kenntlich machen
- Wie soll mit aggregierten Blockwerten von Wärmeauskopplung umgegangen werden?
 - Keine eindeutige Lösung
 - Anmerkung: Vorsicht, alle als Wärmekraftwerk auszugeben, dadurch könnten bei einzelnen Blöcken falsche Must-run-Restriktion entstehen
- Anmerkung: An BNetzA und UBA herantreten und vorschlagen wie man diese Tabellen verbessern kann.
- Werden Daten auch aufgefüllt wenn es Lücken gibt?

b. Bestandsdaten Windkraft, PV, Biomasse; Wetterdaten (Wolf-Dieter Bunke und Martin Jahn)

Teil I: Bestandsdaten Wind, PV, Biomasse

Wolf-Dieter Bunke stellt das Vorgehen vor (Datenarten, Quellen, Jupyter Notebook etc.) -> siehe Folien Fragen/Kommentare von Teilnehmenden

- Dringende Empfehlung von Gesprächen mit BNetzA über deren gegenwärtige Planungen zur Änderung der Datenstruktur der Stammdaten und zukünftige Datenbankstrukturen. Gerade bei Stammdaten sind erhebliche Veränderungen zu erwarten. In der BNetzA sind derzeit viele Puzzleteile "in Bewegung" beim Aufbau der professionellen Datenbank. Bestandsdaten werden in System integriert. Alle Anlagenbetreiber sollen z.B. angeschrieben werden, ihre Daten zu aktualisieren. Es sollten in OPSD also am besten die schon von der BNetzA validierten Daten statt die Eingangsdaten der ÜNBs verwendet werden. Frau Böde im EEG-Daten-Referat wäre eine Ansprechpartnerin.

- Vor vielen Jahren wurden schon Modelle entwickelt, die eine Ertragskontrolle der gemeldeten Daten abgleichen. Geschätzter/selbst berechneter Jahresertrag vs. gemeldeter Ertrag +/-5-8% wäre z.B. eine gute Validierungsgrenze für die Solardatenbank.
- Kooperationsangebot zur gemeinsamen Validierung/Plausibilisierung an BNetzA sinnvoll?
- Sollten Anlagen, die im Register sind, aber entweder ungemeldet außer Betrieb sind oder praktisch einfach nicht mehr laufen/keinen Ertrag gemeldet haben, "gleichberechtigt" in unserer Datenbank stehen? (Oder besonders gekennzeichnet werden?)
- Bei der neuen Anlagenregisterverordnung hat der Datenschutzbeauftragte des Bundes schärfere Anforderungen gestellt. Daher dürfen nicht mehr alle Informationen (zur personenscharfen Zuordnung) veröffentlicht werden, um besonders kleine Anlagen nicht mehr genau zuordnen zu können.
- Wir sollten uns ggf. auch in den Prozess zum Marktstammdatenregister einbringen, um Datenstrukturen etc. abzugleichen/anzupassen.
- Anmerkung: EEG-Register wird wohl 1:1 ins MSR eingebracht werden und somit sind in der Hinsicht wenige Änderungen zu erwarten. Herausfordernd wird es, die konventionellen Kraftwerke auf denselben Datenstand/Detailgrad zu bekommen.
- Anmerkung: Eventuell auf unserer Datenplattform auch ein Dokument anbieten, das nur die Dokumentation beinhaltet. Wenn Code länger wird, ist Jupyter Notebook u.U. zu unübersichtlich.
- Frage in die Runde: Niemand kennt/nutzt Jupyter Notebooks bisher, Python ist aber weit verbreitet und anerkannt.
- Wie detailliert sollte die Dokumentation werden?
 - Antwort eines Teilnehmenden: Sehr wichtig ist, die entstandenen Probleme und mögliche Fallstricke bei der Datenverarbeitung aufzunehmen (Sensibilisierung der Nutzer für mögliche Fehler). Auch ein Kommentarfeld zum Feedback der Nutzer an die Skriptersteller vorsehen. Aber nicht jeden kleinen Bug öffentlich diskutieren sondern eher übergeordnete "Tipps & Tricks" bei Skript-/Datennutzung ermöglichen (Forum?).
 - Antwort eines Teilnehmenden: Muss drin sein: Welche Basisdaten sind verändert worden? Formate, Zeitversionen, Kommas etc.
 - Antwort eines Teilnehmenden: Vielleicht auch zwei verschiedene Versionen anbieten: Eine Version mit fehlerhaften aber gekennzeichneten Daten und eine komplett "gesäuberte" Version.

Wolf-Dieter Bunke erläutert die möglichen zu erhebenden übergreifenden Parameter für Wind, PV, Biomass

Fragen/Kommentare von Teilnehmenden

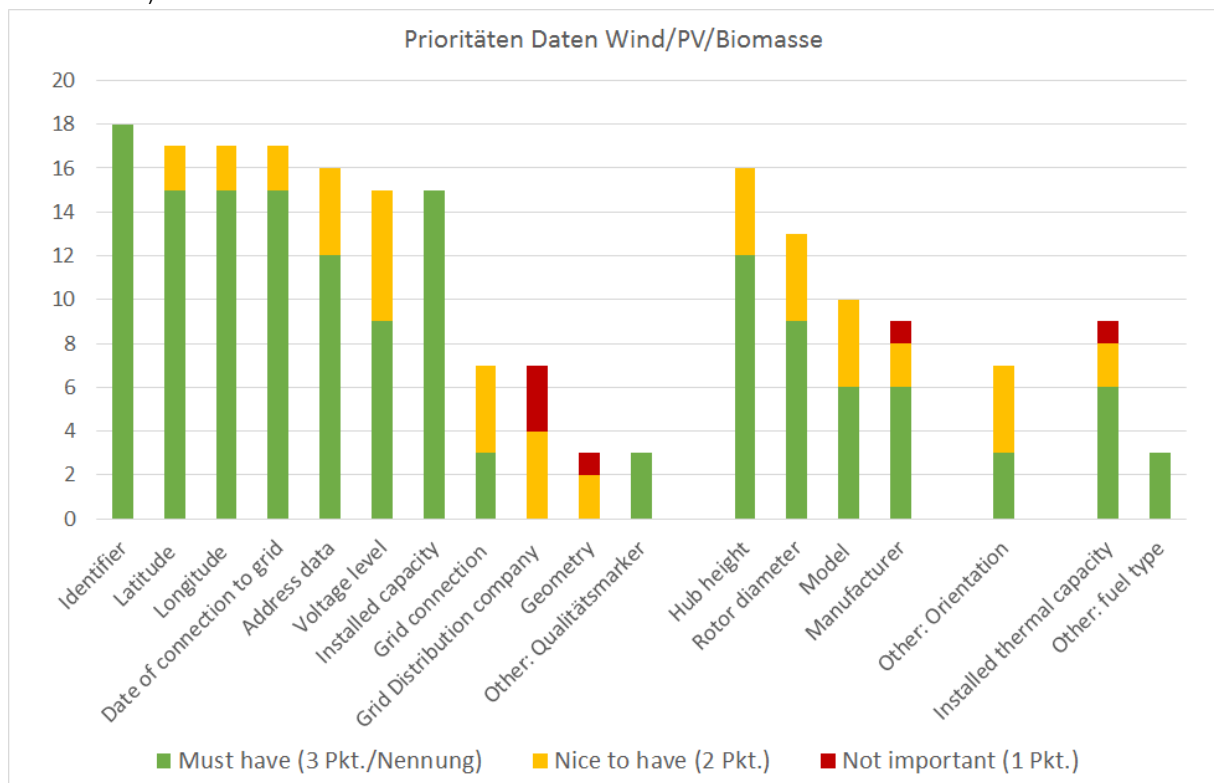
- bei den Geometrien: Welchen Punkt der PLZ nimmt man? Mittelpunkt der Region? Muss klar sein!
- Evtl. Fehler/Radius um PLZ-Geometrie-Punkt angeben?
- Integer & Numeric nicht unterscheiden, lieber gleich Float nutzen.
- Hinweis auf WKA-Datenbank für Notfälle (Feuerwehr, Rettungszwecke) mit Geokoordinaten vorhanden. Kann man das nutzen? (Allerdings keine Enercon-Anlagen drin).
- Andere Datenbank für Luftfahrthindernisse (mit Nabenhöhe, Standort) ist noch eine Möglichkeit.
- Nutzen windpower.net, für jedes Land einzelne Windparks vorhanden mit Anzahl Anlage, Hersteller, Standort etc. verfügbar. U.U. nicht vollständig?
- Zuordnung der Adressdaten von WKA zu "Meldeadressen" und tatsächlichem Standort schwierig.
- Ähnliche Problematik bei Freiflächen-PV!

- Alle Felder in OPSD-Datenbank vorsehen, die im neuen Register vorgesehen sind (auch wenn sie noch leer sind). Anlagenbetreiber sollen später Post bekommen, die Daten nachzuliefern.

Wolf-Dieter Bunke erläutert die möglichen zu erhebenden spezifischen Parameter für Wind, PV, Biomasse

Anmerkungen von Teilnehmenden:

- PV - was macht man mit Anlagen, die in mehreren Teilen aufgebaut sind (z.B. mit unterschiedlichen Ausrichtungen) - gibt man nur Hauptrichtung an? Macht man zwei Anlagen draus?
- Für PV evtl. noch Eigenverbrauch ja/nein interessant. Oder auch das verwendete Vermarktungsmodell. Zusätzlich bei Biomasse: Welcher Energieträger (feste Biomasse, Biogas etc.)?



weitere Fragen:

- How to offer the data sets? - a huge list, by country / generation type ...?
 - Anmerkungen von Teilnehmenden:
 - Bei ihnen getrennt nach Anlagentypen.
 - getrennt ist sinnvoll, da eh unterschiedliche Parameter wichtig sind.
 - Letztendlich nicht so der große Unterschied, da individueller Anpassungsaufwand jedes Nutzers/Filterung recht gering ist.
 - Was spricht dagegen, gleiche Tabellenstruktur/Datenspalten für EE-Anlagen und konventionelle Anlagen anzulegen, auch wenn dann einzelne Felder leer bleiben?
 - Verwenden von mandatory Spalten und optionale Spalten empfehlenswert.
- How should a validation look like? And where is the information needed? (e.g. Zahlenkodierung als Marker in Extra-Spalte)
 - Anmerkungen von Teilnehmenden:

- erkennbar sollte mindestens sein, wenn ein Datensatz verändert wurde. Was genau verändert wurde, ist eine zweite Ebene der Frage.
- Anzahl der geänderten Datensätze wäre auch interessant.
- DLR: Nutzen History-Felder in Metadaten bei deren Meteo-Datenbank. Was wurde von wem gemacht/geändert? Wird z.T. als Kommentarfeld/Zeile eingefügt.
- Information über Art und Anzahl der geänderten Informationen im Header würde ausreichen. Das genügt zumindest, um die Daten einschätzen zu können.
- U.U. mehrere Versionen mit unterschiedlichen "Anpassungstiefen" anbieten?
- How important is the geometry information for you?
 - Anmerkungen von Teilnehmenden:
 - Spielt für Marktsimulation keine große Rolle. Wind sensibler als PV für exakte Standorte. V.a. bei Wind ist z.B. eine Genauigkeitsinformation wichtig. Am besten ist Geokoordinate, aber das ist anwendungsabhängig.

Teil II: Wetter

Wolf-Dieter Bunke stellt das Vorgehen vor (Verfügbare Datenarten, Quellen, Aufbereitung etc.) -> Siehe Folien

Fragen/Kommentare dazu:

- Aus Kopernikus-Daten gibt es Solardaten in 5km-Raster ¼-stündlich, frei verfügbar.
- Auf Folien genannte Wetterdaten (50x50 km) sind z.B. für genaue Berechnung von Windeinspeisung zu wenig genau!
- DLR: Nutzen in ihren Modellen sehr unterschiedliche Datenquellen. Windgeschwindigkeit ist sehr sensitiv was Punktgenauigkeit angeht. Strahlungswerte sollten eher die direkten (statt indirekter wie auf Folie) Strahlungsanteile sein. Solardaten in Reanalyse-Daten sind nicht so genau, Temperatur kann man aber daraus gut nehmen. Datenquellen kann man auch mischen, statt nur eine Quelle mit z.T. unzureichenden Datenqualitäten/Auflösungen etc. zu haben.

Wolf-Dieter Bunke erläutert die möglichen zu erhebenden Wetterdaten

Diskussion der Wichtigkeit

- Welche wichtig sind, hängt sehr stark vom Anwendungsfall ab. Allgemeine Priorisierung ist deshalb schwierig.
- Ein Teilnehmer: Sie gucken sich über 20 Jahre durchschnittliche Windgeschwindigkeit an und schauen sich Jahresprofile an. Frage, wie man (Windjahres-?) Profil und durchschnittliche Windgeschwindigkeit in Wetterdaten/Datentabellen darstellen kann.
- Am wichtigsten sind Windgeschwindigkeit, Solardaten und Temperatur für stündlich simulierende Modelle.
- Wenn tatsächlich nur grobe Auflösung mit frei verfügbaren Daten machbar ist, dann sollte man darauf vlt. nicht zu viel Aufwand darauf verwenden.
- Dann lieber Hinweise auf bessere Quellen und so kann man sie selber aufbereiten (Formate, Raster etc.) -> Skripte dafür zur Verfügung stellen?
- Wetterdaten sind ein sehr großes Fass, eventuell erstmal beiseite stellen?
- Stattdessen Kategorisierung der Modelle/Datenquellen und mögliche Anwendungsfälle bereitstellen.

c. Zeitreihen: Last Einspeisung Windkraft und PV (Lion Hirth & Jonathan Mühlenpfordt)

Vorstellung Jupyter Notebook

- Frage eines Teilnehmenden: Versionierung: Wenn neue Daten gezogen werden, ist das eine neue Version?
 - Lion Hirth: ja, ist aber nur ca. halbjährlich geplant
- Lion Hirth: wahrscheinlich kann man 4 Arten von Nutzern unterscheiden:
 1. Nur csv Download
 2. schaut sich Script an
 3. verändert Skript
 4. Lädt ihr Skript auf GitHub
- Frage eines Teilnehmenden: Gibt es einen maintainer für die Skripte auf Github?
 - Lion Hirth: Noch nicht sicher
- Wer verwendet noch iPython Notebook?
 - Einige
- Anmerkung: Scripte bedeuten viel Arbeit, wenn sich Websites verändern
- Kann man das auch ohne das Notebook als Pythonscript ausführen?
 - Ja, man kann es exportieren.
- Frage eines Teilnehmenden: Wo sind die Metadaten? extra Datei oder oben in der CSV
 - Vorstellung des Data-Package Formats (Metadaten separat in .json-Datei)
 - Nachteil: Gefahr das Metadaten verloren gehen, wenn die csv weitergegeben wird
 - Vorteil: bessere Verarbeitbarkeit
 - Präferenz für Metadaten in externer Datei
- Welche Zeitreihen-Daten wären noch nützlich?
 - Anmerkung: Kapazitäten wären nützlich

Zeitstempel, Dateigröße

- Fragen/Anmerkungen von Teilnehmenden:
- Wirklich alles in einer CSV? Oder lieber Separate Dateien für verschiedenen Zeitzonen?
 - Lion Hirth: Es sind nur 2 Zeitzonen
- Grundsätzliche Frage: Wieviel Redundanz in den Daten? Präferenz für weniger, Lokale Zeitstempel kann sich jede selber generieren
- Man könnte im Script eine optionale Funktion für eigene Zeitstempel einbauen
- wenn es nur eine Zeitzone ist sollte die Information es in die Metadaten, sind es mehrere sollten die Lokalzeiten in eigenen Spalten mitgeführt werden.
- Wird das Notebook nicht zu lang, wenn alles in eine CSV soll?
- 30 Länder (EU 28+CH+NO) x 20 Zeitreihen = 600 Spalten
- Lion Hirth: Findet ihr es nützlich, Länderspezifische Dateien runterladen zu können?
- Vorschlag: bei OPDS gibt es nur die große CSV zum Download, Skripte sind aber Länderweise ausführbar, um sie schlanker zu halten. Nutzer können sich dann einzelne Länderzeitreihen selber runterladen.
- Was ist die passende Einheit? Land, Bidding Zone, Control Areas, TSOs?
- Ist Start und Endzeit überall gleich definiert?
 - Lion: Ja
- Ist der Zeitpunkt der letzten verfügbaren Beobachtung fix codiert? Muss irgendwo transparent gemacht werden.
- Wie werden missing observations in der CSV gekennzeichnet?
 - Optionen (Es gibt kein Richtig und Falsch):
 - Leer (also zwei aufeinanderfolgende Trennzeichen). Problem: Wird von Excel tw. als 0 behandelt
 - . (Punkt)
 - Zwei Anführungszeichen mit Leerzeichen ' '
 - Text, z.B. n/a

Ländernamen

- Wie geht OPSD mit unterschiedlicher Abgrenzungssystematik der Datenquellen um (Beispiel Österreich: Regelzone (APG) ≠ Bundesgebiet (E-Control))
 - Lion Hirth: OPSD übernimmt die Abgrenzung, die die Datenbereitsteller anbieten
- Bei ENTSO-E entweder Bidding Zones oder Control Areas
- Load: Total Load ≠ Vertical Load

Variablenamen

- Anmerkung: sollten über die Jahre konstant bleiben. Bei Enwida: Kurzname der gleich bleibt, ("TSO1") Langname der sich ändern darf: "EON" wurde zu "TenneT" und eine Erklärung. Empfehlung: Orientierung an vorhandenen Standards. Hinweis auf Java Jsigns Quantifiers
- Anmerkung: evtl. einheitliche Länge der Variablenamen?
- Zeichen die nicht genutzt werden sollten: Keine **Minus**, Plus, Leerzeichen, Sonderzeichen, Umlaute
- Unterstriche als Trennzeichen
- Einheiten: Stehen in den Metadaten
- Excel Template zusätzlich anbieten, das die richtigen Formatierungen enthält

Zeitliche Auflösung

- Was tun wenn 60' / 15' Zeitreihen gemischt?
 - Vorschlag eines Teilnehmenden): 60' in großer CSV, Wenn 15' verfügbar darauf in Metadaten hinweisen
 - Vorschlag eines teilnehmenden: Zwei Datensätze anbieten: ein großer mit 60' Auflösung mit allen Daten inkl. 15' resampelt, ein kleinerer mit 15' Auflösung wo Verfügbar
- Vorschlag eines Teilnehmenden: Wie ist die Qualitätssicherung? Wie geht OPSD mit der Verantwortung um, wenn viele die Daten nutzen?
 - Lion Hirth: Es kommt darauf an, gut zu dokumentieren

Correcting Data

- Lion Hirth: Lastdaten skaliert anbieten?
 - Pro: direkte Verwendbarkeit
 - Contra: Mehrarbeit, "Rechtfertigungsdruck" für OPSD, warum dieser und nicht jener Skalierungsfaktor verwendet
- Anmerkung eines Teilnehmenden Seit sie umgerechnete Daten rausschicken, haben sie weniger Nachfragen von Bachelor Studies und Journalisten.
- Anmerkung: Es wäre toll, wenn man mit OPSD-Daten ein Strommarktmodell füllen könnte und es stimmt halbwegs

7. Berichte aus den drei Arbeitsgruppen

a. Workshop Kraftwerksdaten

- Frage: wurde sich zu Wirkungsgraden entschieden? kommen die rein?
 - Ja, hier brechen wir ein wenig mit der Grundphilosophie des Projekts (nur valide, offizielle, öffentlich verfügbare Daten zu nehmen) und füllen Lücken ggf. auch durch statistische Methoden auf. Allerdings alles transparent dokumentiert im Skript.
- Frage: Was ist die Entscheidung zu Identifiern?
 - Was in BNetzA-Liste vorhanden ist wird genommen, bei unbenannten Aggregaten wird Nummer eingefügt, ggf. auch NEP-Nummer und EIC; Nummern für das Ausland werden ggf. definiert

b. Workshop Bestandsdaten Wind/PV/Biomasse, Wetter

- Frage: Gibt es Georeferenzierung?
 - Ja, über die Adresse (Postleitzahl)
- Anmerkung: Vernetzung mit Projekt open_eGo (Workshop am folgenden Tag) wichtig, um klarzustellen wer was macht. Präferenz: OPSD stellt nur gesicherte Daten zur Verfügung

c. Workshop Zeitreihen

- Zeitstempel: ein eindeutiger, weitere Lokale
- Metadaten: In separater Datei
- Anmerkung: es wäre doch gut wenn in der Datei die Einheiten mitaufgeführt wären.
 - Lion Hirth: Auf Grund von wenig Variation sollte es genügen wenn das in der separaten Datei aufgeführt ist.
 - Ingmar Schlecht: Bei viertelstündlichen Daten ist das nicht mehr so trivial. (wurde notiert)
- Zeitliche Auflösung: Die meisten Modelle sind stündlich aufgelöst, darauf liegt auch unser Fokus, Daten, die auch viertelstündlich verfügbar sind in Extra Datei anbieten
- Länderbezeichnungen
 - Gibt es eine eindeutige Kodierung an die man sich anlehnen kann?
 - Ländergrenzen/Biddingzones/Control Areas: OPSD orientiert sich an den bei ENTSO-E veröffentlichten Abgrenzungen
- Frage eines Teilnehmenden: Was ist mit Skalierung der Lastdaten?
- Vorschlag eines Teilnehmenden: Extra Skript/ Verknüpfung zu open_eGo: Kaskade von Post-Processing Prozeduren
 - Lion Hirth: dann Nachteil, dass Laien die unbearbeiteten Daten verwenden
- Modularisierung der Skripte
- Frage: Wie geht man mit Stundenwerten um? Anfang, Mittel, Ende?
 - Lion Hirth: Das ist für die Strommarktmodellierung nicht so wichtig. Bei Wetterdaten relevanter.
 - Anmerkung: In der Meteorologie bedeutet 0:00 Uhr die Stunde von 23:00-23:59.
 - Anmerkung: Bei den Strommarktdaten ist es wohl andersherum, muss man aber bei den Anbietern nachschauen.

8. Nächste Schritte (Lion Hirth)

- Zeitplan
- Betanutzer-Konzept
- Andere Projekte / Initiativen
- Nächster Workshop

- Frage: Wie wird mit Fragen die nicht eindeutig zu beantworten sind umgegangen. Wie wird das dokumentiert? Oder auf Fallstricke aufmerksam gemacht?
 - Lion: OPSD versucht möglichst häufig auf Standards/bekanntere Formate zurückgreifen
- Vorschlag: Liste mit "issues"(Probleme die nicht lösbar sind) auf Website veröffentlichen
 - Lion: Gelöste Probleme werden dokumentiert, Rest bleibt offen (?)
- Anmerkung: Mehrwert des Projekts: Wenn viele verschiedene Modelle mit einem "Referenzdatensatz" modellieren würden könnte man Ergebnisse anders (besser) vergleichen. Deshalb sollte man vlt. doch Lösungen anbieten für genannte Probleme.

- Anmerkung: so etwas in der Art gibt es schon, dieses Projekt nicht überfrachten
- Open Energy Modelling Initiative
- Projekt kann auch Katalysator für den Austausch mit Daten-Bereitstellern dienen.