

Dateinamen für die tägliche Datenverteilung (Download)

1. Namen der ZIP-Dateien

Die Namen der zum Download bereitgestellten ZIP-Archive sind folgendermaßen aufgebaut:

<Praefix>_group_<Gruppe>_<Datenart>.zip

Beispiele:

- 20100518_group_8008_tracks-daily.zip
- 20100518ch-16_group_-8208_covers-0300px-001.zip

Dateinamen bestehen ausschließlich aus druckbaren (7-bit-)ASCII-Zeichen, enthalten also keine Umlaute. Die einzelnen Abschnitte für <Praefix>, <Gruppe> und <Datenart> enthalten ausschließlich alphanumerische Zeichen und Bindestriche, jedoch keine Unterstriche "_" (RegExp/[[:alnum:]-]+/).

Für die einzelnen Abschnitte gilt folgendes:

<Praefix>

Bestehend aus zwei Angaben:

<Datum> = ("20100518")

Der Dateinamen beginnt mit einem Datum in 8 Ziffern als **JJJJMMTT** (ISO-8601-Format ohne Trennzeichen, "%Y%m%d") formatiert und bezieht sich auf den Empfangszeitraum der enthaltenen Daten.

[<Disambiguation> = ("de", "nl-1", "nl-2", ...)]

Ergänzend können an das <Datum> weitere Bezeichner angehängt werden. Folgende Ergänzungen sind definiert:

<Land>, ein 2 Zeichen langer *Country Code* nach ISO 3166-1 ("at", "be", "ch", "de" oder "nl").

-<Lfd.Nr.>, laufende Nummer von einer oder mehr Ziffern; für den Fall, dass eine einzelne ZIP-Datei für zu groß erachtet wird und daher in mehrere aufgeteilt werden soll (oder u.U. auch "-0").

Der <Praefix> garantiert bei alphabetischer Sortierung die richtige Reihenfolge und ermöglicht die Aufteilung der Daten in mehreren Archiven. Aufgrund der Variabilität sollten **Dateinamen-Muster so allgemein wie möglich** verwendet werden. Siehe Bemerkungen zur automatischen Verarbeitung der Dateien.

<Gruppe> ("8008")

Als Gruppenbezeichner dient jeweils eine Folge von Buchstaben, Ziffern und Bindestrichen (RegExp: /[[[:alnum:]-]+/), die mindestens ein Zeichen lang ist. *Üblicherweise* steht hier "all". Bei Einsatz von Gruppenfilter, können hier eine (oder mehrere mit Bindestrich getrennte) Phononet-Nummer(n) sowie ein beliebiger Wert stehen.

<Datenart> ("**tracks-daily**")

Die Bezeichner der Datenarten sind jeweils eine Folge von Buchstaben, Ziffern und Bindestrichen (RegExp: `/(cover_)?[[:alnum:]-]+/`), wobei "cover_" nur in Legacy-Dateinamen verwendet wird). Bindestriche dienen dabei der weiteren Untergliederung einer Datenart, wobei der erste Teil die Hauptdatenart bestimmt. Definierte Datenarten sind:

- **articles-**: Artikelstammdaten
 - Artikelstammdaten mit Löschrzeichen (`articles-daily-clr`); im legacy Format als `pri` bezeichnet.
 - Artikelstammdaten ohne Löschrzeichen (`articles-daily`)
- **tracks-**: Trackdaten
 - Die Datenart `tracks-daily`; im legacy Format als `trk` bezeichnet.
- **covers-**: Coverdaten
 - Coverdatenarten werden ergänzt um die Angabe zur Auflösung (maximale Kantenlänge) der Bilder in Pixeln ("px") und die Seitennummer (001 = Frontcover, 002 = Backcover, other = weitere Bilder). Beispiele:
 - `covers-0300px-001`
 - `covers-0300px-002`
 - `covers-0300px-other`
 - Die Datenarten `covers-0300px-001` und `covers-0300px-002` schließen die jeweiligen Thumbnails (080px) mit ein; im legacy Format wurden sie zudem als `cover_small` zusammengefasst.
 - Die Datenarten `covers-0600px-001`, `covers-0600px-002` wurden im legacy Format als `cover_big` zusammengefasst.
 - Die Datenarten `covers-0300px-other`, `covers-0600px-other` und `covers-1200px-other` enthalten weitere Bilder.
- **xmldata-**: Katalogdaten im XML Format.
 - Der zweite Teil bezeichnet die jeweilige Untermenge zusätzlicher Datenarten, die im Dokument enthalten sein können.

2. Namen der enthaltenen Dateien

Artikeldaten- und Trackdaten-Dateien

Die Namen der Artikelstammdatendateien bestehen wie bisher aus einer **Phono-Nummer und acht weiteren Ziffern**. Letztere entsprechen dem achtstelligen Tagesdatum (im ISO-8601-Format ohne Trennzeichen, "JJJJMMTT"). Falls die Daten mehrere Lieferanten betreffen, wird als Phono-Nummer "8095" angegeben.

Die Dateiendung ist ".pri" für Artikelstammdaten und ".trk" für Trackdaten.

Coverdateien

Entsprechend dem aktuellen Standard-Layout wird der Basisname einer Cover-Datei in drei Felder unterteilt, die durch Bindestriche ("-") getrennt sind:

1. **GTIN-13** (also immer 13-stellig mit Prüfziffer, einem UPC-12 wird eine "0" vorangestellt)
2. **Kantenlänge** als vierstellige Zahl plus "px" (die *größere* der beiden Kantenlängen in Pixeln)
3. **Seite** als dreistellige Zahl: "001" für Frontcover, "002" für Backcover. Zahlen ab "003" gelten für weitere Bilder (z.B. Szenenbilder oder Produktfotos, sofern verfügbar).

Die Dateiendung ist immer ".jpg".

Entsprechend dem alten Layout wird eine Kodierung mit maximal 8 Stellen im Dateinamen verwendet. Da Dateisysteme mit dieser technischen Limitierung nicht mehr üblich sind, stellen sie bitte möglichst bald auf das aktuelle Layout um.

Beispiele:

Dateiname aktuell	Dateiname alt
0123456789012-0080px-001.jpg	3rduqx0c.j01
1234567890128-0300px-001.jpg	3rduqx0c.j31
0123456789012-0600px-002.jpg	3rduqx0c.j62
1234567890128-1200px-002.jpg	3rduqx0c.jc2

Beispiele

Dateiname(n) aktuell	Dateiname alt	Enthaltene Datei(en)
20100518_group_all_articles-daily-clr.zip	20100518.pri.zip	809520100518.pri (Änderungsdatei)
20100518_group_8008_articles-daily-clr.zip	20100518_group_8008_pri.zip	800820100518.pri (Änderungsdatei)
20100518nl_group_8008_articles-daily.zip	--"--	800820100518.pri (Soll-Zustand)

20100518_group_8008_tracks-daily.zip	20100518_group_8008_trk.zip	800820100518.trk
--------------------------------------	-----------------------------	------------------

20100518_group_8008_covers-0300px-001.zip	--"--	0123456789012-0080px-001.jpg, 1234567890128-0080px-001.jpg, 0123456789012-0300px-001.jpg, 1234567890128-0300px-001.jpg, ...
20100518_group_8008_covers-0300px-002.zip	--"--	0123456789012-0080px-002.jpg, 1234567890128-0080px-002.jpg, 0123456789012-0300px-002.jpg, 1234567890128-0300px-002.jpg, ...
20100518_group_8008_covers-0600px-001.zip	--"--	0123456789012-0600px-001.jpg, 1234567890128-0600px-001.jpg, ...
20100518_group_8008_covers-0600px-002.zip	--"--	0123456789012-0600px-002.jpg, 1234567890128-0600px-002.jpg, ...
20100518_group_8008_covers-1200px-001.zip	--"--	0123456789012-1200px-001.jpg, 1234567890128-1200px-001.jpg, ...
20100518_group_8008_covers-1200px-002.zip	--"--	0123456789012-1200px-002.jpg, 1234567890128-1200px-002.jpg, ...

20100518_group_8008_covers-0300px-001.zip 20100518_group_8008_covers-0300px-002.zip	20100518_group_8008_cover_small.zip	3rduqx0c.j01 3rduqx0c.j31 ...
20100518_group_8008_covers-0600px-001.zip 20100518_group_8008_covers-0600px-002.zip	20100518_group_8008_cover_big.zip	3rduqx0c.j62 3rduqx0c.jc2 ...

Bemerkungen zur automatischen Verarbeitung der Dateien

Grundsätzlich sollten **Dateinamen-Muster immer so allgemein wie möglich** gehalten werden. So ist es beispielsweise nicht notwendig bei der Abholung von Dateien einen bestimmten Gruppen-Namen (das ist der Namensteil hinter "_group_") vorauszusetzen. Dieser dient im Wesentlichen zur internen Verwendung durch die PHONONET-Software, um einzelne Datenpakete den jeweils berechtigten Empfängern zuordnen zu können. Die konkret verwendeten Namen können sich aus verschiedenen Gründen ändern und sollten nicht als konstant angesehen werden.

Außerdem kann leider **kein fester Zeitpunkt** garantiert werden, zu dem die Dateien auf jeden Fall bereitstehen werden, auch wenn dies in aller Regel in den frühen Morgenstunden der Fall sein wird. Falls die Daten also nur einmal täglich automatisiert abgeholt werden, kann es vorkommen, dass die Dateien zum entsprechenden Zeitpunkt noch nicht fertig sind. In diesem Fall muss der Download später (oder am nächsten Tag) nachgeholt werden.

Um beispielsweise alle neuen Trackdaten-Updates zu bekommen, sollten daher einfach alle Dateien "*_group_*_track*.zip" abgeholt und die erfolgreich heruntergeladenen Dateien anschließend auf dem Server gelöscht werden. (Der Download war erfolgreich, wenn die ZIP-Datei sich ohne Fehler öffnen/entpacken lässt.) Durch das Format des Datums-Präfixes genügt es, die ZIP-Dateien alphabetisch zu sortieren, um so die richtige chronologische Reihenfolge für das Einlesen der Daten herzustellen.

Folgende Dateinamen-Muster sind in den allermeisten Fällen ausreichend, um die Dateien finden und zuordnen zu können:

Allgemein (ggf. inklusive Land)	mit Datum (in Ausnahmefällen, ggf. inklusive Land)	alte Bezeichnung
<u>*_group_*_article*.zip</u>	20160901*_group_*_article*.zip	
<u>*be_group_*_article*.zip</u> <u>*nl_group_*_article*.zip</u>	20160901be*_group_*_article*.zip 20160901nl*_group_*_article*.zip	<u>*_group_*_pri*.zip</u>
<u>*_group_*_track*.zip</u> ...	20160901*_group_*_track*.zip ...	<u>*_group_*_trk*.zip</u>
<u>*_group_*_cover*.zip</u> ...	20160901*_group_*_cover*.zip ...	
<u>*_group_*_xmldata*.zip</u> ...	20160901*_group_*_xmldata*.zip ...	