

# **HIRLAM-i andmestruktuurid ja GRIB formaat.**

23. märts 1998. a.

HIRLAM-i puhul on tegemist erinevat tüüpi andmestruktuuridega, millest igalühel neist on oma eesmärk ja andmesalvestuse moodus.

- 1) Meteoroloogilised väljad.
- 2) Vaatlusandmed.
- 3) Statistika failid.

Lisaks neile kolmele põhilisele andmestruktuurile on võimalik HIRLAMI-is tekitada aegridade ja kliima andmete faile.

## **1. Meteoroloogiliste väljade salvestamine.**

Kogu meteoroloogiliste väljade infot hoitakse HIRLAMis pakitud kujul. Pakkimisel kasutatakse GRIB formaati. GRIB formaat on tõhus vahend andmete salvestamiseks pakitult, muutes seejuures kergemaks infovahetuse eri süsteemide arvutite vahel.

Meteoroloogiliste väljade fail koosneb järjestikku kirjutatud GRIB teadetest. Seejuures on teadete järjekord ära määratud andmekirjega. Andmekirjes paikneb kogu informatsioon, mis on vajalik väljade faili ja selle sisu määratlemiseks.

Igat eraldiseisvat välja iseloomustavad kolm parameetrit: välja tüüp, parameetri väärtus ja paiknemise tase. Lisaks sellele on GRIB teadetes mitmesugust lisainformatsiooni tekitamise aja, võrgustiku jms. kohta. See tekitab küll mõnesugust informatsiooni dubleerimist, kuid tasub lõpuks ennast ära, kuna GRIB teadet on siis võimalik lugeda ilma lisainfot kasutamata. Seetõttu räägitakse GRIB teadete nn. enesekirjeldatavusest.

### **Andmekirje (DDR - data description record).**

Andmekirje koosneb INTEGER või REAL tüüpi muutujatest või massiividest, mis määravad ära GRIB teadete parameetrid ja väljafailis paiknemise järjekorra. Andmekirje tekitatakse programmi töö käigus ning tema sisu on toodud järgnevas tabelis.

	Muutuja	tüüp/dim.	Tähendus
1	NCODHL	I	koodi defineerimise number = 1000
2	NLDRHL	I	kirje pikkus 1000
3	NLNXHL	I	järgmise deskriptori pikkus (0 kui pole)
4	MRCLHL	I	maksimum kirje pikkus
5	MDRLHL	I	maksimaalne andme kirje pikkus
6	NTYPHL	I	failikorraldus tüüp
7	NEXPHL	I	eksperimendi/mudeli identifikaator
8	NMDIHL	I	puuduva andmestiku identifikaator
9	NIDFHL	I(12)	varu

Aja sektsioon (INTEGER tüüp)

21	NDTVHL	I	kontrollkuupäev/aeg
22	NSCVHL	I	kontrollaeg, sekundid peale NDTVHL
23	NDTBHL	I	andmebaasi kuupäev/aeg
24	NSCBHL	I	andmebaasi aeg, sekundid peale NDTBHL
25	NFLHHL	I	ennustuse kestus tundides
26	NFLSHL	I	ennustuse kestus, sekundid peale NFLHHL
27	NDORHL	I	andmete päritolu 0:analüüs 1:initsialiseeritud analüüs ja ennustus
28	NIDTHL	I(13)	kuupäevade ja aegade defineerimise varu

Horisontaalse võrgu defineerimise sektsioon (INTEGER tüüp)

41	NPRJHL	I	projektsiooni tüüp
42	NPRCHL	I	võrgu deskriptorite olemasolu identifikaator
43	NLONHL	I	võrgupunktide arv pikki x (tsonaalset) telge
44	NLATHL	I	võrgupunktide arv pikki y (meridionaalset) telge

Vertikaalse võrgu defineerimise sektsioon (INTEGER tüüp)

45	NLTPHL	I	vertikaalse taseme tüüp
46	NLEVHL	I	tasemete arv
47	NPPLHL	I	parameetrite arv taseme kohta
48	NRFLHL	I	algtasemete arv
49	NLPTHL	I(40)	viidad mitmetasemeliste väljade eri tasemetele

Mitmetasemeliste väljade defineerimise sektsioon (INTEGER tüüp)

89	NMLFHL	I	mitmetasemeliste väljade arv
90	NWMMHL	I(40)	mitmetasemeliste väljade WMO koodid
130	NMPTHL	I(40)	viidad mitmetasemeliste väljade esimesele tasemele

Ühetasemeliste väljade defineerimise sektsioon (INTEGER tüüp)

170	NSLFHL	I	ühetasemeliste väljade arv
171	NWMSHL	I(40)	ühetasemeliste väljade WMO koodid
211	NSPTHL	I(40)	viidad ühetasemeliste väljadele
251	NSLTHL	I(40)	ühetasemeliste väljade väljatüüp

Horisontaalse võrgu defineerimise sektsioon (REAL tüüp)

291	APLOHL	R	nihutatud pooluse pikkuskraad (kraadides)
292	APLAHL	R	nihutatud pooluse laiuskraad (kraadides)
293	AWESHL	R	ala lääneserv (kraadides)
294	AEASHL	R	ala idaserv (kraadides)
295	ALALHL	R	ala viimane laius (kraadides)
296	ALAFHL	R	ala esimene laius (kraadides)
297	DLONHL	R	meridiaanide võrgusamm
298	DLATHL	R	paralleelide võrgusamm
299	GRIDHL	R(28)	varu teiste võrkude defineerimiseks

Vertikaalse võrgu defineerimise sektsioon (REAL tüüp)

327	ALEVHL	R(40,4)	tasemete defineerimise parameetrid
487	RLEVHL	R(4)	algtasemed

Mitmetasemeliste väljade defineerimise sektsioon (REAL tüüp)

491	STMXHL	R(40)	mitmetasemeliste väljade x-nihutus
531	STMYHL	R(40)	mitmetasemeliste väljade y-nihutus
571	STMZHL	R(40)	mitmetasemeliste väljade z-nihutus

Ühetasemeliste väljade defineerimise sektsioon (REAL tüüp)

611	STSXHL	R(40)	ühetasemeliste väljade x-nihutus
651	STSYHL	R(40)	ühetasemeliste väljade y-nihutus
691	SLEVHL	R(40,4)	ühetasemeliste väljade taseme parameetrid

GRIB teadete pikkused.

851 LRECHL R(150) GRIB teadete pikkused nende kirjutamise järjekorras

## **GRIB formaat.**

GRIB (GRIdded Binary) on WMO poolt loodud standard võrgustikul paiknevate andmete esitamiseks katkematu binaarse jadana.

GRIB “teated” on HIRLAM-i puhul kõige olulisemad andmestruktuuri osad. Selles formaadis antakse nii initsialiseerimisprotsessis kasutatavad meteoroloogilised algväljad kui ka ennustuse lõppväljund. GRIB “teated” üheks olulisemaks iseloomujooneks nn. eneskirjeldatavus. Sellise lähemise eeliseks on kerge kohanevus muutustega.

## **Andmete pakkimise meetodid.**

HIRLAMis on kasutusel kaks alternatiivset meetodit andmete pakkimiseks:

- fikseeritud bitisõne kasutamine
- muutuva bitisõne kasutamine.

Sõltumata kasutatud pakkimise meetodist peab korralik GRIB dekooder failist kätte saama õiged numbrid.

GRIB teade koosneb kuuest osast millest kaks pole kohustuslikud.

- Indikaatorseksioon (0)
- Faili määratluseksioon (1)
- Võrgu kirjelduse seksioon (2) - pole kohustuslik
- Bitikaardi (Andmekaaardi) seksioon (bit map section) (3) - pole kohustuslik, HIRLAM-is puudub
- Binaarsete andmete seksioon (4)
- Lõpu seksioon (5)

Võrgu kirjelduse seksioon pole küll kohustuslik, kuid on rangelt soovitatav, et ei tekiks küsimusi geograafilisest võrgustikust mingi välja puhul.

**Oktett** - kaheksast järjestikusest bitist koosnev info ühik. kõik numbrid on sobivalt skaleeritud täisarvud.

### **Indikaatorseksioon.**

Indikaatorseksioon aitab saada infot faili alguse kohta inimesele loetavas formaadis ning määrata faili kogupikkust ja kasutatava GRIB standardi numbrit.

Okteti nr.	Tähendus
1-4	String 'GRIB' ASCII koodis
5-7	GRIB "teate" kogupikkus oktettides (koos kõikide seksioonidega)
8	Standardi number (1)

### **Faili määratlusseksioon.**

Seksioon sisaldab mitmesugust infot faili tekitamise kuupäeva, ilmaennustuskeskuse, meteoroloogiliste suuruste, mudeli tasemete jms. kohta. Tavaliselt on toote kirjelduse seksioon 28 okteti pikkune, kuid võib olla ka vastavalt ilmaennustuskeskuse vajadustele pikem. Seepärast on oluline, et esimeste oktettidega määrataks ära bloki pikkus.

Okteti nr.	Tähendus
1-3	Seksiooni pikkus oktettides
4	Parameetrite tabeli versiooni number. HIRLAMil on see 1
5	Ilmaennustuskeskuse identifikaator (82)
6	Faili tekitamise protsessi identifikaator või tekitanud mudel (20)
7	Võrgustiku identifikaator (255 - viitab sellele, et kaart defineeritakse võrgukirjelduse seksioonis.)
8	Identifikaator, mis näitab võrgukirjelduse või andmekaardi seksiooni olemasolu (10000000) Koodide tabel 1.
9	Parameetri ja ühiku identifikaator Koodide tabel 2.
10	Taseme või kihi identifikaator Koodide tabel 3.

11-12	Taseme või kihi kõrgus, rõhk jne Koodide tabel 3.
13	Aasta
14	Kuu
15	Päev
16	Tund
17	Minut
18	Ennustuse ajaühik (1) Koodide tabel 4.
19	Ennustusperiood
20	Ennustusperiood
21	Ajalise ulatuse indikaator Koodide tabel 5.
22-23	Arv keskmistamisel
24	Keskmistamisel või akumulereerimisel puuduv arv
25	Sajand
26	Allkeskuse identifikaator (pole veel sanktsioneeritud)
27-28	Detsimaalse skaleerimise faktor
29-40	Reserveeritud (ei pea olema kasutuses)
41-...	Reserveeritud ennustava keskuse tarbeks

### **Võrgu kirjelduse sektsioon.**

Selle bloki eesmärk on kirjeldada modelleerimisel kasutatavat võrgustikku, kui see pole defineeritud standardi järgi.

Okteti nr.	Tähendus
1-3	Sektsiooni pikkus
4	Vertikaalse koordinaadi parameetrite arv
5	Vertikaalsekoordinaadi parameetrite loetelu asukoht (okteti number)
6	Andmeesitustüüp (10)
7-32	Võrgu kirjeldus vastavalt andmesitustüübile
7-42	Kirjeldus Lambertt'i või Mercatori projektsioonis võrgule
7-44	Kirjeldus Space View perspektiivis võrgule
PV	Vertikaalse koordinaadi parameetrite loeteelu
PL	Igas reas olevate punktide loetelu (kasutatakse kvaasiregulaarsete võrkude puhul)

### **Bitikaardi (Andmekaaardi) sektsioon.**

Bitikaart (bit map) koosneb vastavalt võrgustiku kirjeldusele üksteisegakülnevatest bittidest. Bitt, mis on võrdne 1 näitab andmestiku olemasolu antud võrgupunkti jaoks ja 0 selle puudumist. Bitikaardi bloki eesmärk on see kaart defineerida või viidata keskuse poolt eelnevalt määratule. Sektsioon on kasulik kui on vaja kokku suuruda väljasid milles on küllalt suured alad määramata.

Okteti nr.	Tähendus
1-3	Andmekaaardi bloki pikkus
4	Kasutamata bittide arv selle bloki lõpus
5-6	Number: =0 järgneb andmekaaart mingi muu – viide keskuse poolt määratud kaardile
7-nnn	Andmekaaart

HIRLAM jätab selle bloki kasutamata.

### **Binaarne andmete sektsioon.**

See sektsioon sisaldab pakitud kujul andmeid ja binaarse skaleerimise infot, et rekonstrueerida andmete algset kuju. Andmete lahtipakkimiseks vajalik kümnendskaala faktor on antud tootekirjeldusblokis.

Okteti nr.	Tähendus
1-3	Sektsiooni pikkus oktettides
4	Bitid 1 kuni 4 Koodide tabel 11. Bitid 5-8 kasutamata bittide arv sektsioonis 4
5-6	Binaarne skaleerimisfaktor
7-10	Ujukoma esituse miinimumväärtus.
11	Bittide arv, millesse andmestiku punktid pakitakse
12 -nnn	Andmed
14	4. okteti laiendus pole kohustuslik

### **Lõpu sektsioon.**

See sektsioon koosneb 4 oktettist ja sisaldab inimesele loetavas ASCII koodis numbreid '7777'. Lõpustringi võib kasutada ka arvutis faili terviklikkusekontrolliks. Siiski ei tohiks seda kasutada stringi järgi otsimises, kuna kombinatsioon võib esineda ka mujal binaarses andmevoos.

GRIB-failidega tööks on HIRLAMil alamprogrammide raamatukogu libgrib.a.

## **2. Vaatlusandmed ja BUFR formaat.**

BUFR (Binary Universal Form for the Representation of meteorological data) on WMO standard, mis on mõeldud võrgustikul mitte paiknevate andmete esitamiseks katkematu binaarse jadana. Ta spetsiaalselt kohandatud meteoroloogiliste vaatlusandmete kodeerimiseks. Nagu ka GRIB formaadi puhul on BUFR kirjete omapäraks enesekirjeldatavus. Ka BUFR "teade" koosneb kuuest osast, mida nimetatakse sektsioonideks.

- Indikaatorsektsioon. (0)



- Identifitseerimise sektsioon. (1)
- Mittekohustuslik sektsioon. (2)
- Andmete kirjeldamise sektsioon (3)
- Andmete sektsioon. (4)
- Lõpu sektsioon. (5)

HIRLAM-is on olemas tarkvara ainult BUFR-failide andmete dekodeerimiseks, mis ei võimalda vaatlusandmete salvestamist selles formaadis. BUFR failidega opereerimiseks on olemas HIRLAMil raamatukogu BUFR.

### **3. Statistika failid.**

Andmehõive (data assimilation) süsteemi ühe osana on HIRLAMis kasutusel statistika failid. Nendes failides paiknevad vaatlusandmed ja kvaliteedi kontrolli võimalused, mis edastatakse analüüsisüsteemi poolt iga anmekogumi jaoks. Samuti sisaldavad nad vaadeldud väärtuste hälbeid ennustatud väärtusest mingis vaatluspunktis.

Statistika faile võib kasutada mitmetel eesmärkidel. Vaatlus miinus analüüs statistikat kasutatakse süsteemi üldise suutlikkuse hindamiseks. Sellist statistikat võib arvutada eraldi iga vaatlussüsteemi jaoks. Sellised arvutused võivad välja tuua väiksemaid süstemaatilisi vigu vaatlustes, mida muul moel on raske määrata. Samuti võib nii avastada nõrku kohti mõningate vaatluste käsitlemisel andmehõivesüsteemis eneses. Seda infot võib kasutada ka eraldi iga vaatlusjaama kohta. Seega saab välja selgitada jaamu, mis teevad süstemaatiliselt halbu vaatlusi. Nende arvutuste juures on oluline, et kasutataks kõiki kvaliteedi kontrolli võimalusi erinevates analüüsi faasides. Seega annavad need andmed väärtuslikku informatsiooni andmehõivesüsteemi omaduste ja samuti vaatluste kvaliteedi kohta.

Statistika faile võib samuti kasutada sisendist sõltuvate struktuurifunktsioonide loomiseks ja arendamiseks. Praeguse analüüsi süsteemi juures kasutatavate struktuurifunktsioonide nõrkuseks on nende sõltumatus geograafilistest koordinaatidest ja voolamisest. Seega pole neid võimalik kasutada näiteks atmosfääri barokliinse struktuuri hindamiseks. Statistika failide andmeid on võimalik kasutada realistlikemate struktuurifunktsioonide empiiriliseks väljaarendamiseks.

### **4. Aegread.**

Ilmaennustusmudel HIRLAM võimaldab ka arvutada meteoroloogiliste suuruste aegridasid mingi geograafilise punkti jaoks. Aegridade failides saab salvestada mudeli arvutatud välju ja parameetreid vajalikes punktides suure (kuni iga ajasammu tagant) ajalise lahutusega. Failis paiknevad

aegread identifitseeritakse nende deskriptorite järgi. Deskriptor on ühilduv BUFR-formaadiga ning sisaldab punkti geograafilistkoordinaati, ja määrab, kas väljad on ühetasemelised või mitmetasemelised. Mitmetasemeliste väljade puhul peab olema antud ka vastava taseme vertikaalnekoordinaat. Selliseid aegridasid on võimalik kasutada näiteks meteogrammide tegemiseks. Alamprogrammid  
INITSF FINTSF PUTTSF WRITRF

## **5. Kliima andmete failid.**

HIRLAM kasutab kliima andmestikuga töötamiseks ECMWFi kliima andmete süsteemi ilma oluliste muutusteta. HIRLAMi kliima andmed salvestatakse kuude kaupa kasutades samuti GRIB formaati. Kliima failidega tööks on alamprogramm CLGRIB.