



Fedora i18n

Past, Present, Future

Jens Petersen

Flock 2014, Prague

About

- Jens Petersen
- Red Hat since 2001
- i18n Engineering Team
- Born in Denmark
- Lived 15 years now in Japan
- Fedora Haskell SIG

FOSS and i18n

Freedom to read and write in
one's own native language

OS i18n

- Encodings and Locales
- Fonts
- Input Methods
- Packaging



Internationalization vs Localization

i18n:

- Encodings and locales
- International fonts and complex text rendering
- Input methods for complex text input
- Gettext
- Language tools

l10n:

- Software translation
 - strings extracted using gettext, etc
- Documentation translation
 - Publican, etc
- Transifex, Zanata, etc



Encodings and default locales

Past

*Native charset
encodings and locales*

iso8859-{1,2,...},
EUC, S_JIS, GB2312,
etc, etc, etc

*Rpm build still defaults to
LANG=C !*

Present

*Unicode encoding
and locales*

UTF-8

*Note: Fedora console
does not use CIJK
locales*



Installer



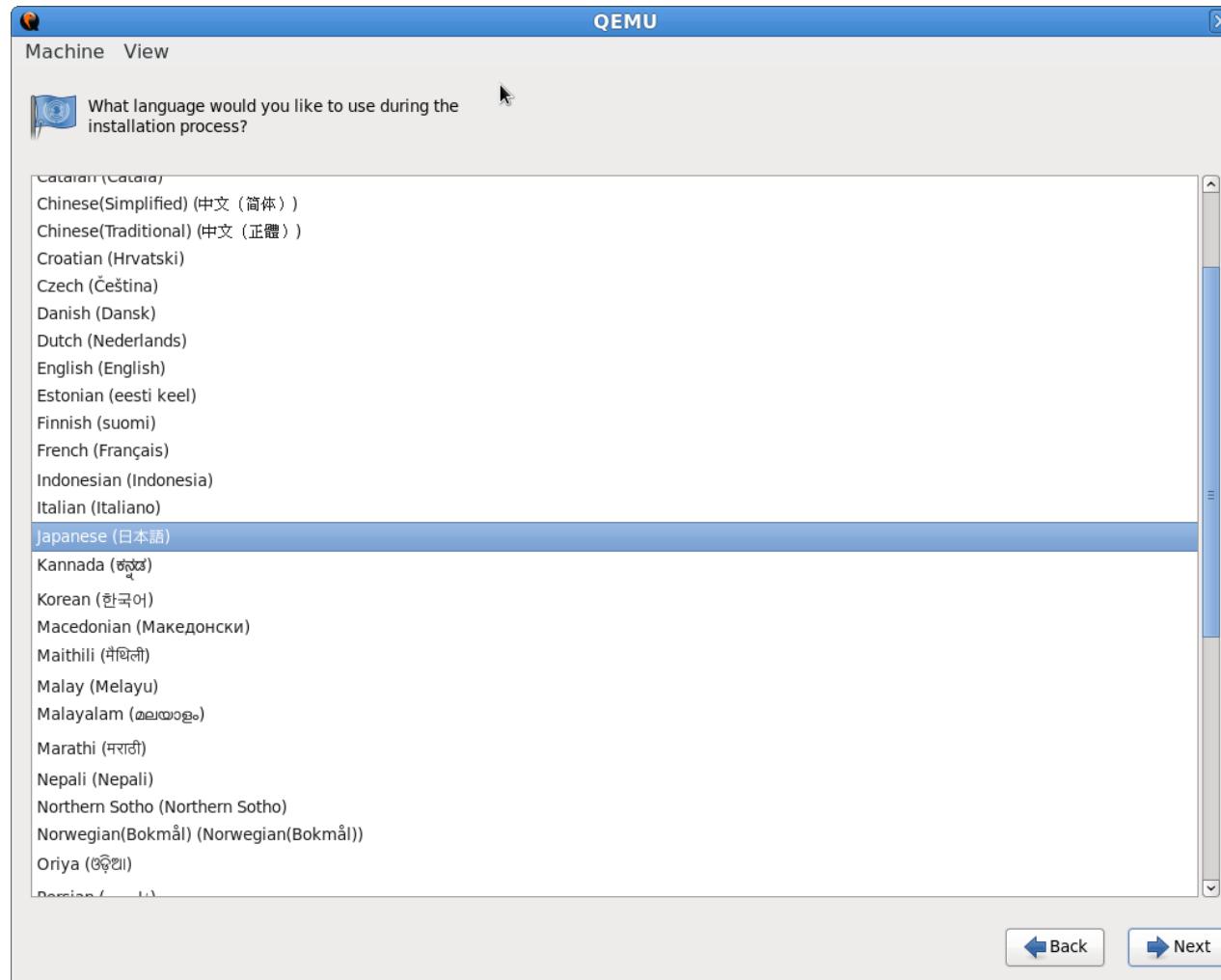
Installer language (EL5 era)



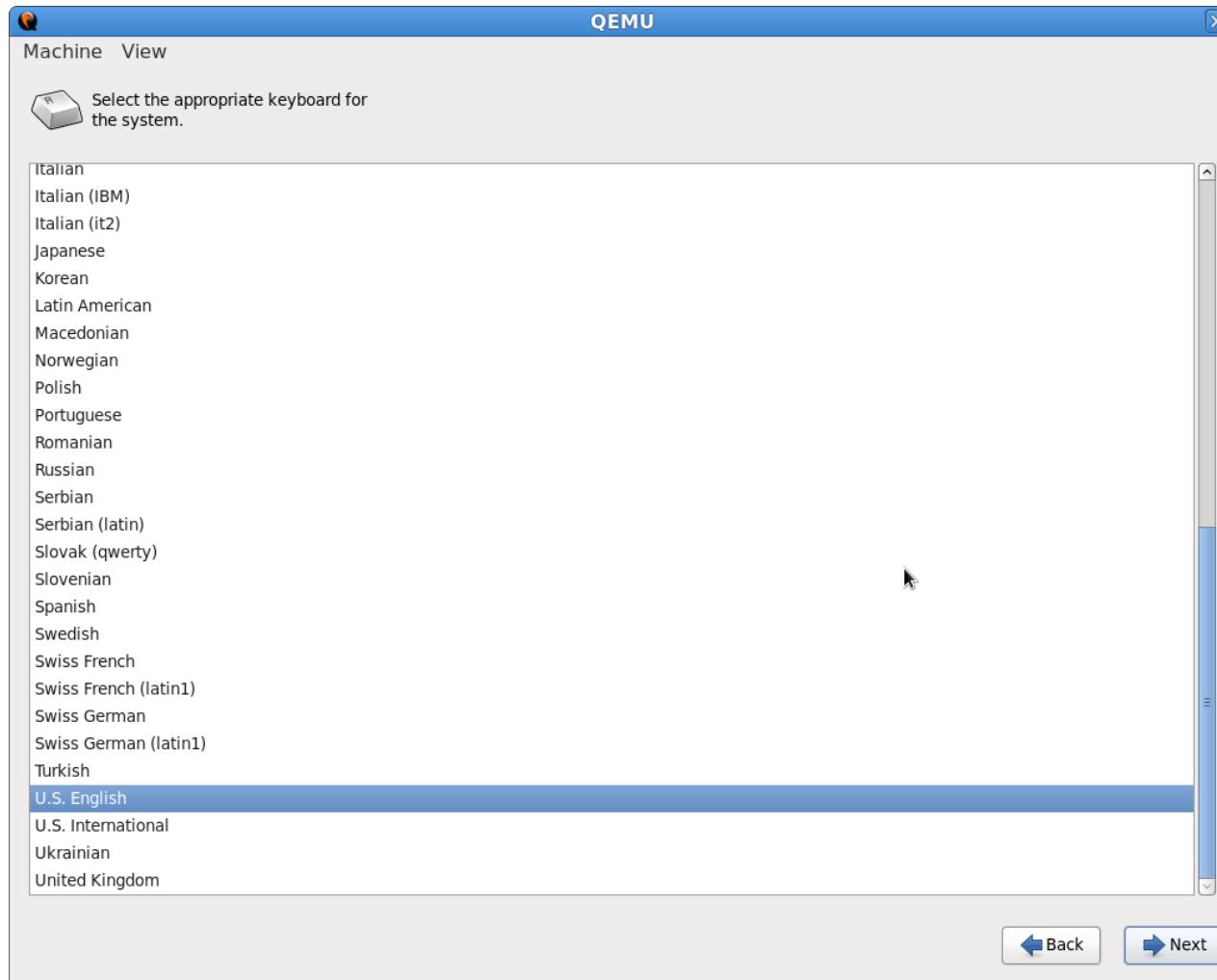
Installer keyboard selection (EL5 era)



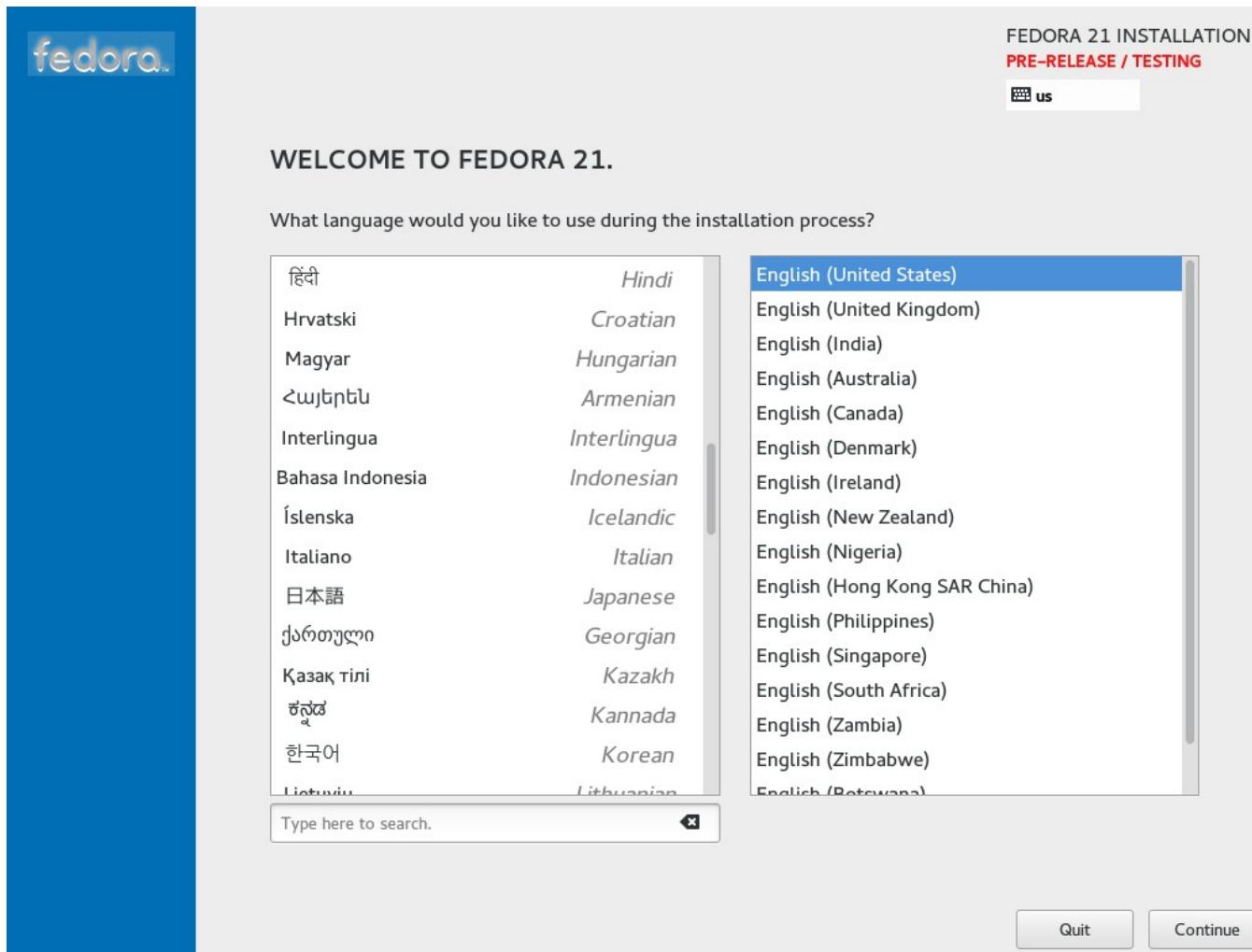
Installer language (EL6 era)



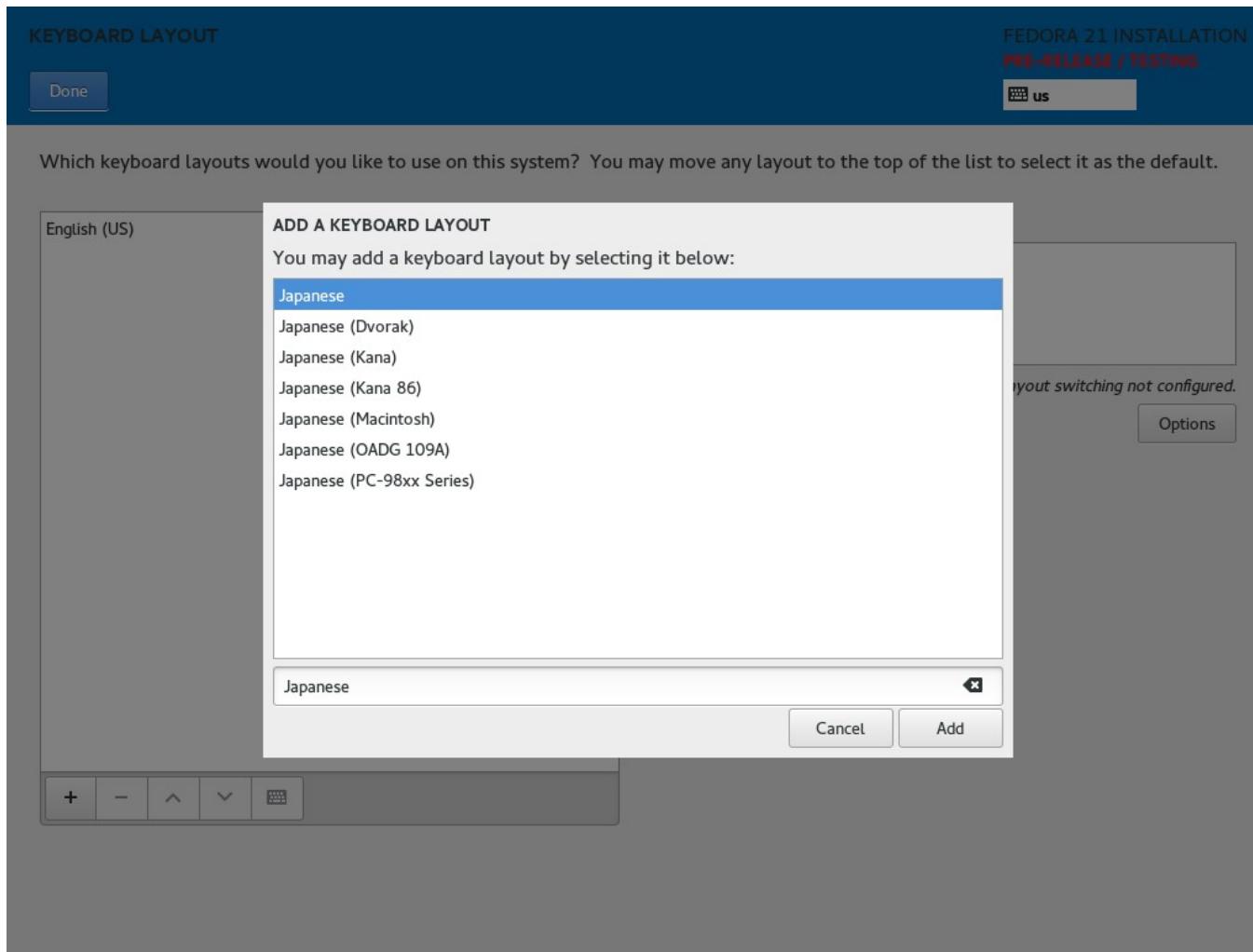
Installer keyboard selection (EL6 era)



Installer language selection (F20/F21)



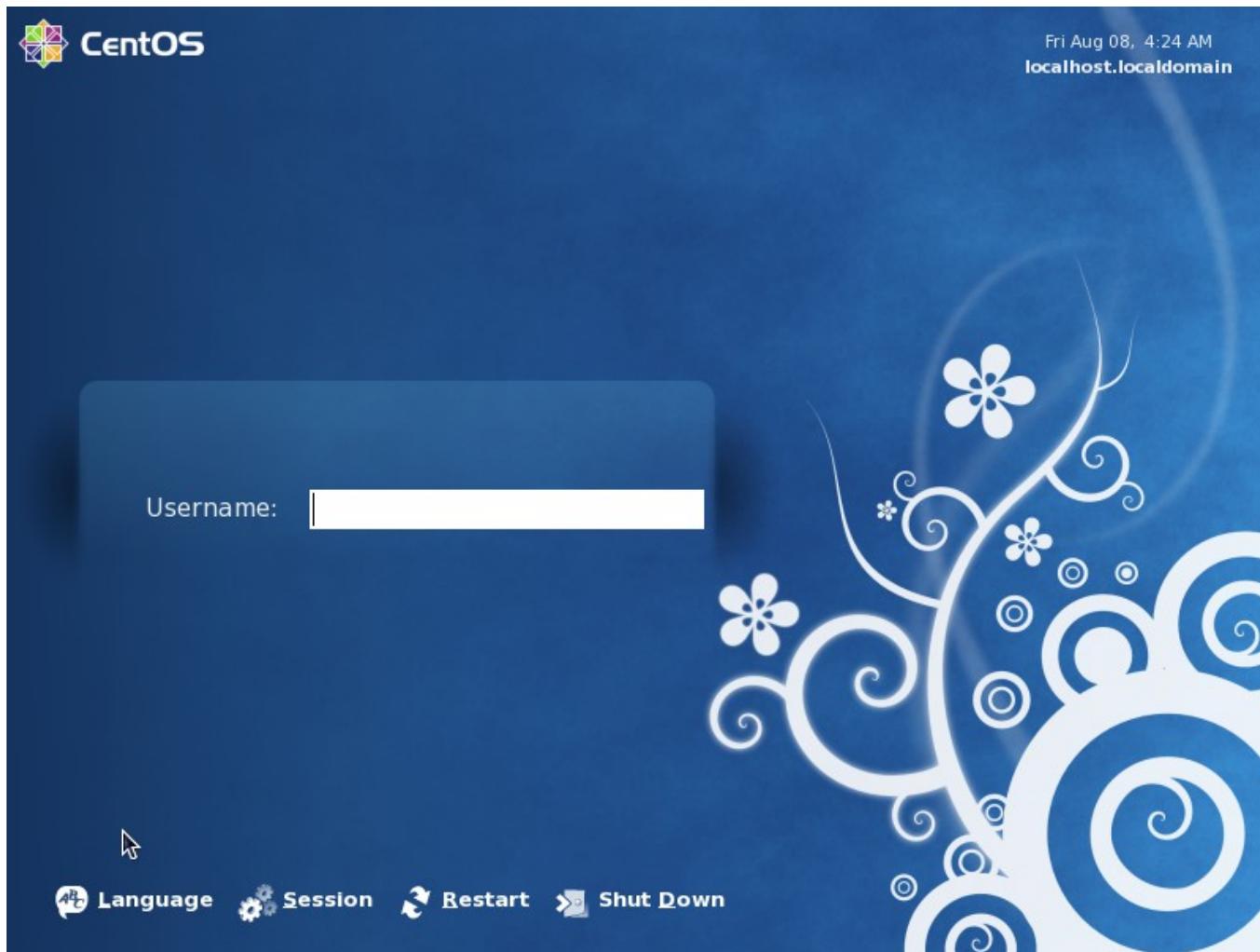
Installer keyboard selection (F20/F21)



GDM



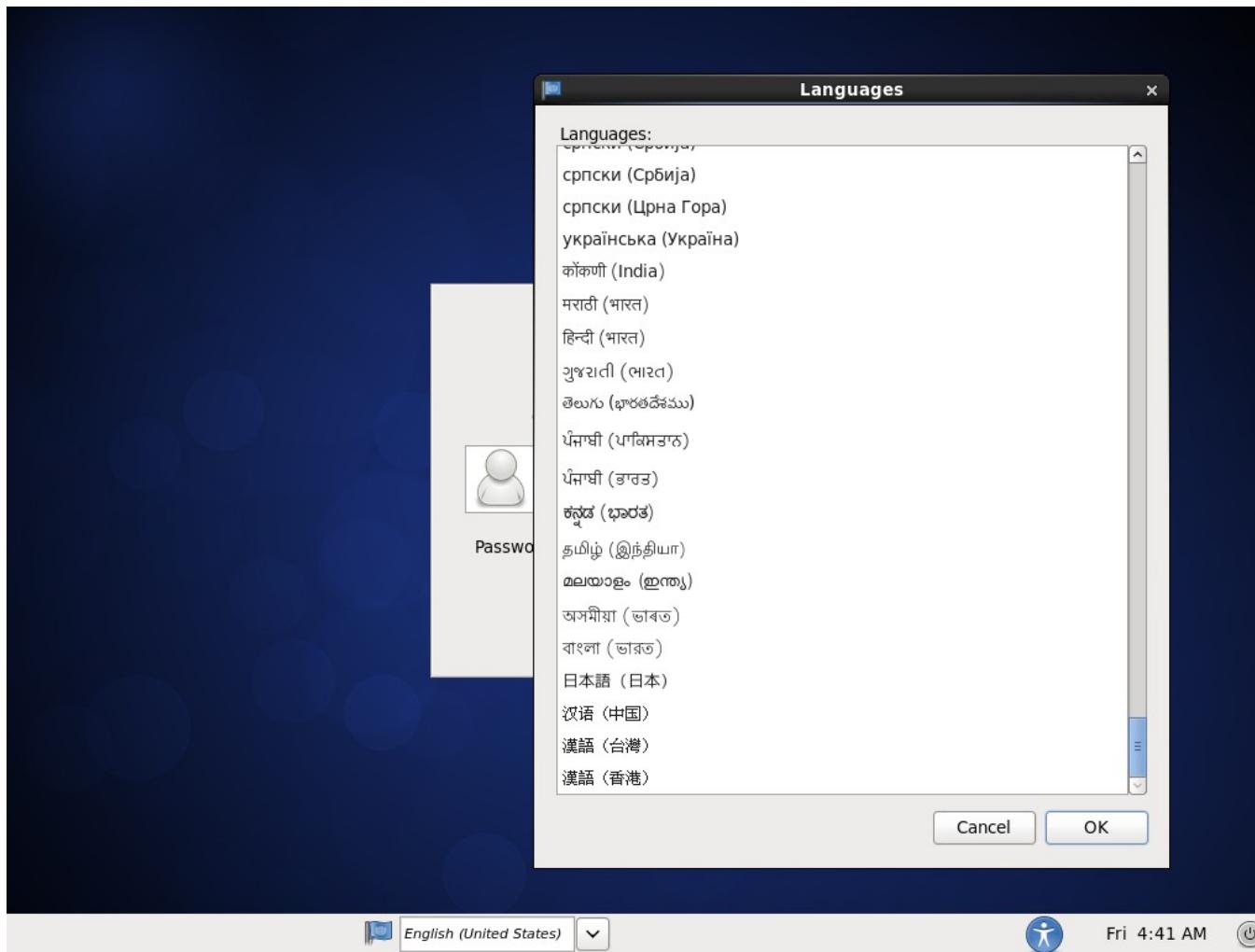
GDM (EL5 era)



GDM language (EL5)



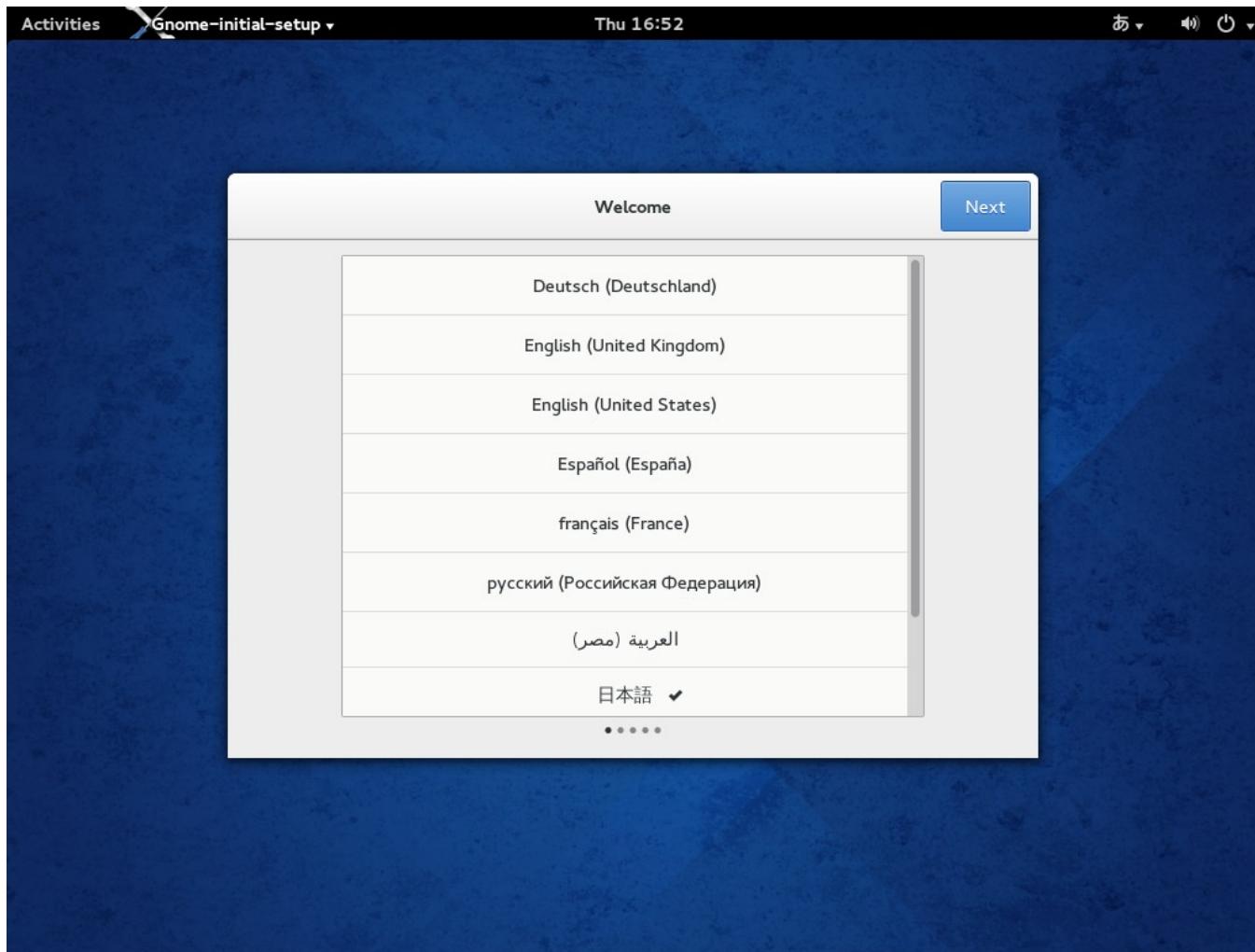
GDM (EL6 era)



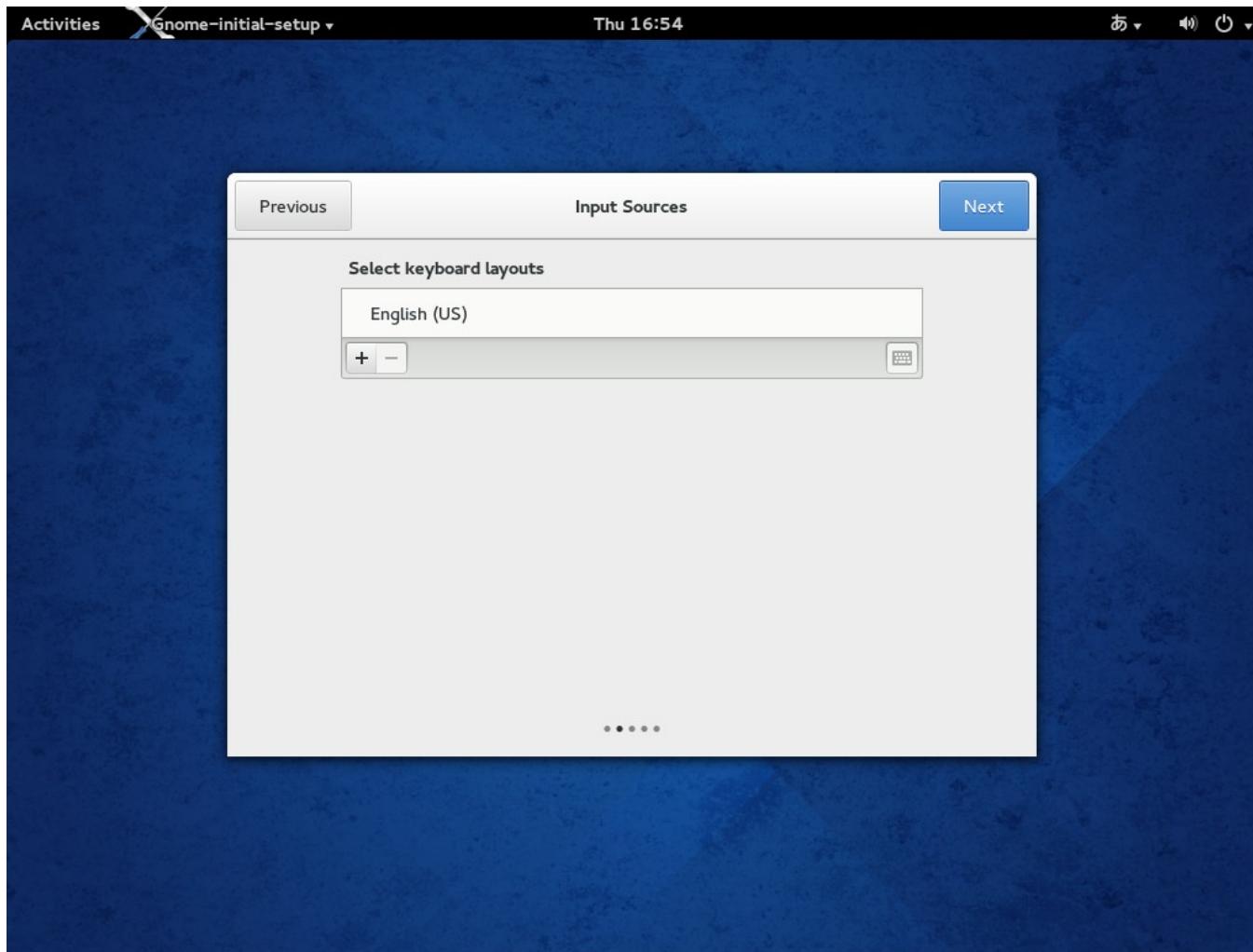
Current GDM



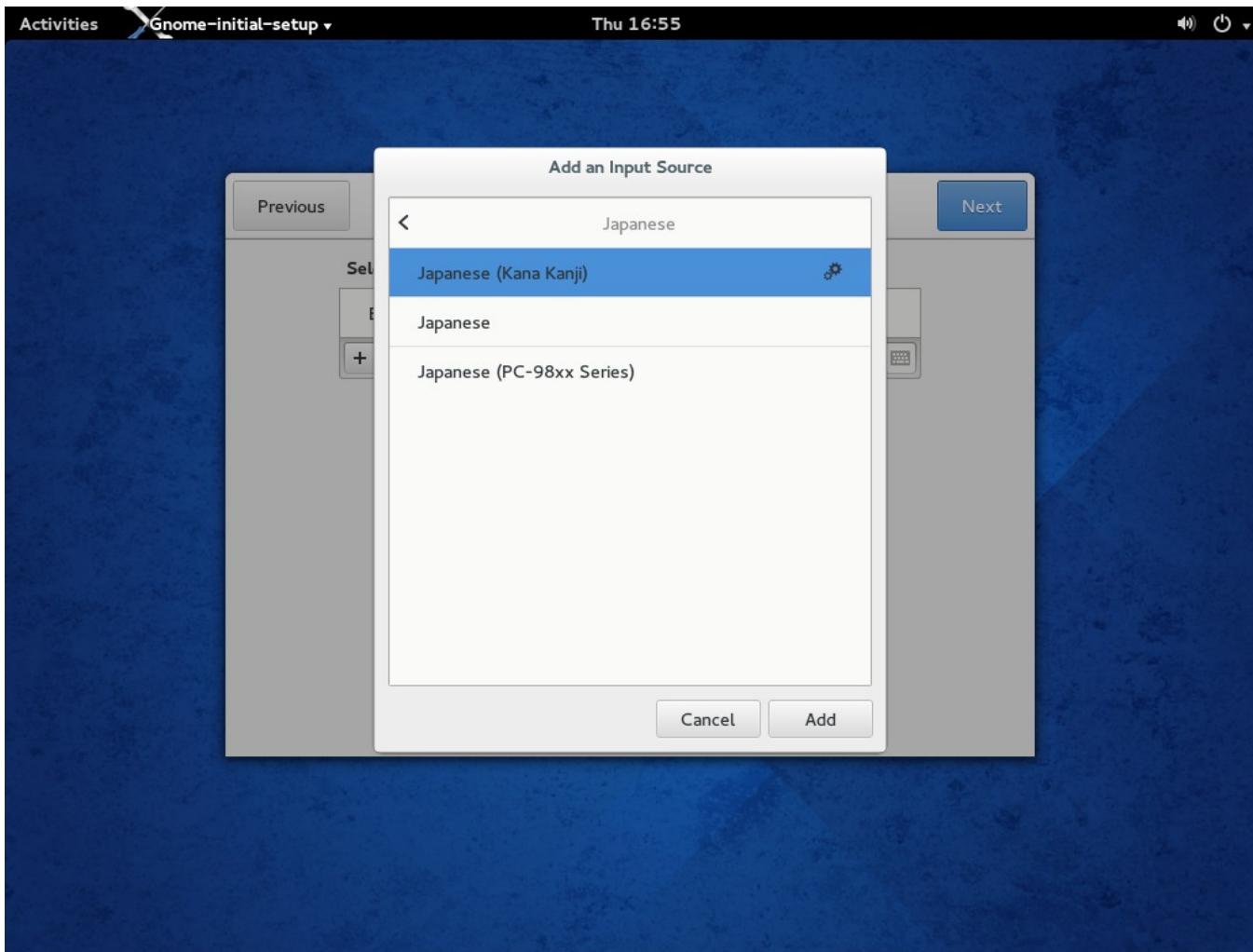
Gnome-initial-setup



Gnome-initial-setup Input Sources



Gnome-initial-setup IME



Fonts and Rendering



Evolution of fonts and rendering

- Bitmap fonts
- Outline fonts
 - TrueType, OpenType

xfs

Fontconfig

X fontsets

Pango

FreeType

Harfbuzz



Fonts coverage

	Qafář af	Ағысшәа	Acèh	Afrikaans	Akan	Alemannisch	አማርኛ	Aragonés	Ænglisc	عەپتى	مَصْبَحٌ	অসমীয়া	Asturianu	Ӑвар	Aymar aru	Azərbaycanca	Башкортса	Boarisch	Žemaitėška	Bikol
Central	Беларуская	Беларуская (тарашкевіца)		Български	ମୋଜପୁରୀ	Bislama	Bahasa Banjar	Bamanankan	ବାଙ୍ଲା	ଶ୍ରୀମଦ୍ଭଗବତ	ବିଷ୍ଣୁଲିଙ୍ଗ ମଣିପୁରୀ	Brezhoneg	Bosanski	ନୋଟାର	Буряад	Català	Chavacano de			
Zamboanga	Mìng-děng-ngū	Нохчийн	Cebuano	Chamoru	Choctaw	CWY	Tsetsêhestâhese	كُورُك	Corsu	Nēhiyawēwin / Ꮚ'අଲ୍ଲାଡ଼	Qırımtatarca	Čeština	Kaszëbszli	Словѣньскъ / ছାନ୍ଦମାର୍ଗବନ୍ଧ						
Чӑвашла	Cymraeg	Dansk	Deutsch	Zazaki	Dolnoserbski	କୁର୍ରାଷ୍ଟାର୍	ଶ୍ରୀମଦ୍ଭଗବତ	Ewëgbe	Ελληνικά	Emiliàn e rumagnòl	English	Esperanto	Español	Eesti	Euskara	Estremeñu	فُل	Fulfulde	Suomi	
Võro	Na Vosa Vakaviti	Føroyskt	Français	Arpetan	Nordfriisk	Furlan	Frysk	Gaeilge	Gagauz	贛語	Gàidhlig	Galego	ગેંગી	Avañe'ẽ	ମୁତ୍ସିକ	ગુજરାતୀ	Gaelg	Hausa	客家語/Hak-kâ-ngî	
Hawai'i	ହାଵାଇ	ହିନ୍ଦୀ	Fiji Hindi	Hiri Motu	Hrvatski	Hornjoserbsce	Kreyòl ayisyen	Magyar	କ୍ଷୁଣ୍ଟପ୍ରତ୍ତୁଳୁ	Otsisherero	Interlingua	Bahasa Indonesia	Interlingue	Igbo	ଶିଖ	Iñupiaq	Ilokano	Ido	Íslenska	
Italiano	Δακτυλικό/Inuktitut	日本語	Lojban	Basa Jawa	ქართველო	Qaraqalpaqsha	Taqbaylit	Адыгэбзэ	Kongo	ଶିଖୁୟ	Kwanyama	Қазақша	Kalaallisut	ଶାସନୀୟ	ଶିଖୁୟ	한국어	Перем Коми			
Kanuri	Къарачай-малкъар	କୁର୍ରାଷ୍ଟାର୍	କୁର୍ରାଷ୍ଟାର୍	Ripoarisch	Kurdî	Коми	Kernowek	କୁର୍ରାଷ୍ଟାର୍	Latina	Ladino	Lëtzebuergesch	ଲାକ୍କୁ	ଲେଝି	Luganda	Limburgs	Ligure	Lumbaart	Lingála		
ଲାଓ	Lietuviai	Latgalu	Latviešu	Basa Banyumasan	ମୋକଶେନ୍	Malagasy	Ebon	Олык марий	Māori	ବସୋ ମିନଙ୍କବାୟ	ମାକେଡୋନ୍ସ୍କି	ମର୍ବିଯୁଣ୍ୟୋ	ମହାରାଜୀ	ମହାରାଜୀ	ମହାରାଜୀ	ମହାରାଜୀ	ମହାରାଜୀ	ମହାରାଜୀ	ମହାରାଜୀ	ମହାରାଜୀ
Melayu	Malti	Myskoke	Mirandés	ଓର୍ଫିଲ୍ଲାମ୍ବାର୍	ଶ୍ରୀମଦ୍ଭଗବତ	ଶ୍ରୀମଦ୍ଭଗବତ	ଶ୍ରୀମଦ୍ଭଗବତ	ଶ୍ରୀମଦ୍ଭଗବତ	Dorerin Naoero	Náhuatl	Napulitano	Plattdüütsch	Nedersaksies	ନେପାଲୀ	ନେପାଲ ଭାଷା	Oshiwambo	Nederlands	Norsk nynorsk	Norsk bokmål	
Novial	Nouormand	Sesotho sa Leboa	Diné bizaad	Chi-Chewa	Occitan	Oromoo	ଓଡ଼ିଆ	ଶିଖ	ଶିଖ	ଶିଖ	Pangasinan	Kapampangan	Papiamentu	Picard	Deitsch	ଶିଖ	ଶିଖ	ଶିଖ	ଶିଖ	
Piemontèis	پیمونتے	Понтиака	پونتیاکا	Português	Runa Simi	Rumantsch	Romani	Kirundi	Română	ଅର୍ମାନେଶ୍ଚେ	Tarandíne	ରୁସ୍କୀନ୍	ରୁସିନ୍ୟକ୍ୟ	Kinyarwanda	ଶିଖ	ଶିଖ	ଶିଖ	ଶିଖ	ଶିଖ	
Scots	سنڌي	Sámeigiella	Sängö	Srpskohrvatski / српскохрватски				Simple English	Slovenčina	Slovenščina	Gagana Samoa	ChiShona	Soomaaliga	Shqip	Српски / srpski	Sranantongo	SiSwati			
Sesotho	Seeltersk	Basa Sunda	Svenska	Kiswahili	Ślůnski	தமிழ்	ତୁଳାରୁ	Tetun	ତୋଚିକ୍କ	ତୈଥ	ନୀର୍ଦ୍ଦିତ	Türkmençe	Tagalog	Setswana	Lea faka-Tonga	Tok Pisin	Türkçe	Xitsonga	Татарча/	
tatarça	ChiTumbuka	Twi	Reo tahiti	Үдмурт	ତୁମିନ୍ଦା	Uyghurche	Українська		O'zbekcha	Tshivenda	Véneto	Vepsän kel	Tiếng Việt	West-Vlams	Volapük	Walon	Winaray	Wolof	吴语	
Хальмг	IsiXhosa	ମାର୍ଗାଲ୍ଗୁର୍ହି	ଶାଦି	Yorùbá	Vahcuengh	Zeêuws	中文	文言	ବାନ୍ଦାମ୍ବାଗ୍	ଶିଖ	ଶିଖ	ଶିଖ	ଶିଖ	ଶିଖ	ଶିଖ	ଶିଖ	ଶିଖ	ଶିଖ		

----- Collected from Wikipedia ULS -----

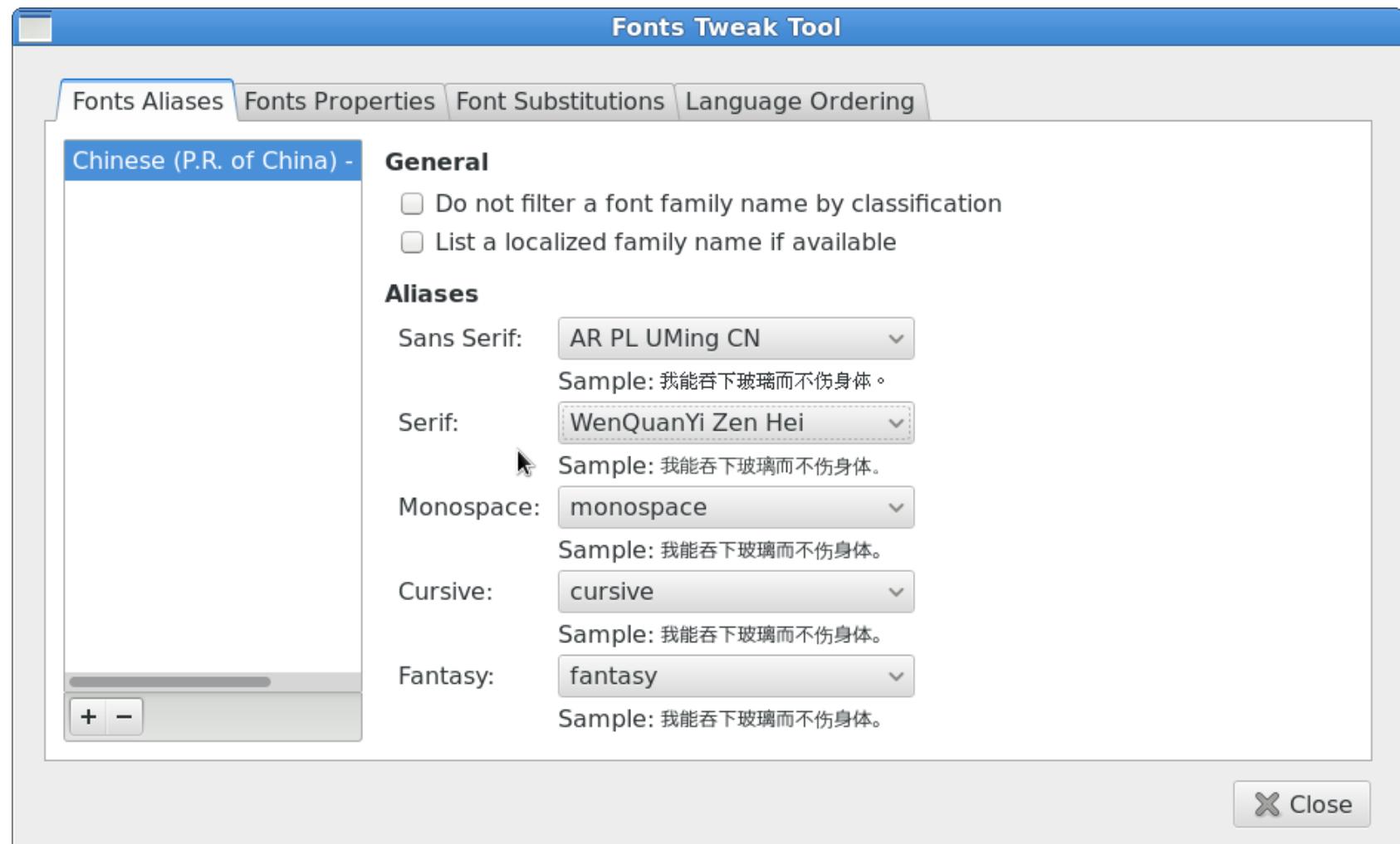


Asian Fonts

- Japanese: vggothic (also IPA)
- Chinese: WQY Zenhei, cjkuni
- Korean: Baekmuk → Nanum
- Indic: Lohit
 - Rewriting of OpenType tables and rules
- *New fonts* appearing from Google, etc
 - Eg Noto, Caladia, Carlito, etc



Fonts-tweak-tool



Input methods



Input methods evolution

XIM → iiimf → scim → ibus → ...

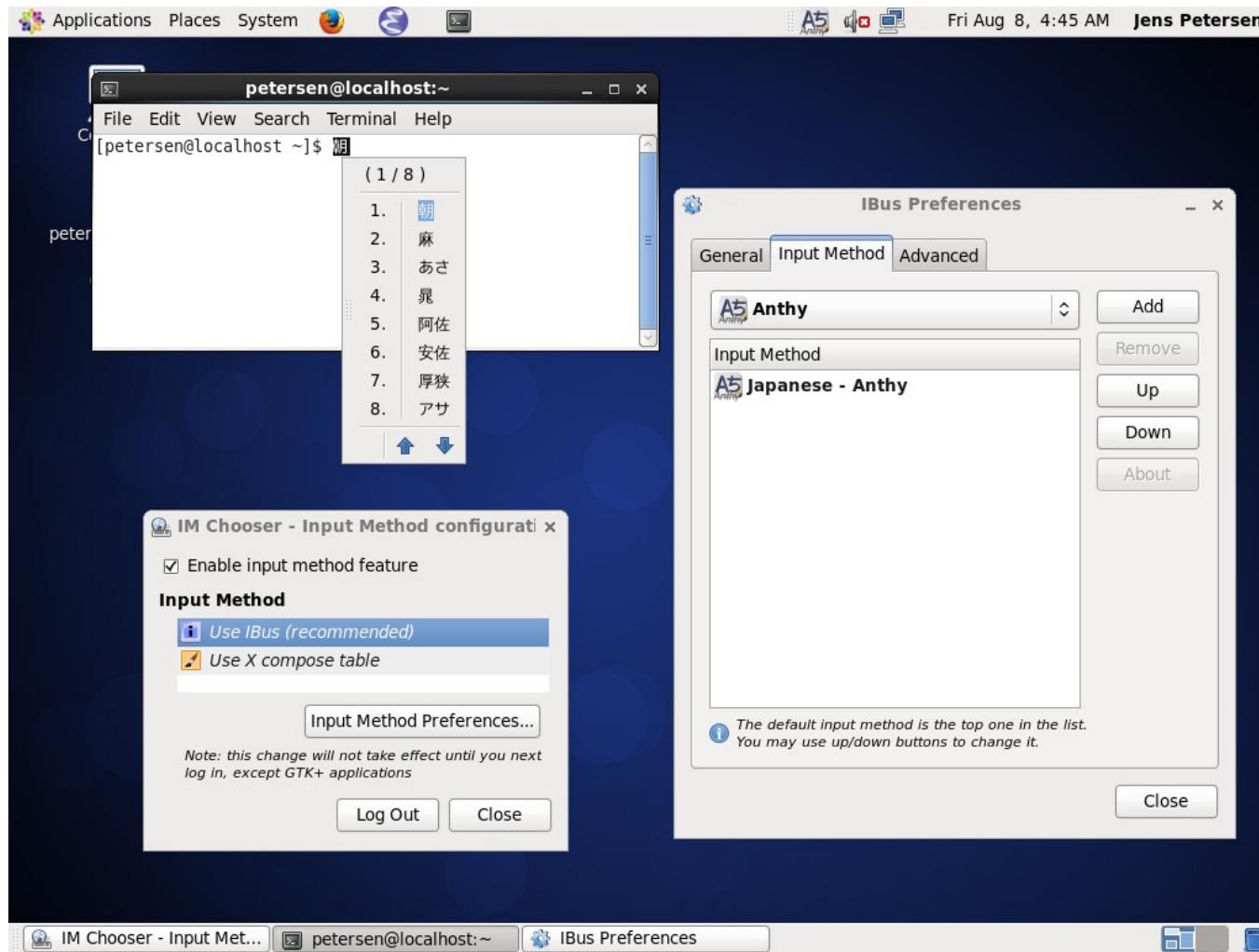


e.g. Japanese input method backend evolution

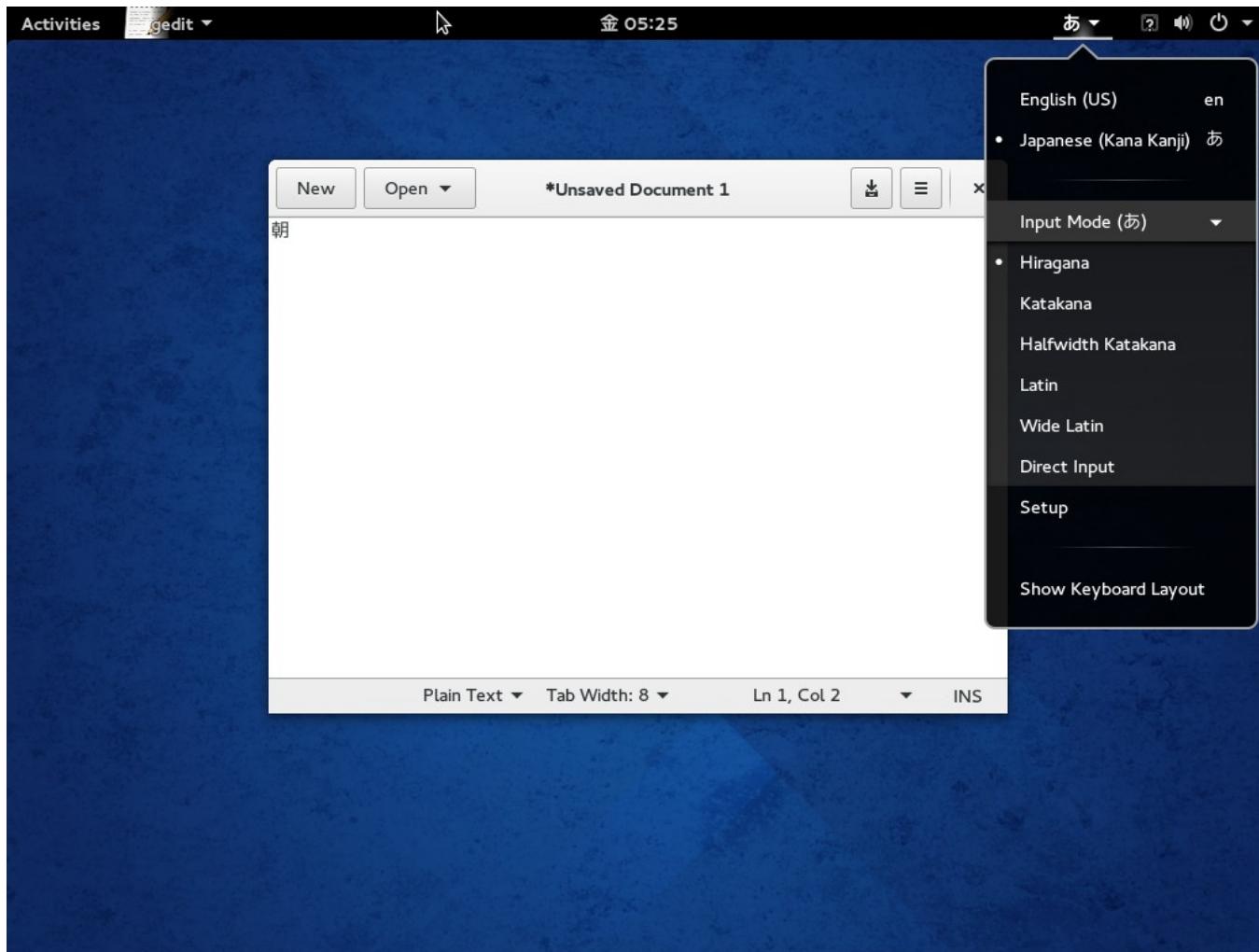
Canna → anthy → libkkc



ibus under GNOME 2



GNOME 3 ibus integration



Input Methods

- Japanese: *ibus-kkc/libkkc*
- Chinese: *ibus-libpinyin/libpinyin*
 - *ibus-chewing*, *ibus-table*, *ibus-cangjie*
 - *ibus-libzhuyin* in development
- Korean: *ibus-hangul*
- M17n: for Indic and many other languages
 - Including input without native keyboard
- *ibus-typing-booster* provides suggestions and basic predictions

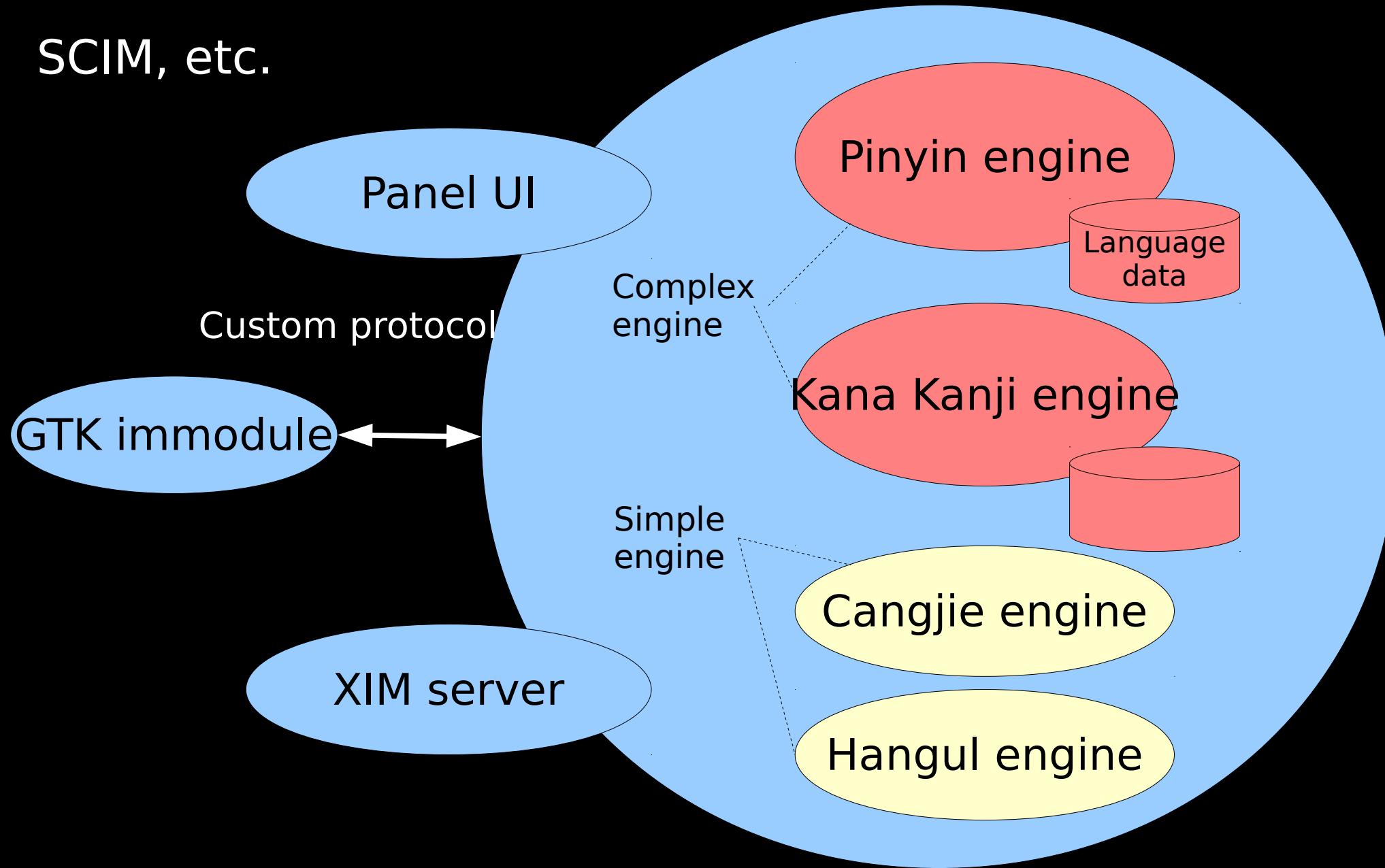


Next Generation Input method (for Wayland etc)

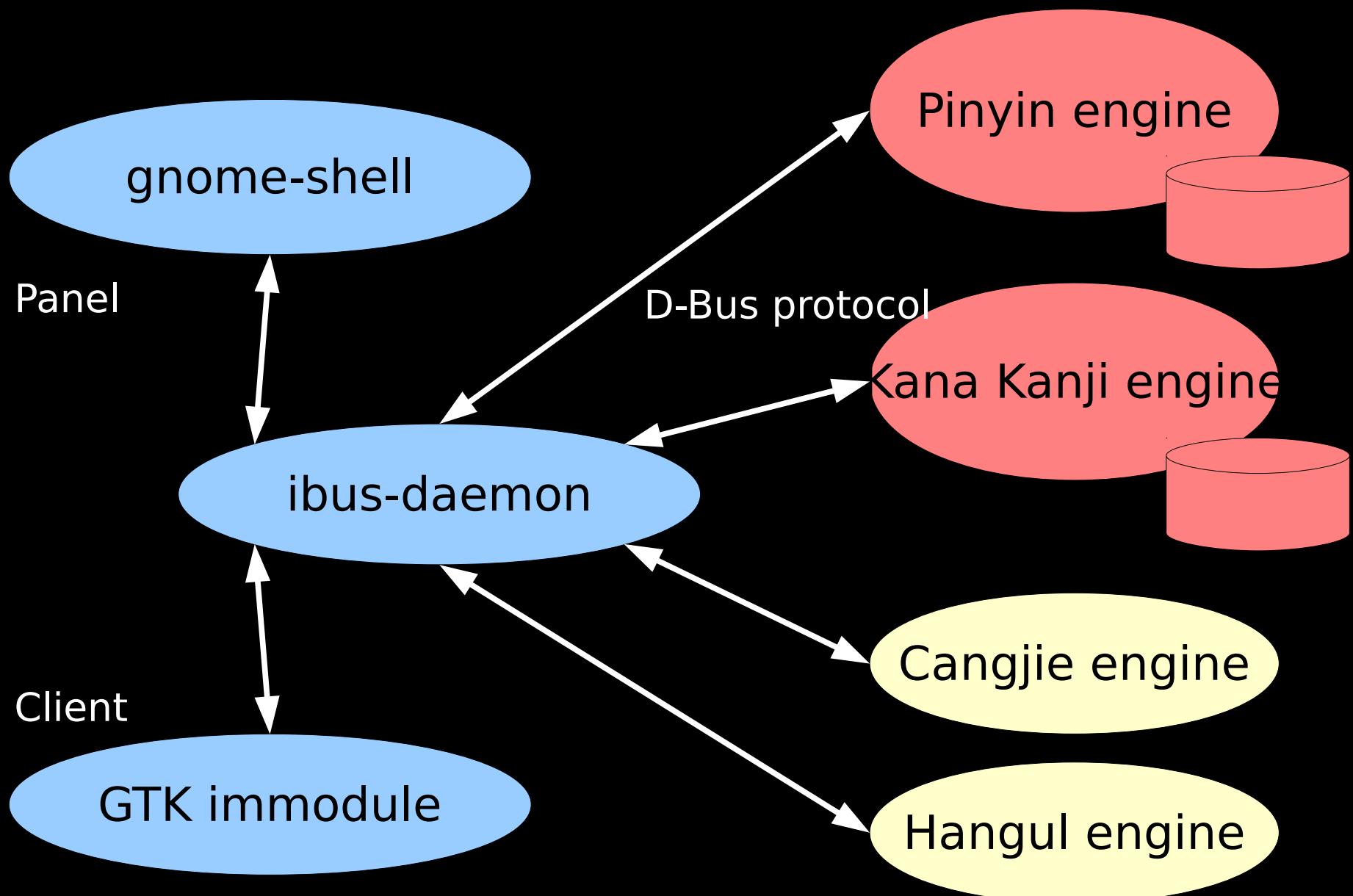


Traditional IM architecture

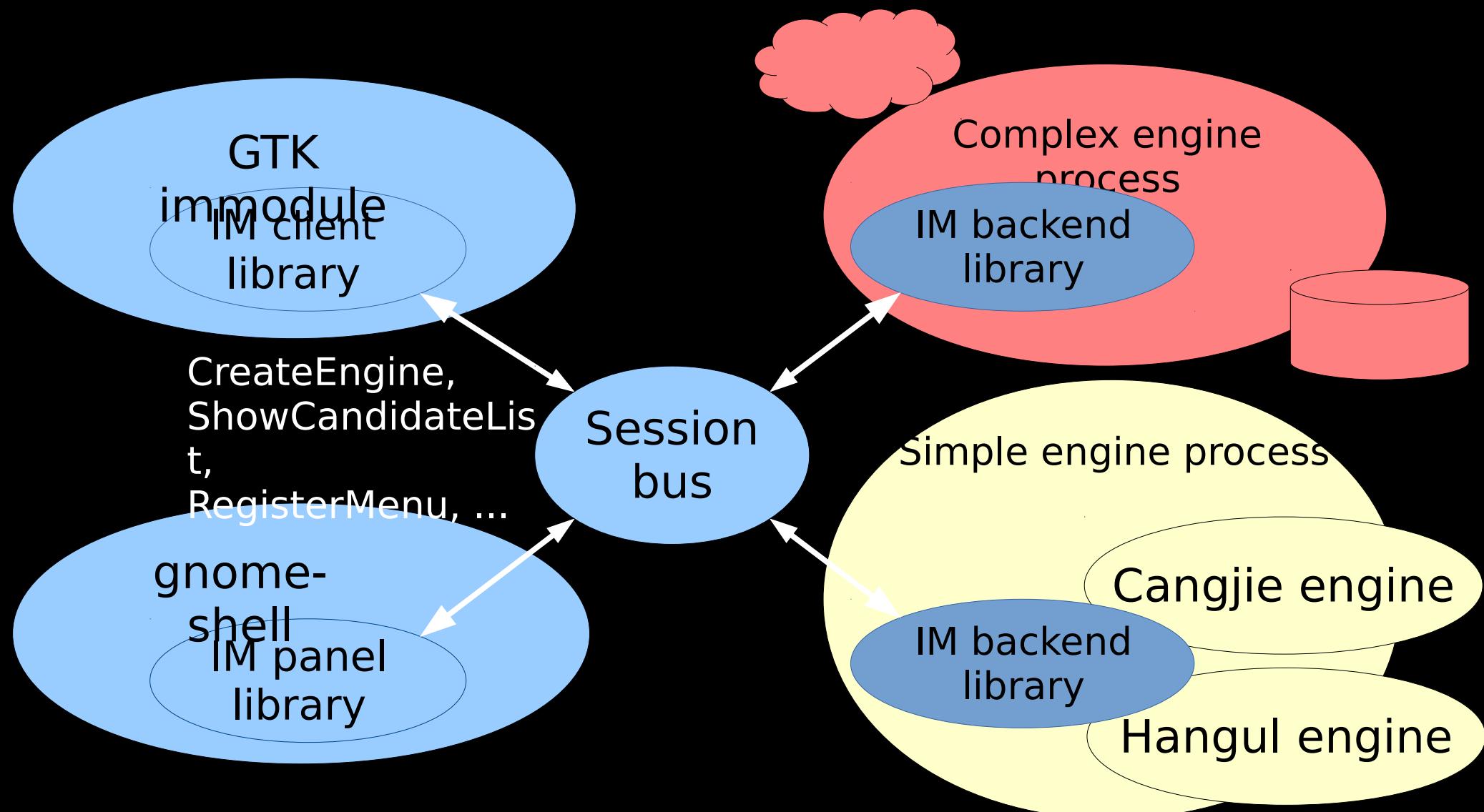
SCIM, etc.



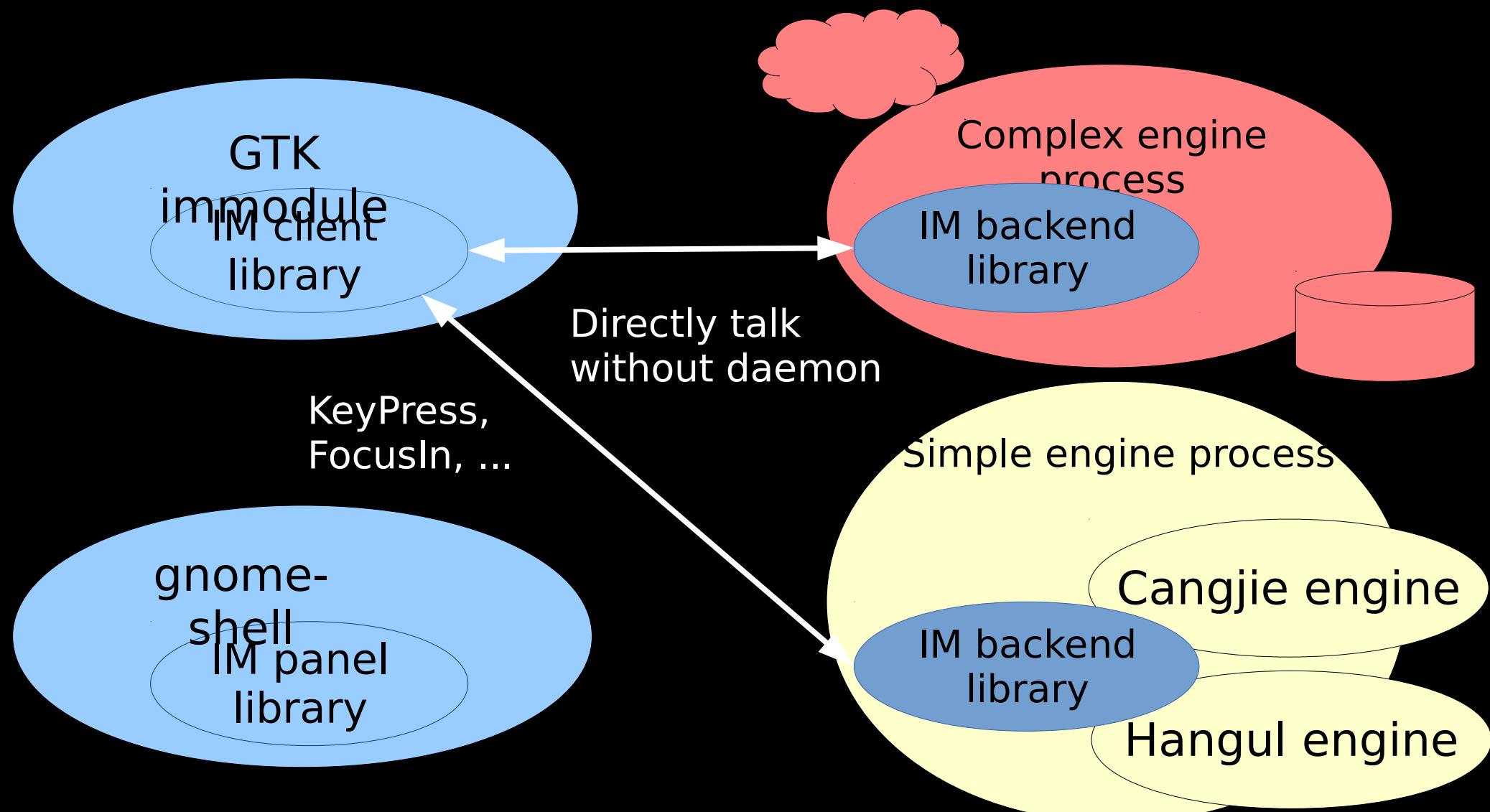
IBus architecture



Proposed architecture



Proposed architecture



Packaging



Comps

- @fonts
- @input-methods
- Lang-packs
 - Language groups → yum-langpacks
 - system-config-language



Looking Ahead

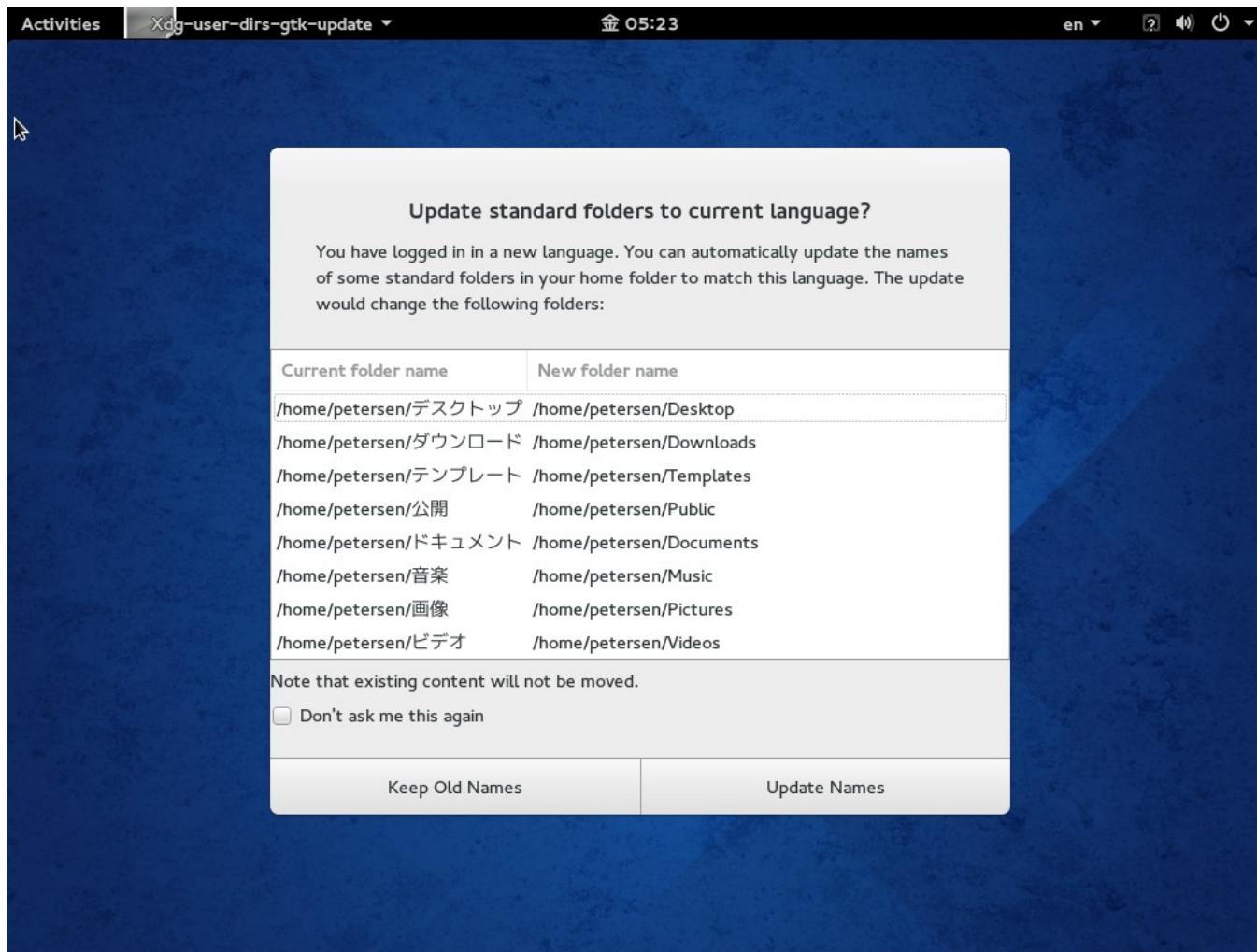


Wanted or nice to have

- **Subpackaging of glibc locales?**
 - Needed for Cloud, small Server, etc
- Bilingual searching
- Better spellchecking
 - Eg Hunspell: aa → å, but not oe → ø and ae → æ
- Restricting keyboard layout choices by geometry?



XDG folders i18n



Plans

- Wayland input method integration
- Enhancements to smart input methods
- Revamping localized man-pages and hunspell
- Glibc unicode and locales refreshing
- dnf-langpacks



Fedora i18n sub-project



FAS i18n-team pseudouser

- 255 packages in pkgdb
- i18n-bugs mailing-list for bugzilla mail



Summary of what we do

- Package maintenance and packaging
 - Fonts
 - Input methods
 - Other: yum-langpacks, langtable, comps, etc
- Upstream development
 - ibus
 - fontconfig
 - gettext
 - Contribute to gnome, glibc, etc



Fedora i18n Communication

- Fedora i18n mailing-list
- Freenode #fedora-i18n IRC channel
- Weekly meeting in #fedora-meeting (APAC time)
- <http://fedoraproject.org/wiki/I18N>
- Welcome more Fedora community collaboration and interaction



Thank you



Fedora i18n Project

<http://fedoraproject.org/wiki/I18N>

Contact me:

petersen@redhat.com

juhp on freenode, github, twitter

