

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: 4 September 2025

L. Dusseault, Ed.
tbd
H.J. Happel, Ed.
audriga
A. Melnikov, Ed.
Isode Ltd
3 March 2025

Personal Data Portability Archive
draft-happel-mailmaint-pdparchive-00

Abstract

This document proposes the Personal Data Portability Archive format (PDPA), suitable for import/export, backup/restore, and data transfer scenarios for personal data.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 4 September 2025.

Copyright Notice

Copyright (c) 2025 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction	3
2. Conventions Used in This Document	3
3. Goals	4
3.1. Use cases	4
3.1.1. Data portability	4
3.1.2. Incremental backup	4
3.1.3. Synchronization	5
3.1.4. Dataset exchange	5
3.1.5. Data persistence	6
3.2. Technical Goals	6
3.2.1. Email standards compatibility	6
3.2.2. Interoperability	6
3.2.3. Extensibility	6
3.2.4. Flexible granularity	7
3.2.5. Accessible for local tooling	7
3.2.6. Efficiency	7
3.3. Related work	7
4. Approach	8
4.1. Using JSON (mostly)	8
4.2. Approach to partial updates	9
4.3. Approach to synchronization	9
5. Solution Requirements	10
6. File format	11
6.1. Index file(s)	11
6.2. Folder structure	11
6.3. Data formats	14
6.3.1. Email	14
6.3.2. Contacts	16
6.3.3. Calendars	16
6.3.4. Tasks	17
6.3.5. Notes	17
6.3.6. Files	17
6.3.7. Other	17
6.4. Synchronization requirements	18
6.4.1. Always include 'uid' and 'updated'	18
6.4.2. Synchronizing address books	21
6.4.3. Synchronizing mailbox folders	21
6.4.4. Blobs and files?	21
7. Implementation status	21
8. Security considerations	22
9. Privacy considerations	22
10. IANA Considerations	22
10.1. File extension	22
11. Informative References	22
Authors' Addresses	24

1. Introduction

As part of communication protocols, the IETF has standardized a number of data formats such as the Internet Message Format [RFC5322], vCard [RFC6350], iCalendar [RFC5545], or, more recently, JSContact [RFC9553] and JSCalendar [RFC8984].

While mainly designed for interoperability, many of these data formats have also become popular for data portability, i.e., the import/export of data across different services.

The growing importance of data portability however demands for an open standard archive format which can deal with different types of personal data in a homogeneous fashion.

To this end, this document proposes the Personal Data Portability Archive format (PDPA), suitable for import/export, backup/restore, and data transfer scenarios for personal data. It is compatible with both IMAP and JMAP and should be suitable as an interchange format between related software and services such as for email, contacts, calendaring, tasks, or files.

The approach is to define JSON formats (using CDDL), folder structure, and a common compression format. Additional specifications will likely define how these files can be requested from, imported into, or transferred between servers.

2. Conventions Used in This Document

The term "personal data" refers to persistent data which users created and managed within applications or services. Classic examples are emails, contacts, calendars, tasks, notes, or files. Other examples might be fitness tracking records, energy bills, or location history.

The term "data portability" refers to the right or technical procedure to use or transfer personal data across different applications or services.

The terms "message" and "email message" refer to "electronic mail messages" or "emails" as specified in [RFC5322]. The term "Message User Agent" (MUA) denotes an email client application as per [RFC5598].

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Goals

The core goal is to provide an open and extensible archive and transfer format for personal data. This includes data types that are subject of existing IETF protocols, such as email and groupware data (e.g., contacts, calendars, tasks) but may also include further types of personal data.

Goals will be refined by use cases and cross-cutting technical goals in the following sub-sections.

JSON is used as a widely interoperable base format that systems can easily translate their internal data representation into. Wherever possible, fields, values and data structures are derived from existing IMAP/JMAP specifications to remain compatible with both.

3.1. Use cases

3.1.1. Data portability

A main use case for the novel format is to allow exporting the full user data managed by services or software products into a simple file (or set of files) which is under full control of the user.

The user might use such export for backup, archiving, or for importing when switching to another service or software (i.e., migration).

Depending on the type of data, exporting/importing can be a time-consuming process. Particularly for the case of switching services, PDPA should allow to minimize the time period during which a user cannot use the origin system but also the destination system is not yet ready.

3.1.2. Incremental backup

Beyond snapshot backups/exports, the format should optionally allow for incremental backups.

There are at least two scenarios for this:

- * Software which wants to keep a permanent, incremental mirror of user data (e.g., for instant export or restore)
- * Users regularly exporting changes in data managed by services or software products

3.1.3. Synchronization

Data portability does not just allow users to switch from one service to another one, but to let users benefit from 3rd party services getting a copy of their data (at the request of the user). Simple synchronization features could make this much better.

For example, current online systems that allow importing contacts are not often suited to maintaining one's address book on two systems. Re-importing a contact into a system that already has that contact often results in duplicating the exact same contact, whether or not there have been edits, making repeated synchronization practically infeasible. It should be easy to do a significantly better job of this with some attention to object IDs and modification timestamps.

We however do not attempt to solve two-way synchronization via export files. It would require significant additional work to allow two systems, neither of which is agreed-upon to be the source of truth, to reliably synchronize changes from both. In comparison, solving one-way synchronization only requires agreed-upon usage of existing fields and values.

3.1.4. Dataset exchange

PDPA should be usable to exchange and share larger data sets than just one user, or to share a single user's data outside the context where the user knows what it is and where it came from.

Potential applications of this are:

- * The ability to exchange test data and known mailstore state, e.g. for conformance testing or internal functional tests
- * Legal discovery and forensics use cases may benefit from a standard export format, such that investigators can expect a great deal of consistency when collecting data from different systems.
- * Researchers may be able to collect archive files through data donations and use as input to research.

While these use cases may suggest small features that would help these use cases be successful, supporting more advanced features required by these use cases is not a priority. For example, we do not attempt to cryptographically solve provenance for use in legal and forensics use cases.

3.1.5. Data persistence

The format MAY be used as a development-time active persistence layer for user data in, e.g., email clients or applications. It is not intended as, or suitable for, a production-level persistence layer.

3.2. Technical Goals

Besides actual use cases, there are a number of side requirements and goals for PDPA.

3.2.1. Email standards compatibility

Data formats should aim for compatibility with JMAP data formats for the sake of interoperability and synergies in software libraries.

Dedicated JMAP API methods for exporting and importing the format described here, or for related server-to-server transport protocols are out of the scope of this document.

Due to its specifics and ubiquitous usage, the Internet Message Format [RFC5322]; latest revision of [RFC2822]/[RFC822] should be the core of representing individual email data.

This specification should ideally describe mappings between PDPA and existing mailbox persistence schemes such as Maildir or MBOX [RFC4155].

3.2.2. Interoperability

It should be mostly possible to use personal data exports from one system with different software or services. When a source exports personal data it can include all the information it would need for a fully-functional import, _however_ destination systems running different software may not be able to import all of that information (especially if it includes non-standard features) and use it exactly the same way. This specification does not attempt to achieve perfect interoperability between diverse systems, but instead to make reasonable tradeoffs.

3.2.3. Extensibility

This format should be extensible to accommodate types of personal data not explicitly mentioned or foreseen when writing these specs.

3.2.4. Flexible granularity

This format should allow flexible granularity in two ways:

- 1) It should enable easy access to separate types of data (e.g., emails vs. contacts), e.g. to allow for partial imports or exports
- 2) While ideally representable as a single file, archives may also span several files due to reasons such as file size restrictions or incremental generation logic.

(The ability for a user to export and/or backup an entire email account requires some accomodation of large amounts of data and risks of interruptions in downloads. Splitting exports into multiple files during export is one possible solution.)

3.2.5. Accessible for local tooling

PDPA should allow easy access for local tools (e.g., CLIs). While this may sound obvious, it is a key factor for the intended versatility of the format.

3.2.6. Efficiency

Since certain kinds of personal data might involve large quantities of data, major use cases for PDPA should be realizable in an efficient manner.

For now, this is stated as an abstract guiding principle. Its actual dimensions and trade-offs need to be refined while evolving this specification.

3.3. Related work

Many email server implementors have found it desirable to have one or more file formats for storing email in a file system even when the primary active email storage is more commonly a database. Examples include [PST] files (Outlook), NSF (Notes), [GoogleTakeout], Maildir, MBOX. File formats are already used for interoperability in many cases even when not standardized.

This specification follows that pattern in order to build on these partial successes. By standardizing one format, we expect to be able to satisfy use cases that are harder to satisfy with a plurality of formats, such as use cases for server-to-server transfer of email account data during account migrations. Specifications that explain how to create these archives in different situations can refer to this specification.

4. Approach

4.1. Using JSON (mostly)

JSON is used in this spec for new metadata and for objects including contacts, tasks, events and notes. However, the Email Message Format [RFC5322] is used for email message content. Individual items are stored in individual files, which are referenced in collection metadata. Finally these JSON and other file formats are packaged and compressed together in a standard but flexible way.

Our reasoning for using JSON as much as is reasonable:

- * We envision an export format being used not just by developers of full IMAP servers but also by developers building task management systems, calendar systems that don't include email, etc.
- * We should minimize requiring multiple libraries to parse different formats. If the Metadata is going to be in JSON, it would really help to have the item data in JSON.
- * Personal data formats must be extensible and extensibility for JSON is well-understood.

Using JSON for all the data *_except_* EML files was carefully considered. EML files are rather specialized and more challenging to replace. Much of email is not structured data but content and involves MIME. EML is more likely to be a system's native data store, unlike VCARD and VTOD0 which are most commonly transformed for use in a relational data store for active use. Finally, email signature implementations like S/MIME and OpenPGP would be less disrupted by keeping EML.

Information not covered by these existing file formats, but still necessary for migration or backup/restore of an email account, is packaged into JSON files. JSON structure and values are defined with CDDL [RFC8610]. The JSON files for email folders and other collections contain references to individual resources by unique ID and filename.

4.2. Approach to partial updates

Our use case requirements above (#use_cases) included some very common personal use cases that motivate exporting only a time-limited set of items, but retaining the ability to use that subset export to update a previously acquired or maintained snapshot. These are partial updates - partial exports able to update a full copy. Our approach is to define a version of the archive format that includes a subset of a repository's content, and additionally some change markers necessary to maintain consistency.

A partial-update export

- * Contains new items just like a full export.
- * Omits unchanged items from before the time cutoff.
- * Does NOT list unchanged items in folder listings either.
- * Allows updated items to be identified so they can replace previous versions
- * Allows deleted items to be identified so they could be removed in the updated snapshot

The existing definitions for UIDs and 'updated' in JMAP and IMAP should make this quite possible. Also IMAP MODSEQs [RFC7162] can be used for flag changes.

4.3. Approach to synchronization

Email and calendaring services appear to already do synchronization just fine, but this is via a client-server model. The client drives the read and write requests until content is synchronized, using mostly UIDs and timestamps.

Archive and export formats aren't part of this client-server model, and the work to make server-to-server or peer-to-peer synchronization work perfectly is substantial (involving features such as version numbers or change logs -- features that aren't commonly standardized for the objects handled in this spec). Still, there are some things possible with archive formats and the current object definitions that are sensible. Our approach is to describe what is possible with the fields that exist, and mandate those fields be used, so that users aren't left with multiple locations for their data and no way to repeatedly synchronize them.

As an illustrative use case, let the user have contacts stored in one email service and also in one mobile device platform doing online backups. The email platform creates contacts when the user emails new recipients, or receives contact information over email. The mobile device platform synchronizes contact information from a phone.

The email service is not a client of the mobile device platform, nor is the mobile device platform a client of the email service, so client-to-server protocols cannot directly synchronize the data between these two services. A user attempting to solve this by repeatedly importing contacts into one system from the other may find this works poorly - for example, it might create new contact objects over and over for the same contact data even if it is unchanged, and deleted objects may re-animate after being re-copied.

Both servers ought to allow exporting of contact data (along with any other data covered in this specification), including especially the UID and updated (timestamp) fields. This would allow at least some sensible personal workflows or for 3rd party tools to make synchronization work better.

When importing contact data (along with any other data covered in this specification), a service needs to take note of the UID in the import, and use it as the UID for creating a new object, so that it may be later avoid being recreated. If the service importing already has the said UID, the service should compare 'updated' timestamps and use that to decide how to update the object with new values. An object that hasn't changed since last imported should remain unchanged and not updated.

This approach to synchronization is imperfect. Let the user have performed one synchronization by exporting from the email service and importing to the mobile device platform. If the user updates a contact on the email service (e.g adding a street address) and the mobile platform also updates the contact (e.g. adding an avatar link), then the user attempts to repeat the synchronization, both services will have a new 'updated' timestamp on the same object UID, and the earlier of the two changes will get wiped out. Nevertheless, a disciplined user can remember to only make changes on the email service and always copy them over to the mobile platform, and avoid many such lost updates.

Based on the approach described here, this specification standardizes some behavior for both exporters and importers, to maximize potential success.

5. Solution Requirements

These technical requirements on the solution are intended to meet the goals above, and to add more specifics about how those goals are intended to be met with the architecture chosen. This section does not attempt to translate the solution requirements into implementor requirements.

Folder format requirements

- * can include partial results, showing only a subset of objects within the folder
- * can include a human-readable representation of the meaning of that subset (e.g. a date filter, or recipient filter)

Email object format requirements

- * can maintain full fidelity including preserving character sets, content transfer encoding of body parts and exact MIME structure

Compression and packaging of resources

- * Servers can bundle resources together in different ways to be flexible in handling size and network limitations. Servers must be able to choose optimal file size and organization of information within files.

Synchronization requirements

- * Can see which items are identical in two systems, e.g. one system previously imported an item exported by the other.
- * Can detect changes made since the last time an item was imported, in a way that supports replacing an older version previously synchronized with a newer version that has edits.

6. File format

6.1. Index file(s)

(index.json)

- * Data source (user, shared)
 - Which account / which server/service
- * Nature (dump, ...)
- * Dates (created, last update)
- * Sync id?
- * Generator (tool)

6.2. Folder structure

- * Folders can be nested. Any kind of content here can be within nested folders -- this is a necessary feature for email, but extends to content like contacts that aren't always represented in nested folders. (TODO: elsewhere, describe the requirements for an importing system to preserve folders or not.)

- * Names of files do not have to be globally unique. Indexes and folder contents listings can name files relatively to their location in the archive structure, which means that references may not be resolvable if that context is lost.
- * Individual content items are individual files. This may not always be the easiest choice for exporters who must generate a large number of files for individually small items (contrast to a JSON stream including all objects) but as an archive format, the individual files allow more clarity in individual handling, transactions and errors.

```
index.json
/mail/
  /Archive/
    folder.json
    m1.eml
    m2.eml
    m3.eml
    ...
  /Archive/2023/
    folder.json
    m1.eml
    m2.eml
    m3.eml
    ...
  /Archive/2024/
    folder.json
    m1.eml
    m2.eml
    m3.eml
    ..
  /INBOX/
    folder.json
    m1.eml
    m2.eml
    m3.eml
    ...
  /Sent Mail/
    folder.json
    m1.eml
    m2.eml
    ...
/contacts/
  contact1.json
  contact2.json
  ...
/calendars/
  /calendar2/
    event1.json
    event2.json
/sieve/
/blob/
  ...?

| TODO: I18n?  Special-use?
```

Folder names are defined in RFC9051 (IMAP v4 rev2) with great freedom for servers. Servers may or may not treat mailbox names as case sensitive. Folder names may even include non-graphic characters, "%" and "*". Hierarchy separators may even differ among IMAP servers although "/" is probably most common.

Since this specification is new, it is possible to be more constrained. This specification only supports "/" as a folder separator.

6.3. Data formats

6.3.1. Email

Each IMAP/JMAP folder is represented as subdirectory under "mail" directory. For example, the folder INBOX would be represented as "mail/INBOX", and the folder "Archive/2024/2024-12" would be represented as "mail/Archive/2024/2024-12".

(In examples above "/" is the IMAP hierarchy delimiter)

Folder names are encoded in UTF-8.

TODO: encoding of characters in folder names not allowed by filesystem.

Each folder metadata is described by "folder.json", which has the following format:

```
;; /// Or possibly use ranges for 2 types below?
u32 = uint .size 4
u64 = uint .size 8
```

```
; uid => message filename map
uid_map = { * uid => filename }
uid = u32
filename = tstr
```

```
flags_map = { * uid => flags }
flags = [tstr]
```

```
modseq_map = { * uid => modseqs }
modseqs = u64
```

```
folder_info = {
  ? allowed_keywords: [tstr],
  last_uid: u32,
```

```
? highest_modseq: u64,  
? recent_uid: u32,  
uidvalidity: u32,  
  
is_subscribed: bool,  
  
; See RFC 6154 (SPECIAL-USE). E.g. "inbox", "sent", "drafts", "junk", etc.  
? role: tstr,  
  
; See Section 2, RFC 8621.  
? sort_order: u32,  
  
  
; See Section 3.5 of RFC 4314. For example "rwiptsldaex"  
? myrights: tstr,  
  
; Maps UIDs to the corresponding relative file names (names of .eml files)  
uids: uid_map,  
  
; Maps UIDs to the corresponding array of flags/keywords (as strings)  
flags: flags_map,  
  
; Maps UIDs to the corresponding modseqs (u64). See RFC 7162.  
? modseqs: modseq_map,  
  
; Can include information about partial export or filter used  
; in human readable UTF-8 text  
? comment: tstr,  
}
```

EML (.eml) file for each message, in order to avoid reconstructing them. Names of EML files are referenced from the "folder.json" file.

Example of folder.json:

```
{
  "allowed_keywords": ["$Forwarded", "$MDNSent", "$ismailinglist"],
  "last_uid": 3360940,
  "highest_modseq": 6371729,
  "recent_uid": 3360940,
  "uidvalidity": 1107190787,
  "is_subscribed": true,
  "role": "sent",
  "sort_order": 1,
  "myrights": "rwiptsldaex",

  "uids": {
    1: "msg-1.eml",
    3: "msg-3.eml",
    15: "imported-ABC.eml",
  },

  "flags": {
    1: ["$seen"],
    3: ["$seen", "$flagged"],
    15: ["$answered", "$forwarded"],
  }
}
```

6.3.2. Contacts

JSContact [RFC9553] * uid property uniquely identifies - * updated property gives some hint at merging changes but requires additional specification to be usable for sync * we also need to reference RFC6350 which defines uid and 'rev'

vCard [RFC6350]

6.3.3. Calendars

JSCalendar [RFC9553]

iCalendar [RFC5545]

Although CalDAV servers are fairly common, they support the older VEVENT and VTODO syntax to represent calendar events and tasks in protocol messages. This specification requires the JSCalendar syntax instead. Either way, a server building an archive is likely transforming an internal implementation-specific relational data format to an export format.

CalDAV servers use the event's UID to identify the same object received more than once (e.g. once in an invitation and again in an import). CalDAV servers use ETags to identify changed events, so that a CalDAV client may make sure it has the version a server has before it updates an item, solving the lost-update problem.

Since this specification doesn't attempt to solve the lost-update problem as well as client-server protocols do, and since JSCalendar does not include the ETag of a calendar event in any way, this specification does not include any requirements for ETags.

(Alexey & Hans-Joerg - what's your opinion on this? JSCalendar doesn't include ETag and obviously it's just fine?)

6.3.4. Tasks

JSCalendar [RFC9553] defines the JSON format used for tasks.

```
{
  "@type": "Task",
  "uid": "2a358cee-6489-4f14-a57f-c104db4dc2f2",
  "updated": "2020-01-09T14:32:01Z",
  "title": "Do something"
}
```

6.3.5. Notes

- * VJournal is first defined in iCalendar (now [RFC5545]).
- * VJournal also used in CalDAV (<https://datatracker.ietf.org/doc/html/rfc4791>)
- * However, they are NOT used in <https://jmap.io/spec-calendars.html> (<https://jmap.io/spec-calendars.html>)

Do we even have a JSON format for notes defined?

6.3.6. Files

6.3.7. Other

(LMD Note: I think this might better fit in an out of scope section - I think out of scope sections are useful for statements that explain why scope is limited. that's assuming we all agree that groups, freebusy and timezones are left out.)

Groups as defined in JSCalendar are NOT part of this archive format. Groups in JSCalendar can combine events and tasks in a container. This specification, for consistency and simplicity, uses folders and requires individual objects to be in separate files.

VFREEBUSY objects are not part of this archive format. Calendar software can calculate freebusy time from event data. Use cases that are not satisfied by this limitation could extend this archive format but understanding what it means to backup, restore, export or import freebusy data would need to be fleshed out.

VTIMEZONE objects are not part of this archive format. Timezones are more likely to be system objects referred to by calendar objects in modern calendar systems, than objects which are exchanged across systems.

6.4. Synchronization requirements

This section describes the requirements to achieve repeated one-way synchronization via export/import operations between software by different vendors. While limited, this still provides better functionality than what many end-users experience with their groupware software and services today.

Supporting `_repeated_` synchronization means that the export from system A and import to system B can happen over and over again without needlessly duplicating items. Supporting `_one-way_` synchronization means that changes in the system with the exporter role propagate reliably to the system with the importer role, but not in the reverse direction. Some of the constraints here arise from the fact that the two systems may not directly connect, and the import may be time-delayed from the export.

This limited solution for export/import synch may also be used for more direct system-to-system transfers such as service-to-service data transfers, repeated data access requests or data migrations, although some of those use cases could be solved much better with direct negotiation of features.

6.4.1. Always include 'uid' and 'updated'

These requirements apply to JSContact [RFC9553], JSTask and JSCalendar [RFC8984] objects when exported or imported using the formats in this specification, because these all have 'uid' and 'updated' values.

Requirements:

- * exporters MUST include the UID and updated fields.
- * The 'updated' field MUST be exported in UTC and interpreted in UTC. Accurate system time is important.
- * importers MUST use the UID in an imported object, if the importer is creating a new object, rather than invent a new UID.

- * importers MUST search for existing objects with the same UID, and if the object in storage is `_similar enough_` (see Note below) to the import data, the importer SHOULD NOT change the object and MUST NOT update its 'updated' timestamp.

Recommendations:

- * Importers SHOULD use caution with fields that are system-updated, especially frequently updated. Such fields SHOULD NOT change the value of 'updated' that is exported or used to decide whether to update an object during an import operation. See note below on 'updated'.
- * Importers SHOULD apply common sense in updating internal or implementation-specific fields. This specification does not require the importer to include, omit, handle or disregard values for fields that it believes are internally-generated or implementation-specific. For example, a system in the role of exporter might export an event object with a video-conference room ID in a custom field. It can decide that it is sensible to export that value as a URL for external use. Later, the same system or one with code written for compatibility could import that event with the video-conference URL, and it would be sensible to avoid overwriting its own knowledge of the room ID with the URL.
- * When importing a `_changed_` or `_new_` object with a UID and 'updated' value, the importer SHOULD set the 'updated' value to the one imported. Thus, if a Contact is updated on Jan 1, exported on Jan 2 and imported on Jan 3, the new or updated imported contact would show an 'updated' value of Jan 1.

Note on `_similar enough_`: This specification requires nuance in order to allow both reasonably consistent synchronization and reasonable behavior in a wide variety of use cases and implementations. The language above is intended to give implementors both guidance and wiggle room. For example, the importer could convert a DTSTART time from UTC to the user's local time and save it as the displayed start time. Later, reimporting the same object with the same UID, the importing code could be smart enough to realize that the time hasn't `_actually_` changed, and avoid changing the 'updated' timestamp or creating a conflicting event. This logic could be implemented by saving separate fields (imported time vs display time), by keeping a log of updates (log entry stating that the system auto-converted start time from X to Y), or by other clever algorithms. Thus, the clever implementation can avoid the appearance of an object that changes every time the calendar is synchronized.

Note on `_updated_`: The definition of 'updated' in JSContact [RFC9553] is not rigorous or nuanced. "when the data in the Card was last modified" could refer to several instances of the card -- its

internal implementation, its representation in an email share, its representation in an HTTP GET response [RFC6352]. It's not specified whether 'updated' is the same as REV in VCard [RFC6350], which is defined differently. Neither definition explicitly covers vendor-specific fields. Thus, this specification makes additional recommendations for handling 'updated':

- * The value of 'updated' SHOULD only change when two conditions hold: the end-user makes a decision to change a value of a user-visible field, AND the export of the JSContact shows a different value.
- * Thus, non-user-visible fields like 'version' could be changed without causing the 'updated' value to change. A value such as 'language' could be set without changing 'updated' (if an implementation infers the language tag and begins to include 'fr-CA' as the language value in exports instead of no language, nevertheless this doesn't change the user-visible content).
- * If implementations need to manage the synchronization of vendor-specific fields, a vendor-specific field like 'example.com:updated' can be used rather than affect the user-visible synchronization made possible by 'updated'. Implementations could possibly also handle 'updated' differently when used for export/import using the formats in this specification, than when the same field is handled in other code paths.

We recognize that this understanding of 'updated' is highly judgement-dependent. The same field can change in one way and cause a change to 'updated' and in another way may not (the example of server inferring the language is 'fr-CA' vs the user explicitly setting it). It is likely to be frustrating to protocol designers and implementors (as it is to the authors of this specification) that the definition is so wobbly. We'd love to know of better solutions that work with the status quo.

6.4.1.1. Synchronizing from and to CalDAV servers

CalDAV [RFC4791] uses URLs, ETags and UIDs for synchronizing changes between two systems reliably, but it relies upon client-server architecture, where the server is the "source of truth" and the client must manage its local history and decide which things to update from the server and which things to tell the server to update. If a user is setting up synchronization or an implementor is building a system that involves synchronization, it may be best to use CalDAV if that is a feasible solution.

Nevertheless, we believe some of the use cases in our use case section (#use_cases) motivate not only including calendar data in these archives for backup purposes, but also for partial updates. This works the same way it does for JSContact and JSCard objects.

6.4.2. Synchronizing address books

Build on [RFC9610]

6.4.3. Synchronizing mailbox folders

Because servers may differ in which characters they support in folder names, how many levels deep folders may be created, and even in what separator character is used to indicate folder hierarchy, difficulties in synchronizing folder names will definitely arise. Folder names that are not likely to be widely supported in other systems should be translated for export, because if the exporting system has a consistent translation algorithm, then even if the mailbox name looks different in the importing system it will still be imported consistently.

Systems that support mailbox IDs MUST include them in exports. Systems that do not (though it's strongly encouraged) SHOULD use the full mailbox name as the unique identifier value.

TODO: Also it would be good to include a "display name" in case the server has had to translate the mailbox name for compatibility. E.g. a server that has a mailbox named "%L33T%", but knows the "%" should not be exported because many servers forbid the "%", would translate the name consistently to _pc_L33T_pc_ or another set of safe characters and include a display name of "%L33T%" for reference and debugging.

6.4.4. Blobs and files?

Reference [RFC9404]?

7. Implementation status

< RFC Editor: before publication please remove this section and the reference to [RFC7942] >

This section records the status of known implementations of the protocol defined by this specification at the time of posting of this Internet-Draft, and is based on a proposal described in [RFC7942]. The description of implementations in this section is intended to assist the IETF in its decision processes in progressing drafts to RFCs. Please note that the listing of any individual implementation

here does not imply endorsement by the IETF. Furthermore, no effort has been spent to verify the information presented here that was supplied by IETF contributors. This is not intended as, and must not be construed to be, a catalog of available implementations or their features. Readers are advised to note that other implementations may exist.

According to [RFC7942], "this will allow reviewers and working groups to assign due consideration to documents that have the benefit of running code, which may serve as evidence of valuable experimentation and feedback that have made the implemented protocols more mature. It is up to the individual working groups to use this information as they see fit".

8. Security considerations

Archive files can be maliciously edited before being uploaded to servers. There is no guarantee that the content of the archive file accurately represents email that an email box received or sent, and so on.

9. Privacy considerations

tbd.

10. IANA Considerations

10.1. File extension

register .pdpa?

11. Informative References

[GoogleTakeout]

Google, "Google Takeout",
<<https://takeout.google.com/settings/takeout>>.

[PST]

Microsoft, "[MS-PST]: Outlook Personal Folders (.pst) File Format", <https://learn.microsoft.com/en-us/openspecs/office_file_formats/ms-pst/141923d5-15ab-4ef1-a524-6dce75aae546>.

[RFC2119]

Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.

- [RFC2822] Resnick, P., Ed., "Internet Message Format", RFC 2822, DOI 10.17487/RFC2822, April 2001, <<https://www.rfc-editor.org/info/rfc2822>>.
- [RFC4155] Hall, E., "The application/mbox Media Type", RFC 4155, DOI 10.17487/RFC4155, September 2005, <<https://www.rfc-editor.org/info/rfc4155>>.
- [RFC4791] Daboo, C., Desruisseaux, B., and L. Dusseault, "Calendaring Extensions to WebDAV (CalDAV)", RFC 4791, DOI 10.17487/RFC4791, March 2007, <<https://www.rfc-editor.org/info/rfc4791>>.
- [RFC5322] Resnick, P., Ed., "Internet Message Format", RFC 5322, DOI 10.17487/RFC5322, October 2008, <<https://www.rfc-editor.org/info/rfc5322>>.
- [RFC5545] Desruisseaux, B., Ed., "Internet Calendaring and Scheduling Core Object Specification (iCalendar)", RFC 5545, DOI 10.17487/RFC5545, September 2009, <<https://www.rfc-editor.org/info/rfc5545>>.
- [RFC5598] Crocker, D., "Internet Mail Architecture", RFC 5598, DOI 10.17487/RFC5598, July 2009, <<https://www.rfc-editor.org/info/rfc5598>>.
- [RFC6350] Perreault, S., "vCard Format Specification", RFC 6350, DOI 10.17487/RFC6350, August 2011, <<https://www.rfc-editor.org/info/rfc6350>>.
- [RFC6352] Daboo, C., "CardDAV: vCard Extensions to Web Distributed Authoring and Versioning (WebDAV)", RFC 6352, DOI 10.17487/RFC6352, August 2011, <<https://www.rfc-editor.org/info/rfc6352>>.
- [RFC7162] Melnikov, A. and D. Cridland, "IMAP Extensions: Quick Flag Changes Resynchronization (CONDSTORE) and Quick Mailbox Resynchronization (QRESYNC)", RFC 7162, DOI 10.17487/RFC7162, May 2014, <<https://www.rfc-editor.org/info/rfc7162>>.
- [RFC7942] Sheffer, Y. and A. Farrel, "Improving Awareness of Running Code: The Implementation Status Section", BCP 205, RFC 7942, DOI 10.17487/RFC7942, July 2016, <<https://www.rfc-editor.org/info/rfc7942>>.

- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.
- [RFC822] Crocker, D., "STANDARD FOR THE FORMAT OF ARPA INTERNET TEXT MESSAGES", STD 11, RFC 822, DOI 10.17487/RFC822, August 1982, <<https://www.rfc-editor.org/info/rfc822>>.
- [RFC8610] Birkholz, H., Vigano, C., and C. Bormann, "Concise Data Definition Language (CDDL): A Notational Convention to Express Concise Binary Object Representation (CBOR) and JSON Data Structures", RFC 8610, DOI 10.17487/RFC8610, June 2019, <<https://www.rfc-editor.org/info/rfc8610>>.
- [RFC8984] Jenkins, N. and R. Stepanek, "JSCalendar: A JSON Representation of Calendar Data", RFC 8984, DOI 10.17487/RFC8984, July 2021, <<https://www.rfc-editor.org/info/rfc8984>>.
- [RFC9404] Gondwana, B., Ed., "JSON Meta Application Protocol (JMAP) Blob Management Extension", RFC 9404, DOI 10.17487/RFC9404, August 2023, <<https://www.rfc-editor.org/info/rfc9404>>.
- [RFC9553] Stepanek, R. and M. Loffredo, "JSContact: A JSON Representation of Contact Data", RFC 9553, DOI 10.17487/RFC9553, May 2024, <<https://www.rfc-editor.org/info/rfc9553>>.
- [RFC9610] Jenkins, N., Ed., "JSON Meta Application Protocol (JMAP) for Contacts", RFC 9610, DOI 10.17487/RFC9610, December 2024, <<https://www.rfc-editor.org/info/rfc9610>>.

Authors' Addresses

Lisa Dusseault (editor)
tbd
tbd
tbd , tbd
United States of America
Email: tbd
URI: tbd

Hans-Joerg Happel (editor)
audriga
Alter Schlachthof 57
76137 Karlsruhe
Germany
Email: hans-joerg@audriga.com
URI: <https://www.audriga.com>

Alexey Melnikov (editor)
Isode Ltd
14 Castle Mews
Hampton, Middlesex
TW12 2NP
United Kingdom
Email: Alexey.Melnikov@isode.com