Suggested SQL Schema based on Parts Registry DAS & FASTA output

The Parts Registry allows developers to access certain annotations via a DAS server, and FASTA files (more info here: http://partsregistry.org/Registry_API). These sources do not contain all the information available within the database (such as the name of the designer, group, accompanying comments, etc). The information that is available has been split into the following three tables: PARTS_INFO, seq_feat & feat_info.

The illustrative tables below have been populated with data extracted from BioBrick BBa_R0050.

TABLE: parts_info

Part_ID	Status	Part_no	Part_type	Description	Sequence	Size
BBa_R0050	A	188	Regulatory	Promoter (HK022 cI regulated)	TTAGGTATTGACTGTACTATCAGTTCCG TCATAATATGAACCATAAGTTCACCAC	188

Part_ID: unique identifier for each part in the registry.
Status: Single letter denoting the availability of the part (more info: http://partsregistry.org/Help:Availability_and_usefulness)
Part_no: Not sure what this represents.
Part_type: Listed here: http://partsregistry.org/Part_Types
Description: Short description of the part.
Sequence: DNA sequence.
Size: Length of DNA sequence.

Candidate Key: {Part_ID}. Also, possibly {PART NO}.
Primary Key: {Part_ID}

TABLE: seq_feat

Part_ID	Feat_ID	Start	End
BBa_R0050	2012	8	13
BBa_R0050	2013	13	27
BBa_R0050	2014	30	35
BBa_R0050	2015	37	51
BBa_R0050	2018	43	43
BBa_R0050	7066	1	54

Part_ID: unique identifier for each part in the registry.
Feat_ID: unique identifier for each feature of the part.
Start: Start site of the feature in the part's sequence.
End: End site of the feature in the part's sequence.

Candidate Key: {Part_ID, Feat_ID} – because you could potentially have the same feature occurring in more than one part. See notes at end.
Primary Key: {Part_ID, Feat_ID}
Foreign Key: {Part_ID}, reference: parts_info{Part_ID}

TABLE: feat_info

Feat_ID	Label	Feat_type	Orientation	Phase
2012	-35	promoter	Null	-
2013	OR2	RBS	Null	-
2014	-10	promoter	Null	-
2015	OR1	RBS	Null	-
2018	putative	Start	Null	-
7066	BBa_R0050	BioBrick	Null	-

Feat_ID: unique identifier for a feature. Have to confirm whether the same feature can be represented by more than one identifier.
Label: User-generated label for each sequence feature.
Feat_type: Standardised feature type. More info here: http://partsregistry.org/Help:Part_Features
Orientation: The orientation of the feature within the part. It does not always have a value.
Phase: Need to clarify the meaning of this.

Candidate Key: {Feat_ID}
Primary Key: {Feat_ID}
Foreign Key: {Feat_ID}, reference: seq_feat{Feat_ID}.

SQL

```
CREATE DATABASE toy_database;

create table parts_info (
  Part_ID varchar(20) not null,
  Status char(1) null,
  Part_no integer null,
  Part_type varchar(50) null,
  Description varchar(500) null,
  Sequence varchar(10000) null,
  Size integer null
  PRIMARY KEY (Part_ID),
);

create table seq_feat (
  Part_ID varchar(20) not null,
  Feat_ID integer not null,
  Start integer null,
  End integer null,
  PRIMARY KEY (Part_ID,Feat_ID),
  FOREIGN KEY (Part_ID) REFERENCES parts_info(Part_ID)
  );

create table feat_info (
  Feat_ID integer not null,
  Label varchar(50) null,
  Feat_type varchar(50) null,
  Orientation integer null,
  Phase char(1) null,
  PRIMARY KEY (Part_ID),
  FOREIGN KEY (Feat_ID) REFERENCES seq_feat(Feat_ID)
);
```

Notes

Tables seq_feat and feat_info could be merged if the same feature is represented by more than on Feat_ID. For example, part BBa_123 and BBa_345 could share the feature, BioBrick: BBa_678. However, this same feature may have two different Feat_IDs assigned according to the part context – or it could only have one Feat_ID which is then used to annotate both parts. More info is needed on how Feat_ID is generated.